Imperial College London

Department of Materials

MSE 101:
Mathematics and Computer Programming

Data Analysis

Prof. David Dye, Dr Sam Cooper

2016–17

Comments and corrections to david.dye@imperial.ac.uk or sjc08@ic.ac.uk

Lecture notes may be found on Blackboard (learn.imperial.ac.uk) or dyedavid.com

Contents

About the Course

This course is about data analysis. First we do some preparatory work on logarithms and curve sketching, and finding the roots of equations - which leads us into finding roots by iteration. Then, we look at approximating functions using power series, which is a very powerful method for handling equations that are otherwise difficult to work with.

Then we move onto errors, and how to mathematically treat errors when we are analysing data, in order to give good estimates of results, include the estimation of uncertainties. Very commonly, we will want to deal with data that is distributed about some central value, for which the classic distribution in the normal or Gaussian distribution.

That then leads us onto fitting data, first to a straight line - and very often, we will be able to manipulate our governing equation such that the data obeys a straight line fit. Then we move onto non-linear curve fitting, and finally onto a general method for non-linear least squares fitting of data to an arbitrary function, such as a Gaussian.

Course Support and Assessment

The course is primarily delivered through these notes, and the videos that can be found on YouTube via `dyedavid.com`. In addition, two tutorial sheets are provided of questions; these are divided into exercises, problems and example exam questions. The course will be assessed in the main June MSE101 exam. In the class sessions (lectures), we will use learning catalytics to support the learning process, by asking and then answering small conceptual problems. Prior to the lectures, students must read the notes and/or watch the videos; there is a 1:1 correspondence between the lecture and chapter numbers.

Further Resources

Much of the material on logarithms and series can be found in your favourite A-level Maths and Further Maths textbooks. The classic textbook is L Bostock and FS Chandler, *Core Maths Advanced Level*, 3rd Ed., Nelson Thornes, 2013.

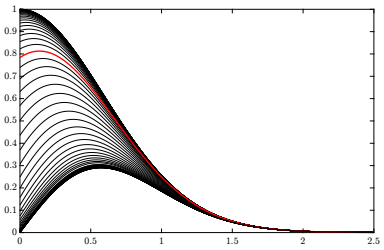
KL Stroud and DJ Booth, *Engineering Mathematics*, 7th Ed., Macmillan, 2013 (Imperial library 510.246STR). This is probably *the* core text for 1st year Maths; series and curve fitting are treated extensively. ML Boas, *Mathematical Methods in the Physical Sciences*, 3rd Ed., Wiley, 2006 (Imperial library: 530.15BOA). Mary Boas' book is a bit less wordy and goes into 2nd year material as well; again, data fitting isn't treated but it is good for series (Ch1) and the normal distribution (Ch15). It also looks at series convergence, which is an important topic we won't have time to cover in this course. Personally, I prefer Boas to Stroud.

GL Squires, *Practical Physics*, 4th Ed., Cambridge University Press, 2001. This short book is an excellent guide to experimental procedure and data analysis, particularly for lectures 5-6 on the treatment of errors and linear regression. It is in the Imperial library at 530.028SQU.

WH Press, SA Teukolsky, WT Vetterling and BP Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd Ed., Cambridge University Press, 2007. Numerical Recipes is the

grand-daddy of scientific computing, although it is rather indigestible. Data Fitting is treated in Chapter 15. It is in the Imperial library at 519.402NUM.

Simple linear regression and data fitting *can* be performed in Excel. However, you will want to move beyond Excel to use some real software that can handle data and perform nonlinear least squares fitting quickly and easily. *R* is a good way to go for the statistical analysis of data, and the *Numerical Recipes* routines can all be called from *Python/Scipy*. Another way to go is to use *Matlab*, which is a very powerful and flexible tool for all sorts of data analysis, programming and plotting - and is a very good front-end for programming in *Python* or *C++*. Matlab is avaialble on the cluster computers and student licenses can be downloaded via the Imperial ICT software shop.



Chapter 1

Functions

This chapter should serve as a reminder of some important concepts and skills that you will all have already encountered earlier in the course. The hope is that it will fill any gaps in your knowledge and give you greater confidence when manipulating functions.

1.1 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.

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

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 shows 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 refereed to as the *base* of the logarithm.

$$\log_b(a) = n$$

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.

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: $\log_2(243^{-0.2}) = \log_2(243^{-\frac{1}{5}}) = \log_2(\frac{1}{\sqrt[5]{243}}) = \log_2(\frac{1}{3})$

1.1.1 Change of base

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

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

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 $\log_3(x)/\log_3(9) = \log_3(x)/2 = \log_3(\sqrt{x})$

1.1.2 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 have been converted into a log scale. The following figure plots the same three functions in each of the three graphs.

Figure 1.1 shows the functions on linear axes with which you are familiar. Figure 1.2, plotted with a semi-log (base 10) y-axis, makes the first two functions into straight lines. Finally, in fig. 1.3, 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.

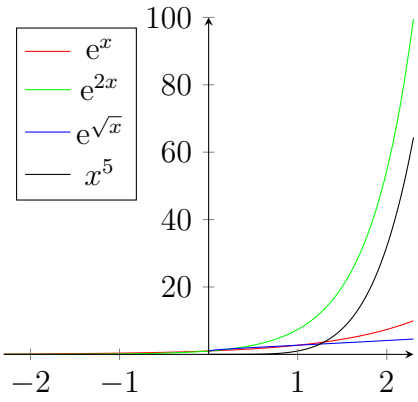


Figure 1.1: Linear Cartesian axes

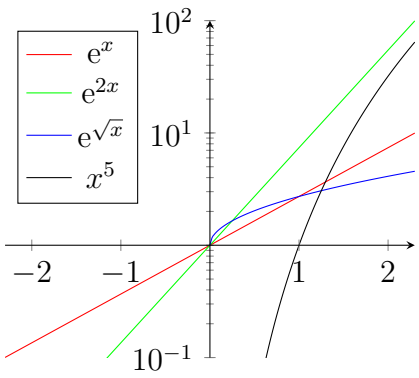


Figure 1.2: Semi-log axes in base 10

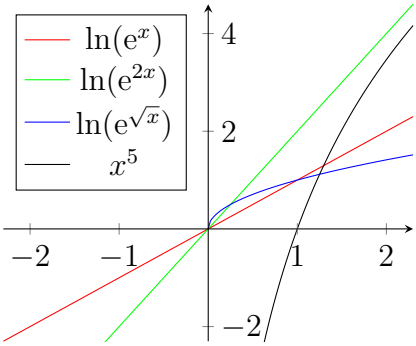


Figure 1.3: Linear axes plotting natural log of the functions

1.2 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 ($\frac{dy}{dx} = 0$)
5. Inflection Points ($\frac{d^2y}{dx^2} = 0$)
6. Domain and Range

1.2.1 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!

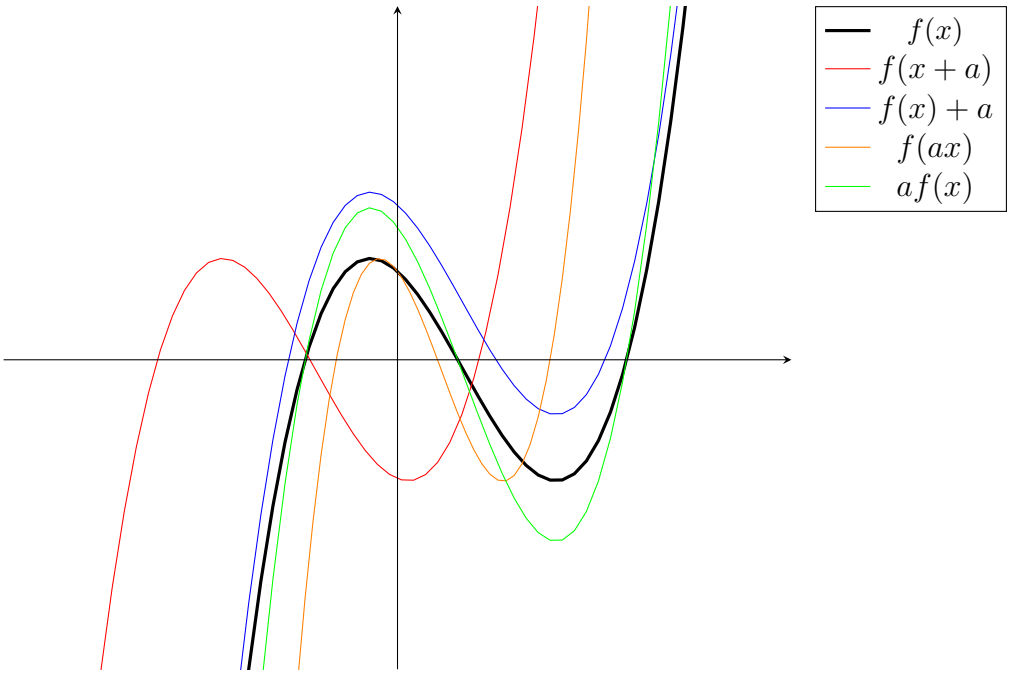


Figure 1.4: Transforms of a cubic function, using a constant factor $a > 0$.

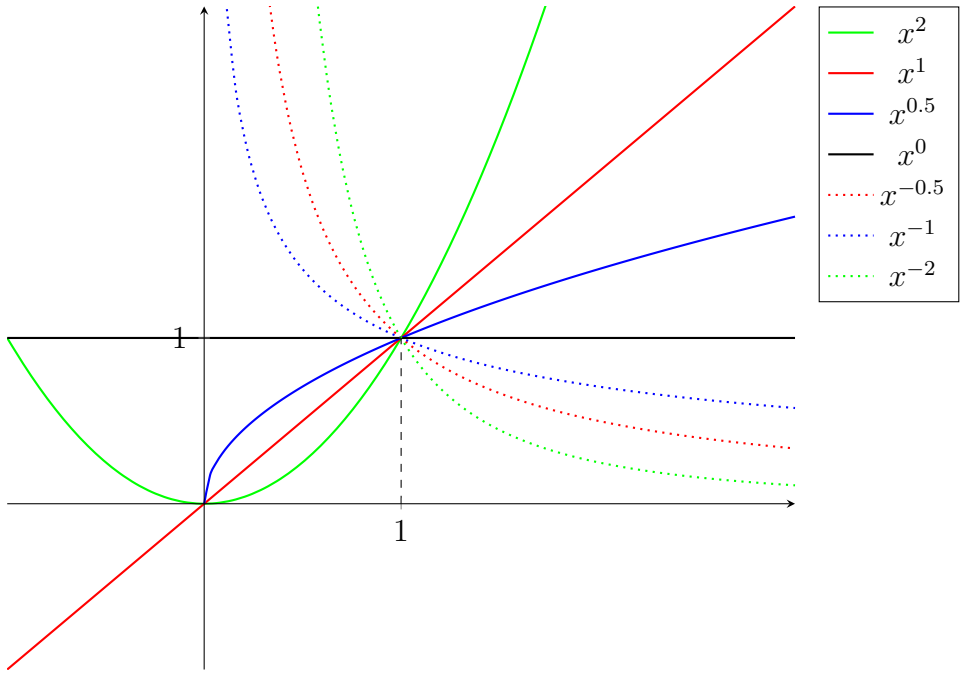


Figure 1.5: Shows the effect of varying the index

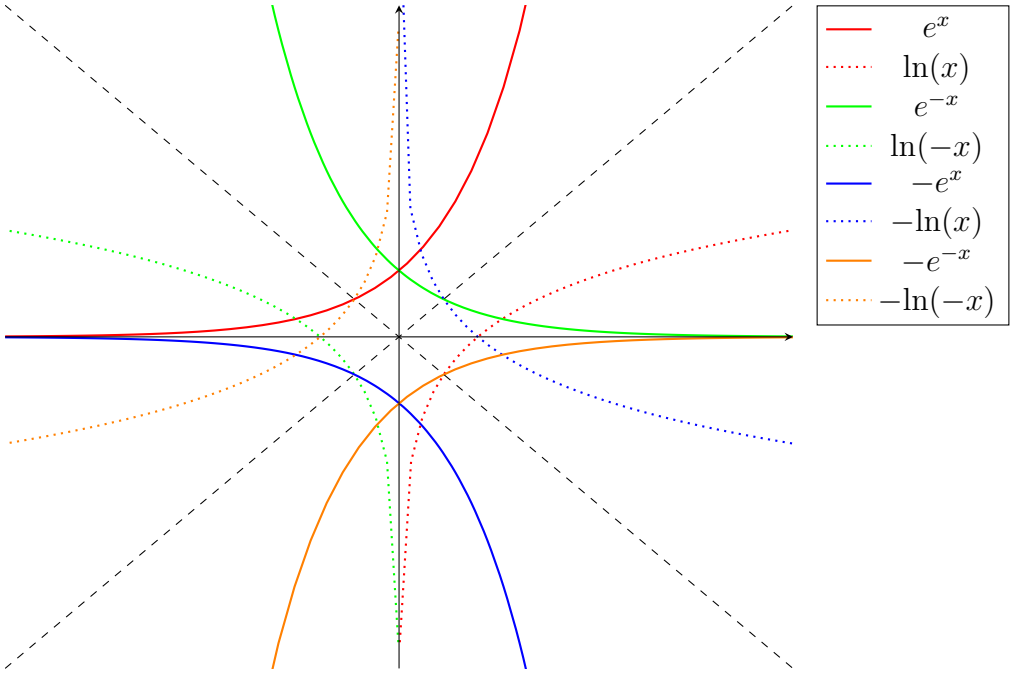


Figure 1.6: Natural functions, illustrating that inverse functions can be constructed with simple reflections across the line $y = x$

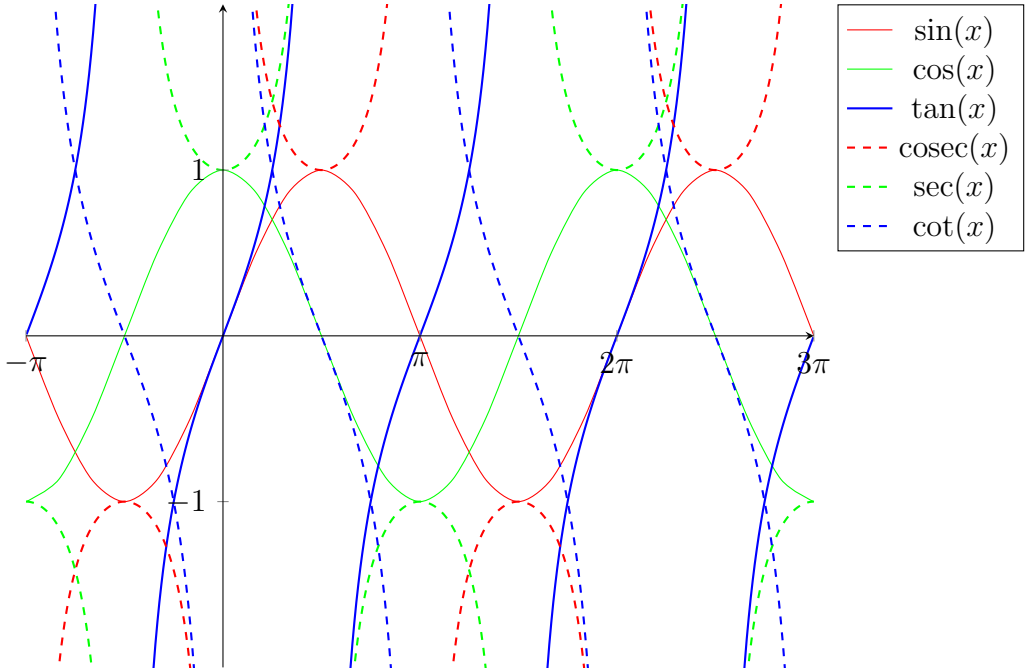


Figure 1.7: Trigonometric functions

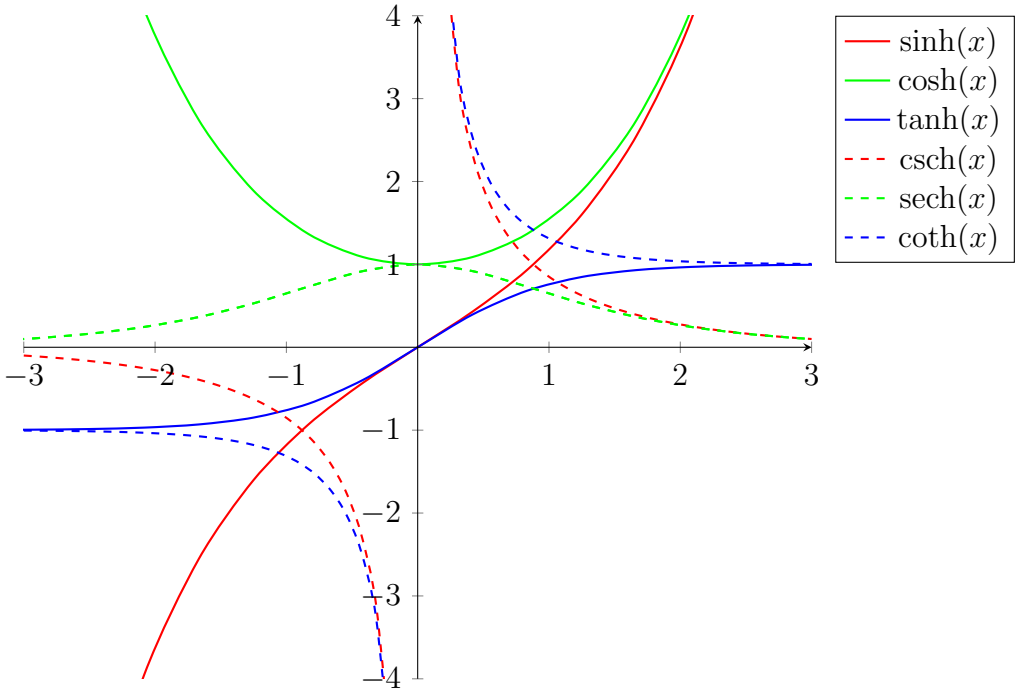


Figure 1.8: Hyperbolic functions

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

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 $x = 1$.

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".

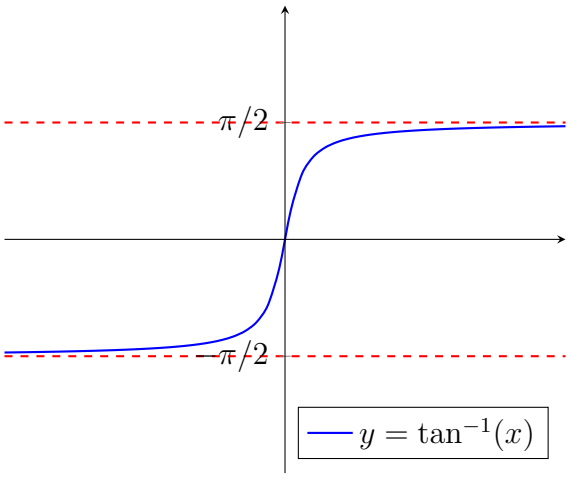
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

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 the following chapter.

1.2.3 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.



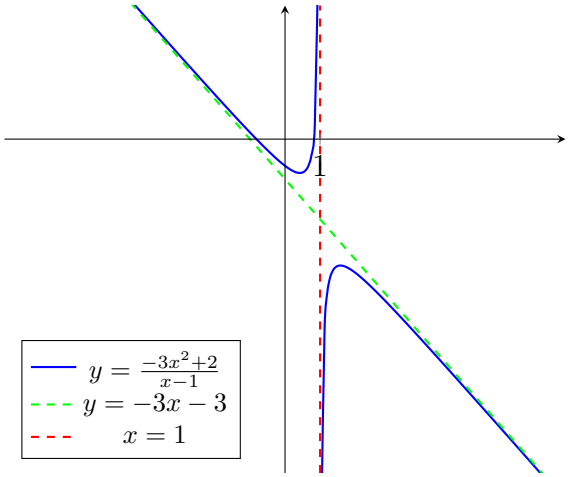
Two horizontal asymptotes

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.

$$
\begin{array}{r}
-3x-3 \\
x-1\,\overline{)\,-3x^2 \quad +2} \\
\underline{\;3x^2-3x\;\;} \\
-3x+2 \\
\underline{\;\;3x-3\;} \\
-1
\end{array}
$$

Which tells us to expect an asymptote on the line $y = -3x - 3$.



Vertical and slant asymptotes

1.2.4 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*.

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.

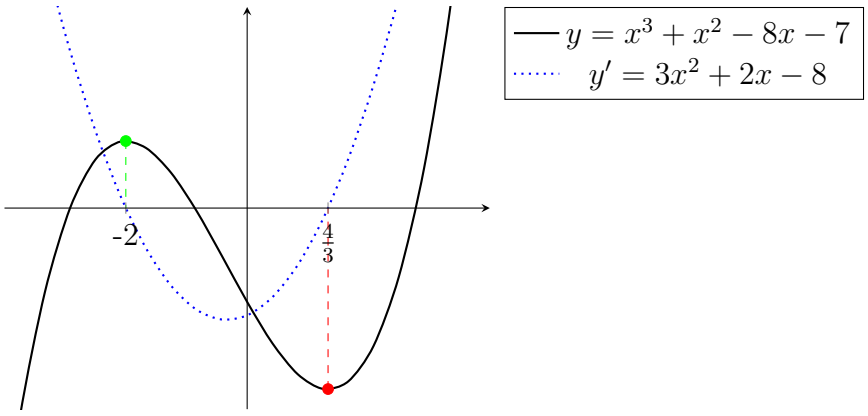


Figure 1.11: Plot showing a polynomial, its derivative and the stationary points.

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!).

1.2.5 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 are not local maxima or local minima. They can be located by finding where the second derivative of a function equals zero.

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 $x = -1/3$.

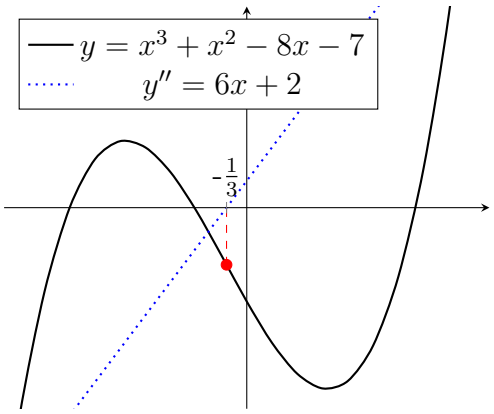


Figure 1.12: Plot showing a polynomial, its second derivative and the inflection point.

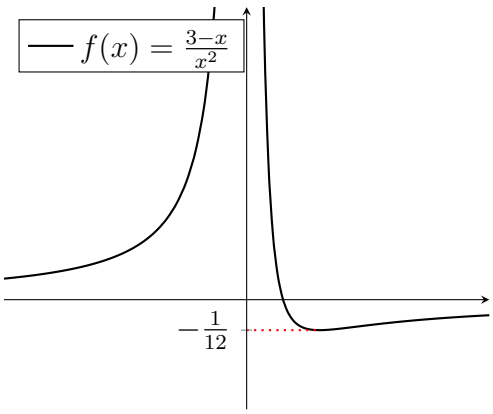
1.2.6 Domain and Range

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

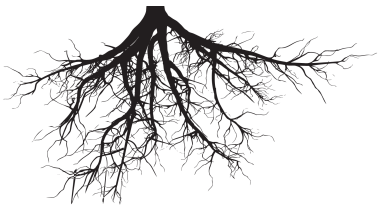
The *range* is the set of all y coordinates that have a corresponding x coordinate.

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 "$\cup$".

Example - For the function $y = \frac{3-x}{x^2}$, as shown in fig. 1.13, 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 $[\frac{-1}{12}, \infty)$



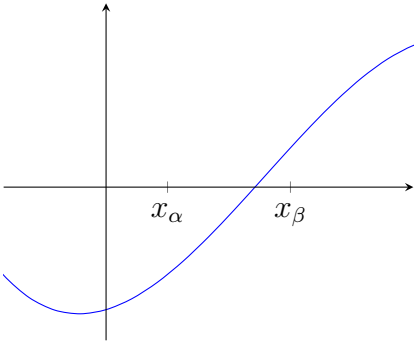
Chapter 2

Root Finding

As discussed in the previous chapter, 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 formular). However, in many case there is no direct approach to finding 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.

2.1 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 with x_n so that the new interval contains a zero crossing.

Example - The function $f(x) = -3x^3 + 7x^2 + 2x - 4$ has three distinct roots. We would like to find an approximation to 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. 2.1).

$$\begin{aligned} f(x_\alpha) &= -3x_\alpha^3 + 7x_\alpha^2 + 2x_\alpha - 4 \\ &= -3(0.5)^3 + 7(0.5)^2 + 2(0.5) - 4 \\ &= -1.625 \quad \text{(negative)} \end{aligned}$$

$$\begin{aligned} f(x_\beta) &= -3x_\beta^3 + 7x_\beta^2 + 2x_\beta - 4 \\ &= -3(1.5)^3 + 7(1.5)^2 + 2(1.5) - 4 \\ &= 4.625 \quad \text{(positive)} \end{aligned}$$

The first bisection point occurs at
$x_1 = \frac{1}{2}(x_\alpha + x_\beta) = \frac{1}{2}(0.5 + 1.5) = 1.$

$$\begin{aligned} f(x_1) &= -3x_1^3 + 7x_1^2 + 2x_1 - 4 \\ &= -3(1)^3 + 7(1)^2 + 2(1) - 4 \\ &= 2 \quad \text{(positive)} \end{aligned}$$

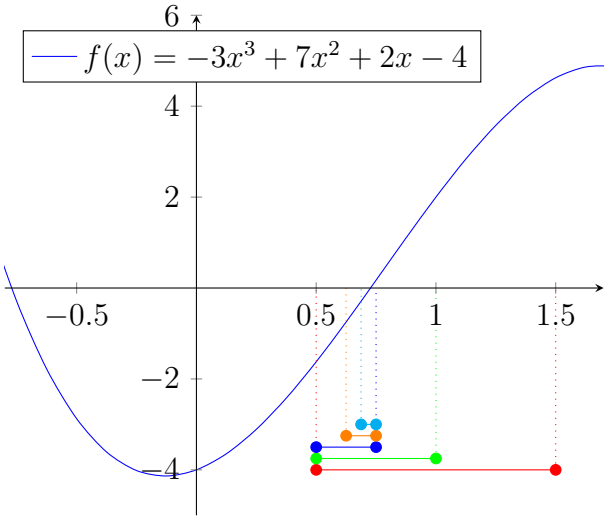


Figure 2.1: Shows successively smaller intervals resulting from each iteration of the bisection method.

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, but this may be acceptably close for our application. We can also say with confidence that the true root is in our final range of $[x_4, x_2] = [0.6875, 0.75]$.

2.1.1 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 2.2 shows an interval containing a discontinuity. If the bisection method was pursued it would locate the discontinuity as if it were a root.

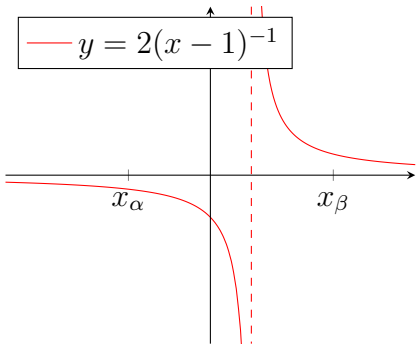


Figure 2.2: Function with discontinuity in the initial interval.

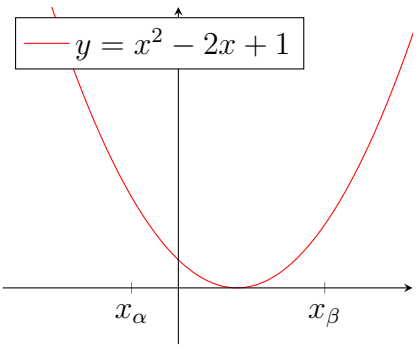


Figure 2.3: Function with coincident roots.

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. 2.3, 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 2.4 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)_/¯

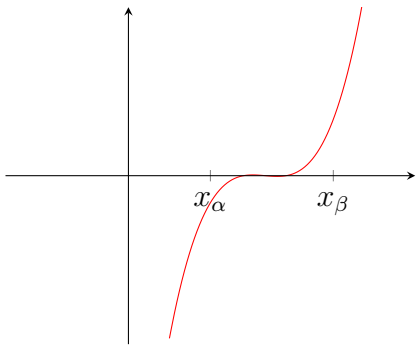


Figure 2.4: Cubic function with multiple close roots.

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 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 is 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.

2.2 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, NR 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,

$$f(x_0) = f'(x_0)x_0 + c, \tag{2.1}$$

which can be rearranged to find c

$$c = f(x_0) - f'(x_0)x_0. \tag{2.2}$$

We can now write the following expression for the tangent line at x_0

$$y = f'(x_0)x + f(x_0) - f'(x_0)x_0, \tag{2.3}$$

Finally, by setting $y = 0$ and rearranging to make x he subject, we get

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}, \tag{2.4}$$

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.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{2.5}$$

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.

2.2.1 NR Example

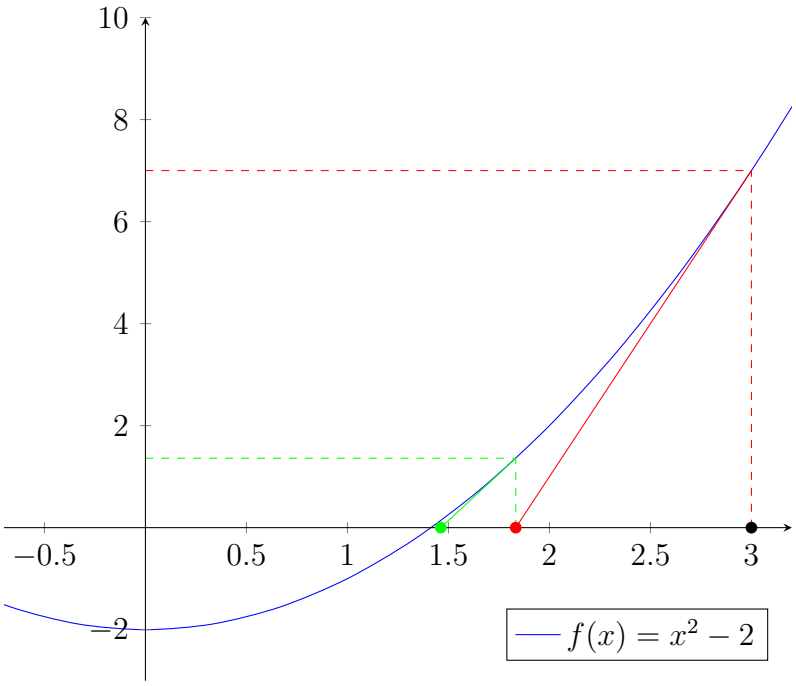
Find an approximation for the square root of 2 by using the NR method to find the root of the equation $f(x) = x^2 - 2$.

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,

$$x_2 = 0.5\left(\frac{11}{6}\right) + \frac{6}{11} = \frac{193}{132} = 1.46\dot{2}\dot{1} \quad ,$$

which gets us to the green point, $x_2 = 1.46\dot{2}\dot{1}$. 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.

Chapter 3

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.

3.1 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 fig. 3.1, that this guess function is not great...
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 $f'(0)$.

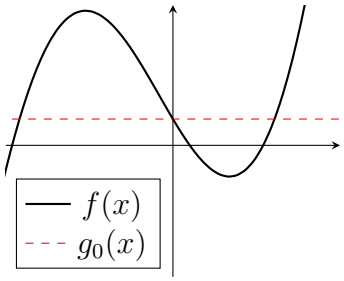


Figure 3.1: 0^{th} order.

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 fig. 3.2 and is clearly a significant improvement.

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

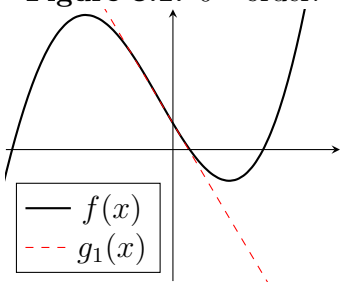


Figure 3.2: 1^{st} order.

Repeating this process, we will now also use the second derivative of our function to help improve our guess function (fig. 3.3). 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 fig. 3.4. Again, to understand where the $(3 \times 2)^{-1}$ term comes from, try differentiating $g_3(x)$ three times. 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.

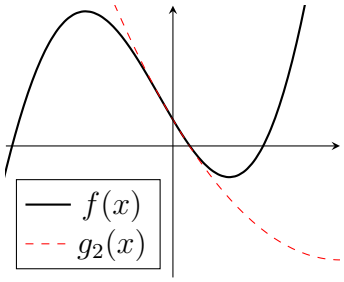


Figure 3.3: 2^{nd} order.

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

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 $0! = 1$).

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

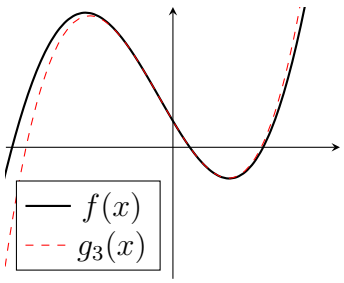


Figure 3.4: 3^{rd} order.

3.1.1 Maclaurin Examples

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

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

Therefore the Maclaurin series expansion is

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

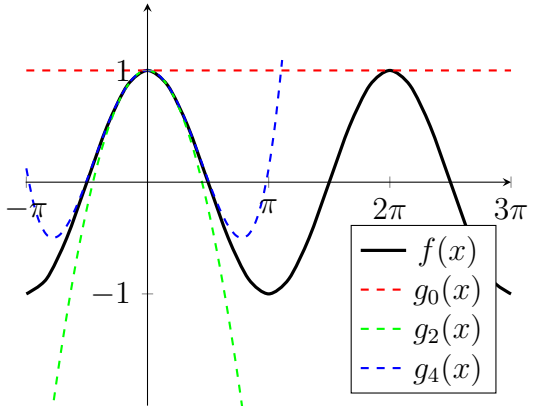


Figure 3.5: Maclaurin expansion of $\cos(x)$.

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:

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

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

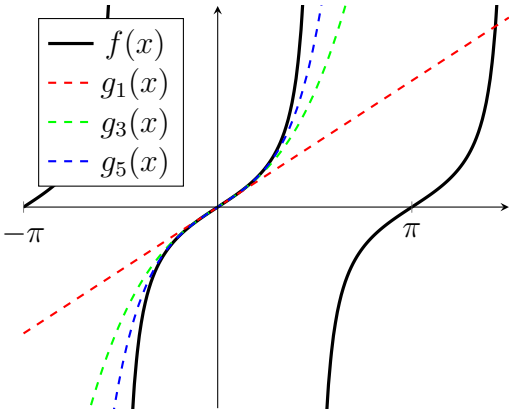


Figure 3.6: Maclaurin expansion of $\tan(x)$.

Looking at fig. 3.6 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.

3.2 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.

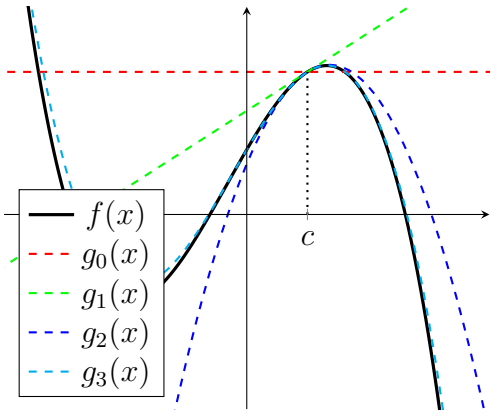


Figure 3.7: Taylor series expansions of an arbitrary function about the point c

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

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

$$\begin{aligned} f(1) &= \mathrm{e}^1 = \mathrm{e} \\ f'(1) &= \mathrm{e}^1 = \mathrm{e} \\ f''(1) &= \mathrm{e}^1 = \mathrm{e} \\ f^{(3)}(1) &= \mathrm{e}^1 = \mathrm{e} \end{aligned}$$

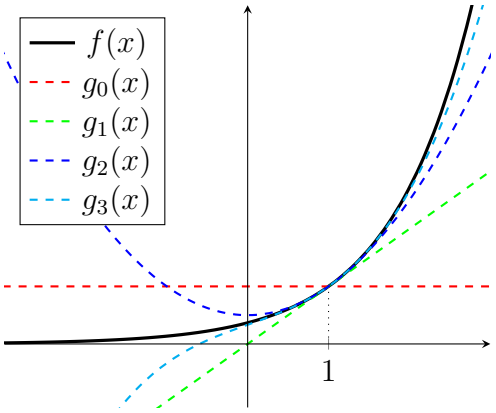


Figure 3.8: Taylor series expansions of the function $f(x) = \mathrm{e}^x$ at the point $x = 1$.

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), where we have just expanded around $x = 1$.

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

It is important to select a sensible point from which to expand you 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 $f(x) = \mathrm{e}^x$, is that seemingly moving our approximation from $x = 0$ to $x = 1$ caused us to multiple every term by e. But when you consider that $\mathrm{e}^{x+1} = \mathrm{e}\mathrm{e}^x$ it should suddenly make a lot of sense!

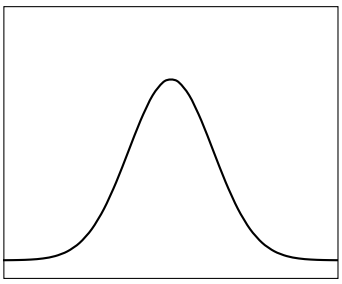
Chapter 4

The Gaussian

Gaussians, named after the mathematician Carl Friedrich Gauss, are functions based on the form

$$f(x) = a\mathrm{e}^{-b(x-c)^2} \quad ,$$

where e is Euler's number and a, b and c are arbitrary positive constants.
These functions have a characteristic "bell curve" shape and are asymptotic to the x-axis in both directions; however, they do also have a finite area. We do not know how to solve the integral of this function between the limits $x = -\infty$ and $x = +\infty$ directly, but can find the answer by making use of a few clever substitutions (tricks!).



4.1 The Gaussian Integral

The following method (alternative approaches exist) uses the following three tricks to find the integral:

1. **Squared**
2. **Polar**
3. **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 without too much difficulty!
We wish to find the area, A, under the the standard Gaussian (*i.e.* where a and b are both 1).

$$A = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\, dx$$

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!).

$$A = \int_{-\infty}^{\infty} \mathrm{e}^{-y^2}\, dy$$

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.

$$\begin{aligned} A^2 &= \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} \mathrm{e}^{-y^2} \mathrm{e}^{-x^2} \, dy \, dx \\ &= \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} \mathrm{e}^{-(x^2+y^2)} \, dy \, dx \end{aligned}$$

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 fig. 4.1

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \end{aligned}$$

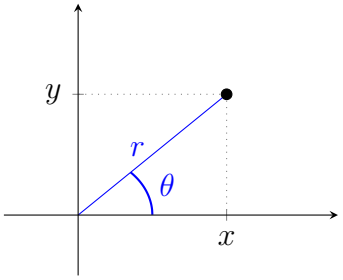


Figure 4.1: Cartesian and polar coordinates.

To convert between coordinate systems, we calculate the determinant of the Jacobian matrix. This is because integration in $(x - y)$ space does not map directly on to $(r - \theta)$ space (*i.e.* $dx\,dy \neq dr\,d\theta$).

The Jacobian matrix is constructed by finding each four possible partial derivative combinations, as in the following table.

	dr	$d\theta$
dx	$\cos(\theta)$	$-r\sin(\theta)$
dy	$\sin(\theta)$	$r\cos(\theta)$

Once we have found the determinant of this matrix, we can then complete the convert our equation.

$$\begin{aligned} |J| &= \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} \\ &= r\cos^2(\theta) + r\sin^2(\theta) \\ &= r(\cos^2(\theta) + \sin^2(\theta)) \\ &= r \end{aligned}$$

Therefore,

$$dx\,dy = r\,dr\,d\theta$$

We can now transform our equation for A^2, which we'll do in three steps. First by using our Jacobian,

$$A^2 = \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} \mathrm{e}^{-(x^2+y^2)} r\,dr\,d\theta \quad .$$

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 $\theta = 2\pi$ (refer to fig. 4.1 to see why).

$$A^2 = \int_{0}^{r=\infty} \int_{0}^{\theta=2\pi} \mathrm{e}^{-(x^2+y^2)} r\,dr\,d\theta$$

Finally, we also know from Pythagoras that $x^2 + y^2 = r^2$, which leads to

$$A^2 = \int_{0}^{r=\infty} \int_{0}^{\theta=2\pi} \mathrm{e}^{-r^2} r\,dr\,d\theta$$

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

$$A^2 = 2\pi \int_{0}^{r=\infty} \mathrm{e}^{-r^2} r\,dr$$

The last stage is to make the **substitution** $s = -r^2$, which differentiates to

$$ds = -2r\,dr$$

Take care to substitute the limits correctly!

$$\begin{aligned} A^2 &= -\pi \int_{0}^{s=-\infty} \mathrm{e}^{s}\,ds \\ &= -\pi \left[\mathrm{e}^{-\infty} - \mathrm{e}^{0}\right] \\ &= -\pi \left[0 - 1\right] \\ &= \pi \end{aligned}$$

Having completed this integration, we simply take the square root of our answer to find the area, A, under our standard Gaussian... easy as π.

$$A = \sqrt{\pi}$$

4.2 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 $y = \mathrm{e}^{-x^2}$ 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 simple divide by its current area.

$$f(x) = \frac{1}{\sqrt{\pi}}\mathrm{e}^{-x^2} \tag{4.1}$$

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. 4.1, 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 and the function itself both by a factor of $\sqrt{2}$, which give what is usually referred to as the *standard normal distribution*

$$f(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$$

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 fig. 4.2.

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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).

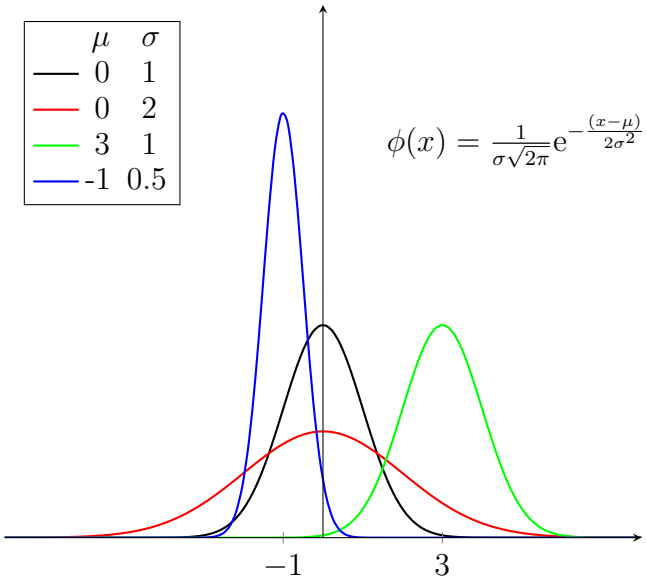


Figure 4.2: Normal distributions.

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.
Later in this course you will be taught how to "fit" a normal distribution to some data, which in this case would mean varying μ and σ until the the curve corresponded closely with the data. 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 evaluated the probability of random 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.

$$P(x_a < X < x_b) = \int_{x_a}^{x_b} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

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.

4.2.1 The Error function

The Gauss error function, written $\mathrm{erf}(x)$, 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.

$$\begin{aligned}\mathrm{erf}(x) &= \frac{1}{\sqrt{\pi}}\int_{-x}^{x}\mathrm{e}^{-t^2}\,dt \\ &= \frac{2}{\sqrt{\pi}}\int_{0}^{x}\mathrm{e}^{-t^2}\,dt\end{aligned}$$

Also, notice that the variable t is what we call a "dummy" variable, as it does not appear outside the expression and is there only to allow us to manipulate x in a certain way.

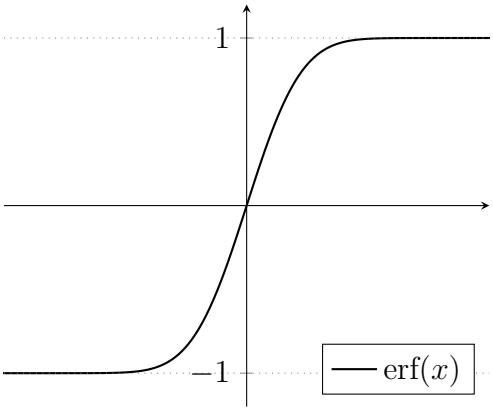


Figure 4.3: Plot of the error function .

We still do not know how to evaluate $\mathrm{erf}(x)$ directly (in fact, it can't be done!), but we do know how to differentiate it... Returning to what we learned in our chapter on power series, we can use a truncated MacLaurin series to approximate the function, which is how your computer would do it (or using an alternative numerical method)! If you've forgotten how to generate a MacLaurin series, then go back and refresh your memory now. Otherwise, the first five derivatives are:

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

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

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

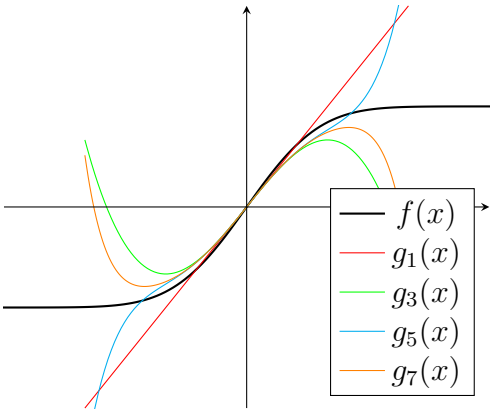


Figure 4.4: Plot of the error function as well as the Maclaurin series expansions up to 7^{th} order. *N.B.* The "true" curve (black) was of course itself also calculated with a (high order) power series.

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. 4.5).

$$P(X < x_b) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right)\right]$$

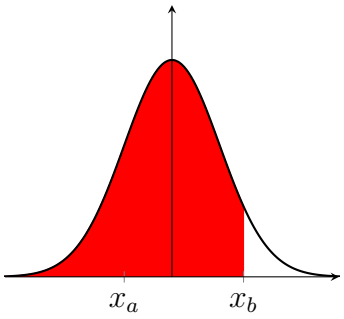


Figure 4.5: Area under a Gaussian up to x_b.

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.

$$\begin{aligned} P(X > x_a) &= 1 - P(X < x_a) \\ &= 1 - \frac{1}{2}\left[1 + \text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right)\right] \\ &= \frac{1}{2}\left[1 - \text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

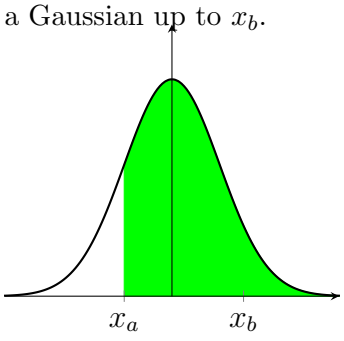


Figure 4.6: Area under a Gaussian beyond x_b.

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. 4.7.

$$\begin{aligned} P(x_a < X < x_b) &= P(X < x_b) - P(X < x_a) \\ &= \frac{1}{2}\left[\text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

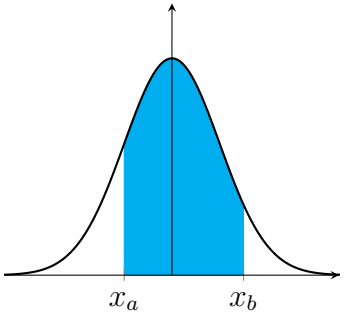


Figure 4.7: Area under a Gaussian from x_a to x_b.

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.

$$\begin{aligned} P(X < x_a) + P(X > x_b) &= 1 - P(x_a < X < x_a) \\ &= 1 - \frac{1}{2}\left[\text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

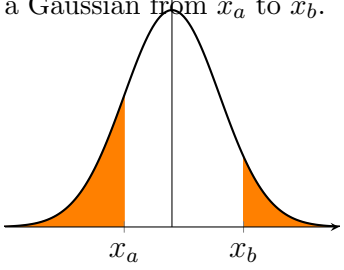


Figure 4.8: Area under a Gaussian outside x_a to x_b.

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?

$$\begin{aligned} P(1.25 < X < 1.45) &= P(X < 1.45) - P(X < 1.25) \\ &= \frac{1}{2}\left[\operatorname{erf}\left(\frac{1.45 - 1.7}{0.2\sqrt{2}}\right) - \operatorname{erf}\left(\frac{1.25 - 1.7}{0.2\sqrt{2}}\right)\right] \\ &= \frac{1}{2}\left[(-0.7887...) - (-0.9756...)\right] \\ &= 0.093... \end{aligned}$$

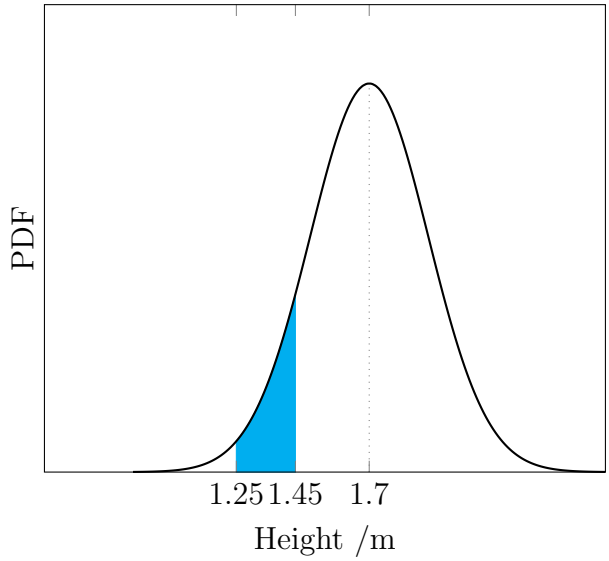


Figure 4.9: PDF of the heights of students in a class.

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.

Chapter 5

Error analysis

In science, we often need to decide whether our different sets of data are *significantly* different. That is, its not just the number that is important but the uncertainty or expected error in that number.

At the most basic level, when you write down a number you need to know the uncertainty in order to know what level of precision is appropriate. If you measure the carbon content of an alloy as 1.035 wt.%, but the uncertainty is 0.6 wt.%, then we should say that the composition is 1.0 ± 0.6 wt.%.

Once an experimental result gets published, people will use it for all sorts of purposes; some you will have thought of, and some you won't. Whatever use they make of it, they will want to know if it is sufficiently precise for their purpose.

In addition, if your experiment is evaluating an hypothesis or theory, you need to know if your result is sufficiently precise to confirm or deny the hypothesis. It might be tempting to suggest that every experiment should be as precise as is possible, but that wouldn't be financially or economically wise - our experiment only needs to be good enough to be fit for purpose.

If we have several ($i = 1...n$) repeated measurements x we will often take some kind of numerical average. The three most commonly used are the *mean* ($\Sigma_i x_i/n$), the middle value or *median* of the numerically sorted distribution, and the most frequently occurring or *mode*. If we make the *ansatz* (initial assumption or guess) that the data are centrally distributed, then the median, mode and mean will be the same. For many types of data this won't be true (think earnings), but nevertheless, most often we will using the arithmetic mean $\overline{x}$, which we will define as

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{5.1}$$

The standard deviation of a population is given by $\sigma_x^2 = \sum_i((x_i - \overline{x})^2)/n$; *i.e.* the variance, or σ^2 is the average of the squares of the differences from the mean. However, if the data we have are only a sample of the data in a population, then we should divide by $n - 1$ rather than n.

$$\sigma_x^2 = \sum_i \frac{(x_i - \overline{x})^2}{n-1} \tag{5.2}$$

If the data are normally distributed (according to a Gaussian or normal distribution), then 68% of the data will fall within ± 1 standard deviation (s.d.) of the mean. If instead, we want the probability of a data point lying within $\pm 2\sigma$, this is 95% and 99.7% for $\pm 3\sigma$.

This is useful; if we want a quick estimate of the standard deviation, then we can simply take the range of the central 2/3rds of the data - so if we have 6 measurements, we can take the range of the central four data points.

A further point is that the standard deviation of the sample (the uncertainty in a single measurement) is not the same as the uncertainty in the mean of a data population. This is given by

$$\sigma_{\overline{x}} = \sum_i \frac{(x_i - \overline{x})^2}{n(n-1)} = \frac{\sigma_x}{\sqrt{n}} \tag{5.3}$$

We can prove this by considering taking the average of a serise of populations; a proof is given on pp.12-15 of *Squires*.

5.1 Propagation of errors

Very often, we will want to know how errors propagate. For example, say we want to estimate the uncertainty in some measured quantity, like a yield stress. This is given by a force divided by an area, so if we can estimate the uncertainty in the force, and the uncertainty in the area of our test sample, we should be able estimate the uncertainty in our measurement of a stress. This will be lower than the variance we would get from making multiple samples, but will be an acceptable lower bound for the purposes of evaluating whether the difference between two alloys or material conditions might be significant.
Before we do that, if we consider a function w of x and y, as well as the variance σ_x and σ_y, and mean $\overline{x}$ and $\overline{y}$, we also need to consider their covariance

$$\sigma_{xy} = \frac{1}{n-1} \sum_i (x_i - \overline{x})(y_i - \overline{y}) \tag{5.4}$$

Consider the function $w = x + y$. Trivially therefore, $\overline{w} = \overline{x} + \overline{y}$. We can show this as follows

$$\overline{w} = \frac{1}{n} \sum_i w_i = \frac{1}{n} \sum_i x_i + y_i = \sum_i \frac{x_i}{n} + \sum_i \frac{y_i}{n} = \overline{x} + \overline{y} \tag{5.5}$$

N.B. since we are estimating the population variance, we use n rather than $n-1$. Similarly, we can figure our how the variances propagate;

$$\sigma_w^2 = \frac{1}{n} \sum_i (w_i - \overline{w})^2 = \frac{1}{n} \sum_i \left(x_i + y_i - (\overline{x} + \overline{y})\right)^2 \tag{5.6}$$

$$= \frac{1}{n} \sum_i \left((x_i^2 - 2x_i\overline{x} + \overline{x}^2) + (y_i^2 - 2y_i\overline{y} + \overline{y}^2) + 2(x_iy_i - x_i\overline{y} - \overline{x}y_i + \overline{xy})\right) \tag{5.7}$$

$$= \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} \tag{5.8}$$

So if the covariance of x and y is small or zero, that is, they are linearly independent variables, then simply $\sigma_w^2 \simeq \sigma_x^2 + \sigma_y^2$. That is, when adding variables together, we can find the uncertainty in their sum by simply adding together the uncertainty in each of the variables, in quadrature.
Referring to Equation 5.2, we can also find that if $w = Kx$, where K is some constant, then $\sigma_w^2 = K^2\sigma_x^2$, or $\sigma_w = K\sigma_x$.
If we continue in this vein, we eventually find that the general equation for the uncertainty in a function $f(x_i)$, for $i = 1...n$ is

$$\sigma_f^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 \tag{5.9}$$

This gives us some interesting results, as follows;

function f	error
$f = x + y$ $f = x - y$	$\sigma_f^2 = \sigma_x^2 + \sigma_y^2$
$f = xy$ $f = x/y$	$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$
$f = \ln x$	$\sigma_f = \frac{\sigma_x}{x}$
$f = e^x$	$\frac{\sigma_f}{f} = \sigma_x$

That is, for addition and subtraction the squares of the absolute errors add, whilst for multiplication and division the squares of the fractional errors add.

So, for example, if we are interested in a *change* in a measurement when we introduce some experimental variation, we must be careful because although the change might be small the errors very often will swamp our measurement, making the deduction of any effect rather difficult.

The other practical implication is that if the uncertainty in one of the variables (say y) is small, then pracitcally speaking the uncertainty in the result will be dominated by the other (x).

We also need to think about how our errors arise. For example, the uncertainty in measuring a distance with a ruler by eye is probably worse than 0.25 mm. If we want to measure an agle by using a tape measure placed on the circumference of a circle, this means that the fractional error will be minimised by making the circle as large as possible - by using a very big protractor, in effect!

Example

In diffraction strain measurement, we use the spacing of the lattice planes in a crystal as an interatomic strain gauge to measure the elastic strain. So we use our definition of strain ε as the fractional change in length $\varepsilon = \partial d/d$ and apply that to Bragg's Law $\lambda = 2d\sin\theta$, where d is the plane spacing, θ is half the diffraction angle and λ is the wavelength of the radiation used. First we must find ∂d by total partial differentiation

$$\partial\lambda = 2\partial d\sin\theta + 2d\partial\theta\ \cos\theta \tag{5.10}$$

But, in our experiment we will use the same radiation each time, so $\partial\lambda = 0$. Therefore we can say that

$$\partial d/d = -\partial\theta\cot\theta \tag{5.11}$$

And if we pick the radiation such that $2\theta = 90^\circ = \pi/2$, then $\cot\theta = 1$ and we obtain the convenient result that $\varepsilon = -\partial\theta$ - where we need to remember to use radians! Therefore if the strain-free lattice spacing is found at a diffraction angle of θ_0, we can say that the strain is given by

$$\varepsilon = \theta_0 - \theta \tag{5.12}$$

where θ is the diffraction angle for the strained specimen. The uncertainty in the strain is then simply given by

$$\sigma_\varepsilon = \sqrt{\sigma_\theta^2 + \sigma_{\theta_0}^2} \tag{5.13}$$

This emphasizes that significant experimental effort should be expended on finding a good measurement of the strain-free lattice spacing.

5.2 Some Notes on Experimental Practice

Back at the dawn of modern science, Francis Bacon (1521-1626) developed some guideline for scientific experimentation, which still stand. They are our major defences against scientific fraud and malpractice. Firstly, there should be a single, clear running record kept of what is done in the lab. When recording data, the actual numbers read should be recorded *immediately* and *directly*, without any manipulation, so that any errors in calculation can be corrected later. Everything should be dates and time stamped.

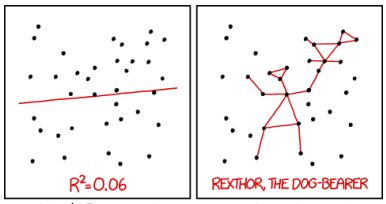
Notes should not be taken loose in the lab and then copied into 'neat' notebook later. (i) This is a waste of human effort, (ii) it introduces mistakes and (iii) it is impossible to avoid being selective. At which point, selectivity will be driven by attemtping to get the 'right' results - those that conform to the hypothesis at hand.

It may be right and proper to discard experimental artefacts, where there is a known reason why those data items are flawed. But the original data should be retained and recorded, in case our ideas about what was happening change after reflection and further thought.

Finally, we need to distringuish systematic and random errors. A systematic error occurs identically to every measurement in a set, wereas a random error is one that varies and may be positive or negative. For example, many rulers have an offset at either end - the scale does not begin at the end of the rule. If this were overlooked, this would result in a systematic error. On the other hand, the error in rading from the scale of $\pm 0.25\,\mathrm{mm}$ will be a random error. One benefit of setting up experiments where a difference is measured (subtraction) is that many systematic errors of this type will subtract away and can therefore be disregarded.

As a guide, uncertainties should not normally be quoted to more than one significant figure (1sf), and measurements should not be reported to greater precision than the uncertainty in the measurement - such additional precision is a waste of ink.

In general, the student is direction to GL Squires, *Practical Physics*, 4th Ed, Cambridge Universtiy Press, 2001. Chapters 6 and 11-13 will be a rewarding evening's reading, almost to any expeirmentalist at any stage in their career. There are many copies in the Imperial library.



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Chapter 6

Linear Regression

Very often, we will have hypothesised that some phyiscal phenomenon will behave according to some algebraic function, the constants of which tell us about the underlying physics such as the frequency of atomic jumps (in diffusion), for example. We will then perform an experiment aimed at confirming the hypothesis, and, along the way, measuring those physical constants. Therefore we will often manipulate the equation so as to enable us to plot the data graphically such that the underlying relationships become obvious to the eye.

We can then fit a line or surface to the data that corresponds to the hypothesized phyiscal law, with the aim of coming up with a good measurement of the underlying constants, including an uncertainty estimate.
We will start out with the simplest case, that of a straight line where all the data are equally (un)reliable. Consider the function

$$y = mx + c \tag{6.1}$$

where the varied quantity is x (placed on the horizontal axis), the measured quantity is y, and the phyiscal constants are m and c.
For many experiments we will be able to reformulate the governing equation such that a straight line is obtained; e.g is $y = \exp\{Q/RT\}$, we could take natural logs and find $\ln y = (Q/R)(1/T)$, and plot a straight line intersecting the origin of $\ln y$ vs. $1/T$, with gradient Q/R.
The standard method for doing this is a least sqaures minimisation method, and it works as follows. We first define the residual or deviation for each reading (x_i, y_i);

$$s_i = y_i - mx_i - c \tag{6.2}$$

Thus, s_i is the deviation of the measured value of y_i from the expected value $mx_i + c$. This approach assumes that the error is entirely in the measurement of y and that none arises from x. We then say (after Legendre, 1806) that the best value of m and c will be those for which the sum of the squares of the deviations s_i are a minimum; so we find

$$S = \sum_i (y_i - mx_i - c)^2 \tag{6.3}$$

We then differentiate and set to zero to find m and c

$$\frac{\partial S}{\partial m} = -2\sum x_i(y_i - mx_i - c) = 0 \tag{6.4}$$

$$\frac{\partial S}{\partial c} = -2\sum (y_i - mx_i - c) = 0 \tag{6.5}$$

Taking the equation for c, 6.5, we find that

$$m\sum x_i + cn = \sum y_i \tag{6.6}$$

$$c = \overline{y} - m\overline{x} \tag{6.7}$$

since the average of x is given by $\overline{x} = \sum x_i/n$; the same holds for $\overline{y}$. We can then insert Equation 6.5 into Equation 6.4 to find

$$m\sum x^2 + c\sum x = \sum xy \tag{6.8}$$

$$m\sum x^2 + (\overline{y} - m\overline{x})\sum x = \sum xy \tag{6.9}$$

we then collect the m terms to obtain

$$m(\sum x^2 - \overline{x}^2 n) = \sum xy - \overline{x}\,\overline{y}n \tag{6.10}$$

again, since $\sum x/n = \overline{x}$. We therefore find m;

$$m = \frac{\sum xy - \overline{x}\,\overline{y}n}{\sum x^2 - \overline{x}^2 n} = \frac{\sum (x-\overline{x})y}{\sum (x-\overline{x})^2} \tag{6.11}$$

Hoo-Rah! As they say in the US Marines. Or something like that. But wait, we're not done yet - we can estimate the eror is m and c Proving this is rather complicated but the result is as follows. First we define D as the sum of the squares of the x deviations from the mean;

$$D = \sum (x - \overline{x})^2 \tag{6.12}$$

Then it is found that

$$\sigma_m^2 \simeq \frac{S}{D(n-2)} \tag{6.13}$$

$$\sigma_c \simeq \sigma_m \sqrt{\frac{D}{n} + \overline{x}^2} \tag{6.14}$$

There is one final technical point; really, the best fit line should be written as

$$y = (m \pm \sigma_m)(x - \overline{x}) + (b \pm \sigma_b) \tag{6.15}$$

That is, the constant b should be thought of as the vertical offset at the centre of mass (average) of the measurement points x. This is instead of writing $y = (m \pm \sigma_m)(x) + (c \pm \sigma_c)$. This is because c and m are not independent quantities - c is related to m. Instead, Equation 6.15 provides a straight line that pivots around the centre of mass of the measured data, and then b is independent of m. b is given by

$$b = \overline{y} \tag{6.16}$$

$$\sigma_b^2 \simeq \frac{S}{n(n-2)} \tag{6.17}$$

While m is unaffected.

6.1 Example - linear fitting

Say we perform a tensile test where, at each strain x (dimensionless) we find a stress y (in GPa). We obtain the following data ($n = 20$);

	x (−)	y (GPa)	$x-\overline{x}$ $\times 10^{-3}$	$y(x-\overline{x})$ (MPa)	$(x-\overline{x})^2$ $\times 10^{-6}$	s (GPa)	# s.d. (-)
	0.001	0.077	-4.75	-0.37	22.56	-0.1692	1.42
	0.0015	0.293	-4.25	-1.25	18.06	-0.0606	0.51
	0.002	0.468	-3.75	-1.76	14.06	0.0071	0.06
	0.0025	0.599	-3.25	-1.95	10.56	0.0306	0.26
	0.003	0.675	-2.75	-1.86	7.56	-0.0010	0.01
	0.0035	0.912	-2.25	-2.05	5.06	0.1293	1.09
	0.004	0.778	-1.75	-1.36	3.06	-0.1123	0.94
	0.0045	1.111	-1.25	-1.39	1.56	0.1129	0.95
	0.005	1.160	-0.75	-0.87	0.56	0.0546	0.46
	0.0055	1.439	-0.25	-0.36	0.06	0.2270	1.91
	0.006	1.256	0.25	0.31	0.06	-0.0636	0.53
	0.0065	1.453	0.75	1.09	0.56	0.0255	0.21
	0.007	1.598	1.25	2.00	1.56	0.0638	0.54
	0.0075	1.403	1.75	2.46	3.06	-0.2386	2.00
	0.008	1.680	2.25	3.78	5.06	-0.0692	0.58
	0.0085	1.950	2.75	5.36	7.56	0.0935	0.78
	0.009	2.078	3.25	6.75	10.56	0.1142	0.96
	0.0095	2.186	3.75	8.20	14.06	0.1154	0.97
	0.01	2.064	4.25	8.77	18.06	-0.1140	0.96
	0.0105	2.140	4.75	10.17	22.56	-0.1454	1.22
Sum	0.115	25.32		35.68	166.3		
SumSq			$D = 0.166$			$S = 0.270$	
Avg.	0.0058	1.27					0.00
Stdev						0.1191	

We obtain

$$m = 1000\frac{35.68}{166.3} = 215 \tag{6.18}$$

$$c = 1.27 - 215 \times 0.0058 = 0.03 \tag{6.19}$$

$$b = 1.27 \tag{6.20}$$

and uncertainties

$$\sigma_m = \sqrt{\frac{0.270}{0.000166 \times 18}} = 9 \tag{6.21}$$

$$\sigma_c = 215\sqrt{215\frac{0.000166}{18} + 0.0058^2} = 0.06 \tag{6.22}$$

$$\sigma_b = \sqrt{0.270/(20 \times 18)} = 0.03 \tag{6.23}$$

So the straight line fit can be stated as

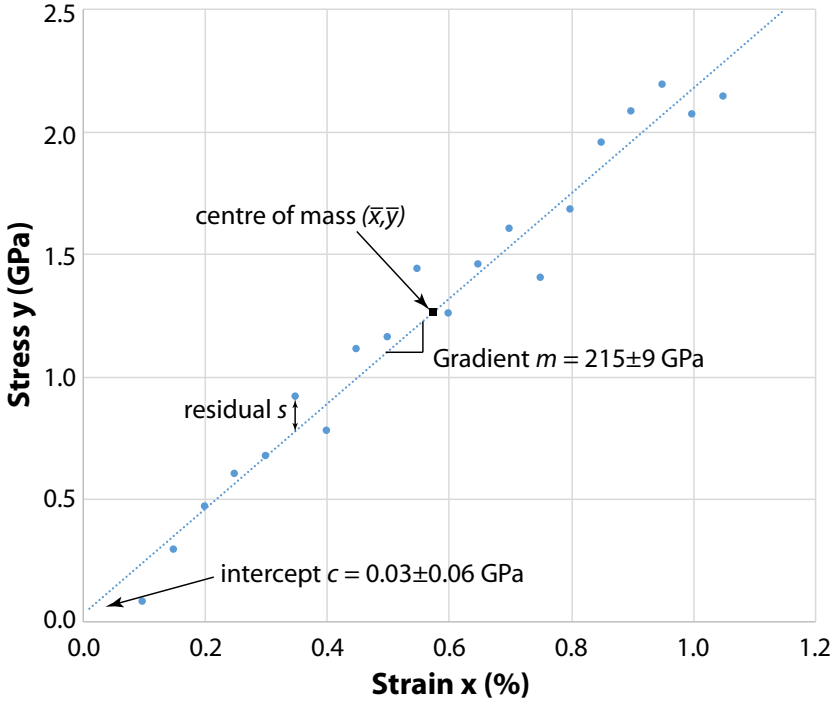
$$y = (215 \pm 9)x + (0.03 \pm 0.06) \tag{6.24}$$

or, alternatively and probably preferably

$$y = (215 \pm 9)(x - 0.0058) + (1.27 \pm 0.03) \tag{6.25}$$

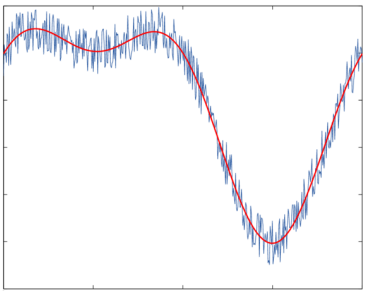
In the table, the deviations that each data point are from the line s are calculated. The root mean square of this $S/(n-1)$, or standard deviation of s) is calculated also. This allows the

number of standard deviations that each point is vertically in y from the line to be obtained in the last column. Pleasingly, these average zero, and around 1/3rd of the data points are greater than 1 s.d. from the line[1].
The data can be plotted as follows



Of course, there are many enhancements that could be made to this procedure; we should take account of what is known about the uncertainties in x and y, for example. In addition, we could or should take care to remove obviously spurious data from the fit, because these can move the fit quite badly - when this is done automatically then this is known as *robust fitting*.

[1] which they should be, since I generated this data using a normal distribution and a random number generator.



Chapter 7

Nonlinear Curve Fitting

7.1 Linearising equations

Diffusion can be taken, at least phenomenologically, to obey an equation like

$$D = D_0 \exp\{-Q/RT\} \tag{7.1}$$

Where D is the diffusivity of a species, D_0 is the intrinsic diffusivity which is reduced by the exponential factor as temperature decreases, by a factor related to the temperature T. Here Q is an activation energy, in $\mathrm{kJ\,mol^{-1}}$, and R is the Gas Constant, $8.314\,\mathrm{kJ\,K^{-1}mol^{-1}}$.

The nickel superalloys used in jet engines often use molybdenum (Mo) as a solid solution strengthener of the matrix *fcc* Ni phase, where Mo also acts to reduce the average of effective bulk diffusivity and hence slow down the rate of dislocation climb, which is a diffusion-activated process. This then acts to reduce the rate of creep in these alloys, which is a desired alloy design behaviour.
If we examine VD Divya et al, *Interdiffusion in the Ni–Mo system*, Scripta Materialia, 62(8):621-624, 2010, we find the following data for the variation in diffusivity of Mo in Ni with temperature;

T, °C	1050	1100	1150	1200	1225
D, $10^{-15}\,\mathrm{m^2\,s^{-1}}$	2.35	3.90	7.80	16.8	21.7

If we refer to our governing equation and take logs, we find

$$\log D = \log D_0 + \frac{-Q}{R \ln 10}\frac{1}{T} \tag{7.2}$$

At this point is is worth noting that $R \ln 10 = 19.14$.
If we re-cast our data into Kelvin and then find $1/T$, and take logs of the D data, we obtain

$1/T$,$10^{-4}\,\mathrm{K^{-1}}$	7.5630	7.2846	7.0282	6.7936	6.6754
$\log_{10} D$	-14.63	-14.41	-14.11	-13.77	-13.66

We can then plot this and find the straight line, to obtain $Q = 220 \pm 10\,\mathrm{kJ\,mol^{-1}}$.
There are several things to note when making a log plot. First, you will often see people plot data where the log scale is shown explicitly; but the actual distances still correspond to the $\log(y)$ values - this is just a plotting trick to make the data easier to read by eye.

Another important feature here is: how should we plot error bars in a log plot? Generally, where at all possible, you should always estimate the uncertainty in each point and plot error bars. In a log plot, the error bars will, in principle, not be symmetric any more. This reflects an

underlying problem, which is that the uncertainties will no longer be normally distributed when logs are taken (or any other function, like $1/x$, for that matter). But, as we said in Chapter 4, the equation for the propagation of errors tells us that the uncertainty in $\ln x$ is σ_x/x, and so this is the result normally used when logs are plotted. Which of these is 'better' is actually still a matter of debate - as pointed out in R Sundberg, Chemometrics and intelligent laboratory systems, 41:249-252, 1998.

7.2 Fitting to a Generalised Function

In principle, we would like to generalise our least squares fitting approach to handle any function $f(x)$, not just straight lines. For example, we could have a power series

$$f(x, a_j) = a_0 + a_1 x + a_2 x^2 + ... + a_m x^m = \sum_{j=0}^{m} a_j x^j \tag{7.3}$$

Or, the functions could be sines and cosines (called a Fourier series), or any other function, like a Gaussian - for example, to describe a diffraction peak;

$$f(x) = \frac{A}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} + b \tag{7.4}$$

This function has four fitting parameters, A which is the area of the peak, σ which is related to its width, μ which is the peak centre on the x-axis, and b which is the background intensity far away from the peak.
To fit this, we define our least-squares error estimator, Chi-squared

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{n} \left[\frac{y_i - f(x_i, a_j)}{\sigma_i}\right]^2 \tag{7.5}$$

where $\mathbf{a}$ is the vector (array, set) of the fitting parameters a_j. Each value in the data series is given by (x_i, y_i, σ_i), that is, every value of y is assumed to have an accompanying uncertainty σ. In our definition of χ^2, we penalise (down-weight) data points with large uncertainties, by dividing the residual by σ. If all the points have the same value of the uncertainty, this will drop out of the analysis naturally (*i.e.* they can all be assumed to be 1, or whatever).

Then, we pick our best parameters a_j as being those which result in the lowest sum of the squares of the weighted residuals, termed χ^2.

We can then perform our normal trick of minimising χ^2 by finding the point in parameter space where the derivative of χ^2 with respect to all the a_j vanishes.

In the case where the model function $f(x, a_j)$ depends linearly on the a_j, that is, they are something like a power series or some sum of functions each of just a_j, then this minimisation can still be performed exactly. However, with functions like our Gaussian that depends on several a_j together the minimisation must be performed iteratively.

We first differentiate our expression to χ^2 to find j equations for the the gradients with respect to the a_j, which will be zero at the minimum;

$$\frac{\partial \chi^2}{\partial a_j} = -2 \sum_{i=1}^{n} \frac{y_i - f(x_i, \mathbf{a})}{\sigma_i^2} \frac{\partial f(x_i, \mathbf{a})}{\partial a_j} \qquad j = 1...n \tag{7.6}$$

We can then write these down conveniently in a vector β_j, and remove the irritating factor of -2 by defining

$$\beta_j = -\frac{1}{2}\frac{\partial \chi^2}{\partial a_j} \tag{7.7}$$

In each iteration, we want to adjust a_j until we find the minimum. Therefore the simplest approach is to move by a vector ∂a_j in the $\mathbf{a}$ parameter space given by,

$$\partial a_j = \text{constant} \times \beta_j \tag{7.8}$$

This is called the *steepest descent* method; it adjusts parameters a_j the most if they have a big effect (gradient) on χ^2. The a_j are then adjusted iteratively until some convergence criterion for the change in χ^2 between successive iterations is met.
More sophisticated approaches make use of the second derivative of the gradient. They do this by computing the matrix of second derivative whose elements (j, k) are

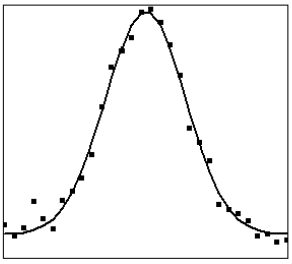
$$\frac{\partial^2 \chi^2}{\partial a_j \partial a_k} \tag{7.9}$$

This matrix is called the Hessian matrix; the Jacobian you met before was the matrix of first derivatives. We then solve a set of simultaneous equations using the Hessian (by inverting it), to find an estimate for the set of a_k which would solve for $\chi^2 = 0$.

The assumption being made in this approach is that the function $f(x, \mathbf{a})$ cna be approximated by a quadratic (first three terms of the Taylor series), and therefore from the second derivatives the root can be found.

The two most popular versions are the Gauss-Newton and Marquardt methods, which are widely implemented in software packages, *e.g.* in Matlab. Further details (for the brave) can be found in *Numerical Recipes*, specifically section 15.5.

At this stage, you most likely lack the confidence with matrix maths to be able to write your own Hessian non-linear least squares minimiser, but in the next Chapter we will have a go at making a steepest descent minimiser for a Gaussian in Excel, before showing how to perform such fitting quickly and easily in Matlab.



Chapter 8

Nonlinear Least Squares Fitting II: Fitting a Gaussian

We now consider how to go about fitting a Gaussian peak to some diffraction data. In X-ray or neutron diffraction, the peak has a breadth and intensity related to the instrumental parameters, sample and crystal structure.

You have already met in MSE104 the idea that the intensity is proportional to the structure factor, the sum of the scattering from the different atoms in the unit cell, which means that for some structures, some peaks can be systematically absent. The grains in the sample will also have a distribution of orientations that may not be random, due to *e.g.* plasticity during processing operations such as casting, forging and rolling. There may also be a variation in the lattice parameter of the crystals due to the straining of the lattice near the grain boundaries, around precipitates, due to defects such as vacancies and dislocations and potentially even due to variations in composition. This will give rise to a breadth to the diffraction peak.

In addition, there will be some variation in the wavelength of the radiation used, due to the physical processes related to its generation. For example, the wavelength of the X-rays generated by a Cu filament will vary depending on its temperature. In addition, the incident beam will often have a spread, called a divergence, related to the extent of collimation used to define the beam. These will also result in a variation in peak width.

There will also often be a background level of radiation present in the measured data, due to noise in the detector physics, and due to stray radiation within the experimental apparatus.

There are many peak shapes that can be used to fit such a diffraction peak, but the simplest is a Gaussian with a constant background. Its equation is given by

$$f(x, \sigma, x_p, I, b) = b + \frac{I}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x - x_p)^2}{2\sigma^2}\right\} \tag{8.1}$$

Here, x is the position in the detector and x_p is the position of the centre of the peak. I is the integrated area of the peak, σ is the peak width - the deviation from x_p at which the intensity reduces by a factor of $e^{-0.5} = 0.606$ from its maximum, and b is the background intensity. Comparing to the last chapter, $\mathbf{a} = \{\sigma, x_p, I, b\}$, where $\mathbf{a}$ is the vector of fitting parameters.

To find Equation 7.6, we need to find the derivative of the Gaussian with respect to each of the

fitting parameters. These are

$$\frac{\partial f}{\partial b} = 1 \tag{8.2}$$

$$\frac{\partial f}{\partial I} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-x_p)^2}{2\sigma^2}\right\} \tag{8.3}$$

$$\frac{\partial f}{\partial \sigma} = \frac{I}{\sqrt{2\pi}}\left(\frac{(x-x_p)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) \exp\left\{\frac{-(x-x_p)^2}{2\sigma^2}\right\} \tag{8.4}$$

$$\frac{\partial f}{\partial x_p} = \frac{I}{\sqrt{2\pi}}\frac{(x-x_p)}{\sigma^3} \exp\left\{\frac{-(x-x_p)^2}{2\sigma^2}\right\} \tag{8.5}$$

We can then form Equation 7.6, find our β_j and obtain our next guess of our four fitting parameters. In practice it is often useful when the fitting parameters are very different in magnitude to adjust using individual tuning constants. These won't change our ability to find the parameters that minimise χ^2, but they will change how quickly the algorithm gets there. These are demonstrated with some diffraction data in the accompanying Excel file. The procedure assumes that you can provide an initial guess of the parameters to be fitted. If this guess is too far away, *i.e.* the model and data peaks do not overlap, then the algorithm will be unable to sensibly calculate whether moving the peak left or right improves the fit, and so the algorithm will fail. Secondly, the procedure is rather inefficient and relies quite strongly on the tuning parameters that are chosen - if they are too big, the fit will be unstable and will bounce around and diverge, whilst if they are too small, then it won't converge in a reasonable number of iterations.

The fitting procedure can be improved using the curvature of the function $\chi^2(\mathbf{a})$ to find the minimum, but this requires us to form an invert the matrix of the second partial derivatives - the Hessian. This is rather much at this stage in your evolution as mathematicians - we are already, in effect, finding the vector of first derivatives to find the steepest way to descend the χ^2 function, which would require us to know about grad or ∇, which are introduced at the end of the MSE101 course.

Jon von Neumann famously remarked that *"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."* This is rather a big problem in experimental science - very often we will have a physical model in mind that might have anywhere up to 20 physical parameters, some of which will be impossible to measure independently and must be fitted to data. And some of those we can measure, like an activation energy for diffusion, will have a large enough uncertainty that they are effectively free fitting parameters. Therefore when making physically based models, we should take care to minimise the number of fitting parameters and to maximise the extent to which they can be independently measured with minimal error.

Finally, many very good general nonlinear least squares minimisers exist and are accessible in standard packages like Matlab, Igor, R and Python, and therefore there is no need to code your own. We will demonstrate fitting the same diffraction data in Matlab in the videos. Often, these will also come with error estimators that evaluate, from the second derivative of χ^2, the likely statistical uncertainty in the fit parameters, subject to assumptions that these are normally distributed, and so on. It should be remarked that this is probably a lower bound estimate of the uncertainty, not taking into account sample-to-sample variations.

As a last caution, given that iterative fitting is necessarily a somewhat brittle process, the fits should also always be evaluated visually so as to avoid the generation of "garbage data".