Ministry of Higher Education
Higher Future Institute of Engineering and Technology in Mansoura
Department of Mathematics and Basic Sciences

معهد المستقبل العالي للهندسة و التكنولوجيا بالمنصورة
Higher Future Institute of Engineering and Technology at Mansoura

---

Mathematics (I)

Dr. Htem Abd Elaziz Elagamy

Mathematics and statistics

Ministry of Higher Education
Higher Future Institute of Engineering and Technology in Mansoura
Department of Mathematics and Basic Sciences

معهد المستقبل العالي للهندسة و التكنولوجيا بالمنصورة
Higher Future Institute of Engineering and Technology at Mansoura

---

Mathematics plays an important role in our practical and scientific life and its applications that have changed many interpretations of natural phenomena. In this course, we will study mathematics 1 and its engineering applications, and its importance lies in linking it to working life.

Dr. Htem Abd Elaziz Elagamy

Mathematics and statistics

Contents

List of Figures

Chapter 1

Revision

Most of you will already be familiar with this material. It is collected here to remind you of some of the things we rely on before starting our discussions.

Remember also that *your next Engineering Mathematics Course will also have a list of prerequisites*. It will include the list here, but will also include the material we are about to cover. There is a real advantage in mastering this material during this course rather than have it hanging over you later.

1.1 Powers

For any real number x and any positive integer n, x^n is taken to mean the product $x.x.x.\dots.x$ (n terms).

This is just a matter of notation, introduced for convenience. Observation and counting now give us the following rules, known as the **laws of indices**.

$$x^m x^n = x^{m+n}$$
$$(x^m)^n = x^{mn}$$

These are important. You must know and be able to use them.

The notation x^n is so compact and convenient that we extend it to cover situations when n is not an integer. This is done in such a way as to ensure that the laws of indices continue to hold.

1.1. Definition. x^0 is defined to be 1 for all x.

Note that this definition is forced on us if we are to keep the first law.

1.2. Definition. For any non-zero real number x and any positive integer n, x^{-n} is defined to be $\dfrac{1}{x^n}$.

So, for example, x^{-2} is the same as $\dfrac{1}{x^2}$. Again, this definition is necessary if you are to retain the first law.

1.3. Definition. For any positive real number x and any positive integer n, $x^{\frac{1}{n}}$ is defined to be the n-th root of x. More precisely, it is that positive real number whose n-th power is x.

$x^{-\frac{1}{n}}$ is just the inverse of $x^{\frac{1}{n}}$.

So, for example, $x^{\frac{1}{2}} = \sqrt{x}$ and $x^{-\frac{1}{2}} = \dfrac{1}{\sqrt{x}}$ Note that $\sqrt{x}$ is always the *positive* square root. If you want the negative square root you write $-\sqrt{x}$. Once more, the definition is forced on us by the second law. The restriction to positive x is to avoid problems with things like square roots of negative numbers.

For other fractional powers you just assemble in the obvious way, e.g. $x^{\frac{2}{3}}$ is the square of $x^{\frac{1}{3}}$ or the cube root of x^2. In general for positive real x

$$x^{\frac{m}{n}} = (x^m)^{\frac{1}{n}}$$

or

$$x^{\frac{m}{n}} = \left(x^{\frac{1}{n}}\right)^m$$

These are equal - but sometimes one is easier to compute than the other, e.g. $9^{\frac{3}{2}} = \left(9^{\frac{1}{2}}\right)^3 = 3^3 = 27$ is easier than $9^{\frac{3}{2}} = \left(9^3\right)^{\frac{1}{2}} = \sqrt{729}$ $(= 27)$.

Here are some examples of working with powers to practice with.

1.4. Example. Simplify the following expressions: $x^{2/3}x^{-3/2}$, $(x^{-1}\sqrt{y})^{3/2}$.

1.5. Example. Expand out the following brackets and simplify the results: $x^{1/3}(x^{2/3}+y^{1/3})$, $(x^{1/2} + x^{-1/2})(x^{1/2} - x^{-1/2})$.

1.2 Algebraic Manipulation

Manipulating indices is just one of the skills we shall rely on you having. The remainder of this Chapter reminds you of more background we assume you have. In addition, you are expected to be able to *manipulate* expressions — write them in equivalent ways which may be simpler, are more convenient.

Here are some examples of basic algebraic manipulation to practice with.

1.6. Example. Multiply out the following brackets and tidy up your results.

$$(x+1)(x-2), \qquad (2+x)^2(1-x), \qquad (1-x)(1-y)(1-z)$$

1.7. Example. Simplify the following expressions by taking each of them over a common denominator and then tidying up.

$$\frac{1}{x+1} - \frac{1}{x-1}, \qquad \frac{1+x}{1-x} - \frac{1-x}{1+x}$$

1.3 Summations

Notation can be a help or a hindrance, depending on how familiar you are with it. Here we discuss the **summation notation**.

The Greek capital letter Σ is used to indicate a summation. The terms that come after the Σ describe the form of the terms to be added together, and the decoration top and bottom of the Σ specifies a range.

1.8. Example. Write out $\sum_{k=1}^{6} k^2$ explicitly.

Solution The sum is shorthand for $1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2$ — the sum from $k = 1$ to $k = 6$ of all terms of the form k^2.

1.9. Example. Write out $\sum_{r=0}^{7} (2r + 1)^3$ explicitly.

Solution The given sum is the sum from $r = 0$ to $r = 7$ of all terms of the form $(2r + 1)^3$ and so is $1^3 + 3^3 + 5^3 + 7^3 + 9^3 + 11^3 + 13^3 + 15^3$.

The way in which the range is described can vary.

1.10. Example. Write out $\sum_{1 \leq k \leq 6} k^2$ explicitly.

Solution The sum is read as " the sum for all k between 1 and 6 of k^2". So it is just another way of writing the sum from the Example 1.8.

In more advanced work you can get "double sums". Here there will be two control parameters rather than one to describe the pattern, and you will be given a summation range for each.

1.11. Example. Write out $\sum_{r=1}^{3} \sum_{s=1}^{4} rs^2$ explicitly.

Solution Here we have the sum of all terms of the form rs^2 for the given ranges of r and s. To expand it, first fix r at its lowest value and let s run over the range described; then move r to the next value and repeat; and so on. With this example the expansion is

$$(1.1^2 + 1.2^2 + 1.3^2 + 1.4^2) + (2.1^2 + 2.2^2 + 2.3^2 + 2.4^2) + (3.1^2 + 3.2^2 + 3.3^2 + 3.4^2).$$

With double sums the range descriptions can get quite fancy. You just have to read it to yourself and think about what is being said.

1.4 The Binomial Theorem

The **Binomial Theorem** is an important theorem, useful across a wide range of mathematics. It gives you a fast method of expanding powers of sums. If for example you are faced with an expression such as $(x + y)^6$, you can write down the expansion without going through the long process of multiplying out.

1.12. Example. $(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$.

The theorem enables you to calculate all the terms of this expansion in your head. To describe it in detail we need more notation. For any positive integer n, **factorial** n, written as $n!$ denotes the product of all the integers from 1 up to n.

1.13. Example. $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, and so on.

It is convenient also to have a meaning for $0!$; we define $0!$ is to be 1. There is no real mystery here. It just turns out to be a good idea; with this convention formulae can be written in a more uniform way with fewer special cases.

1.14. Theorem (The Binomial Theorem). *For any positive integer n*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

where $\binom{n}{k}$ is shorthand for $\dfrac{n!}{k!\,(n-k)!}$.

The quantities $\binom{n}{k}$ are known as **binomial coefficients**. Here are some properties which can help you calculate then:

- the binomial expansion is symmetrical, with $\binom{n}{k} = \binom{n}{n-k}$; in other words, the coefficient of $x^{n-k}y^k$ is the same as that of $x^k y^{n-k}$; this means the definition of the binomial coefficients conceals some cancellations that always take place;
- the quickest way to calculate $\binom{n}{k}$ is via the formula

$$\binom{n}{k} = \frac{\text{the product of the first } k \text{ integers counting down from } n}{k!}$$

  which holds whenever k is positive; when $k = 0$, the answer is 1 for all n;
- the expression $\binom{n}{k}$ is the number of ways of choosing k objects from a pool of n, and so it occurs a lot in books on Statistics, where it is often referred to as "n choose k"; and
- the Binomial Theorem generalises to the situation where n is a number other than a positive integer, but this is outside the scope of this course.

1.15. Example. Use the binomial theorem to expand $(x+y)^5$.

Solution We first put $n = 5$ in the theorem and write out the summation in full.

$$(x+y)^5 = \binom{5}{0}x^5y^0 + \binom{5}{1}x^4y^1 + \binom{5}{2}x^3y^2 + \binom{5}{3}x^2y^3 + \binom{5}{4}x^1y^4 + \binom{5}{5}x^0y^5.$$

We now start calculating the binomial coefficients

$$\binom{5}{0} = 1, \qquad \binom{5}{1} = \frac{5}{1!} = 5, \qquad \binom{5}{2} = \frac{5.4}{2.1} = 10.$$

This gets us halfway along the row. Symmetry then takes over. Therefore

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

1.16. Example. Go through the calculation for $(x+y)^6$, checking your answer against the one given in Example 1.12.

Solution I'm not always going to give solutions; if you are still having difficulty, you need to ask.

1.4.1 Pascal's Triangle

This is an alternative way of working out binomial coefficients. It is quite efficient when n is small, but not so good when it isn't.

$$\begin{array}{c|ccccccccccccc}
n=0 & & & & & & & 1 & & & & & & \\
n=1 & & & & & & 1 & & 1 & & & & & \\
n=2 & & & & & 1 & & 2 & & 1 & & & & \\
n=3 & & & & 1 & & 3 & & 3 & & 1 & & & \\
n=4 & & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
n=5 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
n=6 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

Within the triangle each number is the sum of the two immediately above it. The n-th row gives the binomial coefficients for $(x+y)^n$. The drawback is that in order to get at a row you have to calculate all its predecessors.

1.4.2 Using the Binomial Theorem

We now show how the theorem can help in situations which on first sight don't seem to be a good fit for the Theorem. This is characteristic of a good theorem; I hope you enjoy seeing it being used.

There is an important point here about variable *names*. Even though you *learn* the Binomial Theorem with variables called x and y, it is not restricted to being used *just* with those variables. A more accurate statement of the theorem might start by saying "given any two variables, which just for the rest of the statement we denote by x and y, we have ..."

1.17. Example. Expand $\left(x+\dfrac{1}{x}\right)^5$.

Solution Write down the expansion for $(x+y)^5$ and then put $y=\dfrac{1}{x}$

1.18. Example. Expand $\left(x^2-\dfrac{1}{x}\right)^6$.

Solution Write down the expansion for $(a+b)^6$ and then put $a=x^2$ and $b=-\dfrac{1}{x}$

1.19. Example. Expand $(1+x+x^2)^4$.

Solution Think of this as $(a+b)^4$ where $a=1+x$ and $b=x^2$. Expand, substitute, and then use the Binomial Theorem again to deal with the powers of $(1+x)$.

1.20. Example. What is the coefficient of x^5 in $\left(x^2+\dfrac{2}{x}\right)^{10}$?

Solution This time it is not necessary to write down the full expansion. The general term in the expansion of $(a+b)^{10}$ is $\dbinom{10}{k}a^{10-k}b^k$. Replacing a by x^2 and b by $\dfrac{2}{x}$ turns this into $\dbinom{10}{k}x^{20-2k}2^kx^{-k}$, i.e. into $2^k\dbinom{10}{k}x^{20-3k}$.

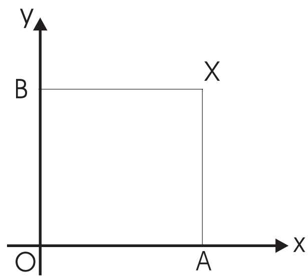
Now work out which k you want. We have $20 - 3k = 5$ when $k = 5$. So the required coefficient is the one with $k = 5$. So the required coefficient is

$$2^5 \binom{10}{5} = 32.\frac{10.9.8.7.6}{5.4.3.2.1} = 32.9.4.7 = 8064$$

1.5 Cartesian Coordinates

There are many ways of describing the position of a point in the plane, and hence of passing from geometry — the picture — to algebra. The advantage of doing this is that we have the whole apparatus of algebra to enable us then to make deductions about the positions of the points. The most common way of referring to the position of points in a plane is to use Cartesian Coordinates which we describe now. There are lots of other coordinate systems in use as well, particularly in specialised applications; in particular we discuss Polar Co-ordinates in section 1.9

We choose an *origin* O and two perpendicular *axes* OX and OY. Then any point P is assigned coordinates (x, y) where x and y are the (signed) distances AP and BP.



1.5.1 Distance between points

By Pythagoras' theorem the distance between the two points $P(x_1, y_1)$ and $Q(x_2, y_2)$ is given by

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1.1}$$

For example, the distance between the points $(1, 2)$ and $(5, 7)$ is $\sqrt{(5-1)^2 + (7-2)^2} = \sqrt{4^2 + 5^2} = \sqrt{41}$. The distance between the points $(-2, -5)$ and $(-3, 7)$ is

$$\sqrt{(-3 - (-2))^2 + (7 - (-5))^2} = \sqrt{(-1)^2 + 12^2} = \sqrt{145}$$

1.5.2 Straight Lines

An important concept for a straight line in the (x, y)-plane is its slope.

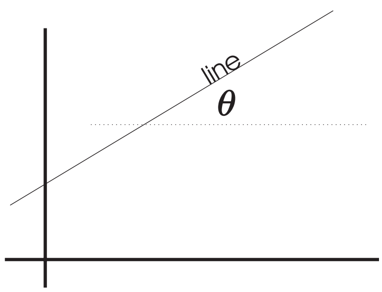


Figure 1.1: The slope of a line.

The *slope*, or *gradient*, of a straight line is the tan of the angle that the line makes with the positive direction of the x-axis

This makes sense for any line other than those lines that are perpendicular to the x-axis. We do not give a slope for those lines.

Two lines with the same slope are parallel. The condition for two lines of slopes m_1 and m_2 to be **perpendicular** is that

$$\boldsymbol{m_1 m_2 = -1} \tag{1.2}$$

The slope of the line joining $P(x_1, y_1)$ to $Q(x_2, y_2)$ is

$$\boldsymbol{m = \frac{y_2 - y_1}{x_2 - x_1}}$$

The Cartesian equation of a straight line having slope m is of the form

$$\boldsymbol{y = mx + c} \tag{1.3}$$

The constant c tells us where the line meets the y-axis.

This equation can be re-written in a couple of other useful forms.

The equation of the straight line through the point (a, b) with slope m is

$$\boldsymbol{y - b = m(x - a)} \tag{1.4}$$

This is clear. The coefficient of x is m, so the line has slope m. When $x = a$ the RHS is zero, so $y = b$. Therefore the line goes through (a, b).

The equation of the straight line through (x_1, y_1) and (x_2, y_2), if $x_1 \neq x_2$, is

$$\boldsymbol{y - y_1 = \frac{y_1 - y_2}{x_1 - x_2}(x - x_1)} \tag{1.5}$$

All the forms so far have had the slight disadvantage that, since they are written in terms of the gradient, they do not apply to the case ($x =$ constant) of a line perpendicular to the x-axis.

This is corrected as follows. The fully general equation of a straight line is

$$\boldsymbol{ax + by + c = 0} \tag{1.6}$$

where a and b are not both zero.

If $b \neq 0$ then we can re-arrange the equation as $y = -\frac{a}{b}x - \frac{c}{b}$ which is of the form $y = mx + c$ as before. If, on the other hand, $b = 0$ then the equation becomes

$$x = -\frac{c}{a} = \text{constant}$$

which covers the case of lines perpendicular to the x-axis.

1.21. Example. What is the equation of the straight line through $(1, 2)$ with slope -1?

Solution Using the above formula the equation is $y-2 = -1(x-1)$ or, tidying up, $y = 3-x$.

1.22. Example. What is the equation of the straight line through $(4, 5)$ which is perpendicular to the line $y = 2x - 3$?

Solution The second line has slope 2. So, by the formula $m_1m_2 = -1$ the required slope of our line is $-1/2$. So the equation of the line is

$$y - 5 = -\tfrac{1}{2}(x - 4) \quad \text{or} \quad x + 2y = 14$$

1.23. Example. What is the equation of the straight line through the two points $(1, 3)$ and $(-3, 8)$?

Solution We can do this by the earlier formula, but it is probably easier to do it in two stages. First, the slope of the line must be

$$m = \frac{8 - 3}{-3 - 1} = -\frac{5}{4}$$

So the equation is

$$y - 3 = -\tfrac{5}{4}(x - 1) \quad \text{or} \quad 5x + 4y = 17$$

1.5.3 Finding the intersection of two lines

Suppose two lines intersect at the point (x, y). Then this point must lie on both lines, so x and y must satisfy both equations. So we are need to treat the equations of thetwo lines as a pair of simultaneous equations

1.24. Example. Find the point of intersection of the line which passes through $(1, 1)$ and $(5, -1)$ and the line which passes through $(2, 1)$ and $(3, -3)$.

Solution Line one has equation $y - 1 = \left(\frac{-1-1}{5-1}\right)(x-1)$ i.e. $2y + x = 3$. Line two has equation $y - 1 = \left(\frac{-3-1}{3-2}\right)(x-2)$ i.e. $y + 4x = 9$. Solving these two equations simultaneously gives us the point $(\frac{15}{7}, \frac{3}{7})$.

1.25. Example. Where do the two lines $y = 3x - 2$ and $y = 5x + 7$ meet?

Solution The point (x, y) where they meet must lie on both lines, so x and y must satisfy both equations. So we are looking to solve the two simultaneous equations

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

Well, in that case, $3x - 2 = 5x + 7$, so $2x = -9$ and $x = -9/2$. Putting this back in one of the equations we get $y = -27/2 - 2 = -31/2$.

Here are some examples of working with lines to practice with.

1.26. Example. What is the equation of the straight line passing through the point $(-3, 5)$ and having slope 2?

1.27. Example. What is the equation of the line passing through the points $(-4, 2)$ and $(3, 8)$? What are the equations of the lines through $(5, 5)$ parallel and perpendicular to this line?

1.28. Example. Where do the lines $y = 4x - 2$ and $y = 1 - 3x$ meet? Where does the line $y = 5x - 6$ meet the graph $y = x^2$?

1.29. Example. Let l_1 be the line of slope 1 through $(1, 0)$ and l_2 the line of slope 2 through $(2, 0)$. Where do they meet and what is the area of the triangle formed by l_1, l_2 and the x-axis?

1.6 Angles and Trigonometry

Mathematicians normally measure angles in two ways, or systems: degrees and radians. Degrees are easy and familiar. There are 360 degrees in a single rotation. A right angle is therefore $90°$. Radians are less familiar, but you must get used to them because most of mathematical theory involving angles is expressed in terms of radians.

The definition of a *radian* is that it is *the angle subtended at the centre of a circle by a piece of the circumference of the circle of length equal to the radius of the circle.*

This means that, since the circumference of a circle of radius r is $2\pi r$, a single rotation is 2π radians.

Outside mathematics the use of degrees is universal, because the numbers are nicer. Degrees are also the older system, being in use by Greek mathematicians in the second century BC, with throwbacks from there to the Babylonian astronomers and mathematicians with their base 60 number system. Radians, along with the trigonometric functions such as sine, were introduced by Indian mathematicians in the sixth century AD. The numbers involved are more awkward, but the system gives a neater relationship between angle and arc length on the circle, and this makes for tidier formulas elsewhere. This is especially true of calculus formulas; calculus using degrees is a mess, and so in this context radians are preferred by everyone.

When you are doing calculus you *always* use radians. A lot of the formulas for standard derivatives and integrals are false if you don't.

Warning: Most calculators wake up in degree mode. You must switch to radians before doing any calculation for this course which involves a trig function. If you don't, you will get wrong answers, because the formulas you are using are built on radian measurement.

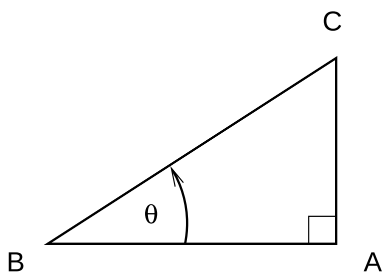
We have the following obvious conversions:

$$360^\circ = 2\pi \text{ radians} \quad 180^\circ = \pi \text{ radians} \quad 90^\circ = \tfrac{1}{2}\pi \text{ radians}$$

$$45^\circ = \frac{1}{4}\pi \text{ radians} \quad 60^\circ = \frac{1}{3}\pi \text{ radians} \quad 30^\circ = \frac{1}{6}\pi \text{ radians}$$

Angles are traditionally measured *anticlockwise*. An angle measured in the clockwise direction is taken to be negative (thinking of it as a 'turn' rather than a 'separation').

1.6.1 Trig Functions



You will have met sine, cosine and tan. They are usually introduced in the context of a right-angled triangle, as shown.

$$\sin\theta = \frac{\text{CA}}{\text{CB}}, \qquad \cos\theta = \frac{\text{AB}}{\text{CB}}, \qquad \tan\theta = \frac{\text{CA}}{\text{AB}}$$

We also sometimes use other combinations of these functions:

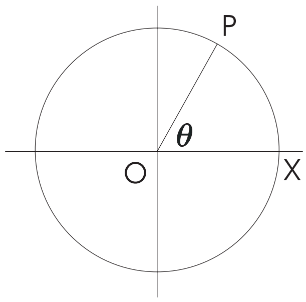
$$\operatorname{cosec}\theta = \frac{1}{\sin\theta}, \qquad \sec\theta = \frac{1}{\cos\theta}, \qquad \cot\theta = \frac{1}{\tan\theta}$$

The weakness of the triangle definition of the trig functions is that it only makes sense for positive angles up to 90°. We want a much more general definition than that, because we want to be able to take trig functions of almost any angle.

A better definition goes as follows. Draw the circle

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

of radius 1 with centre at the origin. Let OX be the radius along the positive x-axis.



Let θ be any angle (positive or negative). Turn OX through angle θ so that it ends up at OP. Then

$\cos\theta$ is the x-coordinate of P and $\sin\theta$ is the y-coordinate of P. i.e.

$$P = (\cos\theta, \sin\theta)$$

This now gives us a definition that works for all possible angles. We still define $\tan\theta$ as $\sin(\theta)/\cos(\theta)$.

We can draw some immediate and important conclusions from this picture.

- Rotating OP through a full rotation (in either direction) brings us back to where we started. So the trig functions are *periodic* with period 2π:

$$\cos(x + 2\pi) = \cos(x) \qquad \sin(x + 2\pi) = \sin(x)$$

- Rotating through half a rotation takes us from (x, y) to the opposite point $(-x, -y)$. So

$$\sin(x + \pi) = -\sin(x) \qquad \cos(x + \pi) = -\cos(x)$$

- Rotating through 90° has a more complicated effect. If you think about it you will see that the effect is to send (x, y) to $(-y, x)$. So

$$\cos(x + \tfrac{1}{2}\pi) = -\sin(x) \qquad \sin(x + \tfrac{1}{2}\pi) = \cos(x)$$

- Rotating through a negative angle is like rotating through the corresponding positive angle and then reflecting in the x-axis. i.e. if a turn of OX through negative angle θ gets us to (x, y) then a turn through the corresponding positive angle will take us to $(x, -y)$. So

$$\sin(-x) = -\sin(x) \qquad \cos(-x) = \cos(x)$$

It is also worth knowing that

$$\sin(\pi/2 - x) = \cos(x) \qquad \cos(\pi/2 - x) = \sin(x)$$

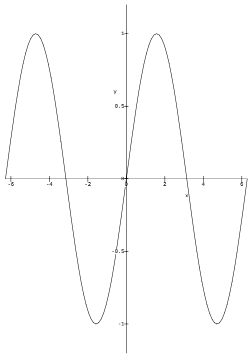
Because the point P lies on the circle we have the very important relationship

$$\mathbf{\sin^2(\theta) + \cos^2(\theta) = 1} \tag{1.7}$$

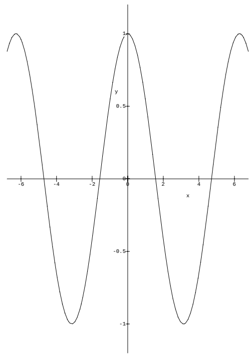
and from this we get

$$1 + \tan^2(\theta) = \sec^2(\theta)$$

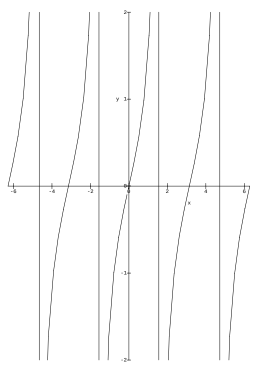
The graphs of the three basic trig functions, sin, cos and tan, look like this:



$y = \sin x$



$y = \cos x$



$y = \tan x$

The values of the trig functions for some special angles are worth knowing. They are given in Table 1.1.

Deg	Rad	cos	sin	tan
0	0	1	0	0
180	π	-1	0	0
90	$\pi/2$	0	1	-
45	$\pi/4$	$1/\sqrt{2}$	$1/\sqrt{2}$	1
60	$\pi/3$	$1/2$	$\sqrt{3}/2$	$\sqrt{3}$
30	$\pi/6$	$\sqrt{3}/2$	$1/2$	$1/\sqrt{3}$

Table 1.1: Trigonometric function at some special angles.

There are important *addition formulas* for the trig functions:

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b) \tag{1.8}$$
$$\sin(a-b) = \sin(a)\cos(b) - \cos(a)\sin(b) \tag{1.9}$$
$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b) \tag{1.10}$$
$$\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b) \tag{1.11}$$

and from these we get the special cases

$$\sin(2a) = 2\sin(a)\cos(a) \tag{1.12}$$
$$\cos(2a) = \cos^2(a) - \sin^2(a) \tag{1.13}$$
$$= 2\cos^2(a) - 1 \tag{1.14}$$
$$= 1 - 2\sin^2(a) \tag{1.15}$$

Here are some examples of working with trigonometric formulae to practice with.

1.30. Example. Starting from the trig formulas (1.8–1.11) show that

$$\sin 2x = 2\sin x\cos x, \qquad \cos 2x = 2\cos^2 x - 1 = 1 - 2\sin^2 x,$$

$$\sin^2 x = \frac{1}{2}(1-\cos 2x), \qquad \cos^2 x = \frac{1}{2}(1+\cos 2x),$$

1.31. Example. Do these in the same way.

$$\tan(x+y) = \frac{\sin(x+y)}{\cos(x+y)} = \frac{\tan x + \tan y}{1 - \tan x \tan y}, \qquad \tan 2x = \frac{2\tan x}{1-\tan^2 x}$$

$$1 + \tan^2 x = \sec^2 x, \qquad 1 + \cot^2 x = \operatorname{cosec}^2 x.$$

1.32. Example. Use the fact that you know the values of the trig functions at π and $\pi/2$ to show that

$$\sin(\pi - x) = \sin x, \qquad \cos(\pi - x) = -\cos(x), \qquad \sin(\pi/2 - x) = \cos(x),$$

$$\sin(x + 2\pi) = \sin x, \qquad \cos(x + 2\pi) = \cos x$$

1.7 Circles

The circle with centre P and radius r is the set of points in a plane that are at distance r from P. By our formula for the distance between two points, we can say that the distance between (x, y) and (a, b) is r precisely if

$$\boldsymbol{(x-a)^2+(y-b)^2=r^2}$$

This is therefore the equation of a circle of radius r having centre (a, b).

We get a simple special case if the centre happens to be at the origin. Then the equation becomes

$$x^2+y^2=r^2$$

Our definition of the trigonometric functions really boils down to the statement that any point on this circle can be expressed in the form

$$x(\theta)=a+r\cos\theta \qquad y(\theta)=b+r\sin\theta$$

The most general form for the equation of a circle is

$$x^2+y^2+2ax+2by+c=0$$

Why is this a circle and, if it is, what are its centre and radius?

This is an exercise in 'completing the square'.

Write x^2+2ax as $(x+a)^2-a^2$ and y^2+2by as $(y+b)^2-b^2$. Then our equation becomes

$$(x+a)^2+(y+b)^2=a^2+b^2-c$$

This is a circle with centre at $(-a, -b)$, provided that the RHS of the equation is positive. In that case, the radius is $r=\sqrt{a^2+b^2-c}$.

1.7.1 Tangents

The tangent line to a circle, with centre O, at a point P is the straight line through P that touches the circle. By the geometry of the circle this line is perpendicular to the radius OP. So we can write down the equation of the tangent: we know it goes through P and we know its slope from the formula $m_1m_2=-1$.

Here are some examples of working with circles to practice with.

1.33. Example. What is the equation of the circle with centre $(1, 2)$ and radius 3? Find the two points where this circle meets the x-axis and the two points where this circle meets the y-axis. What is the area of the quadrilateral formed by these four points?

1.34. Example. What are the centre and the radius of the circle with equation

$$x^2+y^2-2x+4y=4 \quad ?$$

1.35. Example. Simplify the following expressions: $x^{2/3}x^{-3/2}$, $(x^{-1}\sqrt{y})^{3/2}$. Expand out the following brackets and simplify the results: $x^{1/3}(x^{2/3}+y^{1/3})$, $(x^{1/2}+x^{-1/2})(x^{1/2}-x^{-1/2})$.

1.8 Parameters

With other curves you sometimes work with an equation and sometimes with parameters.

Curves are sometimes best thought of as paths traced out by a moving point. When you do this the most convenient way of specifying the path is to give the x-coordinate and the y-coordinate of the point "at time t". So you end up with both x and y as functions of a parameter, which you can think of as time.

1.36. Example. The path traced out by a point on the circumference of a wheel that is rolling along in a straight line is given (with respect to suitable coordinates) by

$$x = a(t - \sin t)\,, \qquad y = a(1 - \cos t)$$

where t is the parameter and a is the radius of the wheel.

1.37. Example. The circle centre (a, b) and radius r can be described parametrically by

$$x = a + r\cos\theta\,, \qquad y = b + r\sin\theta$$

This time θ is the parameter.

This is often the simplest way to deal with complicated curves, and is clearly the approach to adopt if you are plotting a curve using a computer.

1.9 Polar Co-ordinates

We have always used Cartesian Coordinates to describe points in a plane. There are lots of other coordinate systems in use as well, particularly in specialised applications. One well-known coordinate system is that known as Polar Coordinates. It is especially useful in situations where information is most conveniently expressed in terms of distance from the origin.

Let OX and OY be Cartesian axes in the plane. Let P be a point in the plane, other than the origin.

The **Polar Coordinates** of P are (r, θ) where $r = OP > 0$ and θ is the angle from OX round to OP anticlockwise.

By convention, we take the angle range to be $-\pi < \theta \leq \pi$. (some people take $0 \leq \theta < 2\pi$).

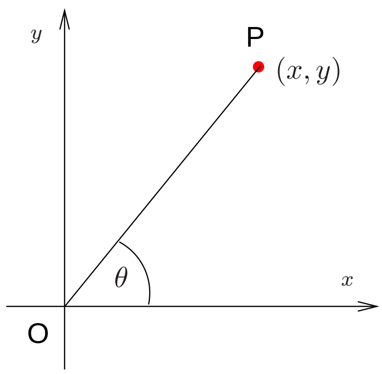


Figure 1.2: Polar co-ordinates (r, θ) of a point.

Note that we do not give polar coordinates for the origin, because the angle does not make sense there.

Now consider the problem of converting between Cartesian and polar coordinates. One way round is easy:

$$x = r\cos\theta \qquad y = r\sin\theta$$

(these are really just the definitions of $\sin\theta$ and $\cos\theta$.)

The other way round is more tricky and needs some understanding of inverse trig functions. First of all, without any difficulty,

$$r = \sqrt{x^2 + y^2}$$

The real problem is to get θ in terms of x and y. Dividing the above equations gives

$$\tan\theta = \frac{y}{x}$$

So we are tempted to write

$$\theta = \arctan(\frac{y}{x})$$

This is wrong.

First of all, we have to be careful about the case $x = 0$ (which corresponds to the y-axis) because we don't want to divide by zero.

Inspection of Figure 1.2 tells us that

if $x = 0$ then
  if $y > 0$ then $\theta = \pi/2$
  if $y < 0$ then $\theta = -\pi/2$

Our problems don't end there. Recall that arctan is *defined* as taking values in the range

$$-\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

So it is not good enough in itself to solve our problem. The obvious trouble is that, for example, $(1, 1)$ and $(-1, -1)$ give the same value for y/x even though they have different angles ($\pi/4$ and $-3\pi/4$). We have to take account of the quadrant in which the point lies.

The correct conversion rules are these—which you should check through as an exercise.

if $x > 0$ then $\theta = \arctan(y/x)$
if $x < 0$ then
  if $y \geq 0$ then $\theta = \arctan(y/x) + \pi$
  if $y < 0$ then $\theta = \arctan(y/x) - \pi$

Chapter 2

The Derivative

2.1 Introduction

One of the oldest problems in mathematics, going back at least as far as the ancient Egyptians, is that of determining area: tax gatherers needed to know how much land people had so that they could tax them on it. So when, in the 17th century, mathematicians came up with the idea of coordinate geometry, and with it the idea of a curve represented by an equation $y = f(x)$, one of the questions that obsessed them was "How do you find the area under the curve $y = f(x)$?".

A less obviously useful but still interesting problem is "Given a curve $y = f(x)$ and a point (x_0, y_0) on the curve, what is the equation of the tangent?".

What the inventors of calculus discovered was

- a method for solving the tangent problem.
- that the tangent problem was the key to the area problem.
- that their technique for solving the tangent problem enabled them to deal with problems concerning changing situations — movement etc.

Previously mathematicians had handled static problems; now they could tackle dynamic ones. And of course it is this that has made calculus so important in applications of mathematics, not least in engineering.

2.2 The Derivative

The tangent problem is the easiest place to start. For the equation of a line you need to know either two points or one point and the slope. Here we have one point, and so somehow we have to calculate the slope.

We have a curve $y = f(x)$ and a point $P = (x_0, y_0)$ on it. To get the equation of the tangent at P we need to calculate the slope.

For example, let Q be a point on the curve close to P. Then the slope of the tangent is close to that of the chord PQ. Moreover, the closer Q gets to P the better the approximation will be. So write down the slope of PQ, and then see what happens when Q is pushed towards P.

P has x-coordinate x_0 and y-coordinate $y_0 = f(x_0)$. Q is close to P and so will have x-coordinate $x_0 + h$, where h is some small quantity. The smaller h is the closer Q will be to P.

The y-coordinate of Q is $f(x_0 + h)$.

It follows that the slope of PQ is

$$\frac{f(x_0 + h) - f(x_0)}{h}$$

The hope is that we will be able to tell what happens to this quotient as h gets smaller and smaller.

The only way to see if this will work is to try an example.

2.1. Example. Consider the curve $y = x^2$.

$$\begin{aligned}\frac{f(x_0 + h) - f(x_0)}{h} &= \frac{(x_0 + h)^2 - (x_0)^2}{h}, \\ &= \frac{2x_0 h + h^2}{h}, \\ &= 2x_0 + h.\end{aligned}$$

It is working. We can see what is going to happen as h gets smaller and smaller.

It is clear that, as h shrinks, the slope of the chord will approach a limiting value of $2x_0$. So the slope of the tangent at $P(x_0, x_0^2)$ is $2x_0$. Consequently the equation of the tangent is $y - x_0^2 = 2x_0(x - x_0)$, which simplifies to $y = 2x_0 x - x_0^2$.

2.2. Definition. $\dfrac{f(x_0 + h) - f(x_0)}{h}$ is called the **Newton quotient** of f at x_0.

2.3. Definition. The limiting value of the Newton quotient as h shrinks to zero is denoted by $\lim_{h\to 0}\left(\dfrac{f(x_0 + h) - f(x_0)}{h}\right)$ and is called the **derivative** of f at x_0.

Notation The derivative of f at x_0 is denoted by $\dfrac{df}{dx}(x_0)$ and also by $f'(x_0)$.

Calculating the limiting values of Newton quotients is not usually as easy as in the example above. We will go through one important derivative using it but quote others without proof.

Geometrical Observation

When the derivative is *positive* the slope of the tangent is positive, and the function is *increasing*.

When the derivative is *negative* the slope of the tangent is negative, and the function is *decreasing*.

When the derivative is zero the tangent is parallel to the x-axis.

It is often useful to know whether a function is increasing or decreasing. The derivative gives you a means of finding out.

2.3 The Derivative as a Rate of Change

Suppose we were to plot distance covered against time elapsed for some journey, getting a curve

The curve is $x = f(t)$ for some function f. Let P be the point on the curve corresponding to the time $t = t_0$, and let Q be that corresponding to $t = t_0 + h$.

To work out the average speed for the section of the journey between the points represented by P and Q we take the distance covered and divide it by the time taken. The time elapsed is $(t_0 + h) - t_0 = h$, and the distance covered is $f(t_0 + h) - f(t_0)$.

I am assuming that the journey is as shown, with no backtracking.

Therefore, the average velocity is

$$\frac{f(t_0 + h) - f(t_0)}{h}$$

And, of course, this is just the Newton quotient.

What about the velocity *at* time t_0?

There is only one sensible answer

$$\lim_{h \to 0} \left(\frac{f(t_0 + h) - f(t_0)}{h} \right)$$

The instantaneous velocity is the limiting value of the average velocity measured over a very small interval.

So the derivative at P gives the velocity at time t_0.

This works more generally. If you plot one quantity against another, the derivative picks up the instantaneous rate at which one is changing relative to the other.

2.4 Three Standard Derivatives

The Newton quotient and calculations concerning its limiting values as h gets small provide the foundation on which calculus is built. However, it is not the way derivatives are usually calculated in practice. What we do instead is build up some theory that will enable us to come in at a later stage of the process. It works as follows:

The most efficient way to calculate derivatives is

- to build up a library of the derivatives of commonly occurring functions.
- to develop a set of rules which enable you to build up the derivatives of more complicated functions by using the standard list as a kit.

2.4.1 The derivative of x^n

2.4. Example. Let $f(x) = x^4$. Then we have

$$\begin{aligned}\frac{f(x+h)-f(x)}{h} &= \frac{(x+h)^4 - x^4}{h}\\ &= \frac{(x^4+4x^3h+6x^2h^2+4xh^3+h^4)-x^4}{h}\\ &= \frac{4x^3h+6x^2h^2+4xh^3+h^4}{h}\\ &= 4x^3 + h(6x^2+4xh+h^2)\end{aligned}$$

As h get small, $h(6x^2+4xh+h^2)$ shrinks to nothing, and so the derivative is $4x^3$.

This argument works for any positive integer n. You use the Binomial Theorem to expand $(x+h)^n$, the h from the bottom line will then cancel into the top, leaving you with an expression of the form $nx^{n-1} + h(\textit{something})$. The $h(\textit{something})$ then becomes negligible as h shrinks to nothing, leaving nx^{n-1} as the derivative. So we have

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

The Binomial Theorem argument only establishes this for n a positive integer n. However, the formula does in fact work for *all* powers – positive, negative and fractional.

2.5. Example. The formula can be used to get the derivative of $\sqrt{x}$.

$$\begin{aligned}\frac{d}{dx}\left(\sqrt{x}\right) &= \frac{d}{dx}\left(x^{\frac{1}{2}}\right)\\ &= \frac{1}{2}x^{\frac{1}{2}-1}\\ &= \frac{1}{2}x^{-\frac{1}{2}}\\ &= \frac{1}{2\sqrt{x}}\end{aligned}$$

2.6. Example. Likewise for inverse powers.

$$\begin{aligned}\frac{d}{dx}\left(\frac{1}{x^2}\right) &= \frac{d}{dx}\left(x^{-2}\right)\\ &= (-2)x^{-2-1}\\ &= -2x^{-3}\\ &= -\frac{2}{x^3}\end{aligned}$$

2.4.2 The derivatives of sine and cosine

These are quite tricky. You need to use some of the trig formulas and then to do some delicate arguments with areas of segments of circles. We will quote the results without proof.

Provided x is measured in radians

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

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

2.5 Rules for Differentiation

These are the means by which you work out the derivatives of functions which have been built up from the basic ones by adding, multiplying, and so on. There are four of them, and they are used so often that you should learn them and not just rely on your handbooks.

2.5.1 SUMS

Let f and g be functions of x, and suppose that the derivatives of both exist. Then

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

Equivalently, using the dash notation,

$$(f+g)' = f' + g'$$

This is the **sum rule**

2.7. Example. The derivative of $x^3 + \sin x$ is $3x^2 + \cos x$.

Here $f(x) = x^3$, which has derivative $3x^2$, and $g(x) = \sin x$, which has derivative $\cos x$.

The rule extends in an obvious way to the sums of more than two functions. For example, the derivative of $x^3 + \sin x + \cos x$ is $3x^2 + \cos x - \sin x$.

2.5.2 PRODUCTS

Let f and g be functions of x, and suppose that the derivatives of both exist. Then

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

Equivalently, using the dash notation,

$$(fg)' = f'g + fg'$$

This is the **product rule**.

2.8. Example. The derivative of $x^3 \sin x$ is $3x^2 \sin x + x^3 \cos x$.

Here $f(x) = x^3$, which has derivative $3x^2$, and $g(x) = \sin x$, which has derivative $\cos x$. So $f'g$ is $3x^2 \sin x$, while fg' is $x^3 \cos x$.

2.9. Remark. The product rule tells us how to deal with functions such as $cg(x)$, where c is a constant.

Think of them as a product fg, where f is the constant function $f(x) = c$ for all x. The graph of $y = f(x)$ is a horizontal straight line, and so has slope zero. So f' is zero. Consequently the expression $f'g + fg'$ collapses down to fg', i.e. to cg'. So we have

$$\frac{d}{dx}(cg) = c\,\frac{dg}{dx} \qquad \text{if } c \text{ is a constant}$$

2.10. Example. The derivative of $2\sin x$ is $2\cos x$. The derivative of $2x^4$ is $8x^3$.

2.11. Remark. If you have a product of more than two functions to deal with, you take it in stages.

2.12. Example. Find the derivative of $x^3 \sin x \cos x$.

Think of the function as the product of $f(x) = x^3$ and $g(x) = \sin x \cos x$, and apply the product rule to get

$$\frac{d}{dx}\left(x^3 \sin x \cos x\right) = 3x^2(\sin x \cos x) + x^3 \frac{d}{dx}(\sin x \cos x)$$

Now use the product rule again, this time to get the derivative of $\sin x \cos x$.

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

Substituting this into the earlier expression gives

$$\frac{d}{dx}\left(x^3 \sin x \cos x\right) = 3x^2(\sin x \cos x) + x^3(\cos^2 x - \sin^2 x)$$

It is not difficult; you just have to take it in steady stages.

2.5.3 QUOTIENTS

Let f and g be functions of x, suppose that the derivatives of both exist and that $g'(x) \neq 0$. Then

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

Equivalently, using the dash notation,

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$$

This is the **quotient rule**.

Again, this is something that you should commit to memory. Note that the minus sign on the top row goes with the derivative of the function from the bottom line — i.e. with the derivative of the function that occurs to the power minus one.

Using the quotient rule we can expand our list of standard derivatives by getting the derivatives of the other trig functions.

2.13. Example. Find the derivative of $\tan x$.

$$\tan x = \frac{\sin x}{\cos x}$$

Therefore

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

The derivatives of $\cot x$, $\sec x$ and $\operatorname{cosec} x$ can be calculated in similar fashion. The result is the following table:

$$\frac{d}{dx}(\tan x) = \sec^2 x \qquad \frac{d}{dx}(\cot x) = -\operatorname{cosec}^2 x$$

$$\frac{d}{dx}(\sec x) = \sec x \tan x \qquad \frac{d}{dx}(\operatorname{cosec} x) = -\operatorname{cosec} x \cot x$$

2.5.4 THE CHAIN RULE

If you think of a functions as a machine which has numbers as both input and output, then the chain rule is designed to cope with those machines that are made up of a "chain" of simpler machines each of which takes as input the output of its predecessor in the chain.

The following examples are situations covered by this rule.

2.14. Example. $f(x) = \sin(x^3)$

2.15. Example. $f(x) = \sqrt{1 + \sin x}$

In both of these you have a succession of functions, with the final answer being got by applying them one after the other. So with the first one you take the given number, you cube it and then you apply the sine function. The **chain rule** tells you how to differentiate such composite functions.

The Chain Rule: Version 1

If $f(x) = g(h(x))$ and g, h are functions that can be differentiated, then

$$f'(x) = g'(h(x)).h'(x)$$

With the first of the examples above, $h(x) = x^3$ and g is the function sine. Since g is sine, g' is cosine. We also have that $h'(x) = 3x^2$. So $f'(x) = \cos(x^3).3x^2$.

With the second example, $h(x) = 1 + \sin x$ and g is the function "square root". So

$$f'(x) = \frac{1}{2}(1 + \sin x)^{-\frac{1}{2}}.\cos x = \frac{\cos x}{2\sqrt{1 + \sin x}}$$

This is the traditional, Newtonian, way of describing the rule. However, I think that the alternative formulation, due to Leibniz, is both easier to remember and easier to use.

The Chain Rule: Version 2

This uses the $\frac{dy}{dx}$ notation, and introduces a subsidiary variable z.

If $y = g(z)$ where $z = h(x)$, then

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

To remember the rule you just think of dy, dz and dx as being small quantities in their own right and then visualise the "cancellation" of the dz.

2.16. Example. $y = \dfrac{1}{\sqrt{\sin x}}$.

Think of y as $\dfrac{1}{\sqrt{z}} = z^{-\frac{1}{2}}$ where $z = \sin x$. Then

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{dz}.\frac{dz}{dx} \\ &= -\frac{1}{2}z^{-\frac{3}{2}}.\cos x \\ &= -\frac{1}{2}(\sin x)^{-\frac{3}{2}}.\cos x \\ &= -\frac{\cos x}{2(\sin x)^{\frac{3}{2}}}\end{aligned}$$

2.17. Example. $y = (x^2 + \tan x)^8$

Think of y as z^8 where $z = x^2 + \tan x$. Then

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{dz}.\frac{dz}{dx} \\ &= 8z^7.(2x + \sec^2 x) \\ &= 8(x^2 + \tan x)^7(2x + \sec^2 x)\end{aligned}$$

The Leibniz version also makes it easy to deal with longer chains.
If $y = g(u)$ where $u = h(v)$ and $v = k(x)$, then

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

This time you can imagine both the du and the dv being cancelled.

2.18. Example. $y = \sin(\sin(1 + x^2))$

Break this down as

$$y = \sin u$$

where

$$u = \sin v$$

and

$$v = 1 + x^2$$

Then

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du}.\frac{du}{dv}.\frac{dv}{dx} \\ &= \cos u.\cos v.2x \\ &= \cos(\sin v).\cos(1 + x^2).2x \\ &= \cos(\sin(1 + x^2)).\cos(1 + x^2).2x\end{aligned}$$

With complicated functions you often have to use several of the four rules in combination.

2.19. Example. Differentiate $\dfrac{x^2 \cos\left(x^{\frac{3}{2}}\right)}{\tan x}$.

Chapter 3

Rates of Change etc.

3.1 Introduction

Now we consider a certain function f of a certain variable t (usually time) and ask how f changes as t changes. Consider the following examples.

3.1. Example. A stone is thrown vertically upwards into the air so that its height in metres is given by $h(t) = 5 + 20t - 5t^2$, where t is the time elapsed since launch (t is in seconds).

(1) How fast is the stone travelling when $t = 1$ and when $t = 3$?

(2) How high does it get?

(3) What is its average speed between $t = 1$ and $t = 3$?

Note that the sign of h' picks out the direction of travel. If h' is positive, h is increasing, and so the stone is travelling upwards. If h' is negative, the stone is travelling downwards.

In that example we were dealing with a situation that we could picture in our heads and the expression that we were interested in — the height — was given to us explicitly in terms of the other key variable — time. Normally things aren't quite so simple.

3.2. Example. A ball is thrown so that its position t seconds after launch is given by $x = 32t$, $y = 32t - 16t^2$. (This time x and y are in feet and air resistance is being ignored.)

(1) How far is the ball from the start at the top of its flight?

(2) At what rate is the distance between the ball and its position of launch increasing when the ball is at the top of its flight?

Solution This time the situation is more complicated, and so we need a picture.

The axes have been chosen so that the ball begins its journey at the point $(0, 0)$.

The picture shows the position at some unspecified time t. *This is important, because we want to be quite clear in our minds about which quantities are varying and which are fixed.*

Let s be the distance of the ball from its launch point.

The next stage is to decide just what it is that we need to calculate.

(1) *How do we recognise when the ball is at the top of its flight?*

The ball is at the top of its flight when it has stopped going up but has not yet started to come down. So y' has stopped being positive, but has not yet gone negative. That being so, we can see that this is the point when $y' = 0$.

So that is the strategy: calculate y', find the value of t for which it is 0, and then calculate s for this value of t. There is a general point to note here: if you are tackling a problem, inside mathematics or out of it, it helps if you first think about strategy.

$y' = 32 - 32t$, which is 0 when $t = 1$.

When $t = 1$, $x = 32$ and $y = 16$. So

$$s^2 = x^2 + y^2 = 32^2 + 16^2 = 16^2(2^2 + 1) = 16^2.5$$

Therefore $s = 16\sqrt{5}$.

(2) For this we need to calculate s' when $t = 1$. The procedure is to get s in terms of t, to differentiate and then to put $t = 1$.

Note that putting in the particular value of t comes last in the process. It is no use fixing t and **then** *trying to differentiate.*

$$\begin{aligned} s^2 &= x^2 + y^2 \\ &= (32t)^2 + (32t - 16t^2)^2 \\ &= 1024t^2 + 1024t^2 - 1024t^3 + 256t^4 \\ &= 2048t^2 - 1024t^3 + 256t^4 \\ &= 256(8t^2 - 4t^3 + t^4) \qquad (1) \end{aligned}$$

There are now two ways to proceed.

Method 1: This is the obvious one. Take square roots to get s as a function of t, differentiate and then put $t = 1$.

Method 2: This removes the mess from the differentiation process.

Go back to equation (1) and differentiate both sides.

The two sides were equal at the start, we have done the same thing to both, and so they must still be equal.

This gives

$$2ss' = 256(16t - 12t^2 + 4t^3)$$

We know that $s = 16\sqrt{5}$ when $t = 1$. So all we have to do is put in these values for s and t and we have the required value for s'.

Now we draw up a general plan for these problems.

General Plan:

- Draw a picture of the situation at *general time t*.
- Get clear in your mind what you are trying to find and just what it is you have been given in the way of information.
- Express the quantity you are interested in in terms of a single variable. This may involve you in doing some geometry to get rid of surplus variables. Then differentiate. Then put in the values that relate to the special situation that you are interested in.

3.3. Example. Water is being poured into a cone at the rate of 5cc/sec. The cone is 10cm deep, and the angle between the side of the cone and the vertical is thirty degrees ($\pi/6$ radians). How fast is the level rising when the depth of water in the cone is 5cm?

Solution First draw the picture at general time t. Having done this, give names to all the variable quantities that look likely to be of relevance.

Get clear in your mind what you are trying to find and what you have been given. In this case we want $\frac{dh}{dt}$ and have been given information about $\frac{dV}{dt}$. So a sensible strategy is to look for the connection between h and V. The handbook tells us that $V = \frac{1}{3}\pi r^2 h$.

This formula contains three variables, which is one too many for our purposes. So look for a way of getting rid of one of them, and it is clearly r that we would like to see go. The geometry enables us to express r in terms of h. The result is

$$V = \frac{1}{9}\pi h^3$$

This doesn't give h as a function of t, as written it doesn't even give h as an explicit function of V, but none of this matters. If we differentiate both sides of the equation with respect to t, $\frac{dh}{dt}$ will appear, and that is likely to be good enough.

Differentiating both sides with respect to t gives

$$\frac{dV}{dt} = \frac{1}{3}\pi h^2 \frac{dh}{dt}$$

We know V', we know the value of h that interests us, and so if we plug in these values, we can get the required value of h'.

3.4. Remark. The differentiation technique that we used here, and also in method two in the previous example is known as **implicit differentiation**. The definition of the derivative works with explicitly defined functions. So to find $\frac{dh}{dt}$ you begin with h as a function of t. However, as our two examples show, the way the chain rule operates means that you don't have to have things in this form. If you have an equation involving h, this defines h *implicitly*, and this is enough. Differentiate through the equation and h' will appear.

Here is another example

3.5. Example. An observer stands 100 metres from the launch pad of a rocket, which then blasts off so that its height at time t is given by $h = 25t^2$. How quickly is the angle of elevation observer to rocket) increasing two seconds after launch?

Solution The picture produces the formula $h = 100 \tan\theta$. We want $\frac{d\theta}{dt}$.

With our existing repertoire of functions, we can't even get θ as an explicit function of t, but if we just differentiate straight through the formula (implicit differentiation), we shall get the answer we are looking for.

$$100 \tan\theta = 25t^2$$

Therefore

$$4 \tan\theta = t^2$$

Differentiating

$$4\sec^2\theta\frac{d\theta}{dt} = 2t$$

Now calculate $\tan\theta$ and hence $\sec^2\theta$ for the required value of t, substitute, and you have the answer.

3.2 Higher Derivatives

Differentiation is something you do to functions. The result of differentiating a function is another function. That being so, we can differentiate this other function, thereby getting the derivative of the derivative. These secondary and later derivatives are called **higher derivative**s of the original function.

Let f be a function of x.

The derivative of f is denoted by f' (Newton) or $\frac{df}{dx}$ (Leibniz).

Differentiating the derivative gives us $(f')'$ (Newton) or $\frac{d}{dx}\left(\frac{df}{dx}\right)$ (Leibniz), which notations are shortened to f'' and $\frac{d^2f}{dx^2}$ respectively.

These derivatives of the derivative are called the **second derivative** of f.

You can then differentiate the second derivative to get the **third derivative**, and so on.

Notation: The n-th derivative of f is denoted by $f^{(n)}$ (Newton) and by $\frac{d^nf}{dx^n}$ (Leibniz). In the Newtonian version Roman letters are sometimes used in place of (n) for the first few: so we have f'', f''', f^{iv}, f^{v}, etc.

3.6. Example. Calculate the first five derivatives of $f(x) = x^3 + x + \sin x$.

$$\begin{aligned} f'(x) &= 3x^2 + 1 + \cos x \\ f''(x) &= 6x - \sin x \\ f^{(3)} &= 6 - \cos x \\ f^{(4)} &= \sin x \\ f^{(5)} &= \cos x \end{aligned}$$

So there is nothing mysterious. You know how to differentiate, and what you can do once you can do repeatedly.

3.2.1 The meaning of the second derivative

The most common every day use is in the notion of **acceleration**.

$$\text{acceleration} = \text{rate of change of velocity with time} = \frac{dv}{dt}$$
$$\text{velocity} = \text{rate of change of distance with time} = \frac{dx}{dt}$$

Substituting

$$\text{acceleration} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}$$

So acceleration is the second derivative of distance with respect to time.

3.7. Example. The thrown ball: Here we had $x = 32t$, $y = 32t - 16t^2$.

$$\frac{d^2x}{dt^2} = 0\,; \qquad \frac{d^2y}{dt^2} = -32$$

What this tells us that is that there is no horizontal acceleration ($x'' = 0$), and that the vertical acceleration is a constant -32 ft/sec^2.

This latter is one of Newton's laws of motion.

3.8. Example. The rocket problem: Here we had $h = 25t^2$. Again the acceleration is constant, but this time it is positive as a result of the thrust from the engine.

The technique of implicit differentiation can be used to find higher derivatives.

3.9. Example. A curve is given by the equation

$$16x^2 = (x-1)^2(x^2+y^2)$$

Find $\dfrac{dy}{dx}$ and $\dfrac{d^2y}{dx^2}$ at the point $(3, 3\sqrt{3})$.

Check the point is on the curve!

Differentiate straight through the equation with respect to x

$$32x = 2(x-1)(x^2+y^2) + (x-1)^2\left(2x + 2y\frac{dy}{dx}\right) \qquad (*)$$

That was a combination of the product rule and the chain rule.

Put in the values $x = 3$, $y = 3\sqrt{3}$.

$$32.3 = 2.2.36 + 4\left(6 + 6\sqrt{3}\frac{dy}{dx}\right)$$

Therefore

$$96 = 144 + 24 + 24\sqrt{3}\frac{dy}{dx}$$

Therefore

$$-72 = 24\sqrt{3}\frac{dy}{dx}$$

To get the second derivative just go back to equation $(*)$ and differentiate straight through once again. The result will be an equation involving x, y, y' and y''. Plugging the given and calculated values in to this will give y''.

First cancel the 2 which appears throughout $(*)$.

$$16x = (x-1)(x^2+y^2) + (x-1)^2\left(x + y\frac{dy}{dx}\right) \qquad (*)$$

Then differentiate both sides of the equation to get

$$16 = (x^2+y^2) + (x-1)\left(2x + 2y\frac{dy}{dx}\right) + 2(x-1)\left(x + y\frac{dy}{dx}\right) + (x-1)^2\left(1 + \left(\frac{dy}{dx}\right)^2 + y\frac{d^2y}{dx^2}\right)$$

Into this you can now put $x = 3$, $y = 3\sqrt{3}$ and $y' = -\sqrt{3}$ and can solve for y'' at the point in question.

It is not pretty, but with a complicated equation you can't expect things to be pretty. The point is that we could cope, and all we needed to cope was care and the ability to avoid panic.

3.3 Parametric Differentiation

This is a technique for finding derivatives when, instead of a link between x and y, you have both x and y expressed in terms of a parameter.

It is a situation that often occurs with complicated curves.

3.10. Example. A curve plotting device traces out a curve which is described parametrically by means of the equations

$$x = \cos^3 t\,, \qquad y = \sin^3 t$$

What is the equation of the tangent at the point where $t = \dfrac{\pi}{3}$?

Solution We need to find $\dfrac{dy}{dx}$.

The option of getting y as a function of x is not attractive. You can do it, but the result will be a mess, and not the sort of function you relish differentiating.

The best way is to use the **chain rule**.

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

And so

$$\frac{dy}{dx} = \frac{dy}{dt} \quad \text{divided by} \quad \frac{dx}{dt}$$

In general, if f is a function of x and x a function of t we have

$$\frac{df}{dx} = \frac{\frac{df}{dt}}{\frac{dx}{dt}}$$

Finding the derivatives of x and y with respect to t is easy enough.

$x = \cos^3 t$ and so $\dfrac{dx}{dt} = 3\cos^2 t(-\sin t)$ (chain rule).

$y = \sin^3 t$ and so $\dfrac{dy}{dt} = 3\sin^2 t(\cos t)$.

Therefore

$$\frac{dy}{dx} = \frac{3\sin^2 t\cos t}{3\cos^2 t(-\sin t)} = -\tan t$$

This is typical. The derivative will be a function of the parameter.

When $t = \pi/3$, $\tan t = \sqrt{3}$, and so the gradient of the tangent at this point is $-\sqrt{3}$.

We also have that

$$\text{When } t = \frac{\pi}{3}, \quad y = \frac{3\sqrt{3}}{8}, \quad x = \frac{1}{8}$$

So the tangent has equation

$$\left(y - \frac{3\sqrt{3}}{8}\right) = -\sqrt{3}\left(x - \frac{1}{8}\right)$$

3.11. Example. The **cardioid** is a curve which can be described by the equations

$$x = 2\cos t + \cos 2t\,, \qquad y = 2\sin t + \sin 2t$$

Find the tangent at the point where $t = \pi/4$.

The chain rule enables you to calculate higher derivatives for these parametrically described curves.

3.12. Example. A point moving round in a circle of radius r with constant angular velocity has x and y coordinates given by

$$x = r\cos\omega t\,, \qquad y = r\sin\omega t$$

Find $\dfrac{dy}{dx}$ and $\dfrac{d^2y}{dx^2}$ in terms of t.

Solution First get the derivatives of y and x with respect to t.

$$\frac{dy}{dt} = r\omega\cos\omega t\,, \quad \frac{dx}{dt} = -r\omega\sin\omega t\,, \quad \text{and so} \quad \frac{dy}{dx} = -\cot\omega t$$

For the second derivative use the chain rule

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{d}{dx}\left(\frac{dy}{dx}\right)\\ &= \frac{d}{dx}(-\cot\omega t)\end{aligned}$$

Replace f by $-\cot\omega t$ in our earlier formula

$$\frac{d}{dt}(-\cot\omega t) = \frac{d}{dx}(-\cot\omega t)\frac{dx}{dt}$$

Therefore

$$\omega\,\mathrm{cosec}^2\,\omega t = \frac{d}{dx}(-\cot\omega t)(-r\omega\sin\omega t)$$

Therefore

$$\frac{d^2y}{dx^2} = -\frac{1}{r}\,\mathrm{cosec}^3\,\omega t$$

Chapter 4

More Functions

4.1 Introduction

This week we complete our collection of basic functions. We shall look at their properties, shall compute their derivatives and shall look at the sort of problems that make them necessary.

4.1.1 The inverse trig functions

4.1. Q. In the rocket launch problem that we looked at briefly in example 4 in the section on rates of change we had the following picture.

This produced the formula

$$\tan\theta = \frac{h}{100} = \frac{t^2}{4}$$

The question then was to do with how quickly θ was changing relative to t, and the fact that we didn't have θ as a function of t was handled by means of implicit differentiation. However, there are other related questions that one could ask, such as

4.1. Example. What is θ when $t = 6$?

Solution In words it is "the angle whose tangent is 9", more precisely it is "the angle between 0 and $\pi/2$ whose tangent is 9".

This is an answer of sorts, but it is not really the answer we were looking for. What we would like is an answer in radians or degrees. How are we to get it?

In such cases we can't get a proper answer to the question using just our existing repertoire of functions. So if you want to be able to answer such questions you need some new functions, functions whose properties you can analyse, whose values you can compute and which will deliver answers of the type you are looking for.

The functions that enable one to get the answers sought are the **inverse trig functions**.

The word "inverse" is used not in the sense of "one over" but in the sense of reversing the processes carried out by the original.

Remember the "black box" picture we had of a function.

x was the input to the box and $f(x)$ the output. What the inverse function for f does is reverse the process. It takes $f(x)$ as input and turns it back into x.

The natural way to start trying to do this is to begin with the graph of the original. To use this to find $f(b)$ for a given b, you locate b on the x-axis, move from there up to the curve and then from there to the y-axis. The picture is

To reverse the process you are trying to get from $d = f(c)$ back to c. So locate d on the y-axis, move to the curve and then from there to the x-axis.

However, when you try this with the sine curve you realise there is a snag. When you try to move to the curve from the y-axis you find that you have a choice. Functions aren't allowed to make choices; they have to be consistent in operation. So the choice has to be made when the function is defined, and that is why we set

$$\arcsin(x) = \text{the angle between } -\frac{\pi}{2} \text{ and } \frac{\pi}{2} \text{ whose sine is } x$$

Without the range specification the definition would be ambiguous. So we make a choice and then stick to it.

Similar choices and definitions are made for arccos and arctan. For arccos the range is $(0, \pi)$ and for arctan it is $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Alternative Notation: The functions arcsin, arccos and arctan are often called $\sin^{-1}$, $\cos^{-1}$ and $\tan^{-1}$, respectively. It is a confusing piece of notation, but it is long established and so you need to be aware of it. Your calculators probably use it.

4.1.2 Differentiating the inverse trig functions

We shall begin with the function arctan, because this is the one with fewest complications.

Arctan:

Suppose that

$$y = \arctan x$$

Then, by definition

$$\tan y = x$$

Differentiating both sides we get

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

And so

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

This gives us an expression for $\frac{dy}{dx}$, but not in a form we can use very easily. y is a function of x. So we want y' as a function of x also. What we have got is y' as a function of y. So we need to carry the calculation further.

$$\begin{aligned} \sec^2 y &= 1 + \tan^2 y \\ &= 1 + x^2 \end{aligned}$$

Substituting this into our formula

$$\frac{d}{dx}(\arctan x) = \frac{1}{1+x^2}$$

This can now be added to our list of standard derivatives.

Arcsin:

Suppose that

$$y = \arcsin x$$

Then, by definition

$$\sin y = x$$

Differentiating both sides we get

$$\cos y \frac{dy}{dx} = 1$$

And so

$$\frac{dy}{dx} = \frac{1}{\cos y}$$

Again we have to get from this to an expression in x rather than y.

$$\begin{aligned} \cos^2 y &= 1 - \sin^2 y \\ &= 1 - x^2 \end{aligned}$$

Therefore

$$\cos y = \pm\sqrt{1-x^2}$$

Now we have to make a decision. Only one of the plus and minus sign can be right. To see which is, look at the graph of $\arcsin x$. The slope is always positive, and so it is the plus sign we want here. Therefore

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

The argument for arccos parallels that for arcsin, except for the end where this time the graph makes it clear that we are to choose the minus sign. So

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

4.1.3 Log and Exp

4.2. Example. When sound passes through a glass window its intensity is reduced according to the formula

$$I_0 - I_1 = I_0\left(1 - 10^{-k}\right)$$

where I_0 and I_1 are the intensities outside and inside respectively, and k is some constant which depends on the particular window. Suppose that the outside noise (loud street traffic) is measured at 6.5×10^{-5} Wm^{-2} and that we want this reduced to a conversational level 5.0×10^{-6} Wm^{-2}. What should k be?

Solution Rearranging the formula gives

$$1 - 10^{-k} = \frac{I_0 - I_1}{I_0}$$

Therefore

$$\begin{aligned} 10^{-k} &= 1 - \frac{I_0 - I_1}{I_0} \\ &= \frac{I_1}{I_0} \end{aligned}$$

So

$$\begin{aligned} 10^{k} &= \frac{I_0}{I_1} \\ &= \frac{6.5 \times 10^{-5}}{5.0 \times 10^{-6}} \\ &= 13 \end{aligned}$$

So we want that value of k for $10^k = 13$.

This is another inverse function game. This time we need an inverse function for the function defined by $f(x) = 10^x$.

The name given to the inverse function in this case is called **logarithm** (to the base 10).

Any $a > 1$ can be used as a base for logarithms. However, only three values of a are important in practice:

1. $a = 10$ used in the measurement of sound and in the measurement of earthquake intensities.

2. $a = 2$ used in computer timings (2 because computers do their arithmetic in base 2).

3. $a = e$ — adopted because it gives the logarithm with the nicest general properties.

All the log functions have nice properties in connection with multiplication. A list is given at the start of the handout.

The graph of $y = e^x$ is

4.1.4 The derivative of log

To get this you need to go back to the Newton quotient, and the argument is not particularly easy. It is sketched in the handout. The result is that

$$\frac{d}{dx}\left(log_a x\right) = \frac{M(a)}{x}$$

where $M(a)$ is some constant depending on a.

The so-called **Natural Logarithms** are those for which this constant is 1. The corresponding value of a is the number e given in the handout. This is the only log function we shall use in the course. The notation for it is ln.

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

Learn this: it is important.

The graph of $y = \log_e x$ is important. Note it is only defined for $x > 0$. This corresponds to the fact that $e^x > 0$ always (see above graph).

4.1.5 The Exponential Function

Let e be the number we introduced in connection with the natural logarithms. The **exponential function** is defined by

$$exp(x) = e^x$$

Some useful properties of exp come from the standard laws of indices.

$$e^x e^y = e^{x+y}\,, \quad e^{-x} = \frac{1}{e^x}\,, \quad e^0 = 1$$

And so

$$\exp x \exp y = \exp(x+y)\,, \quad \exp(-x) = \frac{1}{\exp x}\,, \quad \exp 0 = 1$$

It follows from the definitions that exp is the inverse function for ln and that ln is the inverse function for exp. In other words

$$\ln(\exp(x)) = x\,, \qquad \exp(\ln(x))$$

So applying one function and then the other gets you back to where you started.

4.1.6 The derivative

We can get this from the derivative of ln by using the standard method for dealing with inverse functions.

If

$$y = \exp x$$

then

$$x = \ln y$$

Differentiating with respect to x

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

So

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

Substituting for y

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

4.1.7 The scientific importance of exp

Whenever we have a situation concerning growth or decay exp is likely to put in an appearance. The reason for this is to be found in the relation between the function and its derivative. It is common in growth/decay situations for the rate of change to be proportional to the current state. In mathematical terms this means a derivative which is a constant times the function being differentiated, in other words $y' = ky$, and the function with that property is the exponential.

4.3. Example. Consider a population of animals or plants. Without predators to keep things in check, the colony will grow at a rate proportional to its current size — the rate of proportionality k being dependent on the birth and death rate. So we have

$$\frac{dP}{dt} = kP$$

where $P = P(t)$ is the size of the population at time t.

4.4. Example. (Newton's Law of Cooling)

A body heats up or cools down at a rate proportional to the temperature difference between it and its surroundings. So

$$\frac{d\theta}{dt} = k\theta$$

where θ is the temperature difference and k a constant.

4.5. Example. Substances dissolve (e.g. sugar in water) at a rate proportional to the current mass of undissolved material. So we have the same sort of law as in the other two.

4.6. Example. Radioactive decay follows the same pattern.

With radioactive substances you speak of the "half-life" of the substance, and what this tells you is how long it will take the substance to decay to half its present mass. So it is telling you what k is.

The only function which fits this $y' = ky$ pattern is $y = Ae^{kt}$ where A and k are constants. A is the initial state and k the constant of proportionality.

A nice feature of this type of situation is that only two readings are necessary to enable you to calculate A and k and thereby get the complete behaviour pattern.

4.7. Example. A steel ingot is cooling down to room temperature (which is presumed constant at 15°C). After 3 hours its temperature is 150°C and after 10 it is 30°C. What is the temperature as a function of time?

Solution Let the temperature difference at time t be θ. Then

$$\theta = Ae^{kt} \qquad \text{for some } A \text{ and } k$$

From the given data

$$135 = Ae^{3k} \tag{1}$$
$$15 = Ae^{10k} \tag{2}$$

Dividing

$$9 = e^{-7k}$$

Therefore

$$\ln 9 = -7k$$

Therefore

$$k = -\frac{1}{7}\ln 9$$

This gives k, and putting the value into (1) will give A.

4.1.8 The hyperbolic trig functions

These are certain combinations of e^x and e^{-x} which occur often enough to justify giving them names of their own.

4.8. Definition.

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2}, \qquad \tanh x = \frac{\cosh x}{\sinh x}$$

The justification for giving them similar names to the trig functions is that they have similar properties. In particular:

$$\frac{d}{dx}(\sinh x) = \cosh x$$
$$\frac{d}{dx}(\cosh x) = \sinh x$$
$$\frac{d}{dx}(\tanh x) = \frac{1}{\cosh^2 x}$$
$$\cosh^2 x - \sinh^2 x = 1$$

Chapter 5

Maxima and Minima

5.1 Introduction

This is another batch of applications. Once again we are faced with a situation which is changing, but instead of asking how quickly, the question is going to be how big or small a quantity is going to get. Typical situations are minimum costs, maximum stresses, and so on.

5.1. Definition. Let $y = f(x)$. A **critical point** (sometimes called a **stationary point**) is a point on the curve $y = f(x)$ where $\dfrac{dy}{dx} = 0$.

$\dfrac{dy}{dx} = 0$ precisely when the tangent to the curve is horizontal.

There are three possibilities:

1. The shape of the curve near the critical point P is

   Here the slope is positive to the left of the critical point and negative to the right. Such a point is called a **local maximum**, because the curve has reached its highest point — at least its highest in this locality.

2. The shape of the curve near the critical point P is

   Here the slope is negative to the left of the critical point and positive to the right. Such a point is called a **local minimum**.

3. The shape of the curve near the critical point P is

   Here the slope has the same sign on both sides of the critical point. Such points are called **points of inflexion**. They are important geometrically, but won't concern us much in this course.

5.1.1 Finding critical points and determining their nature

Finding them is easy. Calculate the derivative, equate it to zero and solve the resulting equation.

There are two main methods of determining the nature of a critical point:

1. Look at the sign of the slope on either side of the critical point. This always works and is not usually difficult.

2. The second derivative test. In the right circumstances this can be very fast. The drawbacks are that sometimes it can be messy to calculate the second derivative, and there are times when the test is inconclusive. The test says

   (a) If $\dfrac{d^2y}{dx^2}$ is positive, then $\dfrac{dy}{dx}$ is increasing. To increase through a value where it is zero it must be going from negative to positive, and if we check that against our list we see that we have a local minimum. So $y'' > 0$ means that we have a local minimum.

   (b) In the same way we can see that $y'' < 0$ means that we have a local maximum.

   (c) If $y'' = 0$, the test gives no information. You could, in fact, have anything. The curves $y = x^3$, $y = x^4$, $y = 1 - x^4$ all have critical points at $x = 0$ and all have $y'' = 0$. The first is a point of inflexion, the second a local minimum and the third a local maximum. You check these for yourselves.

5.2. Example. Find and determine the nature of the critical points on the curve $y = x^2(x-1)$.

Solution Differentiating gives

$$\frac{dy}{dx} = 2x(x-1) + x^2 = x(3x-2)$$

So there are critical points at $x = 0$ and at $x = \dfrac{2}{3}$.

Method 1:

Slightly to the left of $x = 0$ we have that x and $3x - 2$ are both negative. So their product is positive. Thus y' is positive slightly to the left of $x = 0$. Slightly to the right it is negative. Therefore $x = 0$ is a local maximum. With the other critical point we are negative slightly to the left and positive slightly to the right. So this one is a local minimum.

Method 2:

The second derivative is $6x-2$. This is negative when $x = 0$ and positive when $x = 2/3$.

On this occasion the second derivative test was quicker because the second differentiation was easy and because we didn't hit the awkward case. You aren't always so lucky.

5.1.2 Global Maxima and Minima

In practical problems we are more likely to be interested in the greatest and smallest values that a quantity can take over a range rather than just in the vicinity of a point. For these it isn't quite enough to use the derivative to find the critical points, but it nearly is.

5.1. Q. Given a function $y = f(x)$ and a range $a \le x \le b$ we want to find the largest and smallest values attained by f.

The range is there because real life problems are like that. In mathematics it is O.K. to talk of all possible x from minus infinity to plus infinity, but in real situations there will be commonsense restrictions.

So we have pictures such as those shown below.

In the diagram E is the global maximum (and is an endpoint), B is the global minimum (and is not an endpoint), C is a local maximum and D a local minimum. The theorem that governs things is:

5.3. Theorem. *If f can be differentiated, the greatest and least values of f occur either at critical points or at one of the endpoints of the range.*

This leads to the following procedure:

1. Find the critical points of f.
2. Calculate f at each of these critical points and (if applicable) at the endpoints. Select the greatest and smallest values from the list you have calculated.

5.4. Example. For a belt drive the power transmitted is a function of the speed of the belt, the law being

$$P(v) = Tv - av^3$$

where T is the tension in the belt and a some constant. Find the maximum power if $T = 600$, $a = 2$ and $v \leq 12$. Is the answer different if the maximum speed is 8?

Solution First find the critical points.

$$P = 600v - 2v^3$$

And so

$$\frac{dP}{dv} = 600 - 6v^2$$

This is zero when $v = \pm 10$.

Commonsense tells us that $v \geq 0$, and so we can forget about the critical point at -10.

So we have just the one relevant critical point to worry about, the one at $x = 10$. The two endpoints are $v = 0$ and $v = 12$.

We don't hold out a lot of hope for $v = 0$, since this would indicate that the machine was switched off, but we calculate it anyway.

Next calculate P for each of these values and see which is the largest.

$$P(0) = 0\,, \qquad P(10) = 6000 - 2000 = 4000\,, \qquad P(12) = 7200 - 3456 = 3744$$

So the maximum occurs at the critical point and is 4000.

When the range is reduced so that the maximum value of v is down to 8, neither of the critical points is in range. That being the case, we just have the endpoints to worry about. The maximum this time is $P(8) = 4800 - 1024 = 3776$.

5.5. Example. A box of maximum volume is to be made from a sheet of card measuring 16 inches by 10. It is an open box and the method of construction is to cut a square from each corner and then fold.

Solution

Let x be the side of the square which is cut from each corner. Then $AB = 16 - 2x$, $CD = 10 - 2x$ and the volume, V, is given by

$$\begin{aligned} V &= (16 - 2x)(10 - 2x)x \\ &= 4x(8 - x)(5 - x) \\ &= 4(x^3 - 13x^2 + 40x) \end{aligned}$$

And so

$$\frac{dV}{dx} = 4(3x^2 - 26x + 40)$$

The critical points occur when

$$3x^2 - 26x + 40 = 0$$

i.e. when

$$\begin{aligned} x &= \frac{26 \pm \sqrt{676 - 480}}{6} \\ &= \frac{26 \pm 14}{6} \end{aligned}$$

The commonsense restrictions are $5 \geq x \geq 0$. So the only critical point in range is $x = 2$.

Now calculate V for the critical point and the two endpoints.

$$V(0) = 0\,, \qquad V(2) = 144\,, \qquad V(5) = 0$$

So the maximum value is 144, occurring when $x = 2$.

5.6. Example. The illumination at P from the light source L is given by

$$y = \frac{100 \cos \alpha}{x^2}$$

where x is the distance from L to P and α is the angle that LP makes with the vertical. The distance of P from the point on the floor directly below L is 5 feet. At what height above the ground should L be if y is to be maximised?

5.7. Example. A simple model for the flow of cars along a straight level road is

$$f(v) = \frac{v}{L + vT + \frac{v^2}{2a}}$$

where v is the speed of the cars, L the car length, a is the maximum acceleration of a car and T is the thinking time of a driver. With $L = 4$ metres, $T = .8$ seconds and $a = 7$ metres/sec^2 find the car speed which gives the maximum traffic flow.

Chapter 6

Integration

6.1 Introduction

If you are going to understand a process, it helps if you know what the process is for. And in mathematics that means knowing what problem people were trying to solve when they came up with a particular set of ideas. In the case of integration the basic problem was this:

6.1. Q. How do you calculate the area under the curve $y = f(x)$?

This is not just of geometrical interest, because in applied mathematics and physics the area under a curve often measures something.

6.1. Example. Suppose that we plot speed against time for a car journey. In the simple case there will be periods where the speed will be constant. Let t_0 to t_1 be such an interval. We then have the graph

The speed maintained over this time interval is v, and it is not difficult to see that the distance travelled is $v(t_1 - t_0)$ and that this is also the area under this section of the curve.

Now let us consider the general case of a journey in which the speed varies.

It is no longer clear how we can get an accurate figure for the distance simply by looking at the graph, but it is not difficult to see how we could get a very good estimate.

Chop the time axis up into small intervals, say of a minute each.

In each minute the speed will vary, but not by much. So the speed at the end of the minute will be a good estimate of the average speed during the minute. So it times the time gives a good estimate of the distance travelled during this minute. Do this for each minute and add up.

In area terms we have computed the sum of the areas of the shaded rectangles, and this area is clearly a reasonable approximation to the area under the graph. So area under the graph is at worst a good approximation to the distance covered.

It is also clear that if we shrink the lengths of the time intervals the approximation will get better — both to the distance covered and to the area under the curve.

So it is at least plausible that

$$\text{distance covered} = \text{area under the curve}$$

In a similar sort of way, if you plot force against distance for something like a piston you find that the area under the curve is the total work done.

So areas aren't just about geometry.

Now consider the curve given by $y = f(x)$.

We want a method of calculating the area under the curve between $x = a$ and $x = b$. Let $F(x)$ be the area bounded by the curve, the x-axis and verticals dropped from the points $(a, f(a))$ and $(x, f(x))$.

As with the speed/time curve it seems a good idea to begin with just a small section, say that between $x = x_0$ and $x = x_0 + h$.

In terms of the area function F this is $F(x_0 + h) - F(x_0)$.

On the diagram it is the shaded area.

We are supposing that h is small, and so a reasonable approximation to this area is got by taking the rectangle of height $f(x_0)$ and width h. This rectangle has area $f(x_0)h$, and so

$$\text{Shaded Area} = F(x_0 + h) - F(x_0) \approx f(x_0)h$$

And so

$$\frac{F(x_0 + h) - F(x_0)}{h} \approx f(x_0)$$

And the smaller h is the better this approximation will be.

This suggests that

$$\lim_{h \to 0} \frac{F(x_0 + h) - F(x_0)}{h} = f(x_0)$$

But the expression on the left is just the derivative of F at x_0. So

$$F'(x_0) = f(x_0)$$

We are trying to find the area under the curve. If we had an explicit expression for the function F, our problems would be over. What the link between F and f gives us is a very large clue as to what F is. We know f, and so all we have to is work out what function has it as derivative.

So to find F we start from f (which is known) and we try to reverse the differentiation process. If we succeed, we shall have F and can use it to calculate the areas we require.

It probably sounds tricky, but it needn't be. We have done enough differentiation by now to be able to recognise quite a lot of situations. If I ask you which function has derivative $\cos x$, it shouldn't take very long for you to answer $\sin x$. That is the sort of reversal problem we are setting ourselves.

This process of "anti-differentiation" is known as **integration**.

6.1.1 Indefinite Integrals

6.2. Definition. If F is a function with the property that $F' = f$, we say that F is an **indefinite integral** for f.

Some books prefer to say that F is a **primitive** for f.

Both terms mean simply that $F' = f$.

Notation: Since the starting point in the integration process is f rather than F, it is convenient to have a notation that focuses just on f.

The standard notation for "indefinite integral of f" is

$$\int f(x)\,dx$$

Terminology: When speaking of $\int f(x)\,dx$, f is referred to as the **integrand**.

"integrand" = function to be integrated

x is the **variable of integration**.

6.1.2 The Constant of Integration

The reason for the word "indefinite" is that integrals are not uniquely specified in the way that derivatives are.

For example, the functions $\sin x$, $1 + \sin x$ and $300 + \sin x$ all have derivative $\cos x$. So all are equally valid as answers to the question "What function has derivative $\cos x$?"

Therefore, all are indefinite integrals for $\cos x$.

There is nothing that can be done about this, but fortunately this as far as the arbitrariness goes.

Indefinite integrals are specified to within an arbitrary additive constant known as the **constant of integration**. Many books choose to stress this by always writing

$$\int f(x)\,dx = F(x) + c$$

rather than

$$\int f(x)\,dx = F(x)$$

and give the impression that lightning will strike if you forget to mention c.

The truth is more sensible. Sometimes this constant is important and sometimes it isn't. When it isn't nothing is lost by leaving it out. The guidelines are:

1. If you are solving a differential equation, the c is likely to be important; so include it.
2. If you are computing a "definite integral", it is going to cancel, and so you might as well miss it out.
3. If you are just doing integration practice — "learn how to integrate by doing the following page of integrals" — it is irrelevant; so please yourself. The rightness or wrongness of your answer is unaffected.

Common sense will get you through; but if in doubt, include it.

6.1.3 How to Integrate

The strategy is the same as that for differentiation:

1. We collect together a set of standard integrals.

2. We develop a set of techniques for reducing complicated integrals to these standard ones.

The standard integrals are got from the table of standard derivatives by reading that tables from right to left instead of from left to right. For example, since the derivative of sine is cosine it follows that the integral of cosine is sine.

This produces the following list:

$$\frac{d}{dx}(\sin x) = \cos x \quad \text{and so} \quad \int \cos x \, dx = \sin x$$

$$\frac{d}{dx}(\cos x) = -\sin x \quad \text{and so} \quad \int (-\sin x) \, dx = \cos x$$

which tidies up as

$$\int \sin x \, dx = -\cos x$$

$$\frac{d}{dx}(\tan x) = \sec^2 x \quad \text{and so} \quad \int \sec^2 x \, dx = \tan x$$

$$\frac{d}{dx}(x^{n+1}) = (n+1)x^n \quad \text{and so} \quad \int (n+1)x^n \, dx = x^{n+1}$$

which tidies up as

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} \quad (\text{ provided } n \neq -1)$$

$$\frac{d}{dx}(\ln x) = \frac{1}{x} \quad \text{and so} \quad \int \frac{1}{x} \, dx = \ln x \quad \text{if } x > 0$$

The restriction to $x > 0$ is because $\ln x$ is not defined when $x < 0$.

$$\int \frac{1}{x} \, dx = \ln(-x) \quad \text{if } x < 0$$

$$\frac{d}{dx}(e^x) = e^x \quad \text{and so} \quad \int e^x \, dx = e^x$$

$$\frac{d}{dx}(\arctan x) = \frac{1}{1+x^2} \quad \text{and so} \quad \int \frac{1}{1+x^2} \, dx = \arctan x$$

$$\frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}} \quad \text{and so} \quad \int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin x$$

And a few others that you can add for yourself.

There is a much longer list in section 11.3 of the handbook, but the above are the ones that are worth carrying round in your head.

The rules for getting beyond this standard list are also got from our work on differentiation. The first two are fairly obvious:

Sums:

Because $(F + G)' = F' + G'$ it follows that

$$\int (f(x) + g(x)) \, dx = \int f(x) \, dx + \int g(x) \, dx$$

Scalar Products:

Because $(\lambda F)' = \lambda F'$ when λ is a constant we have

$$\int \lambda f(x)\,dx = \lambda \int f(x)\,dx$$

6.3. Example.

$$\int (x^2 + \cos x)\,dx = \int x^2\,dx + \int \cos x\,dx = \frac{x^3}{3} + \sin x$$

6.4. Example.

$$\begin{aligned}\int \left(\frac{1}{x} + 3e^x - \sin x\right) &= \int \frac{1}{x}\,dx + \int 3e^x\,dx + \int (-\sin x)\,dx \\ &= \int \frac{1}{x}\,dx + 3\int e^x\,dx - \int \sin x\,dx \\ &= \ln x + 3e^x + \cos x\end{aligned}$$

Try it and see:

This is not a rule but a method.

If an integral looks like a minor variant of one that you could do in your head, try a "guess, test, make adjustments and retest" approach.

6.5. Example. $\int \cos 3x\,dx$

Your first thought is probably $\sin 3x$. (This is the guess.)

Test your guess by differentiating. You get $3\cos 3x$.

This is out by a factor of 3, but is close enough to enable you to see what adjustment you need to make. Making it gives you a second guess of

$$\frac{1}{3}\sin 3x$$

Testing this gives what you want. So

$$\int \cos 3x\,dx = \frac{1}{3}\sin 3x$$

N.B. When using this approach your final step should always be to check your answer by differentiating it.

6.6. Example.

$$\int \frac{dx}{1+4x^2}$$

Solution Apart from the 4, this is $\arctan x$. $4x^2 = (2x)^2$, and so our first try is $\arctan(2x)$. Differentiate this and see what you get.

6.7. Example.

$$\int \frac{dx}{\sqrt{9-x^2}}$$

Solution Again we have something which looks like a standard integral. The standard integral that we nearly have this time is arcsin. That leads to $\sqrt{1-x^2}$, rather than $\sqrt{9-x^2}$. So begin by taking the 9 outside a bracket to get $\sqrt{9(1-x^2/9)}$. That should suggest $\arcsin(x/3)$ as a first try.

6.1.4 The Definite Integral

The indefinite integral is a **function**; the definite integral is a **number** got by putting values into that function.

Let

$$F(x) = \int f(x)\,dx$$

So F is a function with the property that $F' = f$.

6.8. Definition. The definite integral of f from a to b is denoted by $\int_a^b f(x)\,dx$ and is defined to be $F(b) - F(a)$.

6.9. Definition. a and b are called the **limits of integration**.

Notation: $[F(x)]_a^b$ is used to mean $F(b) - F(a)$.

6.10. Example.

$$\int_1^2 x\,dx = \left[\frac{x^2}{2}\right]_1^2 = 2 - \frac{1}{2} = \frac{3}{2}$$

6.11. Remark. Note that the constant of integration is of no importance when you are calculating a definite integral. Include it and you get

$$(F(b) + c) - (F(a) + c) = F(b) - F(a)$$

— the same answer as you would have got by ignoring the constant in the first place.

6.12. Remark. The definite integral is the one that is used to calculate areas, volumes, lengths, moments of inertia and a host of other things.

Properties of the Definite Integral

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx \tag{1}$$

So reversing the limits of integration multiplies the answer by -1.

Proof. Let $F(x)$ be an indefinite integral for $f(x)$. Then

$$\begin{aligned}\int_b^a f(x)\,dx &= F(a) - F(b) \\ &= -(F(b) - F(a)) \\ &= -\int_a^b f(x)\,dx\end{aligned}$$

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx \tag{2}$$

Proof. $F(x)$ as before. Then

$$\begin{aligned}\int_a^b f(x)\,dx + \int_b^c f(x)\,dx &= (F(b) - F(a)) + (F(c) - F(b)) \\ &= F(c) - F(a) \\ &= \int_a^c f(x)\,dx\end{aligned}$$

Using the Definite Integral to Calculate Areas

This is slightly more complicated than one might hope, but not by much. The problem is caused by the fact that definite integrals can be either positive or negative, while areas are always positive.

Case 1: The curve is entirely above the x-axis

This is complication free.

Choose a reference point $(r, 0)$ on the x-axis, and let $F(x)$ be the area under the section of the curve between $(r, f(r))$ and $(x, f(x))$.

As we saw earlier, $F'(x) = f(x)$, and so $F(x) = \int f(x)\, dx$.

Now consider the area under the curve between $x = a$ and $x = b$.

$F(b)$ is the area between $x = r$ and $x = b$; $F(a)$ is that between $x = r$ and $x = a$.

The cross-hatched area is the difference, i.e. $F(b) - F(a)$. Therefore the area sought is $\int_a^b f(x)\, dx$.

6.13. Example. Find the area under the curve $y = \sqrt{x}$ between $x = 0$ and $x = 4$.

Solution The curve does not go below the x-axis and so

$$\begin{aligned} \text{Area} &= \int_0^4 \sqrt{x}\, dx \\ &= \left[\frac{2}{3}x^{3/2}\right]_0^4 \\ &= \frac{16}{3} \end{aligned}$$

Case 2: The curve is entirely below the x-axis

This time $y = f(x)$ is negative, and so the height of the rectangle we looked at when considering the difference between $F(x_0 + h)$ and $F(x_0)$ is not $f(x_0)$ but $-f(x_0)$ (since heights can't be negative). So $F'(x) = -f(x)$, and so *for areas under the x-axis*

$$\text{Area} = -\int_a^b f(x)\, dx$$

Case 3: The curve is part above and part below the x-axis

You have to split the problem up and consider the two parts separately. Use case 1 to deal with the section above the x-axis and case 2 to deal with the section below the x-axis.

6.14. Example. Find the area bounded by the curve $y = \sin x$, the x-axis and the lines $x = 0$ and $x = 2\pi$.

Solution Draw a picture so that you can see what is happening.

From 0 to π the curve is above the axis and from π to 2π it is below.

Above the axis we have

$$\begin{aligned} \text{Area} &= \int_0^\pi \sin x\, dx \\ &= [-\cos x]_0^\pi \\ &= -\cos \pi + \cos 0 \\ &= 2 \end{aligned}$$

And below we have

$$\begin{aligned}\text{Area} &= -\int_{\pi}^{2\pi} \sin x\,dx \\ &= [\cos x]_{\pi}^{2\pi} \\ &= \cos 2\pi - \cos \pi \\ &= 2\end{aligned}$$

Therefore the total area is 4.

Note what would have happened had we not split the area into two sections. The two parts would have cancelled, giving us an area of zero — which is clearly silly. In this case we could have noted, by symmetry, that the required area $= 2\int_0^{\pi} \sin x\,dx = 4$.

The same "assemble the bits" approach works on areas between curves.

6.15. Example. Find the area between the curve $y = \sin x$ and the line $y = \dfrac{2x}{\pi}$ and above the x-axis.

Solution Again the best way to begin is by drawing a picture.

Here $A = \left(\dfrac{\pi}{2}, 1\right)$ and lies on both $y = \sin x$ and $y = \dfrac{2x}{\pi}$.

The next step is to work out where the line and curve cross. This will give us the limits for the integrals.

In this case they cross when $x = 0$ and when $x = 1$.

From the picture we can see that the area we want is the area under the curve minus that under the line (shaded). Therefore

$$\begin{aligned}\text{Area} &= \int_0^{\frac{\pi}{2}} \sin x\,dx - \int_0^{\frac{\pi}{2}} \frac{2x}{\pi}\,dx \\ &= [-\cos x]_0^{\frac{\pi}{2}} - \left[\frac{x^2}{\pi}\right]_0^{\frac{\pi}{2}} \\ &= 1 - \frac{\pi^2}{4}\frac{1}{\pi} \\ &= 1 - \frac{\pi}{4}\end{aligned}$$

6.16. Example. Find the area between the curve $y = x^2$ and the line $y = 2$.

6.17. Example. Find the area between the curves $y = x^2$ and $y = x^4$.

Chapter 7

Techniques of Integration

7.1 Introduction

There are two main methods for dealing with integrals which are too complicated either to do in your head or by the guess and test method. They are

1. Integration by substitution

2. Integration by parts

In addition there is a trick involving algebraic manipulation called "partial fractions" which is quite often useful.

Over the next few lectures we shall be seeing how to use these techniques and how to tell which one to try.

7.2 Integration by Substitution

This is based on the chain rule.

Let f be a function of a function, e.g. $\sin(x^2+1)$ or $\exp(\tan x)$. So f is a function of u where u is a function of x.

If you write down the Leibniz version of the chain rule and do some manipulations you get the formula:

$$\int f(u)\,du = \int f(u)\,\frac{du}{dx}\,dx$$

The right hand side looks complicated; the left much less so. We get from one to the other by working in terms of the new variable u and by pretending that the dx has cancelled.

And this is what integration by substitution comes down to: change the variable you are working with and hope that this will result in a friendlier integral.

7.1. Example. Let

$$I = \int \frac{\ln x}{x}\,dx$$

It looks complicated, but watch what happens if we introduce a new variable by setting $u = \ln x$.

$$\frac{du}{dx} = \frac{1}{x} \qquad \text{and so} \qquad du = \frac{1}{x}\,dx$$

That last step of writing du in terms of dx is a piece of notational sleight of hand, but it works and is the easiest way to see what to do; so use it.

Therefore when we substitute for x and for dx we get

$$I = \int \ln x \, \frac{1}{x} \, dx = \int u \, du$$

The original integral has been transformed into one we can do.

$$\begin{aligned} I &= \int u \, du \\ &= \frac{u^2}{2} \\ &= \frac{1}{2}(\ln x)^2 \end{aligned}$$

7.2.1 The technique

1. Decide what function of x the new variable u is to be.
2. Differentiate to get $\frac{du}{dx}$ and then separate the du and the dx via $du = \frac{du}{dx}dx$.
3. Substitute for x and for dx to get a new integral entirely in terms of u and du.
4. Do the integration.
5. Substitute back from u to x so as to get an answer in terms of the original variable.

The tricky part is (1), and here the two questions you should ask are

- Does part of the function you are integrating (often called the **integrand**) look like a function of a function, i.e. are you looking at $f(g(x))$? If so then $u = g(x)$ is worth a try.
- Is part of the integrand the derivative of another part? If it is, you may well have spotted u' and can figure out u from that.

In example 1 it was question 2 that provided the clue. We had a $\ln x$ and its derivative x^{-1}. That suggested $u' = x^{-1}$ and hence $u = \ln x$.

7.2. Example. $I = \int x \sin(x^2 + 1) \, dx$

An obvious "yes" to question 1; so try $u = (x^2 + 1)$.

7.3. Example.

$$I = \int \frac{\sec^2 x}{\sqrt{1 + \tan x}} \, dx$$

"Yes" to both questions. We have a function of a function in $\sqrt{1 + \tan x}$ and also $\sec^2 x$ as the derivative of $\tan x$. Both $u = 1 + \tan x$ and $u = \tan x$ will work.

7.4. Example.

$$I = \int \frac{\cos(\ln x)}{x} \, dx$$

Again we have a "yes" to both questions, and this tells us to try $u = \ln x$.

7.2.2 How to handle definite integrals

Be clear about which variable the limits of integration refer to.

The best way to do this is to keep writing them down as a reminder to yourself.

7.5. Example.

$$I = \int_{1/2}^{1} \frac{\arcsin x}{\sqrt{1-x^2}}\,dx$$

The obvious substitution is $u = \arcsin x$. (Hint supplied by question 2.)

$$du = \frac{dx}{\sqrt{1-x^2}}$$

and so

$$\begin{aligned} I &= \int_{x=1/2}^{x=1} u\,du \\ &= \left[\frac{1}{2}u^2\right]_{x=1/2}^{x=1} \\ &= \left[\frac{1}{2}(\arcsin x)^2\right]_{x=1/2}^{x=1} \\ &= \frac{1}{2}\left(\frac{\pi}{2}\right)^2 - \frac{1}{2}\left(\frac{\pi}{6}\right)^2 \\ &= \frac{\pi^2}{8} - \frac{\pi^2}{72} \\ &= \frac{\pi^2(9-1)}{72} \\ &= \frac{\pi^2}{9} \end{aligned}$$

7.3 Integration by Parts

This is based on the rule for differentiating products. There are two versions of it. Choose one and stick with it.

Don't try to use both; it will lead to confusion if you do.

Version 1 (Evans, Stroud and CALM)

Let u and v be functions of x. If you differentiate the product uv, rearrange and then integrate you get the formula

$$\int u\frac{dv}{dx}\,dx = uv - \int v\frac{du}{dx}\,dx$$

To use this formula you take the function you are being asked to integrate and split it up as a product of something you are going to think of as u and something you are going to think of as v'. For example if the integrand were $x\cos x$, you would take $u = x$ and $v' = \cos x$. Putting these into the right hand side, and noting that if $v' = \cos x$ $v = \sin x$ you get that

$$\int x\cos x\,dx = x\sin x - \int \sin x\,dx = x\sin x + \cos x$$

If you have met this version before and are happy with it then stick with it. If not then skip it and use version 2 instead.

Version 2

This time we label things slightly differently, do the manipulation and get

$$\int u(x)v(x)\,dx = v(x)\int u(x)\,dx - \int\left(\int u(x)\,dx\right)\frac{dv}{dx}\,dx$$

As a formula it is nothing like as pretty, but I find it easier to use — largely because it begins with the integral you are actually facing, rather than with a v *which exists only in your head.* Don't remember the formula; instead remember the recipe in words.

Break the integrand up as a product u times v. Then the integral of uv is v times the integral of u minus the integral of (the integral of u times the derivative of v).

So we have two sections to the integrand and are going to integrate one and differentiate the other. The gain is that many functions simplify on differentiating.

With the example from above we have $uv = x\cos x$. x is the part that will simplify on differentiating. So we take this as v and the other part ($\cos x$) as u. Then $v' = 1$, $\int u\,dx = \sin x$ and so when we put these into the formula we have

$$\int x\cos x\,dx = x\sin x - \int 1.\sin x\,dx = x\sin x + \cos x$$

Essentially the same calculation as before, but a slightly different way of thinking about it. The gains are a more natural start position and a process which breaks up into simple steps.

7.6. Example. $I = \int x\ln x\,dx$

Solution The integrand is clearly a product. We are going to have to integrate one bit and differentiate the other. We know how to differentiate $\ln x$ but don't know how to integrate it. So this is clearly the best choice for v, leaving $u = x$.

$$\frac{dv}{dx} = \frac{1}{x} \qquad \text{and} \qquad \int u(x)\,dx = \frac{x^2}{2}$$

Therefore

$$I = \frac{x^2}{2}.\ln x - \int \frac{x^2}{2}\frac{1}{x}\,dx = \frac{x^2}{2}.\ln x - \int\frac{x}{2}\,dx = \frac{x^2}{2}\ln x - \frac{x^2}{4}$$

The point is that the routine got me from the initial integral that I couldn't see how to do to the easy $\int \frac{x}{2}\,dx$.

7.3.1 Guidelines on the choice of u and v

1. Whatever we choose for u is going to have to be integrated at the start in order to get the routine moving. We don't want this first stage to be hard. So ask "What is the largest chunk of the integrand that I can integrate in my head?" If there is an obvious answer, try taking this chunk for u and letting v be the rest.

2. If (1) doesn't give a clear lead, switch your attention to v and ask "Is there part of the integrand which will simplify on differentiating?" If so, you have a possible candidate for v.

Functions which simplify on differentiating

x^n for $n \geq 1$, $\ln x$, $\arctan x$, $\arcsin x$.

7.7. Example. $I = \int x \sin x \, dx$

Solution Both x and $\sin x$ can be integrated in your head. However, x simplifies on differentiating, while $\sin x$ does not. So take $u = \sin x$ and $v = x$.

7.8. Example. $I = \int x \arctan x \, dx$

Solution Although both components are on the "simplify on differentiating" list, x is something we can integrate in our heads, while $\arctan x$ is not. So take $u = x$ and $v = \arctan x$. Then

$$\int u(x)\, dx = \frac{x^2}{2} \qquad \text{and} \qquad \frac{dv}{dx} = \frac{1}{1+x^2}$$

Therefore

$$I = \frac{x^2}{2} \arctan x - \frac{1}{2} \int \frac{x^2}{1+x^2}\, dx$$

Not home yet, but we have made progress.

$$\int \frac{x^2}{1+x^2}\, dx = \int \frac{(1+x^2)-1}{1+x^2}\, dx = \int \left(1 - \frac{1}{1+x^2}\right) dx = \int dx - \int \frac{dx}{1+x^2}$$

7.9. Example. $I = \int \sin^3 x \, dx$

Solution Apply the criterion. The largest part of the integrand that you can integrate in your head is $\sin x$. So take this as u, the rest as v and trust to providence. The result is

$$I = -\cos x \sin^2 x + 2 \int \sin x \cos^2 x \, dx$$

Again there is still work to do before we get an answer.

However, if you put $u = \cos x$ into the righthand integral, nice things happen.

7.10. Example. $I = \int \arctan x \, dx$

Solution Another hard one. However, $\arctan x$ is still a function we'd be happier differentiating. So think of this as v. If this is v and the whole thing is uv, then u must be 1.

Try it and see what happens.

Note that integration by parts doesn't always get you all the way to the answer. All it hopes to achieve is to give you an easier integral than the one you started with. How you deal with this easier integral is your problem.

7.4 Partial Fractions

This is an algebraic technique used on integrals of the form $\int \frac{f(x)}{g(x)}\,dx$ where $f(x)$ and $g(x)$ are polynomials.

In case you have forgotten, a **polynomial** is a function of the form $a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$. where the a_i are numbers and x is a variable.

So $x^3 + x - 4$, $2x^4 + 1$ and $3x^2 + 2x - 5$ are all polynomials.

The idea is to put the polynomial fraction into a form which makes the integration easier to manage.

The **degree** of a polynomial is the highest power of x that occurs. So the examples listed have degrees 3, 4 and 2 respectively.

7.11. Example.

$$I = \int \frac{dx}{(x-1)(x-2)}\,dx$$

Solution Observe that

$$\begin{aligned}\frac{1}{x-2} - \frac{1}{x-1} &= \frac{(x-1)-(x-2)}{(x-1)(x-2)}\\ &= \frac{1}{(x-1)(x-2)}\end{aligned}$$

And so

$$\int \frac{dx}{x-2} - \int \frac{dx}{x-1} = \int \frac{dx}{(x-1)(x-2)}$$

On the right we have the original integral — which we didn't know how to do as it stood — and on the left we have two integrals which are close enough to one on the standard list for us to be able to write down the answer.

Let $u = x - 2$. Then du=dx and the first integral from the left hand side turns into $\frac{du}{u}$, which is $\ln u$ (provided $u > 0$). So the first integral on the left is $\ln(x-2)$ (provided $x > 2$). Likewise, the second is $\ln(x-1)$ (provided $x > 1$). So

$$\int \frac{dx}{(x-1)(x-2)}\,dx = \ln(x-2) - \ln(x-1) = \ln\left(\frac{x-2}{x-1}\right)$$

(provided $x > 2$).

By "expanding the fraction" we have turned the integral into one we know how to do.

Partial Fractions is a method for expanding fractions.

7.12. Example. Expand the fraction $\frac{1}{(x-1)(x-3)}$.

Solution The expansion will take the form

$$\frac{A}{x-1} + \frac{B}{x-3}$$

for some numbers A and B.

This is obvious when you think about it. The process of combining fractions consists of putting them over a common denominator and then tidying up. The common denominator in this case has turned out to be $(x-1)(x-3)$. So the components that went into it must have been $x-1$ and $x-3$.

What we need to do now is find A and B.

To do this take the expression

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

and clear fractions. This gives

$$1 = A(x-3) + B(x-1)$$

The quickest way to proceed for here is to note that this expression has to be true for *all* values of x. That being so, it is true when $x = 1$ and when $x = 3$.

Putting $x = 1$ gives the equation $1 = (-2)A$, and so $A = -\frac{1}{2}$.

Putting $x = 3$ gives the equation $1 = 2B$, and so $B = \frac{1}{2}$.

So these must be the values of A and B, and we have the expression

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

If we wished we could now go ahead and find the integral, since the routine has converted it into the sum of two integrals each of which we can do easily.

7.4.1 The Partial Fractions Routine

The main routine will work on an integrand of the form $\frac{f(x)}{g(x)}$ where $f(x)$ and $g(x)$ are polynomials and where the degree of $f(x)$ is less than the degree of $g(x)$. So $\frac{x^2+1}{(x-1)(x^2-4)}$ is OK, since we have degree 2 on the top and degree 3 on the bottom. However, $\frac{x^3+1}{(x-1)(x^2-4)}$ is not OK, since now we have degree 3 top and bottom.

Step 1:

This is only necessary when the condition on the degrees of f and g is not met, i.e. when the degree of f is greater than or equal to that of g.

It won't happen in questions I set, but you might need it in integrals which arise in other courses.

If the degree of f is not less than that of g, divide $f(x)$ by $g(x)$ to get $f(x) = g(x)h(x) + r(x)$ where $h(x)$ is the quotient and $r(x)$ the remainder. Then divide through to get

$$\frac{f(x)}{g(x)} = h(x) + \frac{r(x)}{g(x)}$$

$h(x)$ is a simple polynomial and so will be easy to integrate. The second component, $\frac{r(x)}{g(x)}$ meets the condition about degrees and so can be carried forward to the main routine.

7.13. Example. $\int \frac{x^2}{x^2+1}$ is an integral we have already met and which needs this preliminary treatment.

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

So

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

Therefore

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

Step 2:

Factorise $g(x)$ as a product of linear and quadratic factors. This can always be done. *Again I shan't ask you to do this, except in very simple cases such as the break up of* (x^2-a^2) *as* $(x-a)(x+a)$.

Step 3:

Here again, I'll spare you the general case. Look at the factorisation of $g(x)$ and proceed as follows:

1. For each linear factor which occurs only to the first power write down a term $\frac{A}{x-a}$.
2. For each linear factor which occurs to the power 2 write down a term $\frac{B}{x-b} + \frac{C}{(x-b)^2}$.
3. For each quadratic factor which occurs only to the first power write down a term $\frac{Dx+E}{x^2+cx+d}$.

It can get more complicated than this, but it won't in this course.

So, in the example we did earlier, $g(x)$ was $(x-1)(x-3)$. This is two linear factors, each to the power 1. $(x-1)$ leads to the term $\frac{A}{x-1}$ and $(x-3)$ to the term $\frac{B}{x-3}$.

7.14. Example.

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

This time $g(x)$ was the product of a linear factor to the power one, another linear factor to the power two and a quadratic factor. We just worked through systematically, writing down an appropriate term for each.

Clearly, we have to move through the alphabet as we do so, since there is no reason why the various constants $A, B, \ldots$ *should be equal.*

Step 4:

Having written down the appropriate break up, clear fractions by multiplying through on both sides by $g(x)$.

The job then is to calculate $A, B, \ldots$.

There are two techniques you can use here:

1. putting in special values for x,
2. comparing coefficients for the various powers of x.

The first of these will not always get you all the way to the answer, but it will get you off to a fast start. So begin with (1). The values of x to try are those from the linear factors in $g(x)$.

7.15. Example. $\dfrac{x+1}{(x-1)^2(x-2)}$

Our rules say that

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

for some constants A, B, C.

Clearing fractions gives us

$$x+1 = A(x-1)(x-2) + B(x-2) + C(x-1)^2$$

This is true for all x.

Putting $x = 1$ gives $2 = B(-1)$, and so $B = -2$.

Putting $x = 2$ gives $3 = C(1)^2$, and so $C = 3$.

That is the end of the fast progress from putting in special values. To complete the job compare the coefficients of some suitable power of x.

On the lefthand side the coefficient of x^2 is zero; on the right it is $A + C$. These must be equal, and so we must have $A + C = 0$. We know that $C = 3$, and so conclude that $A = -3$. Therefore we have

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

7.16. Example. $\dfrac{x^2+x+4}{(x-2)(x^2+2x+2)}$ This time our rules give us the starting point

$$\frac{x^2+x+4}{x-2)(x^2+2x+2)} = \frac{A}{x-2} + \frac{Bx+C}{x^2+2x+2}$$

Clearing fractions we get

$$x^2+x+4 = A(x^2+2x+2) + (Bx+C)(x-2)$$

Putting $x = 2$ gives $10 = 10A$, and so $A = 1$.

For the rest we equate coefficients.

Coefficient of x^2: $1 = A + B$. So $B = 0$.

Constant term: $4 = 2A - 2C$. So $C = -1$. Therefore

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

7.4.2 The Integration Stage

All you ever handle at the integration stage in a partial fractions routine are:

1. Terms such as $\dfrac{k}{x-a}$ where k and a are constants. These integrate as logs.

$$\begin{aligned}\int \frac{k}{x-a}\,dx &= k\ln(x-a) \quad \text{if } x>a \\ &= k\ln(a-x) \quad \text{if } x<a\end{aligned}$$

   You need to be aware of the two possibilities, because you will often be dealing with a definite integral, and there the range of the function really does matter.

2. Terms such as $\dfrac{k}{(x-a)^2}$ where k and a are constants. Here there is no complication at all.

$$\int \frac{k}{(x-a)^2}\,dx = -\frac{k}{x-a}$$

   With terms which have a quadratic on the bottom line, the best first step is to "complete the square", turning the quadratic into something of the form $(x+c)^2+d$. d will be positive if you have carried the factorisation of $g(x)$ as far as you should have. There are then two cases.

3. Terms such as $\dfrac{k}{(x+c)^2+d}$. Putting $u=(x+c)$ shows that these integrate as arctan of something (see your handbook).

4. Terms such as $\dfrac{x+k}{(x+c)^2+d}$. Here you have to be slightly clever. Begin by setting $u=(x+c)$. (So $x=u-c$.) This turns the expression into something of the form $\dfrac{u+m}{u^2+d}$, and this you break up further as $\dfrac{u}{u^2+d}+\dfrac{m}{u^2+d}$. The first integrates as $\dfrac{1}{2}\ln(u^2+d)$ and the second as an arctan.

7.17. Example. $I=\displaystyle\int_2^3 \frac{2x}{(x^2+1)(x+1)}\,dx$

Solution

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

Therefore

$$\begin{aligned}
I &= \int_2^3 \frac{x+1}{x^2+1}\,dx - \int_2^3 \frac{1}{x+1}\,dx \\
&= \int_2^3 \frac{x}{x^2+1}\,dx \int_2^3 \frac{1}{x^2+1}\,dx - \int_2^3 \frac{1}{x+1}\,dx \\
&= \left[\frac{1}{2}\ln(x^2+1)\right]_2^3 + [\arctan x]_2^3 - [\ln(x+1)]_2^3 \\
&= \frac{1}{2}\ln(10) - \frac{1}{2}\ln 5 + \arctan 3 - \arctan 2 - \ln 4 + \ln 3 \\
&= \ln\left(\frac{3\sqrt{2}}{4}\right) + \arctan 3 - \arctan 2
\end{aligned}$$

Chapter 8

Applications of Integration

8.1 Introduction

We have already used the definite integral to calculate areas bounded by curves. In this set of lectures we are going to look at four more applications:

1. The calculation of certain volumes
2. The calculation of the lengths of curves
3. The location of centres of gravity of planar regions
4. The calculation of the surface areas of certain solids.

8.2 Volumes of Revolution

To calculate general volumes you need a more general version of calculus, one that handles functions of more than one variable. For that you must wait until next year.

8.1. Q. An area is rotated about a line. Calculate the volume of the solid thus created.

I have not been subjecting you to proofs so far, because, as engineers, they aren't something you are likely to need in the future. However, this time I am going to give you the derivation of the formula we shall be using. And the reason for this is that the way of thinking that produces the formula is, in the long term, more important than the formula itself. For the purposes of the exam questions you will face the formula is sufficient; but for the problems you will come across outside this course, the thought process is more valuable.

8.2.1 The Basic Case

The section of the curve $y = f(x)$ between $x = a$ and $x = b$ is rotated about the x-axis.

The idea is to look at a small slice, δV, of the volume, lying between x and $x + \delta x$.

We calculate δV in terms of x and δx and then use integration to add up the slices and thereby get the whole volume.

The small slice is a disc of radius $y = f(x)$ and width δx. (This is not quite exact, but it is a very good approximation.)

Therefore, its volume, $\delta V = \pi y^2 \delta x$.

Dividing through by δx gives

$$\frac{\delta V}{\delta x} = \pi y^2$$

And as δx tends to zero this turns into

$$\frac{dV}{dx} = \pi y^2$$

And when we integrate both sides of this with respect to x we have

$$V = \int_a^b \pi y^2\,dx$$

8.1. Example. Find the volume that results when the portion of the curve $y^2 = 8x$ between $x = 0$ and $x = 2$ is rotated about the x-axis.

The curve is as shown.

$$V = \int_0^2 \pi y^2\,dx = \int_0^2 8\pi x\,dx = \left[4\pi x^2\right]_0^2 = 16\pi$$

The formula only works for rotations about the x-axis, but the thinking that produced it allows you to handle rotations about any line you choose.

8.2. Example. Same piece of curve, but this time it is rotated about the line $x = 2$.

Solution Draw the picture showing a cross section taken down the axis of rotation and consider a small slice at right angles to this axis. Calculate its volume and then integrate.

This time the small volume is a disc of radius $(2 - x)$ and thickness δy.

So

$$\delta V = \pi(2 - x)^2\,dy$$

Therefore

$$\frac{\delta V}{\delta y} = \pi(2 - x)^2$$

Therefore

$$\frac{dV}{dy} = \pi(2 - x)^2$$

Therefore

$$V = \int \pi(2 - x)^2\,dy$$

Now substitute for x so that the integrand is a function of y (necessary if we are going to integrate with respect to y) and work out the limits of integration. $x = \dfrac{y^2}{8}$ and the limits run from $y = -4$ to $y = 4$ (see diagram). Therefore

$$V = \int_{-4}^{4} \pi\left(2 - \frac{y^2}{8}\right)^2 dy = \int_{-4}^{4} \pi\left(4 - \frac{y^2}{2} + \frac{y^4}{64}\right) dy = \pi\left[4y - \frac{y^3}{6} + \frac{y^5}{320}\right]_{-4}^{4} = \frac{256\pi}{15}$$

8.3. Example. The region between the curve $y = x^2$ and the line $y = x$ is rotated about the x-axis. Find the resulting volume.

Method: The line and curve intersect when $x = 0$ and when $x = 1$. Calculate the volume swept out by the curve, the volume swept out by the line and take the difference.

8.2.2 Curves given parametrically

This doesn't happen very often in volume problems. However, if it does, you use the same method, but at the integration stage express the integrand in terms of the parameter and substitute for dx or dy (as the case may be) by using $dx = \frac{dx}{dt}dt$ (or the corresponding expression for dy).

e.g. with the unit circle we have $x = \cos t$, $y = \sin t$ etc.

8.3 Lengths of Curves

8.2. Q. Find the length of the path along the curve $y = f(x)$ from $x = a$ to $x = b$.

Again the routine is to consider the small length, δs between the points (x, y) and $(x + \delta x, y + \delta y))$.

The length of the curve is approximately equal to that along the chord. So

$$\delta s^2 \approx \delta x^2 + \delta y^2 \tag{*}$$

Therefore

$$\left(\frac{\delta s}{\delta x}\right)^2 \approx 1 + \left(\frac{\delta y}{\delta x}\right)^2$$

Therefore

$$\left(\frac{ds}{dx}\right)^2 = 1 + \left(\frac{dy}{dx}\right)^2$$

Therefore

$$\left(\frac{ds}{dx}\right) = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}$$

Integrating we get

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

Parameters are quite common in length problems. To handle such problems go back to $(*)$ and divide through by $(\delta t)^2$ rather than $(\delta x)^2$. The result is the formula

$$\left(\frac{ds}{dt}\right)^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2$$

and then

$$s = \int_r^s \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dx}\right)^2}\, dt$$

where r and s are the values of t that give the limits of integration.

Because of the square root the integrals in these length calculations can be quite hard unless the curve is carefully chose. This doesn't make the method useless, but does mean that you often have to resort to numerical integration.

8.4. Example. Calculate the length of the section from $t = 0$ to $t = 2\pi$ of the curve which is given parametrically by

$$x = a(t - \sin t)\,, \qquad y = a(1 - \cos t)$$

(This curve is called the **cycloid** and is the path traced out by a point on a circle of radius a as the circle rolls along a straight line.)

We are finding the length of one of the arches.

$$\frac{dx}{dt} = a(1 - \cos t)\,, \quad \frac{dy}{dt} = a \sin t$$

Therefore

$$\begin{aligned}\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 &= a^2(1 - 2\cos t + \cos^2 t) + a^2 \sin^2 t \\ &= a^2(2 - 2\cos t) \\ &= 2a^2(1 - \cos t)\end{aligned}$$

Therefore

$$s = \int_0^{2\pi} \sqrt{2a^2(1 - \cos t)}\, dt$$

The square root is an obvious problem, but we can get rid of it by using the trig formula

$$\cos 2\theta = 1 - 2\sin^2 \theta$$

Rearranging this as $1 - \cos 2\theta = 2\sin^2 \theta$, and then putting $2\theta = t$ we get

$$1 - \cos t = 2\sin^2\left(\frac{t}{2}\right)$$

Therefore

$$\begin{aligned}s &= \int_0^{2\pi} \sqrt{4a^2 \sin^2\left(\frac{t}{2}\right)}\, dt \\ &= \int_0^{2\pi} 2a \sin\left(\frac{t}{2}\right) dt \\ &= \left[-4a\cos\left(\frac{t}{2}\right)\right]_0^{2\pi} \\ &= 8a\end{aligned}$$

8.5. Example. Find the length of the curve which is given parametrically by the equations

$$x = \cos^3 t\,, \qquad y = \sin^3 t$$

The picture is as shown.

The full curve is from $t = 0$ to $t = 2\pi$, but the obvious symmetry means that we can just compute the length of the portion in the first quadrant and then multiply by 4 to get the full length.

(Ans=6)

Exercise: Find the length of the curve which is given parametrically by the equations

$$x = 3\cos t + \cos 3t\,, \qquad y = 3\sin t + \sin 3t$$

(Total length is four times that from $t = 0$ to $t = \pi/2$.
(The curve is the nephroid.)

8.4 Centroids

The geometrical centre of a body is known as its **centroid**. If the body has uniform density, the centroid is also the **centre of gravity**.

We can locate the centroids of planar areas using integration. The approach is similar to the one used in the volumes of revolution — that is we consider small strips, and then use integration to add them up.

The moment exerted by a body is the same as that that would be exerted were the whole mass to be concentrated at the centre of gravity.

What we shall do is calculate this moment in two ways and thereby locate the centre of gravity.

In these problems the planar area is to be thought of as a thin metal plate of uniform density. So the mass is proportional to the area. We already know how to calculate areas, and so we can get the total mass.

8.4.1 Calculating M_x, the moment with respect to the x-axis.

The top edge of the area is the curve $y = f(x)$ and the bottom edge is $y = g(x)$. The limits are $x = a$ and $x = b$.

As is usual, we begin by looking at a thin strip, as shown.

The strip lies between x and $x + \delta x$.

Let the contribution of this strip to M_x be δM_x.

The strip may be thought of as a rectangle.

Its top edge is at height $f(x)$ and its bottom edge at $g(x)$. So the height of the rectangle is $f(x) - g(x)$.

The width is δx, and so the area (="the mass") is $(f(x) - g(x))\delta x$.

In calculating δM_x we think of this mass as being located at the rectangle's centre of gravity.

The centre of gravity of a rectangle is half way between the bottom and the top. The top is at $f(x)$, and the bottom at $g(x)$. Halfway between the two is $\dfrac{f(x) + g(x)}{2}$. The moment of the strip about the x-axis is the mass times the distance of the c. of g. from the x-axis.

Therefore,

$$\delta M_x = \frac{1}{2}(f(x) + g(x))(f(x) - g(x))\delta x$$

Therefore

$$\frac{\delta M_x}{\delta x} = \frac{1}{2}(f(x) + g(x))(f(x) - g(x))$$

and so

$$\frac{dM_x}{dx} = 12(f(x) + g(x))(f(x) - g(x))$$

Therefore

$$M_x = \int_a^b \frac{1}{2}(f(x) + g(x))(f(x) - g(x))\, dx \qquad (1)$$

8.4.2 Calculating M_y, the moment with respect to the y-axis.

The distance of the centre of gravity of the rectangle from the y-axis is (to a first approximation) x. Therefore

$$\delta M_y = x(f(x) - g(x))\delta x$$

Therefore

$$M_y = \int_a^b x(f(x) - g(x))\, dx \qquad (2)$$

8.4.3 Special Case

If the lower boundary of the region is the x-axis, we have $g(x) = 0$, and a consequent simplification of the formulas, which become

$$M_x = \int_a^b \frac{1}{2}y^2\, dx \qquad (3)$$

$$M_y = \int_a^b xy\, dx \qquad (4)$$

(writing y for $f(x)$.)

8.4.4 The Centroid

We now have expressions for M_x and M_y. To get the coordinates of the centroid we use the fact that these moments are equal to the ones you would get by concentrating the whole mass at the centre of gravity.

Let the centre of gravity be at $(\bar{x}, \bar{y})$, and let the area of the region be A.

The distance of the centre of gravity from the x-axis is $\bar{y}$, and so the moment about the x-axis is $A\bar{y}$. Similarly that about the y-axis is $A\bar{x}$. Therefore we have the formulas

$$A\bar{x} = M_y\,, \qquad A\bar{y} = M_x$$

We know how to calculate A, M_x and M_y, and so using these formulas we can calculate $\bar{x}$ and $\bar{y}$.

8.6. Example. Find the centroid of the region bounded by $y = \sqrt{x}$, the x-axis and the line $x = 1$.

Solution Calculate A, M_x, M_y and then put them into the formulas.

$$A = \int_0^1 \sqrt{x}\,dx = \left[\frac{2}{3}x^{3/2}\right]_0^1 = \frac{2}{3}$$

For the two moments we can use the simplified formulas, since the lower boundary is the x-axis.

$$M_x = \int_0^1 \frac{1}{2}(\sqrt{x})^2\,dx = \int_0^1 \frac{1}{2}x\,dx = \left[\frac{1}{4}x^2\right]_0^1 = \frac{1}{4}$$

$$M_y = \int_0^1 x\sqrt{x}\,dx = \int_0^1 x^{3/2}\,dx = \left[\frac{2}{5}x^{5/2}\right]_0^1 = \frac{2}{5}$$

Therefore

$$\bar{x} = \frac{M_y}{A} = \frac{2}{5}\frac{3}{2} = \frac{3}{5}$$

and

$$\bar{y} = \frac{M_x}{A} = \frac{1}{4}\frac{3}{2} = \frac{3}{8}$$

Therefore the centroid is at $\left(\dfrac{3}{5}, \dfrac{3}{8}\right)$.

8.7. Example. Find the centroid of the region bounded by the curve $y = x^2$ and the line $y = x$.

8.8. Remark. Non-uniform density. Into the calculation of the three integrals you have to incorporate a density function $\sigma(x, y)$. The principle remains the same, but to handle the integrals you now need the calculus of severable variables.

8.5 Surfaces of Revolution

This time we rotate a section of a curve about a line and consider the **surface area** of the solid created.

e.g. if we take the circle $x^2 + y^2 = r^2$ and rotate it about the x-axis, we get a sphere. The calculation would give us the surface area of that sphere.

As always, the plan is to look at a small section and then to integrate to get the whole thing.

The section of the curve $y = f(x)$ between $x = a$ and $x = b$ is rotated about the x-axis.

Consider the small segment lying between x and $x + \delta x$.

The rotation of the small segment produces a ring with a curved edge and radius y (approximately). To get the surface area of the ring we multiply the distance along the curved edge by 2π times the radius. ($2\pi r$ being the circumference of a circle of radius r.)

We know from our earlier work on curve length that the approximate distance along the curved edge is given by

$$\delta s = \sqrt{\delta x^2 + \delta y^2}$$

So the small element, δA, of surface area is given by

$$\begin{aligned}\delta A &\approx 2\pi y \delta s \\ &\approx 2\pi y \sqrt{\delta x^2 + \delta y^2}\end{aligned}$$

Therefore

$$\frac{\delta A}{\delta x} \approx 2\pi y \sqrt{1 + \left(\frac{\delta y}{\delta x}\right)^2}$$

and so

$$\frac{dA}{dx} = 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}$$

Integrating we get

$$A = \int_a^b 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

As in the problem of curve length we sometimes have to deal with curves given by parameters. This time the formula is

$$A = \int_a^b 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

This is the most complicated-looking of the formulas we have developed in this section. However, appearances deceive. In practice, the integrals you get are often easier to deal with than the length ones: that extra y term seems to help, often making for an integration by substitution.

8.9. Example. The circle $x^2 + y^2 = r^2$ is rotated about the x-axis. Calculate the surface area of the resulting sphere.

Solution We could express y as $\sqrt{r^2 - x^2}$ and differentiate, but this is another case where implicit differentiation is easier.

$$2x + 2y\frac{dy}{dx} = 0 \qquad \text{and so} \quad \frac{dy}{dx} = -\frac{x}{y}$$

The limits of integration are r and $-r$. Therefore

$$\begin{aligned} A &= \int_{-r}^{r} 2\pi y\sqrt{1 + \frac{x^2}{y^2}}\, dx \\ &= \int_{-r}^{r} 2\pi\sqrt{y^2 + x^2}\, dx \\ &= \int_{-r}^{r} 2\pi\sqrt{r^2}\, dx \\ &= \int_{-r}^{r} 2\pi r\, dx \\ &= [2\pi r x]_{-r}^{r} \\ &= 4\pi r^2 \end{aligned}$$

So the surface area of a sphere of radius r is $4\pi r^2$.

8.10. Example. A parabola is given parametrically by the equations $x = 5t^2$, $y = 10t$. The curve is as shown.

The section that runs from $t = 0$ to $t = 2$ is rotated about the x-axis. Calculate the surface area of the resulting solid.

Solution

$$\frac{dx}{dt} = 10t\,, \qquad \frac{dy}{dt} = 10$$

Therefore

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 100(t^2 + 1)$$

Therefore

$$\begin{aligned} A &= \int_0^2 2\pi(10t)\sqrt{100(t^2 + 1)}\, dt \\ &= \int_0^2 200\pi t\sqrt{t^2 + 1}\, dt \end{aligned}$$

This illustrates how the extra y in the integral helps. Had it not been there we should have had to integrate $\sqrt{t^2 + 1}$, and that involves using the substitution $t = \sinh\theta$ (see the tutorial sheet Integration 2). With it there we can use the much more straightforward substitution $u = t^2 + 1$. With this we have $du = 2tdt$ and we get

$$\begin{aligned} A &= \int_0^2 200\pi t\sqrt{t^2+1}\,dt \\ &= \int_{t=0}^{t=2} 100\pi\sqrt{u}\,du \\ &= 100\pi\left[\frac{2}{3}u^{3/2}\right]_{t=0}^{t=2} \\ &= 100\pi\left[\frac{2}{3}u^{3/2}\right]_{u=1}^{u=5} \\ &= \frac{200\pi}{3}\left(5\sqrt{5}-1\right) \end{aligned}$$

Chapter 9

Reduction Formulas

9.1 Reduction Formulas

This is refinement of the technique of integration by parts. It is a labour saving device designed to cope with the situation where a simple-minded approach would leave us integrating by parts many times over before reaching an answer.

9.1. Example. Evaluate $\int_0^1 x^6 e^x \, dx$.

Solution x^6 is a function that simplifies on differentiation, e^x is a function we can integrate in our heads, and so we use integration by parts.

Let $u = e^x$ and $v = x^6$. Then $\int u \, dx = e^x$ and $v' = 6x^5$. And so

$$\begin{aligned} \int_0^1 &= \left[x^6 e^x\right]_0^1 - \int_0^1 6x^5 e^x \, dx \\ &= e - 6 \int_0^1 x^5 e^x \, dx \end{aligned}$$

It is an improvement, since we have an integral involving a lower power of x, but we are still a long way from the answer. What we must do next is use parts again. This will get us down to $\int_0^1 x^4 e^x \, dx$. Then again to get down to $x^3 e^x$. And so on.

We shall reach an answer, but it is going to take time, time spent doing a lot of repetitive processes. What a reduction formula does is take the work out of the repeats. Instead of doing lots of integrations by parts, we shall do one, with an n in place of the 6. This will give us a formula which we can then use over and over with different values for n.

Let n be a positive integer, and let

$$I_n = \int_0^1 x^n e^x \, dx$$

Integrate by parts with $u = e^x$ and $v = x^n$. The result is

$$\begin{aligned} I_n &= [x^n e^x]_0^1 - \int n x^{n-1} e^x \, dx \\ &= e - n \int_0^n x^{n-1} e^x \, dx \end{aligned}$$

and so

$$I_n = e - nI_{n-1} \tag{*}$$

This is what is meant by a **reduction formula**. It gives I_n in terms of a simpler integral of the same type (I_{n-1}).

Now let us use it on our earlier problem, which was to calculate I_6.

(*) with $n = 6$ gives $I_6 = e - 6I_5$. Now use it again with $n = 5$ to get $I_5 = e - 5I_4$ and substitute to get

$$I_6 = e - 6I_5 = e - 6(e - 5I_4) = -5e + 30I_4$$

Now use the formula again, but with $n = 4$. Substitute once more, and then keep going in the same sort of way.

$$\begin{aligned} I_6 &= -5e + 30I_4 \\ &= -5e + 30(e - 4I_3) \\ &= 25e - 120I_3 \\ &= 25e - 120(e - 3I_2) \\ &= -95e + 360I_2 \\ &= -95e + 360(e - 2I_1) \\ &= 265e - 720I_1 \\ &= 265e - 720(e - I_0) \end{aligned}$$

At that point we have to pause, because our formula was only valid for $n > 0$ and we are now down to $n = 0$. However, I_0 is an integral we need no help with. $I_0 = \int_0^1 e^x \, dx = e - 1$. Therefore

$$I_6 = 265e - 720e + 720(e - 1) = 265e - 720$$

9.2. Example. Let $I_n = \int \sin^n x \, dx$ where n is an integer ≥ 2. Find a reduction formula for I_n, and then use it to calculate $\int_0^{\pi/2} \sin^6 x \, dx$.

Solution With all these questions we use integration by parts, and the aim is to recover an integral of the same shape but with a smaller n.

The largest section of the integrand that we can integrate in our heads is $\sin x$. So set $u = \sin x$ and $v = \sin^{n-1} x$ (the bit that is left after we have removed u).

$$\int u \, dx = -\cos x \,, \qquad \frac{dv}{dx} = (n-1)\sin^{n-2} x \cos x$$

Therefore

$$\begin{aligned}
I_n &= -\sin^{n-1}\cos x - \int (n-1)\sin^{n-2} x\cos x(-\cos x)\,dx \\
&= -\sin^{n-1}\cos x + (n-1)\int \sin^{n-2} x\cos^2 x\,dx \\
&= -\sin^{n-1}\cos x + (n-1)\int \sin^{n-2} x(1-\sin^2 x)\,dx \\
&= -\sin^{n-1}\cos x + (n-1)\int \left(\sin^{n-2} x - \sin^n x\right)\,dx \\
&= -\sin^{n-1}\cos x + (n-1)\left[\int \sin^{n-2} x\,dx - \int \sin^n x\,dx\right] \\
&= -\sin^{n-1} x\cos x + (n-1)\left[I_{n-2} - I_n\right]
\end{aligned}$$

Taking all the I_n terms on to the left we get

$$nI_n = -\sin^{n-1} x\cos x + (n-1)I_{n-2}$$

To use this on a definite integral we just put in the limits.

So if we had that $I_n = \int_0^{\pi/2} \sin^n x\,dx$, the formula would become

$$\begin{aligned}
nI_n &= \left[-\sin^{n-1} x\cos x\right]_0^{\pi/2} + (n-1)I_{n-2} \\
&= (n-1)I_{n-2} \qquad \text{provided } n \geq 2
\end{aligned}$$

Repeated application of this formula will bring you down to $I_1 = \int \sin x\,dx$ or $I_0 = \int\,dx$, both of which you can do without difficulty.

For example, with the definite integral and $n = 6$, we have

$$I_6 = \frac{5}{6}I_4 = \frac{5}{6}\frac{3}{4}I_2 = \frac{5}{6}\frac{3}{4}\frac{1}{2}I_0$$

and since $\displaystyle\int_0^{\pi/2} dx = \frac{\pi}{2}$, this leads to

$$I_6 = \frac{5}{6}\frac{3}{4}\frac{1}{2}I_0 = \frac{5}{6}\frac{3}{4}\frac{1}{2}\frac{\pi}{2} = \frac{5\pi}{32}$$

Reduction formulas for $\int \cos^n x\,dx$ and $\int x^n e^{ax}\,dx$ are to be found in your handbook. Others that are sometimes useful are

$$\int \sin^m x\cos^n x\,dx = \frac{\sin^{m+1} x\cos^{n-1} x}{m+n} + \frac{n-1}{m+n}\int \sin^m x\cos^{n-2} x\,dx$$

and

$$\int \sin^m x\cos^n x\,dx = -\frac{\sin^{m-1} x\cos^{n+1} x}{m+n} + \frac{m-1}{m+n}\int \sin^{m-2} x\cos^n x\,dx$$

Faced with an integral such as $\int \sin^8 x\cos^4 x\,dx$ you use the first of these to bring down the powers of $\cos x$ and the second to bring down those of $\sin x$.

Chapter 10

Complex Numbers

10.1 Introduction

You are familiar with the following situation with regard to quadratic equations:

> The equation $x^2 = 1$ has two roots $x = 1$ and $x = -1$.
> The equation $x^2 = -1$ has no roots because you cannot take the square root of a negative number.

Long ago mathematicians decided that this was too restrictive. They did not like the idea of an equation having no solutions — so they invented them. They proved to be very useful, even in practical subjects like engineering.

Consider the general quadratic equation $ax^2 + bx + c = 0$ where $a \neq 0$. The usual formula. obtained by "completing the square" gives the solutions

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

If $b^2 \geq 4ac$ we are happy. If $b^2 < 4ac$ then the number under the square root is negative and you would say that the equation has no solutions. In this case we can write $b^2 - 4ac = (-1)(4ac - b^2)$ and $4ac - b^2 > 0$. So, in an obvious formal sense,

$$x = \frac{-b \pm \sqrt{-1}\,\sqrt{4ac - b^2}}{2a},$$

and now the only 'meaningless' part of the whole formula is $\sqrt{-1}$.

So we might say that any quadratic equation either has "real" roots in the usual sense or else has roots of the form $p + q\sqrt{-1}$, where p and q belong to the real number system $\mathbb{R}$.

The expressions $p + q\sqrt{-1}$ do not make any sense as real numbers, but there is nothing to stop us from playing around with them *as symbols*. In fact, playing around with them proves to be very useful for applications to problems in differential equations, electrical circuit theory and fluid mechanics.

Although we don't discuss it formally here, there is a number system larger than $\mathbb{R}$ containing a special number j such that $j^2 = -1$, called the complex numbers, and written $\mathbb{C}$. This number system can be put on just as proper or correct a foundation as $\mathbb{R}$, and so, although we introduce it as a device to do calculations, there is no logical objection to its use. Informally, you can think of $j = \sqrt{-1}$ but remember that $(-j)^2 = -1$ too.

We call these numbers *complex numbers*; the special number j is called an *imaginary* number, even though j is just as "real" as the real numbers and complex numbers are probably simpler in many ways than real numbers.

Engineering usage is different from that of mathematicians or physicists. One of the important early uses in engineering is in connection with electrical circuits, and in particular, in calculating current flows. The symbol i is reserved for current, and so j is used for $\sqrt{-1}$, while mathematicians use i for this *imaginary* number.

For manipulations, remember that

$$\boxed{j^2 = -1}$$

10.1. Definition. A **complex number** is any expression of the form $x + jy$, where x and y are ordinary real numbers. The collection of all complex numbers is denoted by $\mathbb{C}$.

Note that all real numbers are complex numbers as well: $x = x + 0j$.

10.2 The Arithmetic of Complex Numbers

Using the 'rule' $j^2 = -1$ we can build up an 'arithmetic' of complex numbers which is very similar to that of ordinary numbers. In this Section we define what we mean by the sum, difference, product and ratio of complex numbers. All the definitions are derived by assuming that the ordinary rules of arithmetic work with the addition that $j^2 = -1$.

In what follows let a, b, c, d are ordinary real number, and let $z = a + jb$ and $w = c + jd$ be two complex numbers.

$$z = w \quad \text{iff} \quad a = c \quad \text{and} \quad b = d \qquad (\textbf{Equality})$$

$$z + w = (a + c) + j(b + d) \qquad (\textbf{Addition})$$

$$z - w = (a - c) + j(b - d) \qquad (\textbf{Subtraction})$$

$$zw = (ac - bd) + j(bc + ad) \qquad (\textbf{Multiplication})$$

$$\frac{z}{w} = \frac{(ac + bd) + j(bc - ad)}{c^2 + d^2} \qquad (\textbf{Division})$$

Equality, Addition and Subtraction are fairly obvious. Note that the definition of equality enables us to 'equate real parts' and 'equate imaginary parts' when we have two complex numbers that are equal. The definitions of multiplication and division are more complicated. They come about as follows.

Multiplication Multiply out $(a + jb)(c + jd)$ just as you usually would

$$zw = (a + jb)(c + jd) = ac + jbc + ajd + j^2bd.$$

Now add in the information that $j^2 = -1$ and get

$$zw = ac + jbc + jad - bd = (ac - bd) + j(bc + ad),$$

as given above.

Division This is even more complicated. Start by noting that

$$(c + jd)(c - jd) = c^2 - j^2d^2 = c^2 + d^2.$$

Now use this to rearrange the quotient as follows:

$$\frac{z}{w} = \frac{a + jb}{c + jd} = \frac{(a + jb)(c - jd)}{(c + jd)(c - jd)} = \frac{(ac + bd) + j(bc - ad)}{c^2 + d^2},$$

again as given above.

This arithmetic works in the same way as ordinary arithmetic. You use the usual rules.

It is not really worth remembering the 'formula' for the product and quotient. It is better to calculate them in the same way that I derived the formulae.

10.2. Example. If $z = 2 + 3j$ and $w = 1 - 2j$ what are $z + w$, $z - w$, zw and z/w?

Solution We compute using these rules.

$$\begin{aligned}
z + w &= (2 + 3j) + (1 - 2j) = (2 + 1) + (3 - 2)j = 3 + j. \\
z - w &= (2 + 3j) - (1 - 2j) = (2 - 1) + (3 + 2)j = 1 + 5j. \\
zw &= (2 + 3j)(1 - 2j) = 2 - 4j + 3j - 6j^2 = 2 - 4j + 3j + 6 = 8 - j. \\
\frac{z}{w} &= \frac{2 + 3j}{1 - 2j} = \frac{(2 + 3j)(1 + 2j)}{(1 - 2j)(1 + 2j)} = \frac{2 + 3j + 4j - 6}{1 + 4} = \frac{-4 + 7j}{5} = -\frac{4}{5} + \frac{7}{5}j.
\end{aligned}$$

10.3. Example. If $z = 5 - 6j$, $w = 2 + 7j$ and $x = -1 - j$ what are $2z + w - x$, $z^2 - x^2$, xzw, $1/z$, $z(w - 2x)$?[1]

Solution Again computing, we have:

$$\begin{aligned}
2z + w - x &= 2(5 - 6j) + (2 + 7j) - (-1 - j) = 10 - 12j + 2 + 7j + 1 + j = 13 - 4j; \\
z^2 - x^2 &= (5 - 6j)(5 - 6j) - (-1 - j)(-1 - j), \\
&= 25 - 60j - 36 - (1 - 1 + 2j) = -11 - 62j; \\
xzw &= (-1 - j)(5 - 6j)(2 + 7j) = (-1 - j)(10 - 12j + 35j + 42), \\
&= (-1 - j)(52 + 23j) = -52 - 52j - 23j + 23 = -29 - 75j; \\
\frac{1}{z} &= \frac{5 + 6j}{(5 - 6j)(5 + 6j)} = \frac{5 + 6j}{25 + 36} = \frac{5 + 6j}{61}; \\
z(w - 2x) &= (5 - 6j)(2 + 7j + 2 + 2j) = (5 - 6j)(4 + 9j), \\
&= 20 - 24j + 45j + 54 = 74 + 21j.
\end{aligned}$$

10.4. Example. What are the roots of the quadratics $x^2 + x + 1 = 0$, $x^2 - 2x + 3 = 0$?

Solution Using the usual formula, we get for roots as follows:

$$\frac{-1 \pm \sqrt{1 - 4}}{2} = \frac{-1 \pm j\sqrt{3}}{2}$$

and

$$\frac{2 \pm \sqrt{4 - 12}}{2} = \frac{2 \pm j\sqrt{8}}{2} = 1 \pm j\sqrt{2}.$$

[1]You see here the difficulty of trying to reserve particular symbols for particular meanings. Usually x is a real number, but here we have no need of symbols for real numbers, but need three different ones for complex numbers. So x gets temporarily used as a complex number.

Powers of j. What are the powers of j? Starting from $j^2 = -1$ we get

$$j^2 = -1, \quad j^3 = -j, \quad j^4 = 1, \quad j^5 = j, \quad \dots$$

The powers go in a cycle of length 4.

10.2.1 Square Roots

The square root of a complex number z is any complex number w such that $w^2 = z$. Given a particular z, it is not too hard to calculate the square roots of z.

10.5. Example. Let $z = 1 - 5j$; calculate the square roots of z.

Solution Write $w = a + jb$ where a and b are real. Then

$$1 - 5j = (a + jb)(a + jb) = (a^2 - b^2) + 2jab.$$

So our problem reduces to that of solving two simultaneous *real* equations;

$$a^2 - b^2 = 1 \quad \text{and} \quad 2ab = -5$$

The second one gives $b = -5/(2a)$. Put this into the first and get

$$a^2 - \frac{25}{4a^2} = 1 \quad \text{or} \quad 4a^4 - 4a^2 - 25 = 0$$

This looks a bit worrying because it is an equation of degree 4. The trick is to notice that it is actually just a quadratic for a^2. So, by the usual formula for quadratics,

$$a^2 = \frac{4 \pm \sqrt{16 + 400}}{8} = \frac{1 \pm \sqrt{26}}{2}$$

Now a is definitely a real number, so its square cannot be negative. So the only possibility is that

$$a^2 = \frac{1}{2}(1 + \sqrt{26}) \quad \text{and} \quad a = \pm\sqrt{\frac{1 + \sqrt{26}}{2}}$$

This gives us two possible values for a. The corresponding values of b are then obtained from $b = -5/(2a)$. This gives **two** values for w, differing by a factor of -1.

10.2.2 Complex Conjugates

Let $z = x + jy$ be a complex number. We say that z is *real* if $y = 0$, and *purely imaginary* if $x = 0$. The real number x is called the *real part* of z and written $x = \Re z$. The real number y is called the *imaginary part* of z and written $y = \Im z$.

The *complex conjugate* of the complex number $z = x + jy$ is the complex number $\bar{z} = x - jy$. Thus z and $\bar{z}$ have the same real part, while $z + \bar{z}$ has 0 as its imaginary part. Note that

$$z\bar{z} = x^2 + y^2, \qquad z + \bar{z} = 2x, \qquad z - \bar{z} = 2jy.$$

10.6. Example. Let $z_1 = 2 + 3j$, $z_2 = 4j$ and $z_3 = -j$. Give the real and imaginary parts, and the complex conjugates of z_1, z_2 and z_3.

Solution If $z_1 = 2 + 3j$ then z_1 has real part 2, imaginary part 3 and complex conjugate $\bar{z}_1 = 2 - 3j$.

If $z_2 = 4j$ then z_2 is purely imaginary and $\bar{z}_2 = -4j$.

If $z_3 = -j$ then z_3 is purely imaginary and $\bar{z}_3 = j$.

10.3 The Argand Diagram

You are familiar with the representation of real numbers as points along a line:

Figure 10.1: The real line.

A complex number $z = x + jy$ is specified by *two* real numbers x and y. So it is often useful to think of a complex number as being represented by the point in a plane with Cartesian coordinates (x, y). This representation is called the **Argand diagram** or the **complex plane**.

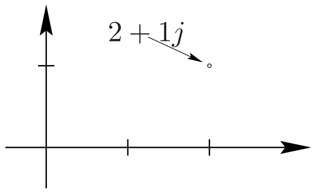


Figure 10.2: The Argand diagram or complex plane.

10.4 Modulus and Argument

Thinking in terms of the Argand diagram we can specify the position of the complex number $z = x + jy$ on the plane by giving the polar coordinates of the point (x, y).

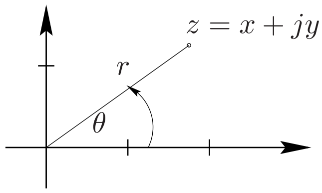


Figure 10.3: The modulus - argument representation of z.

The polar coordinate r is the distance from O to P and is called the *modulus* of the complex number z and written as $|z|$.

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

The polar coordinate θ is called *an argument* of z. If we take θ in the range $-\pi < \theta \leq \pi$ then we call it *the (principal) argument* of z and we denote it by $\arg(z)$. Note that any argument of z differs from $\arg(z)$ by an integer multiple of 2π (working in radians) or of $360°$ (working in degrees)[2].

[2]You are reminded that there is a very good reason for working in radians: the derivative of $\sin x$ is $\cos x$

Since $x = r\cos\theta$ and $y = r\sin\theta$ we can write z in terms of its modulus and argument as

$$\boxed{z = r(\cos\theta + j\sin\theta)} \qquad r \geq 0, \quad -\pi < \theta \leq \pi.$$

This is called writing z in *polar form* or *modulus - argument form.* Any non-zero complex number can be written in this form. The point 0 is a slightly special case, it has $r = 0$ but the angle θ is not defined.

10.7. Example. Give the real and imaginary parts, complex conjugate and the modulus and argument of each of the complex numbers $z = 1 + j$, $z = 1 - j$, $z = -4j$, $z = -3$, $z = -1 - 3j$.

Solution The given complex numbers are plotted in the complex plane in Fig 10.4.

$z = 1 + j$ has real part 1, imaginary part 1, complex conjugate $\bar{z} = 1 - j$ and modulus $|z| = \sqrt{1+1} = \sqrt{2}$. The argument of z is $\pi/4$.

$$z = \sqrt{2}(\cos\frac{\pi}{4} + j\sin\frac{\pi}{4}).$$

$z = 1 - j$ has real part 1, imaginary part -1, complex conjugate $\bar{z} = 1 + j$ and modulus $|z| = \sqrt{2}$. The argument of z is, according to our conventions, $-\pi/4$.

$$z = \sqrt{2}(\cos\frac{\pi}{4} - j\sin\frac{\pi}{4}).$$

Clearly $z = -4j$ has real part 0, imaginary part -4, complex conjugate $\bar{z} = 4j$ and modulus $|z| = \sqrt{0 + (-4)^2} = 4$. The argument of z is $-\pi/2$, and

$$z = 4\left(\cos -\frac{\pi}{2} + j\sin -\frac{\pi}{2}\right).$$

$z = -3$ has real part -3, imaginary part 0, complex conjugate $\bar{z} = -3 = z$ and modulus $|z| = 3$. The argument of z is, according to our conventions, π so $z = 3(\cos\pi + j\sin\pi)$.

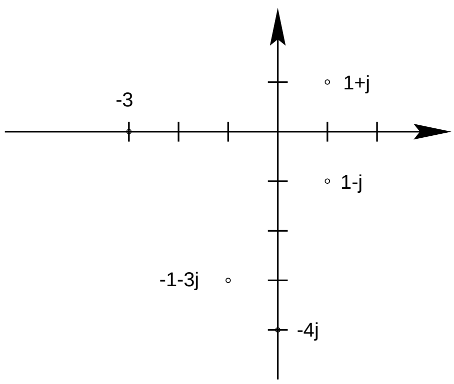


Figure 10.4: Plotting points: Example 10.7.

$z = -1 - 3j$ has real part -1, imaginary part -3, complex conjugate $\bar{z} = -1 + 3j$ and modulus $|z| = \sqrt{1+9} = \sqrt{10}$. The argument of z has to be found with the aid of a calculator. It lies in the range $-\pi < \theta < -\pi/2$ (third quadrant) and has value

$$\theta = \arctan\frac{-3}{-1} - \pi = -1.8925.$$

It is perhaps of interest that this problem with arctan is quite common; so common that many computer languages, starting with Fortran, have two version of the function, typically called `atan` and `atan2`. The first one genuinely computes the inverse tangent function, and returns an angle between $-\pi/2$ and $\pi/2$; the second function is the "proper" one in our context andit takes the *two* arguments need to compute the angle to within 2π.

only when the angle is measured in radians; if degrees are used there is a constant $\pi/180$ in the formula: there is a similar reason for measuring arguments in radians.

10.5 Products

Let $z = r(\cos\theta + j\sin\theta)$ and $w = s(\cos\varphi + j\sin\varphi)$ be two complex numbers in polar form. Thus $r = |z|$ and $\theta = \arg(z)$, while $s = |w|$ and $\varphi = \arg(w)$.

Consider the product of z and w:

$$\begin{aligned} zw &= rs(\cos\theta + j\sin\theta)(\cos\varphi + j\sin\varphi) \\ &= rs((\cos\theta\cos\varphi - \sin\theta\sin\varphi) + j(\sin\theta\cos\varphi + \cos\theta\sin\varphi)) \\ &= rs(\cos(\theta+\varphi) + j\sin(\theta+\varphi)) \end{aligned}$$

This tells us that the modulus of zw is just the product of the moduli of z and w:

$$|zw| = |z|\,|w|$$

and, provided we adjust the angles to the correct range by adding or subtracting multiples of 2π, the argument of the product is the sum of the arguments:

$$\arg(zw) = \arg(z) + \arg(w) \qquad \text{(modulo } 2\pi).$$

For example, if z has argument $120°$ and w has argument $150°$ then an argument of zw is $120 + 150 = 270$, which is not in the right range, so we subtract $360°$ and get the principal argument, which is $-90°$ (or $-\pi/2$ radians).

Similarly, for $w \neq 0$,

$$\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$$

and

$$\arg\left(\frac{z}{w}\right) = \arg(z) - \arg(w) \qquad \text{(modulo } 2\pi).$$

10.6 De Moivre's Theorem

If we repeat the process of the above section over and over again we can show that if $z \neq 0$ and n is a positive whole number then $|z^n| = |z|^n$ and $n\arg(z)$ is an argument of z^n.

Since $\left|\frac{1}{z}\right| = \frac{1}{|z|}$ and $\arg(1/z) = -\arg(z)$ we also get the same results if n is a negative integer. So we have the above formulae for all integer values of n ($n = 0$ is easy — check).

The result is often put in the following useful form, which is known as **de Moivre's Theorem**. If n is any whole number then

$$\boxed{z = r(\cos\theta + j\sin\theta) \quad \Rightarrow \quad z^n = r^n(\cos n\theta + j\sin n\theta)}$$

Let me tell you of one other notation at this point, which looks a bit obscure at the moment but which you will meet a lot in later years:

$$\boxed{e^{j\theta} = \cos\theta + j\sin\theta}$$

So

$$z = r(\cos\theta + j\sin\theta) = re^{j\theta} \quad \text{and} \quad z^n = r^n e^{nj\theta}.$$

10.7 The Roots of Unity

The problem here is to solve the equation $z^n = 1$, where n is usually a positive whole number.

Write *both* sides of the equation in polar form. Let z have polar form

$$z = r(\cos\theta + j\sin\theta) \quad \text{so that} \quad z^n = r^n(\cos n\theta + j\sin n\theta)$$

We know that

$$1 = 1(\cos 0 + j\sin 0)$$

So our equation becomes

$$r^n(\cos n\theta + j\sin n\theta) = 1(\cos 0 + j\sin 0)$$

Now two complex numbers in standard polar form are equal if and only if their moduli and arguments are equal. In the case of the argument this statement has to be handled with care. It means 'are equal if reduced to the proper range'. So, for example, 10° and 370° count as equal from this point of view.

So we can say that $r^n = 1$ and that $n\theta$ and 0 are equal up to the addition of some multiple of 2π radians.

$$r^n = 1 \qquad n\theta = 0 + 2k\pi$$

where k is some whole number.

Since r is real and positive, the only possibility for r is $r = 1$.

The other equation gives us

$$\theta = 0 + 2\pi\frac{k}{n}.$$

This, in principle, gives us infinitely many answers! One for each possible whole number k. But not all the answers are different. Remember that changing the angle by 2π does not change the number z.

The distinct solutions, of which there are n, are given by $r = 1$ and

$$\theta = 2\pi\frac{k}{n} \qquad k = 0, 1, 2, 3, \dots, n-1$$

and we can write these solutions as

$$z_k = \cos\theta_k + j\sin\theta_k \quad \text{where} \quad \theta_k = 2\pi\frac{k}{n} \qquad k = 0, 1, 2, \dots n-1$$

That looks rather complicated. It becomes a lot simpler if you think in terms of the Argand diagram. All the solutions have modulus 1 and so lie on the circle of radius 1 centred at the origin. The solution with $k = 1$ is just $z = 1$. The other solutions are just $n - 1$ other points equally spaced round this circle, with angle $2\pi/n$ between one and the next. This is illustrated in Fig 10.5 in fact for the case $n = 17$.

Let's look at some specific examples. The cube roots of unity are the solutions to $z^3 = 1$. There are three of them and they are

$$z_0 = 1, \qquad z_1 = \cos 2\pi/3 + j\sin 2\pi/3, \qquad z_2 = \cos 4\pi/3 + j\sin 4\pi/3$$

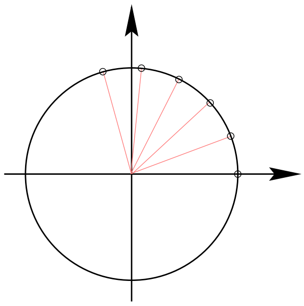


Figure 10.5: The n^{th} roots of 1.

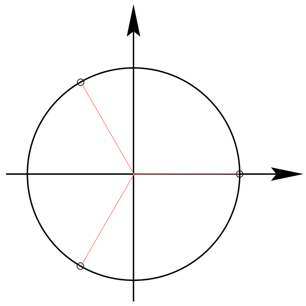


Figure 10.6: The three cube roots of 1.

Note that $z_2 = \bar{z}_1$, $z_2 = z_1^2$ and $1 + z_1 + z_2 = 0$. the roots are shown in Fig 10.6.

Similarly the fourth roots of unity are the solutions of $z^4 = 1$ and these are

$$z = 1, \quad z = j, \quad z = -1, \quad z = -j.$$

A picture for $n = 4$ together with those for $n = 5$ and $n = 6$ is given in Fig 10.7

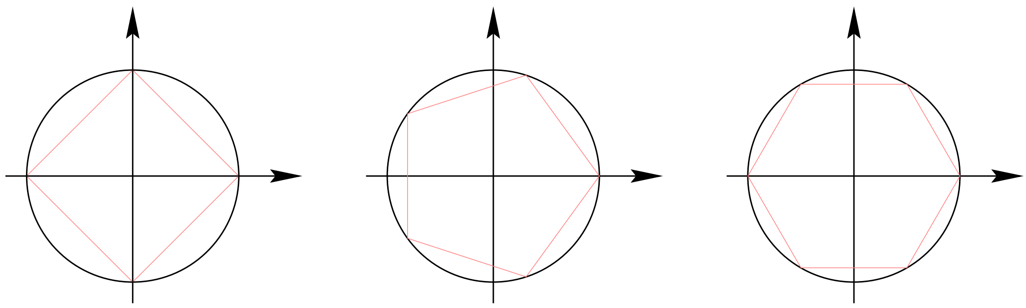


Figure 10.7: The n^{th} roots of 1 for $n = 4, 5, 6$.

We can do other equations like this in much the same way.

10.8. Example. Find the solutions of the equation $z^4 = j$.

Solution Put $z = r(\cos\theta + j\sin\theta)$. Then $z^4 = r^4(\cos 4\theta + j\sin 4\theta)$. We know that $j = 1(\cos\pi/2 + j\sin\pi/2$. So our equation becomes

$$r^4(\cos 4\theta + j\sin 4\theta) = 1(\cos\pi/2 + j\sin\pi/2)$$

Therefore

$$r = 1 \quad \text{and} \quad 4\theta = \frac{\pi}{2} + 2k\pi \quad \text{or} \quad \theta = \frac{\pi}{8} + k\frac{\pi}{2}$$

There are 4 distinct solutions, given by $k = 0, 1, 2, 3$. They form a square on the unit circle.

10.8 Polynomials

We have learned how to manipulate complex numbers, and suggested that they will prove valuable in Engineering calculations. The original motivation for introducing them was to give the equation $x^2 = -1$ two roots, namely j and $-j$, rather than it having no roots. It turns out that this is *all* we have to do to ensure that every polynomial has the *right* number of roots. We now discuss this, and a number of other basic results about polynomials, that are quite useful to know.

A *polynomial* in x is a function of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where the a's are (real or complex) numbers and $a_n \neq 0$. For example

$$p(x) = x^3 - 2x + 4, \quad q(t) = 5t^8 - t^4 + 6t^3 - 1.$$

The highest power in the polynomial is called the *degree* of the polynomial. The above examples have degrees 3 and 8.

A number a (real or complex) is said to be a *root* of the polynomial $p(x)$ if $p(a) = 0$. Thus $x = 1$ is a root of $x^2 - 2x + 1$.

The first important result about polynomials is that a number a (real or complex) is a root of the polynomial $p(x)$ if and only if $(x - a)$ is a *factor* of $p(x)$, in the sense that we can write $p(x)$ as

$$p(x) = (x - a)q(x)$$

where $q(x)$ is another polynomial. This result is often called the *remainder theorem* . For example, $x = 2$ is a root of $p(x) = x^3 + x^2 - 7x + 2$ and it turns out that

$$p(x) = (x - 2)(x^2 + 3x - 1).$$

Note that necessarily the polynomial q has degree one less than the degree of p.

It may be the case that you can pull more than one factor of $(x-a)$ out of the polynomial. For example, 2 is a root of $p(x) = x^3 - x^2 - 8x + 12$ and it turns out that

$$p(x) = (x - 2)(x - 2)(x + 3).$$

In such cases a is said to be a *multiple root* of $p(x)$. The *multiplicity* of the root is the number of factors $(x - a)$ that you can take out. In the above example, 2 is a root of multiplicity 2, or a *double root.* A root is called a *simple root* if it produces only one factor. Multiple roots are a considerable pain in the neck in many applications, but they have the advantage that the Fundamental Theorem of Algebra, Theorem 10.9 takes a simple form.

There is a simple test for multiplicity. Suppose a is a root of $p(x)$, so that $p(a) = 0$. If, in addition, $p'(a) = 0$ (derivative) then a is a multiple root. To take the above example: $p(x) = x^3 - x^2 - 8x + 12$. We have $p'(x) = 3x^2 - 2x - 8$ and $p(2) = 0$ and we have $p'(2) = 0$, so we know that 2 is a multiple root.

Let me prove this result. Since a is a root of $p(x)$, we can write $p(x) = (x - a)q(x)$ where $q(x)$ is another polynomial. By the product rule,

$$p'(x) = (x - a)q'(x) + q(x).$$

So $p'(a) = 0.q'(a) + q(a) = q(a)$. Since $p'(a) = 0$ we have $q(a) = 0$. But this means that $q(x)$ has $(x - a)$ as a factor — and hence that $p(x)$ has $(x - a)$ as a factor more than once.

You should check the converse: if a is a multiple root of $p(x)$ (so that $p(x) = (x-a)^2 q(x)$ for some polynomial $q(x)$) then $p(a) = p'(a) = 0$.

The next result is fundamental. I am not going to attempt to prove it in detail; it requires some rather fancy mathematics!.

10.9. Theorem (Fundamental Theorem of Algebra). *Let p be* any *polynomial of degree n. Then p can be factored into a product of a constant and n factors of the form $(x-a)$, where a may be real or complex.*

Also, the factorisation is unique; you cannot find two essentially different factorisations for the same polynomial. The factors need not all be different because of multiple roots. The fact that there cannot be *more than n* such factors is fairly obvious, since we would have the wrong degree. What is not at all obvious is that we have all the factors that we want. Note that this result does not tell you how to find these factors; just that they must be there!

The result is often stated loosely as: a polynomial of degree n must have exactly n roots. You have to allow complex roots or the theorem is not true. For example $p(x) = x^2 + 1$ has no real roots at all. Its roots are $x = \pm j$ and it factorises as $p(x) = (x - j)(x + j)$.

We have already seen this result in action when solving equations earlier in the Chapter. I told you then that you can take it for granted that an equation like $z^7 = 2 + j$ will have exactly 7 solutions. In fact, if $w \neq 0$ then $p(z) = z^n - w$ $(n \geq 1)$ always has exactly n *distinct* roots because we know that it must have n roots in all and it cannot have any multiple roots because $p'(z) = nz^{n-1}$ has only 0 as a root and 0 is not a root of $p(z)$.

There is one other result about roots of polynomials that is worth knowing. Suppose we have a polynomial with *real*, as opposed to complex, coefficients. Suppose that the complex number z is a root of the polynomial. Then the complex conjugate $\bar{z}$ is also a root. So you get two roots for the price of one. You can see this in the example of the previous paragraph. $x^2 + 1$ has j as a root, so it automatically must have $-j$ as a root as well.

10.10. Example. Let $p(z) = z^4 - 4z^3 + 9z^2 - 16z + 20$. Given that $2 + j$ is a root, express $p(z)$ as a product of real quadratic factors and list all four roots, drawing attention to any conjugate pairs.

Solution Since p has real coefficients, and complex roots occur in pairs consisting of a root and its complex conjugate. Given that $2 + j$ is a root, it follows that $2 - j$ must also be a root, and so the quadratic

$$(z - (2 + j))(z - (2 - j)) = z^2 - 4z + 5$$

must be a factor. Dividing the given polynomial by this factor gives

$$p(z) = z^4 - 4z^3 + 9z^2 - 16z + 20 = (z^2 - 4z + 5)(z^2 + 4).$$

The roots of $z^2 + 4$ are $2j$ and its complex conjugate, $-2j$. Thus the given polynomial, of degree four, has two pairs of complex conjugate roots.

Having seen how useful the result can be in practice, let me give a proof, because it is really a very simple manipulation with complex conjugates.

10.11. Proposition. *Let P be a polynomial with real coefficients, and assume that $p(z_0) = 0$. Then $p(\bar{z}_0) = 0$.*

Proof. Let

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$$

and assume that $a_0,\ a_1, \dots a_n \in R$. Thus p is a polynomial with real coefficients. Let $p(z_0) = 0$, so that

$$a_0 + a_1 z_0 + a_2 z_0^2 + \cdots + a_n z_0^n = 0.$$

We have

$$p(\bar{z}_0) = a_0 + a_1 \bar{z}_0 + a_2 \bar{z}_0^2 + \cdots + a_n \bar{z}_0^n$$

and since each coefficient is real,

$$\begin{aligned} &= \bar{a}_0 + \bar{a}_1 \bar{z}_0 + \bar{a}_2 \bar{z}_0^2 + \cdots + \bar{a}_n \bar{z}_0^n \\ &= p(\bar{z_0}) = 0 \quad \text{since } z_0 \text{ is a root of } p. \end{aligned}$$

Thus $\bar{z}_0$ is a root of p as claimed.

Of course if $z_0 \in \mathbb{R}$, the result tells us nothing, since in that case $z_0 = \bar{z}_0$. But as we saw in the example, if we have found one complex root, we can immediately get hold of another one; the complex roots come in pairs.

10.12. Example. Express $z^5 - 1$ as a product of real linear and quadratic factors.

Solution We rely on our knowledge of the n^{th} roots of unity from Section 10.7. Let

$$\alpha = \exp\left(\frac{2\pi j}{5}\right) = \cos\left(\frac{2\pi}{5}\right) + j \sin\left(\frac{2\pi}{5}\right)$$

Then the roots of $z^5 - 1 = 0$ are α, α^2, α^3, α^4 and 1 and

$$z^5 - 1 = (z-1)(z^4 + z^3 + z^2 + z + 1) = (z-1)(z-\alpha)(z-\alpha^2)(z-\alpha^3)(z-\alpha^4).$$

For convenience, write $\beta = \alpha^2$, and note that $\bar{\beta} = \alpha^3$ while $\bar{\alpha} = \alpha^4$. Our problem is to factorise $z^4 + z^3 + z^2 + z + 1$ as a product of real quadratic factors. We know the roots are α, $\bar{\alpha}$, β and $\bar{\beta}$. Now construct the quadratic with roots α and $\bar{\alpha}$. We have

$$(z-\alpha)(z-\bar{\alpha}) = z^2 - (\alpha + \bar{\alpha}) + \alpha.\bar{\alpha} = z^2 - 2\Re(\alpha) + 1$$

where $\Re(\alpha)$ is the real part of α. Since $(z-\beta)(z-\bar{\beta})$ behaves in the same way, we have

$$\begin{aligned} z^5 - 1 &= (z-1)(z^2 - 2\Re(\alpha) + 1)(z^2 - 2\Re(\beta) + 1), \\ &= (z-1)\left(z^2 - 2\cos\left(\frac{2\pi}{5}\right) + 1\right)\left(z^2 - 2\cos\left(\frac{4\pi}{5}\right) + 1\right). \end{aligned}$$

and this is a product of real linear and quadratic factors.

Questions 1 (Hints and solutions start on page 159.)

10.1. Q. a) Let $z = 1 + 3j$ and $w = 2 - j$ be complex numbers. Express each of the following complex numbers in the form $x + jy$ where x and y are real numbers.

$$z - 3w, \quad \frac{1}{z}, \quad \left|\frac{w + \bar{w}}{w - \bar{w}}\right|.$$ **[5 marks]**

b) Express the complex number

$$z = 1 - \sqrt{3}j$$

exactly in modulus - argument form. Hence find the modulus and principal argument of z^4.
[5 marks]

c) Find all solutions w to the equation

$$w^3 = -27j$$

and mark them on an Argand diagram. **[5 marks]**

10.2. Q. a) Let $z = 1 - 2j$ and $w = 3 + j$ be complex numbers. Express each of the following complex numbers in the form $x + jy$ where x and y are real numbers.

$$zw, \quad \frac{w}{z + 2 + j}, \quad |1 + 3j - z\bar{z}|.$$ **[5 marks]**

b) Express the complex number $-2 + 2j$ exactly in modulus - argument form. Hence find all solutions w to the equation

$$w^3 = -2 + 2j$$

and mark them on an Argand diagram. **[5 marks]**

10.3. Q. a) Let $z = 3 + j$ and $w = 1 - 7j$. Express

$$\frac{w}{w + \bar{z}}$$

in the form $x + jy$ where x and y are real. Find also

$$|z|, \quad |w|, \quad \left|\frac{w}{z}\right|.$$ **[5 marks]**

b) Express the complex number $-2+2j$ in polar form. Hence solve the equation

$$z^3 = -2 + 2j,$$

expressing the solutions in polar form and marking them in the Argand Diagram. **[5 marks]**

10.4. Q. a) Let

$$p(z) = z^5 - 5z^4 + 8z^3 - 2z^2 - 8z + 8.$$

Show that $p(2) = 0$. Show also that $z^2 - 2z + 2$ is a factor of $p(z)$. Hence write p as a product of linear factors. **[10 marks]**

10.5. Q. Show that $z - (1 + j)$ is a factor of the real polynomial

$$p(z) = z^3 + 2z^2 - 6z + 8.$$

Hence write p as a product of linear factors. **[5 marks]**

10.6. Q. Let

$$p(z) = z^4 - 3z^3 + 5z^2 - 27z - 36.$$

Show that $p(3j) = 0$. Hence write p as a product of linear factors. **[5 marks]**

Chapter 11

Matrices

11.1 Introduction

A **matrix** is a rectangular array of numbers. For example

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 4 & 6 \\ 2 & -3 & \pi \end{pmatrix}, \qquad \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix}.$$

You will find both round and square brackets being used.

If a matrix has n **rows** and m **columns** then it is said to be an $n \times m$ matrix. The above examples are 2×2, 2×3 and 2×1.

We usually denote the *element* in the i^{th} row and j^{th} column of a matrix A by A_{ij}, or sometimes a_{ij}. So the elements of an $n \times m$ matrix are labelled as follows

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1m} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{1m} \\ A_{31} & A_{32} & A_{33} & \cdots & A_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{n1} & A_{n2} & A_{n3} & \cdots & A_{nm} \end{pmatrix}$$

and we often write things like $A = (a_{ij})$ to indicate that the elements of the matrix A are called a_{ij}. For example

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}.$$

At the moment a matrix is just a *symbol* or a *diagram*. However, it turns out that we can usefully combine matrices in various ways. In fact we can develop an *algebra* for matrices.

11.2 Terminology

There a number of words that are used to describe matrices of certain shapes and properties. I will gather a number of them together here for reference.

Two matrices are said to have the same size if they have the same number of rows and the same number of columns.

A matrix with the same number of rows as columns is called a **square matrix**, other types tend to be called **rectangular**.

An $n \times 1$ matrix (n rows and 1 column) is called a **column vector** and a $1 \times n$ matrix (1 row and n columns) is called a **row vector** . Vectors are considered in their own right in the "Engineering Science" course EG1007.

The **diagonal** of a square matrix is made up of the elements down the diagonal from top left to bottom right: $a_{11}, a_{22}, \dots, a_{nn}$. For example

$$\begin{pmatrix} \mathbf{1} & 2 & 1 \\ 5 & \mathbf{2} & 2 \\ 4 & 3 & \mathbf{6} \end{pmatrix}$$

A square matrix is said to be a **diagonal matrix** if all the elements not on the diagonal are zero. So the following are diagonal matrices:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

A **zero matrix** is a matrix with all its elements zero. We write such a matrix as 0. The size usually has to be deduced from the context.

The $n \times n$ **identity matrix** is the $n \times n$ (square) diagonal matrix with 1's down the diagonal. This is denoted by I_n, or just I if the size is known from the context. For example

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The **transpose** A^{T} of a matrix A is the matrix obtained by switching over the rows and the columns of A, so the first row of A becomes the first column of A^{T} and so on. In symbols we have

$$(A^{\mathrm{T}})_{ij} = A_{ji}.$$

If A is an $n \times m$ matrix then A^{T} is an $m \times n$ matrix. Examples:

$$\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 2 & 6 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 2 & 1 & 2 \\ 3 & 4 & 6 \end{pmatrix}$$

In particular, the transpose of a row vector is a column vector and *vice versa.*

A square matrix A is said to be **symmetric** if $a_{ij} = a_{ji}$ and **skew-symmetric** or **anti-symmetric** if $a_{ij} = -a_{ji}$. The following matrix is symmetric

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 5 \\ 3 & 5 & 4 \end{pmatrix}, \qquad A = A^{\mathrm{T}}.$$

The following matrix is skew-symmetric:

$$A = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 3 \\ -2 & -3 & 0 \end{pmatrix}, \qquad A = -A^{\mathrm{T}}.$$

Note that the diagonal elements of a skew-symmetric matrix must all be zero, since $a_{ii} = -a_{ii}$.

That's enough jargon for the time being. Note that most of these definitions are just referring to the *shape* of the matrix.

11.3 Matrix Algebra

It turns out to be very useful to combine matrices in various ways. We are going to set up an *algebra* for matrices by defining rules for adding, subtracting, multiplying matrices etc. The algebra that we get is quite similar to that of ordinary numbers but there are also some very important differences.

11.3.1 Addition of Matrices

Suppose that A and B are two matrices *of the same size*. Then their **sum** $C = A + B$ is defined by the rule

$$C_{ij} = A_{ij} + B_{ij}$$

In other words, each element of the sum is the sum of the corresponding elements of A and B. For example

$$\begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 7 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 11 & 10 \end{pmatrix},$$

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 2 \\ 6 \end{pmatrix} \quad \text{is not allowed.}$$

11.3.2 Subtraction of Matrices

This follows the same idea as for addition. Suppose that A and B are two matrices *of the same size*. Then their **difference** $C = A - B$ is defined by the rule

$$C_{ij} = A_{ij} - B_{ij}.$$

For example,

$$\begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} - \begin{pmatrix} 2 & 3 \\ 7 & 4 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -3 & 2 \end{pmatrix},$$

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

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{is not allowed.}$$

11.3.3 Multiplication by a Number

If λ is a number and A is a matrix then λA is the matrix, of the same size as A, given by

$$(\lambda A)_{ij} = \lambda A_{ij}$$

In other words, just multiply each element of A by λ.

For example,

$$2\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix} = \begin{pmatrix}2 & 4\\6 & 8\end{pmatrix}, \qquad -3\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}-3\\-6\end{pmatrix}.$$

$$\pi\begin{pmatrix}1 & 2\end{pmatrix} = \begin{pmatrix}\pi & 2\pi\end{pmatrix}, \qquad -\begin{pmatrix}2 & 3\\4 & 5\end{pmatrix} = \begin{pmatrix}-2 & -3\\-4 & -5\end{pmatrix}$$

Putting some of these rules together we now get things like this: if $A = \begin{pmatrix}1 & 0 & 2\\3 & 1 & 0\end{pmatrix}$ and $B = \begin{pmatrix}4 & 0 & 0\\5 & 2 & 1\end{pmatrix}$ then

$$3A - 2B = \begin{pmatrix}3 & 0 & 6\\9 & 3 & 0\end{pmatrix} - \begin{pmatrix}8 & 0 & 0\\10 & 4 & 2\end{pmatrix} = \begin{pmatrix}-5 & 0 & 6\\-1 & -1 & -2\end{pmatrix}.$$

11.3.4 Multiplication of Matrices

This is where the fun starts. The rules up to now have been fairly obvious. The rule for multiplication is *not* the obvious one. I will give the definition, explain how to use it and then try to explain where it comes from.

Suppose that A is an $n \times l$ matrix and B an $l \times m$ matrix, so that A has the same number of *columns* as B has *rows*. Then the matrix $C = AB$ is defined by

$$C_{ij} = \sum_{k=1}^{l} A_{ik}B_{kj}$$

and is an $n \times m$ matrix (same number of rows as A and same number of columns as B). Spelling the definition out a bit more, we get

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + A_{i3}B_{3j} + \cdots + A_{il}B_{lj}.$$

You can think of this as saying that the (i, j) element of C is the result of multiplying that i row of A into the j column of B.

Some examples may make this clearer.

$$\begin{pmatrix}2 & 1\\3 & 4\end{pmatrix}\begin{pmatrix}4 & 2\\6 & 1\end{pmatrix} = \begin{pmatrix}2.4 + 1.6 & 2.2 + 1.1\\3.4 + 4.6 & 3.2 + 4.1\end{pmatrix} = \begin{pmatrix}14 & 5\\36 & 10\end{pmatrix},$$

$$\begin{pmatrix}1 & 2 & 1\\2 & 0 & 1\end{pmatrix}\begin{pmatrix}4 & 2\\0 & 1\\2 & 1\end{pmatrix} = \begin{pmatrix}1.4 + 2.0 + 1.2 & 1.2 + 2.1 + 1.1\\2.4 + 0.0 + 1.2 & 2.2 + 0.1 + 1.1\end{pmatrix} = \begin{pmatrix}6 & 5\\10 & 5\end{pmatrix},$$

$$\begin{pmatrix}2 & 3 & 4\end{pmatrix}\begin{pmatrix}1\\2\\1\end{pmatrix} = \begin{pmatrix}2.1+3.2+4.1\end{pmatrix} = (12) \quad \text{a} \quad 1\times 1 \quad \text{matrix} \quad ,$$

$$\begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}3 & 4\end{pmatrix} = \begin{pmatrix}1.3 & 1.4\\2.3 & 2.4\end{pmatrix} = \begin{pmatrix}3 & 4\\6 & 8\end{pmatrix}.$$

You will need some practice to get used to this definition.

11.3.5 Origins of the Definition

Where does such a complicated definition come from?

Suppose we change coordinates in a plane by a transformation of the form

$$x' = a_{11}x + a_{12}y,$$
$$y' = a_{21}x + a_{22}y.$$

This is what is known as a *linear transformation*. We can represent this transformation by giving its coefficients, arranged as a matrix:

$$A = \begin{pmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\end{pmatrix},$$

which contains all the necessary information about the transformation. If we write

$$X' = \big(x', y'\big) \text{ and } X = \big(x, y\big), \text{ we see that } X' = AX, \tag{11.1}$$

and we represent the action of this linear transformation by multiplication with the column vector of co-ordinates.

Now suppose that we do a second transformation of the same form after the first one, say the transformation:

$$x'' = b_{11}x' + b_{12}y',$$
$$y'' = b_{21}x' + b_{22}y',$$

represented by the matrix

$$B = \begin{pmatrix}b_{11} & b_{12}\\b_{21} & b_{22}\end{pmatrix}$$

What now is the transformation from (x, y) to (x'', y'')? Doing the substitutions gives

$$x'' = b_{11}(a_{11}x + a_{12}y) + b_{12}(a_{21}x + a_{22}y)$$
$$y'' = b_{21}(a_{11}x + a_{21}y) + b_{22}(a_{21}x + a_{22}y),$$

and if we tidy this up we get

$$x'' = (b_{11}a_{11} + b_{12}a_{21})x + (b_{11}a_{12} + b_{12}a_{22})y$$
$$y'' = (b_{21}a_{11} + b_{22}a_{21})x + (b_{21}a_{12} + b_{22}a_{22})y.$$

So the transformation from (x, y) to (x'', y'') is also a linear transformation and its coefficient matrix C (say) is given by

$$C = \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{pmatrix} = BA$$

So, according to our definition of matrix multiplication, the matrix of the combined transformation is just the product of the matrices of the individual transformations. That is why multiplication is defined the way that it is.[1]

The result is true in general. Multiplication of matrices is defined so as to agree with what happens when we combine, or compose linear transformations.

11.4 Properties of Matrix Algebra

Much of matrix algebra follows the same rules as ordinary algebra. The main difference comes with the behaviour of matrix multiplication.

The thing to remember about matrix multiplication is that, in general, $AB \neq BA$. So the order in which you multiply matrices *matters*. If it does happen to be the case that $AB = BA$ then we say that these two matrices *commute*. An example in which $AB \neq BA$ is given by

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

$$\begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}.$$

A square matrix certainly commutes with itself, so it does make sense to talk about A^2, A^3 etc, if A is a square matrix. But note the following difficulty. Suppose that A and B are two square matrices of the same size. Then

$$(A + B)^2 = (A + B)(A + B) = A(A + B) + B(A + B) = \\ AA + AB + BA + BB = A^2 + AB + BA + B^2$$

since the multiplying out of brackets is allowed. Now note that this cannot be simplified further. It need not be equal to $A^2 + 2AB + B^2$, because BA need not be the same thing as AB.

The message here is that you have to be very careful when using standard algebraic techniques on matrices in case you are accidentally assuming that matrices commute. You have to particularly careful when multiplying out brackets. But this is the *only* rule that goes wrong; things like $A(BC) = (AB)C$, the *associativity* of multiplication, still hold whenever they make sense.

[1] You may wish to read BA as B *follows* A; it is then clear that $C = BA$ has come out the correct way round

11.4.1 The Identity and Zero Matrices

We defined these earlier on. Now note their basic properties. Let us stick to square $n \times n$ matrices. Let I and 0 be the $n \times n$ identity and zero matrices. Then, for any $n \times n$ matrix A,

$$AI = IA = A, \qquad A0 = 0A = 0, \qquad A + 0 = A.$$

So the identity matrix works like a '1' and the zero matrix works like a '0'. For example

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}$$

11.4.2 Relating Scalar and Matrix Multiplication

We now have two sorts of multiplication that can be used on a matrix A; we can multiply it by a number or by another matrix. At the moment there is no problem about what we mean by a number, but in Section 10 we meet another sort of number called a *complex* number, and it turns out that these work just as well. This sort of multiplication is given a special name; it is **scalar multiplication**, because the numbers are sometimes called **scalars**. This contrasts with **matrix multiplication** just discussed, when we take the product of two matrices.

How are they related? In the obvious way. If λ is a scalar, or number, and A is an $n \times m$ matrix, then

$$\lambda A = (\lambda I_n)A = A(\lambda I_m),$$

which is exactly what you would expect. In other words, multiplication by a scalar is the same as matrix multiplication by a diagonal matrix (of the right size) with the scalar on the diagonal.

And as you expect, multiplying by a 1×1 matrix (a_{11}) when it is allowed is exactly the same as multiplying by the scaler a_{11}.

11.4.3 Transpose of a Product

This is worth a mention, though I will not prove anything. The transpose of a product is related to the transposes of the individual terms of the product by

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}.$$

In other words, the transpose of the product is the product of the transposes in the reverse order. And we have just seen that the order in which a product is written is important.

This works for any length of product:

$$(ABC \ldots Z)^{\mathrm{T}} = Z^{\mathrm{T}} \ldots C^{\mathrm{T}}B^{\mathrm{T}}A^{\mathrm{T}}.$$

11.4.4 Examples

Let us now use what we know to do some calculations with matrices.

11.1. Example. If $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ evaluate $C = A(A+B) + 2B^2$.

Solution

$$A + B = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \qquad A(A+B) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 6 \\ 2 & 2 \end{pmatrix},$$

$$B^2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}.$$

So

$$C = A(A+B) + 2B^2 = \begin{pmatrix} 6 & 6 \\ 2 & 2 \end{pmatrix} + 2\begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 6 \\ 10 & 4 \end{pmatrix}$$

11.2. Example. Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \end{pmatrix}$. Which of the following expressions make sense?

$$AB, \quad BA, \quad A+B, \quad AB^{\mathrm{T}}, \quad BB^{\mathrm{T}}, \quad B^{\mathrm{T}}B.$$

Solution AB certainly does not work because A has two columns and B has only one row — they don't fit.

BA make sense because B has two columns and A has two rows.

$A + B$ does not make sense because A and B have different sizes.

AB^{T} does work because B^{T} has two rows and A has two columns.

Both BB^{T} and $B^{\mathrm{T}}B$ work (they always work). Note that BB^{T} is a 1×1 matrix whereas $B^{\mathrm{T}}B$ is a 2×2 matrix. So, in fact, something like $A + 2B^{\mathrm{T}}B$ will also work.

11.3. Example. Prove that the square of a symmetric matrix or an anti-symmetric matrix is a symmetric matrix.

Solution Recall that A is symmetric if $A^{\mathrm{T}} = A$ and anti-symmetric if $A^{\mathrm{T}} = -A$. Now

$$(A^2)^{\mathrm{T}} = (AA)^{\mathrm{T}} = A^{\mathrm{T}}A^{\mathrm{T}}$$

and this last expression is equal to $AA = A^2$ in both cases. So $(A^2)^{\mathrm{T}} = A^2$ and hence A^2 is symmetric.

11.4. Example. We know that in general $AB \neq BA$. Consider the matrix $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Which 2×2 matrices B satisfy $AB = BA$?

Solution Write $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We want to find the possibilities for a, b, c and d. Now

$$AB = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 2a+c & 2b+d \end{pmatrix}, \quad \text{and}$$
$$BA = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} a+2b & b \\ c+2d & d \end{pmatrix}.$$

For these two matrices to be equal a, b, c and d must satisfy the simultaneous equations

$$\begin{aligned} a + 2b &= a, & b &= b, \\ c + 2d &= 2a + c, & d &= 2b + d. \end{aligned}$$

The second equation tells us nothing. The first says that $b = 0$. The last equation now tells us nothing and, finally, the third equation tells us that $a = d$. So, putting it all together, b must be 0, a must equal d and c can be anything. So the matrices that we are looking for are those of the form

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

where a and c are arbitrary.

11.5 Inverses of Matrices

What about *dividing* matrices? This is much more difficult.

Let's start by thinking about ordinary numbers. The problem of dividing a by b is just the same as the problem of multiplying a by $1/b$. So, granted that we know how to do multiplication, the problem of division boils down to the problem of finding 'inverses': the inverse of b is $1/b$ — the unique number which, when multiplied by b gives 1. Note that this is really the same problem as that of solving the equation $bx = 1$.

We can always find the inverse of an ordinary number so long as that number is not zero.

Now consider matrices. We are going to consider square matrices only and are going to ask this question: given a square matrix A can we find a square matrix B such that $AB = I$? If so, there can be only one such B; we will call B the *inverse matrix* of A and write it as A^{-1}.

11.5. Example. If we know a *little* more about B, it is easy to show it is unique: suppose that $AB = I$ and $CA = I$. Show that $B = C$.

Solution From the definitions and the associativity of multiplication, we have

$$B = IB = (CA)B = C(AB) = CI = C.$$

The inverse of a 2×2 matrix. If A^{-1} is the inverse matrix for A, we have

$$AA^{-1} = I \quad \text{and} \quad A^{-1}A = I.$$

Before considering the general case let's look at 2×2 matrices in detail.

Suppose that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Can we find a matrix B such that $AB = I$? Let $B = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$, and suppose it is the inverse for A. Then

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & y \\ z & t \end{pmatrix} = \begin{pmatrix} ax + bz & ay + bt \\ cx + dz & cy + dt \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This gives us the following simultaneous equations to solve for x, y, z and t:

$$\begin{aligned} 1 &= ax + bz & \qquad 0 &= ay + bt \\ 0 &= cx + dz & \qquad 1 &= cy + dt \end{aligned}$$

Solving the first and third equations and then the second and fourth gives:

$$x = \frac{d}{\Delta}, \qquad y = \frac{-b}{\Delta}, \qquad z = \frac{-c}{\Delta}, \qquad t = \frac{a}{\Delta}$$

where $\Delta = ad - bc$. Which is fine and gives us a unique answer, *provided* that Δ is not zero! In fact, it can be shown that:

> The matrix 2×2 matrix A will have an inverse so long as $ad - bc \neq 0$. In that case the inverse is unique and is given by
>
> $$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

11.6. Example. Show that the matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ has an inverse and find the inverse.

Solution The matrix A has an inverse because $\Delta = 1.3 - 2.2 = -1 \neq 0$. The inverse is

$$A^{-1} = \frac{1}{-1} \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}.$$

You can now easily check that

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

11.7. Example. The matrix $A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ does *not* have an inverse.

Solution The matrix A does not have an inverse because $\Delta = 1.2 - 2.1 = 0$. This shows up the important fact that it is not only the zero matrix that does not have an inverse — lots of others don't either.

11.8. Example. When does the diagonal matrix $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ have an inverse?

Solution The diagonal matrix $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ has an inverse precisely when both a and b are not zero. The inverse is then

$$A^{-1} = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b} \end{pmatrix} = \begin{pmatrix} a^{-1} & 0 \\ 0 & b^{-1} \end{pmatrix}.$$

Let me now tell you what happens in general, without proving anything and leaving one important concept to the next section.

11.9. Theorem. *Let A be a square $n \times n$ matrix. There may or may not be an $n \times n$ matrix B such that $AB = I$ (or $BA = I$). If there is such a matrix B then it is unique and is called the* inverse A^{-1} *of A. In this case we say that A is* invertible *or* non-singular *and we have not only $AA^{-1} = I$ but also*

$$A^{-1}A = I \quad \textit{and} \quad (A^{-1})^{-1} = A.$$

There is a number associated to each $n \times n$ matrix, called the determinant, *and a square matrix has an inverse if and only if this determinant is not zero.*

The determinant of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $ad - bc$.

One useful result is that an $n \times n$ *diagonal* matrix is invertible precisely when none of the diagonal elements is zero. In that case the inverse is the diagonal matrix whose elements are the inverses of the original elements. So for example

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

11.5.1 Complex Numbers and Matrices

You may be surprised to find that our definition of matrix multiplication already gives us the arithmetic of complex numbers. Define two matrices by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and consider all matrices of the form $aI + bJ$ for $a, b \in \mathbb{R}$. Given that I is the identity matrix, it may not be too confusing to write such matrices as $a + bJ$. What may be surprising is that the matrix product of $a + bJ$ with $c + dJ$ is *exactly* what it should be if we thought of each matrix as a complex number and replaced J by j. That isn't all; you can easily check that $a + bJ$ has a matrix inverse if and only if $a^2 + b^2 > 0$ — the same condition you need to invert the corresponding complex number — and that the inverse is what you expect.

In other words, you don't need to *invent* j as soon as you have the idea of a matrix. I hope this dispells any myth that j does not exist!

11.6 Determinants

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the number $ad - bc$ is called the **determinant** of A. We write it as $\det(A)$, or $|A|$ or $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

More generally, associated with any $n \times n$ matrix $A = (a_{ij})$ we have a number, called the **determinant** of A, denoted as above.

The definition of this number is rather complicated. I have given it for 2×2 matrices. The definition for 3×3 matrices is given in terms of 2×2 matrices as follows:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

For an $n \times n$ matrix A the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and j^{th} column of A is called the (i, j)**-minor** of A. We denote it by M_{ij}.

We can now write the above definition of the determinant of a 3×3 matrix as

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13},$$

which looks a bit more tidy.

I can now give you the definition of the determinant of an $n \times n$ matrix A. It is just the same as the above, expressing $\det(A)$ in terms of the minors of the top row of A.

$$\det(A) = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} - \cdots \pm a_{1n}M_{1n}.$$

Note that the signs are alternating $+ - + - +-$ etc.

Here is an example.

$$\begin{aligned} \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 1 & 1 & 3 & 1 \\ 1 & 0 & 4 & 3 \end{vmatrix} &= 1. \begin{vmatrix} 1 & 0 & 2 \\ 1 & 3 & 1 \\ 0 & 4 & 3 \end{vmatrix} - 0 + 0 - 1. \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 3 \\ 1 & 0 & 4 \end{vmatrix}, \\ &= 1. \left(1. \begin{vmatrix} 3 & 1 \\ 4 & 3 \end{vmatrix} - 0 + 2. \begin{vmatrix} 1 & 3 \\ 0 & 4 \end{vmatrix} \right) - 1. \left(0 - 1. \begin{vmatrix} 1 & 3 \\ 1 & 4 \end{vmatrix} + 0 \right), \\ &= \begin{vmatrix} 3 & 1 \\ 4 & 3 \end{vmatrix} + 2 \begin{vmatrix} 1 & 3 \\ 0 & 4 \end{vmatrix} + \begin{vmatrix} 1 & 3 \\ 1 & 4 \end{vmatrix}, \\ &= (9 - 4) + 2(4 - 0) + (4 - 3), \\ &= 14. \end{aligned}$$

That's the definition. We don't often work out determinants in this way if we can help it. It gets to be very hard work if n is much bigger than 4. It can be shown that, using the above method, it takes in all about $(e-1)n!$ multiplications to work out an $n \times n$ determinant. The number of multiplications needed to evaluate a 20×20 determinant is $4,180,411,311,071,440,000$. If a computer can do a million multiplications per second, and we don't count the time for the additions etc., then the evaluation of a 20×20 determinant will take about $130,000$ years by this method. This is not practical! There are better methods which will reduce the time to a matter of seconds. These methods are consequences of the basic properties of determinants that I will now explain.

11.6.1 Properties of Determinants

I am going to state a number of properties of determinants without proof. They can all be proved from the above definition.

I will adopt the usual habit of using the word 'determinant' to refer both to the *value* and to the array of numbers. So I will talk about a 'row of the determinant' when what I really mean is a row of the matrix that produces the determinant.

Here are some rules:

1. Interchanging two rows of A just changes the sign of $\det(A)$.
2. Interchanging two columns of A just changes the sign of $\det(A)$.
3. If A has a complete row, or column, of zeroes then $\det(A) = 0$.
4. $\det(A) = \det(A^{\mathrm{T}})$.
5. To any row of A we can add any multiple of any other row without changing $\det(A)$.
6. To any column of A we can add any multiple of any other column without changing $\det(A)$.
7. A common factor of all the elements of a row of A can be 'taken outside the determinant', in the following sense:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ p.a_{21} & p.a_{22} & p.a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = p \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

8. The same applies to columns.
9. If all the elements of A below (or above) the diagonal are zero then the determinant is equal to the product of the diagonal elements. In particular, the determinant of a diagonal matrix is equal to the product of the diagonal elements. For example

$$\begin{vmatrix} a & b & c & d \\ 0 & p & q & r \\ 0 & 0 & s & t \\ 0 & 0 & 0 & u \end{vmatrix} = a.p.s.u.$$

10. The determinant of a product is the product of the determinants. In symbols,

$$\det(AB) = \det(A)\det(B).$$

These give us ways to manipulate a determinant into a more manageable form for calculation. Let me do one or two examples to give you the idea. I will not be very systematic about it at this stage.

11.10. Example. Show that

$$\Delta = \begin{vmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 3 & 2 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{vmatrix} = 3.$$

Solution We aim to produce as many zeros as possible and, ideally to produce a matrix in which all the elements below (or above) the diagonal are zero.

$$\begin{vmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 3 & 2 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & -1 \\ 3 & 2 & -2 & -1 \\ 1 & 1 & 1 & 0 \end{vmatrix}, \qquad [C_3' = C_3 - C_1];\ [C_4' = C_4 - C_1].$$

Here we have subtracted column 1 from column 3 and from column 4.

$$\Delta = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & -1 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & 0 \end{vmatrix}. \qquad [R_3' = R_3 - R_2].$$

Here we have taken row 2 from row 3. Now switch over rows 2 and 4, which changes the sign:

$$\Delta = -\begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & -2 & 0 \\ 2 & 1 & 0 & -1 \end{vmatrix}. \qquad [R_2' = R_4];\ [R_4' = R_2].$$

Finally, subtract column 2 from column 3 to get:

$$\Delta = -\begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & -3 & 0 \\ 2 & 1 & -1 & -1 \end{vmatrix} \qquad [C_3' = C_3 - C_1].$$

Now all the elements above the diagonal are zero, so the value of the determinant is the product of the diagonal elements. So

$$\det(A) = -(1 \times 1 \times -3 \times -1) = -3.$$

11.11. Example. Prove that $\begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix} = (x-y)(y-z)(z-x).$

Solution Before we start, remember that $a^2 - b^2 = (a-b)(a+b)$. We are going to use this a lot.

Start by subtracting row 1 from both row 2 and row 3 to get:

$$\det(A) = \begin{vmatrix} 1 & x & x^2 \\ 0 & y-x & y^2-x^2 \\ 0 & z-x & z^2-x^2 \end{vmatrix}.$$

All the terms in the second row now have common factor $(y-x)$ and all the terms in the third row have common factor $(z-x)$. So use the rules to pull these out:

$$\det(A) = (y-x)(z-x)\begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 1 & z+x \end{vmatrix}.$$

Next we subtract row 2 from row 3 and get a matrix in which all the terms below the diagonal are zero:

$$\begin{aligned}\det(A) &= (y-x)(z-x)\begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 0 & z-y \end{vmatrix} = (y-x)(z-x).1.1.(z-y) \\ &= (x-y)(y-z)(z-x).\end{aligned}$$

11.7 The Fourier Matrix

In the remainder of these notes we show how matrices are of use in describing and solving systems of equations. However before we do that, we present an application which may be a little surprising. The **Fourier Matrix** is used almost everywhere in modern computer graphics; almots every television program you watch will use these ideas somewhere in the special effects or titles.

The Fourier matrix is a **complex matrix**; one whose entries are *complex* numbers. So far we have deliberatly restricted out attention to **real matrices** because complex ones introduce no new ideas. However this example is a matrix which is really a "complex rotation" and we need the extra flexibility to have complex coefficients. It sounds odd but is certainly useful.

We need a simple result about complex numbers.

11.12. Lemma. *Let $z^n = 1$, and assume that $z \neq 1$. Then*

$$1 + z + z^2 + \cdots + z^{n-1} = 0.$$

Proof. It is trivial to verify the factorisation

$$z^n - 1 = (z-1)(1 + z + z^2 + \cdots + z^{n-1}).$$

The result follows, since we are given a zero of the left hand side.

Now recall our work on the n^{th} roots of unity in Section 10.7. Let $\omega = \exp(-2\pi j/n)$; then the roots are all of the form ω^k for some k with $0 \le k < n$. Note also that ω satisfies the condition of the lemma; that is why the lemma will be interesting.

We now define the Fourier matrix $\mathcal{F}_n$ for any integer $n \ge 1$ by:

$$\mathcal{F}_n = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{pmatrix}.$$

Note that $\mathcal{F}_n$ is a complex symmetric matrix; so is the same as its transpose. Thus its adjoint (the transposed complex conjugate of the given matrix) is just the complex conjugate. We write this as $\bar{\mathcal{F}}_n$. It is given simply by replacing ω in the above by $\bar{\omega} = \exp(2\pi j/n)$.

A word about notation: we shall have quite a lot to do with ω. Note the choice of sign in the exponent; it was arbitrary, but this way the definition of $\mathcal{F}_n$ is correct. When we need to emphasise that we have an n^{th} root of unity, we shall write ω_n.

In the study of complex vector spaces, an interesting class of matrices are the *unitary* ones; those whose inverse is the adjoint. They are the complex equivalent of the *orthogonal* matrices, in which the inverse is just the transpose. An orthogonal matrix represents a rotation of given orthonormal axes (possibly with a reflection); a unitary matrix could be considered for this reason, as a complex rotation.

11.13. Theorem. *The Fourier matrix $\mathcal{F}_n$ is a unitary matrix.*

Proof. It is enough to show that $\bar{\mathcal{F}}_n\mathcal{F}_n = I_n$, where I_n is the identity matrix. Since the p^{th} row of $\bar{\mathcal{F}}_n$ is

$$\frac{1}{\sqrt{n}}(1, \bar{\omega}^p, \bar{\omega}^{2p}, \dots, \bar{\omega}^{(n-1)p})$$

while the q^{th} column of $\mathcal{F}_n$ is

$$\frac{1}{\sqrt{n}}(1, \omega^q, \omega^{2q}, \dots, \omega^{(n-1)q}),$$

and since $\bar{\omega}^q = w^{-q}$, the entry in the $(p, q)^{th}$ place of the product is just

$$\frac{1}{n}(1 + \omega^{(p-q)} + \omega^{2(p-q)} + \dots + \omega^{(n-1)(p-q)}).$$

We now distinguish two cases. If $p = q$, this is just $(1 + 1 + \dots + 1)/n = 1$, while if $p \ne q$, it is zero by the lemma, since $\omega^{(p-q)}$ is an n^{th} root of unity, but is not equal to 1. It follows that the product is the identity matrix as claimed.

We can identify a complex valued function defined on $\mathbb{Z}^n$ with an element of $\mathbb{C}^n$ by $f \to \mathbf{f} = (f(0), f(1), \dots, f(n-1))$.[2] The discrete Fourier transform of $\mathbf{f}$ is then defined

[2] We confuse the function f and the vector $\mathbf{f}$; it may have been clearer to drop the bold font vector notation.

to be the action of the Fourier matrix on $(f(0), f(1), \dots, f(n-1))$. It will be convenient to write $\mathbf{f} \in \mathbb{C}^n$, or $\mathbf{f} = (f_1, f_2, \dots, f_{n-1})$ although we think of $\mathbb{C}^n$ in this form as periodic functions on $\mathbb{Z}^n$.

Identifying the Fourier matrix as unitary gives one way of thinking of the discrete Fourier transform; as a complex rotation in $\mathbb{C}^n$. The idea that a rotation of the co-ordinate axes will preserve essential features of a problem, while possibly making the new co-ordinate representation easier to work with, is a familiar one. We thus begin a study which will show there are indeed some things which are easier to describe and manipulate in the transformed representation.

11.1. Q. he Fourier matrices $\mathcal{F}_2$ and $\mathcal{F}_3$ are

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad \text{and} \qquad \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix},$$

where

$$\omega = -\frac{1}{2} - \frac{i\sqrt{3}}{2} = \cos\left(\frac{2\pi}{3}\right) - j\sin\left(\frac{2\pi}{3}\right).$$

Check that these are both unitary matrices directly, and write down $\mathcal{F}_4$ and $\mathcal{F}_5$.

11.14. Corollary. *Let* $\mathbf{f} = (f_0, f_1, \dots, f_{n-1})$ *be a vector in* $\mathbb{R}^n$ *or* $\mathbb{C}^n$*, and define* $\mathbf{F}$ *in* $\mathbb{C}^n$ *by*

$$F_p = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f_k \exp\left(\frac{-2\pi jkp}{n}\right) = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f_k\omega^{jp} \qquad (0 \le p \le n-1), \tag{11.2}$$

Then the transformation can be inverted to give

$$f_k = \frac{1}{\sqrt{n}}\sum_{p=0}^{n-1} F_p \exp\left(\frac{2\pi jkp}{n}\right) = \frac{1}{\sqrt{n}}\sum_{p=0}^{n-1} F_p\omega^{-kp} \qquad (0 \le k \le n-1). \tag{11.3}$$

Proof. Our definition gives $\mathbf{F} = \mathcal{F}_n(\mathbf{f})$. Since the Fourier matrix is unitary, we have $\mathbf{f} = \bar{\mathcal{F}}_n(\mathbf{F})$, which is the required result.

11.7.1 Rapid convolution and the FFT

It so happens that the Fourier transform can be done rapidly for certain values of n. Indeed even if $n = 512 \times 512$ it can be implmented in hardware in such a way that it can be pipelined - essentially it performs a complex rotation is real time. In this arrangement the operation is known as the **fast fourier transfrom** or **FFT**.

Many image manipulations involve an operation called *convolution* in which each pixel has its value modified depending on values of all of its neighbouring pixels. One such eaxmple is that of making an image more blurred, but there are lots of more interesting ones.

Now you see the problem. If you have to look up *lots* of neighbouring values for each pixel in an image, and there are (say) 512×512 pixels to do this on, the operation will be very slow. the cruicial importance of the FFT is that it converts convolution into a very quite pointwise multiplication operation. It is much qucker to do a complex rotation (the FFT), then do a simple manipulation, and then rotate backwards (the inverse FFT) than to implement convolution directly. And that is how it is done every night on television!

11.8 Linear Systems of Equations

.

We have spent some time finding out how to manipulate matrices, but have only hinted above and in Section 11.3.5 as to why they are of interest. Their main use is to describe the real world; the position or motion of an object in space, or the state of a collection of coupled systems. We leave such descriptions to the Engineering courses which rely on you being able to handle matrices. Here we take up a very simple application, to the study of simultaneous linear equations.

The system of n equations in m unknowns

$$\left.\begin{array}{llllllll} a_{11}x_1 & +a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1m}x_m & = b_1 \\ a_{21}x_1 & +a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2m}x_m & = b_2 \\ \vdots & & \vdots & & & \vdots & & & \\ a_{n1}x_1 & +a_{n2}x_2 & + & a_{n3}x_3 & + & \cdots & + & a_{nm}x_m & = b_n \end{array}\right\} \qquad (*)$$

can be written in our matrix notation far more simply as

$$\boxed{A\mathbf{x} = \mathbf{b}}$$

where $A = (a_{ij})$ and $\mathbf{x}$ and $\mathbf{b}$ are the column vectors of the x's and b's; we first met this in Equation (11.1).

The "standard" situation is when there are the same number of equations as there are unknowns: $n = m$. In this case A is a square matrix.

If, in this case, A has an inverse matrix A^{-1} then

$$A\mathbf{x} = \mathbf{b} \quad \text{gives} \quad A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}.$$

So

$$I\mathbf{x} = A^{-1}\mathbf{b} \quad \text{or} \quad \mathbf{x} = A^{-1}\mathbf{b}.$$

So, in this case $(*)$ has the *unique* solution

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Here we started by using matrix notation simply as an abbreviation, but the algebra we introduced turned out to be useful; indeed solving a system of linear equations (of size $n \times n$) is very closely related to the the problem of finding the inverse of the corresponding (square) matrix.

11.15. Example. Check that

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 2 & -2 \\ 1 & -1 & 1 \\ 1 & -2 & 3 \end{pmatrix}. \tag{11.4}$$

Hence solve the system of equations

$$\begin{array}{ccccccc} x & + & 2y & & & = & 1, \\ 2x & + & y & + & z & = & 2, \\ x & & & + & z & = & 3. \end{array}$$

Solution The given system can be written in matrix terms as

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

It is easy to verify that the inverse is correct. Thus

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 & 2 & -2 \\ 1 & -1 & 1 \\ 1 & -2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 2 \\ 6 \end{pmatrix}.$$

So the solution to the system is $x = -3$, $y = 2$ and $z = 6$.

In other cases, when for one reason or another we cannot find an inverse, a variety of things can happen. Let me state the basic possibilities without proof.

$\boxed{n = m}$ If $\det(A) \neq 0$ then we have a unique solution, as above. This is the normal case.

If $\det(A) = 0$ then either (1) the system is 'inconsistent' and has no solutions at all, or (2) the system has infinitely many solutions.

As examples of these two cases consider the systems

$$\left.\begin{array}{ccccc} x & + & y & = 1 \\ x & + & y & = 2 \end{array}\right\}, \qquad \left.\begin{array}{ccccc} x & + & y & = 1 \\ 2x & + & 2y & = 2 \end{array}\right\}.$$

In both systems you can check that the corresponding matrix has zero determinant. In the first one the two equations are obviously inconsistent and can have no solutions. In the second one the second equation is really just the same as the first one so the "solution" is $y = 1 - x$ and x can take any value. So there are infinitely many solutions.

$\boxed{n < m}$ Here we have fewer equations than unknowns. It is usually the 'under-determined' case. The equations should not give enough information to fix a solution. So, generally speaking, there will be infinitely many solutions though there are cases where there are no solutions at all.

As examples of these two cases consider the following systems.

$$\left.\begin{array}{ccccccc} x & + & y & + & z & = 1 \\ x & + & 2y & + & z & = 3 \end{array}\right\}, \qquad \left.\begin{array}{ccccccc} x & + & y & + & z & = & 1 \\ x & + & y & + & z & = & 2 \end{array}\right\}.$$

The first system has infinitely many solutions. It is satisfied by $(x, 2, -1 - x)$ for any value of x. The second system is obviously inconsistent and has no solutions at all.

$\boxed{n > m}$ Here we have more equations than unknowns. It is usually the 'over-determined' case. There should normally be no solutions at all because there are too many conditions for the variables to satisfy. In exceptional cases just about anything can happen — unique solution, infinitely many solutions.

Consider the following three systems

$$\left.\begin{array}{rcrl} x & + & y & = 1 \\ x & - & y & = 2 \\ x & + & y & = 2 \end{array}\right\}, \quad \left.\begin{array}{rcrl} x & + & y & = 1 \\ x & - & y & = 2 \\ x & + & y & = 1 \end{array}\right\}, \quad \left.\begin{array}{rcrl} x & + & y & = 1 \\ x & + & y & = 1 \\ x & + & y & = 1 \end{array}\right\}.$$

The first system has no solutions at all; clearly the first and third equations are inconsistent. The second system has precisely one solution; ignoring the irrelevant third equation it is "really" the case with two equations in two unknowns. the third system has infinitely many solutions; there is really only one equation present

11.9 Geometrical Interpretation

As is often the case, although we use algebra as a calculating tool, we use geometry as a way of understanding what is happening. So let us try to get some feel for the above results by looking at some pictures. We are going to consider systems of equations in two unknowns x and y. One such equation $ax + by = c$ represents a straight line in the (x, y)-plane; we exclude the case when $a = b = 0$.

So the solutions to a set of simultaneous equations

$$\begin{aligned} a_1x + b_1y &= c_1 && (l_1) \\ a_2x + b_2y &= c_2 && (l_2) \\ \vdots \quad & \quad \vdots \\ a_nx + b_ny &= c_n && (l_n) \end{aligned}$$

are the points common to the straight lines $l_1, l_2, \ldots, l_n$.

Now let's look at the problem in terms of the number of equations.

One equation in two unknowns: the equation

$$ax + by = c$$

has infinitely many solutions — all the points on a line, since we excluded the case $a = b = 0$ and $c \neq 0$, in which case there are no solutions. The normal case is (rather boringly) illustrated in Fig. 11.1.

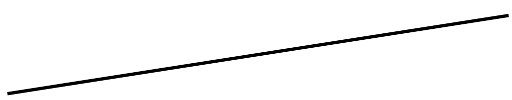


Figure 11.1: Normal solution set of one equation in two unknowns.

Two equations in two unknowns: the equations are

$$\left.\begin{aligned} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{aligned}\right\} .$$

This system represents two straight lines. The *normal* case is when the two lines meet in a single point; when there is precisely one solution to the system, as shown in Fig. 11.2. The *exceptional* cases are:

- when the two lines are parallel, in which case there are no solutions, as illustrated in Fig 11.4; and, even more exceptionally,
- when the two lines are identical so there are infinitely many solutions, as illustrated in Fig 11.3.

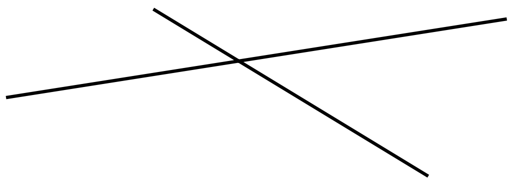


Figure 11.2: Normal solution set of two equations in two unknowns.

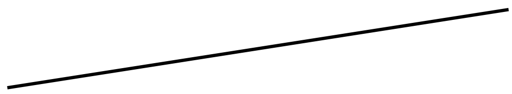


Figure 11.3: Co-incident lines, so many solutions.

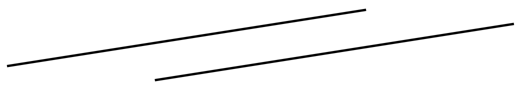


Figure 11.4: Parallel lines so no solutions.

Three equations in two unknowns: here there are three straight lines. You can probably convince yourself quite easily that the *normal* case is that the three lines do not have *any* points in common — so no solutions. Exceptionally the three lines may have just one point in common or otherwise infinitely many points in common. The normal case is illustrated in Fig. 11.5.

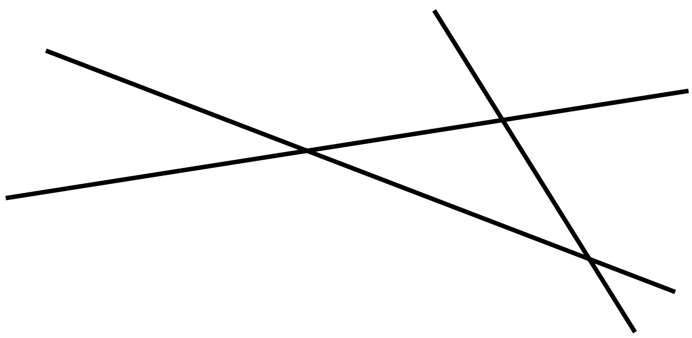


Figure 11.5: Normal solution set of three equation in two unknowns.

11.10 Gaussian Reduction

I now want to show you an efficient routine way for solving systems of simultaneous linear equations. To keep life simple I will concentrate mainly on the normal situation of n equations in n unknowns.

$$\left.\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = b_2 \\ \vdots & & \vdots & & & & \vdots & & & \\ a_{n1}x_1 & + & a_{n2}x_2 & + & a_{n3}x_3 & + & \cdots & + & a_{nn}x_n & = b_n \end{array}\right\}. \qquad (*)$$

The method, known as Gaussian Reduction, is generally much quicker than the method of finding the solution by $\mathbf{x} = A^{-1}\mathbf{b}$. It is very suitable for implementation as a computer routine and is also probably the best method to use by hand unless the system is very small or very simple.

The method is very similar to the kind of thing that we were doing with determinants in Examples 11.10 and 11.11 — trying to get zeros below the diagonal.

There are two stages to the method. The first stage, known as the *reduction*, stage tries to combine the equations so as to get a system with an *upper triangular matrix*. Such systems are easy to solve and the second stage proceeds to solve the system by a process known as *back-substitution*. Another word sometimes used to described the shape of an upper triangular matrix is to say that the system is in *echelon* form.

Let me describe the method in outline with an example. We start with a system like

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = b_1, \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = b_2, \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = b_3, \end{array}$$

and by combining equations in a systematic way we reduce it to the form

$$\begin{array}{cccccc} x_1 & + & A_{12}x_2 & + & A_{13}x_3 & = B_1, \\ & & x_2 & + & A_{23}x_3 & = B_2, \\ & & & & x_3 & = B_3, \end{array}$$

which can now be solved easily by solving the equations in reverse order. The last equation gives x_3. Substituting this value into the second equation gives x_2. And then substituting the values of x_2 and x_3 that we have just found into the first equation gives x_1. This is called *back substitution*.

11.10.1 The Simple Algorithm

The complicated bit is obviously the reduction to upper triangular form. We hope to be able to program this method for a computer, so our approach had better be systematic. Let me first describe the method in its simplest form and then try to improve it a bit and also consider the cases where it fails to work.

Step 1 Start with the first equation. Divide through by a_{11} so as to make the coefficient of x_1 equal to 1. Now, for each equation $i = 2, \ldots, n$, subtract a_{1i} times the first equation from this equation. This will kill off the x_1 term in each of equations $2, \ldots, n$. This is the first stage of the reduction. Note that neither of these operations changes the set of solutions, we are just putting the equations into an equivalent, but more convenient form.

Let me do an example. Consider the system

$$\begin{array}{rcrcrcr} 2x_1 & + & 4x_2 & + & 8x_3 & = & 4, \\ x_1 & + & 3x_2 & - & x_3 & = & 1, \\ 3x_1 & + & 3x_2 & + & 2x_3 & = & 2. \end{array} \tag{11.5}$$

Dividing through the first equation by the coefficient of x_1 we get

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 2, \\ x_1 & + & 3x_2 & - & x_3 & = & 1, \\ 3x_1 & + & 3x_2 & + & 2x_3 & = & 2. \end{array}$$

Now subtract the first equation from the second and 3 times the first from the third:

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 2, \\ & & x_2 & - & 5x_3 & = & -1, \\ & & -3x_2 & - & 10x_3 & = & -4. \end{array} \tag{11.6}$$

Step 2 Now we move on to the second equation and divide through by the coefficient of x_2. We subtract suitable multiples of equation 2 from each of the equations below it so as to kill off their x_2 terms. Because we don't use the first equation, nothing we do brings an x_1 into any of these equations. And again we don't change the set of solutions of the equations

In the example we get, adding 3 times the second equation to the third equation:

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 2, \\ & & x_2 & - & 5x_3 & = & -1, \\ & & & & -25x_3 & = & -7. \end{array} \tag{11.7}$$

Step 3 We carry on in this way. At the i^{th} stage we divide through equation i by the coefficient of x_i and then subtract suitable multiples of the ith equation from the equations below it so as to kill off the x_i term in each of these equations. When we have done $n-1$ stages we will be left with a system in upper triangular form, as in the above example. This upper triangular system is equivalent to the original one $(*)$ in the sense that both sets of equations have the same solutions.

Step 4 Now solve by back-substitution. For the given example we get $x_3 = 7/25$ from the last equation. The second equation now becomes

$$x_2 = 5x_3 - 1 = 2/5$$

and the first equation becomes

$$x_1 = 2 - 2x_2 - 4x_3 = 2 - 4/5 - 28/25 = 2/25.$$

So the solution to our system is $x_1 = 2/25, x_2 = 2/5, x_3 = 7/25$.

11.16. Example. Consider the following system of equations in which t is a parameter.

$$\begin{array}{rcrcrcl} x & - & y & + & z & = & 2, \\ 4x & - & 5y & + & (t+4)z & = & 9, \\ 3x & + & (t-3)y & - & z & = & 8. \end{array}$$

For what values of t does the system have (i) a unique solution, (ii) no solution, and (iii) infinite number of solutions. In case (iii), solve for x and y in terms of z.

Solution We follow the steps just described. Step 1 uses the first equation to eliminate the x - terms in the second and third equations, to give

$$\begin{array}{rcrcrcl} x & - & y & + & z & = & 2, \\ & - & y & + & tz & = & 1, \\ & + & ty & - & 4z & = & 2. \end{array}$$

Now move to Step 2 and the second equation to eliminate y from the subsequent equations, giving

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & 2, \\ & & y & - & tz & = & -1, \\ & & & - & (4-t^2)z & = & 2+t. \end{array}$$

This is now in reduced echelon form. Consider the third equation, $(4-t^2)z = 2+t$. If $t = -2$, this reduces to the trivial equation $0z = 0$ and z is unrestricted. In that case, we use Step 4 to get $y = -1-2z$ and then $x = 1-3z$.

It remains to consider what happens when $t \neq -2$, in which case the third equation becomes $(2-t)z = 1$. If $t = 2$, we must have $0z = 1$, and there are no solutions. However if $t \neq 2$, we have $z = 1/(t-2)$ and then back substituting first for y and then for x as described in Step 4 gives a unique solution.

11.10.2 Complications

One way to find out the difficulties that can occur with this algorithm is to write it as computer (pseudo-) code. Writing the actual code is an interesting exercise, although in Matlab, this has all been done for you.

In the simple case, the reduction step is performed as follows (written in pseudo-code):

```
 for equation i=1 to equation (n-1) :
     divide through equation i by a(i,i)
     for equation j=(i+1) to equation n
         subtract a(j,i) times equation i from equation j
     end-for
end-for
```

The back-substitution is then done as follows

```
x(n) = b(n)/a(n,n)
for i = (n-1) down to 1
      x(i) = b(i) - a(i,i+1).x(i+1) - ... - a(i,n).x(n)
end-for
```

This is an accurate coding of the method described in Section 11.10.1. As it stands it may fail to work on some perfectly reasonable systems. The first place where it can fail is at line `2`, where we might find ourselves dividing by zero (see also line `7`).

To avoid this difficulty we adopt the process known as **partial pivoting**. This means replacing line `2` by the following lines

```
2a.    Let j be such that abs(a(j,i)) is the largest of
2b.    abs(a(i,i)), abs(a(i+1,i), ..., abs(a(n,i))
2c.    If j is not i then swap equations i and j.
2d.    If all these elements are zero then FAIL.
```

In practice we usually replace line `2d` by '`if all these magnitudes are less than EPS`', where `EPS` is some very small quantity like 10^{-10}. This is to cover for the possibility that, due to computer rounding errors, a number that theoretically should be zero actually shows up as a very small quantity.

11.10.3 Solving Systems in Practice

One reason to study this algorithm is because you may find it useful in solving small systems of equations by hand. We have already done this, for example with the system in Equation (11.5).

A second reason is to understand the sort of algorithm that computer algebra systems such as Matlab will use. If you just want to solve a system of equations, you would use Matlab's built-in routine. My Matlab book suggests that

```
>> A = [ 2 4 8;1 3 -1;3 3 2];
>> B = [4;1;2]
>> X = inv(A) * B
```

will produce a solution to the system of equations (11.5). You can even use `X = A\B`, which just applies the Gaussian reduction algorithm, and avoids actually computing the inverse.

A final reason is that you may wish to run through the algorithm yourself on a realistic size of problem, doing the individual calculations by machine. Although not exactly the same, you will find a very similar trace for the system (11.5) starting on page 118.

Here is another example to practice handling systems of equations.

11.17. Example. Consider the following system of equations, in which s and t are real parameters

$$\begin{array}{rcrcrcl} x & - & 2y & + & z & = & 9, \\ 2x & - & 3y & - & z & = & 4, \\ -x & + & (2+t)y & + & 3z & = & s-14t. \end{array}$$

By using row operations on the associated matrix, obtain a row-equivalent triangular form suitable for solution by back substitution.

1. Find the value of t for which the system either has no solution or does not have a unique solution. How does the behaviour of the system depend on s in these cases?

2. Use back substitution to solve the system in the case that $t = -1$ and $s = -5$.

Solution Here is a machine trace of the solution (in MAPLE). This way I know it is correct and the typesetting is eased. You are expected to be able to do questions of this type "by hand".

```
> A:=matrix(3,4,[1,-2,1,9,2,-3,-1,4,-1,2+t,3,s-14*t]);
```

$$A := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 2 & -3 & -1 & 4 \\ -1 & 2+t & 3 & -14\,t+s \end{bmatrix}$$

```
> B:=pivot(A,1,1);
```

$$B := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 0 & 1 & -3 & -14 \\ 0 & t & 4 & 9-14\,t+s \end{bmatrix}$$

```
> C:=addrow(B,2,3,-t);
```

$$C := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 0 & 1 & -3 & -14 \\ 0 & 0 & 3\,t+4 & 9+s \end{bmatrix}$$

From the last line of the reduction, we see there is a unique solution unless the coefficient of z in the last equation becomes zero, in which case, the equation becomes $0.z = 9 + s$. Thus there are no solutions when $t = -4/3$ unless $s = -9$. If $s = -9$, we have an infinite family of solutions of the form

$$y = 3z - 14, \qquad x = 9 + 2y - z = 5z - 19,$$

where z is a free parameter.

If $t = -1$ and $s = -5$ the last equation becomes $z = 4$. Using back substitution, we see first that $y = -2$ and then that $x = 1$.

11.11 Calculating Inverses

An extension of the ideas used in the Gaussian reduction algorithm discussed in Section 11.10.1 for solving a system of simultaneous equations leads almost immediately to an effective algorithm for computing inverses. The first step is to note that we did not need to stop work with the system of equations (11.7) when it was reduced to an upper triangular form. Continuing in the same way, the next step is to divide the last equation by the coefficient of x_3 to give:

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 2, \\ & & x_2 & - & 5x_3 & = & -1, \\ & & & & x_3 & = & -7/25. \end{array}$$

The idea now is to work back up the set of equations, killing off the remaining non-diagonal terms. Thus we add 5 times the third equation to the second, and -4 times the third equation to the first, to get:

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & & = & 22/25, \\ & & x_2 & - & & = & 2/5, \\ & & & & x_3 & = & -7/25. \end{array}$$

The last step is to eliminate the final off-diagonal term by adding −2 times the second equation to the first to give

$$\begin{aligned} x_1 \qquad\qquad &= \ 2/25, \\ x_2 \qquad &= \ 2/5, \\ x_3 &= -7/25. \end{aligned}$$

Of course we can now read off the solution immediately, so this is an alternative (and rather slower) way of doing the back substitution.

There is an alternative way to present this process, which shows that it also calculates the inverse of the coefficient matrix. Our concern so far has been that the allowed *row operations* did not change the solution set of the system of equations. Now that I am less concerned directly with equations, I want to introduce a more concise notation, writing the system of equations (11.5) using the *augmented matrix*

$$(A|b) = \begin{pmatrix} 2 & 4 & 8 & | & 4 \\ 1 & 3 & -1 & | & 1 \\ 3 & 3 & 2 & | & 2 \end{pmatrix}.$$

Our first step was to multiply the first row by 1/2 to get

$$A_1 = \begin{pmatrix} 1 & 2 & 4 & | & 2 \\ 1 & 3 & -1 & | & 1 \\ 3 & 3 & 2 & | & 2 \end{pmatrix}.$$

This can be interpreted in terms of matrix algebra; let me do the same *operation* on the identity matrix I_3 to get the *elementary matrix*

$$E_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A calculation then shows that $E_1(A|b) = A_1$. We see that in this case, in order to do a row operation on a matrix A, we can first do the required operation on the identity matrix, and then pre-multiply by the resulting elementary matrix E. This fact is general; all the row operations we shall need can be obtained by pre-multiplying by an elementary matrix obtained by doing the required row-operations on the identity matrix.

The next step taken was to subtract the first equation from the second and 3 times the first from the third. The resulting elementary matrices are

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}.$$

Computing, we see that

$$E_3.E_2.E_1.A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & -5 \\ 0 & -3 & -10 \end{pmatrix},$$

as expected from equations (11.6). Continuing in this way, we find a sequence of elementary matrices E_n, $E_{n-1}, \ldots, E_1$ such that $E_n.E_{n-1}.\ldots.E_1.A = I_n$. Let

$$B = E_n.E_{n-1}.\ldots.E_1;$$

we have just shown that $B.A = I_n$ and so B is the inverse for A.

It remains to find a convenient way of recording the elementary matrices used to reduce A to the identity I_n, and then computing their product. It turns out that we can do both things at the same time. As before we write an augmented matrix as

$$[A|I_3] = \begin{pmatrix} 2 & 4 & 8 & | & 1 & 0 & 0 \\ 1 & 3 & -1 & | & 0 & 1 & 0 \\ 3 & 3 & 2 & | & 0 & 0 & 1 \end{pmatrix}.$$

We now manipulate A exactly as before; but by having I_3 in the augmented matrix, we record the product of the elementary matrices used to perform the reduction on A. The remainder of the calculation is done, as an illustration, in a computer algebra package; as noted in the forward, I have used MAPLE rather than MATLAB .

```
> with(linalg):
> A:=matrix(3,3,[2,4,8,1,3,-1,3,3,2]);
```

$$A := \begin{bmatrix} 2 & 4 & 8 \\ 1 & 3 & -1 \\ 3 & 3 & 2 \end{bmatrix}$$

```
> A0:=concat(A,diag(1,1,1));
```

$$A0 := \begin{bmatrix} 2 & 4 & 8 & 1 & 0 & 0 \\ 1 & 3 & -1 & 0 & 1 & 0 \\ 3 & 3 & 2 & 0 & 0 & 1 \end{bmatrix}$$

```
> A1:=mulrow(A0,1,1/2);
```

$$A1 := \begin{bmatrix} 1 & 2 & 4 & \frac{1}{2} & 0 & 0 \\ 1 & 3 & -1 & 0 & 1 & 0 \\ 3 & 3 & 2 & 0 & 0 & 1 \end{bmatrix}$$

```
> A2:=pivot(A1,1,1);
```

$$A2 := \begin{bmatrix} 1 & 2 & 4 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -5 & \frac{-1}{2} & 1 & 0 \\ 0 & -3 & -10 & \frac{-3}{2} & 0 & 1 \end{bmatrix}$$

```
> A3:=pivot(A2,2,2);
```

$$A3 := \begin{bmatrix} 1 & 0 & 14 & \frac{3}{2} & -2 & 0 \\ 0 & 1 & -5 & \frac{-1}{2} & 1 & 0 \\ 0 & 0 & -25 & -3 & 3 & 1 \end{bmatrix}$$

```
> A4:=mulrow(A3,3,-1/25);
```

$$A4 := \begin{bmatrix} 1 & 0 & 14 & \frac{3}{2} & -2 & 0 \\ 0 & 1 & -5 & \frac{-1}{2} & 1 & 0 \\ 0 & 0 & 1 & \frac{3}{25} & \frac{-3}{25} & \frac{-1}{25} \end{bmatrix}$$

```
> A5:=pivot(A4,3,3);
```

$$A5 := \begin{bmatrix} 1 & 0 & 0 & \frac{-9}{50} & \frac{-8}{25} & \frac{14}{25} \\ 0 & 1 & 0 & \frac{1}{10} & \frac{2}{5} & \frac{-1}{5} \\ 0 & 0 & 1 & \frac{3}{25} & \frac{-3}{25} & \frac{-1}{25} \end{bmatrix}$$

```
> B:=delcols(A5,1..3);
```

$$B := \begin{bmatrix} \frac{-9}{50} & \frac{-8}{25} & \frac{14}{25} \\ \frac{1}{10} & \frac{2}{5} & \frac{-1}{5} \\ \frac{3}{25} & \frac{-3}{25} & \frac{-1}{25} \end{bmatrix}$$

```
> multiply(A,B);
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that, although the procedure was described as first reducing to triangular form, and then removing the remaining off-diagonal elements using the 1's on the diagonal, in practice it is easy to clear *all* the non-diagonal elements in a given column as they are first met. This operation is known as (full) pivoting.

11.18. Example. Let $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 5 \end{pmatrix}$. Show that $A^{-1} = \begin{pmatrix} 2 & -7 & 1 \\ -1 & 6 & -1 \\ 1 & -5 & 1 \end{pmatrix}$.

Solution We perform row operations on the augmented matrix

$$\begin{pmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ -1 & 3 & 5 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 5 & 6 & 1 & 0 & 1 \end{pmatrix} \qquad [R_3' = R_3 + R_1],$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & -1 & 1 & -2 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -5 & 1 \end{pmatrix} \qquad [R_1' = R_1 - 2R_2;\ R_3' = R_3 - 5R_2],$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 2 & -7 & 1 \\ 0 & 1 & 0 & -1 & 6 & -1 \\ 0 & 0 & 1 & 1 & -5 & 1 \end{pmatrix} \qquad [R_1' = R_1 + R_3;\ R_2' = R_2 - R_3].$$

11.19. Example. Let $A = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -4 & 2 \\ 3 & -1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$. Compute A^{-1} and B^{-1} using row operations. Verify your answer both directly (by calculating AA^{-1} etc.) and using the determinant formula.

Solution Computing with the augmented matrices

$$\begin{pmatrix} 5 & -3 & 1 & 0 \\ -3 & 2 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -3/5 & 1/5 & 0 \\ 0 & 1/5 & 3/.5 & 1 \end{pmatrix} \qquad [R_1' = R_1/5; R_2' = R_2 + 3R_1']$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 3 & 5 \end{pmatrix} \qquad [R_2' = 5R_2; R_1' = R_1 + 3R_2/5].$$

11.20. Example. As a final example, we apply the same technique to the matrix in Equation 11.4 on page 109, and hence obtain the inverse.

Solution

```
> with(linalg):
> A:=matrix(3,3,[1,2,0,2,1,1,1,0,1]):A0:=concat(A,diag(1,1,1));
```

$$A0 := \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

```
> A1:=pivot(A0,1,1);
```

$$A1 := \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & -3 & 1 & -2 & 1 & 0 \\ 0 & -2 & 1 & -1 & 0 & 1 \end{bmatrix}$$

```
> A2:=mulrow(A1,2,-1/3);
```

$$A2 := \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & -2 & 1 & -1 & 0 & 1 \end{bmatrix}$$

```
> A3:=pivot(A2,2,2);
```

$$A3 := \begin{bmatrix} 1 & 0 & \frac{2}{3} & \frac{-1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{-2}{3} & 1 \end{bmatrix}$$

```
> A4:=mulrow(A3,3,3);
```

$$A4 := \begin{bmatrix} 1 & 0 & \frac{2}{3} & \frac{-1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & 1 & 1 & -2 & 3 \end{bmatrix}$$

```
> A5:=pivot(A4,3,3);
```

$$A5 := \begin{bmatrix} 1 & 0 & 0 & -1 & 2 & -2 \\ 0 & 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 & -2 & 3 \end{bmatrix}$$

Questions 2 (Hints and solutions start on page 163.)

11.2. Q. a) Define matrices A, B and C by

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ -1 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 2 & -1 \end{pmatrix}.$$

Calculate those of the following expressions that can be evaluated and explain why the others cannot be:

$$A^{-1}, \qquad AB, \qquad BA, \qquad CB, \qquad C^{-1}.$$ **[5 marks]**

b) What is the *transpose*, A^T of a matrix A? Find all 2×2 matrices A such that

$$A \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot A^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$ **[5 marks]**

c) By first using column operations involving subtracting the first column, find the determinant of the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1+a & 1 & 1 \\ 1 & 1 & 1+b & 1 \\ 1 & 1 & 1 & 1+c \end{pmatrix}$. **[5 marks]**

11.3. Q. a) Matrices A and B are defined by

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

and the transpose of B is written as B^{T}. Calculate those of the following expressions that can be evaluated and explain why the others cannot be evaluated:

$$A^2, \qquad AB, \qquad BA, \qquad B^{-1}, \qquad B^{\mathrm{T}}B.$$ **[5 marks]**

b) Using row operations, compute the inverse of the matrix

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$ **[5 marks]**

11.4. Q. a) Define matrices A, B and C by

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

Calculate those of the following expressions that can be evaluated and explain why the others cannot:

$$AB, \qquad BA, \qquad \det(AB), \qquad BC, \qquad (CB)^{-1}.$$ **[5 marks]**

b) What is the *transpose*, A^T of a matrix A? Find all 2×2 matrices $A = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$ such that

$$A \cdot A^T = \begin{pmatrix} 4 & 6 \\ 6 & 25 \end{pmatrix}.$$ **[5 marks]**

11.5. Q. a) Find all the values of λ for which

$$\begin{vmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1-\lambda \end{vmatrix} = 0.$$ **[10 marks]**

b) Consider the following system of equation, in which α is a real parameter.

$$\begin{aligned} x_1 + \ x_2 \qquad\quad &= 3, \\ 2x_1 + \ x_2 + \ x_3 &= 7 - \alpha, \\ x_1 + 2x_2 + \alpha x_3 &= 10. \end{aligned}$$

By using row operations on the associated matrix, obtain a row-equivalent echelon form.

1. Find the value of α for which the system is *inconsistent*.
2. Solve the system in the case that $\alpha = 2$, explaining carefully how you obtain the solution from the echelon form.

[10 marks]

11.6. Q. a) Find the determinant of the matrix $A = \begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix}$ where λ is a constant. For what values of λ is A invertible? **[4 marks]**

b) Consider the following system of equations, in which s and t are real parameters

$$\begin{array}{rcrcrcl} x & - & 2y & + & z & = & 9, \\ 2x & - & 3y & - & z & = & 4, \\ -x & + & (2+t)y & + & 3z & = & s-14t. \end{array}$$

By using row operations on the associated matrix, obtain a row-equivalent triangular form suitable for solution by back substitution.

1. Find the value of t for which the system either has no solution or does not have a unique solution. How does the behaviour of the system depend on s in these cases?
2. Use back substitution to solve the system in the case that $t=-1$ and $s=-5$.

[11 marks]

11.7. Q. a) By first using column operations, including subtracting the first column from other columns, find the determinant of the matrix $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 7 \\ 1 & 4 & 7 & 10 \\ 1 & 5 & 9 & 13 \end{pmatrix}$. **[5 marks]**

b) Consider the following system of equations in which t is a parameter.

$$\begin{array}{rcrcrcl} x & - & y & + & z & = & 2, \\ 4x & - & 5y & + & (t+4)z & = & 9, \\ 2x & + & ty & - & z & = & 5. \end{array}$$

For what values of t does the system have (i) a unique solution, (ii) no solution, and (iii) infinite number of solutions.

In case (iii), solve for x and y in terms of z. **[10 marks]**

11.8. Q. Use row - reduction of an augmented matrix to calculate the inverse of the matrix

$$A = \begin{pmatrix} 1 & 6 & 2 \\ -1 & 3 & 2 \\ -2 & -1 & 1 \end{pmatrix}.$$

[10 marks]

11.9. Q. Evaluate $(x-y)(x^2+xy+y^2)$ and simplify your answer. By using column operations, show that

$$\begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ x^3 & y^3 & z^3 \end{vmatrix} = k(x-y)(y-z)(z-x)(x+y+z)$$

where k is a constant that you should find. **[10 marks]**

11.10. Q. Using row operations find the inverse of the matrix

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix}.$$

Explain briefly in terms of elementary matrices *why* the method works. **[10 marks]**

Chapter 12

Approximation and Taylor Series

12.1 Introduction

Approximations are very important in mathematics, particularly in its applications. There are very many problems that we do not know how to solve exactly but for which it is comparatively easy to get an approximate answer. These are not ideal from a purely mathematical point of view but in practical work there is very little difference between knowing that the answer to a problem is $\sqrt{2}$ and knowing that it is approximately 1.414214 to 6 decimal places.

For example, there are no formulas that will allow you to write down the exact solutions of even such apparently simple equations as $x^8 + 3x^5 - 2x + 1 = 0$, $\sin x = x \cos x$, $e^x = 5x$. With these it is approximation or nothing.

There is a related problem that is just as important and which we will consider later. How do you actually work out the values of functions like the trig functions or the exponential? If you want to know $e^{2.3}$ or $\sin 2.1$ then you press a button on your calculator. But how does the calculator know how to do it? In a strict sense it doesn't. All it knows is how to obtain the value to the level of accuracy required by the display of the calculator. So this is once more an approximation problem.

12.2 Accuracy

Since we are going to be working a lot with approximations in this chapter I had better say a thing or two about terminology.

I will often make statements like this: *the solution to the equation is* 4.453 *to 3 decimal places*. What exactly does this mean? When I use such a statement I always mean that the result has been *rounded*. In other words, I am saying that the true solution is somewhere between 4.45**25** and 4.45**35**. In other words the expression really refers to a range of numbers rather than a specific number—don't forget this.

For example, the number 23.342643 rounded to 3 decimal places is 23.343 and rounded to 4 decimal places is 23.3426.

To get you more used to this idea let me try to do a simple calculation. Suppose I know that x has the value 2.34 rounded to 2 decimal places and y has the value 0.23 rounded to two decimal places. What can I say about the value of x/y? If I just plug the numbers mindlessly into my calculator I get $x/y = 10.173913$.

This is actually rubbish. Let me now do the calculation more carefully. All that we know is that the value of x lies somewhere between 2.335 and 2.345 and that the value of y lies somewhere between 0.225 and 0.235.

The biggest value that x/y could take is obtained by dividing the biggest possible value of x by the smallest possible value of y. This gives $2.345/0.225 = 10.42222$.

On the other hand, the smallest value that x/y could take is obtained by dividing the smallest value that x can take by the biggest value that y can take. This gives $2.335/0.235 = 9.93617$.

Anything between these two extremes is a possible value for x/y ! So we really do not know very much about the value of x/y and the best that we can say in simple terms is that $x/y = 10$, to zero decimal places. This is known as the problem of dividing by small numbers and can be a pest.

12.3 Linear Approximation

Recall the definition of the derivative of a function:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

As h gets very small we would expect the value of the Newton Quotient to give a good approximation to the value of the derivative:

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}$$

To get an idea of the accuracy look at the following example where I am approximating the value of the derivative of e^x at $x = 0$. The correct answer is 1. The Newton Quotient is

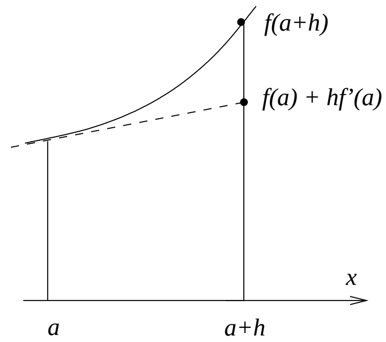


Figure 12.1: Linear approximation

$$\frac{e^{0+h} - e^0}{h} = \frac{e^h - 1}{h}$$

h	Newton Quotient
0.1	1.0517
0.01	1.0050
0.001	1.0005

I am now going to look at this approximation from a rather different point of view. Rather than regarding it as a way of approximating the value of the derivative I am going to assume that we know the derivative and really want to find $f(a+h)$. We can rearrange our approximation as follows:

$$\boxed{f(a+h) \approx f(a) + hf'(a)} \quad \text{Linear Approximation}$$

This now allows us to approximate the value of $f(a+h)$, when h is small, in terms of the values of $f(a)$ and $f'(a)$.

The Fig. 12.1 shows you what the linear approximation is doing. It is estimating the value of $f(a+h)$ by assuming that, near to a, $f(x)$ can be replaced by the straight line through $(a, f(a))$ with slope $f'(a)$.

12.1. Example. What, approximately, is $\sqrt{9.1}$?

Let us assume that we cannot cheat by using a square root button on our calculator!

The trick here is to notice that 9.1 is quite close to 9 and we certainly know $\sqrt{9}$ without using a calculator. In other words, we can write $9.1 = 9+h$, where $h = 0.1$. Let $f(x) = \sqrt{x}$. Then $f'(x) = \dfrac{1}{2\sqrt{x}}$. So the linear approximation formula becomes

$$\sqrt{x+h} \approx \sqrt{x} + \frac{h}{2\sqrt{x}}$$

This will be 'true' for any positive value of x so long as h is small enough. In our case we want to put $x = 9$ and $h = 0.1$. Then we get

$$\sqrt{9.1} \approx \sqrt{9} + \frac{0.1}{2\sqrt{9}} = 3 + \frac{0.1}{6} = 3.01667$$

The answer given by my calculator is $\sqrt{9.1} = 3.0166206$. So our approximation is quite reasonable. The error is 0.00005.

If we use the method to estimate $\sqrt{9.8}$ instead ($h = 0.8$) we get the answer 3.13333 compared to the correct answer 3.130495. Here the error is much greater. We are getting to the limit of reasonable values of h.

The real problem with this method is that we have no way of knowing how accurate the result is. I could only check the above example because I could get the 'exact' answer from my calculator to compare with it. Estimation of accuracy is something that we will have to consider later.

12.2. Example. What, approximately, is the value of $\sin 10°$?

We have to be a bit careful here and remember that all of our calculus work on the trig functions has been done in radians. We had better convert at once. $10°$ is 0.17453 radians.

It is now more clear that we are dealing with a 'small' angle. Let's try to use linear approximation with $f(x) = \sin x$, $x = 0$ and $h = 0.17453$. The derivative of $\sin x$ is $\cos x$ (since we are now working in radians). So

$$\sin 0.17453 = \sin h \approx \sin 0 + h \cos 0 = 0.17453$$

The true answer is 0.173648, to 6 decimal places. So we have a moderately good approximation in this case.

Use the same method to approximate $\sin 5°$ for yourself. Then do $\cos 10°$.

12.3. Example. Show that if x is very small then $\dfrac{1+x}{1-x} \approx 1 + 2x$.

This is a simple example of an important type of problem. We have a rather complicated function and we want to know how it behaves near a certain point. We might be able to do this by finding a much simpler function that is a good approximation to it near this point.

In this example the approximation allows us to estimate that a small change in the (small) value of x will produce roughly double that change in the value of the function (because of the $2x$).

Now let me obtain the desired approximation. Let

$$f(x) = \frac{1+x}{1-x}, \quad \text{then} \quad f'(x) = \frac{2}{(1+x)^2}$$

Our linear approximation formula says that if h is very small then

$$f(0+h) \approx f(0) + hf'(0)$$

In this case $f(0) = 1$ and $f'(0) = 2$. So the linear approximation becomes

$$\frac{1+h}{1-h} \approx 1 + 2h$$

Changing notation by writing x for h gives us the answer that we wanted.

Let's experiment with this approximation, to find out how good it is.

x	$1+2x$	$f(x)$	error
0.2	1.4	1.5	0.1
0.1	1.2	1.22222	0.02222
0.05	1.1	1.10526	0.00526
0.01	1.02	1.02020	0.00020

These results are not untypical. The linear approximation is usually quite rough unless h is very small indeed.

12.3.1 Small Changes

As you have seen above, the linear approximation gives us a way of finding, approximately, the effect on $f(x)$ of a small change in x.

All that I am going to do in this section is change notation.

We often write δx for a small change in x and δf for the corresponding change in $f(x)$. The linear approximation says

$$f(x + \delta x) \approx f(x) + \delta x f'(x)$$

or, even more simply,

$$\delta f = \delta x f'(x)$$

or, for $y = y(x)$,

$$\delta y = \frac{dy}{dx}\delta x$$

Get used to recognising the linear approximation in these various forms. They are all the same—it is only the notation that is different.

12.4. Example. A surveyor wants to measure the distance AB (see Fig 12.2). He does this by moving out distance 100m to C and measuring the angle ACB. He finds that the angle is 70°, but his apparatus is only capable of measuring angles to the nearest degree. How accurately can he find the distance AB?

If the angle measured is θ then the distance AB is $d(\theta) = 100 \tan\theta$. With $\theta = 70°$ this gives $AB = 274.75m$. But we know that the 'real' angle could have been anywhere between 69.5° and 70.5°. How might this affect the answer?

If we use the first approximation we get $\delta d \approx 100 \sec^2\theta\delta\theta$ (θ in radians!). In this case $\theta = 1.22173$ and $\delta\theta$ is one half of a degree or, in radians, $\delta\theta = 0.008727$.

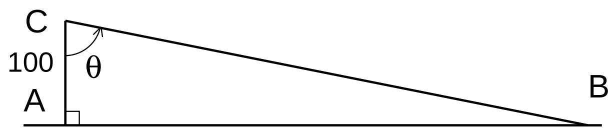


Figure 12.2: A Surveyor measuring distance

So $\delta d \approx 100 \sec^2(1.22173) \times 0.008727 = 7.460$. So the surveyor really ought to quote his result for the distance as $d = 274.75 \pm 7.46$ or, to be more reasonable and honest, $d = 275 \pm 8m$. The half degree error in the angle has produced a quite significant error in the distance.

Let's check our answer by working with the exact formula. What are the actual values of $d(69.5°)$, $d(70°)$ and $d(70.5°)$? Just plug in the numbers and get the answers 267.5, 274.75 and 282.4 respectively. This fit quite nicely with our approximation.

12.4 Solving Equations—Approximately

As I said above, solving equations is a very important process in mathematics and its applications. Indeed, at an elementary level mathematics is almost identified with 'solving equations'.

There are very few kinds of equations for which we know how to write down the 'exact' solution. For the majority of equations that arise in the applications of mathematics the best we can hope for is an approximation to a solution. For this reason the problem of finding approximations to the solutions of equations is a very important branch of mathematics.

In this section I am just going to look at one method—in practice the most important one (in the form I give it and in its generalisations). The method was invented by Newton soon after he invented the calculus and is basically an application of the Linear Approximation.

The basic idea is quite simple. Suppose that we are trying to find a solution of the equation $f(x) = 0$. Suppose that we are able to make a reasonably good guess at the answer, say $x = a$. In other words we are claiming that there is a solution to $f(x) = 0$ quite near to $x = a$. Suppose that we write the solution as $s = a + h$ where our assumption is that h is quite small. Then, using the linear approximation, we say

$$0 = f(s) = f(a + h) \approx f(a) + hf'(a).$$

The thing that we don't know is h. So we 'solve' this equation for h to get

$$h \approx -\frac{f(a)}{f'(a)}$$

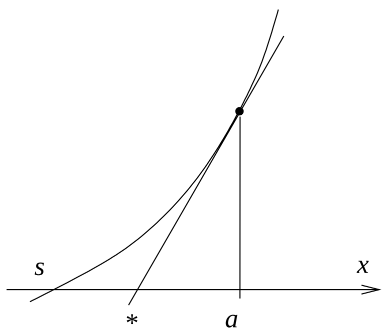


Figure 12.3: The tangent helps find where the curve cuts the axis.

We cannot say that $a + h$ is the required solution, because we are only working with an approximation, but we can reasonably hope that $a + h$ is a better approximation to the solution than a was. So this is really a process for trading in an approximation for a better one.

The process can be shown geometrically as follows. The graph of $f(x)$ is being 'replaced' by the graph of the tangent at $(a, f(a))$. We then find where the tangent cuts the x-axis. If the picture is to be believed, this will give us a point much closer to the solution $x = s$ than a was.

We do not stop there. We start again, taking the new approximation as our starting point, and use the method to produce an even better approximation. This process can be repeated over and over again, producing (we hope) a sequence of steadily improving approximations. Indeed, it seems that if we keep going for ever then we will achieve the true solution as the limiting case of our sequence of approximations. In practice we just repeat the process often enough to get the solution to the level of accuracy that we happen to want.

12.4.1 Newton's Method

Let me now gather together the arguments of the previous section into an explicit method. It is called the Newton (or Newton-Raphson) method.

We want to find an approximation to a solution of the equation $f(x) = 0$. We know that $x = x_0$ is a reasonable approximation to the required answer.

Form a sequence $\{x_1, x_2, \ldots, x_n, \ldots\}$ of approximations by the following scheme:

$$\boxed{x_{n+1} = x_n + h_n} \quad \text{where} \quad \boxed{h_n = -\frac{f(x_n)}{f'(x_n)}}$$

Continue the process until the 'corrections' h_n get so small that you are happy to accept that you have obtained the required solution to the accuracy that you want.

You should usually find that if you are fairly close to the answer the number of correct decimal places will just about *double* at each stage of the process. The method has *quadratic* convergence, except in certain special cases.

It is possible that the method will fail. It may do so, very occasionally, because the function is unsuitable or it may be that your initial guess is too unreasonable. Failure will usually show itself by the fact that the h_n do not shrink rapidly. One particular problem, which is not really a failure but can be a nuisance, is where the derivative of the function is zero at the point where the function becomes zero—the graph *touches* the x-axis rather than *cutting* it. In this situation you should get convergence to the solution, but it will be very slow.

12.5. Example. Find the solution to $x = \cos x$ that lies between $x = 0$ and $x = \pi/2$. Give the answer to 4 decimal places. (x is in radians.)

If we draw the graphs of $y = x$ and $y = \cos x$ on the same axes we can see that there is going to be a single solution (where the graphs cross) in the given range and that it is probably somewhere near $\pi/4$. I will take $x_0 = 1$ as my first guess at the solution. Don't waste effort trying to find a very good first approximation.

The next important point is that Newton's method works for equations of the form $f(x) = 0$—i.e. it is a way of finding where graphs cut the x-axis. Our equation has to be put into this form before we go any further. There are lots of ways to do this. I will use the most obvious, which is to let $f(x) = x - \cos x$.

We now need the derivative of $f(x)$ which is $f'(x) = 1 + \sin x$.

Now we are ready to calculate. The calculation scheme is:

$$x_{n+1} = x_n + h_n \quad \text{where} \quad h_n = -\frac{f(x_n)}{f'(x_n)} = \frac{x_n - \cos x_n}{1 + \sin x_n}$$

and we are starting with $x_0 = 1$.

So

$$h_0 = -(1 - \cos 1)/(1 + \sin 1) = 0.249636 \qquad x_1 = x_0 + h_0 = 0.750364.$$

Do it again,

$$h_1 = -(x_1 - \cos x_1)/(1 + \sin x_1) = 0.011251 \qquad x_2 = x_1 + h_1 = 0.739113.$$

And again,

$$h_2 = -(x_2 - \cos x_2)/(1 + \sin x_2) = 0.000028 \quad x_3 = x_2 + h_2 = 0.739085.$$

Notice that h is beginning to get quite small and the value of x is not changing very much from one step to the next. Let's do another round:

$$h_3 = -(x_3 - \cos x_3)/(1 + \sin x_3) = 1.8 \times 10^{-10} \qquad x_4 = x_3 + h_3 = 0.739085.$$

h really has become very small indeed. Notice that, to 6 decimal places, our approximation did not change at all on the last step. If you go round once more you will find that h has become even more dramatically small—in fact your calculator will probably give it as 0.

Since we only wanted to know the answer to 4 decimal places we can reasonably stop at this point and claim that the required solution is $s = 0.7391$, to 4 decimal places. This is actually correct.

In practice it is usually convenient, when doing the calculations on paper rather than on a computer, to lay out your results in the form of a table:

n	x_n	$-h_n$
0	1.000000	0.249636
1	0.750364	0.011251
2	0.739113	0.000028
3	0.739085	1.8×10^{-10}
4	0.739085	

This makes it easy to see what is happening. Note that I have done all the calculations to 6 decimal places. It is usually wise to do the calculations to one or two places more than you want for the eventual answer.

12.6. Example. Find $\sqrt{90}$ to 5 decimal places.

What has this go to do with solving equations? Well, we can recast the question as: find the positive solution to the equation $x^2 = 90$. Now we are solving equations.

We will use the Newton method with $f(x) = x^2 - 90$. What will we take as first guess? The answer is somewhere between 9 and 10 and probably closer to 9 than to 10. Without thinking much harder about it let's take $x_0 = 9.4$.

The derivative is $f'(x) = 2x$ and our scheme becomes

$$x_{n+1} = x_n + h_n \quad \text{where} \quad h_n = -\frac{x_n^2 - 90}{2x_n}$$

You can check that you agree with the following numbers.

n	x_n	h_n
0	9.400000	0.087234
1	9.487234	-0.000401
2	9.486833	-8.5×10^{-9}
3	9.486833	

So we would expect that the solution, to 5 decimal places, is $\sqrt{90} = 9.48683$.

Perhaps it is worth, for once, checking our answer. We have to be a bit careful how we do this. The fact that $9.48683^2 = 89.999943$ does not, in itself, prove anything. A better argument goes as follows. By saying that the answer is 9.48683 to 5 decimal places we are really saying that the answer is *somewhere* between $a = 9.4868\mathbf{25}$ and $b = 9.4868\mathbf{35}$. Now $a^2 = 89.99985$ and $b^2 = 90.000038$. We know that x^2 increases as x increases, so we were right in claiming that $\sqrt{90}$ had to lie between a and b. So our answer is correct.

12.7. Example. Solve the equation $2x^3 + 3x^2 - 6x - 4 = 0$, giving the results to 4 decimal places.

The real problem that we face here is that an equation like this may have more than one solution. We will have to cope with this problem when we come to it.

Let $f(x) = 2x^3 + 3x^2 - 6x - 4$. We start to experiment a bit. $f(0) = -4$, $f(1) = -5$, $f(2) = 12$. Ah! so there is an answer between 1 and 2. $f(1.5) = 0.5$, so there is an answer between 1 and 1.5 and it is probably closer to 1.5 than to 1. Let us now use the Newton method to find this solution. We may as well take $x_0 = 1.5$ as our first guess.

The derivative is $f'(x) = 6x^2 + 6x - 6$, so the Newton scheme is

$$x_{n+1} = x_n + h_n \quad \text{where} \quad h_n = -\frac{2x_n^3 + 3x_n^2 - 6x_n - 4}{6x_n^2 + 6x_n - 6}$$

You should check the values in the following table.

n	x_n	$-h_n$
0	1.500000	-0.030303
1	1.469697	-0.000695
2	1.469002	-3.6×10^{-7}
3	1.469002	-6.3×10^{-11}

We can now be confident that one solution, to 4 decimal places, is $x = 1.4690$.

That is not the end of the problem. There may well be other solutions. For example, $f(-3) = -13$ and $f(-2) = 4$. So there is certainly another solution between -3 and -2. You can, if you like, go chasing around looking for other possibilities and then using Newton on each one in turn. It is a relief to know that a cubic equation cannot have more than 3 solutions!

Another approach is to use a basic fact about polynomials. If $x = a$ is a solution of the polynomial equation $p(x) = 0$ then $(x - a)$ is a factor of $p(x)$. In our case $(x - 1.4690)$ is, approximately, a factor of $f(x)$. So we can divide out the factor. This will leave us with a quadratic, which we can solve by the usual formula without bothering with Newton's method. I will not go any further with this.

It turns out that our equation has three solutions and they are, to 1 decimal place, -2.4, -0.6 and 1.5. Use Newton's method yourself to improve these results to 4 decimal places

12.8. Example. To give you some idea of the rate of convergence of the method let me do a calculation to a large number of decimal places (I've got software for doing this—you can't really do it on your calculator). Consider the equation $x^3 - 3x + 1 = 0$. It has a solution between 0 and 1 as you can easily see. Find it.

The result of using Newton on this problem, with $x_0 = 1.5$ and the calculations taken to 50 decimal places, is

n	x_n
0	$\underline{1.5}0000000000000000000000000000000000000000000000000$
1	$\underline{1.53}333333333333333333333333333333333333333333333$
2	$\underline{1.5320}906432748538011695906432748538011695906432749$
3	$\underline{1.5320888862}414666691024170140560915916262702002099$
4	$\underline{1.5320888862379560704047}993157913546008732037569433$
5	$\underline{1.532088886237956070404785301110833347871664914}3841$
6	$\underline{1.5320888862379560704047853011108333478716649141608}$

The underlined figures are the correct figures. We have got 46 accurate decimal places by the fifth iteration!

12.5 Higher Approximations

The basic idea behind the linear approximation was that the tangent line to a graph at a point stays reasonably close to the graph near to that point. You can also think of the linear approximation to $f(x)$ at $x = a$ as being the straight line that goes through $(a, f(a))$ and has the same slope at this point as $f(x)$. It happens that this slope is given by the derivative—hence our formula.

12.5.1 Second Approximation

Having got the idea we might think to improve it a bit by replacing the straight line ($y = ax + b$) by a quadratic curve $y = px^2 + qx + r$. This gives us *three* parameters to play with and might allow us to get a better 'fit' to the graph of $f(x)$.

Which quadratic do we use? As before, and using the same notation, we would want the quadratic to go through $(a, f(a))$ and to have the same slope as $f(x)$ at this point. That fixes the values of two of the coefficients but still leaves us with one to play with. Suppose we also ask that the quadratic should have the same *second* derivative at the point as $f(x)$. This now fixes the value of all three coefficients.

It isn't hard to see what that means. Write $g(x) = p(x-a)^2 + q(x-a) + r$. We want $f(x)$ and $g(x)$ to have the same value, the same first derivative and the same second derivative at a. Now

$$\begin{aligned} g(x) &= p(x-a)^2 + q(x-a) + r, & g(a) &= r = f(a), \\ g'(x) &= 2p(x-a) + q, & g'(a) &= q = f'(a), \\ g''(x) &= 2p, & g''(a) &= 2p = f''(a). \end{aligned}$$

Thus our approximating quadratic is

$$g(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a)$$

and we can write down an approximation which is called the **Second Approximation** (the Linear Approximation is often called the First Approximation).

$$\boxed{f(a+h) \approx f(a) + hf'(a) + \frac{1}{2}h^2 f''(a).}$$

Notice that the first two terms are exactly the same as for the linear approximation. We have only added a 'correction term' on the end.

To compare the accuracy of the linear and second approximations let me do some calculations to evaluate e^x for small values of x. The first and second derivatives of e^x at $x = 0$ are both 1. So, for small values of h, the two approximations are

$$e^h \approx 1 + h \quad \text{First Approximation}$$

$$e^h \approx 1 + h + \frac{1}{2}h^2 \quad \text{Second Approximation}$$

h	First	Second	Exact
0.2	1.2	1.22000	1.22140
0.1	1.1	1.10500	1.10517
0.05	1.05	1.05125	1.05127

You can see that the second approximation performs much better than the first.

12.9. Example. Use the second approximation to estimate the value of $\sqrt{9.8}$.

We have already used the first approximation and got the result 3.13333.

To use the second approximation we need the second derivative of $f(x) = \sqrt{x}$. This is $f''(x) = -\frac{1}{4}x^{-3/2}$. So $f''(9) = -1/108$. The second approximation therefore gives

$$\sqrt{9.8} = f(9 + 0.8) \approx 3 + \frac{0.8}{6} - \frac{0.8^2}{216} = 3.13037$$

which compares quite favourably with the correct result $\sqrt{9.8} = 3.130495$ (given to 6 decimal places).

12.5.2 Higher Approximations

This argument can be carried further. We can look for approximations that use higher and higher degrees of polynomials. To cut a long story short let me just give you the answers.

The successive approximations to the value of $f(x)$ near $x = a$ are

$$\begin{aligned} f(a+h) &\approx f(a) + hf'(a) \\ &\approx f(a) + hf'(a) + \frac{h^2}{2!}f''(a) \\ &\approx f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \frac{h^3}{3!}f'''(a) \\ &\approx f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \frac{h^3}{3!}f'''(a) + \frac{h^4}{4!}f^{iv}(a) \\ &\approx f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \frac{h^3}{3!}f'''(a) + \frac{h^4}{4!}f^{iv}(a) + \frac{h^5}{5!}f^{v}(a) \end{aligned}$$

and so on. The pattern is simple. Each approximation differs from the previous one by the addition of one further correction term. All the correction terms have the same form. Note that $n!$ stands for the 'factorial' of n, i.e. the product of all the numbers up to n, so $5! = 5.4.3.2.1 = 120$.

Following the above pattern we get the n^{th} approximation as

$$\boxed{f(a+h) \approx f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \frac{h^3}{3!}f'''(a) + \cdots + \frac{h^n}{n!}f^{(n)}(a).}$$

As you would expect, these approximations tend to get better and better as you take more terms. To give you some idea of their behaviour here are the first few approximations to the value of e^1 obtained by taking $f(x) = e^x$, $a = 0$ and $h = 1$:

order	approx
1	2.0000000
2	2.5000000
3	2.6666667
4	2.7083333
5	2.7166667
6	2.7180556
7	2.7182540
8	2.7182788
9	2.7182815
10	2.7182818
exact	2.7182818

So, by the time we have got to the 10^{th} approximation we have got the value accurate to at least 7 decimal places—though the earlier approximations are pretty hopeless.

12.10. Example. By using the fourth approximation show that if x is small then

$$\cos x \approx 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4$$

We are told that x is small. So we should take our approximation about $a = 0$. To get the fourth approximation we need the first four derivatives of $f(x) = \cos x$. This is easy: $f'(x) = -\sin x$, $f''(x) = -\cos x$, $f'''(x) = \sin x$ and $f^{iv}(x) = \cos x$. We now evaluate these at $x = 0$ to get $f(0) = 1$, $f'(0) = 0$, $f''(0) = -1$, $f'''(0) = 0$ and $f^{iv}(0) = 1$. So the fourth approximation is

$$\begin{aligned} f(0+h) &\approx f(0) + hf'(0) + \frac{h^2}{2}f''(0) + \frac{h^3}{3!}f'''(0) + \frac{h^4}{4!}f^{iv}(0) \\ &\approx 1 + 0 - \frac{h^2}{2} + 0 + \frac{h^4}{24} \\ &\approx 1 - \frac{1}{2}h^2 + \frac{1}{24}h^4 \end{aligned}$$

So, replacing h by x (just a change of notation), we get the required result:

$$\cos x \approx 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 \quad \text{if } x \text{ small}$$

12.6 Taylor Series

I have now got to try to answer the question that I asked at the start of the chapter. How do we actually evaluate functions like e^x or $\sin x$? Note that all we usually want is the value to a certain accuracy. The successive approximations in the previous section have shown us a way to approach this problem and we will follow it up in this section.

12.6.1 Infinite Series

Before doing the real work of this section I need to tell you a little bit about *infinite series*.

A *finite series* is just the sum of a finite number of terms

$$s = a_1 + a_2 + a_3 + \cdots + a_n$$

An *infinite series* is the sum of an infinite number of terms

$$s = a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

(we usually use $\cdots$ at the end of a series to indicate that it goes on for ever).

What do we mean by 'the sum of an infinite series'. You know how to add up a finite number of terms, so the sum of a finite series is no problem. But how on earth do we propose to add up an *infinite* number of terms? It took mathematicians a long time to sort out their ideas on this. The definition that we use goes as follows. Call the sum of the first n terms of the series s_n (called the n^{th} *partial sum*). So

$$\begin{aligned} s_1 &= a_1 \\ s_2 &= a_1 + a_2 \\ s_3 &= a_1 + a_2 + a_3 \\ s_4 &= a_1 + a_2 + a_3 + a_4 \end{aligned}$$

and so on. If this infinite sequence of values $s_1, s_2, s_3, \ldots$ tends to a limiting value than that value is called the sum of the infinite series, and we say that the infinite series *converges*.

If the sequence of partial sums does not tend to a limiting value then we say that the infinite series *diverges* and we do not give a sum for it. Here are two simple examples of infinite series that 'do not add up':

$$1 + 2 + 3 + 4 + 5 + \cdots + n + \cdots$$

$$1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \cdots$$

The first one is divergent because the partial sums are steadily increasing towards infinity and do not tend to a finite limiting value. You might get away with saying that the sum of the infinite series is infinite, but that is not really very helpful.

The second one is more interesting. The partial sums are $1, 0, 1, 0, 1, 0, 1, \ldots$. This sequence is not going off to infinity, but it is not tending to a limiting value either. So the series does not have a sum. (You will never believe the tangles that 18th century mathematicians got into with this series. They managed to produce all sorts of arguments to claim that the series actually added up to $\frac{1}{2}$—a sort of 'average' of the partial sums.)

12.6.2 Geometric Series

This is an example that you may already know something about. We have a formula for the sum of a simple finite geometrical series:

$$1 + x + x^2 + x^3 + x^4 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1}$$

(if $x \neq 1$). Look up the proof of this if you do not know it already.

What about the infinite geometrical series?

$$1 + x + x^2 + x^3 + x^4 + \cdots + x^n + \cdots$$

According to our definition the sum of this series is the limiting value of its partial sums, if they have a limiting value. We already have a formula for the n^{th} partial sum:

$$s_n = \frac{x^{n+1} - 1}{x - 1}$$

The question is: what happens to this as $n \to \infty$? The answer depends very much on the value of x. The only bit of the formula that depends on n is the x^{n+1} in the numerator.

If $x > 1$ then x^{n+1} tends to infinity as $n \to \infty$, so the partial sums also tend to infinity. So the series is divergent and does not have a sum.

If $x < -1$ then the situation is even worse in a sense. Not only are the partial sums getting bigger and bigger, they are also switching sign. Certainly divergent.

If $-1 < x < 1$ then $x^{n+1} \to 0$ as $n \to \infty$ because the successive powers of a number between -1 and 1 get closer and closer to zero. So in this case our formula for the partial sums *does* tend to a finite limiting value as $n \to \infty$. So the infinite geometrical series *is* convergent in this case and has the sum

$$1 + x + x^2 + x^3 + \cdots + x^n + \cdots = \frac{1}{1 - x}$$

The remaining two possibilities are $x = 1$ and $x = -1$. I leave it to you to see that the series does not converge in either of these cases.

The end result is that the infinite geometric series given above only converges (has a sum) if $-1 < x < 1$.

12.6.3 Taylor Series

I have shown you a sequence of approximations to the value of a function. It is *usually* the case that the higher the approximation you use the more accurate a result you get (not always true). Some of the numerical evidence given above demonstrates this.

This opens up the possibility that if we imagine the sequence of successive approximations going on for ever then the error will eventually dwindle away to nothing and, in the limiting case, we will not have an approximation but will get the exact result.

For elementary functions this is usually true, at least for some values of a and h.

The resulting infinite series is called the *Taylor Series* of the function $f(x)$ expanded about the point a. If it converges to the value of the function we get

$$\boxed{f(a + h) = f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \cdots + \frac{h^n}{n!}f^{(n)}(a) + \cdots}$$

If $a = 0$ then the series is known as the *Maclaurin Series* of the function f, which we might write as:

$$\boxed{f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0) + \cdots}$$

Taylor (and Maclaurin) series are tremendously important in many areas of mathematics. They are often used to *define* functions.

12.11. Example. What is the Maclaurin series of $f(x) = e^x$?

To write down the Maclaurin series we need to know the value at $x = 0$ of every derivative of the function. This is usually the practical problem that we face in working out Taylor series. In this case it is easy since every derivative of e^x is e^x and this has value 1 at $x = 0$. So the Maclaurin series becomes

$$1 + x.1 + \frac{x^2}{2!}.1 + \cdots + \frac{x^n}{n!}.1 + \cdots$$

It turns out that this is actually equal to the value of e^x for any value of x (I cannot prove that here). So we have the famous result that

$$\boxed{e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots + \frac{x^n}{n!} + \cdots}$$

You will often find this given as the *definition* of e^x.

12.12. Example. What are the Maclaurin series of $\sin x$ and $\cos x$?

This is another case where it is reasonably easy to work out the values of all the derivatives at $x = 0$. The first few derivatives of $\sin x$ are

$$\cos x, \quad -\sin x, \quad -\cos x, \quad \sin x, \quad \cos x$$

and so on. You can see that they go round and round in a cycle of four. The values of $\sin x$ and its derivatives at $x = 0$ are therefore

$$0, \quad 1, \quad 0, \quad -1, \quad 0, \quad 1, \quad 0$$

and so on.

It turns out that the Maclaurin series of $\sin x$ is valid for all values of x. So we have

$$\boxed{\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \cdots}$$

Using the same results you can show that the Maclaurin series of $\cos x$, which is also valid for all values of x is

$$\boxed{\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots}$$

12.13. Example. Find the value of $\sin 1.23$ to 6 decimal places.

Put $x = 1.23$ into the Maclaurin series for $\sin x$ and start adding up terms. Carry on doing this until the terms being added on 'obviously' don't affect the answer to 6 decimal places. Stop there.

Term	value	sum so far
x	+1.23	1.23
$-x^3/3!$	-0.31014450	0.9198555
$x^5/5!$	+0.02346088	0.9433164
$-x^7/7!$	-0.00084509	0.9424713
$x^9/9!$	+0.00001776	0.9424890
$-x^{11}/11!$	-0.00000024	0.9424888
$x^{13}/13!$	+0.00000000	0.9424888

Which gives us $\sin 1.23 = 0.942489$ to 6 decimal places (and is correct).

Note that this is a slightly dangerous game and is not quite like what was happening with the Newton method. By stopping the calculation early I have missed out an *infinite* number of terms (because the series goes on for ever). The individual terms are certainly getting very small, but an awful lot of very small terms can add up to something enormous. You do need to be a bit careful—especially if the terms are getting small quite slowly.

As a warning, consider the infinite series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

The individual terms in this series are tending to zero, yet it can be shown that this series is divergent. If you add up enough of the terms you can get a sum as big as you like!

12.14. Example. A Maclaurin series for the logarithm.

Here we have to be a bit careful. $\ln x$ is not defined at $x = 0$, so there is no hope of a Maclaurin series for $\ln x$.

Instead, it is usual to look at the Maclaurin series of $f(x) = \ln(1+x)$ (or the Taylor series of $\ln x$ expanded about $x = 1$ if you prefer).

The first few derivatives are:

$$f'(x) = \frac{1}{1+x}, \qquad f''(x) = -\frac{1}{(1+x)^2}, \qquad f'''(x) = \frac{2}{(1+x)^3}$$

You can probably see the pattern beginning to develop. The n^{th} derivative is

$$f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{(1+x)^n}, \qquad f^{(n)}(0) = (-1)^{n-1}(n-1)!$$

The Maclaurin series now becomes

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

This is not valid for all values of x. Obviously it does not make sense for $x \le -1$ because the logarithm is not defined for negative values. Less obviously, it does not work for $x > 1$ either. The Maclaurin series for $\ln(1+x)$ is only valid for $-1 < x \le 1$.

12.6.4 The Binomial Series

As I have said, it is difficult to work out Taylor series for functions if we do not have an easy way to work out all the derivatives of the function. That is why I did not try to work out the Maclaurin series for $\tan x$—just you try to work out the 20^{th} derivative of $\tan x$.

One more class of functions that we can cope with are those of the form $f(x) = (1+x)^a$. If we work out the first few derivatives we get

$$f'(x) = a(1+x)^{a-1}, \qquad f''(x) = a(a-1)(1+x)^{a-2},$$
$$f'''(x) = a(a-1)(a-2)(1+x)^{a-3}$$

The pattern is obvious and there is no problem in writing down the general derivative.

Notice that if a is a positive whole number the derivatives are eventually all zero. This is not true if a is any other number (apart from $a = 0$).

We now have the Maclaurin series

$$\boxed{(1+x)^a = 1 + ax + \frac{a(a-1)}{2!}x^2 + \frac{a(a-1)(a-2)}{3!}x^3 + \frac{a(a-1)(a-2)(a-3)}{4!}x^4 + \cdots}$$

If a is not a non-negative integer then this expansion is only valid for $-1 < x < 1$.

This expansion is called the Binomial Series.

Putting $a = -1$ we get the special case

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 + \cdots$$

which is just the usual geometric series.

Putting $a = \frac{1}{2}$ we get

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \frac{7}{256}x^5 - \cdots$$

Chapter 13

Differential Equations

This is a very brief introduction to a very important topic. If you meet mathematics again after this year then you will probably meet differential equations again as well.

13.1 Introduction

A **differential equation** is an equation for an unknown function, say $y(x)$, which involves derivatives of the function.

For example

$$\frac{dy}{dx} = x, \qquad y'' - 3y' + 2y = \sin x, \qquad \frac{y''' - y'}{y^2} = x + y'$$

The **order** of a differential equation is the order of the highest derivative occurring in it. In the above examples the orders are 1,2,3.

[*Technically, these are known as Ordinary Differential Equations (odes) because the unknown function is a function of one variable. Differential equations involving functions of several variables and their partial derivatives are called Partial Differential Equations (pdes).*]

Many laws in science and engineering are statements about the relationship between a quantity and the way in which it changes. The change is often measured by a derivative and therefore the mathematical expression of these laws tends to be in terms of differential equations. Our earlier example of Malthus' Law was a case in point.

Given a differential equation the obvious reaction is to try to solve it for the unknown function. As with integrals, and for much the same reason, this is easier said than done. In this chapter I will have a quick look at one very simple class of differential equations. The aim is to give you some feel for the way in which differential equations behave.

Consider the differential equation $y''(x) = x$, where the dash denotes differentiation wrt x. Integrating both sides of this equation wrt x we get

$$y' = \frac{1}{2}x^2 + C$$

where C is an arbitrary constant of integration. This is one place where it is crucially important to include the constant of integration!

Now integrate once more and get

$$y = \frac{1}{6}x^3 + Cx + D$$

where D is a further constant of integration.

We now have the **General Solution** of the differential equation with arbitrary constants C and D. Note that this is really an infinite class of solutions. If we give C and D particular values then we get a **Particular Solution**. For example, $y = x^3/6+x$ and $y = x^3/6+2x-3$ are particular solutions of the equation.

The values of the arbitrary constants that we almost invariably acquire when solving a differential equation are usually determined by giving conditions that the solution is required to satisfy. The most common kind of conditions are **Initial Conditions**, where the values of y and some of its derivatives are given for a specific value of x.

13.1. Example. Find the solution to $y'' = x$ then satisfies $y(0) = 1$ and $y'(0) = 0$.

We know that the general solution is

$$y = \frac{1}{6}x^3 + Cx + D$$

The condition $y(0) = 1$ says that $1 = 0 + 0 + D$, so $D = 1$. The condition $y'(0) = 0$ says that $0 = 0 + C$, so $C = 0$. So the required solution is

$$y = \frac{1}{6}x^3 + 1$$

Note that an equation of order n generally requires n integrations to get the general solution, so the general solution can be expected to contain n unknown constants and you would expect to have to give n conditions to fix these constants.

Those of you who know some physics will know that Newton's law of motion relates acceleration (second derivative) to force (which is usually a function of position and time). This yields a second order differential equation. In the simple case of a particle moving in a straight line the general solution should contain two constants of integration. A particular solution can be specified by giving the initial position and velocity of the particle, and this corresponds well with our physical intuition.

13.2. Example. I drop a stone from height H above the ground. It falls under gravity and there is no air-resistance. When does it hit the ground?

Measure x vertically upwards from the ground. Let $x(t)$ be the height of the stone above the ground at time t. Suppose that the particle is dropped at time $t = 0$.

Physics tells us that $x(t)$ satisfies the differential equation

$$\ddot{x} = -g \quad \text{where } g \text{ is a constant}$$

Our Initial Conditions are that $x(0) = H$ and that $\dot{x}(0) = 0$.

Integrating the equation gives $\dot{x} = -gt + C$ and integrating once more gives

$$x(t) = -\frac{gt^2}{2} + Ct + D$$

—this is the general solution.

The condition $x(0) = H$ says that $D = H$ and the condition $\dot{x}(0) = 0$, applied to the previous equation, says that $C = 0$. So the required solution is

$$x(t) = -\frac{gt^2}{2} + H$$

The stone hits the ground when $x = 0$. This happens when

$$H = \frac{1}{2}gt^2 \quad \text{or} \quad t = \sqrt{\frac{2H}{g}}$$

13.2 Separable Equations

This is the only class of differential equations that I am going to treat in general.

A differential equation is said to be *separable* if it can be manipulated into the form

$$f(y)\frac{dy}{dx} = g(x)$$

You can, in principle, solve this equation by integrating both sides with respect to x. You get

$$\int f(y)\,dy = \int g(x)\,dx + C$$

It may not be possible to express y simply in terms of x.

The following are examples of separable equations

$$y^2\frac{dy}{dx} = x + 1, \qquad e^v\frac{dv}{du} = 2u, \qquad (1+y)\frac{dy}{dx} = 1$$

So are the following, though they need some rearrangement:

$$\begin{array}{rcc} x^2\frac{dy}{dx} = e^y & \rightarrow & e^{-y}\frac{dy}{dx} = \frac{1}{x^2} \\ \frac{dy}{dx} = \frac{1+x}{1+y} & \rightarrow & (1+y)\frac{dy}{dx} = 1 + x \\ xy' = \frac{1+x}{1+y} & \rightarrow & (1+y)y' = 1 + 1/x \end{array}$$

Let's do some examples.

13.3. Example. $yy' = x^2$

This equation rearranges to give

$$\int y\,dy = \int x^2\,dx + C \quad \text{or} \quad \frac{1}{2}y^2 = \frac{1}{3}x^3 + C$$

The explicit solution comes in two cases

$$y(x) = \sqrt{\frac{2}{3}x^3 + 2C} \quad \text{and} \quad y(x) = -\sqrt{\frac{2}{3}x^3 + 2C}$$

(Notice that this example shows that the solution of a differential equation may not exist for all values of x. In this case there will be values of x for which the term under the square root is negative.)

13.4. Example. $y' = e^y x$

This can be rewritten as $e^{-y}y' = x$ and gives

$$\int e^{-y}\,dy = \int x\,dx + C \quad \text{or} \quad -e^{-y} = \frac{1}{2}x^2 + C$$

The explicit solution is then

$$y(x) = -\ln(-C - \frac{1}{2}x^2)$$

There are values of C for which this solution does not exist. Note that C cannot be allowed to be positive.

13.5. Example. $\dfrac{dy}{dx} = \dfrac{x+1}{y+1}$

This can be rewritten as $(y+1)y' = x + 1$. So integrating it we get

$$\frac{1}{2}(y+1)^2 = \frac{1}{2}(x+1)^2 + C$$

13.2.1 The Malthus Equation

Let us go back to the population model that we developed in the section on the exponential function.

$$\frac{dP}{dt} = kP, \qquad P(0) = P_0$$

where k is a constant.

This is a separable equation and we can rearrange it to get

$$\int \frac{dP}{P} = \int k\,dt + C$$

or

$$\ln|P| = kt + C$$

Exponentiate both sides and get

$$P(t) = \pm e^C e^{kt}$$

now impose the initial condition $P(0) = P_0$ and get the result

$$P(t) = P_0 e^{kt}$$

13.3 Generalities

Even the simple examples that we have done so far highlight most of the basic points about the behaviour of differential equations.

1. Since the process of solving a differential equation involves integration so as to get rid of the derivatives we always acquire arbitrary constants in our solution. This means that, in general, differential equations have infinitely many solutions.

2. An expression for the arbitrary solution of a differential equation, involving all the unknown constants of integration, is called the **General Solution** to the equation. A solution that you obtain by giving values to the constants of integration is called a **Particular Solution**. For example, the General Solution to the differential equation $y' = x$ is $y(x) = \frac{1}{2}x^2 + C$. The functions $y(x) = \frac{1}{2}x^2$ ($C = 0$) and $y(x) = \frac{1}{2}x^2 + 2$ ($C = 2$) are Particular Solutions.

3. We haven't done many higher-order equations yet but, as you will realise, an equation of order n contains an n^{th} order derivative of the unknown function y that has to be reduced down to y. This requires n integrations. Each integration produces a constant of integration. So, in general, the general solution to an equation of order n will involve n unknown constants.

4. To obtain a particular solution from a general solution we need to be told something about the required solution in order to fix the values of the constants of integration. A fairly standard approach to this, very common in dynamics, is to give information about the solution and its derivatives at a particular value of x. This is called, for historical and practical reasons, specifying Initial Values for the problem.

5. There are lots of other ways to determine a particular solution. Sometimes it is done, for second order equations, by giving the value of y at two different values of x. This is called giving Boundary Values for the problem.

13.4 Linear First-Order Equations

13.6. Example. Solve the differential equation

$$x\frac{dy}{dx} + y = \frac{1}{x}.$$

Solution This may be a first order equation, but the variables aren't separable. But we notice something about the left hand side. The differential equation can be rewritten as

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

and we can then set about integrating both sides of the equation. We get

$$\int \frac{d}{dx}(xy)\,dx = \int x^2\,dx + C, \qquad \text{so} \qquad xy = \frac{x^3}{3} + C.$$

and the required solution is

These are differential equations of the form

$$a(t)\frac{dx}{dt} + b(t)x = c(t) \tag{13.1}$$

They can always be solved, in principle, by a method known as the **Integrating Factor Method**.

If we start by dividing through by $a(t)$ we can simplify the equation down to

$$\dot{x} + \beta(t)x = \gamma(t) \tag{13.2}$$

Now, suppose we can find a function $f(t)$ such that

$$\dot{f} = \beta(t)f \tag{13.3}$$

Multiply both sides of (13.2) by f and get

$$f\gamma = f\dot{x} + \beta f x = f\dot{x} + \dot{f}x = \frac{d}{dt}(f.x)$$

as a consequence of the rule for differentiating a product.

We can now integrate and get

$$x = \frac{1}{f}\int f\gamma\,dt$$

The only remaining problem is to find an f to satisfy (13.3). But (13.3) is just an ordinary separable equation and we get the solution

$$\ln f = \int \beta\,dt$$

Let me do one or two examples to show you that it is easy (in principle).

Consider the equation $t\dot{x} + 2x = 1$. If we multiply through by t we get

$$t = t^2\dot{x} + 2tx = \frac{d}{dt}(t^2 x)$$

So, integrating,

$$t^2 x = \frac{1}{2}t^2 + C \quad \text{so} \quad x(t) = \frac{1}{2} + fracCt^2$$

Now consider the equation $\dot{x} + tx = t$. In this case $\beta = t$, so

$$f = \exp(\int \beta\,dt) = \exp(\frac{t^2}{2})$$

(Note that we don't have to bother with a constant of integration because we are just looking for *something* which satisfies (13.3).)

Multiplying through by f our equation becomes

$$e^{t^2/2}\dot{x} + te^{t^2/2}x = te^{t^2/2}$$

or

$$\frac{d}{dt}(e^{t^2/2}x) = te^{t^2/2}$$

Integrate and get

$$e^{t^2/2}x = \int te^{t^2/2}\,dt = e^{t^2/2} + C$$

So the solution is

$$x(t) = 1 + Ce^{-t^2/2}$$

One more example for luck. Consider the equation $\dot{x} + \sin(t)x = \cos(t)$. Here we have $\beta(t) = \sin(t)$ and we get f by

$$\ln f = \int \beta \, dt = -\cos(t) \quad \text{so} \quad f = e^{-\cos t}$$

Multiply through the equation by f and get

$$f\dot{x} + f\sin(t)x = \frac{d}{dt}(fx) = \cos(t)e^{-\cos t}$$

Thus

$$e^{-\cos t}x = \int \cos(t)e^{-\cos t}\, dt + C$$

and

$$x(t) = e^{\cos t}\int \cos(t)e^{-\cos(t)}\, dt + Ce^{\cos t}$$

Unfortunately, as often happens when using this method, I don't think that this integral can be done (prove me wrong if you can).

13.5 Linear Differential Equations

Solving general differential equations is complicated and can usually only be done numerically; certainly it is way beyond the scope of this course. Even if we restrict to a very special case; that of a **linear differential equation** with **constant coefficients** in which we restrict to equations of the form

$$a_n\frac{d^ny}{dx^n} + a_{n-1}\frac{d^{n-1}y}{dx^{n-1}} + \ldots + a_2\frac{d^2y}{dx^2} + a_1\frac{dy}{dx} + a_0y = q(x)$$

with each of $a_n, \ldots, a_0$ constant, rather than a function of x, we still have to work quite hard. However, the particular case in which $n = 2$, so we are dealing with a **second order** occurs a lot in mechanics and electronics and you should know something about how to solve it.

A differential equation of the form

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = q(x)$$

where a, b, c are constants and $q(x)$ is a function of x is called a second order linear differential equation with constant coefficients. Snappy title, eh?

If $q(x)$ is replaced by 0, you add the word **homogeneous** in front of "second". Otherwise such an equation is known as **inhomogeneous**.

The homogeneous case is significantly simpler and helps us with the general case, so we start by concentrating on the homogeneous case; the one in which the right hand side of the equation is 0. Thus we are looking at the equation

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0.$$

It is easy to summarise this method:

- Form the quadratic equation $am^2+bm+c=0$. This is done by simply replacing $\dfrac{d^2y}{dx^2}$ by m^2 and $\dfrac{dy}{dx}$ by m. This quadratic is called the *auxiliary equation*.
- Use the quadratic formula to solve the auxiliary equation. This gives $m = \dfrac{-b \pm \sqrt{D}}{2a}$ where $D = b^2 - 4ac$. There are now three possible cases:
  - This occurs if $D > 0$, i.e. if $b^2 > 4ac$. The auxiliary equation will have two distinct real numbers as roots. Call them α_1 and α_2. Then the general solution of the equation is
  $$y = Ae^{\alpha_1 x} + Be^{\alpha_2 x}$$
  where A and B are constants. As with first order equations, there will often be extra information in the question that will enable you to calculate A and B.
  - This occurs if $D = 0$, i.e. if $b^2 = 4ac$. This time the auxiliary equation just has one root. Call it α. Then the general solution of the differential equation is
  $$y = Ae^{\alpha x} + Bxe^{\alpha x}$$
  where A and B are constants.
  - This occurs if $D < 0$, i.e. if $b^2 < 4ac$. The auxiliary equation has no real roots. However it does have "complex roots" (see next term, and the theory of complex numbers produces the following as roots of the differential equation:
  $$y = e^{\alpha x}\left(A\sin\beta x + B\cos\beta x\right)$$
  where $\alpha = -\dfrac{b}{2a}$ and $\beta = \sqrt{\dfrac{-D}{2a}}$ and A and B are some constants.

13.7. Example. Find the general solution of the differential equation
$$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 4y = 0.$$
Solution The auxiliary equation is $m^2+5m+4=0$. $5^2 > 4.1.4$ and so we are in the first case. The roots of the auxiliary equation are -4 and -1, giving us a solution of
$$y = Ae^{-4x} + Be^{-x}$$
where A and B are some constants.

13.8. Example. Find the general solution of the differential equation
$$\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + 4y = 0\,.$$
Solution The auxiliary equation is $m^2-4m+4=0$. Since $4^2 = 4.1.4$ the discriminant vanishes and so we are in case two. The auxiliary equations just the one (repeated) root $m=2$. So the general solution is
$$y = Ae^{2x} + Bxe^{2x}$$
where A and B are some constants.

13.9. Example. If you neglect air resistance, the equation of motion for small oscillations of a simple pendulum is

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0$$

where θ is the angular displacement from the vertical, l the length of the pendulum, g the gravitational constant and t time.

Solution This time we are in case 3. Calculating α and β we find that $\alpha = 0$ and $\beta = \sqrt{\frac{g}{l}}$. So the solution is

$$\theta = A\cos\left(\sqrt{\frac{g}{l}}\,t\right) + B\sin\left(\sqrt{\frac{g}{l}}\,t\right)$$

where A and B are some constants.

Typically one would be told the values of θ and θ' when $t = 0$, and this would make it possible to find A and B. For example, it might be the case that when $t = 0$ the pendulum was at rest and $\theta = \theta_0$. You would then find that $A = \theta_0$ and $B = 0$.

If we didn't neglect air resistance, we'd get a situation in which the equation of motion would have a $\frac{d\theta}{dt}$ term. This would lead to a value for α which was negative and a solution which was of the form

$$\theta = e^{\alpha t}\left(A\cos\gamma t + B\sin\gamma t\right)$$

producing a situation known as "damped oscillation" (see diagram).

The sine and cosine part of the solution produces the oscillations and the $e^{\alpha t}$ part the damping.

13.10. Remark. Another situation, particularly common in situations concerning electrical oscillations, is one where the simple, damped or undamped oscillation is disturbed by some external stimulus. This leads us on to non-homogeneous equations, and I shall say a little about those next time.

13.5.1 The Inhomogeneous Case

This time the equation is

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = q(x)$$

where a, b, c are constants and $q(x)$ is some function of x.

13.11. Example. An electrical network consists of an inductance L, a] resistance R and a condenser of capacity C connected in series with an applied voltage $E(t) = a \sin \omega t$. The physics tells us that the charge q on a plate of the condenser satisfies the equation

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = a \sin \omega t$$

There are two stages to the solution process.

- Solve the homogeneous equation got by replacing $q(x)$ by 0. Do this using the method described last time. The result is something called the **Complementary Function**.
- This consists of finding one particular solution to the given equation. This, when you find it, is called the **Particular Integral**. The general solution to the equation then consists of the sum of the Complementary Function and the Particular Integral.

The new skill to be acquired is the method of finding particular integrals. I don't have time to give you an exhaustive account, which will cover all the possibilities. However, I can try to show you the general approach.

The trick is to begin with something called a **trial function**. To do this you begin with something that has the same general form as $q(x)$. By this I mean the following sort of thing:

- If $q(x)$ is a quadratic in x, try a general quadratic in x, i.e. try $y = \alpha x^2 + \beta x + \gamma$.
- If $q(x)$ is of the form $d \sin \omega x + e \cos \omega x$ for some d, e, try $y = C \sin \omega x + D \cos \omega x$.
- If $q(x) = de^{\alpha x}$, try $y = Ce^{\alpha x}$.

The main complication comes when the thing you want to try as a trial function is already part of the complementary function. Then the general rule is to take what you would have tried but for this and multiply it by x.

If the situation facing you is not covered by this, go and find a book!

13.12. Example. Find the general solution of the differential equation

$$2\frac{d^2y}{dx^2} - 9\frac{dy}{dx} - 35y = \sin 2x\,.$$

Solution First find the complementary function by solving

$$2\frac{d^2y}{dx^2} - 9\frac{dy}{dx} - 35y = 0\,.$$

The auxiliary equation is $2m^2 - 9m - 35 = 0$. This has roots

$$m = 9 \pm \sqrt{\frac{81 + 280}{4}} = 9 \pm \sqrt{\frac{361}{4}} = \frac{9 \pm 19}{4} = 7 \quad \text{or} \quad -2.5$$

So the complementary function is $y = Ae^{7x} + Be^{-2.5x}$.

The function on the right is not part of this complementary function, and so for the trial function we can just take something of the form $C \sin 2x + D \cos 2x$. To complete the process we have to calculate the particular values of C and D that will work. This is a matter of putting our trial function into the equation and doing the sums.

$$y = C\sin 2x + D\cos 2x$$

Therefore

$$\frac{dy}{dx} = 2C\cos 2x - 2D\sin 2x$$

and

$$\frac{d^2y}{dx^2} = -4C\sin 2x - 4D\cos 2x$$

Therefore

$$2\frac{d^2y}{dx^2} - 9\frac{dy}{dx} - 35y = -8C\sin 2x - 8D\cos 2x - 18C\cos 2x + 18D\sin 2x - 35C\sin 2x - 35D\cos 2x$$

Therefore

$$\sin 2x = (-8C + 18D - 35C)\sin 2x + (-8D - 18C - 35D)\cos 2x$$

Equating coefficients tells us that

$$-43C + 18D = 1 \qquad\qquad -18C - 43D = 0$$

So

$$C = \frac{-43}{2173} \qquad \text{and} \qquad D = \frac{18}{2173}$$

Therefore the general solution of the given equation is

$$y = Ae^{7x} + Be^{-2.5x} + \frac{1}{39}\sin 2x + \frac{1}{26}\cos 2x\,.$$

13.13. Example. Find the general solution of the differential equation

$$\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 4y = 4e^{4x}\,.$$

Solution The auxiliary equation is $m^2 - 5m + 4 = 0$. This has roots 1 and 4, and so the complementary equation is $y = Ae^{4x} + Be^{x}$.

The righthand side of the given equation, being a constant times e^{4x} is part of the complementary function. So for a trial function we have to try Cxe^{4x} instead.

$$\begin{aligned} y &= Cxe^{4x} \\ \frac{dy}{dx} &= Ce^{4x} + 4Cxe^{4x} \\ \frac{d^2y}{dx^2} &= 8Ce^{4x} + 16Cxe^{4x} \end{aligned}$$

Therefore

$$\begin{aligned}\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 4y &= 8Ce^{4x} + 16Cxe^{4x} - 5Ce^{4x} - 20Cxe^{4x} + 4Cxe^{4x} \\ &= 3Ce^{4x}\end{aligned}$$

and so

$$4e^{4x} = 3Ce^{4x}.$$

Therefore

$$C = \frac{4}{3}$$

and the equation has general solution

$$y = Ae^{4x} + Be^{x} + \frac{4}{3}xe^{4x}$$

where A and B are arbitrary constants.

Appendix A

Numbers

It is a fact of life that almost all the numerical quantities that you deal with are, for one reason or another, not exactly correct. I mean that the number that you are working with does not give the true value of the thing that it measures. There are two obvious reasons for this. Firstly, the number may come from experimental data which tends to be approximate by its nature. Secondly, things like calculators and computers are very limited in the precision to which they can store and use numbers.

It is important that you know how to work with this inaccuracy, rather than just ignoring it.

A.1 Measures of Error

Suppose the exact value of some quantity is T and that A is an approximation to this value. There are two standard ways in which to measure the 'error' in using A as an approximation to T:

- Absolute Error $\quad E = |A - T|;$

- Relative Error $\quad E = \left|\dfrac{A - T}{T}\right|.$

Percentage error is relative error times 100.

For numbers close to 1 both measures are roughly the same. For T very large or very close to zero they are very different. Which one gives the appropriate measure of error depends very much on context. There are many situations in which your main concern is the percentage error. If you order 1 ton of sand and only receive 0.99 of a ton then you might not complain too loudly. On the other hand, if the two bits of the channel tunnel had been out by one percent in direction the result would have been extremely expensive.

Decimal Places

Let $x = 23.46285$, $y = 0.002145$ and $z = 12468.23$ be exact values.

x	rounded to 3 decimal places (3D) is	23.463
x	truncated to 3 decimal places is	23.462
y	rounded to 2 decimal places is	0.00
z	truncated to 1 decimal place is	12468.2

The rule for *truncation* is easy — just forget everything after the final decimal place. The rule for *rounding* is a bit more complicated. If the next digit is bigger than 5 then add one to the last digit. If the next digit is less than 5 don't do anything. If the next digit is 5 and there are further digits other than zero then add 1. If the next digit is 5 and there are no further non-zero digits then toss a coin. That takes longer to say than to understand! Note that the 'adding 1' may have lots of consequences: 2.3999 rounded to 3D is 2.400.

Note that the statement 'a = 1.23 rounded to 2D' means that

$$1.225 \leq a \leq 1.235$$

In other words a rounded, or truncated, value really represents not a single number but a range of numbers. This is the important fact to bear in mind.

A.2 Scientific Notation

This is best explained at first by examples. There are a number of different conventions in use, mainly concerning the position of the decimal point.

If we are writing numbers in *scientific notation* with 4 *significant digits* then

12.12342	becomes	0.1212×10^2
-6.2321	becomes	-0.6232×10^1
0.002132	becomes	0.2132×10^{-2}
234221.23	becomes	0.2342×10^6
0	becomes	0.0000×10^0

Again, you have the options of rounding or truncating. In general, a number is written in scientific notation with k significant digits if it is put into the form

$$x = \pm 0.d_1 d_2 d_3 \ldots d_k \times 10^a$$

where $d_1 \neq 0$ unless $x = 0$ (zero is always a problem since it alone cannot be scaled in the desired way).

234.12321	rounded to 4 significant figures is	0.2341×10^3
-23.67	rounded to 2 significant figures is	-0.24×10^2
123231.123	rounded to 2 significant figures is	0.12×10^6
0.0000123	rounded to 1 significant figure is	0.1×10^{-4}

Accuracy of Calculations

Let me do a few examples to warn you of the difficulties.

Suppose that x rounded to 2D is 1.23 and that y rounded to 4D is 21.4628. Then the statement

$$x + y = 1.23 + 21.4628 = 22.6928$$

is dangerous nonsense. The result implies that we know more about the value of x than we actually do. Let me do the calculation properly. Because of the rounding *all* we know is that

$$1.225 \leq x \leq 1.235 \quad \text{and} \quad 21.46275 \leq y \leq 21.46285.$$

So all that we can say is that

$$1.225 + 21.46275 \leq x + y \leq 1.235 + 21.46285;$$

in other words:

$$22.68775 \leq x + y \leq 22.69785$$

and, if you want a simple answer, the best you can say is that $x + y = 22.7$ to 1 decimal place.

Suppose that x rounded to 2D is 34.21 and that y rounded to 3D is 0.005. Then the statement

$$\frac{x}{y} = \frac{34.21}{0.005} = 6842.$$

is, once more, dangerous nonsense. Let's do this calculation carefully as well:

$$34.205 \leq x \leq 34.215 \quad \text{and} \quad 0.0045 \leq y \leq 0.0055.$$

So

$$\frac{34.205}{0.0055} \leq \frac{x}{y} \leq \frac{34.215}{0.0045}.$$

In other words:

$$6219.091 \leq \frac{x}{y} \leq 7603.333. \tag{*}$$

You can say

$$\frac{x}{y} = 0.1 \times 10^5 \quad \text{to 1 significant digit} \quad ,$$

but this gives much less information than (*).

Appendix B

Solutions to Exercises

Solutions for Questions 1 (page 89).

Solution 10.1:

a) Adding, we get $1 + 3j - 3(2 - j) = -5 + 6j$. We have

$$\frac{1}{z} = \frac{1}{1+3j} = \frac{1-3j}{(1+3j)(1-3j)} = \frac{1-3j}{10}.$$

Thus $1/z$ has real part 0.1 and imaginary part -0.3.

Finally

$$\left|\frac{w+\bar{w}}{w-\bar{w}}\right| = \left|\frac{2-j+2+j}{2-j-2+j}\right| = \left|\frac{4}{-2j}\right| = 2.$$

This has real part 2 and imaginary part 0.

Again

b) Recall the "30, 60, 90" triangle, with sides 1, $\sqrt{3}$, and 2. Then

$$z = 1 - \sqrt{3}j = 2\exp\left(-j\frac{\pi}{3}\right)$$

and so

$$z^4 = 16\exp\left(-\frac{4j\pi}{3}\right) = 16\exp\left(\frac{2j\pi}{3}\right),$$

where the second form is in terms of the modulus (16) and *principal* argument ($2\pi/3$) of z^4.

c) We have $w = -27j = 3^3\exp\left(-\frac{1}{2}\right)\pi j = 3^3\exp\left(2k - \frac{1}{2}\right)\pi j$ for any integer k. Thus by de Moivres theorem, we have

$$w = 3\exp\left(\frac{2k}{3} - \frac{1}{6}\right)\pi j,$$

with distinct solutions occurring when $k = 0$, 1 and 2. Thus the three solutions are

$$w = 3\exp\left(-\frac{1}{6}\right)\pi j, \quad 3\exp\left(\frac{2}{3}-\frac{1}{6}\right)\pi j, \quad \text{and} \quad 3\exp\left(\frac{4}{3}-\frac{1}{6}\right)\pi j.$$

These can be written, in simplified form as

$$w = 3\exp\left(-\frac{\pi j}{6}\right), \quad 3j, \quad \text{and} \quad 3\exp\left(-\frac{5\pi j}{6}\right).$$

The three roots are shown in Fig. B.1.

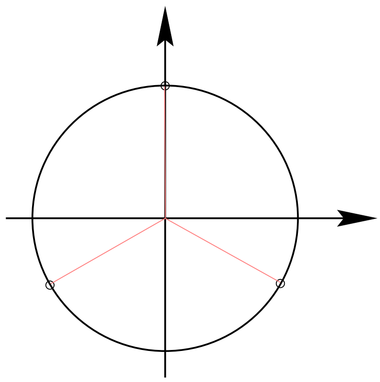


Figure B.1: The three solutions of $w^3 = -27j$.

Solution 10.2:

a) We use the usual rules for addition and multiplication:

$$\begin{aligned} z.w &= (1-2j)(3+j) = 3 - 6j + 2 + j = 5 - 5j \\ \frac{w}{z+2+j} &= \frac{3+j}{1-2j+2+j} = \frac{3+j}{3-j} = \frac{(3+j)(3+j)}{3-j)(3+j)} \\ &= \frac{9+6j-1}{9+1} = \frac{4+3j}{5}. \\ |1+3j-z\bar{z}| &= |1+3j-(1-2j)(1+2j)| = |1+3j-(1+4)| \\ &= |-4+3j| = 5. \end{aligned}$$

b) Working directly, or sketching on the Argand diagram shows that

$$-2+2j = \sqrt{8}\exp\left(\frac{3\pi j}{4}\right).$$

To solve the equation $w^3 = -2+2j$, write $w = r\exp(j\theta)$. Then by de Moivre's theorem,

$$r^3\exp(3j\theta) = \sqrt{8}\exp\left(\frac{3\pi j}{4}\right).$$

Thus $r^3 = \sqrt{8}$ and $3\theta = \frac{3\pi}{4} + 2k\pi$ for some integer k. Since $r > 0$, we see that $r = 8^{1/6}$ and

$$\theta = \frac{\pi}{4} + \frac{2k\pi}{3} \qquad (k = 0, 1, 2)$$

gives the three distinct solutions. These are shown in Fig. B.2; the angles involved are $\pi/4$, $11\pi/12$ and $-5\pi/12$, while the circle has radius $r = 2^{1/6}$.

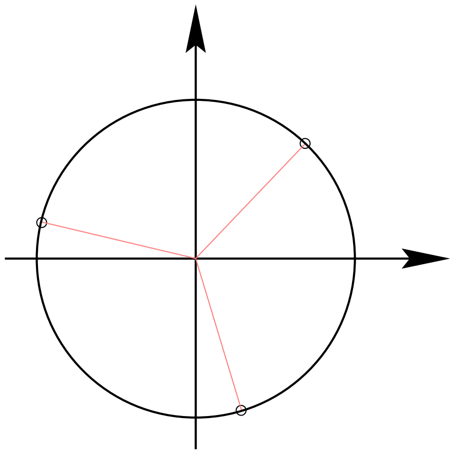


Figure B.2: Three roots of the equation $w^3 = -2 + 2j$.

Solution 10.3:

a) We substitute the given values.

$$\frac{w}{w+\bar{z}} = \frac{(1-7j)}{(1-7j)+(3-j)} = \frac{(1-7j)(4+8j)}{(4-8j)(4+8j)} = \frac{4-28j+8j+56}{16+64} = \frac{60-20j}{80} = \frac{3-j}{4}.$$

Using the definition of $|z|$, $|z| = \sqrt{9+1} = \sqrt{10}$. Similarly, $|w| = \sqrt{1+49} = \sqrt{50}$. Finally,

$$\left|\frac{w}{z}\right| = \frac{|w|}{|z|} = \sqrt{\frac{50}{10}} = \sqrt{5}.$$

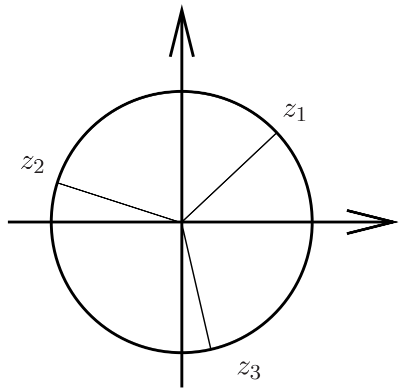
b) Write $w = -2 + 2j$. Then $|w| = \sqrt{4+4}$ and $\arg w = 3\pi/4$. Since $z^3 = w$, we have

$$z = \sqrt{2}\left(\cos\left(\frac{3\pi}{12} + \frac{2k\pi}{3}\right) + j\sin\left(\frac{3\pi}{12} + \frac{2k\pi}{3}\right)\right)$$

for $k = 0, 1, 2$. Thus the three roots are

$$z_1 = \sqrt{2}\left(\cos\frac{\pi}{4} + j\sin\frac{\pi}{4}\right) = 1 + j,$$
$$z_2 = \sqrt{2}\left(\cos\frac{11\pi}{12} + j\sin\frac{11\pi}{12}\right),$$
$$z_3 = \sqrt{2}\left(\cos\frac{19\pi}{12} + j\sin\frac{19\pi}{12}\right).$$

The diagram is as shown.



Solution 10.4:

a) Doing the long division of polynomials, we see that

$$p(z) = (z^2 - 2z + 2) \cdot (z^3 - 3z^2 + 4) = (z^2 - 2z + 2) \cdot q(z) \quad \text{(say)}.$$

Also $q(2) = 0$, so $p(2) = 0$. The remainder theorem now shows that $(z - 2)$ is a factor of $q(z)$. Again doing the division, we have

$$p(z) = (z^2 - 2z + 2) \cdot (z - 2) \cdot (z^2 - z - 2).$$

Finally we factor each of the quadratics. Using the quadratic formula on the first, the roots of $z^2 - 2z + 2$ are seen to be $1 + j$ and $1 - j$. Thus

$$z^2 - 2z + 2 = (z - 1 - j) \cdot (z - 1 - j).$$

The same method works on the second quadratic, or it can be factored directly to give

$$(z^2 - z - 2) = (z - 2) \cdot (z + 1).$$

Putting this all together shows that

$$p(z) = (z - 1 - j) \cdot (z - 1 - j) \cdot (z - 2)^2 \cdot (z + 1).$$

Solution 10.5:

a) Since the given polynomial has real coefficients, we know that if $z - (1 + j)$ is a factor of $p(z)$ then so is $z - (1 - j)$. The product of these is $z^2 - 2z + 2$. We have

$$p(z) = (z^2 - 2z + 2) \cdot (z + 4).$$

Thus the given linear term *was* a factor of $p(z)$ and we have

$$p(z) = z^3 + 2z^2 - 6z + 8 = (z - 1 - j) \cdot (z - 1 + j) \cdot (z + 4),$$

which expresses $p(z)$ as a product of linear factors.

Solution 10.6: Since the polynomial has real coefficients, if $z - 3j$ is a factor, then so is $z + 3j$ and hence $z^2 + 9$ will be a quadratic factor. Then

$$p(z) = z^4 - 3z^3 + 5z^2 - 27z - 36 = (x^2 + 9)(z^2 - 3z - 4)$$
$$= (z^2 + 9)(z - 4)(z + 1) = (z - 3j)(z + 3j)(z - 4)(z + 1).$$

and we have p as a product of four linear factors.

Solutions for Questions 2 (page 122).

Solution 11.2:

a) Since $\det(A) = 2 \neq 0$, the inverse exists. From the formula

$$A^{-1} = \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.$$

Since A has two columns and B has three rows, AB does not exists, while $BA = \begin{pmatrix} 0 & 2 \\ 1 & 1 \\ -1 & 1 \end{pmatrix}$.

We can compute $CB = (3, 2)$, while C^{-1} does not exists, since it is not square.

b) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then we have

$$A \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot A^T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & c \\ -b & -d \end{pmatrix} = \begin{pmatrix} a^2 - b^2 & ac - bd \\ ac - bd & c^2 - d^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus we have the simultaneous equations

$$a^2 = b^2 \qquad\qquad ac = bd \qquad\qquad c^2 = d^2.$$

Thus $b = \pm a$ and $d = \pm c$. If $a = b$, the remaining equation shows we must have $c = d$, while if $b = -a$ we must have $d = -c$. So the only possible form for A is

$$A = \begin{pmatrix} a & a \\ c & c \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} a & -a \\ c & -c \end{pmatrix},$$

where a and c are any numbers.

c) The value of the determinant is unchanged by subtracting the first column from the remaining one:

$$\det(A) = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1+a & 1 & 1 \\ 1 & 1 & 1+b & 1 \\ 1 & 1 & 1 & 1+c \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 1 & 0 & b & 0 \\ 1 & 0 & 0 & c \end{vmatrix}.$$

Thus $\det(A) = a \cdot b \cdot c$.

Solution 11.3:

a) Calculating, $A^2 = \begin{pmatrix} 7 & 2 \\ 3 & 10 \end{pmatrix}$.

The matrix A does not have the same number of columns as B has rows, so AB does not exist.

We have $BA = \begin{pmatrix} -7 & 6 \\ 6 & 4 \\ 5 & -2 \end{pmatrix}$.

Only square matrices have inverses; thus B^{-1} does not exist.

We have $B^{\mathrm{T}}B = \begin{pmatrix} 20 & -6 \\ -6 & 6 \end{pmatrix}$.

b) We perform row operations on the augmented matrix

$$\begin{pmatrix} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -1 & 3 & -2 & 1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 1 \end{pmatrix} \quad \begin{matrix} [R_2' = R_2 - 2R_1]\,, \\ [R_3' = R_3 - R_1]\,, \end{matrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 5 & -3 & 2 & 0 \\ 0 & 1 & -3 & 2 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 \end{pmatrix} \quad \begin{matrix} [R_2' = -R_2]\,, \\ [R_1' = R_1 - 2R_2']\,, \\ [R_3' = R_3 + R_2']\,. \end{matrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 2 & -3 & 5 \\ 0 & 1 & 0 & -1 & 2 & -3 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{pmatrix} \quad \begin{matrix} [R_3' = -R3]\,, \\ [R_1' = R_1 - 5R_3']\,, \\ [R_2' = R_2 - 3R_3']\,. \end{matrix}$$

Thus $A^{-1} = \begin{pmatrix} 2 & -3 & 5 \\ -1 & 2 & -3 \\ -1 & 1 & -1 \end{pmatrix}$.

Solution 11.4:

a) The product exists provided the first matrix as the same number of columns as the second has rows. Thus

$$AB = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad BC = \begin{pmatrix} -1 & -2 & 0 \\ 2 & 4 & 0 \\ 1 & 2 & 0 \end{pmatrix}.$$

Since the sizes are wrong, BA does not exists, while $\det(AB)$ does not exist since AB is not square. Finally, $CB = 3$, a 1×1 matrix, so $CB^{-1} = \frac{1}{3}$.

b) The transpose of a matrix is the matrix obtained by interchanging rows and columns. In this case,

$$A \cdot A^T = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & c \\ 0 & d \end{pmatrix} = \begin{pmatrix} a^2 & ac \\ ac & c^2 + d^2 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 25 \end{pmatrix}.$$

Thus $a^2 = 4$, $ac = 6$ and $c^2 + d^2 = 25$. We examine each solution of $a^2 = 4$ in turn. If $a = 2$, then $c = 3$ and $9 + d^2 = 25$ so $d = \pm 4$. If $a = -2$, $c = -3$ and again $d = \pm 4$. This gives four possible values for A, namely

$$\begin{pmatrix} 2 & 0 \\ 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 \\ 3 & -4 \end{pmatrix}, \quad \begin{pmatrix} -2 & 0 \\ -3 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -2 & 0 \\ -3 & -4 \end{pmatrix}.$$

Solution 11.5:

a) We subtract the first column from columns 2 and 4, and then expand along the first row.

$$\begin{vmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 1 & -1 & 1 & -\lambda \end{vmatrix},$$

$$= -1\begin{vmatrix} 0 & 1 \\ 1 & -\lambda \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ -1 & -\lambda \end{vmatrix} - 1\begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix},$$

$$= -1)(-1) - 1(-\lambda + 1) - 1.1 = 1 - \lambda.$$

The determinant thus vanishes if and only if $\lambda = 1$.

b) Here is the MAPLE.

```
> with(linalg):
> A:=matrix([[1,1,0,3],[2,1,1,7-alpha],[1,2,alpha,10]]);
```

$$A := \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & 1 & 7-\alpha \\ 1 & 2 & \alpha & 10 \end{bmatrix}$$

```
> B:=gausselim(A);
```

$$B := \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & 1 & 1-\alpha \\ 0 & 0 & \alpha+1 & 8-\alpha \end{bmatrix}$$

```
> x:=alpha->backsub(B):x(alpha);subs(alpha=2,x(alpha));
```

$$\left[-\frac{-4\,\alpha+4+\alpha^2}{\alpha+1}, \frac{-\alpha+7+\alpha^2}{\alpha+1}, -\frac{-8+\alpha}{\alpha+1}\right]$$

$$[0,\ 3,\ 2]$$

From the last line of the reduction, we see there are no solutions if the coefficient of x_3 in the last equation becomes zero, in which case, the equation becomes the inconsistent $0.x_3 = 9$. Thus there are no solutions when $\alpha = -1$.

If $\alpha = 2$, the last equation becomes $3x_3 = 6$, so $x_3 = 2$. Back substitution gives $x_2 = 3$ and then $x_1 = 0$ as the MAPLE shows.

Solution 11.6:

a) Calculating, we have

$$\det(A) = (2-\lambda)(1-\lambda) - 6 = \lambda^2 - 3\lambda - 4 = (\lambda-4)(\lambda+1).$$

The matrix A is invertible precisely when $\det(A) \neq 0$. Thus A is invertible unless $\lambda = 4$ or $\lambda = -1$.

b) Here is the MAPLE.

```
> A:=matrix(3,4,[1,-2,1,9,2,-3,-1,4,-1,2+t,3,s-14*t]);
```

$$A := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 2 & -3 & -1 & 4 \\ -1 & 2+t & 3 & -14t+s \end{bmatrix}$$

```
> B:=pivot(A,1,1);
```

$$B := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 0 & 1 & -3 & -14 \\ 0 & t & 4 & 9-14t+s \end{bmatrix}$$

```
> C:=addrow(B,2,3,-t);
```

$$C := \begin{bmatrix} 1 & -2 & 1 & 9 \\ 0 & 1 & -3 & -14 \\ 0 & 0 & 3t+4 & 9+s \end{bmatrix}$$

From the last line of the reduction, we see there is a unique solution unless the coefficient of z in the last equation becomes zero, in which case, the equation becomes $0.z = 9 + s$. Thus there are no solutions when $t = -4/3$ unless $s = -9$. If $s = -9$, we have an infinite family of solutions of the form

$$y = 3z - 14, \qquad x = 9 + 2y - z = 5z - 19,$$

where z is a free parameter.

If $t = -1$ and $s = -5$ the last equation becomes $z = 4$. Using back substitution, we see first that $y = -2$ and then that $x = 1$.

Solution 11.7:

a) The value of the determinant is unchanged by subtracting the first column from the remaining one:

$$\det(A) = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 7 \\ 1 & 4 & 7 & 10 \\ 1 & 5 & 9 & 13 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 4 & 6 \\ 1 & 3 & 6 & 9 \\ 1 & 4 & 8 & 12 \end{vmatrix}.$$

Now subtracting twice the second column from the third gives a column consisting entirely of zeros, and hence $\det(A) = 0$.

b) We use the first equation to eliminate the x - terms in the second and third equations, to give

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & 2, \\ & & y & - & tz & = & -1, \\ & + & (2+t)y & - & 3z & = & 1. \end{array}$$

Now use the second equation to eliminate y from the subsequent equations, giving

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & 2, \\ & & y & - & tz & = & -1, \\ & & & & (t^2+2t-3)z & = & 3+t. \end{array}$$

This is now in reduced echelon form. Consider the third equation, $(t^2+2t-3)z = (t+3)(t-1)z = 3+t$. If $t=-3$, this reduces to the trivial equation $0z = 0$ and z is unrestricted. In that case, we get $y = -3z - 1$ and then $x = 1 - 4z$.

It remains to consider what happens when $t \neq -3$, in which case the third equation becomes $(t-1)z = 1$. If $t = 1$, we must have $0z = 1$, and there are no solutions. However if $t \neq 1$, we have $z = 1/(t-1)$ and then back substituting first for y and then for x as described above gives a unique solution.

Solution 11.8: The matrix inversion has the following MAPLE trace:

```
> with(linalg):
> A:=matrix([[1,6,2],[-1,3,2],[-2,-1,1]]);
```

$$A := \begin{bmatrix} 1 & 6 & 2 \\ -1 & 3 & 2 \\ -2 & -1 & 1 \end{bmatrix}$$

```
> A1:=concat(A,diag(1,1,1));
```

$$A1 := \begin{bmatrix} 1 & 6 & 2 & 1 & 0 & 0 \\ -1 & 3 & 2 & 0 & 1 & 0 \\ -2 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

```
> A2:=pivot(A1,1,1);
```

$$A2 := \begin{bmatrix} 1 & 6 & 2 & 1 & 0 & 0 \\ 0 & 9 & 4 & 1 & 1 & 0 \\ 0 & 11 & 5 & 2 & 0 & 1 \end{bmatrix}$$

```
> A3:=pivot(A2,2,2):A3:=mulrow(A3,2,1/9);
```

$$A3 := \begin{bmatrix} 1 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{-2}{3} & 0 \\ 0 & 1 & \frac{4}{9} & \frac{1}{9} & \frac{1}{9} & 0 \\ 0 & 0 & \frac{1}{9} & \frac{7}{9} & \frac{-11}{9} & 1 \end{bmatrix}$$

```
> A4:=pivot(A3,3,3):A4:=mulrow(A4,3,9);
```

$$A4 := \begin{bmatrix} 1 & 0 & 0 & 5 & -8 & 6 \\ 0 & 1 & 0 & -3 & 5 & -4 \\ 0 & 0 & 1 & 7 & -11 & 9 \end{bmatrix}$$

Solution 11.9: We have

$$(x-y)(x^2+xy+y^2) = x^3 - yx^2 + x^2y - xy^2 + xy^2 - y^3 = x^3 - y^3.$$

We first subtract the last column from the remaining two:

$$\begin{vmatrix} 1 & 1 & 1 \\ x & y & z \\ x^3 & y^3 & z^3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ x & y-x & z-x \\ x^3 & y^3-x^3 & z^3-x^3 \end{vmatrix} = (y-x)(z-x)\begin{vmatrix} 1 & 0 & 0 \\ x & 1 & 1 \\ x^3 & y^2+xy+x^2 & z^2+zx+x^2 \end{vmatrix}.$$

Now subtract the second column form the third,

$$\begin{aligned} &= (y-x)(z-x)\begin{vmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ x^3 & y^2+xy+x^2 & z^2-y^2+zx-xy \end{vmatrix}, \\ &= (y-x)(z-x)(z-y)\begin{vmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ x^3 & y^2+xy+x^2 & z+y+x \end{vmatrix}, \\ &= (y-x)(z-x)(z-y)(x+y+z). \end{aligned}$$

Thus $k = 1$.

Solution 11.10: Here is the row reduction

```
> A := matrix(3,3,[1,-1,-1,-2,1,2,2,-1,-1]);
```

$$A := \begin{bmatrix} 1 & -1 & -1 \\ -2 & 1 & 2 \\ 2 & -1 & -1 \end{bmatrix}$$

```
> B:=concat(A,diag(1,1,1)):B:=pivot(B,1,1);
```

$$B := \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & -2 & 0 & 1 \end{bmatrix}$$

```
> C:=mulrow(B,2,-1):C:=pivot(C,2,2);
```

$$C := \begin{bmatrix} 1 & 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -2 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

```
> E:=pivot(C,3,3);
```

$$E := \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -2 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$