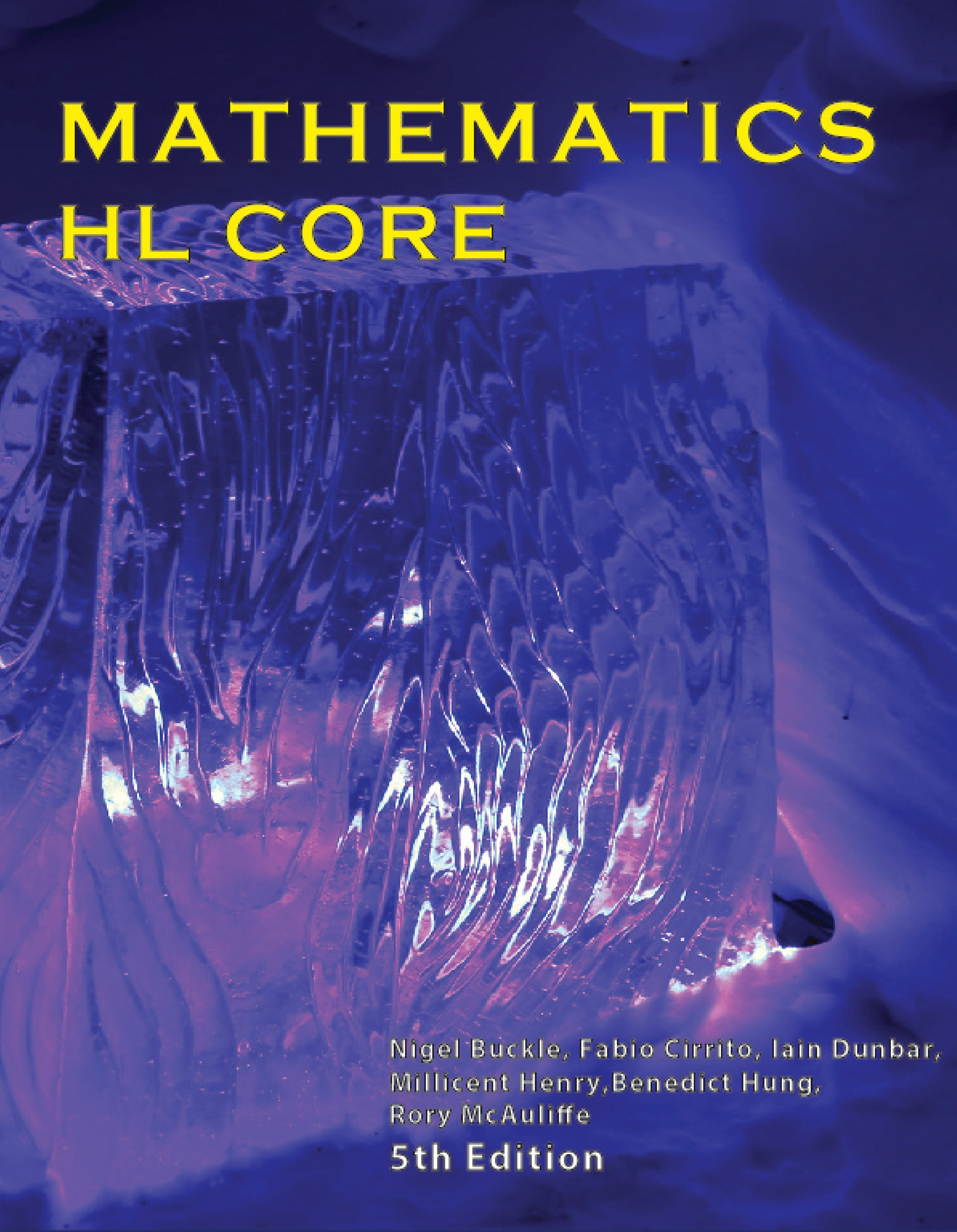
MATHEMATICS HL CORE



Nigel Buckle, Fabio Cirrito, Iain Dunbar,
Millicent Henry, Benedict Hung,
Rory McAuliffe

5th Edition

FOR USE WITH THE I.B. DIPLOMA PROGRAMME

MATHEMATICS HL CORE

Nigel Buckle,

Fabio Cirrito,

Iain Dunbar,

Millicent Henry,

Benedict Hung,

Rory McAuliffe.

5th Edition

www.ibid.com.au

First published in 2017 by IBID Press, Victoria

Library Catalogue:

Buckle, et. al.

1. Mathematics

2. International Baccalaureate.

Series Title: International Baccalaureate in Detail

ISBN-978-1-921784-81-1

IBID Press express their thanks to the International Baccalaureate Organization for permission to reproduce its intellectual property.

Cover design by Key-Strokes.

Published by IBID Press, 36 Quail Crescent, Melton, 3337, Australia.

Printed in China through Red Planet Print Management

PREFACE

This essential text for the Maths HL course has been prepared to closely align with the current course.

It has concise explanations, clear diagrams and calculator references.

Appropriate, graded exercises are provided throughout.

Also relating to International Perspectives and the Theory of Knowledge, it provides more than just the basics. It is an 'essential' resource for those teachers and students who are looking for a reliable guide for their HL course.

This is a re-worked and revised edition of the Higher Level text first published by IBID Press in 1997.

2nd Edition published in 1999

3rd Edition published in 2004

4th Edition published 2012

5th Edition 2017 – compact and better than ever!

QR Codes

There are many of these throughout the book. They link to additional resources that can be accessed through the internet using tablet, smart phone or computer. Readers will need to acquire a QR Reader 'app' if they do not already have one.

There are four main types of QR Files:

Extra Questions

Many exercises are provided with additional questions in pdf form.

Answers

Each chapter has a pdf version of the answers to all the questions (including 'extras').

Video

YouTube videos that relate to some examples.

3d files

These are Collada files (.dae). Readers may need to obtain Collada Reader to access these.

Calculators

Students who are thoroughly familiar with the capabilities of their model of calculator place themselves at a considerable advantage over students who are not.

In preparing a text such as this, we cannot provide an exhaustive account of every place in which a calculator can help. Or, for that matter, an explanation of how each model works!

This text uses examples from advanced Casio and Texas Instruments graphic calculators.

The manufacturers all provide extensive 'manuals'. These can be intimidating.

We suggest that a good strategy is to take each topic and, as you are learning it, take some time to discover your model's capability in that topic.

For example, Section 1.3 deals with counting principles. It is highly likely your calculator will be very helpful here. A good strategy can be to 'Google' or 'Bing' your model plus the topic.

There are now a number of training videos available on YouTube.

Answers

Answers to the Exercises are available using QR codes or as a download from the publisher's website:

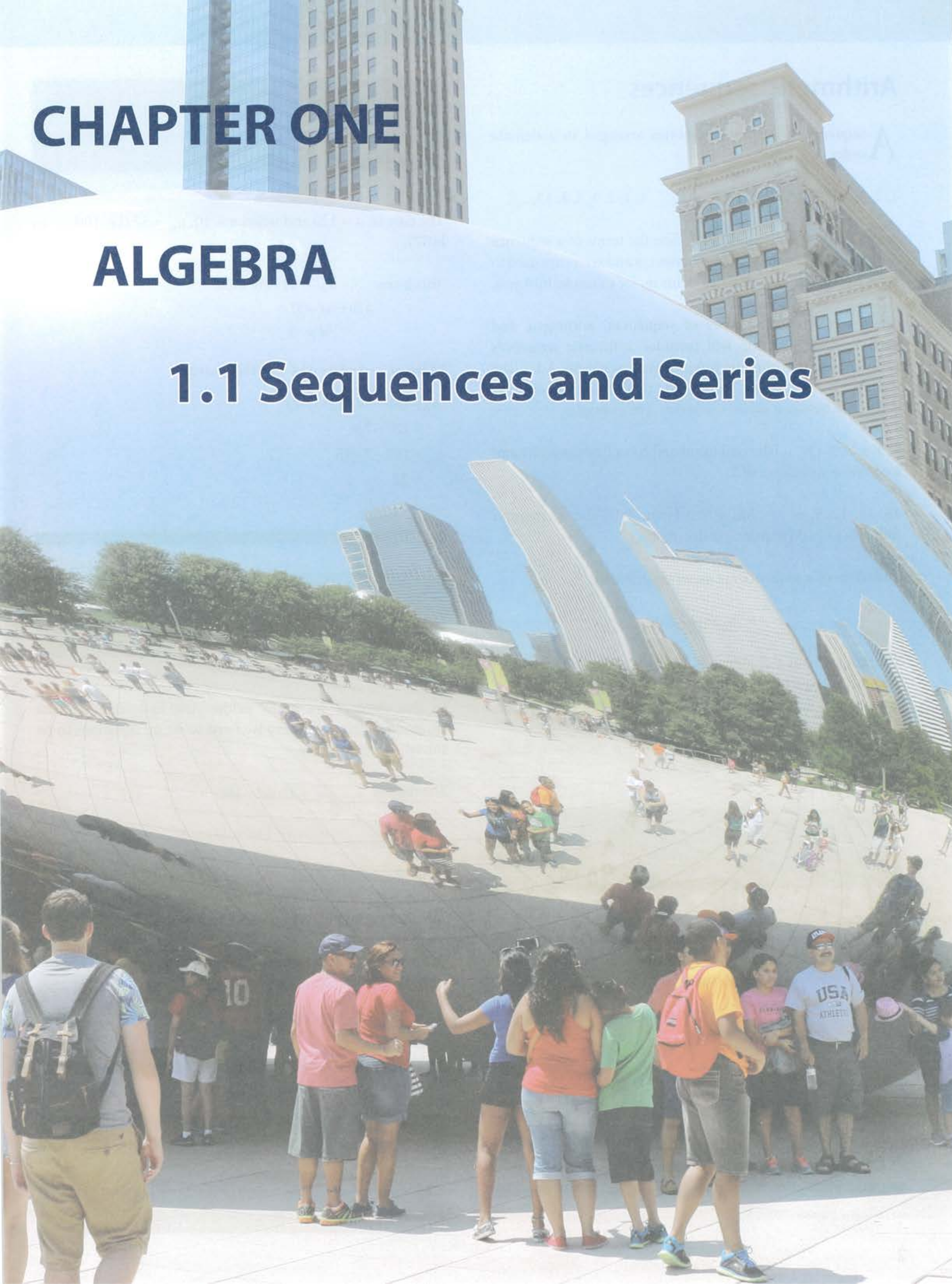
www.ibid.com.au

TABLE OF CONTENTS

CHAPTER ONE

ALGEBRA

1.1 Sequences and Series



Arithmetic sequences

A sequence is a set of quantities arranged in a definite order.

1, 2, 3, 4, 5, 6, ... −1, 2, −4, 8, −16, ... 1, 1, 2, 3, 5, 8, 13, ...

are all examples of sequences. When the terms of a sequence are added, we obtain a series. Sequences and series are used to solve a variety of practical problems in, for example, business.

There are two major types of sequences, arithmetic and geometric. This section will consider arithmetic sequences (also known as arithmetic progressions, or simply A.P.). The characteristic of such a sequence is that there is a common difference between successive terms. For example:

1, 3, 5, 7, 9, 11, . . . (the odd numbers) has a first term of 1 and a common difference of 2.

18, 15, 12, 9, 6, . . . has a first term of 18 and a common difference of −3 (sequence is decreasing).

The terms of a sequence are generally labelled:

$$u_1, u_2, u_3, u_4, \ldots, u_n .$$

The 'nth term' of a sequence is labelled u_n. In the case of an arithmetic sequence which starts with a and has a common difference of d, the nth term can be found using the formula:

$u_n = a+(n-1)d$ where $d = u_2 - u_1 = u_3 - u_2 = \ldots$ n = 1, 2,, 3...

Example 1.1.1

For the sequence 7, 11, 15, 19, ... , find the 20th term.

In this case, a = 7 and d = 4 because the sequence starts with a 7 and each term is 4 larger than the one before it, i.e. d = 11 − 7 = 4. Therefore the nth term is given by

$$u_n = 7+(n-1)4$$

That is, $u_n = 4n+3$

$$u_{20} = 4\times 20+3 = 83$$

(n = 20 corresponds to the 20th term)

Example 1.1.2

An arithmetic sequence has a first term of 120 and a 10th term of 57. Find the 15th term.

The data is: a = 120 and when n = 10, u_{10} = 57 (i.e. 10th term is 57).

This gives,
$$\begin{aligned} u_{10} &= 120+(10-1)d \\ 120+9d &= 57 \\ d &= -7 \end{aligned}$$

Using $u_n = a+(n-1)d$, we then have:

$$\begin{aligned} u_n &= 120+(n-1)\times(-7) \\ &= 127-7n \\ u_{15} &= 127-7\times 15 \\ &= 22 \end{aligned}$$

Example 1.1.3

An arithmetic sequence has a 7th term of 16.5 and a 12th term of 24. Find the 24th term.

In this instance we know neither the first term nor the common difference and so we need to set up equations to be solved simultaneously.

The data is: $u_7 = a+6d = 16.5$ – (1)

$u_{12} = a+11d = 24$ – (2)

We first solve for 'd':(2) − (1) : $5d = 7.5 \Leftrightarrow d = 1.5$

Substituting into (1): $a+6\times 1.5 = 16.5 \Leftrightarrow a = 7.5$

To find the 24th term we use the general term: $u_n = a+(n-1)d$ with n = 24:

$u_{24} = 7.5+(24-1)\times 1.5 = 42$

Example 1.1.4

A car whose original value was \$25 600 decreases in value by \$90 per month. How long will it take before the car's value falls below \$15 000?

The values can be seen as a sequence: \$25 600, \$25 510, \$25 420 etc.

In this case a = 25 600 and d =25 510 − 25 600 = −90 so that:

$$\begin{aligned} u_n &= 15600+(n-1)\times(-90) \\ &= 25690-90n \\ 15000 &= 25690-90n \\ 90n &= 25690-15000 \\ n &= 118.77 \end{aligned}$$

The car will be worth less than \$15 000 after 119 months.

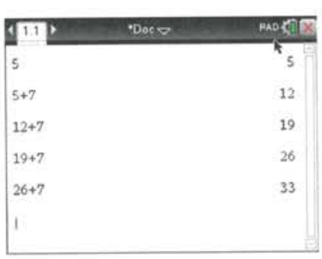
On 'final approach' a pilot aims to hold the airspeed constant and to descend in a straight line. As a result, if measured at regular intervals, range to the 'piano keys' and altitude form arithmetic sequences.

Using a graphics calculator

Most graphic calculators have an automatic memory facility (often called **Ans**) that stores the result of the last calculation as well as an ability to remember the actual calculation. This can be very useful in listing a sequence.

Example 1.1.5

List the arithmetic sequence 5, 12, 19, 26, ...



The sequence has a first term of 5. Enter this and press ENTER or EXE.

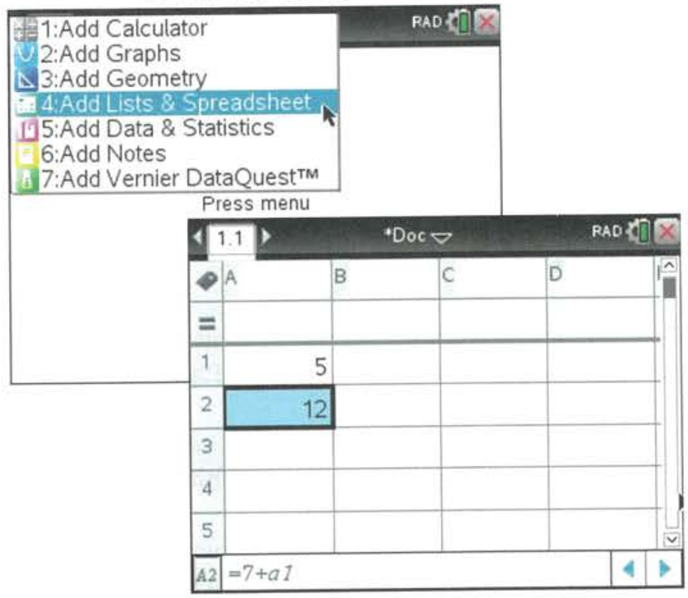
The common difference of the sequence is 7 so enter + 7.

The display will show Ans + 7 which means 'add 7 to the previous answer'.

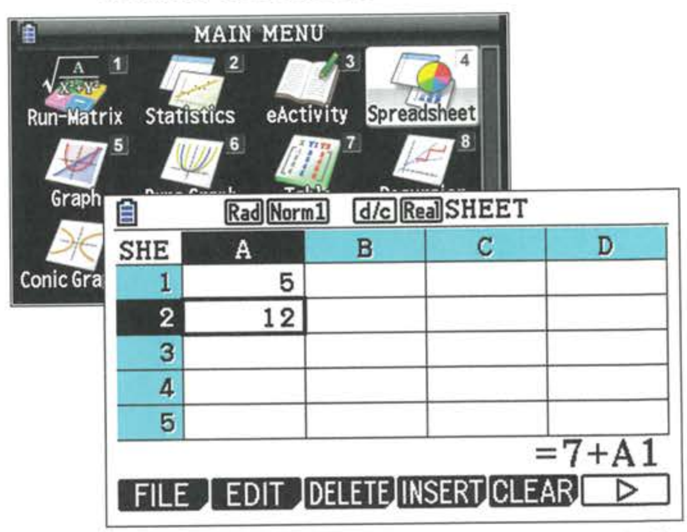
From here, every time you press ENTER (or EXE), you will repeat the calculation, generating successive terms of the sequence.

However most calculators are more sophisticated than this. It is possible to set up a spreadsheet type file. These can be used much as one might use Excel to make repetitive calculations.

TI models



Casio models



We now consider Example 8.2, where we obtained the sequence $u_n = 127-7n$ and wished to determine the 15th term.

Most calculators have many features that can be used with sequences. Become familiar with all of them for your model.

Exercise 1.1.1

1.

a Show that the following sequences are arithmetic.

b Find the common difference.

c Define the rule that gives the nth term of the sequence.

i $\{2, 6, 10, 14, \dots\}$

ii $\{20, 17, 14, 11, \dots\}$

iii $\{1, -4, -9, \dots\}$

iv $\{0.5, 1.0, 1.5, 2.0, \dots\}$

v $\{y+1, y+3, y+5, \dots\}$

vi $\{x+2, x, x-2, \dots\}$

2. Find the 10th term of the sequence whose first four terms are 8, 4, 0, –4.

3. Find the value of x and y in the arithmetic sequence $\{5, x, 13, y, \dots\}$.

4. An arithmetic sequence has 12 as its first term and a common difference of –5. Find its 12th term.

5. An arithmetic sequence has –20 as its first term and a common difference of 3. Find its 10th term.

6. The 14th term of an arithmetic sequence is 100. If the first term is 9, find the common difference.

7. The 10th term of an arithmetic sequence is –40. If the first term is 5, find the common difference.

8. If $n + 5$, $2n + 1$ and $4n - 3$ are three consecutive terms of an arithmetic sequence, find n.

9. The first three terms of an arithmetic sequence are 1, 6, 11.

a. Find the 9th term.

b. Which term will equal 151?

10. Find x and y given that $4-\sqrt{3}, x, y$ and $2-\sqrt{3}$ are the first four terms of an arithmetic sequence.

11. For each of the following sequences, determine:

i. its common difference

ii. its first term

a. $u_n = -5+2n, n \geq 1$

b. $u_n = 3+4(n+1), n \geq 1$

12. The third and fifth terms of an A.P. are $(x + y)$ and $(x - y)$ respectively. Find the twelfth term.

13. The sum of the fifth term and twice the third of an arithmetic sequence equals the twelfth term. If the seventh term is 25 find an expression for the general term, u_n.

14. For a given arithmetic sequence, $u_n = m$ and $u_m = n$. Find:

a. the common difference.

b. u_{n+m}.

Arithmetic series

If the terms of a sequence are added, the result is known as a series.

The **sequence:** 1, 2, 3, 4, 5, 6, . . .

gives the **series:** 1 + 2 + 3 + 4 + 5 + 6 + . . .

and the **sequence:** –1, –2, –4, –8, –16 . . .

gives the **series:** (–1) + (–2) + (–4) + (–8) + (–16) + . . .

(or – 1– 2 – 4 – 8 – 16 – . . .)

The sum of the terms of a series is referred to as S_n, the **sum of n terms of a series.**

For an arithmetic series, we have:

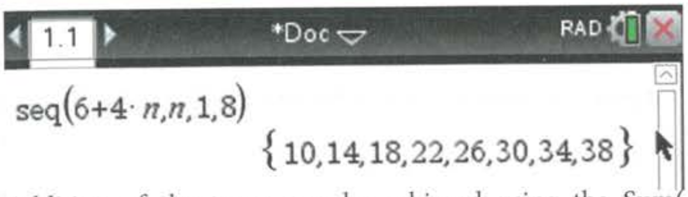
$$\begin{aligned} S_n &= u_1 + u_2 + u_3 + \dots u_n \\ &= a + (a+d) + (a+2d) + (a+3d) + \dots + (a+(n-1)d) \end{aligned}$$

For example, if we have a sequence defined by $u_n = 6+4n, n \geq 1$ then the sum of the first 8 terms is given by:

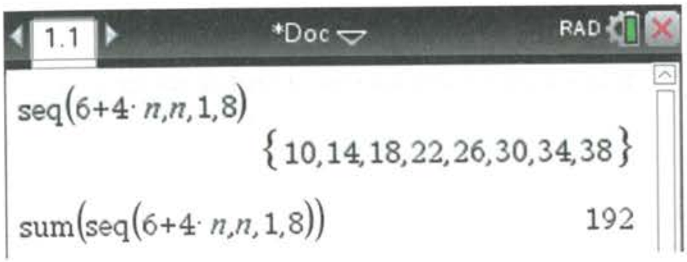
$$\begin{aligned} S_8 &= u_1 + u_2 + u_3 + \dots u_8 \\ &= 10+14+18+\dots+38 \\ &= 192 \end{aligned}$$

Most calculators have several ways of handling sequences and series. We have already referred to the LIST (spreadsheet) feature.

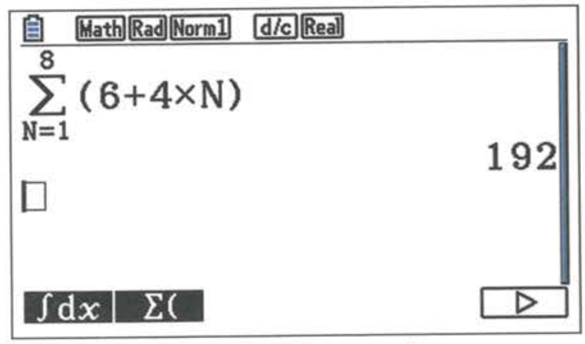
An alternative (TI) is to use MENU, 6 (STATISTICS), 4 (LIST OPERATIONS),5 (SEQUENCE) - the exact calculator method will vary - consult the manual!:



Addition of the terms can be achieved using the **Sum(** function. This can be found under the 'book' key:



Casio provide a summation function in run mode. This can be found under F4-MATH, F6-▷, F2-Σ. Fill the boxes and press EXE:



The sigma notation is discussed later in this section.

There will be many cases in which we can add the terms of a series in this way. If, however, there are a large number of terms to add, a formula is more appropriate.

There is a story that, when the mathematician Gauss was a child, his teacher was having problems with him because he always finished all his work long before the other students. In an attempt to keep Gauss occupied for a period, the teacher asked him to add all the whole numbers from 1 to 100. '5050' Gauss replied immediately.

It is probable that Gauss used a method similar to this:

1	2	3	4	5	6	......,	96	97	98	99	100
100	99	98	97	96	95	......,	5	4	3	2	1
101	**101**	**101**	**101**	**101**	**101**	......,	**101**	**101**	**101**	**101**	**101**

Adding each of the pairings gives 100 totals of 101 each. This gives a total of 10100. This is the sum of two sets of the numbers $1 + 2 + 3 + ... + 98 + 99 + 100$ and so dividing the full answer by 2 gives the answer 5050, as the young Gauss said.

It is then possible to apply the same approach to such a sequence, bearing in mind that the sequence of numbers must be arithmetic.

Applying this process to the general arithmetic series we have:

a	$a+d$	$a+2d$	...	$a+(n-3)d$	$a+(n-2)d$	$a+(n-1)d$
$a+(n-1)d$	$a+(n-2)d$	$a+(n-3)d$	...	$a+2d$	$a+d$	a

Each of the pairings comes to the same total.

Here are some examples:

1st pairing: $a + a + (n-1)d = 2a + (n-1)d$

2nd pairing: $a + d + a + (n-2)d = 2a + (n-1)d$

3rd pairing: $a + 2d + a + (n-3)d = 2a + (n-1)d$

There are n such pairings so: $2\times S_n = n\times\left[2a+(n-1)d\right]$ That is, $S_n = \frac{n}{2}\times\left[2a+(n-1)d\right]$

Giving the formula, for the sum of n terms of a sequence: $\boxed{S_n = \frac{n}{2}\times\left[2a+(n-1)d\right]}$

Note also that $S_n = \frac{n}{2}(u_1+u_n)$ because $S_n = \frac{n}{2}(u_1+u_n) = \frac{n}{2}\left(u_1+(u_1+(n-1)d)\right) = \frac{n}{2}(2u_1+(n-1)d)$ as above - and $(a = u_1)$.

This formula can now be used to sum large arithmetic series:

Example 1.1.6

Find the sum of 20 terms of the series $-2 + 1 + 4 + 7 + 10 + \ldots$

We have the following information: $a = u_1 = -2$

and $d = u_2 - u_1 = 1 - (-2) = 3$.

Then, the sum to n terms is given by: $S_n = \frac{n}{2} \times [2a + (n-1)d]$

So that the sum to 20 terms is given by

$$\begin{aligned} S_{20} &= \frac{20}{2} \times [2 \times (-2) + (20-1) \times 3] \\ &= 10[-4 + 19 \times 3] \\ &= 530 \end{aligned}$$

Example 1.1.7

Find the sum of 35 terms of the series: $-\frac{3}{8} - \frac{1}{8} + \frac{1}{8} + \frac{3}{8} + \frac{5}{8} + \ldots$

We have the following information: $a = u_1 = -\frac{3}{8}$

and $d = u_2 - u_1 = -\frac{1}{8} - \left(-\frac{3}{8}\right) = \frac{1}{4}$.

Then, with $n = 35$ we have

$$\begin{aligned} S_{35} &= \frac{35}{2}\left[2 \times \left(-\frac{3}{8}\right) + (35-1) \times \frac{1}{4}\right] \\ &= 17.5\left[-\frac{3}{4} + 34 \times \frac{1}{4}\right] \\ &= 135\frac{5}{8} \end{aligned}$$

Example 1.1.8

An arithmetic series has a third term of 0. The sum of the first 15 terms is –300. What is the first term and the sum of the first ten terms?

From the given information we have: $u_3 = a + 2d = 0$ – (1)

and: $S_{15} = \frac{15}{2}[2a + 14d] = -300$

i.e. $5a + 105d = -300$

$\therefore a + 7d = -20$ – (2)

The pair of equations can now be solved simultaneously:

(2) – (1): $5d = -20 \Leftrightarrow d = -4$

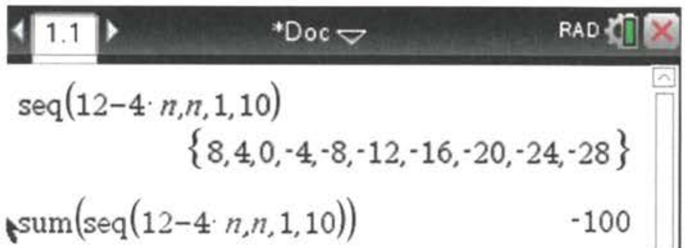
Substituting into (1) we have: $a + 2 \times -4 = 0 \Leftrightarrow a = 8$

This establishes that the series is $8 + 4 + 0 + (-4) + (-8) + \ldots$

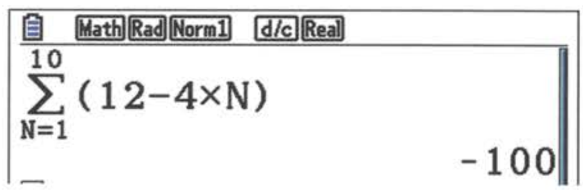
So the first term is 8 and the sum of the first ten terms is:

$S_{10} = \frac{10}{2}[16 + 9 \times -4] = -100$

Using the TI_NSpire we have, with the general term $12 - 4n$:



Using a Casio model, this can be evaluated:



Example 1.1.9

A new business is selling home computers. They predict that they will sell 20 computers in their first month, 23 in the second month, 26 in the third and so on, in arithmetic sequence. How many months will pass before the company expects to sell their thousandth computer?

The series is: $20 + 23 + 26 + \ldots$..

The question implies that the company is looking at the total number of computers sold, so we are looking at a series, not a sequence.

The question asks how many terms (months) will be needed before the total sales reach more than 1000. From the given information we have: $a = 20$, $d = 23 - 20 = 3$.

Therefore, we have the sum to n terms given by:

$$\begin{aligned} S_n &= \frac{n}{2}[2 \times 20 + (n-1) \times 3] \\ &= \frac{n}{2}[3n + 37] \end{aligned}$$

Next, we determine when $S_n = 1000$:

$$\frac{n}{2}[3n+37]=1000$$

$$3n^2+37n=2000$$

$$3n^2+37n-2000=0$$

Solving for n can be done using several methods:

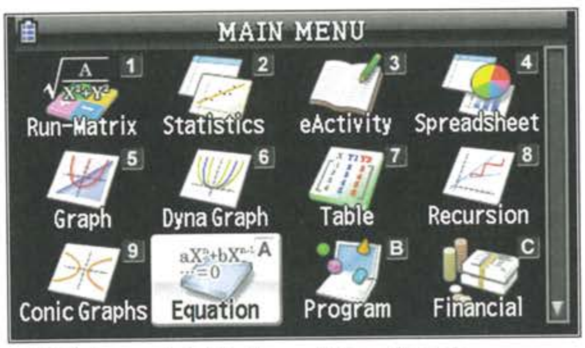
Method 1: Quadratic formula

$$n=\frac{-37\pm\sqrt{37^2-4\times3\times-2000}}{2\times3}$$

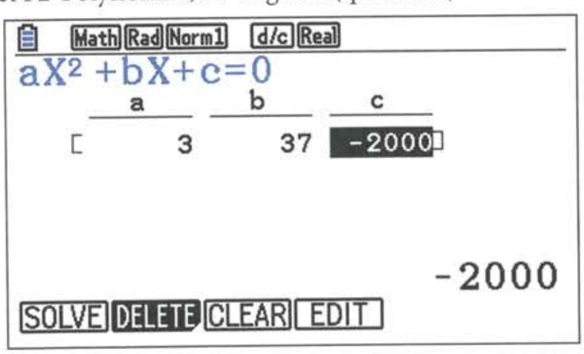
$$=20.37...[-32.7]$$

Method 2: Graphics Calculator Solve function

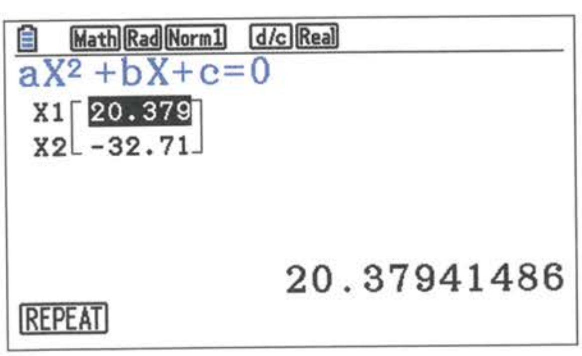
If using Casio, select the Equations module:



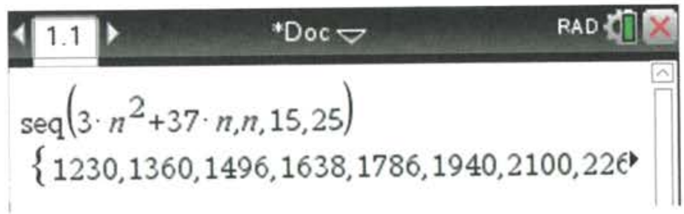
Select F2-Polynomial, F1-degree2 (quadratic)



Then F1 will initiate solve:

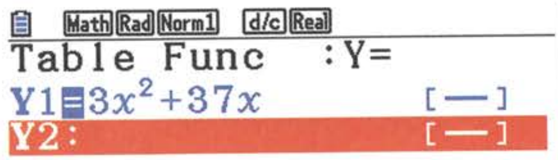


Method 3: Table of values

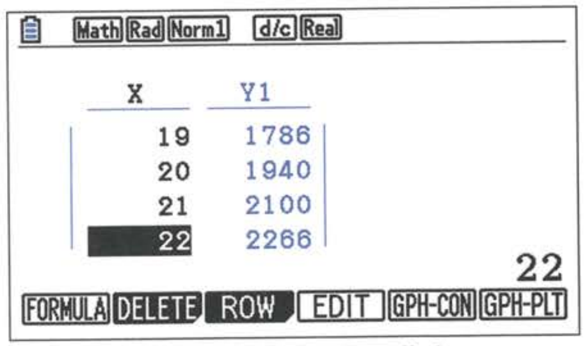


Casio: Using the Table Module (7)

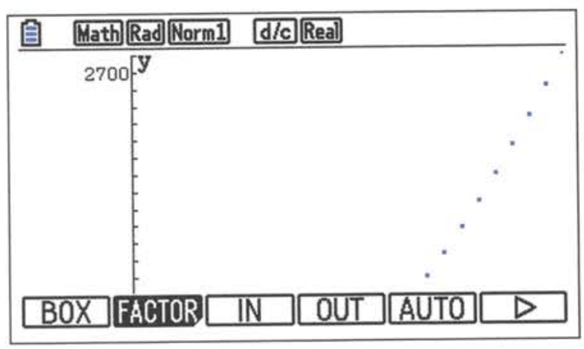
Enter the rule (using x as the variable can save time):



Use F5-SET to set the values of the variable (15 to 25) and then EXE F6-Table:



A nice feature of the Casio is that it will show you a graph. Press F6 and Shift F2-ZOOM, F5-AUTO (to find the points),



Notice that we have entered the expression for S_n as the sequence rule. In fact, the series itself is made up of terms in a sequence of so-called partial sums, often called a sum sequence.

That is, we have that $\{S_1, S_2, S_3, ...\} = \{15, 33, 54, ...\}$ forms a sequence.

The answer then, is that the company will sell its thousandth computer during the 20th month.

Exercise 1.1.2

1. Find the sum of the first ten terms in the arithmetic sequences

   a $\{1, 4, 7, 10, \dots\}$

   b $\{3, 9, 15, 21, \dots\}$

   c $\{10, 4, -2, \dots\}$.

2. For the given arithmetic sequences, find the sum, S_n, to the requested number of terms.

   a $\{4, 3, 2, \dots\}$ for $n = 12$

   b $\{4, 10, 16, \dots\}$ for $n = 15$

   c $\{2.9, 3.6, 4.3, \dots\}$ for $n = 11$

3. Find the sum of the following sequences:

   a $\{5, 4, 3, \dots, -15\}$

   b $\{3, 9, 15, \dots, 75\}$

   c $\{3, 5, 7, \dots, 29\}$

4. The weekly sales of washing machines from a retail store that has just opened in a new housing complex increase by 2 machines per week. In the first week of January 1995, 24 machines were sold.

   a How many were sold in the last week of December 1995?

   b How many machines did the retailer sell in 1995?

   c When was the 500th machine sold?

5. The fourth term of an arithmetic sequence is 5 while the sum of the first 6 terms is 10. Find the sum of the first nineteen terms.

6. Find the sum of the first 10 terms for the sequences defined by:

   a $u_n = -2 + 8n$

   b $u_n = 1 - 4n$

7. The sum of the first eight terms of the sequence $\{\ln x, \ln x^2 y, \ln x^3 y^2, \dots\}$ is given by $4(a\ln x + b\ln y)$. Find a and b.

Sigma notation

There is a second notation to denote the sum of terms. This other notation makes use of the Greek letter Σ as the symbol to inform us that we are carrying out a summation.

In short, Σ stands for 'The sum of . . .'.

This means that the expression

$$\sum_{i=1}^{n} u_i = u_1 + u_2 + u_3 + \dots + u_{n-1} + u_n$$

For example, if $u_i = 2 + 5(i-1)$, i.e. an A.P. with first term $a = 2$ and common difference $d = 5$, the expression:

$$S_n = \sum_{i=1}^{n} [2 + 5(i-1)]$$

would represent the sum of the first n terms of the sequence.

So, the sum of the first 3 terms would be given by:

$$\sum_{i=1}^{3} [2 + 5(i-1)]$$

$$= [2 + 5(1-1)] + [2 + 5(2-1)] + [2 + 5(3-1)]$$

Term 1   Term2   Term 3

$= 2 + 7 + 12$

$= 21$

Properties of Σ

1. $\sum$ is distributive. That is, $\sum_{i=1}^{n} [u_i + v_i] = \sum_{i=1}^{n} u_i + \sum_{i=1}^{n} v_i$.

2. $\sum_{i=1}^{n} k u_i = k \sum_{i=1}^{n} u_i$, for some constant value k.

3. $\sum_{i=1}^{n} k = kn$, i.e. adding a constant term, k, n times is the same as multiplying k by n.

Example 1.1.10

Given that $u_i = 5 + 2i$ and $v_i = 2 - 5i$, find:

a $\sum_{i=1}^{5} u_i$ b $\sum_{i=1}^{5} [2u_i - v_i]$

c $\sum_{i=1}^{1000} [5u_i + 2v_i]$

a
$$\begin{aligned}\sum_{i=1}^{5} u_i &= u_1 + u_2 + u_3 + u_4 + u_5 \\ &= [5+2]+[5+4]+[5+6]+[5+8]+[5+10] \\ &= 7+9+11+13+15 \\ &= 55\end{aligned}$$

b
$$\begin{aligned}\sum_{i=1}^{5} [2u_i - v_i] &= \sum_{i=1}^{5} [2u_i] + \sum_{i=1}^{5} [-v_i] \\ &= 2\sum_{i=1}^{5} [u_i] - \sum_{i=1}^{5} [v_i]\end{aligned}$$

Now,
$$2\sum_{i=1}^{5} [u_i] = 2 \times 55 = 110$$

and (Using properties)
$$\begin{aligned}\sum_{i=1}^{5} v_i &= \sum_{i=1}^{5} (2-5i) \\ &= \sum_{i=1}^{5} (2) - 5\sum_{i=1}^{5} (i) \\ &= 2\times 5 - 5[1+2+3+4+5] \\ &= -65\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{5} [2u_i - v_i] &= 110 - (-65) \\ &= 175\end{aligned}$$

c
$$\begin{aligned}\sum_{i=1}^{1000} [5u_i + 2v_i] &= \sum_{i=1}^{1000} [5(5+2i) + 2(2-5i)] \\ &= \sum_{i=1}^{1000} [25 + 10i + 4 - 10i] \\ &= \sum_{i=1}^{1000} 29 \\ &= 29 \times 1000 \\ &= 29000\end{aligned}$$

In this example we have tried to show that there are a number of ways to obtain a sum. It is not always necessary to enumerate every term and then add them. Often, an expression can first be simplified.

Exercise 1.1.3

1. Find the twentieth term in the sequence 9, 15, 21, 27, 33, . . .

2. Fill the gaps in this arithmetic sequence: –3, _, _, _, _, _, 12.

3. An arithmetic sequence has a tenth term of 17 and a fourteenth term of 30. Find the common difference.

4. If $u_{59} = \frac{1}{10}$ and $u_{100} = -1\frac{19}{20}$ for an arithmetic sequence, find the first term and the common difference.

5. Find the sum of the first one hundred odd numbers.

6. An arithmetic series has twenty terms. The first term is –50 and the last term is 83, find the sum of the series.

7. Thirty numbers are in arithmetic sequence. The sum of the numbers is 270 and the last number is 38. What is the first number?

8. How many terms of the arithmetic sequence: 2, 2.3, 2.6, 2.9, . . . must be taken before the sum of the terms exceeds 100?

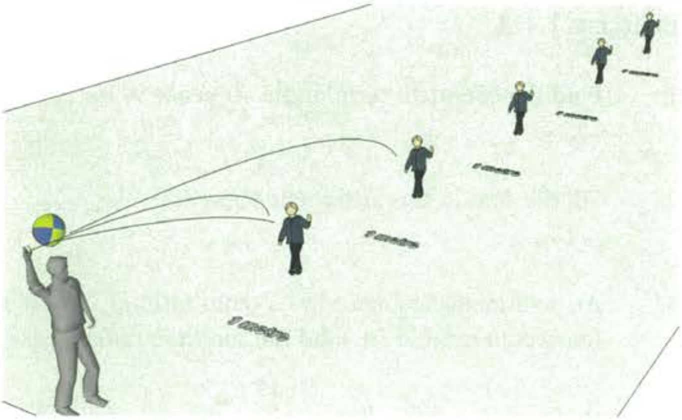
9. Sandip and Melissa save \$50 in the first week of a savings program, \$55 in the second week, \$60 in the third and so on, in arithmetic progression. How much will they save in ten weeks? How long will they have to continue saving if their target is to save \$5000?

10. A printing firm offers to print business cards on the following terms: \$45 for design and typesetting and then \$0.02 per card.

    a What is the cost of 500 cards from this printer?

    b How many cards can a customer with \$100 afford to order?

11. A children's game consists of the players standing in a line with a gap of 2 metres between each. The child at the left-hand end of the line has a ball which s/he throws to the next child in the line, a distance of 2 metres. The ball is then thrown back to the first child who then throws the ball to the third child in the line, a distance of 4 metres. The ball is then returned to the first child, and so on until all the children have touched the ball at least once.



a If a total of five children play and they make the least number of throws so that only the leftmost child touches the ball more than once:

i What is the largest single throw?

ii What is the total distance travelled by the ball?

b If seven children play, what is the total distance travelled by the ball?

c If n children play, derive a formula for the total distance travelled by the ball.

d Find the least number of children who need to play the game before the total distance travelled by the ball exceeds 100 metres.

e The children can all throw the ball 50 metres at most.

i What is the largest number of children that can play the game?

ii What is the total distance travelled by the ball?

12. Find each sum.

a $\sum_{i=1}^{100} k$ b $\sum_{i=1}^{100}(2k+1)$ c $\sum_{i=1}^{51}(3k+5)$

13. If $-3 + 4i$ and $12 - 3i$ find:

a $\sum_{i=1}^{10}[u_i+v_i]$ b $\sum_{i=1}^{10}[3u_i+4v_i]$ c $\sum_{i=1}^{10}[u_iv_i]$

14.a Show that for an arithmetic sequence, $u_n = S_n - S_{n-1}$, where u_n is the nth term and S_n is the sum of the first n terms.

b Find the general term, u_n, of the A.P given that

$$\sum_{i=1}^{10} u_i = \frac{n}{2}(3n-1).$$

Geometric sequences

Attempts to understand the sizes of animal populations have often used sequences and series.

Sequences such as 2, 6, 18, 54, 162, ... and 200, 20, 2, 0.2, ... in which each term is obtained by multiplying the previous one by a fixed quantity are known as **geometric sequences.**

The sequence: 2, 6, 18, 54, 162, ... is formed by starting with 2 and then multiplying by 3 to get the second term, by 3 again to get the third term, and so on.

For the sequence 200, 20, 2, 0.2, ..., begin with 20 and multiply by 0.1 to get the second term, by 0.1 again to get the third term and so on.

The constant multiplier of such a sequence is known as the **common ratio.**

The common ratio of 2, 6, 18, 54, 162,... is 3 and of 200, 20, 2, 0.2,... it is 0.1.

The nth term of a geometric sequence is obtained from the first term by multiplying by $n-1$ common ratios.

This leads to the formula for the nth term of a geometric sequence:

$$u_n = a \times r^{n-1}$$

where $r = \frac{u_2}{u_1} = \ldots = \frac{u_n}{u_{n-1}}$ and n is the term number, a the first term and r is the common ratio.

Example 1.1.11

Find the tenth term in the sequence 2, 6, 18, 54, 162, ...

The first term is $a = 2$. The common ratio $r = 3 = {}^6/_2 = {}^{18}/_6$ and n, the required term, is 10.

Use the formula to solve the problem:

$$\begin{aligned} u_n &= a \times r^{n-1} \\ u_{10} &= 2 \times 3^{10-1} \\ &= 2 \times 3^9 \\ &= 39366 \end{aligned}$$

Example 1.1.12

Find the fifteenth term in the sequence 200, 20, 2, 0.2, ...

In this case, $a = 200$, $r = \frac{20}{200} = \frac{1}{10} = 0.1$ and $n = 15$.

Using the general term $u_n = a \times r^{n-1}$, the 15th term is given by:

$$\begin{aligned} u_{15} &= 200 \times 0.1^{(15-1)} \\ &= 200 \times 0.1^{14} \\ &= 2 \times 10^{-12} \end{aligned}$$

Example 1.1.13

Find the eleventh term in the sequence: $1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \frac{1}{16}, \ldots$

The sequence $1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \frac{1}{16}, \ldots$, has a common ratio of $r = -\frac{1}{2}$.

Using the general term , we have:

$$\begin{aligned} u_{11} &= 1 \times \left(-\frac{1}{2}\right)^{11-1} \\ &= \left(-\frac{1}{2}\right)^{10} \\ &\approx 0.000977 \end{aligned}$$

Many questions will be more demanding in terms of the way in which you use this formula. You should also recognise that the formula can be applied to a range of practical problems.

Many of these involve growth and decay and can be seen as similar to problems studied in 1.2.

Example 1.1.14

A geometric sequence has a fifth term of 3 and a seventh term of 0.75. Find the first term, the common ratio and the tenth term.

From the given information we can set up the following equations:

$$u_5 = a \times r^4 = 3 \qquad -(1)$$

and $$u_7 = a \times r^6 = 0.75 \qquad -(2)$$

As with similar problems involving arithmetic sequences, the result is a pair of simultaneous equations. In this case these can best be solved by dividing (2) by (1) to get:

$$\frac{a \times r^6}{a \times r^4} = \frac{0.75}{3} \Leftrightarrow r^2 = 0.25 \Leftrightarrow r = \pm 0.5$$

Substituting results into (1) we have: $a\left(\pm\frac{1}{2}\right)^4 = 3 \Leftrightarrow a = 48$

Therefore, the 10th term is given by: $u_{10} = 48 \times (\pm 0.5)^9 = \pm\frac{3}{32}$

There are two solutions: 48, 24, 12, 6, . . . (for the case $r = 0.5$) & 48, −24, 12, −6, . . . ($r = -0.5$).

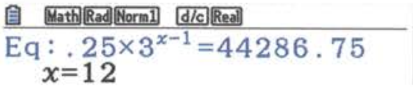
Example 1.1.15

Find the number of terms in the geometric sequence: 0.25, 0.75, 2.25, . . . , 44286.75.

The sequence 0.25, 0.75, 2.25, ..., 44286.75 has a first term $a = 0.25$ and a common ratio $r = \frac{0.75}{0.25} = 3$. In this problem it is n that is unknown. Substitution of the data into the formula gives: $u_n = 0.25 \times 3^{n-1} = 44286.75$

The equation that results can be solved using a calculator or logarithms (see Sec. 1.2)

If using a Casio calculator, select module A-Equation. Then choose F3-Solver. Enter the equation and press EXE, F6-SOLVE.



For those of you who have encountered logarithms, the analytic solution is:

$$
\begin{aligned}
3^{(n-1)} &= \left(\frac{44286.75}{0.25}\right) \\
&= 177147 \\
\log_{10}\left(3^{(n-1)}\right) &= \log_{10}(177147) \\
(n-1)\log_{10}(3) &= \log_{10}(177147) \\
n-1 &= \frac{\log_{10}(177147)}{\log_{10}(3)} \\
n-1 &= 11 \\
n &= 12
\end{aligned}
$$

Example 1.1.16

A car originally worth \$34 000 loses 15% of its value each year.

a Write a geometric sequence that gives the year-by-year value of the car.

b Find the value of the car after 6 years.

c After how many years will the value of the car fall below \$10 000?

a If the car loses 15% of its value each year, its value will fall to 85% (100% – 15%) of its value in the previous year. This means that the common ratio is 0.85 (the fractional equivalent of 85%). Using the formula, the sequence is: $u_n = 34000 \times 0.85^{(n-1)}$, i.e. \$34000, \$28900, \$24565, \$20880.25, . . .

b The value after six years have passed is the seventh term of the sequence. This is because the first term of the sequence is the value after no years have passed. $u_7 = 34000 \times 0.85^6 \approx 12823$ or \$12 823.

c

$$
\begin{aligned}
10000 &= 34000 \times 0.85^n \\
0.85^n &= 0.2941 \\
\log_{10}\left(0.85^n\right) &= \log_{10}(0.2941) \\
n\log_{10}(0.85) &= \log_{10}(0.2941) \\
n &= \frac{\log_{10}(0.2941)}{\log_{10}(0.85)} \\
n &\approx 7.53
\end{aligned}
$$

This means that the car's value will fall to \$10000 after about 7 years 6 months.

Example 1.1.17

The number of people in a small country town increases by 2% per year. If the population at the start of 1970 was 12500, what was the population at the start of the year 2010?

A quantity can be increased by 2% by multiplying by 1.02. Note that this is different from finding 2% of a quantity which is done by multiplying by 0.02. The sequence is 12500, 12500 ×1.02, 12500×1.02² etc. with $a = 12500$, $r = 1.02$.

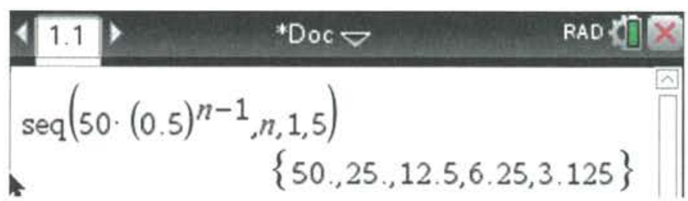
It is also necessary to be careful about which term is required. In this case, the population at the start of 1970 is the first term, the population at the start of 1971 the second term, and so on. The population at the start of 1980 is the eleventh term and at the start of 2010 we need the forty-first term:

$$
\begin{aligned}
u_{41} &= 12500 \times 1.02^{40} \\
&\approx 27600
\end{aligned}
$$

In all such cases, you should round your answer to the level given in the question or, if no such direction is given, round the answer to a reasonable level of accuracy.

Using a graphics calculator

As with arithmetic sequences, geometric sequences such as 50, 25, 12.5, . . . can be listed using a graphics calculator. For this sequence we have $a = 50$ and $r = 0.5$, so,



Exercise 1.1.4

1. Find the common ratio, the 5th term and the general term of the following sequences.

a 3, 6, 12, 24, ... b 3, 1, $\frac{1}{3}, \frac{1}{9}$

c $2, \frac{2}{5}, \frac{2}{25}, \frac{2}{125}, \ldots$ d –1, 4, –16, 64, ...

e $ab, a, \frac{a}{b}, \frac{a}{b^2}, \ldots$ f $a^2, ab, b^2, \ldots$

2. Find the value(s) of x if each of the following are in geometric sequence.

a 3, x, 48 b $\frac{5}{2}, x, \frac{1}{2}$

3. The third and seventh terms of a geometric sequence are 0.75 and 12 respectively.

   a Find the 10th term.

   b What term is equal to 3072?

4. A rubber ball is dropped from a height of 10 metres and bounces to reach $^5/_6$ of its previous height after each rebound. Let u_n be the ball's maximum height before its nth rebound.

   a Find an expression for u_n.

   b How high will the ball bounce after its 5th rebound.

   c How many times has the ball bounced by the time it reaches a maximum height of $^{6250}/_{1296}$ m.

5. The terms $k+4$, $5k+4$, $k+20$ are in a geometric sequence. Find the value(s) of k.

6. A computer depreciates each year to 80% of its value from the previous year. When bought the computer was worth \$8000.

   a Find its value after: i 3 years ii 6 years.

   b How long does it take for the computer to depreciate to a quarter of its purchase price?

7. The sum of the first and third terms of a geometric sequence is 40 while the sum of its second and fourth terms is 96. Find the sixth term of the sequence.

8. The sum of three successive terms of a geometric sequence is $^{35}/_2$, while their product is 125. Find the three terms.

9. The population in a town of 40000 increases at 3% per annum. Estimate the town's population after 10 years.

10. Following new government funding it is expected that the unemployed workforce will decrease by 1.2% per month. Initially there are 120000 people unemployed. How large an unemployed workforce can the government expect to report in 8 months time.

11. The cost of erecting the ground floor of a building is \$44000, for erecting the first floor it costs \$46200, to erect the second floor costs \$48510 and so on. Using this cost structure, how much will it cost to erect the 5th floor? What will be the total cost of erecting a building with six floors?

Geometric series

Wall Street in New York - where interest becomes a geometric series

When the terms of a geometric sequence are added, the result is a geometric series.

For example:

The sequence 3, 6, 12, 24, 48, . . . gives rise to the series: 3 + 6 + 12 + 24 + 48 + . . .

and, the sequence 24, −16, $10^2/_3$, $-7^1/_9$... leads to the series 24 −16 + $10^2/_3$ $-7^1/_9$+...

Geometric series can be summed using the formula that is derived by first multiplying the series by r:

$$S_n = a + ar + ar^2 + ar^3 ... + ar^{n-3} + ar^{n-2} + ar^{n-1}$$
$$rS_n = \quad ar + ar^2 + ar^3 ... + ar^{n-3} + ar^{n-2} + ar^{n-1} + ar^n$$

Subtracting the second equation from the first:

$$S_n - rS_n = a - ar^n$$
$$S_n(1-r) = a(1-r^n)$$
$$S_n = \frac{a(1-r^n)}{1-r}$$

This formula can also be written as: $S_n = \dfrac{a(r^n-1)}{r-1}, r \neq 1$.

It is usual to use the version of the formula that gives a positive value for the denominator. And so, we have:

The sum of the first n terms of a geometric series, S_n, where $r \neq 1$ is given by:

$$S_n = \frac{a(1-r^n)}{1-r}, |r| < 1 \text{ or } S_n = \frac{a(r^n-1)}{r-1}, |r| > 1$$

Example 1.1.18

Sum the following series to the number of terms indicated.

a $2 + 4 + 8 + 16 + \ldots$ 9 terms

b $5 - 15 + 45 - 135 + \ldots$ 7 terms

c $24 + 18 + 27 + {}^{27}/_{2} + {}^{81}/_{8}$ 12 terms

d $20 - 30 + 45 - 67.5 + \ldots$ 10 terms

a In this case $a = 2$, $r = 2$ and $n = 9$.

Because $r = 2$ it is convenient to use: $S_n = \dfrac{a(r^n - 1)}{r - 1}$

$$S_9 = \frac{2(2^9 - 1)}{2 - 1}$$
$$= 1022$$

Using this version of the formula gives positive values for the numerator and denominator. The other version is correct but gives negative numerator and denominator and hence the same answer.

b $a = 5$, $r = -3$ and $n = 7$.

$$S_n = \frac{a(1 - r^n)}{1 - r} \quad \text{or} \quad S_n = \frac{a(r^n - 1)}{r - 1}$$

$$S_7 = \frac{5(1 - (-3)^7)}{1 - (-3)} \qquad S_7 = \frac{5((-3)^7 - 1)}{(-3) - 1}$$

$$= 2735 \qquad = 2735$$

c $a = 24$, $r = 0.75$ and $n = 12$.

$$S_n = \frac{a(1 - r^n)}{1 - r}$$

$$S_{12} = \frac{24\left(1 - \left(\frac{3}{4}\right)^{12}\right)}{1 - \left(\frac{3}{4}\right)}$$

This version gives the positive values.

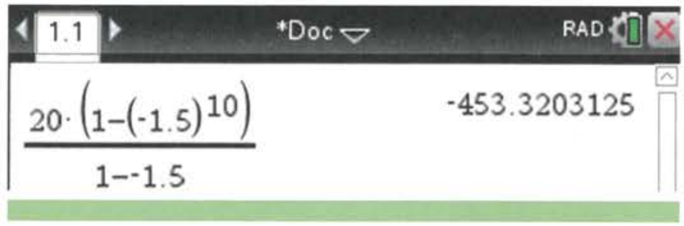
$$= 92.95907$$

d $a = 20$, $r = -1.5$ and $n = 10$.

$$S_n = \frac{a(1 - r^n)}{1 - r}$$

$$S_{10} = \frac{20(1 - (-1.5)^{10})}{1 - (-1.5)}$$
$$= -453.32031$$

When using a calculator to evaluate such expressions, it is advisable to use brackets to ensure that correct answers are obtained. For both the graphics and scientific calculator, the negative common ratio must be entered using the +/– or (–) key.



Other questions that may be asked in examinations could involve using both formulae. A second possibility is that you may be asked to apply sequence and series theory to some simple problems.

Example 1.1.19

The second term of a geometric series is –30 and the sum of the first two terms is –15. Find the first term and the common ratio.

From the given information we have:

$u_2 = -30 \therefore ar = -30$ – (1)

$S_2 = -15 \therefore \dfrac{a(r^2 - 1)}{r - 1} = -15$ – (2)

The result is a pair of simultaneous equations in the two unknowns. The best method of solution is substitution:

From (1): $a = \dfrac{-30}{r}$.

Substituting into (2): $\dfrac{\frac{-30}{r}(r^2 - 1)}{r - 1} = -15$

$$\frac{-30(r^2 - 1)}{r(r - 1)} = -15$$
$$\frac{-30(r + 1)(r - 1)}{r(r - 1)} = -15$$
$$-30(r + 1) = 15r$$
$$-30r - 30 = -15r$$
$$r = -2$$
$$\therefore a = \frac{-30}{r} = \frac{-30}{-2} = 15$$

The series is 15 – 30 + 60 – 120 + 240 – ... which meets the conditions set out in the question.

Example 1.1.20

A family decide to save some money in an account that pays 9% annual compound interest calculated at the end of each year. They put \$2500 into the account at the beginning of each year. All interest is added to the account and no withdrawals are made. How much money will they have in the account on the day after they have made their tenth payment?

The problem is best looked at from the last payment of \$2500 which has just been made and which has not earned any interest.

The previous payment has earned one lot of 9% interest and so is now worth 2500×1.09.

The previous payment has earned two years' worth of compound interest and is worth 2500×1.09^2.

This process can be continued for all the other payments and the various amounts of interest that each has earned. They form a geometric series:

Last payment First payment

$2500 + 2500 \times 1.09 + 2500 \times 1.09^2 + \ldots + \ldots + 2500 \times 1.09^9$.

The total amount saved can be calculated using the series formula:

$$S_n = \frac{a(r^n - 1)}{r - 1}$$

$$S_{10} = \frac{2500(1.09^{10} - 1)}{1.09 - 1}$$

$$= 37982.32$$

The family will save about \$37 982.

Exercise 1.1.5

1. Find the common ratios of these geometric sequences:

   a 7, 21, 63, 189, ... b 12, 4, $^4/_3$, $^4/_9$, ...

   c 1, −1, 1, −1, 1, ... d 9, −3, 1, $-^1/_3$, $^1/_9$, ...

   e 64, 80, 100, 125, ... f 27, −18, 12, −8, ...

2. Find the term indicated for each of these geometric sequences.

   a 11, 33, 99, 297, ... 10th term.

   b 1, 0.2, 0.04, 0.008, ... 5th term.

   c 9, −6, 4, $-^8/_3$, ... 9th term.

   d 21, 9, $^{27}/_7$, $^{81}/_{49}$, ... 6th term.

   e $-^1/_3$, $-^1/_4$, $-^3/_{16}$, $-^9/_{64}$, ... 6th term.

3. Find the number of terms in each of these geometric sequences and the sum of the numbers in each sequence:

   a 4, 12, 36, ... , 236196 b 11, −22, 44, ... , 704

   c 100, −10, 1, ... , -10^{-5} d 48, 36, 27, ... , $\frac{6561}{1024}$

   e $\frac{1}{8}, -\frac{9}{32}, \frac{81}{128}, \ldots \frac{6561}{2048}$ f 100, 10, 1, ... , 10^{-10}

4. Write the following in expanded form and evaluate.

   a $\sum_{k=1}^{7}\left(\frac{1}{2}\right)^k$ b $\sum_{i=1}^{6} 2^{i-4}$ c $\sum_{j=1}^{4}\left(\frac{2}{3}\right)^j$

   d $\sum_{s=1}^{4}(-3)^s$ e $\sum_{t=1}^{6} 2^{-t}$

5. The third term of a geometric sequence is 36 and the tenth term is 78 732. Find the first term in the sequence and the sum of these terms.

6. A bank account offers 9% interest compounded annually. If \$750 is invested in this account, find the amount in the account at the end of the twelfth year.

7. When a ball is dropped onto a flat floor, it bounces to 65% of the height from which it was dropped. If the ball is dropped from 80 cm, find the height of the fifth bounce.

8. A computer loses 30% of its value each year.

   a Write a formula for the value of the computer after n years.

   b How many years will it be before the value of the computer falls below 10% of its original value?

9. A geometric sequence has a first term of 7 and a common ratio of 1.1. How many terms must be taken before the value of the term exceeds 1000?

10. A colony of algae increases in size by 15% per week. If 10 grams of the algae are placed in a lake, find the weight of algae that will be present in the lake after 12 weeks. The lake will be considered 'seriously polluted'

when there is in excess of 10 000 grams of algae in the lake. How long will it be before the lake becomes seriously polluted?

11. A geometric series has nine terms, a common ratio of 2 and a sum of 3577. Find the first term.

12. A geometric series has a third term of 12, a common ratio of $-^1/_2$ and a sum of $32^1/_{16}$. Find the number of terms in the series.

13. A geometric series has a first term of 1000, seven terms and a sum of $671^7/_8$. Find the common ratio.

14. A geometric series has a third term of 300, and a sixth term of 37500. Find the common ratio and the sum of the first fourteen terms (in scientific form correct to two significant figures).

15. A \$10 000 loan is offered on the following terms: 12% annual interest on the outstanding debt calculated monthly. The required monthly repayment is \$270. How much will still be owing after nine months.

16. As a prize for inventing the game of chess, its originator is said to have asked for one grain of wheat to be placed on the first square of the board, 2 on the second, 4 on the third, 8 on the fourth and so on until each of the 64 squares had been covered. How much wheat would have been the prize?

Combined arithmetic and geometric sequences and series

There will be occasions on which questions will be asked that relate to both arithmetic and geometric sequences and series.

Example 1.1.21

A geometric sequence has the same first term as an arithmetic sequence. The third term of the geometric sequence is the same as the tenth term of the arithmetic sequence with both being 48. The tenth term of the arithmetic sequence is four times the second term of the geometric sequence. Find the common difference of the arithmetic sequence and the common ratio of the geometric sequence.

When solving these sorts of questions, write the data as equations, noting that a is the same for both sequences. Let u_n denote the general term of the arithmetic sequence and v_n the general term of the geometric sequence.

We then have:

$u_{10} = a + 9d$ $\qquad$ $v_3 = ar^2 = 48$

i.e. $a + 9d = ar^2 = 48$ – (1)

$u_{10} = 4v_2 \Rightarrow a + 9d = 4ar$ – (2)

(1) represents the information 'The third term of the geometric sequence is the same as the tenth term of the arithmetic sequence with both being 48'.

(2) represents 'The tenth term of the arithmetic sequence is four times the second term of the geometric sequence'.

There are three equations here and more than one way of solving them. One of the simplest is:

From (1) $a + 9d = 48$ and so substituting into (2):

$48 = 4ar \Leftrightarrow ar = 12$ – (3)

Also from (1) we have: $ar^2 = 48 \Leftrightarrow (ar)r = 48$ – (4)

Substituting (3) into (4): $12r = 48 \Leftrightarrow r = 4$

Substituting result into (1): $a \times 16 = 48$ $\; a = 3$

Substituting result into (1): $3 + 9d = 48 \Leftrightarrow d = 5$

The common ratio is 4 and the common difference is 5.

It is worth checking that the sequences are as specified:

Geometric sequence: 3, 12, 48

Arithmetic sequence: 3, 8, 13, 18, 23, 28, 33, 38, 43, 48

Exercise 1.1.6

1. Consider the following sequences:

Arithmetic: 100, 110, 120, 130, . . .

Geometric: 1, 2, 4, 8, 16, . . .

Prove that:

a The terms of the geometric sequence will exceed the terms of the arithmetic sequence after the 8th term.

b The sum of the terms of the geometric sequence will exceed the sum of the terms of the arithmetic after the 10th term.

2. An arithmetic series has a first term of 2 and a fifth term of 30. A geometric series has a common ratio of –0.5. The sum of the first two terms of the geometric series is the same as the second term of the arithmetic series. What is the first term of the geometric series?

3. An arithmetic series has a first term of –4 and a common difference of 1. A geometric series has a first term of 8 and a common ratio of 0.5. After how many terms does the sum of the arithmetic series exceed the sum of the geometric series?

4. The first and second terms of an arithmetic and a geometric series are the same and are equal to 12. The sum of the first two terms of the arithmetic series is four times the first term of the geometric series. Find the first term of each series, if the A.P. has $d = 4$.

5. Bo-Youn and Ken are to begin a savings program. Bo-Youn saves \$1 in the first week \$2 in the second week, \$4 in the third and so on, in geometric progression. Ken saves \$10 in the first week, \$15 in the second week, \$20 in the third and so on, in arithmetic progression. After how many weeks will Bo-Youn have saved more than Ken?

6. Ari and Chai begin a training program. In the first week Chai will run 10km, in the second he will run 11km and in the third 12km, and so on, in arithmetic progression. Ari will run 5km in the first week and will increase his distance by 20% in each succeeding week.

a When does Ari's weekly distance first exceed Chai's?

b When does Ari's total distance first exceed Chai's?

7. The Fibonacci sequence: 1, 1, 2, 3, 5, 8, 13, 21, . . . in which each term is the sum of the previous two terms is neither arithmetic nor geometric. However, after the eighth term (21) the sequence becomes approximately geometric. If we assume that the sequence is geometric:

a What is the common ratio of the sequence (to four significant figures)?

b Assuming that the Fibonacci sequence can be approximated by the geometric sequence after the eighth term, what is the approximate sum of the first 24 terms of the Fibonacci sequence?

Convergent series

If a geometric series has a common ratio between –1 and 1, the terms get smaller and smaller as n increases.

The sum of these terms is still given by the formula

$$S_n = \frac{a(1-r^n)}{1-r}, r \neq 1$$

For $-1 < r < 1$, $r^n \to 0$ as $n \to \infty$, $S_n \to \dfrac{a}{1-r}$

If $|r| < 1$, the infinite sequence has a sum given by: $S_\infty = \dfrac{a}{1-r}$

This means that if the common ratio of a geometric series is between –1 and 1, the sum of the series will approach a value of $\dfrac{a}{1-r}$ as the number of terms of the series becomes large,

i.e. the series is convergent.

Example 1.1.22

Find the sum to infinity of the series:

a $16 + 8 + 4 + 2 + 1 + \ldots$

b $9 - 6 + 4 - {}^8/_3 + {}^{16}/_9 + \ldots$

a $16 + 8 + 4 + 2 + 1 + \ldots$

In this case: a $a = 16, r = \dfrac{1}{2} \Rightarrow S_\infty = \dfrac{a}{1-r} = \dfrac{16}{1-\dfrac{1}{2}} = 32$

b $9 - 6 + 4 - {}^8/_3 + {}^{16}/_9$

$$a = 9, r = -\frac{2}{3} \Rightarrow S_\infty = \frac{a}{1-r} = \frac{9}{1-\left(-\dfrac{2}{3}\right)} = 5.4$$

There are many applications for convergent geometric series. The following examples illustrate two of these.

Example 1.1.23

Use an infinite series to express the recurring decimal $0.\dot{4}6\dot{2}$ as rational number.

$0.\dot{4}6\dot{2}$ can be expressed as the series:

$$0.462 + 0.000462 + 0.000000462 + \ldots$$

or $\frac{462}{1000}+\frac{462}{1000000}+\frac{462}{1000000000}+\ldots$

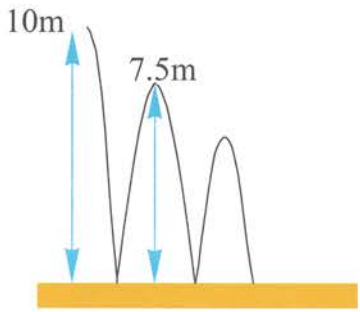
This is a geometric series with: $a=\frac{462}{1000}, r=\frac{1}{1000}$.

It follows that: $S_{\infty}=\frac{a}{1-r}=\frac{\frac{462}{1000}}{1-\frac{1}{1000}}=\frac{\frac{462}{1000}}{\frac{999}{1000}}=\frac{462}{999}$

Example 1.1.24

A ball is dropped from a height of 10 metres. On each bounce the ball bounces to three quarters of the height of the previous bounce. Find the distance travelled by the ball before it comes to rest (if it does not move sideways).

The ball bounces in a vertical line and does not move sideways. On each bounce after the drop, the ball moves both up and down and so travels twice the distance of the height of the bounce.



Distance = $10+15+15\times\frac{3}{4}+15\times\left(\frac{3}{4}\right)^{2}+\ldots$

All terms, except the first, are geometric.

Distance $10+S_{\infty}=10+\frac{15}{1-\frac{3}{4}}=70$ m

Exercise 1.1.7

1. Evaluate:

   a $27+9+3+\frac{1}{3}+\ldots$

   b $1-\frac{3}{10}+\frac{9}{100}-\frac{27}{1000}+\ldots$

   c $500 + 450 + 405 + 364.5 + \ldots$

   d $3 - 0.3 + 0.03 - 0.003 + 0.0003 - \ldots$

2. Use geometric series to express the recurring decimal 23.232323… as a mixed number.

3. Biologists estimate that there are 1000 trout in a lake. If none are caught, the population will increase at 10% per year. If more than 10% are caught, the population will fall. As an approximation, assume that if 25% of the fish are caught per year, the population will fall by 15% per year. Estimate the total catch before the lake is 'fished out'. If the catch rate is reduced to 15%, what is the total catch in this case? Comment on these results.

4. Find the sum to infinity of the sequence 45, –30, 20, ...

5. The second term of a geometric sequence is 12 while the sum to infinity is 64. Find the first three terms of this sequence.

6. Express the following as rational numbers:

   a $0.3\dot{6}$ b $0.\dot{3}\dot{7}$ c $2.1\dot{2}$

7. A swinging pendulum covers 32 centimetres in its first swing, 24 cm on its second swing, 18 cm on its third swing and so on. What is the total distance this pendulum swings before coming to rest?

8. The sum to infinity of a geometric sequence is $^{27}/_{2}$ while the sum of the first three terms is 13. Find the sum of the first 5 terms.

9. Find the sum to infinity of the sequence: $1+\sqrt{3}, 1, \frac{1}{\sqrt{3}+1}, \ldots$.

10. a Find: i $\sum_{i=0}^{n}(-t)^{i}, |t|<1$, ii $\sum_{i=0}^{\infty}(-t)^{i}, |t|<1$.

    b i Hence, show that ,

    $$\ln(1+x)=x-\frac{1}{2}x^{2}+\frac{1}{3}x^{3}-\frac{1}{4}x^{4}+\ldots, |x|<1$$

    ii Using the above result, show that:

    $$\ln 2=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots$$

11. a Find: i $\sum_{i=0}^{n}\left(-t^{2}\right)^{i}, |t|<1$, ii $\sum_{i=0}^{\infty}\left(-t^{2}\right)^{i}, |t|<1$.

    i Hence, show that:

    $$\arctan x=x-\frac{1}{3}x^{3}+\frac{1}{5}x^{5}-\frac{1}{7}x^{7}+\ldots, |x|<1$$

Using the above result, show that $\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\ldots$.

Exercise 1.1.8

1. $2k + 2$, $5k + 1$ and $10k + 2$ are three successive terms of a geometric sequence. Find the value(s) of k.

2. Evaluate $\dfrac{1+2+3+...+10}{1+\frac{1}{2}+\frac{1}{4}+...+\frac{1}{512}}$.

3. Find a number which, when added to each of 2, 6 and 13, gives three numbers in geometric sequence.

4. Find the fractional equivalent of:

   a $2.3\dot{8}$ b $4.\dot{6}\dot{2}$ c 0.41717...

5. Find the sum of all integers between 200 and 400 that are divisible by 6.

6. Find the sum of the first 50 terms of an arithmetic progression given that the 15th term is 34 and the sum of the first 8 terms is 20.

7. Find the value of p so that $p + 5$, $4p + 3$ and $8p - 2$ will form successive terms of an arithmetic progression.

8. For the series defined by $S_n = 3n^2 - 11n$, find t_n and hence show that the sequence is arithmetic.

9. How many terms of the series $6 + 3 + {}^3/_2 + ...$ must be taken to give a sum of $11^{13}/_{16}$?

10. If $1 + 2x + 4x^2 + \ldots = {}^3/_4$, find the value of x.

11. Logs of wood are stacked in a pile so that there are 15 logs on the top row, 16 on the next row, 17 on the next and so on. If there are 246 logs in total,

    a how many rows are there?

    b how many logs are there in the bottom row?

12. The lengths of the sides of a right-angled triangle form the terms of an arithmetic sequence. If the hypotenuse is 15 cm in length, what is the length of the other two sides?

13. The sum of the first 8 terms of a geometric series is 17 times the sum of its first four terms. Find the common ratio.

14. Three numbers a, b and c whose sum is 15 are successive terms of a G.P, and b, a, c are successive terms of an A.P. Find a, b and c.

15. The sum of the first n terms of an arithmetic series is given by:

$$S_n = \frac{n(3n+1)}{2}.$$

    a Calculate S_1 & S_2.

    b Find the first three terms of this series.

    c Find an expression for the nth term.

16. An ant walks along a straight path. After travelling 1 metre it stops, turns through an angle of 90° in an anticlockwise direction and sets off in a straight line covering a distance of half a metre. Again, the ant turns through an angle of 90° in an anticlockwise direction and sets off in a straight line covering a quarter of a metre. The ant continues in this manner indefinitely.

    a How many turns will the ant have made after covering a distance of ${}^{63}/_{32}$ metres?

    b How far will the ant eventually travel?

APPLICATIONS

Sequences and series are used in a range of areas.

Founded as a refuge from lawlessness, Venice was a great trade and financial centre in the 13th Century.

In the Middle Ages, the charging of interest was forbidden to Christians and some other religions.

Banking was the speciality of Jews. Many modern banking dynasties, such as the Rothschilds, are Jewish.

Some cultures do not allow the charging of interest. How does money lending work in the Arab world?

Financial District, Dubai, UAE.

This diver is taking extra nitrogen into his blood because of the elevated pressure of the air he is breathing - one extra atmosphere for every 10 metres of depth. This nitrogen is released when he surfaces. If he surfaces too rapidly, this may cause bubbles to form in his blood. This results in the painful and sometimes fatal condition known as 'the bends'.

To prevent this, he carries a 'dive computer' that records his depth as a time sequence and uses a mathematical model (a time series, as nitrogen is progressively added) to predict the nitrogen uptake and advise as to the safe ascent procedure.

Compound interest

We have already come across some practical examples of the use of G.P.s in the area of finance. In this section we further develop these ideas and look at the area of compound interest and superannuation.

Example 1.1.25

Find what \$600 amounts to in 20 years if it is invested at 8% p.a. compounding annually.

End of year 1 value $= \$600 + 8\% \times \$600$
$= \$600(1.08)$

End of year 2 value $= \$600(1.08) + 8\% \times \$600(1.08)$
$= \$600(1.08) + 0.08 \times \$600(1.08)$
$= \$600(1.08)[1 + 0.08]$
$= \$600(1.08)^2$

End of year 3 value $= \$600(1.08)^2 + 8\% \times \$600(1.08)^2$
$= \$600(1.08)^2 + 0.08 \times \$600(1.08)^2$
$= \$600(1.08)^2[1 + 0.08]$
$= \$600(1.08)^3$
⇩

End of year 20 value $= \$\,600(1.08)^{20}$

Thus, after 20 years the \$600 amounts to \$2796.57.

Looking closely at the terms of the sequence, they form a G.P.:

$$600(1.08), 600(1.08)^2, 600(1.08)^3, \ldots, 600(1.08)^{20}$$

where $a = 600$ and $r = 1.08$.

Developing a formula for compound interest

In general, if $\$P$ is invested at $r\%$ p.a. compound interest, it grows according to the sequence:

$$P\left(1+\frac{r}{100}\right), P\left(1+\frac{r}{100}\right)^2, P\left(1+\frac{r}{100}\right)^3, \ldots, P\left(1+\frac{r}{100}\right)^n$$

where $a = P\left(1+\frac{r}{100}\right)$ and $r = \left(1+\frac{r}{100}\right)$ so that

$$A_n = P\left(1+\frac{r}{100}\right)^n$$

where A_n is the amount after n time periods.

Superannuation

Superannuation is a common way in which working people attempt to provide for themselves in retirement. In many cases, workers save a fixed amount from each pay-packet into an interest bearing account.

Example 1.1.26

A woman invests \$1000 at the beginning of each year in a superannuation scheme. If the interest is paid at the rate of 12% p.a. on the investment (compounding annually), how much will her investment be worth after 20 years?

t_1 = the 1st \$1000 will be invested for 20 years at 12% p.a.

t_2 = the 2nd \$1000 will be invested for 19 years at 12% p.a.

t_3 = the 3rd \$1000 will be invested for 18 years at 12% p.a.

$\Downarrow$

t_{20} = the 20th \$1000 will be invested for 1 year at 12% p.a.

Finding the amount compounded annually using :

$A_n = P\left(1+\frac{r}{100}\right)^n$, we have:

$$t_1 = 1000\left(1+\frac{12}{100}\right)^{20} = 1000(1.12)^{20}$$
$$t_2 = 1000\left(1+\frac{12}{100}\right)^{19} = 1000(1.12)^{19}$$
$$t_3 = 1000\left(1+\frac{12}{100}\right)^{18} = 1000(1.12)^{18}$$
$$\Downarrow$$
$$t_{20} = 1000\left(1+\frac{12}{100}\right)^{1} = 1000(1.12)^{1}$$

To find the total of her investment after 20 years, we need to add the separate amounts:

$$1000(1.12)^{20} + 1000(1.12)^{19} + 1000(1.12)^{18} + \ldots 1000(1.12)^{1}$$

$$= \$80\,698.74$$

Thus her total investment amounts to \$80 698.74

Example 1.1.27

Linda borrows \$2000 at 1% per month reducible interest. If she repays the loan in equal monthly instalments over 4 years, how much is each instalment?

Amount borrowed = \$2000, r = 1% per month = 0.01 and $n = 4 \times 12 = 48$ months.

Let the monthly instalment be = \$ M and the amount owing after n months = \$$A_n$.

Our aim is to find \$$M$ i.e. the amount of each instalment.

After 1 month (after paying the 1st instalment), we have:

$A_1 = 2000 + \text{interest} - M = 2000 + 2000 \times 0.01 - M$

After 2 months,

$$\begin{aligned}A_2 &= A_1 \times 1.01 - M\\ &= \left[2000(1.01) - M\right] \times 1.01 - M\\ &= 2000 \times 1.01^2 - 1.01 \times M - M\\ &= 2000 \times 1.01^2 - M(1.01+1)\end{aligned}$$

After 3 months,

$$\begin{aligned}A_3 &= A_2 \times 1.01 - M\\ &= \left[2000(1.01)^2 - M(1.01+1)\right] \times 1.01 - M\\ &= 2000 \times 1.01^3 - M(1.01+1) \times 1.01 - M\\ &= 2000 \times 1.01^3 - M\left[1.01^2 + 1.01 + 1\right]\end{aligned}$$

After 4 months,

$$\begin{aligned}A_4 &= A_3 \times 1.01 - M\\ &= \left[2000(1.01)^3 - M\left(1.01^2+1.01+1\right)\right] \times 1.01 - M\\ &= 2000 \times 1.01^4 - M\left(1.01^3+1.01^2+1.01\right) \times 1.01 - M\\ &= 2000 \times 1.01^4 - M\left[1.01^3+1.01^2+1.01+1\right]\end{aligned}$$

After n months, we then have

$$A_n = 2000(1.01)^n - M[1+1.01+1.01^2+1.01^3 \ldots 1.01^{n-1}]$$

thus, $A_{48} = 2000(1.01)^{48} - M[1+1.01+1.01^2+1.01^3 \ldots 1.01^{47}]$

Now, the loan is repaid after 48 months, meaning that $A_{48} = 0$, therefore, solving for M, we have

$$2000(1.01)^{48} - M[1+1.01+1.01^2+1.01^3 \ldots 1.01^{47} = 0$$

$$2000(1.01)^{48} = M[1+1.01+1.01^2+1.01^3 \ldots 1.01^{47}$$

$$M = \frac{2000 \times 1.01^{48}}{1+1.01+1.01^2+1.01^3+\ldots+1.01^{47}}$$

The denominator is a G.P. with $a = 1$, $r = 1.01$ and $n = 48$, so that

$$\begin{aligned}1+1.01+1.01^2+1.01^3+\ldots+1.01^{47} &= S_{48}\\ &= \frac{1\left(1-1.01^{48}\right)}{1-1.01}\\ &= 61.22261\end{aligned}$$

Therefore, $M = \dfrac{2000(1.01)^{48}}{61.22261} = 52.67$

That is, each instalment must be \$52.67.

The total paid = $52.67 \times 48 = 2528.16$ so that the interest paid $= 2528.16 - 2000 = 528.16$

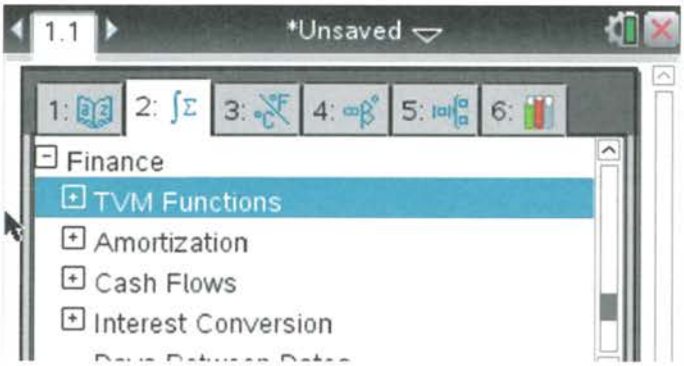
That is, she ends up paying \$528.16 in interest.

Although it is important to understand the process used in the examples shown, it is also important to be able to make use of technology. Calculators can help ease the pain of long calculations.

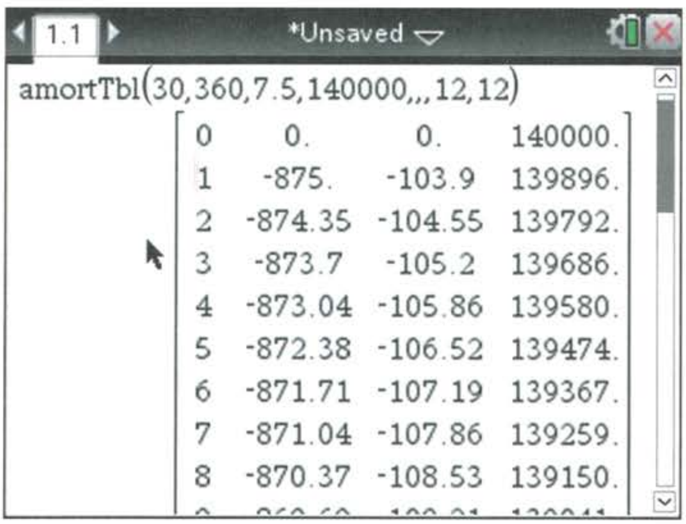
If you purchase a flat for \$200 000, pay 30% deposit, and mortgage the balance at 7.5% interest. You amortize your debt with monthly repayments for 30 years.

What is the repayment?

Using the amortisation table:



and the data given:



In the first month, the payment will be \$875 interest + \$103.90 principal making a total of (to the nearest \$), \$979.

If using Casio, select the Financial Module (C), F4 -Amortization.

Exercise 1.1.9

1. To how much will \$1000 grow to if it is invested at 12% p.a. for 9 years, compounding annually?

2. A bank advertises an annual interest rate of 13.5% p.a. but adds interest to the account monthly, giving a monthly interest rate of 1.125%. Scott deposits \$3500 with the bank. How much will be in the account in 20 months time?

3. To what amount will \$900 grow to if it is invested at 10% p.a. for 7 years, compounding every 6 months?

4. A man borrows \$5000 at 18% p.a. over a period of 5 years, with the interest compounding every month. Find to the nearest dollar the amount owing after 5 years.

5. Find the total amount required to pay off a loan of \$20,000 plus interest at the end of 5 years if the interest is compounded half yearly and the rate is 12%.

6. A man invests \$500 at the beginning of each year in a superannuation fund. If the interest is paid at the rate of 12% p.a. on the investment (compounding annually), how much will his investment be worth after 20 years?

7. A woman invests \$2000 at the beginning of each year into a superannuation fund for a period of 15 years at a rate of 9% p.a. (compounding annually). Find how much her investment is worth at the end of the 15 years.

8. A man deposits \$3000 annually to accumulate at 9% p.a. compound interest. How much will he have to his credit at the end of 25 years? Compare this to depositing \$750 every three months for the same length of time and at the same rate. Which of these two options gives the better return?

9. A woman invests \$200 at the beginning of each month into a superannuation scheme for a period of 15 years. Interest is paid at the rate of 7% p.a. and is compounded monthly. How much will her investment be worth at the end of the 15-year period?

10. Peter borrows \$5000 at 1.5% per month reducible interest. If he repays the loan in equal monthly instalments over 8 years, how much is each instalment, and what is the total interest charged on the loan? Compare this to taking the same loan, but at a rate of 15% p.a. flat rate.

11. Kevin borrows \$7500 to be paid back at 12.5% p.a. monthly reducible over a period of 7 years. What is the amount of each monthly instalment and what is the total interest charged on the loan. Find the equivalent flat rate of interest.

12. Find the possible values of x if $x+1$, $3x+2$, and $2x^2$ are three consecutive terms of an arithmetic sequence.

13. Find k, given that $\sum_{i=1}^{k}(4i-29)=45$.

14. Show that $\sum_{k=3}^{8}(2k-2)=\sum_{k=1}^{6}(3(k+2)-1)$.

15. Given four consecutive terms in a progression, 4, m, n, 49. Find the possible values of m and n, if the first three terms form an arithmetic sequence and the last three terms form a geometric sequence.

Answers to Exercises

Theory of Knowledge

Why did humans develop mathematics?

There is no right answer to this question.

There is little doubt that the need to service trade (count goods etc.) was one powerful motivation.

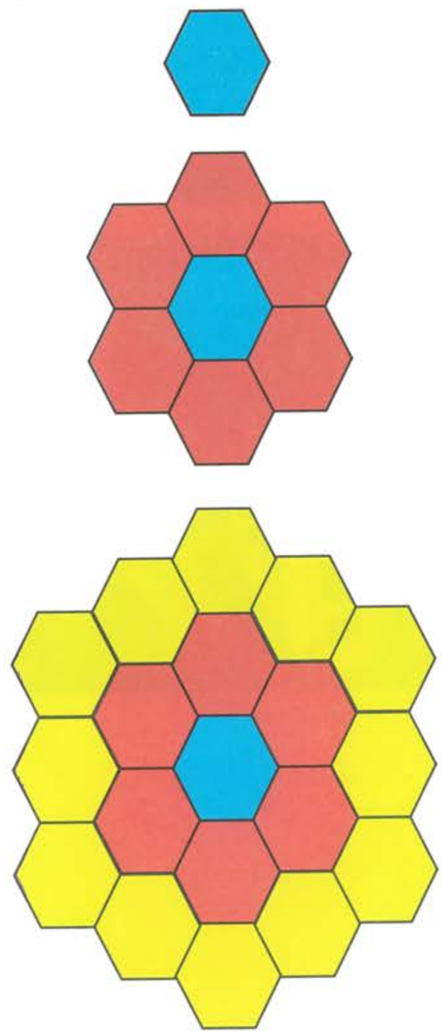
However, it is probable that the patterns evident in nature played their part in stimulating the thought processes.

You have probably noticed the patterns evident in plants:

But were you aware similar patterns are present in rocks?

These hexagonal basalt columns on a beach in Iceland are natural. They result from crystallisation from molten rock.

If we assume that the crystallisation of the rocks starts in one place (is this reasonable?), the pattern of hexagons will develop as follows:



What sequence is implied here?

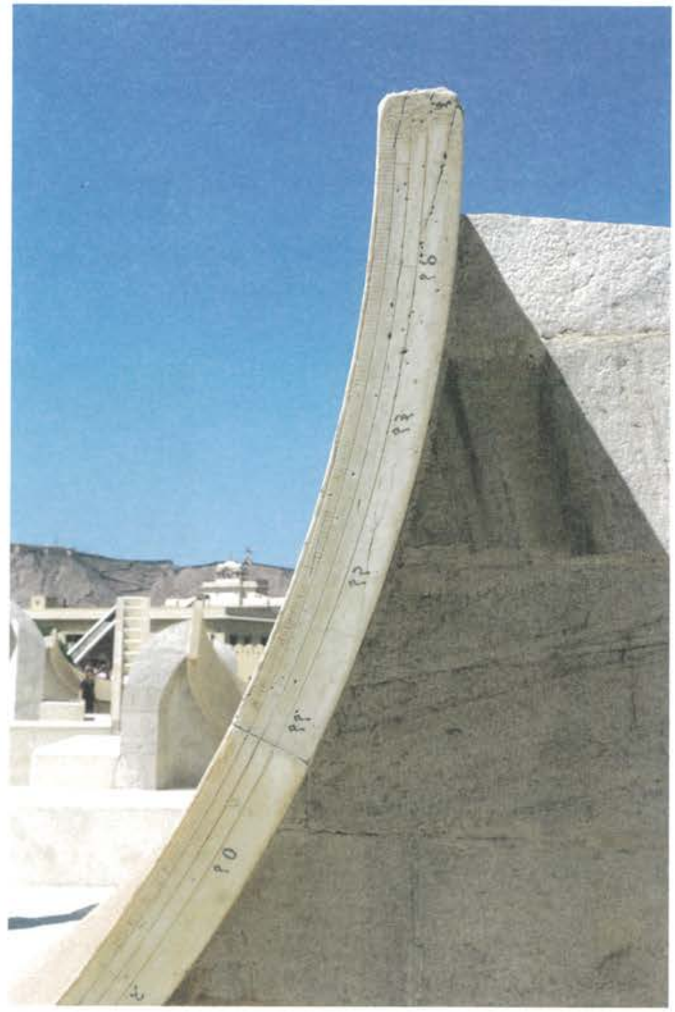
The study of these patterns would have gone nowhere without the ability to record numbers.

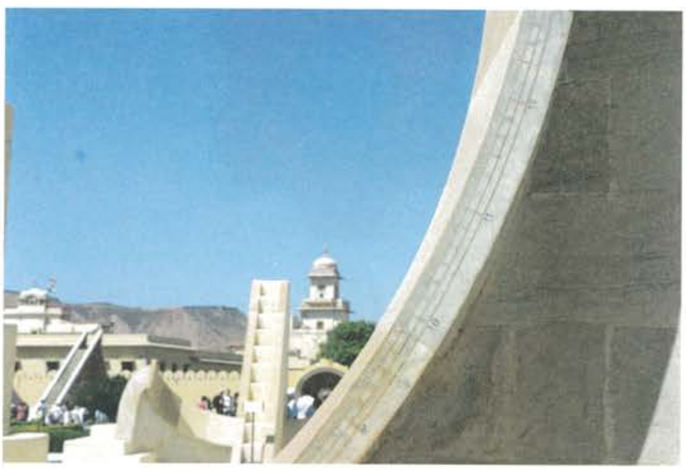
Many counting systems were invented, but the system that has survived into the modern world is Hindu numeration.

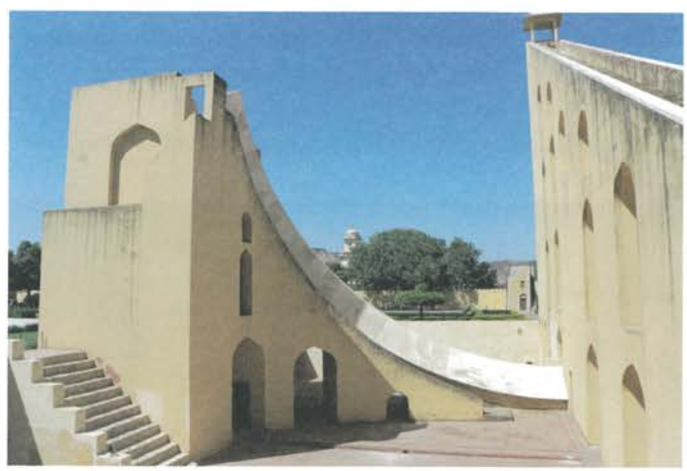
The original system uses symbols that are similar, but not identical to modern symbols.

These pictures show the Jaipur observatory in Rajahstan India.

Note the numeric notation.







1.2 Exponents and Logarithms

Exponents

Basic rules of indices

We start by looking at the notation involved when dealing with indices (or exponents).

The expression:

$$a \times a \times a \times a \times a \times \ldots \times a$$

$$\longleftarrow \quad n \text{ times} \quad \longrightarrow$$

can be written in index form, a^n, where n is the index (or power or exponent) and a is the base.

This expression is read as "*a to the power of n*". or more briefly as "*a to the n*". For example, we have that $3^5 = 3 \times 3 \times 3 \times 3 \times 3$ so that 3 is the base and 5 the exponent (or index).

The laws for positive integral indices are summarized below. If a and b are real numbers and m and n are positive integers, we have:

	Law	Rule	Example
1	Multiplication [same base]	$a^m \times a^n = a^{m+n}$	$3^4 \times 3^6 = 3^{4+6}$ $= 3^{10}$
2	Division [same base]	$a^m \div a^n = \dfrac{a^m}{a^n} = a^{m-n}$	$7^9 \div 7^5 = 7^{9-5}$ $= 7^4$
3	Power of a power [same base]	$(a^m)^n = a^{m \times n}$	$(2^3)^5 = 2^{3\times5}$ $= 2^{15}$
4	Power of a product	$a^m \times b^m = (ab)^m$	$3^4 \times 7^4 = (3\times7)^4$ $= 21^4$
5	Power of a quotient	$a^m \div b^m = \left(\dfrac{a}{b}\right)^m$	$5^3 \div 7^3 = \left(\dfrac{5}{7}\right)^3$
6	Negative one to a power	$(-1)^n = \begin{cases} -1, n \text{ odd} \\ 1, n \text{ even} \end{cases}$	$(-1)^3 = -1$ $(-1)^4 = 1$

There are more laws of indices that are based on rational indices, negative indices and the zero index.

	Law	Rule	Example
1	Fractional index Type 1 [nth root]	$a^{\frac{1}{n}} = \sqrt[n]{a}, n \in \mathbb{N}$ Note: if n is even, then $a \geq 0$. If n is odd, then $a \in \mathbb{R}$	$8^{\frac{1}{3}} = \sqrt[3]{8}$ $= 2$ $(-27)^{\frac{1}{3}} = \sqrt[3]{-27}$ $= -3$
2	Fractional index Type 2	$a^{\frac{m}{n}} = \sqrt[n]{a^m}, n \in \mathbb{N}$	$16^{\frac{3}{4}} = \sqrt[4]{16^3}$ $= 8$
3	Negative index	$a^{-n} = \dfrac{1}{a^n}, a \neq 0, n \in \mathbb{N}$	$2^{-1} = \dfrac{1}{2}$ $3^{-2} = \dfrac{1}{3^2} = \dfrac{1}{9}$
4	Zero index	$a^0 = 1, a \neq 0$ Note $0^n = 0, a \neq 0$	$12^0 = 1$

We make the following note about fractional indices:

As $\frac{m}{n} = m \times \frac{1}{n} = \frac{1}{n} \times m$, we have that for $b \geq 0$

i $\quad b^{\frac{m}{n}} = b^{m \times \frac{1}{n}} = (b^m)^{\frac{1}{n}} = \sqrt[n]{b^m}$

ii $\quad b^{\frac{m}{n}} = b^{\frac{1}{n} \times m} = \left(b^{\frac{1}{n}}\right)^m = (\sqrt[n]{b})^m$

Then,

If $b \geq 0$, then $b^{\frac{m}{n}} = \sqrt[n]{b^m} = (\sqrt[n]{b})^m, m \in \mathbb{Z}, n \in \mathbb{N}$

If $b < 0$, then $b^{\frac{m}{n}} = \sqrt[n]{b^m} = (\sqrt[n]{b})^m, m \in \mathbb{Z}, n \in \{1, 3, 5, \ldots\}$

Example 1.2.1

Simplify the following.

a $\quad \left(\frac{4x^2}{5y^4}\right)^2 \times (2x^3y)^3$ $\qquad$ b $\quad \frac{3^{n+1} + 3^2}{3}$

a
$$\left(\frac{4x^2}{5y^4}\right)^2 \times (2x^3y)^3 = \frac{4^2x^{2\times 2}}{5^2y^{4\times 2}} \times 2^3x^{3\times 3}y^{1\times 3}$$
$$= \frac{16x^4}{25y^8} \times 8x^9y^3$$
$$= \frac{128}{25}x^{4+9}y^{3-8}$$
$$= \frac{128}{25}x^{13}y^{-5}$$
$$= \frac{128x^{13}}{25y^5}$$

b
$$\frac{3^{n+1} + 3^2}{3} = \frac{3(3^n + 3)}{3} = 3^n + 3$$

Example 1.2.2

Simplify the following:

a $\quad \frac{4x^2(-y^{-1})^{-2}}{(-2x^2)^3(y^{-2})^2}$ $\qquad$ b $\quad \frac{x^{-1} + y^{-1}}{x^{-1}y^{-1}}$

a
$$\frac{4x^2(-y^{-1})^{-2}}{(-2x^2)^3(y^{-2})^2} = \frac{4x^2 \times y^{-1\times -2}}{-8x^{2\times 3} \times y^{-2\times 2}} = -\frac{x^2y^2}{2x^6y^{-4}}$$
$$= -\frac{y^{2-(-4)}}{2x^{6-(2)}} = -\frac{y^6}{2x^4}$$

b
$$\frac{x^{-1} + y^{-1}}{x^{-1}y^{-1}} = \frac{\frac{1}{x} + \frac{1}{y}}{\frac{1}{xy}} = \left(\frac{1}{x} + \frac{1}{y}\right) \times \frac{xy}{1} = \frac{xy}{x} + \frac{xy}{y} = y + x$$

Example 1.2.3

Simplify the following.

a $\quad \frac{2^{n-3} \times 8^{n+1}}{2^{2n-1} \times 4^{2-n}}$ $\qquad$ b $\quad \frac{(a^{1/3} \times b^{1/2})^{-6}}{\sqrt[4]{a^8b^9}}$

a
$$\frac{2^{n-3} \times 8^{n+1}}{2^{2n-1} \times 4^{2-n}} = \frac{2^{n-3} \times (2^3)^{n+1}}{2^{2n-1} \times (2^2)^{2-n}} = \frac{2^{n-3} \times 2^{3n+3}}{2^{2n-1} \times 2^{4-2n}}$$
$$= \frac{2^{n-3+(3n+3)}}{2^{2n-1+(4-2n)}} = \frac{2^{4n}}{2^3} = 2^{4n-3}$$

b
$$\frac{(a^{1/3} \times b^{1/2})^{-6}}{\sqrt[4]{a^8b^9}} = \frac{a^{\frac{1}{3}\times -6} \times b^{\frac{1}{2}\times -6}}{(a^8b^9)^{\frac{1}{4}}} = \frac{a^{-2} \times b^{-3}}{a^2b^{\frac{9}{4}}}$$
$$= a^{-2-2} \times b^{-3-\frac{9}{4}}$$
$$= a^{-4}b^{-\frac{21}{4}}$$
$$= \frac{1}{a^4b^{\frac{21}{4}}}$$

Exercise 1.2.1

1. Simplify the following.

a $\quad \left(\frac{3y^2}{4x^3}\right)^3 \times (2x^2y^3)^3$ $\qquad$ **b** $\quad \left(\frac{2}{3a^2}\right)^3 + \frac{1}{8a^6}$

c $\quad \frac{2^{n+1} + 2^2}{2}$ $\qquad$ **d** $\quad \left(\frac{2x^3}{3y^2}\right)^3 \times (xy^2)^2$

2. Simplify the following.

a $\quad \frac{20^6}{10^6}$ $\qquad$ b $\quad \frac{12^{2x}}{(6^3)^x}$

c $\quad \frac{16^{2y+1}}{8^{2y+1}}$ $\qquad$ d $\quad \frac{(ab)^{2x}}{a^{2x}b^{4x}}$

3. Simplify the following.

a $\quad \left(\frac{x}{y}\right)^3 \times \left(\frac{y}{z}\right)^2 \times \left(\frac{z}{x}\right)^4$ $\qquad$ b $\quad 3^{2n} \times 27 \times 243^{n-1}$

c $\quad \frac{25^{2n} \times 5^{1-n}}{(5^2)^n}$ $\qquad$ d $\quad \frac{9^n \times 3^{n+2}}{27^n}$

4. Simplify $\frac{(x^m)^n(y^2)^m}{(x^m)^{(n+1)}y^2}$.

5. Simplify the following, leaving your answer in positive power form.

   a $\frac{(-3^4)\times 3^{-2}}{(-3)^{-2}}$ b $\frac{9y^2(-x^{-1})^{-2}}{(-2y^2)^3(x^{-2})^3}$

   c $\frac{x^{-1}-y^{-1}}{x^{-1}y^{-1}}$ d $\frac{x^{-2}+2x^{-1}}{x^{-1}+x^{-2}}$

6. Simplify the following.

   a $\frac{(x^{-1})^2+(y^2)^{-1}}{x^2+y^2}$ b $\frac{(x^2)^{-2}+2y}{1+2yx^4}$

   c $\frac{(x+h)^{-1}-x^{-1}}{h}$

   d $(x^2-1)^{-1}\times(x+1)$

Extra questions

Indicial equations

Why do indices turn up so often in applications?

This is a complex question, but it does seem that the Universe has a preference for power laws!

Biologists use power expressions to model animal populations:

Chemists use power laws to track the rates of reactions:

and so on. Is there anything behind this or is nature's liking for indices just chance?

Solving equations of the form $x^{\frac{1}{2}} = 3$, where the **variable is the base**, requires that we square both sides of the equation so that:

$$\left(x^{\frac{1}{2}}\right)^2 = 3^2 \Rightarrow x = 9.$$

However, when the **variable is the power** and not the base we need to take a different approach.

Indicial (exponential) equations take on the general form $b^x = a$, where the unknown (variable), x, is the power.

Consider the case where we wish to solve for x given that $2^x = 8$. In this case we need to think of a value of x so that when 2 is raised to the power of x the answer is 8. Using trial and error, it is not too difficult to arrive at $x = 3$ $(2^3 = 2\times2\times2 = 8)$.

Next consider the equation $3^{x+1} = 27$. Again, we need to find a number such that when 3 is raised to that number, the answer is 27. Here we have that $27 = 3^3$. Therefore we can rewrite the equation as $3^{x+1} = 3^3$.

As the base on both sides of the equality is the same we can then equate the powers, that is,

$$3^{x+1} = 27 \Leftrightarrow 3^{x+1} = 3^3$$
$$\Leftrightarrow x+1 = 3$$
$$\Leftrightarrow x = 2$$

Example 1.2.4

Solve the following:

a $3^x = 81$ b $2 \times 5^u = 250$ c $2^x = \dfrac{1}{32}$

a $3^x = 81 \Leftrightarrow 3^x = 3^4$

$\Leftrightarrow x = 4$

b $2 \times 5^u = 250 \Leftrightarrow 5^u = 125$

$\Leftrightarrow 5^u = 5^3$

$\Leftrightarrow u = 3$

c $2^x = \dfrac{1}{32} \Leftrightarrow 2^x = \dfrac{1}{2^5}$

$\Leftrightarrow 2^x = 2^{-5}$

$\Leftrightarrow x = -5$

Example 1.2.5

Find:

a $\left\{x \mid \left(\frac{1}{2}\right)^x = 16\right\}$ b $\{x \mid 3^{x+1} = 3\sqrt{3}\}$

a $\left(\dfrac{1}{2}\right)^x = 16 \Leftrightarrow (2^{-1})^x = 16$

$\Leftrightarrow 2^{-x} = 2^4$

$\Leftrightarrow -x = 4$

$\Leftrightarrow x = -4$

i.e. solution set is $\{-4\}$.

b $3^{x+1} = 3\sqrt{3} \Leftrightarrow 3^{x+1} = 3 \times 3^{1/2}$

$\Leftrightarrow 3^{x+1} = 3^{3/2}$

$\Leftrightarrow x + 1 = \dfrac{3}{2}$

$\Leftrightarrow x = \dfrac{1}{2}$

i.e. solution set is $\{0.5\}$

Exercise 1.2.2

1. Solve the following equations.

a	$\{x \mid 4^x = 16\}$	b	$\left\{x \mid 7^x = \frac{1}{49}\right\}$
c	$\{x \mid 8^x = 4\}$	d	$\{x \mid 3^x = 243\}$
e	$\{x \mid 3^{x-2} = 81\}$	f	$\left\{x \mid 4^x = \frac{1}{32}\right\}$

2. Solve the following equations.

a	$\{x \mid 7^{x+6} = 1\}$	b	$\left\{x \mid 8^x = \frac{1}{4}\right\}$
c	$\{x \mid 10^x = 0.001\}$	d	$\{x \mid 9^x = 27\}$
e	$\{x \mid 2^{4x-1} = 1\}$	f	$\{x \mid 25^x = \sqrt{5}\}$
g	$\left\{x \mid 16^x = \frac{1}{\sqrt{2}}\right\}$	h	$\{x \mid 4^{-x} = 32\sqrt{2}\}$
i	$\{x \mid 9^{-2x} = 243\}$		

3. Show that: $\frac{1}{16}\left(8^{3x+2}\right) = 2^{9x+2}$

Extra questions

Drag is the force resisting the forward motion of vehicles such as aircraft. It is a 'square law' - the drag force is proportional to the square of the velocity.

LOGARITHMS

What are logarithms?

Consider the following sequence of numbers:

Sequence N	1	2	3	4	5	6	7	8	9	10	...
Sequence y	2	4	8	16	32	64	128	256	512	1024	...

The relationship between the values of N and y is given by $y = 2^N$.

We can the above table, evaluate the product 16×64.

Sequence N	1	2	3	4	5	6	7	8	9	10	...
Sequence y	2	4	8	16	32	64	128	256	512	1024	...

We start by setting up a table of values that correspond to the numbers in question:

N	4	6	Sum(4 + 6) = 10
y	16	64	Product = 1024

From the first table of sequences, we notice that the sum of the 'N sequence' (i.e. 10), corresponds to the value of the 'y sequence' (i.e. 1024).

We next consider the product 16×64, again. Setting up a table of values for the numbers in the sequences that are under investigation we have:

N	3	5	Sum(3 + 5) = 8
y	8	32	Product = 256

What about 4×64? As before, we set up the required table of values:

N	2	6	Sum(2 + 6) = 8
y	4	64	Product = 256

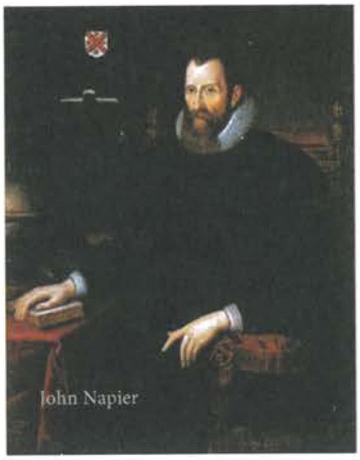
In each case the **product** of two terms of the sequence y corresponds to the **sum** obtained by adding corresponding terms of the sequence N.

Notice that **dividing** two numbers from the sequence y corresponds to the result when **subtracting** the two corresponding numbers from the sequence N,

e.g. for the sequence y: $512 \div 32 = 16$.

for the sequence N: $9 - 5 = 4$

This remarkable property was observed as early as 1594 by **John Napier**. John Napier was born in 1550 (when his father was all of sixteen years of age!) He lived most of his life at the family estate of Merchiston Castle, near Edinburgh, Scotland. Although his life was not without controversy, in matters both religious and political, Napier (when relaxing from his political and religious polemics) would indulge in the study of mathematics and science. His amusement with the study of mathematics led him to the invention of logarithms. In 1614 Napier published his discussion of logarithms in a brochure entitled *Mirifici logarithmorum canonis descriptio* ('A description of the Wonderful Law of Logarithms'). Napier died in1617.


John Napier

It is only fair to mention that the Swiss instrument maker **Jobst Bürgi** (1552–1632) conceived and constructed a table of logarithms independently of Napier, publishing his results in 1620, six years after Napier had announced his discovery.

One of the anomalies in the history of mathematics is the fact that logarithms were discovered before exponents were in use.

In this age of technology, the use of electronic calculators and computers has reduced the evaluation of products and quotients to tasks that involve the simple push of a few buttons. However, logarithms are an efficient means of converting a product to a sum and a quotient to a difference. So, what are logarithms?

Nowadays, a logarithm is universally regarded as an exponent.

Thus, if $y = b^N$ we say that N is the logarithm of y to the base b.

From the sequence table, we have that $2^7 = 128$, so that 7 is the logarithm of 128 to the base 2.

Similarly, $3^4 = 81$, and so 4 is the logarithm of 81 to the base 3.

$y = b^N$ we say that $N = \log_b y$.

That is, N is the logarithm of y to the base b, which corresponds to the power that the base b must be raised so that the result is y.

To determine the number $\log_2 32$, we ask ourselves the following question: "To what power must we raise the number 2, so that our result is 32?"

Letting $x = \log_2 32$, we must find the number x such that $2^x = 32$.

Clearly then, $x = 5$, and so we have that $\log_2 32 = 5$.

One convention in setting out such questions is:

$$\log_2 32 = x \Leftrightarrow 2^x = 32$$

$$\Leftrightarrow x = 5$$

As in part a, we ask the question "To what power must we raise the number 10, so that our result is 1000?"

That is, $x = \log_{10} 1000 \Leftrightarrow 10^x = 1000$

$$\Leftrightarrow x = 3$$

So that $\log_{10} 1000 = 3$.

Now, $x = \log_3 729 \Leftrightarrow 3^x = 729 \quad \Leftrightarrow x = 6$

(This was obtained by trial and error.)

Therefore, $\log_3 729 = 6$.

Although we have a fraction, this does not alter the process:

$$x = \log_2 \frac{1}{16} \Leftrightarrow 2^x = \frac{1}{16}$$

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

$$\Leftrightarrow x = -4$$

Can we find the logarithm of a negative number?

To evaluate $\log_a(-4)$ for some base $a > 0$, we need to solve the equivalent statement:

$$x = \log_a(-4) \Leftrightarrow a^x = -4$$

However, the value of a^x where $a > 0$, will always be positive, therefore there is no value of x for which $a^x = -4$. This means that

we cannot evaluate the logarithm of a negative number.

We can now make our definition a little stronger:

$$N = \log_b y \Leftrightarrow y = b^N \; y > 0$$

Example 1.2.6

Find the value of x given that:

a $\log_e x = 3$ b $\log_e(x-2) = 0.5$ c $\log_e 5 = x$

a $\quad \log_e x = 3 \Leftrightarrow x = e^3 \approx 20.09$

b $\quad \log_e(x-2) = 0.5 \Leftrightarrow x - 2 = e^{0.5}$

$$\Leftrightarrow x = 2 + \sqrt{e}$$

$$\therefore x \approx 3.65$$

c $\quad x = \log_e 5 \approx 1.61$

The Nautilus

The Nautilus is a very ancient (and we can infer, successful) sea creature.

This is the fossil of an ammonite that appears to have been a precursor of the Nautilus. Note the logarithmic spiral shape.

The different sizes of the segments probably result from the natural growth of the animal.

A nautilus in action:

It is jet-propelled.

Exercise 1.2.3

1. Use the definition of a logarithm to determine the following.

   a $\log_6 36$ b $\log_7 49$ c $\log_3 243$

   d $\log_4 64$ e $\log_2\left(\frac{1}{8}\right)$ f $\log_3\left(\frac{1}{9}\right)$

2. Change the following exponential expressions into their equivalent logarithmic form.

   a $10^4 = 10000$ b $10^{-3} = 0.001$

   c $10^y = x+1$ d $10^7 = p$

   e $2^y = x-1$ f $2^{4x} = y-2$

3. Change the following logarithmic expressions into their equivalent exponential form.

   a $\log_2 x = 9$ b $\log_b y = x$ c $\log_b t = ax$

   d $\log_{10} z = x^2$ e $\log_{10} y = 1-x$

4. Solve for x in each of the following.

   a $\log_2 x = 4$ b $\log_3 9 = x$ c $\log_4 x = \frac{1}{2}$

   d $\log_x 3 = \frac{1}{2}$ e $\log_x 2 = 4$ f $\log_5 x = 3$

5. Solve for x in each of the following, giving your answer to 4 d.p.

   a $\log_e x = 4$ b $\log_e 4 = x$ c $\log_e x = \frac{1}{2}$

   d $\log_x e = \frac{1}{2}$ e $\log_x e = 2$ f $\log_x e = -1$

Extra questions

THE ALGEBRA OF LOGARITHMS

The following logarithmic laws are a direct consequence of the definition of a logarithm and the index laws already established.

First law: **The logarithm of a product**

$$\log_a(x \times y) = \log_a x + \log_a y, x > 0, y > 0$$

Proof: Let $M = \log_a x$ and $N = \log_a y$ so that $x = a^M$ and $y = a^N$.

$$\text{Then, } x \times y = a^M \times a^N$$

$$\Leftrightarrow x \times y = a^{M+N}$$

$$\Leftrightarrow \log_a(x \times y) = M + N$$

$$\Leftrightarrow \log_a(x \times y) = \log_a x + \log_a y$$

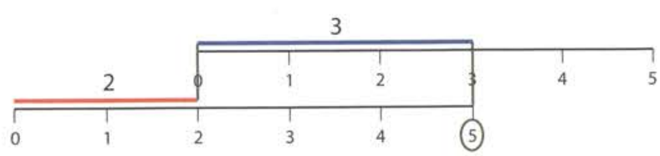
Originally, logarithms were viewed as a series. This led to the logarithmic scale.

Linear Scale

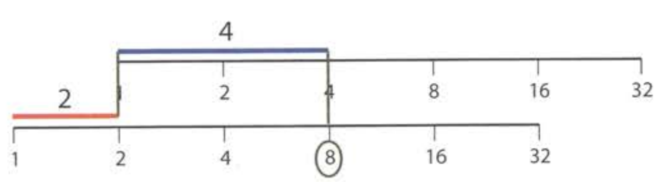
Logarithmic Scale

The linear scale is arithmetic and the logarithmic scale is geometric.

A pair of linear scales can be used to add numbers: In this case 2 + 3 = 5.



A pair of logarithmic scales can be used to multiply numbers. In this case 2 × 4 = 8



Note that other multiplications by 2 can be read from the diagram.

This is a brief demonstration of the use of a slide rule:

Example 1.2.7

Simplify the following expressions:

a $\log_3 x + \log_3(4x)$ b $\log_2 x + \log_2(4x)$

a $\log_3 x + \log_3(4x) = \log_3(x \times 4x)$

$= \log_3 4x^2$

b This time we note that because the base is '2' and there is a '4' in one of the logarithmic expressions, we could first try to remove the '4'.

$$\log_2 x + \log_2(4x) = \log_2 x + (\log_2 4 + \log_2 x)$$
$$= \log_2 x + 2 + \log_2 x$$
$$= 2\log_2 x + 2$$

Example 1.2.8

Given that $\log_a p = 0.70$ and $\log_a q = 2$, evaluate the following.

a $\log_a p^2$ b $\log_a(p^2 q)$ c $\log_a(apq)$

a $\log_a p^2 = \log_a(p \times p) = \log_a p + \log_a p$

$= 2\log_a p$

$= 2 \times 0.70$

$= 1.40$

b $\log_a(p^2 q) = \log_a p^2 + \log_a q$

$= 2\log_a p + \log_a q$

$= 1.40 + 2$

$= 3.40$

c $\log_a(apq) = \log_a a + \log_a p + \log_a q$

$= 1 + 0.70 + 2$

$= 3.70$

Example 1.2.9

Find $\{x| \log_2 x + \log_2(x+2) = 3\}$.

$$\log_2 x + \log_2(x+2) = 3 \Leftrightarrow \log_2[x \times (x+2)] = 3$$
$$\Leftrightarrow x(x+2) = 2^3$$
$$\Leftrightarrow x^2 + 2x = 8$$
$$\Leftrightarrow x^2 + 2x - 8 = 0$$
$$\Leftrightarrow (x+4)(x-2) = 0$$
$$\Leftrightarrow x = -4 \text{ or } x = 2$$

Next, we must check our solutions.

When $x = -4$, substituting into the **original equation**, we have: L.H.S $= \log_2(-4) + \log_2(-4+2)$ – which cannot be evaluated (as the logarithm of a negative number does not exist). Therefore, $x = -4$, is not a possible solution.

When $x = 2$, substituting into the **original equation**, we have:

L.H.S $= \log_2(2) + \log_2(2+2)$

$= \log_2 8$

$= 3$

$=$ R.H.S

Therefore, $\{x| \log_2 x + \log_2(x+2) = 3\} = \{2\}$.

Second law: The logarithm of a quotient

$$\log_a\left(\frac{x}{y}\right) = \log_a x - \log_a y, x > 0, y > 0$$

Proof: Let $M = \log_a x$ and $N = \log_a y$ so that $x = a^M$ and $y = a^N$.

Then, $\frac{x}{y} = \frac{a^M}{a^N}$

$$\Leftrightarrow \frac{x}{y} = a^{M-N}$$
$$\Leftrightarrow \log_a\left(\frac{x}{y}\right) = M - N$$
$$\Leftrightarrow \log_a\left(\frac{x}{y}\right) = \log_a x - \log_a y$$

Example 1.2.10

Simplify: a $\log_{10}100x - \log_{10}xy$

b $\log_2 8x^3 - \log_2 x^2 + \log_2\left(\frac{y}{x}\right)$

a $\log_{10}100x - \log_{10}xy = \log_{10}\left(\frac{100x}{xy}\right) = \log_{10}\left(\frac{100}{y}\right)$

Note: We could then express $\log_{10}\left(\frac{100}{y}\right)$ as:

$\log_{10}100 - \log_{10}y = 2 - \log_{10}y$.

b $\log_2 8x^3 - \log_2 x^2 + \log_2\left(\frac{y}{x}\right) = \log_2\left(\frac{8x^3}{x^2}\right) + \log_2\left(\frac{y}{x}\right)$

$= \log_2 8x + \log_2\left(\frac{y}{x}\right)$

$= \log_2\left(8x \times \frac{y}{x}\right)$

$= \log_2 8y$

Note: We could then express $\log_2 8y$ as:

$\log_2 8 + \log_2 y = 3 + \log_2 y$.

Example 1.2.11

Find $\{x|\ \log_{10}(x+2) - \log_{10}(x-1) = 1\}$.

$\log_{10}(x+2) - \log_{10}(x-1) = 1 \Leftrightarrow \log_{10}\left(\frac{x+2}{x-1}\right) = 1$

$\Leftrightarrow \left(\frac{x+2}{x-1}\right) = 10^1$

$\Leftrightarrow x + 2 = 10x - 10$

$\Leftrightarrow 12 = 9x$

$\Leftrightarrow x = \frac{4}{3}$

Next, we check our answer. Substituting into the original equation, we have:

L.H.S

$\log_{10}\left(\frac{4}{3}+2\right) - \log_{10}\left(\frac{4}{3}-1\right) = \log_{10}\frac{10}{3} - \log_{10}\frac{1}{3} = \log_{10}\left(\frac{10}{3} \div \frac{1}{3}\right) =$

$= \log_{10}10 \quad = 1 = \text{R.H.S}$

Therefore, $\{x|\ \log_{10}(x+2) - \log_{10}(x-1) = 1\} = \left\{\frac{4}{3}\right\}$

Third law: **The logarithm of a power**

$$\log_a x^n = n.\log_a x, x > 0$$

Proof: This follows from repeated use of the First Law or it can be shown as follows:

Let $M = \log_a x \Leftrightarrow a^M = x \quad \Leftrightarrow (a^M)^n = x^n$

(Raising both sides to the power of n)

$\Leftrightarrow a^{nM} = x^n$

(Using the index laws)

$\Leftrightarrow nM = \log_a x^n$

(Converting from exponential to log form)

$\Leftrightarrow n\log_a x = \log_a x^n$

Example 1.2.12

Given that $\log_a x = 0.2$ and $\log_a y = 0.5$, evaluate:

a $\log_a x^3y^2$ b $\log_a\sqrt{\frac{x}{y^4}}$

$\log_a x^3y^2 = \log_a x^3 + \log_a y^2$

$= 3\log_a x + 2\log_a y$

$= 3 \times 0.2 + 2 \times 0.5$

$= 1.6$

$\log_a\sqrt{\frac{x}{y^4}} = \log_a\left(\frac{x}{y^4}\right)^{1/2} = \frac{1}{2}\log_a\left(\frac{x}{y^4}\right)$

$= \frac{1}{2}[\log_a(x) - \log_a y^4]$

$= \frac{1}{2}[\log_a x - 4\log_a y]$

$= \frac{1}{2}[0.2 - 4 \times 0.5]$

$= -0.9$

Fourth law: **Change of base**

$$\log_a b = \frac{\log_k b}{\log_k a}, a, k \in \mathbb{R}^+ \setminus \{1\}$$

Proof: Let $\log_a b = N$ so that $a^N = b$

Taking the logarithms to base k of both sides of the equation we have:

$$\log_k(a^N) = \log_k b \Leftrightarrow N\log_k a = \log_k b$$
$$\Leftrightarrow N = \frac{\log_k b}{\log_k a}$$

However, we have that $\log_a b = N$, therefore, $\log_a b = \dfrac{\log_k b}{\log_k a}$

Other observations include:

1. $\log_a a = 1$
2. $\log_a 1 = 0$
3. $\log_a x^{-1} = -\log_a x, x > 0$
4. $\log_{1/a} x = -\log_a x$
5. $a^{\log_a x} = x, x > 0$

MISCELLANEOUS EXAMPLES

Example 1.2.13

Express y in terms of x if:

a $2 + \log_{10} x = 4\log_{10} y$

b $\log x = \log(a - by) - \log a$

a Given that $2 + \log_{10} x = 4\log_{10} y$

then $2 = 4\log_{10} y - \log_{10} x$

$$\Leftrightarrow 2 = \log_{10} y^4 - \log_{10} x$$
$$\Leftrightarrow 2 = \log_{10}\left(\frac{y^4}{x}\right)$$
$$\Leftrightarrow 10^2 = \frac{y^4}{x}$$
$$\Leftrightarrow y^4 = 100x$$
$$\Leftrightarrow y = \sqrt[4]{100x} \quad (\text{as } y > 0)$$

b Given that $\log x = \log(a - by) - \log a$

then $\log x = \log\dfrac{a - by}{a}$

$$\Leftrightarrow x = \frac{a - by}{a}$$
$$\Leftrightarrow ax = a - by$$
$$\Leftrightarrow by = a - ax$$
$$\Leftrightarrow y = \frac{a}{b}(1 - x)$$

Example 1.2.14

Find x if:

a $\log_x 64 = 3$ b $\log_{10} x - \log_{10}(x - 2) = 1$

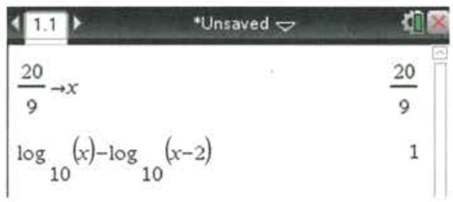
a $\log_x 64 = 3 \Leftrightarrow x^3 = 64$

$$\Leftrightarrow x^3 = 4^3$$
$$\Leftrightarrow x = 4$$

b $\log_{10} x - \log_{10}(x - 2) = 1 \Leftrightarrow \log_{10}\left(\dfrac{x}{x - 2}\right) = 1$

$$\Leftrightarrow \frac{x}{x - 2} = 10^1$$
$$\Leftrightarrow x = 10x - 20$$
$$\Leftrightarrow -9x = -20$$
$$\Leftrightarrow x = \frac{20}{9}$$

Check:



Example 1.2.15

Find $\{x \mid 5^x = 2^{x+1}\}$. Give both an exact answer and one to 2 d.p.

Taking the logarithm of base 10 of both sides $5^x = 2^{x+1}$ gives:

$$5^x = 2^{x+1} \Leftrightarrow \log_{10} 5^x = \log_{10} 2^{x+1}$$
$$\Leftrightarrow x\log_{10} 5 = (x + 1)\log_{10} 2$$
$$\Leftrightarrow x\log_{10} 5 - x\log_{10} 2 = \log_{10} 2$$
$$\Leftrightarrow x(\log_{10} 5 - \log_{10} 2) = \log_{10} 2$$
$$\Leftrightarrow x = \frac{\log_{10} 2}{\log_{10} 5 - \log_{10} 2}$$

And so, $x = 0.75647\ldots = 0.76$ (to 2 d.p).

Exact answer $= \left\{\dfrac{\log_{10} 2}{\log_{10} 5 - \log_{10} 2}\right\}$, answer to 2 d.p = $\{0.76\}$

Example 1.2.16

Find x where $6e^{2x} - 17 \times e^x + 12 = 0$.

We first note that $6e^{2x} - 17 \times e^x + 12$ can be written as $6 \times e^{2x} - 17 \times e^x + 12$.

This in turn can be expressed as $6 \times (e^x)^2 - 17 \times e^x + 12$.

Therefore, making the substitution $y = e^x$, we have that $6 \times (e^x)^2 - 17 \times e^x + 12 = 6y^2 - 17y + 12$ (i.e. we have a 'hidden' quadratic).

Solving for y, we have:
$6y^2 - 17y + 12 = 0 \Leftrightarrow (2y-3)(3y-4) = 0$

So that $y = \frac{3}{2}$ or $y = \frac{4}{3}$

However, we wish to solve for x, and so we need to substitute back:

$$e^x = \frac{3}{2} \text{ or } e^x = \frac{4}{3}$$

$$\Leftrightarrow x = \ln\frac{3}{2} \text{ or } x = \ln\frac{4}{3}$$

Example 1.2.17

Solve for x where $8^{2x+1} = 4^{5-x}$.

Taking logs of both sides of the equation $8^{2x+1} = 4^{5-x}$, we have

$$\log 8^{2x+1} = \log 4^{5-x} \Leftrightarrow (2x+1)\log 8 = (5-x)\log 4$$

$$\Leftrightarrow (2x+1)\log 2^3 = (5-x)\log 2^2$$

$$\Leftrightarrow 3(2x+1)\cancel{\log 2} = 2(5-x)\cancel{\log 2}$$

Therefore, we have that $6x + 3 = 10 - 2x \Leftrightarrow 8x = 7$

$$\therefore x = \frac{7}{8}$$

Exercise 1.2.4

1. Without using a calculator, evaluate the following.

   a $\log_2 8 + \log_2 4$ b $\log_6 18 + \log_6 2$

   c $\log_5 2 + \log_5 12.5$ d $\log_3 18 - \log_3 6$

2. Write down an expression for $\log a$ in terms of $\log b$ and $\log c$ for the following.

   a $a = bc$ b $a = b^2 c$

   c $a = \frac{1}{c^2}$ d $a = b\sqrt{c}$

3. Given that $\log_a x = 0.09$, find:

   a $\log_a x^2$ b $\log_a \sqrt{x}$ c $\log_a\left(\frac{1}{x}\right)$

4. Express each of the following as an equation that does not involve a logarithm.

   a $\log_2 x = \log_2 y + \log_2 z$

   b $\log_{10} y = 2\log_{10} x$

   c $\log_2(x+1) = \log_2 y + \log_2 x$

5. Solve the following equations.

   a $\log_2(x+1) - \log_2 x = \log_2 3$

   b $\log_{10}(x+1) - \log_{10} x = \log_{10} 3$

   c $\log_2(x+1) - \log_2(x-1) = 4$

6. Simplify the following

   a $\log_3(2x) + \log_3 w$ b $\log_4 x - \log_4(7y)$

7. Solve the following

   a $\log_2(x+7) + \log_2 x = 3$

   b $\log_3(x+3) + \log_3(x+5) = 1$

c $\log_{10}(x+7) + \log_{10}(x-2) = 1$

8. Solve for x.

a $\log_2 x^2 = (\log_2 x)^2$ b $\log_3 x^3 = (\log_3 x)^3$

9. Solve the following, giving an exact answer and an answer to 2 d.p.

a $2^x = 14$ b $10^x = 8$

c $3^x = 125$ d $\dfrac{1}{1-2^x} = 12$

10. Solve for x.

a $(\log_2 x)^2 - \log_2 x - 2 = 0$

b $\log_2(2^{x+1} - 8) = x$

11. Solve the following simultaneous equations.

a $\begin{aligned} x^y &= 5x - 9 \\ \log_x 11 &= y \end{aligned}$ b $\begin{aligned} \log_{10} x - \log_{10} y &= 1 \\ x + y^2 &= 200 \end{aligned}$

12. Express each of the following as an equation that does not involve a logarithm.

a $\log_e x = \log_e y - \log_e z$ b $3\log_e x = \log_e y$

13. Solve the following for x.

a $\ln(x+1) - \ln x = 4$ b $\ln(x+1) - \ln x = \ln 4$

14. Solve the following for x.

a $e^x = 21$ b $e^x - 2 = 8$

15. Solve the following for x.

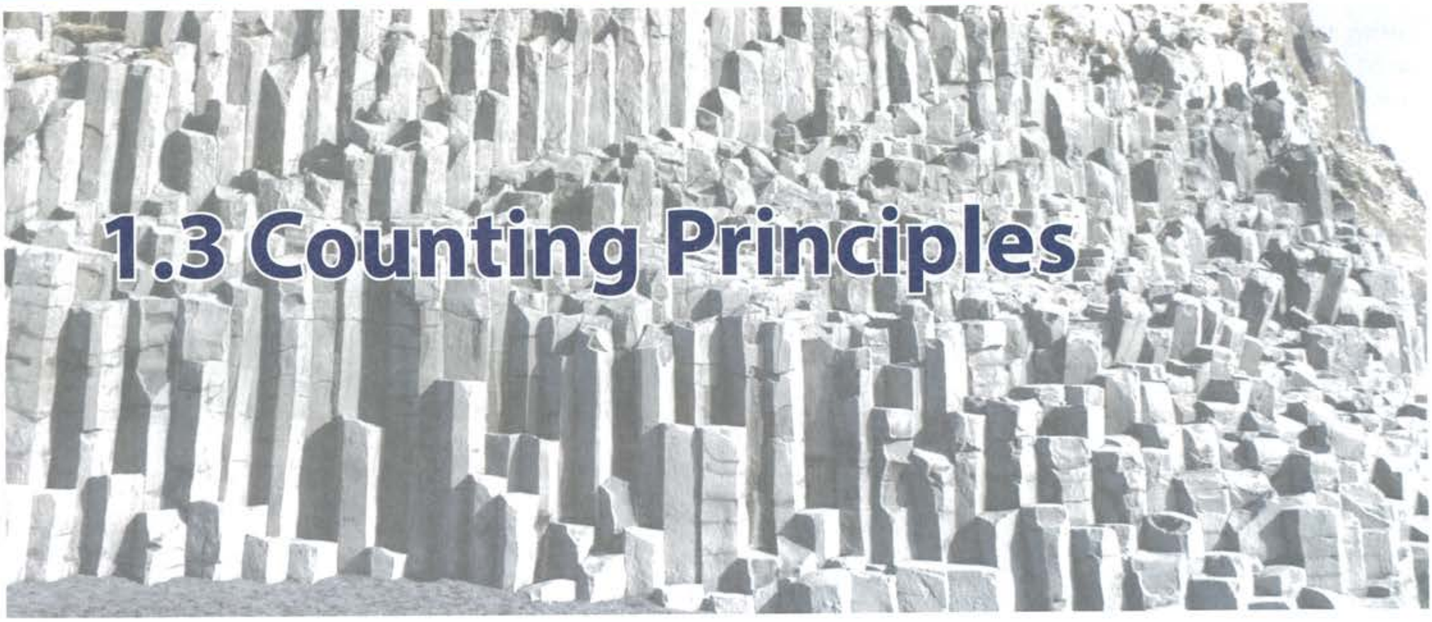
a $e^{2x} - 3e^x + 2 = 0$

b $e^{2x} - 4e^x - 5 = 0$

16. If: $\log_k(x+3) = \log_k x - 4$, prove that:

Extra questions

Answers



1.3 Counting Principles

Permutations

Permutations represent a counting process where the order must be taken into account.

For example, the number of permutations of the letters A, B, C and D, if only two are taken at a time, can be enumerated as

AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, BC

That is, AC is a different permutation from CA (different order).

Instead of permutation the term arrangement is often used.

This definition leads to a number of **Counting Principles.**

Multiplication principle

Rule 1: If any one of n different mutually exclusive and exhaustive events can occur on each of k trials, the number of possible outcomes is equal to n^k.

For example, if a die is rolled twice, there are a total of $6^2 = 36$ possible outcomes.

Rule 2: If there are n_1 events on the first trial, n_2 events on the second trial, and so on, and finally, n_k events on the kth trial, then the number of possible outcomes is equal to $n_1 \times n_2 \times \ldots \times n_k$.

For example, if a person has three different coloured pairs of pants, four different shirts, five different ties and three different coloured pairs of socks, the total number of different ways that this person can dress is equal to $3 \times 4 \times 5 \times 3 = 180$ ways.

Rule 3: The total number of ways that n different objects can be arranged in order is equal to $n \times (n-1) \times (n-2) \times \ldots 3 \times 2 \times 1$.

Because of the common usage of this expression, we use the factorial notation. That is, we write:

$$n! = n \times (n-1) \times (n-2) \times (n-3) \times \ldots \times 3 \times 2 \times 1, n > 0$$

which is read as ***n* factorial**. Notice also that 0! is defined as 1, i.e. $0! = 1$.

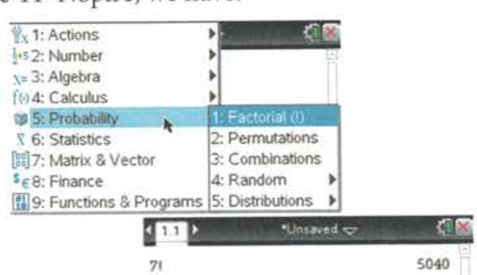
For example, in how many ways can 4 boys and 3 girls be seated on a park bench? In this case any one of the seven children can be seated at one end, meaning that the adjacent position can be filled by any one of the remaining six children, similarly, the next adjacent seat can be occupied by any one of the remaining 5 children, and so on . . .

Therefore, in total there are:

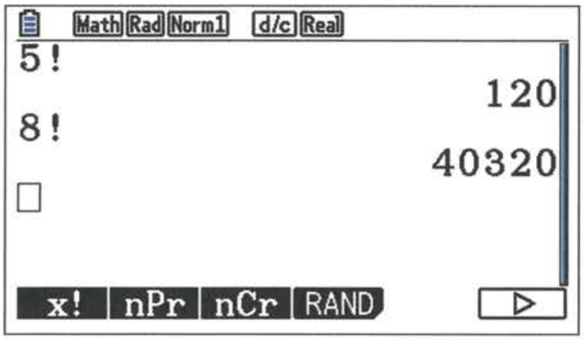
$$7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 7! = 5040$$

possible arrangements.

Using the TI–NSpire, we have:



If using a Casio graphic calculator, the permutation functions can be accessed in run mode by using the OPTN key, F6 for 'more' and F3 for PROB.

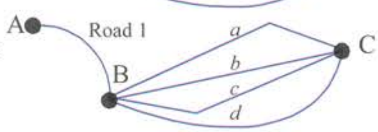


Example 1.3.1

Jai wishes to get from town A to town C via town B. There are three roads connecting town A to town B and 4 roads connecting town B to town C. In how many different ways can Jai get from town A to town C?

We start by visualizing this situation:

Consider the case where Jai uses Road 1 first.



The possibilities are:

Road 1 then a, Road 1 then b,

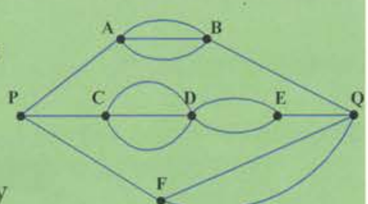
Road 1 then c, Road 1 then d.

That is, there are 4 possible routes. Then, for each possible road from A to B there are another 4 leading from B to C.

All in all, there are $4+4+4 = 3\times 4 = 12$ different ways Jai can get from A to C via B.

Example 1.3.2

Using the following street network, in how many different ways can a person get from point P to point Q if they can only move from left to right?



In travelling from P to Q there are:

$3 = 1 \times 3 \times 1$ paths (along P to A to B to Q)

$6 = 1 \times 3 \times 2 \times 1$ paths (along P to C to D to E to Q)

$2 = 1 \times 2$ paths (along P to F to Q)

In total there are $3 + 6 + 2 = 11$ paths

Example 1.3.3

A golfer has 3 drivers, 4 tees and 5 golf balls. In how many ways can the golfer take his first hit?

The golfer has 3 possible drivers to use and so the first task can be carried out in 3 ways.

The golfer has 4 possible tees to use and so the second task can be carried out in 4 ways.

The golfer has 5 golf balls to use and so the third task can be carried out in 5 ways.

Using the multiplication principle, there are $3 \times 4 \times 5 = 60$ ways to take the first hit.

Permutations

Based on the definition given in §1.3 we have the following rule:

Rule 4:

The total number of ways of **arranging** n objects, taking r at a time is given by:

$$\frac{n!}{(n-r)!}$$

Notation: We use the notation ${}^nP_r$ (read as "n-p-r') to denote $\frac{n!}{(n-r)!}$.

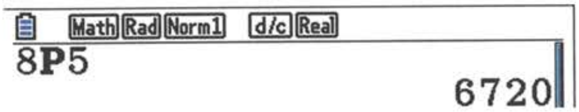
That is, ${}^nP_r = \frac{n!}{(n-r)!}$.

For example, the total number of arrangements of 8 books on a bookshelf if only 5 are used is given by:

$${}^8P_5 = \frac{8!}{(8-5)!} = \frac{8!}{3!} = 6720.$$

When using the TI calculator, use MENU, 5, 2:

or using Casio (F2):



Example 1.3.4

In how many ways can 5 boys be arranged in a row:

a using three boys at a time?

b using 5 boys at a time?

a We have 5 boys to be arranged in a row with certain constraints.

The constraint is that we can only use 3 boys at a time. In other words, we want the number of arrangements (permutations) of 5 objects taken 3 at a time.

From rule 4: $n = 5, r = 3,$

$$\text{Therefore, number of arrangements} = {}^5P_3 = \frac{5!}{(5-3)!} = \frac{120}{2} = 60$$

b This time we want the number of arrangements of 5 boys taking all 5 at a time.

From rule 4: $n = 5, r = 5,$

$$\text{Therefore, number of arrangements} = {}^5P_5 = \frac{5!}{(5-5)!} = \frac{120}{0!} = 120$$

Box method

Problems like Example 1.3.4 can be solved using a method known as "the box method". In that particular example, part (a) can be considered as filling three boxes (with only one object per box) using 5 objects:

Box 1	Box 2	Box 3

The first box can be filled in 5 different ways (as there are 5 possibilities available). Therefore we 'place 5' in box 1:

Box 1	Box 2	Box 3
5		

Now, as we have used up one of the objects (occupying box 1), we have 4 objects left that can be used to fill the second box. So, we 'place 4' in box 2:

Box 1	Box 2	Box 3
5	4	

At this stage we are left with three objects (as two of them have been used). This means that there are 3 possible ways in which the third box can be filled. So, we 'place 3' in box 3:

Box 1	Box 2	Box 3
5	4	3

This is equivalent to saying, that we can carry out the first task in 5 different ways, the second task in 4 different ways and the third task in 3 different ways. Therefore, using the multiplication principle we have that the total number of arrangements is $5 \times 4 \times 3 = 60$.

Comparing this to the expression ${}^5P_3 = \frac{5!}{(5-3)!}$ we have:

$$\frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 5 \times 4 \times 3 = 60$$

i.e. the last step in the evaluation process is the same as the step used in the 'box method'.

Example 1.3.5

Vehicle licence plates consist of two letters from a 26-letter alphabet, followed by a three-digit number whose first digit cannot be zero. How many different licence plates can there be?

We have a situation where there are five positions to be filled:

Letter	Letter	Number	Number	Number

That is, the first position must be occupied by one of 26 letters, similarly, the second position must be occupied by one of 26 letters. The first number must be made up of one of nine different digits (as zero must be excluded), whilst the other two positions have 10 digits that can be used. Therefore, using Rule 2, we have:

Total number of arrangements =
$26 \times 26 \times 9 \times 10 \times 10 = 608400$.

Example 1.3.6

How many 5-digit numbers greater than 40,000 can be formed from 0, 1, 2, 3, 4, and 5 if:

a there is no repetition of digits allowed?

b repetition of digits is allowed?

a Consider the five boxes:

Box 1	Box 2	Box 3	Box 4	Box 5

Only the digits 4 and 5 can occupy the first box (so as to obtain a number greater than 40,000). So there are 2 ways to fill box 1:

Box 1	Box 2	Box 3	Box 4	Box 5
2				

Box 2 can now be filled using any of the remaining 5 digits. So, there are 5 ways of filling box 2:

Box 1	Box 2	Box 3	Box 4	Box 5
2	5			

We now have 4 digits left to be used. So, there are 4 ways of filling box 3:

Box 1	Box 2	Box 3	Box 4	Box 5
2	5	4		

Continuing in this manner we have:

Box 1	Box 2	Box 3	Box 4	Box 5
2	5	4	3	2

Then, using the multiplication principle we have $2 \times 5 \times 4 \times 3 \times 2 = 240$ arrangements.

Otherwise, we could have relied on rule 4 and obtained $2 \times {}^5P_4 = 2 \times 120 = 240$

As in part a, only the digits 4 and 5 can occupy the first box.

b If repetition is allowed, then boxes 2 to 5 can each be filled using any of the 6 digits:

Box 1	Box 2	Box 3	Box 4	Box 5
2	6	6	6	6

Using the multiplication principle there are $2 \times 6 \times 6 \times 6 \times 6 = 2592$ arrangements.

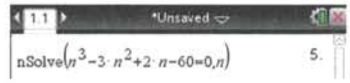
However, one of these arrangements will also include the number 40 000. Therefore, the number of 5 digit numbers greater than 40,000 (when repetition is allowed) is given by $2592 - 1 = 2591$.

Example 1.3.7

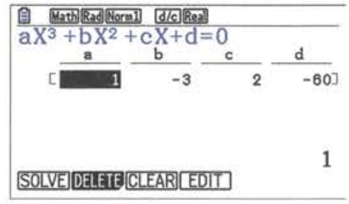
Find n if ${}^nP_3 = 60$.

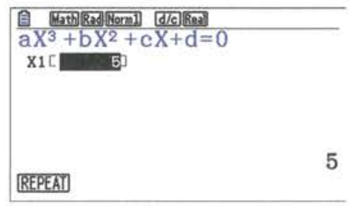
$$
\begin{aligned}
{}^nP_3 = 60 \Leftrightarrow \frac{n!}{(n-3)!} &= 60 \\
\Leftrightarrow \frac{n(n-1)(n-2)(n-3)!}{(n-3)!} &= 60 \\
\Leftrightarrow n(n-1)(n-2) &= 60 \\
\Leftrightarrow n^3 - 3n^2 + 2n &= 60 \\
\Leftrightarrow n^3 - 3n^2 + 2n - 60 &= 0
\end{aligned}
$$

Using the a TI calculator to solve this polynomial, we have: $n = 5$



With a Casio model, use the Equation Module (A), F2 Polynomial:





Example 1.3.8

How many different arrangements are there of the letters of the word HIPPOPOTAMUS?

The word 'HIPPOPOTAMUS' is made up of 12 letters, unfortunately, they are not all different! This means that, although we can swap the three Ps with each other, the word will remain the same.

Now, the total number of times we can rearrange the Ps (and not alter the word) is 3! = 6 times (as there are three Ps). Therefore, if we 'blindly' use Rule 2, we will have increased the number of arrangements 6 fold.

Therefore, we will need to divide the total number of ways of arranging 12 objects by 6.

That is, $\frac{12!}{3!} = 79833600$.

However, we also have 2 Os, and so, the same argument holds. So that in fact, we now have a total of:

$$\frac{12!}{3! \times 2!} = 39916800 \text{ arrangements.}$$

This example is a special case of **permutations with repetitions**:

Rule 5: The number of permutations of n objects of which n_1 are identical, n_2 are identical, . . . , n_k are identical is given by:

$$\frac{n!}{n_1! \times n_2! \times n_3! \times \ldots \times n_k!}$$

Example 1.3.9

In how many ways can 5 apples, 3 oranges, and 2 pears be arranged in a straight line, so that the three oranges are all together?

First we must treat the 3 oranges as one unit and there are 3! ways to arrange the oranges. Hence, there are 8 units which can be permuted in 8! ways when they are arranged in a straight line.

Using Rule 2, we have that the total number of possible arrangement $3! \times 8! = 241\,920$.

Example 1.3.10

In how many ways can 8 apples and 8 oranges be arranged in a straight line, so that no two apples are next to each other?

First we must arrange the oranges first and there are $8! = 40\,320$ ways. Now there are 9 positions to arrange the 8 apples. Hence the apples can be arranged in ${}^9P_8 = 362\,880$ ways.

Using Rule 2, we have that the total number of $8! \times {}^9P_8$ possible arrangements ($\approx 1.46 \times 10^{10}$).

Lotteries

The TattsLotto game has forty-five balls numbered 1 to 45 from which eight are randomly selected. Is this selection with replacement? Does the order matter? What are your chances of winning with a single "game"?

Exercise 1.3.1

1. A, B and C are three towns. There are 5 roads linking towns A and B and 3 roads linking towns B and C. How many different paths are there from town A to town C via town B?

2. In how many ways can 5 letters be mailed if there are:

   a 2 mail boxes available?

   b 4 mail boxes available?

3. There are 4 letters to be placed in 4 letter boxes. In how many ways can these letters be mailed if:

   a only one letter per box is allowed?

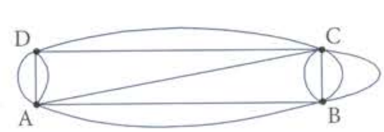
   b there are no restrictions on the number of letters per box?

4 Consider the cubic polynomial: $p(x) = ax^3 + bx^2 - 5x + c$.

   a If the coefficients, a, b and c come from the set $\{-3, -1, 1, 3\}$, find the number of possible cubics if no repetitions are allowed.

   b Find the number of cubics if the coefficients now come from $\{-3, -1, 0, 1, 3\}$ (again without repetitions).

5. The diagram alongside shows the possible routes linking towns A, B, C and D.



A person leaves town A for town C. How many different routes can be taken if the person is always heading towards town C?

6. In how many different ways can Susan get dressed if she has 3 skirts, 5 blouses, 6 pairs of socks and 3 pairs of shoes to choose from?

7. In how many different ways can 5 different books be arranged on a shelf?

8. In how many ways can 8 different boxes be arranged taking 3 at a time?

9. How many different signals can be formed using 3 flags from 5 different flags.

10. Three Italian, two chemistry and four physics books are to be arranged on a shelf.

    In how many ways can this be done if:

    a there are no restrictions?

    b the chemistry books must remain together?

    c the books must stay together by subject?.

11. Find n if ${}^{n}P_{2} = 380$.

12. Five boys and five girls, which include a brother-sister pair, are to be arranged in a straight line. Find the number of possible arrangements if:

    a there are no restrictions.

    b the tallest must be at one end and the shortest at the other end.

    c the brother and sister must be:
    i together **ii** separated.

13. In how many ways can the letters of the word Mississippi be arranged?

14. In how many ways can three yellow balls, three red balls and four orange balls be arranged in a row if the balls are identical in every way other than their colour?

15. In a set of 8 letters, m of them are the same and the rest different. If there are 1680 possible arrangements of these 8 letters, how many of them are the same?

Combinations

Combinations represent a counting process where the order has no importance. For example, the number of combinations of the letters A, B, C and D, if only two are taken at a time, can be enumerated as:

AB, AC, AD, BC,BD, CD,

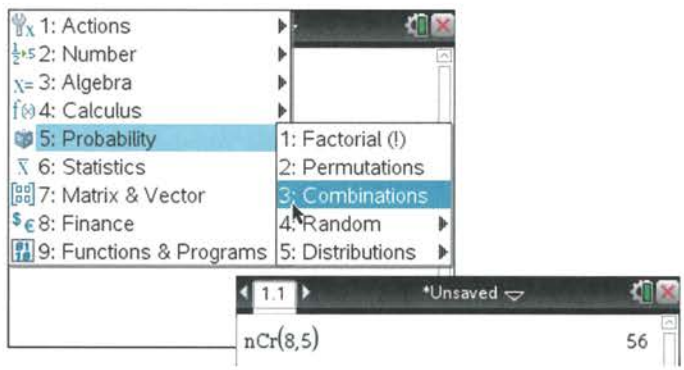
That is, the combination of the letters A and B, whether written as AB or BA, is considered as being the same.

Instead of **combination** the term **selection** is often used.

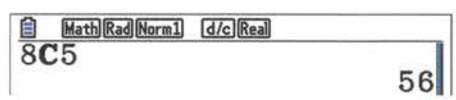
For example, in how many ways can 5 books be selected from 8 different books? In this instance, we are talking about selections and therefore, we are looking at combinations. Therefore we have, the selection of 8 books taking 5 at a time is equal to

$$\binom{8}{5} = \frac{8!}{(8-5)!5!} = \frac{8!}{3!5!} = 56 \ .$$

Using the TI–NSpire we can make use of the **nCr** function:



or using Casio:



Example 1.3.11

A sports committee at the local hospital consists of 5 members. A new committee is to be elected, of which 3 members must be women and 2 members must be men. How many different committees can be formed if there were originally 5 women and 4 men to select from?

First we look at the number of ways we can select the women members (using Rule 6):

We have to select 3 from a possible 5, therefore, this can be done in ${}^5C_3 = 10$ ways.

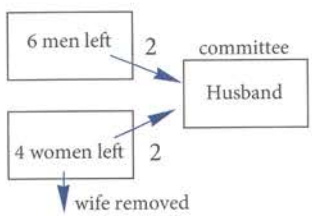
Similarly, the men can be selected in ${}^4C_2 = 6$ ways.

Using Rule 2, we have that the total number of possible committees $= {}^5C_3 \times {}^4C_2 = 60$.

Example 1.3.12

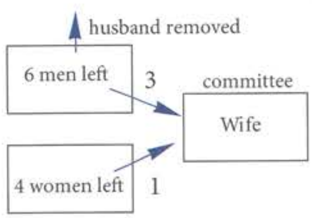
A committee of 3 men and 2 women is to be chosen from 7 men and 5 women. Within the 12 people there is a husband and wife. In how many ways can the committee be chosen if it must contain either the wife or the husband but not both?

Case 1: Husband included



If the husband is included, the wife must be removed (so that she cannot be included). We then have to select 2 more men from the remaining 6 men and 2 women from the remaining 4 women.

This is done in ${}^6C_2 \times {}^4C_2 = 90$ ways



Case 2: Wife included

If the wife is included, the husband must be removed. We then have to select 3 men from the remaining 6 men and 1 woman from the remaining 4 women.

This is done is ${}^6C_3 \times {}^4C_1 = 80$ ways

Therefore there are a total of ${}^6C_2 \times {}^4C_2 + {}^6C_3 \times {}^4C_1 = 90 + 80 = 170$ possible committees.

Example 1.3.13

From 15 cans, in how many ways can a selection of 5 be made, when one particular can is always included?

Since one can is always included in the selection, it can be removed from the original collection of 15 cans. For the remaining 14 cans, we must now select 4 more cans to complete the selection. Hence, we have ${}^{14}C_4 = 1001$ ways to make the selection.

Example 1.3.14

From 15 cans, in how many ways can a selection of 5 be made, when one particular can is always excluded?

Since one can is always excluded in the selection, it can be removed from the original collection of 15 cans. For the remaining 14 cans, we must now select 5 more cans to complete the selection. Hence, we have ${}^{14}C_5 = 2002$ ways to make the selection.

Exercise 1.3.2

1. In how many ways can 5 basketball players be selected from 12 players?

2. A tennis club has 20 members.

   a In how many ways can a committee of 3 be selected.

   b In how many ways can this be done if the captain must be on the committee?

3. In how many ways can 3 red balls, 4 blue balls and 5 white balls be selected from 5 red balls, 5 blue balls and 7 white balls?

4. In how many ways can 8 objects be divided into 2 groups of 4 objects?

5. A cricket training squad consists of 4 bowlers, 8 batsmen, 2 wicket keepers and 4 fielders.

   From this squad a team of 11 players is to be selected. In how many ways can this be done if the team must consist of 3 bowlers, 5 batsmen, 1 wicket keeper and 2 fielders?

6. A class consists of 12 boys of whom 5 are prefects. How many committees of 8 can be formed if the committee is to have:

   a 3 prefects?

   b at least 3 prefects?

7. An equal number of boys and girls are trying out for 8 spots on a sport team. If the team must consist of at least 5 girls, then how many different possible teams can result if 7 boys try out?

8. Show that:

   $1\times 1!+2\times 2!+3\times 3!+...+(n-1)\times(n-1)!+n\times n!$
   $=(n+1)!-1!$

Exercise 1.3.3

1. Five different coloured flags can be run up a mast.

   a How many different signals can be produced if all five flags are used?

   b How many different signals can be produced if any number of flags is used?

2. In how many different ways can 7 books be arranged in a row?

3. In how many different ways can 3 boys and 4 girls be seated in a row? In how many ways can this be done if:

   a no two girls are sitting next to each other?

   b the ends are occupied by girls?

4. In how many different ways can 7 books be arranged in a row if:

   a three specified books must be together?

   b two specified books must occupy the ends?

5. A school council consists of 12 members, 6 of whom are parents and 2 are students, the Principal and the remainder are teachers. The school captain and vice-captain must be on the council. If there are 10 parents and 8 teachers nominated for positions on the school council, how many different committees can there be?

6. A committee of 5 men and 5 women is to be selected from 9 men and 8 women.

   a How many possible committees can be formed?

   b Amongst the 17 people, there is a married couple. If the couple cannot serve together, how many committees could there be?

7. A sports team consists of 5 bowlers (or pitchers), 9 batsmen and 2 keepers (or back-stops). How many different teams of 11 players can be chosen from the above squad if the team consists of:

   a 4 bowlers (pitchers), 6 batsmen and 1 keeper (back-stop)?

   b 6 batsmen (pitchers) and at least 1 keeper (back-stop)?

8. Twenty people are to greet each other by shaking hands. How many handshakes are there?

9. How many arrangements of the letters of the word "MARRIAGE" are possible?

10. How many arrangements of the letters of the word "COMMISSION" are possible?

11. A committee of 4 is to be selected from 7 men and 6 women. In how many ways can this be done if:

    a there are no restrictions?

    b there must be an equal number of men and women on the committee?

    c there must be at least one member of each sex on the committee?

12. Prove that:

    a $\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$

    b ${}^{n+1}P_r = {}^{n}P_r + r\times {}^{n}P_{r-1}$.

13. A circle has n points on its circumference. How many chords joining pairs of points can be drawn?

14. A circle has n points on its circumference. What is the maximum number of points of intersection of chords inside the circle?

15. a Show that $2^n = \sum_{r=0}^{n}\binom{n}{r}$.

    b In how many ways can 8 boys be divided into two unequal sets?

16. Whilst at the library, Patrick decides to select 5 books from a group of 10. In how many different ways can Patrick make the selection?

17. A fish tank contains 5 gold-coloured tropical fish and 8 black-coloured tropical fish.

    a In how many ways can five fish be selected?

    b If a total of 5 fish have been selected from the tank, how many of these contain two gold fish?

Extra questions for Exercises 1.3.2 & 1.3.3

The Binomial Theorem

The binomial theorem is an example of the use of combinatorics.

Bracketed expressions such as $(2x-3)^7$ are said to be 'binomial' because there are two terms in the bracket (the prefix *bi* means two). Such expressions can be expanded using the distributive law. In a simple case such as $(a+b)^2$ the distributive law gives:

$$\begin{aligned}(a+b)^2 &= (a+b)(a+b)\\ &= a^2+ab+ba+b^2\\ &= a^2+2ab+b^2\end{aligned}$$

The distributive law states that each term in the first bracket must be multiplied by each term in the second bracket.

The next most complicated binomial can be evaluated using the previous result:

$$\begin{aligned}(a+b)^3 &= (a+b)(a+b)^2\\ &= (a+b)(a^2+2ab+b^2)\\ &= a^3+2a^2b+ab^2+a^2b+2ab^2+b^3\\ &= a^3+3a^2b+3ab^2+b^3\end{aligned}$$

Similarly, the fourth power of this simple binomial expression can be expanded as:

$$\begin{aligned}(a+b)^4 &= (a+b)(a+b)^3\\ &= (a+b)(a^3+3a^2b+3ab^2+b^3)\\ &= a^4+3a^3b+3a^2b^2+ab^3+a^3b+3a^2b^2+3ab^3+b^4\\ &= a^4+4a^3b+6a^2b^2+4ab^3+b^4\end{aligned}$$

The calculations are already fairly complex and it is worth looking at these results for the underlying pattern. There are three main features to the pattern. Looking at the fourth power example above, these patterns are:

The powers of a.

These start at 4 and decrease: a^4, a^3, a^2, a^1, a^0. Remember that $a^0 = 1$

The powers of b.

These start at 0 and increase: b^0, b^1, b^2, b^3, b^4. Putting these two patterns together gives the final pattern of terms in which the sum of the indices is always 4:

$$\underbrace{a^4}_{4} + \dots \underbrace{a^3b^1}_{3+1=4} + \dots \underbrace{a^2b^2}_{2+2=4} + \dots \underbrace{a^1b^3}_{1+3=4} + \dots \underbrace{b^4}_{4}$$

The coefficients complete the pattern.

These coefficients arise because there is more than one way of producing most of the terms. Following the pattern begun above, produces a triangular pattern of coefficients known as Pascal's Triangle. Blaise Pascal (1623-1662) developed early probability theory but is lucky to have this triangle named after him as it had been studied by Chinese mathematicians long before he was born.

So, if we continue our expansions (up to and including the sixth power) we have the following:

$$\begin{aligned}(x+a)^0 &= 1\\ (x+a)^1 &= 1x+1a\\ (x+a)^2 &= 1x^2+2ax+1a^2\\ (x+a)^3 &= 1x^3+3x^2a+3xa^2+1a^3\\ (x+a)^4 &= 1x^4+4x^3a+6x^2a^2+4xa^3+1a^4\\ (x+a)^5 &= 1x^5+5x^4a+10x^3a^2+10x^2a^3+5xa^4+1a^5\\ (x+a)^6 &= 1x^6+6x^5a+15x^4a^2+20x^3a^3+15x^2a^4+6xa^5+1a^6\\ &\vdots\\ \text{etc.} \quad \text{etc.} & \quad \text{etc.}\end{aligned}$$

Now consider only the coefficients for the above expansions. Writing down these coefficients we reproduce Pascal's triangle:

$$\begin{array}{c}1\\ 1 \quad 1\\ 1 \quad 2 \quad 1\\ 1 \quad 3 + 3 \quad 1\\ 1 \quad 4 \quad 6 \quad 4 \quad 1\\ 1 \quad 5 \quad 10 \quad 10 + 5 + 1\\ 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1\end{array}$$

etc.

The numbers in the body of the triangle are found by adding the two numbers immediately above and to either side.

An alternative to using Pascal's Triangle to find the coefficients is to use combinatorial numbers. If expanding $(a+b)^5$ the set of coefficients are:

$$\binom{5}{0} = 1, \binom{5}{1} = 5, \binom{5}{2} = 10, \binom{5}{3} = 10, \binom{5}{4} = 5, \binom{5}{5} = 1$$

which are the same as those given by Pascal's Triangle.

Most calculators can do this.

Example 1.3.15

Expand $(x+y)^6$.

Making use of Pascal's triangle we first determine the required coefficients:

1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

Write down each term:

$$x^6y^0 \quad x^5y^1 \quad x^4y^2 \quad x^3y^3 \quad x^2y^4 \quad x^1y^5 \quad x^0y^6$$

Combine the two steps:

$$x^6y^0 + 6x^5y^1 + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6x^1y^5 + x^0y^6$$

We can also combine the two steps into one:

$$(x+y)^6 = {}^6C_0x^6y^0 + {}^6C_1x^5y^1 + {}^6C_2x^4y^2 + {}^6C_3x^3y^3 + \ldots$$
$$\ldots + {}^6C_4x^2y^4 + {}^6C_5x^1y^5 + {}^6C_6x^0y^6$$

$$= x^6y^0 + 6x^5y^1 + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6x^1y^5 + x^0y^6$$

With practice, you should be able to expand such expressions as above.

What happens if we have the difference of two terms rather than the sum? For example, what about expanding $(x-a)^5$? The process is the same except that this time we rewrite the expression $(x-a)^5$ as follows:

$$(x-a)^5 = (x+(-a))^5$$

So, how do we proceed? Essentially, in exactly the same way.

Making use of Pascal's triangle we first determine the required coefficients:

1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

Notice that Pascal's triangle does not change due to the change in sign and so the coefficients remain the same whether it is a sum or difference.

Write down each term:

$$(x)^5(a)^0 \quad (x)^4(-a) \quad (x)^3(-a)^2 \quad (x)^2(-a)^3 \quad (x)^1(-a)^4 \quad (x)^0(-a)^5$$

i.e. $x^5 \quad -x^4a \quad x^3a^2 \quad -x^2a^3 \quad xa^4 \quad -a^5$

Combining the coefficients and the terms we have:

$$(x-a)^5 = x^5 - 5x^4a + 10x^3a^2 - 10x^2a^3 + 5xa^4 - a^5$$

Example 1.3.16

Expand

a $(4x-3)^3$ b $\left(2x-\frac{2}{x}\right)^3$.

a Making use of Pascal's triangle we first determine the required coefficients:

1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

Write down each term:

$$(4x)^3 \quad (4x)^2(-3) \quad (4x)(-3)^2 \quad (-3)^3$$

$$64x^3 \quad -48x^2 \quad 36x \quad -27$$

Combining steps 1 and 2 we have:

$$(4x-3)^3 = 1\times 64x^3 + 3\times -48x^2 + 3\times 36x + 1\times -27$$
$$= 64x^3 - 144x^2 + 108x - 27$$

b As with (a), the term pattern must be built on $(2x)$ and $-\frac{2}{x}$:

$$\left(2x-\frac{2}{x}\right)^3 = 1\times(2x)^3\left(-\frac{2}{x}\right)^0 + 3\times(2x)^2\left(-\frac{2}{x}\right)^1 + \ldots$$
$$\ldots + 3\times(2x)^1\left(-\frac{2}{x}\right)^2 + 1\times(2x)^0\left(-\frac{2}{x}\right)^3$$
$$= 8x^3 - 24x + \frac{24}{x} - \frac{8}{x^3}$$

Exercise 1.3.4

1. Expand the following binomial expressions:

a	$(b+c)^2$	b	$(a+g)^3$
c	$(1+y)^3$	d	$(2+x)^4$
e	$(2+2x)^3$	f	$(2x-4)^3$
g	$\left(2+\frac{x}{7}\right)^4$	h	$(2x-5)^3$
i	$(3x-4)^3$	j	$(3x-9)^3$
k	$(2x+6)^3$	l	$(b+3d)^3$
m	$(3x+2y)^4$	n	$(x+3y)^5$
o	$\left(2p+\frac{5}{p}\right)^3$	p	$\left(x^2-\frac{2}{x}\right)^4$
q	$\left(q+\frac{2}{p^3}\right)^5$	r	$\left(x+\frac{1}{x}\right)^3$

The general term

Recall that: ${}^nC_r = \binom{n}{r}$

We have seen the relationship between Pascal's triangle and its combinatorial equivalent. From this relationship we were able to produce the general expansion for $(x+a)^n$.

That is,

$$(x+a)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}a + \ldots + \binom{n}{r}x^{n-r}a^r + \ldots + a^n$$

Where the **first** term, $t_1 = x^n$

the **second** term, $t_2 = \binom{n}{1}x^{n-1}a$

the **third** term, $t_3 = \binom{n}{2}x^{n-2}a^2$

$\vdots$ $\qquad\qquad$ $\vdots$

the **rth** term, $t_r = \binom{n}{r-1}x^{n-(r-1)}a^{r-1}$

The (r +1)th term is also know as the general term.

That is: $t_{r+1} = \binom{n}{r}x^{n-r}a^r$

It is common in examinations for questions to only ask for a part of an expansion. This is because the previous examples are time consuming to complete.

Example 1.3.17

Find the 5th term in the expansion $\left(x+\frac{2}{x}\right)^{10}$, when expanded in descending powers of x.

The fifth term is given by t_5. Using the general term, this means that $r+1 = 5 \Leftrightarrow r = 4$.

For this expansion we have that $n = 10$, therefore,

$$t_5 = \binom{10}{4}(x)^{10-4}\left(\frac{2}{x}\right)^4 = 210 \times x^6 \times \frac{16}{x^4}$$
$$= 3360x^2$$

Therefore the fifth term is $t_5 = 3360x^2$.

Example 1.3.18

Find the term independent of x in the expansion $\left(2x-\frac{1}{x^2}\right)^6$.

In this case we want the term independent of x, that is, the term that involves x^0.

Again, we first find an expression for the general term,

$$t_{r+1} = \binom{6}{r}(2x)^{6-r}\left(-\frac{1}{x^2}\right)^r$$
$$= \binom{6}{r}(2)^{6-r}(x)^{6-r}(-1)^r(x^{-2})^r$$
$$= \binom{6}{r}2^{6-r}(-1)^r x^{6-r-2r}$$
$$= \binom{6}{r}2^{6-r}(-1)^r x^{6-3r}$$

Notice how we had to separate the constants and the x term.

Next, we equate the power of x in the expansion to 0: $6-3r = 0 \Leftrightarrow r = 2$.

We therefore want:

$$t_3 = \binom{6}{2}2^{6-2}(-1)^2x^{6-6} = 15 \times 16 \times 1 \times x^0 = 240.$$

So, the term independent of x is 240.

Exercise 1.3.5

1. Find the terms indicated in the expansions of the following expressions.

	Expression	*Term*
a	$(x+4)^5$	x^3
b	$(x+y)^7$	x^5y^2
c	$(2x-1)^8$	x^3
d	$(3x-2)^5$	x^4
e	$(2-3p^2)^4$	p^4
f	$(2p-3q)^7$	p^2q^5
g	$\left(3p-\frac{2}{p}\right)^7$	p

2. Find the coefficients of the terms indicated in the expansions of the following expressions.

	Expression	*Term*
a	$(2x-5)^8$	x^3
b	$(5x-2y)^6$	x^2y^4
c	$(x+3)^6$	x^3
d	$(2p-3q)^5$	p^4q
e	$\left(2x-\frac{3}{p}\right)^8$	$\frac{x^2}{p^6}$
f	$\left(q+\frac{2}{p^3}\right)^5$	$\frac{q^3}{p^6}$

3. Use the first three terms in the expansion of $(1+x)^4$ to find an approximate value for 1.01^4. Find the percentage error in using this approximation.

4. a Write the expansion of $(5+2x)^6$.

   b Use the first three terms of the expansion to approximate 5.2^6.

   c Find the absolute error in this approximation.

   d Find the percentage error in this approximation.

5. Find the coefficient of x^{-3} in the expansion of $(x-1)^3\left(\frac{1}{x}+x\right)^6$

6. Find the constant term in the expansion of $\left(x-\frac{1}{2x}\right)^{10}$

7. Find the constant term in the expansion of $\left(3x-\frac{1}{6x}\right)^{12}$

8. Find the term independent of x in the expansion of $(2-x)^3\left(\frac{1}{3x}-x\right)^6$

9. Find the term independent of x in the expansion of $\left(2x-\frac{1}{x}\right)^6\left(\frac{1}{2x}+x\right)^6$

10. In the expansion of $\left(x-\frac{a}{x}\right)^5\left(x+\frac{a}{x}\right)^5$, where a is a non-zero constant, the coefficient of the term in x^{-2} is '–9' times the coefficient in x^2. Find the value of the constant a.

11. If the coefficient of the x^2 in the expansion of $(1-3x)^n$ is 90, find n.

12. Three consecutive coefficients in the expansion of $(1+x)^n$ are in the ratio 6 : 14 : 21. Find the value of n.

13. Find the independent term in the following expansions:

    a $\left(y+\frac{1}{y}\right)^3\left(y-\frac{1}{y}\right)^5$ b $\left(2x+1-\frac{1}{2x^2}\right)^6$

14. In the expansion of $(1+ax)^n$ the first term is 1, the second term is $24x$ and the third term is $252x^2$. Find the values of a and n.

15. In the expansion of $(x+a)^3(x-b)^6$, the coefficient of x^7 is –9 and there is no x^8 term. Find a and b.

16. The first three terms in the expansion of $(1+ax)^n$ are given below. Find the value of a and n.

    a $1-18x+135x^2\ldots$

    b $1+\frac{5}{3}x+\frac{10}{9}x^2\ldots$

    c $1+4\sqrt{2}x+12x^2\ldots$

    d $1-\frac{2\sqrt{2}}{3}x+\frac{1}{3}x^2\ldots$

Answers

1.4 Mathematical Induction

The principles of Mathematical Induction and proofs

Induction is an indirect method of proof which is used in cases where a direct method is either not possible or not convenient. It involves the derivation of a general rule from one or more particular cases, i.e. the general rule is *induced*. This is the opposite to deduction, where you use the general rule to provide detail about a particular case. For example, we know that 60 is divisible by 1, 2, 3, 4, 5 and 6, but does it follow that 60 is divisible by all positive integers?

Example 1.4.1

Consider the pattern:

$$\begin{aligned} 1 &= 1^2 \\ 1+3 &= 2^2 \\ 1+3+5 &= 3^2 \\ 1+3+5+7 &= 4^2 \end{aligned}$$

The pattern shows that the sum of the first two positive odd integers is a perfect square, the sum of the first three positive odd integers is a perfect square and the sum of the first four positive odd integers is a perfect square. Can we then say, based on the first few lines of this pattern, that:

$1 + 3 + 5 + 7 + \ldots + (2n - 1) = n^2$

This can be checked, as the positive odd integers form an arithmetic progression (see 1.1) with $a = 1$ and $d = 2$. The sum of the first n terms is given by:

$$\begin{aligned} S_n &= \frac{n}{2}(a+l) \quad \text{where } a \text{ is the first term, } l \text{ is the last term.} \\ &= \frac{n}{2}(1+2n-1) \\ &= n^2 \end{aligned}$$

In this case the general result was easy to guess, but remember that a guess is not a proof. Thankfully in this example we had a method (sum of an AP) to verify our guess. This will not always be the case.

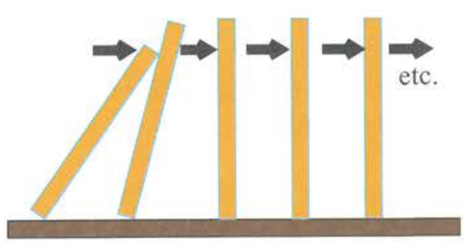
Consider the expression $4n^3 - 18n^2 + 32n - 15$ for values of n from 1 to 4 as shown in the table below:

n	1	2	3	4
$4n^3 - 18n^2 + 32n - 15$	3	9	27	81

The expression appears to produce the successive powers of 3, and so we could assume, based on the results in this table, that $4n^3 - 18n^2 + 32n - 15 = 3^n$.

Can we then say that this will always be the case, and if so, what would you predict the value of the expression to be when $n = 5$? Check to see if your prediction is correct.

Many formulae which we may guess or develop from simple cases can be proved using the **principle of mathematical induction**.



This method of proof relies upon a similar principle to that of 'domino stacking'. In the process of domino stacking, one domino is first pushed over, thus causing a series of dominoes to fall. Before each successive domino will fall, the preceding domino must fall.

With induction, for each expression to be true, the expression before it must also be true. The process can be summarized into four steps:

Step 1: the first expression must be true (the first domino falls)

Step 2: assuming that a general expression is true (*assume that some domino in the series falls*)

Step 3: prove that the next expression is true (*prove that the next domino in the series falls*)

Step 4: if all of these events happen then we know by induction that all of the expressions are true and thus the original formula is true (*all the dominoes will fall*).

Example 1.4.2

Prove by induction that the sum of the first n odd numbers is n^2.

That is, $1 + 3 + 5 + 7 + \dots + (2n - 1) = n^2$ for all $n \geq 1$.

First we need to state what our proposition is. We do this as follows:

Let $P(n)$ be the proposition that $1 + 3 + 5 + 7 + \dots + (2n - 1) = n^2$ for all $n \neq 1$.

Next we proceed with our four steps:

Step 1: test for $n = 1$

LHS $= 1$ AND RHS $= 1^2 = 1$, $\therefore$ LHS = RHS

$\therefore$ the proposition $P(n)$ is true for $n = 1$ (the first domino falls!)

Step 2: assume that $P(n)$ is true for $n = k$ (a general domino falls)

i.e. $1 + 3 + 5 + 7 + \dots + (2k - 1) = k^2$

Step 3: test the proposition for $n = k + 1$ (prove that the next domino falls)

i.e. we wish to prove that $1 + 3 + 5 + \dots$
$\dots + (2k - 1) + \{2(k + 1) - 1\} = (k + 1)^2$

Now, LHS $= 1 + 3 + 5 + \dots + (2k - 1) + (2k + 1)$

$= k^2 + (2k + 1)$ (as $1 + 3 + 5 + \dots + (2k - 1) = k^2$ (from Step 2)

$= (k + 1)^2$

$=$ RHS $\therefore$ $P(n)$ is true for $n = k + 1$

Step 4: Thus, if the proposition is true for $n = k$ (Step 2), then it is true for $n = k + 1$. As it is true for $n = 1$, then it must be true for $n = 1 + 1$ ($n = 2$). As it is true for $n = 2$ then it must hold for $n = 2 + 1$ ($n = 3$) and so on for all positive integers n.

An alternative way of looking at mathematical induction is to think of the problem as a series of assertions. If the first assertion is true, and then each assertion which is true is followed by a true assertion, then all of the assertions in the sequence are true.

Example 1.4.3

Prove that the formula

$$1 + 2 + 3 + \dots + n = \frac{1}{2}n(n + 1)$$

is true for all positive integral n.

Step 1: The formula is actually a series of assertions:

$$n = 1: \quad 1 = \frac{1}{2} \times 1 \times 2$$

$$n = 2: \quad 1 + 2 = \frac{1}{2} \times 2 \times 3$$

$$n = 3: \quad 1 + 2 + 3 = \frac{1}{2} \times 3 \times 4 \text{ etc.}$$

The first assertion is obviously true so we now need to prove that the assertion following each true assertion is itself true.

Step 2: Suppose the k^{th} assertion is true,

i.e. $1 + 2 + 3 + \dots + k = \dfrac{k(k + 1)}{2}$

Step 3: Now add the $(k + 1)^{\text{th}}$ term i.e. $(k + 1)$ to both sides of this equation, obtaining:

$$1 + 2 + 3 + \dots + k + (k + 1) = \frac{k(k + 1)}{2} + (k + 1)$$

$$= \frac{k(k + 1) + 2(k + 1)}{2}$$

$$= \frac{(k + 1)(k + 2)}{2}$$

Step 4: But this is equivalent to the $(k + 1)^{\text{th}}$ assertion, which is true if the k^{th} assertion is true. We have thus shown that the assertion following each true assertion is also true, and thus by mathematical induction the formula given is true for all n.

Exercise 1.4.1

Prove by induction that for all n:

a $\quad 1 + 4 + 7 + \ldots + (3n - 2) = \frac{1}{2}n(3n - 1)$

b $\quad 1 + 5 + 9 + \ldots + (4n - 3) = n(2n - 1)$

c $\quad 2 + 4 + 6 + \ldots + 2n = n(n + 1)$

d $\quad 5 + 10 + 15 + \ldots + 5n = \frac{5}{2}n(n + 1)$

e $\quad 6 + 12 + 18 + \ldots + 6n = 3n(n + 1)$

f $\quad 1 + 2 + 4 + 8 + \ldots + 2^{n-1} = 2n - 1$

g $\quad 1 + r + r^2 + \ldots + r^{n-1} = \frac{1 - r^n}{1 - r}$

h $\quad 1^2 + 3^2 + 5^2 + \ldots + (2n - 1)^2 = \frac{1}{3}n(4n^2 - 1)$

i $\quad 1^2 - 2^2 + 3^2 - \ldots + (-1)^{n-1}n^2 = \frac{1}{2}(-1)^n - 1n(n + 1)$

In this section we consider some propositions involving divisibility and inequalities.

Example 1.4.4

Prove by induction that $9^n - 1$ is divisible by 8 for all $n \geq 1$ where n is an integer.

Let $P(n)$ be the proposition that $9^n - 1|8$ (i.e. $9^n - 1$ is divisible by 8) for all $n \geq 1$.

Step 1: The proposition is true when $n = 1$ since $9^1 - 1 = 8$ which is divisible by 8

Step 2: Assume $P(n)$ holds true for $n = k$, i.e. assume that $9^k - 1 = 8m$ where m is an integer.

Step 3: Prove $P(n)$ is true for $n = k + 1$. i.e. prove that $9^{k+1} - 1$ is divisible by 8.

Now $9^{k+1} - 1 = 9(9^k) - 1$

$= 9(8m + 1) - 1$ (as $9^k = 8m + 1$ (from Step 2))

$= 72m + 8$

$= 8(9m + 1)$ which is divisible by 8

Therefore, $P(n)$ is true for $n = k + 1$.

Step 4: That is, if the proposition holds for $n = k$, it also holds for $n = k + 1$. As it is true for $n = 1$ it is then true for $n = 2$, and so on, and thus the proposition is true for all $n \geq 1$.

Example 1.4.5

Prove by induction that $2^n > n$ for all $n \geq 1$.

Let $P(n)$ be the proposition that $2^n > n$ for all $n \geq 1$.

Step 1: $P(n)$ is true when $n = 1$ since

$$\text{L.H.S} = 2^1 = 2 > 1 = \text{R.H.S}$$

Step 2: Assume that $P(n)$ holds for $n = k$; i.e. that $2^k > k$

Step 3: Prove that $P(n)$ is true for $n = k + 1$

i.e. show that $2^{k+1} > k + 1$.

From Step 2 above, $2^k > k$

$2 \times 2^k > 2k$ (multiplying both sides by 2)

But, $2 \times 2^k = 2^{k+1}$ $\quad \therefore 2^{k+1} > 2k$

Now, $k \geq 1$ so $2k = k + k \geq k + 1$ and hence $2^{k+1} > k + 1$

i.e. $P(n)$ holds for $n = k + 1$ if it holds for $n = k$.

Step 4: Thus as $P(n)$ holds for $n = 1$, it holds for $n = 1 + 1$ and so on for all values of $n \geq 1$.

Exercise 1.4.2

By induction, prove that:

a $\quad 9^{n+2} - 4^n$ is divisible by 5 for all $n \geq 1$

b $\quad n^3 - n$ is divisible by 3 for all $n > 1$

c $\quad n^3 + 2n$ is a multiple of 3 for all $n \geq 1$

d $\quad 7^n + 2$ is divisible by 3 for all $n \geq 1$

e $\quad 9^{n+1} - 8n - 9$ is divisible by 64 for all $n \geq 1$

f $\quad 2^n \geq 1 + n$ for all $n \geq 1$

Extra questions

FURTHER EXAMPLES

We now consider more difficult propositions.

Example 1.4.6

Prove by induction that $n^3 + 5n$ is divisible by 6 for all $n \geq 1$.

Let $P(n)$ be the proposition that $n^3 + 5n$ is divisible by 6 for all $n \geq 1$.

Step 1: Test for $n = 1$

$1^3 + 5 \times 1 = 6$ which is divisible by 6 and so the proposition is true for $n = 1$

Step 2: Let $P(n)$ be true for $n = k$,

i.e. $\frac{k^3 + 5k}{6} = m \Leftrightarrow k^3 + 5k = 6m$, m is an integer.

Step 3: Test for $n = k + 1$

$$\begin{aligned}(k+1)^3 + 5(k+1) &= k^3 + 3k^2 + 3k + 1 + 5k + 5 \\ &= (k^3 + 5k) + 3k^2 + 3k + 6 \\ &= 6m + 3k^2 + 3k + 6 \text{ (from Step 2)} \\ &= 6m + 6 + 3k(k+1)\end{aligned}$$

Now $k(k + 1)$ is an even number and thus it has a factor of 2 (the product of two consecutive integers is even). Thus the product $3k(k + 1)$ can be written as $3 \times 2 \times q$ where q is the quotient of $k(k + 1)$ and 2.

$$\begin{aligned}\therefore \text{ LHS} &= 6m + 6 + 6q \\ &= 6(m + 1 + q) \text{ which is divisible by 6.}\end{aligned}$$

Step 4: Thus, if the proposition is true for $n = k$ then it is true for $n = k + 1$ as proved. As it is true for $n = 1$, then it must be true for $n = 1 + 1$ $(n = 2)$. As it is true for $n = 2$ then it must hold for $n = 2 + 1$ $(n = 3)$ and so on for all positive integers n.

That is, by the principle of mathematical induction $P(n)$, is true.

Example 1.4.7

Prove that: $\sum_{r=1}^{n} (4r - 6) = 2n(n - 2)$ for all $n \in \mathbb{Z}^+$

Let $P(n)$ be the proposition that

$\sum_{r=1}^{n} (4r - 6) = 2n(n - 2)$ for all $n \in \mathbb{Z}^+$.

However, when dealing with sigma notation it can be helpful to write the first few terms of the sequence:

$$\sum_{r=1}^{n} (4r - 6) = -2 + 2 + 6 + 10 + \ldots + (4n - 6) = 2n(n - 2)$$

Step 1: $P(n)$ is true for $n = 1$ since
L.H.S $= 4 \times 1 - 6 = 2 \times (1 - 2) = -2 =$ R.H.S.

Step 2: Assume that $P(n)$ is true for $n = k$,

i.e. $\sum_{r=1}^{k} (4r - 6) = 2k(k - 2)$.

Step 3: Test $P(n)$ for $n = k + 1$:

Adding the $(k + 1)^{\text{th}}$ term, $[4(k + 1) - 6]$ to both sides gives

$$\begin{aligned}\sum_{r=1}^{k} (4r - 6) + [4(k+1) - 6] &= 2k(k - 2) + [4(k + 1) - 6] \quad \text{(from Step 2)} \\ &= 2k^2 - 4k + 4k - 2 \\ &= 2(k^2 - 1) \\ &= 2(k + 1)(k - 1) \\ &= 2(k + 1)[(k + 1) - 2]\end{aligned}$$

which is the $(k + 1)^{\text{th}}$ assertion.

That is, $P(n)$ is true for $n = k + 1$.

Step 4: Thus, if the proposition is true for $n = k$, then it is true for $n = k + 1$. As it is true for $n = 1$, then it must be true for $n = 1 + 1$ $(n = 2)$. As it is true for $n = 2$ then it must hold for $n = 2 + 1$ $(n = 3)$ and so on for all positive integers n.

That is, by the principle of mathematical induction, $P(n)$ is true.

Exercise 1.4.3

Prove the following using the principle of mathematical induction for all $n \in \mathbb{Z}^+$.

a $\quad \displaystyle\sum_{r=1}^{n} \frac{1}{(2r-1)(2r+1)} = \frac{n}{2n+1}$

b $\quad \displaystyle\sum_{r=1}^{n} (2r-1)^3 = n^2(2n^2-1)$

c $\quad \displaystyle\sum_{r=1}^{n} 5^r = \frac{1}{4}(5^{n+1}-5)$

d $\quad 2+2^2+2^3+\ldots+2^n = 2(2^n-1)$

e $\quad 1+3+9+\ldots+3^{n-1} = \frac{1}{2}(3^n-1)$

f $\quad 1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{n-1}} = 2-2^{1-n}$

g $\quad 1.1+2.3+3.5+\ldots+n(2n-1) = \frac{1}{6}n(n+1)(4n-1)$

h $\quad 1.1+3.2+5.4+\ldots+(2n-1)2^{n-1} = 3+2^n(2n-3)$

Extra questions

Mathematics can be considered to be the study of patterns. A useful ability in maths can be forming a rule to describe a pattern. Of course any rule that we develop must be true in all relevant cases and mathematical induction provides one method of proof.

Here are two final examples:

Example 1.4.8
Find the sum to n terms of the series $\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^n}$

Consider $f(n) = \frac{a}{2^n}$ then: $f(n+1) = \frac{a}{2^{n+1}}$.

Now $u_n = f(n+1)-f(n)$, i.e. $\frac{1}{2^n} = \frac{a}{2^{n+1}} - \frac{a}{2^n}$.

Equating and solving for a gives $a = -2$

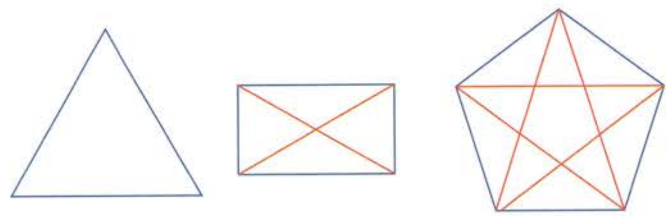
$\therefore f(n+1) = \frac{-2}{2^{n+1}} = -\frac{1}{2^n}$, $f(1) = -1$.

Therefore,

$\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^n} = f(n+1)-f(1) = -\frac{1}{2^n}+1 = 1-\frac{1}{2^n}$.

This result can be proved by induction.

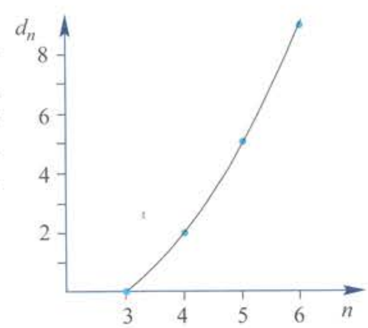
Example 1.4.9
Find the number of diagonals that can be drawn in an n-sided convex polygon.



Let d_n represent the number of diagonals in an n-sided polygon. The value for d_n is shown in the table for values of n up to $n = 6$ (*construct the next few diagrams in the pattern to verify and extend this table*).

n	3	4	5	6
d_n	0	2	5	9

Plotting the points related to the variables n and d_n (*above*) suggests that the relationship between them *could* be quadratic, and so we might assume that



$d_n = an^2+bn+c$.

Substituting the first 3 values for n gives:

$n = 3 \quad \Rightarrow 0 = 9a+3b+c$

$n = 4 \quad \Rightarrow 2 = 16a+4b+c$

$n = 5 \quad \Rightarrow 5 = 25a+5b+c$

Solving these three equations for a, b and c gives

$a = \frac{1}{2}$, $b = -\frac{3}{2}$, and $c = 0$ and thus $d_n = \frac{1}{2}n^2 - \frac{3}{2}n = \frac{n(n-3)}{2}$.

When $n = 6$, $d_6 = \frac{6(6-3)}{2} = 9$, which corresponds to the tabulated value for $n = 6$ above.

So far we have formed a **conjecture** that the number of diagonals in an n-sided convex polygon is given by $d_n = \frac{n(n-3)}{2}$. This formula remains a conjecture until we prove that it is true for values of $n \geq 3$.

Proof:

Let $P(n)$ be the proposition that the number of diagonals that can be drawn in an n-sided convex polygon is given by $d_n = \frac{n(n-3)}{2}$ for $n \geq 3$.

Step 1: $P(n)$ is true for $n = 3$ as $d_3 = \frac{3(3-3)}{2} = 0$ which is the number of diagonals in a 3-sided polygon.

Step 2: Assume that $P(n)$ is true for a k-sided polygon i.e. that $d_k = \frac{k(k-3)}{2}$. We consider the effect that adding an extra side will have on the result.

Step 3: Looking at the tabulated values for n and d_n you should see that adding an extra side to an n-sided polygon produces an extra $(n - 1)$ diagonals, and so we can say that

$$
\begin{aligned}
d_{k+1} &= d_k + \text{the extra diagonals added by the extra side} \\
&= d_k + (k-1) \\
&= \frac{k(k-3)}{2} + (k-1) \\
&= \frac{k(k-3) + 2(k-1)}{2} \\
&= \frac{(k+1)(k-2)}{2} \\
&= \frac{(k+1)[(k+1)-3]}{2}
\end{aligned}
$$

Step 4: Which is the $(k + 1)^{th}$ assertion.

Thus, if the proposition is true for $n = k$, then it is true for $n = k + 1$. As it is true for $n = 3$, then it must be true for $n = 3 + 1$ $(n = 4)$. As it is true for $n = 4$ then it must hold for $n = 4 + 1$ $(n = 5)$ and so on for all integers $n \geq 3$.

By the principle of mathematical induction, $P(n)$ is true.

Exercise 1.4.4

Find the sum to n terms of the sequences below and then prove your results true.

a $\quad 2 + 5 + 10 + 17 + \ldots + (n^2 + 1)$

b $\quad 1 + 8 + 27 + 64 + \ldots + n^3$

c $\quad \frac{1}{5} + \frac{1}{25} + \frac{1}{125} + \ldots + \frac{1}{5^n}$

d $\quad 1^3 + 3^3 + 5^3 + \ldots + (2n-1)^3$

e $\quad 1.3 + 2.4 + 3.5 + \ldots + n(n+2)$

f $\quad \frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \frac{1}{5 \cdot 7} + \ldots + \frac{1}{(2n-1)(2n+1)}$

For questions 2 to 6, find the required general result and then prove your answer using mathematical induction.

2. 1, 3, 6, 10, 15, … are called triangular numbers.

Denoting the nth triangular number as t_n, find a formula for t_n.

3. Find the size of each angle in a regular n-sided polygon.

4. Find the maximum number of pieces that can be formed making n straight cuts across a circular pizza (*pieces don't have to be of equal size*).

5. Find the number of squares of all sizes on an $n \times n$ chess board.

6. Prove that a three digit number is divisible by 3 if the sum of its digits is divisible by 3.

Extra questions

Answers

1.5 Complex Numbers (1)

Introduction

Complex numbers are often first encountered when solving a quadratic equation of the type for which there are no real solutions, e.g. $x^2+1=0$ or $x^2+2x+5=0$ (because for both equations the discriminant, $\Delta = b^2-4ac$, is negative). However, the beginning[1] of complex numbers is to be found in the work of *Girolamo Cardano* (1501–1576), who was resolving a problem which involved the solution to a reduced cubic equation of the form $x^3+ax=b, a>0, b>0$. Although others later improved on the notation and the mechanics of complex algebra, it was the work found in his book, *Ars magna*, that led to the common usage of complex numbers found today.

Notation and $i^2 = -1$

The set of complex numbers is denoted by:

$$C=\{z : z=x+iy, \text{ where } x, y \in \mathbb{R}, i^2=-1\}$$

The complex number, z, is 'made up' of two parts; 'x' and 'iy'. The 'x-term' is called the **real part** and the 'y-term' is the **imaginary part** i.e. the part attached to the 'i', where $i^2=-1$. It is important to note the following:

1. The complex number $z=x+iy$ is a single number (even though there are 'two parts', it is still a single value).

1 See *An Imaginary Tale, The Story of*, by Paul J. Nahim.

2. **The real part** of z, denoted by $Re(z)$ is x.

   The imaginary part of z, denoted by $Im(z)$ is y.

This means that the complex number z can be written as:

$$z=Re(z)+Im(z)i$$

Notice that the imaginary part is not 'iy' but simply 'y'.

Example 1.5.1

For each of the following complex numbers, state the real and imaginary parts of:

a $z=2+3i$ b $w=3-9i$

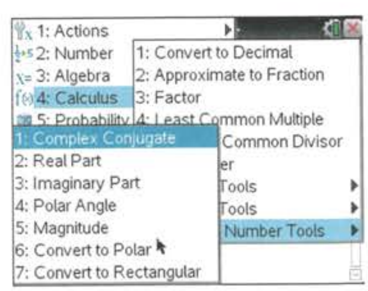
a We have that $Re(z)=Re(2+3i)=2$ and $Im(z)=Im(2+3i)=3$.

Therefore, the real part of z is 2 and the imaginary part of z is 3.

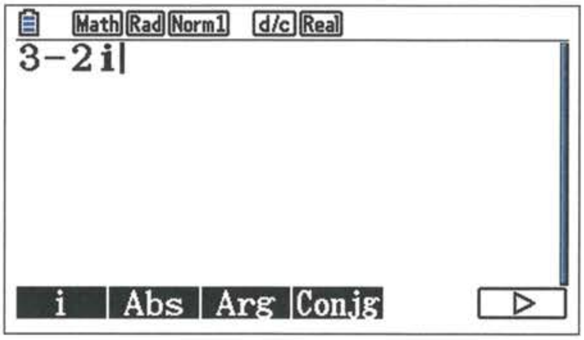
b Similarly, $Re(w)=Re(3-9i)=3$ and $Im(w)=Im(3-9i)=-9$.

That is, for w, the real part is 3 and the imaginary part is -9.

It is important to locate, and become familiar with, the Complex Number part of your calculator (TI).



If using Casio, press the Option key (OPTN) followed by F3 - complex,



Example 1.5.2

If $z = 2xi + y^2 - 1$, find the values of x and y for which $Im(z) = 8$ and $Re(z) = 0$.

We first need to determine what the real and imaginary parts of $z = 2xi + y^2 - 1$ are.

We have that $Im(z) = Im[(2x)i + (y^2 - 1)] = 2x$.

$\therefore Im(z) = 8 \Leftrightarrow 2x = 8 \Leftrightarrow x = 4$.

Similarly, $Re(z) = Re[(2x)i + (y^2 - 1)] = y^2 - 1$

$\therefore Re(z) = 0 \Leftrightarrow y^2 - 1 = 0 \Leftrightarrow y = \pm 1$.

The algebra of complex numbers

Working with 'i'

Since we have that $i = \sqrt{-1}$, then $i^2 = -1$, meaning that $i^3 = i^2 \times i = -1 \times i = -i$.

Similarly, $i^4 = i^2 \times i^2 = -1 \times -1 = 1$, etc.

General results for expressions such as i^n can be determined. We leave this to the set of exercises at the end of this section.

Operations

For any two complex numbers $z_1 = a + ib$ and $z_2 = c + id$, the following hold true:

Equality:

Two complex numbers are equal if and only if their real parts are equal **and** their imaginary parts are equal.

$$z_1 + z_2 = (a+ib) + (c+id) = (a+c) + (b+d)i$$

Addition:

The sum of two (or more) complex numbers is made up of the sum of their real parts plus the sum of their imaginary parts (multiplied by 'i').

$$z_1 + z_2 = (a+ib) = (c+id) = (a+c) + (b+d)i$$

Subtraction:

The difference of two (or more) complex numbers is made up of the difference of their real parts plus the difference of their imaginary parts (multiplied by 'i').

$$z_1 - z_2 = (a+ib) - (c+id) = (a-c) + (b-d)i$$

Multiplication:

When multiplying two (or more) complex numbers, we complete the operation as we would with normal algebra. However, we use the fact that $i^2 = -1$ when simplifying the result.

$$\begin{aligned} z_1 z_2 &= (a+ib)(c+id) \\ &= ac + adi + bci + bdi^2 \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

Conjugate:

The conjugate of $z = x + iy$, denoted by $\bar{z}$ or z^* is the complex number $z^* = x - iy$. Note that:

$$\begin{aligned} zz^* &= (x+iy)(x-iy) \\ &= x^2 - xyi + xyi - y^2 i^2 \\ &= x^2 + y^2 \end{aligned}$$

That is, when a complex number is multiplied with its conjugate, the result is a real number. $z = x + iy$ and $z^* = x - iy$ are known as **conjugate pairs**.

Division:

When dividing two complex numbers, we multiply the numerator and denominator by the conjugate of the denominator (this has the effect of 'realizing' the denominator).

That is,

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{a+ib}{c+id}\\ &= \frac{a+ib}{c+id}\times\frac{c-id}{c-id}\\ &= \left(\frac{ac+bd}{c^2+d^2}\right)+\left(\frac{bc-ad}{c^2+d^2}\right)i\end{aligned}$$

Note: It is important to realise that these results are not meant to be memorised. Rather, you should work through the multiplication or division in question and then simplify the result.

Example 1.5.3

Find the values of x and y if
$z = x+(y-2)i,\ w = 4+i$ and $z = w$.

Recall: Two complex numbers are equal if and only if their corresponding real parts and imaginary parts are equal.

So, $z = w \Leftrightarrow x+(y-2)i = 4+i \Leftrightarrow x = 4$ and $y-2 = 1$.

That is, $z = w$ if and only if $x = 4$ and $y = 3$.

Example 1.5.4

Find the real values x and y given that
$$(3-2i)(x+iy) = 12-5i.$$

As we are equating two complex numbers, we need to determine the simultaneous solution brought about by equating their real parts and imaginary parts:

From $(3-2i)(x+iy) = 12-5i$ we have
$3x+3yi-2xi-2yi^2 = 12-5i$

$\Leftrightarrow (3x+2y)+(3y-2x)i = 12-5i$

$\Leftrightarrow 3x+2y = 12$ – (1) and $3y-2x = -5$ – (2)

Solving simultaneously, we have:

$2\times$(1): $6x+4y = 24$ – (3)

$3\times$(2): $9y-6x = -15$ – (4)

Adding, (3) + (4), we have: $13y = 9$

Therefore, $y = \frac{9}{13}$. Then, substituting into (1) we have:

$3x+2\times\frac{9}{13} = 12 \Leftrightarrow 3x = \frac{138}{13} \Leftrightarrow x = \frac{46}{13}$.

So, we have the solution pair, $x = \frac{46}{13}, y = \frac{9}{13}$.

Example 1.5.5

Given that $z = 3+i$ and $w = 1-2i$, evaluate the following.

a $z+w$ b $2z-3w$

c zw d w^2

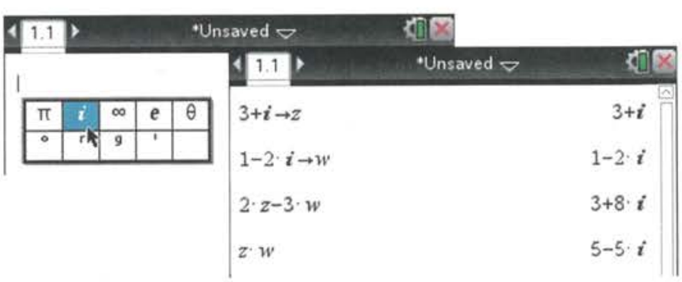
a
$$\begin{aligned}z+w &= (3+i)+(1-2i)\\ &= (3+1)+(i-2i)\\ &= 4-i\end{aligned}$$

b
$$\begin{aligned}2z-3w &= 2(3+i)-3(1-2i)\\ &= (6-3)+(2i+6i)\\ &= 3+8i\end{aligned}$$

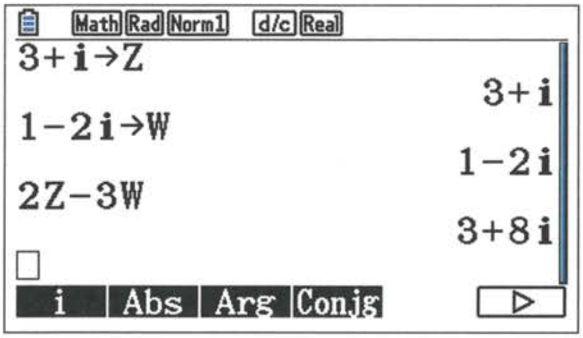
c
$$\begin{aligned}zw &= (3+i)(1-2i)\\ &= 3-6i+i-2i^2\\ &= 3-5i+2\\ &= 5-5i\end{aligned}$$

d
$$\begin{aligned}w^2 &= (1-2i)(1-2i)\\ &= 1-2i-2i+4i^2\\ &= 1-4i-4\\ &= -3-4i\end{aligned}$$

You should be able to perform such calculations both manually and using your calculator. Parts b & c of the previous example are solved as follows (note that you must use the complex number version of 'i', not the variable 'I').



The same calculations can be performed using Casio models:



Example 1.5.6

Find the conjugate of:

a $z = 2 + 6i$ b $w = \sqrt{3}i - 1$

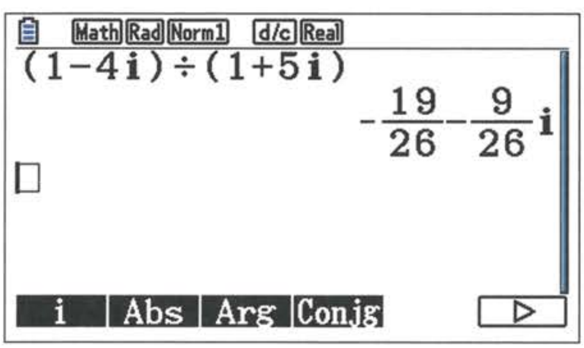
a $z = 2 + 6i \Rightarrow z^* = \overline{2 + 6i} = 2 - 6i$.

b $w = \sqrt{3}i - 1 \Rightarrow w^* = \overline{\sqrt{3}i - 1} = -\sqrt{3}i - 1$.

Example 1.5.7

Express the complex number $\dfrac{1-4i}{1+5i}$ in the form $u + iv$.

$$\begin{aligned}\frac{1-4i}{1+5i} &= \frac{1-4i}{1+5i} \times \frac{1-5i}{1-5i}\\ &= \frac{1-5i-4i+20i^2}{1-5i+5i-25i^2}\\ &= \frac{1-9i-20}{1+25}\\ &= \frac{-19-9i}{26}\\ &= -\frac{19}{26} - \frac{9}{26}i\end{aligned}$$



You may need to use the F↔D key (above 8) to get the answer as a fraction.

Example 1.5.8

If $z = x + iy$, find:

a $z + z^*$ b $z - z^*$

$z + z^* = (x + iy) + (x - iy) = 2x = 2Re(z)$.

Note then, that $Re(z) = \frac{1}{2}(z + z^*)$.

$z - z^* = (x + iy) - (x - iy) = 2yi = 2Im(z)i$

Note then, that $Im(z) = \frac{1}{2i}(z - z^*)$.

Example 1.5.9

If $z = \cos\theta + i\sin\theta$ and $w = \sin\alpha + i\cos\alpha$ express zw in the form $p + qi$, where $p, q \in \mathbb{R}$. Hence find the maximum value of $p^2 + q^2$.

$$\begin{aligned}zw &= (\cos\theta + i\sin\theta)(\sin\alpha + i\cos\alpha)\\ &= \cos\theta\sin\alpha + \cos\theta\cos\alpha i + \sin\theta\sin\alpha i + \sin\theta\cos\alpha i^2\\ &= \cos\theta\sin\alpha + \cos\theta\cos\alpha i + \sin\theta\sin\alpha i - \sin\theta\cos\alpha\\ &= (\cos\theta\sin\alpha - \sin\theta\cos\alpha) + (\cos\theta\cos\alpha + \sin\theta\sin\alpha)i\\ &= \sin(\alpha - \theta) + \cos(\alpha - \theta)i\end{aligned}$$

With $p = \sin(\alpha - \theta)$ and $q = \cos(\alpha - \theta)$
we have $p^2 + q^2 = \sin^2(\alpha - \theta) + \cos^2(\alpha - \theta) = 1$.

As $p^2 + q^2$ will always have a fixed value of 1, its maximum value is also 1.

Exercise 1.5.1

1. Find: a $Re(z)$ b $Im(z)$ c z^* for each of the following.

i $z = 2+2i$ ii $z = -3+\sqrt{2}i$

iii $z = -5i+6$ iv $z = -\frac{2}{5}i$

v $z = \frac{3+i}{2}$ vi $2z = 1-3i-z$

2. If $z = 4-i$ and $w = 3+2i$, find in simplest form (i.e. expressed as $u+iv$), the following.

a $z+w$ b $z-w$ c z^2

d $2z-3w$ e z^*w f iw

3. If $z = 2+i$ and $w = -3+2i$, find in simplest form (i.e. expressed as $u+iv$), the following.

a $z+w$ b $z-w$ c iz^2

d z^2-i^2w e $\bar{z}w$ f $\overline{iw}$

4. For the complex numbers $z = 1-i$ and $w = 2i-3$, express each of the following in the form $u+iv$.

a $\frac{1}{z}$ b $\frac{w}{z}$ c $\frac{z+1}{i}$

d z^{-2} e $\frac{2i}{w+3}$ f $\frac{z^*}{w^*}$

5. Simplify the following.

a $(2+4i)(3-2i)$ b $(1-i)^3$

c $(1+\sqrt{2}i)^2 i$ d $\frac{i}{1+2i}$

e $\frac{1+2i}{i}$ f $\frac{(1-i)i}{(-i+2)}$

6. Given that $z = 3+\sqrt{2}i$ and $w = \frac{1}{1-i}$, find:

a $Re(w)$ b $Im(zw)$ c $Re\left(\frac{z}{w}\right)$

7. Find the real numbers x and y such that:

a $2x+3i = 8-6yi$

b $x+iy = (2+3i)^2$

c $(x+iy)(-i) = 5$

8. **a** Simplify i^n for:

i $n = 0, 1, 2, 3, 4, 5$

ii $n = -1, -2, -3, -4, -5$

Evaluate :

i i^{10} ii i^{15}

iii i^{90} iv i^{74}

9. Find the real numbers x and y, for which $(x+yi)(5-2i) = -18+15i$.

10. Show that for any complex numbers $z = x+iy$ and $w = a+bi$:

a $(z+w)^* = z^*+w^*$ b $(z-w)^* = z^*-w^*$

c $(zw)^* = z^*w^*$ d $(z^2)^* = (z^*)^2$

e $\left(\frac{z}{w}\right)^* = \frac{z^*}{w^*}$ f $(z^*)^* = z$

11.a Prove that $z\bar{w}-\bar{z}w$ is purely imaginary or zero for all complex numbers z and w.

b Prove that $z\bar{w}+\bar{z}w$ is real for all complex numbers z and w.

12. Given that $w = \frac{z-1}{\bar{z}+1}$, where $z = x+iy$,

find the condition(s) under which:

a w is real b w is purely imaginary.

13. a Find the real values of x and y, such that $(x+iy)^2 = 8-6i$.

b Hence, determine $\sqrt{8-6i}$, expressing your answer in the form $u+iv$, where u and v are both real numbers and $u > 0$. Find $\sqrt{3-4i}$, expressing your answer in the form $u+iv$, where u and v are both real numbers and $u > 0$.

14. Simplify the following.

a $(1+i)^3-(1-i)^3$

b $(1+i)^3+(1-i)^3$

c $\frac{(1+i)^3}{(1-i)^3}$

15. Find the real values x and y for which:

a $(x-y)+4i = 9+yi$

b $(2x+3y)-x^3 i = 12-64i$.

16. Find the complex number z given that:

$5z+2i = 5+2iz$,

giving your answer in the form $a + ib$, where a and b are real.

17. Find the complex number z which satisfies the equation $z(1+\sqrt{2}i) = 1-\sqrt{2}i$.

18. The complex number z satisfies the equation $z^2 - i = 2z-1$. If $z = u+iv$ find all real values of u and v.

19. If $z = \frac{2-i}{1+i}$, find:

a $Re(z^2)+Im(z^2)$

b $Re\left(z+\frac{1}{z}\right)+Im\left(z+\frac{1}{z}\right)$

20. a Show that:

i $\frac{1+i}{1-i} = i$

ii $\left(\frac{1+i}{1-i}\right)^2 = -1$

b Show that $\left(\frac{1+i}{1-i}\right)^{4k} = 1$ if k is a positive integer.

21. Find the complex number(s) $z = a+bi$,

satisfying the equation $\frac{1+z^2}{1-z^2} = i$.

22. Express the following in the form $p+qi$, where p and q are real numbers.

a $(\cos\theta + i\sin\theta)(\cos\alpha + i\sin\alpha)$

b $(\cos\theta + i\sin\theta)(\cos\alpha - i\sin\alpha)$

c $(r_1\cos\theta + ir_1\sin\theta)(r_2\cos\alpha + ir_2\sin\alpha)$

d $(x-\cos\theta - i\sin\theta)(x-\cos\theta + i\sin\theta)$

e $(x+\sin\alpha + i\cos\alpha)(x+\sin\alpha - i\cos\alpha)$

23. For the complex number defined as $z = \cos(\theta)+i\sin(\theta)$, show that:

a $z^2 = \cos(2\theta)+i\sin(2\theta)$

b $z^3 = \cos(3\theta)+i\sin(3\theta)$

Assuming now that $z^k = \cos(k\theta)+i\sin(k\theta)$, show that:

c $C+i(S-1) = \frac{1-z^n}{1-z}$,

where $C = 1+\cos(\theta)+\cos(2\theta)+\ldots+\cos((n-1)\theta)$

and

$S = 1+\sin(\theta)+\sin(2\theta)+\ldots+\sin((n-1)\theta)$,

where $0<\theta<\frac{\pi}{2}$.

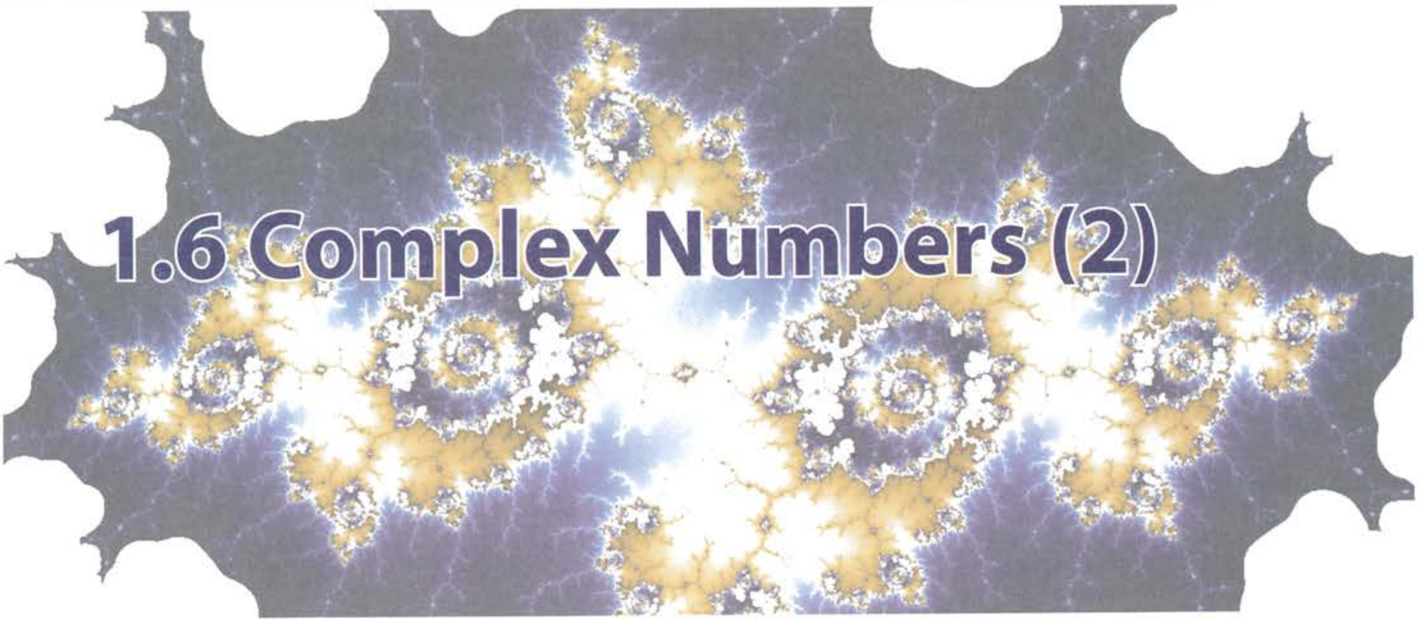
24. a Given that $(x+iy)^2 = 8+6i$, find the values of x and y. Hence, find $\sqrt{8+6i}$.

b If $(2+3i)(3-4i) = p+qi$, find the value of p^2+q^2.

c If $(x+iy)^2 = a+ib$, find an expression for a^2+b^2 in terms of x and y.

Extra questions

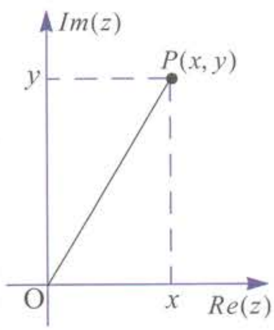
Answers



1.6 Complex Numbers (2)

The Argand Diagram

Unlike real numbers (which can be described geometrically by the position they occupy on a one dimensional number line), complex numbers require the real and imaginary parts to be described. The geometrical representation best suited for this purpose would be two- dimensional. Any complex number $z = x + iy$ may be represented on an **Argand Diagram**, by using either



1. the point $P(x, y)$, or
2. the position vector $\overrightarrow{OP}$

That is, we make use of a plane that is similar to the standard Cartesian plane to represent the complex number $z = x + iy$. This means that the x-axis represents the $Re(z)$ value and the y-axis represents the $Im(z)$ value.

The complex plane has led to the Mandelbrot Set (heading picture by Binette228) and models of tree branching and other elaborate natural forms.

Example 1.6.1

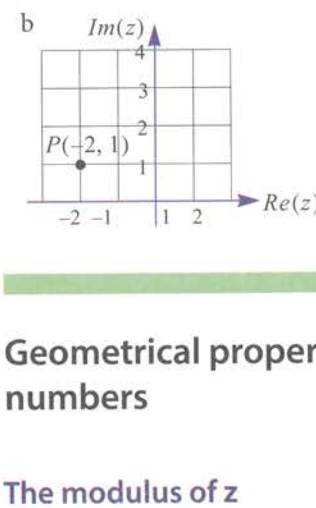
Represent each of the following complex numbers on an Argand diagram.

a $z = 1 + 3i$ b $z = -2 + i$ c $z = -2i$.

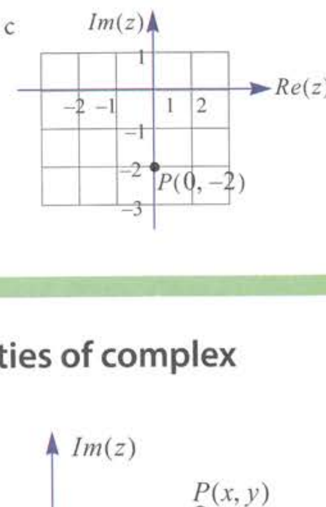
a With $z = 1 + 3i$, we have $x = Re(z) = 1$ and $y = Im(z) = 3$. Therefore, we may represent the complex number $z = 1 + 3i$ by the point $P(1,3)$ on the Argand diagram:

Similarly for parts b and c we have:

b



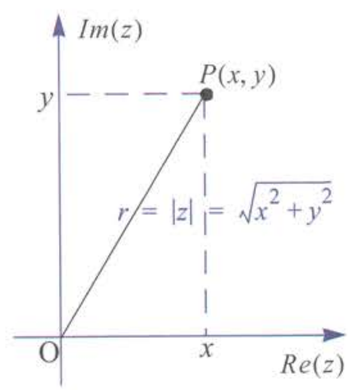
c



Geometrical properties of complex numbers

The modulus of z

The modulus of a complex number $z = x + iy$ is a measure of the length of $z = x + iy$ and is denoted by $|z|$. That is, $mod(z) = |z|$.



The **modulus** of z is also called the **magnitude** of z. We can determine the length by using Pythagoras's theorem:

$$(OP)^2 = x^2 + y^2$$
$$\therefore OP = \sqrt{x^2 + y^2}$$

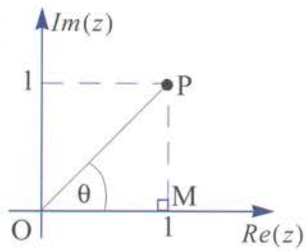
That is, if $x = x + iy$ then $mod(z) = |z| = \sqrt{x^2 + y^2}$.

The modulus of z is also written as r, i.e. $r = |z|$.

Notice then, that $|z| = \sqrt{x^2 + y^2} = \sqrt{zz^*}$

The Argument of z

The **argument** of a complex number $z = x + iy$ is a measure of the angle which $z = x + iy$ makes with the **positive** $Re(z)$ -axis and is denoted by $\arg(z)$ and sometimes by $ph(z)$, which stands for the **phaze** of z. If θ is this angle, we then write, $\theta = \arg(z)$.



If $-\pi < \theta < \pi$, then $\theta = Arg(z)$

Notice the use of capital 'A' rather than lower case 'a'. Using $\theta = Arg(z)$, implies that we are referring to the **Principal argument value**, that is, we have restricted the range in which the angle θ lies.

Example 1.6.2

Find the modulus of the following complex numbers.

a $\quad z = 4 + 3i \quad$ b $\quad z = -1 + 2i$.

a $\quad z = 4 + 3i \therefore |z| = \sqrt{(4)^2 + (3)^2} = \sqrt{25} = 5$

Notice that we only square the real and imaginary parts of the complex number. That is, we do not use $3i$ because this would give $\sqrt{(4)^2 + (3i)^2} = \sqrt{16 - 9} = \sqrt{7}$!

b $\quad$ In the same way we have:

$z = -1 + 2i \therefore |z| = \sqrt{(-1)^2 + (2)^2} = \sqrt{5}$.

Example 1.6.3

Find the principal argument of the following complex numbers.

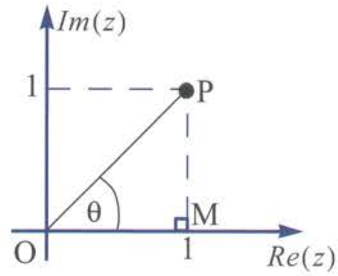
a $z = 1 + i \quad$ b $z = -1 + 2i$. $\quad$ c $-1 - \sqrt{3}i$

When finding the principal argument of a complex number, an Argand diagram can be used as an aid. This will always enable us to work with right-angled triangles. Then we can make use of the diagram to find the restrictions on the required angle, i.e. $-\pi < \theta \le \pi$, then $\theta = Arg(z)$.

a $\quad$ We first represent $z = 1 + i$ on an Argand diagram:

From the triangle OPM, we have:

$$\tan\theta = \frac{PM}{OM} = \frac{1}{1}$$
$$\therefore \theta = Tan^{-1}(1)$$
$$\Rightarrow \theta = \frac{\pi}{4} \text{ (or } 45°)$$

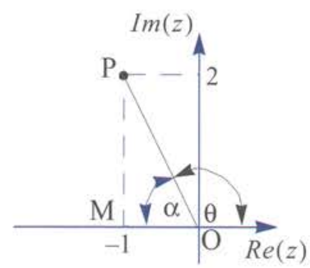


Therefore, the principal argument of z, is $Arg(z) = \frac{\pi}{4}$.

b $\quad$ Again, we start by using an Argand diagram:

From the triangle OPM, we have:

$$\tan\alpha = \frac{PM}{OM} = \frac{2}{1}$$
$$\therefore \alpha = Tan^{-1}(2)$$
$$\Rightarrow \alpha = 63°26'$$



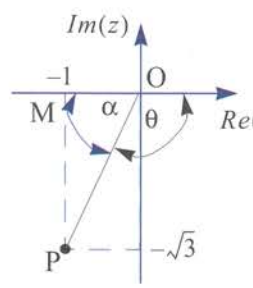
Therefore, $\theta = 180 - 63°26' = 116°34'$.

So that (the principal argument) $Arg(z) = 116°34'$.

c $\quad$ Notice that we only make use of α to help us determine θ [i.e. $\alpha + \theta = \pi$ (or $180°$)]

From the triangle OPM, we have:

$$\tan\alpha = \frac{PM}{OM} = \frac{\sqrt{3}}{1}$$
$$\therefore \alpha = Tan^{-1}(\sqrt{3})$$
$$\Rightarrow \alpha = 60°$$



Therefore, $\theta = 180 - 60° = 120°$.

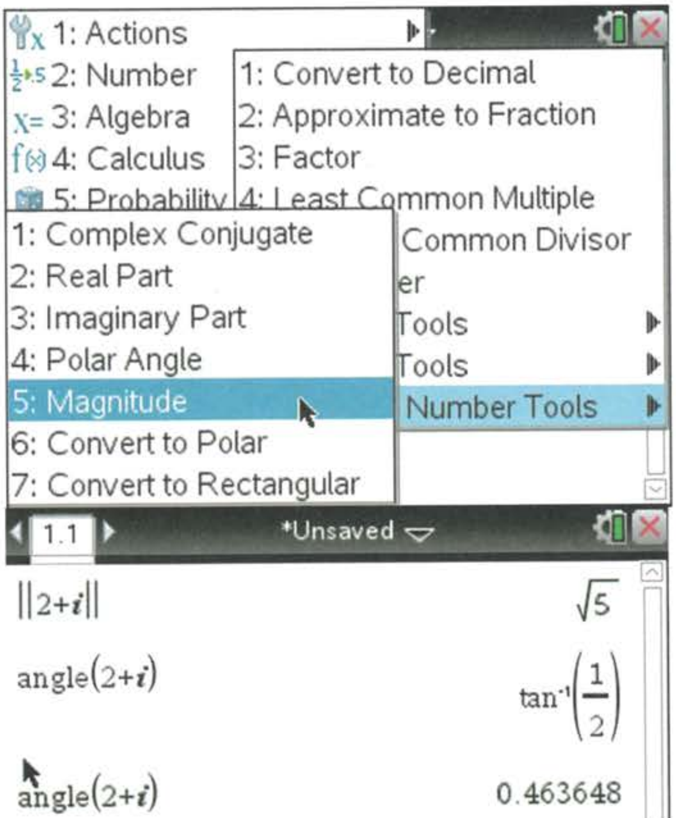
So that (the principal argument) $Arg(z) = -120°$.

Notice that because we are 'moving' in a clockwise direction, the angle is negative.

Notice that in the last example, although $Arg(z) = -120°$, we could have written $\arg(z) = 180° + 60° = 240°$ (using 'small' 'a').

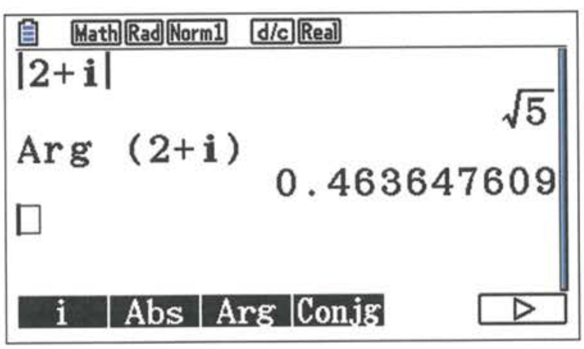
Using a calculator

On TI models, remember to use Menu 2, 9 to access the complex number capabilities.



Use run mode if using Casio.

Press the Option key (OPTN) followed by F3 - complex,



Example 1.6.4

If $z = 1 + 2i$ and $w = x - i$, find:

a $|z + 4|$ b $|z + w|$

a First, we need to determine the complex number $z + 4$

$z + 4 = (1 + 2i) + 4 = 5 + 2i$.

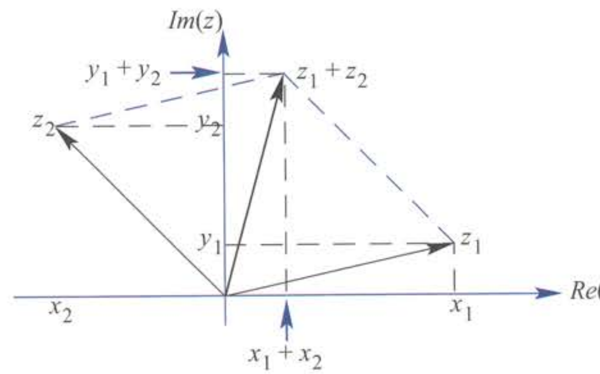
Then we have, $|5 + 2i| = \sqrt{25 + 4} = \sqrt{29}$

b First, we need to determine the complex number $z + w$:

$$
\begin{aligned}
z + w &= (1 + 2i) + (x - i) \\
&= (x + 1) + i \\
&= |(x + 1) + i| \\
&= \sqrt{(x + 1)^2 + 1} \\
&= \sqrt{x^2 + 2x + 2}
\end{aligned}
$$

Adding complex numbers - geometric representation

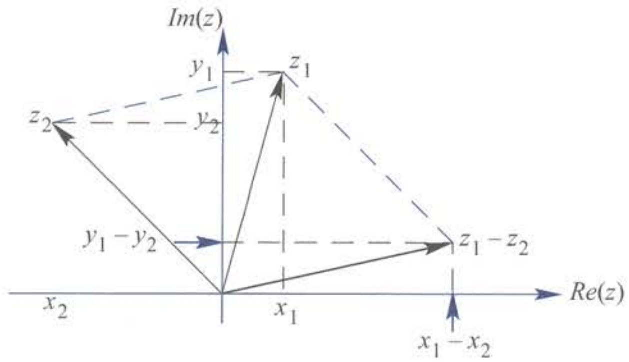
The addition of two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ can be considered in the same way as the addition of two vectors. That is, if $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are represented by directed line segments from the origin $0 + 0i$ their sum, $(x_1 + x_2) + (y_1 + y_2)i$ can also be represented by a directed line segment from the origin $0 + 0i$.



e.g. if $z_1 = 6 + 2i$ and $z_2 = -4 + 4i$ then $z_1 + z_2 = 2 + 6i$

Subtracting complex numbers - geometric representation

Subtracting two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ can be considered in the same way as subtracting two vectors. That is, if $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are represented by directed line segments from the origin $0 + 0i$. Subtracting z_2 from z_1, i.e. $z_1 - z_2$ we obtain $(x_1 - x_2) + (y_1 - y_2)i$ which can also be represented by a directed line segment from the origin $0 + 0i$.



e.g. if $z_1 = 2+6i$ and $z_2 = -4+4i$ then $z_1 - z_2 = 6+2i$

The similarities between complex numbers and vectors in two dimensions (see Chapter 4) make much of the theory interchangeable. Often, complex numbers are represented by the same notation as used in vector theory. For example, if the point P on the Argand diagram represents the complex number $z = 2+3i$ then the vector $\overrightarrow{OP} = [2, 3]$ would represent the same point. However, at this stage we will concentrate on features that deal directly with the complex numbers field.

Exercise 1.6.1

1. Show the following complex numbers on an Argand diagram:

   a $2+i$ b $-6i$

   c $4-3i$ d $2(1-i)$

   e $-3(1-i)$ f $(1+2i)^2$

2. a For the complex number $z = 1+i$, represent the following on an Argand diagram:

   i zi ii zi^2 iii zi^3

   iv zi^4

   b What is the geometrical effect of multiplying a complex number by i?

   i z^* ii $z+z^*$ iii $z-z^*$

   Describe the geometrical significance of each of the operations in part b.

3. If $z_1 = 1+2i$ and $z_2 = 1+i$, show each of the following on an Argand diagram:

   a z_1^2 b $\dfrac{1}{z_2}$ c $z_1 z_2$

   d $2z_1 - z_2$ e $\overline{z_1 z_2}$ f $\overline{z_1} + \overline{z_2}$

4. Find the modulus and argument of:

   a $1+\sqrt{3}i$ b $1-\sqrt{3}i$ c $1+\sqrt{2}i$

5. Consider the two complex numbers $z = a+bi$ and $w = -a+bi$.

   Find $|z|, |w|, |zw|$.

   Find: i $Arg(z+w)$ ii $Arg(z-w)$.

6. If $z = (x-3)+i(x+3)$, find: a $|z|$ b $\{x:|z| = 6\}$

7. If $z = 2+i$ and $w = -1-i$, verify the following.

   a $|z|^2 = zz^*$ b $|zw| = |z||w|$

   c $|w^3| = |w|^3$ d $|z+w| \le |z|+|w|$

   e $Arg(zw) = Arg(z) + Arg(w)$

   What is the geometrical significance of part d?

8. If $w = \dfrac{z+1}{z-1}$ and $|z| = 1$, find $Re(w)$.

9. Given that $|w| = 5$, find

   a $|-3w|$ b $|\overline{w}|$ c $|2iw|$.

10. If $\text{Arg}(z) = 0$, show that z is real and positive.

11. A complex number w is such that w is purely imaginary.

    Show that $\text{Arg}(w) = \pm\dfrac{\pi}{2}$.

12. a If $\arg(z) = \dfrac{\pi}{6}$ and $z = x+iy$, show that $\sqrt{3}y = x$.

    b Find z if $|z-1| = 1$ and $\arg(z-i) = 0$.

13. a If the complex number z satisfies the equations:

    $\arg(z+1) = \dfrac{\pi}{6}$ and $\arg(z-1) = \dfrac{2\pi}{3}$,

    show that $z = \dfrac{1}{2}(1+\sqrt{3}i)$.

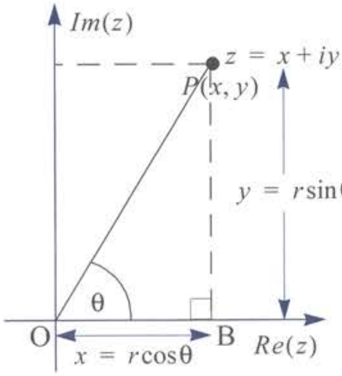
    b If w and z are two complex numbers such that $|z-w| = |z+w|$,

    show that $|\arg(z) - \arg(w)| = \dfrac{\pi}{2}$ or $\dfrac{3\pi}{2}$.

Extra questions

Polar Form

So far we have been dealing with complex numbers of the form $z = x + iy$, where x and y are real numbers. Such a representation of a complex number is known as a **rectangular** representation.



However, the position of a complex number on an Argand diagram has also been described by its magnitude (i.e. its modulus) and the angle which it makes with the positive $Re(z)$ -axis. When we represent a complex number by making use of its modulus and argument, we say that the complex number is in **polar form**.

To **convert from the rectangular form to the polar form**, we make the following observations: From triangle OBP, we have:

1. $\sin(\theta) = \dfrac{BP}{OP} = \dfrac{y}{r} \Rightarrow y = r\sin(\theta)$
2. $\cos(\theta) = \dfrac{OB}{OP} = \dfrac{x}{r} \Rightarrow x = r\cos(\theta)$

Therefore, we can rewrite the complex number z as follows:

$z = x + iy = r\cos(\theta) + ir\sin(\theta)$

$= r(\cos\theta + i\sin\theta)$ - we say that **z is in polar form.**

Often, we abbreviate the expression $z = r(\cos\theta + i\sin\theta)$ to:

$$z = \underbrace{\mathbf{r}}_{r}(\underbrace{\mathbf{c}}_{c}\text{os}\theta + \underbrace{\boldsymbol{is}}_{is}\text{in}\underbrace{\mathbf{q}}_{\theta}) = \boldsymbol{rcis}(\theta)$$

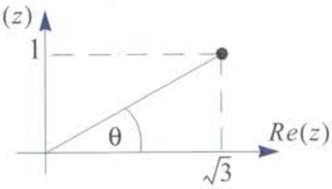
Example 1.6.5

Express the following complex numbers in polar form.

a $\quad z = \sqrt{3} + i \qquad$ b $\quad z = -1 - i$

a When converting from rectangular to polar form, the angle θ refers to the Principal argument.

It is advisable to draw a diagram when converting from rectangular to polar form.



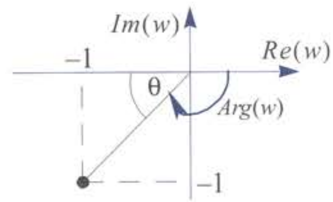
Step 1 $\tan\theta = \dfrac{1}{\sqrt{3}} \Rightarrow \theta = \dfrac{\pi}{6}$ **Step 2** $r = \sqrt{1^2 + (\sqrt{3})^3} = 2$

Therefore, $z = \sqrt{3} + i = 2\left(\cos\dfrac{\pi}{6} + i\sin\dfrac{\pi}{6}\right) = 2cis\left(\dfrac{\pi}{6}\right)$

b

Step 1 $\tan\theta = \dfrac{1}{1} \Rightarrow \theta = \dfrac{\pi}{4}$

Step 2 $r = \sqrt{1^2 + (1)^2} = \sqrt{2}$



$z = -1 - i = \sqrt{2}\left(\cos\left(-\dfrac{3\pi}{4}\right) + i\sin\left(-\dfrac{3\pi}{4}\right)\right) = \sqrt{2}cis\left(-\dfrac{3\pi}{4}\right)$

Example 1.6.6

Convert $\sqrt{2}cis\left(\dfrac{3\pi}{4}\right)$ to Cartesian form.

Let $z = \sqrt{2}cis\left(\dfrac{3\pi}{4}\right)$. Therefore, we have:

$z = \sqrt{2}\left(\cos\left(\dfrac{3\pi}{4}\right) + i\sin\left(\dfrac{3\pi}{4}\right)\right)$ ('expanding' cis-term)

$= \sqrt{2}\left(-\dfrac{1}{\sqrt{2}} + \dfrac{1}{\sqrt{2}}i\right)$

$= -1 + i$

Exercise 1.6.2

1. Express each of the following complex numbers in polar form.

   a $\quad 1 + i \qquad$ b $\quad -1 + i \qquad$ c $\quad -1 - i$

2. Express each of the following complex numbers in polar form.

   a $\quad 2 + 2i \qquad$ b $\quad \sqrt{3} + i \qquad$ c $\quad 4 - 4i$

d $3+4i$ e $-2+i$ f $-2-3i$

g $-\sqrt{3}+i$ h $\frac{1}{2}-\frac{\sqrt{3}}{2}i$ i $3-i$

3. Express each of the following complex numbers in Cartesian form.

a $2cis\left(\frac{\pi}{2}\right)$ b $3cis\left(\frac{\pi}{6}\right)$

c $\sqrt{2}cis\left(-\frac{\pi}{4}\right)$ d $5cis\left(\frac{3\pi}{2}\right)$

e $-8cis\left(-\frac{\pi}{3}\right)$ f $\frac{\sqrt{2}}{3}cis\left(\frac{7\pi}{3}\right)$

4. Simplify the following.

a $\frac{|2+i|}{|1-\sqrt{2}i|}$ b $\frac{zz^*}{|z|^2}$

c $Arg(z)+Arg(z^*)$

5. If $z = \sqrt{2}cis\left(\frac{\pi}{4}\right)$ and $w = 1+\sqrt{3}i$,

find the following, giving your answer in the form $u+iv$.

a w^* b z^* c wz

6. a If $z = x+iy$, show that $z+\frac{|z|^2}{z} = 2Re(z)$.

b If $z = x+iy$, show that:

i $|z| = |\bar{z}|$ ii $z\bar{z} = |z|^2$

7. If $z = 1+i$ and $w = -1+\sqrt{3}i$, find:

a $|z|$ b $|w|$ c $|zw|$

d $Arg(z)$ e $Arg(w)$ f $Arg(zw)$

Answers

Theory of Knowledge

Here are six vultures...

... and six penguins (look carefully, there are three babies - they are not fluffy slippers).

The animals are 'real' but six is an imaginary concept. Discuss!

1.7 Complex Numbers (3)

de Moivre's Theorem

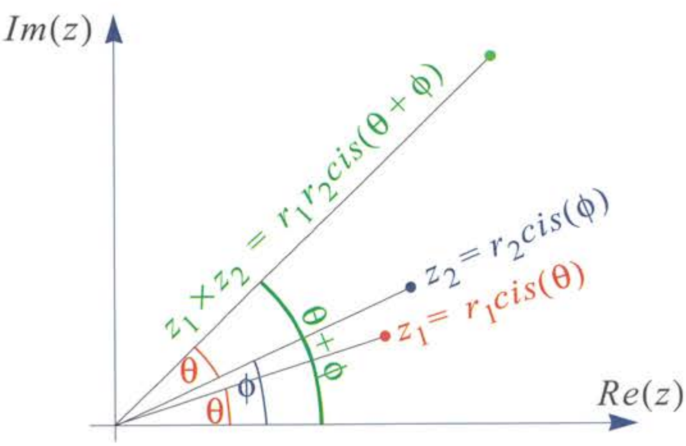
We begin with an important result:

When complex numbers are expressed in polar form, their product can be found by:

1. Multiplying their moduli.
2. Adding their arguments.

Algebraically, this is: If $z_1 = r_1 cis(\theta)$ and $z_2 = r_2 cis(\phi)$, then $z_1 \times z_2 = r_1 r_2 cis(\theta + \phi)$.

Graphically, this becomes:



The powers of a complex number are a special case of this property.

Next, if $z_1 = z_2 = z = rcis(\theta)$, we then have that:

$z^2 = z \times z = rcis(\theta) \times rcis(\theta) = r^2 cis(\theta + \theta)$.

That is, $z^2 = r^2 cis(2\theta)$.

and: $z^3 = z \times z^2 = rcis(\theta) \times r^2 cis(2\theta) = r^3 cis(\theta + 2\theta)$.

That is, $z^3 = r^3 cis(3\theta)$.

In general then, we have that $z^n = r^n cis(n\theta)$.

de Moivre's Theorem states:

$$\boxed{\left(r(\cos\theta + i\sin\theta)\right)^n = r^n\left((\cos n\theta + i\sin n\theta)\right)}$$

Proof: (By mathematical induction)

Let $P(n)$ be the proposition that $(rcis(\theta))^n = r^n cis(n\theta)$.

For $n = 1$, we have that

L.H.S $= (rcis(\theta))^1 = rcis(\theta) = r^1 cis(1 \times \theta) =$ R.H.S

Therefore, $P(n)$ is true for $n = 1$.

Assume now that $P(n)$ is true for $n = k$,

that is, $(rcis(\theta))^k = r^k cis(k\theta)$.

Then, for $n = k + 1$, we have

$$\begin{aligned}(rcis(\theta))^{k+1} &= (rcis(\theta))^k (rcis(\theta)) \\ &= r^k cis(k\theta)(rcis(\theta)) \\ &= r^{k+1} cis(k\theta) cis(\theta) \\ &= r^{k+1} cis(k\theta + \theta) \\ &= r^{k+1} cis((k+1)\theta)\end{aligned}$$

Therefore, we have that $P(k+1)$ is true whenever $P(k)$ is true. Therefore, as $P(1)$ is true, by the Principle of Mathematical Induction, $P(n)$ is true for $n = 1, 2, 3, \ldots$

Note that the case $n = 0$ is the trivial case.

Notice that de Moivre's Theorem holds for all integral values of n, both positive and negative, i.e. $n \in \mathbb{Z} \cup \{0\}$ as well as rational values of n, *i.e.* $n \in \mathbb{Q}$.

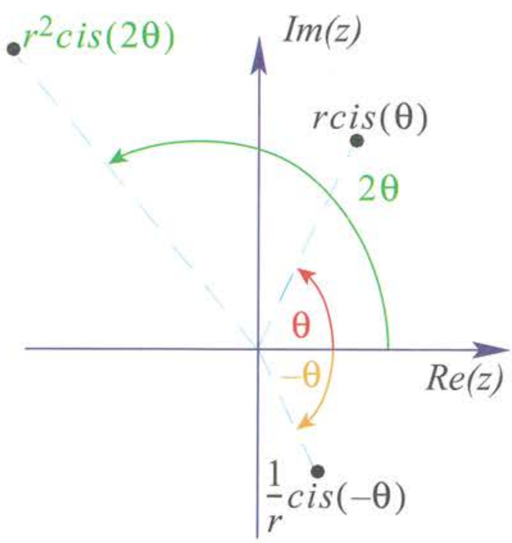
Graphical properties of de Moivre's Theorem

For the complex number $z = rcis(\theta)$, we have

1. $z^{-1} = (rcis(\theta))^{-1} = \frac{1}{r}cis(-\theta)$

   i.e. $|z^{-1}| = |z|^{-1} = \frac{1}{r}$ and $\arg(z^{-1}) = -\theta$.

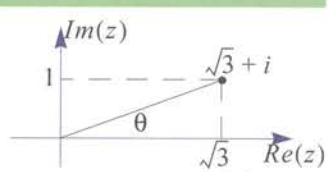
2. $z^2 = (rcis(\theta))^2 = r^2cis(2\theta)$



Example 1.7.1

Find $(\sqrt{3}+i)^5$ using de Moivre's Theorem.

Let $z = \sqrt{3}+i$.



This means:

$r = |\sqrt{3}+i| = \sqrt{(\sqrt{3})^2+1^2} = 2$ and $\theta = Tan^{-1}\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{6}$.

Therefore, we have that $z = \sqrt{3}+i = 2cis\left(\frac{\pi}{6}\right)$.

Using de Moivre's Theorem, we have:

$$(\sqrt{3}+i)^5 = 2^5cis\left(\frac{5\pi}{6}\right) = 32cis\left(\frac{5\pi}{6}\right)$$

$$= 32\left(\cos\left(\frac{5\pi}{6}\right) + i\sin\left(\frac{5\pi}{6}\right)\right)$$

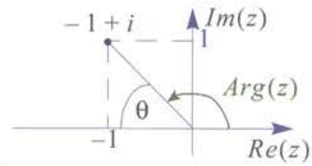
$$= 32\left(-\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)$$

$$= -16\sqrt{3} + 16i$$

Example 1.7.2

Find $(-1+i)^{-4}$ using de Moivre's Theorem.

Let $z = -1+i$.



This means:

$r = |-1+i| = \sqrt{(-1)^2+1^2} = \sqrt{2}$
and

$\theta = Tan^{-1}\left(\frac{1}{1}\right) = \frac{\pi}{4} \therefore Arg(z) = \frac{3\pi}{4}$.

Therefore, we have that $z = -1+i = \sqrt{2}cis\left(\frac{3\pi}{4}\right)$.

Using de Moivre's Theorem, we have

$$(-1+i)^{-4} = (\sqrt{2})^{-4}cis\left(-4\times\frac{3\pi}{4}\right) = \frac{1}{(\sqrt{2})^4}cis(-3\pi)$$

$$= \frac{1}{4}(\cos(-3\pi) + i\sin(-3\pi)) \quad = \frac{1}{4}(-1+0i) \quad = -\frac{1}{4}$$

Example 1.7.3

Express $\frac{1+i}{(1-i)^3}$ in polar form.

We first convert both numerator and denominator into polar form.

$$1+i = \sqrt{2}cis\left(\frac{\pi}{4}\right) \text{ [standard result]}$$

and $1-i = \sqrt{2}cis\left(-\frac{\pi}{4}\right) \therefore (1-i)^3 = (\sqrt{2})^3cis\left(-\frac{3\pi}{4}\right)$

Therefore,

$$\frac{1+i}{(1-i)^3} = \frac{\sqrt{2}cis\left(\frac{\pi}{4}\right)}{2\sqrt{2}cis\left(-\frac{3\pi}{4}\right)} = \frac{1}{2}cis\left[\left(\frac{\pi}{4}\right) - \left(-\frac{3\pi}{4}\right)\right] = \frac{1}{2}cis(\pi)$$

$= \frac{1}{2}\cos\pi + \frac{1}{2}i\sin\pi$.

Example 1.7.4

Simplify:

a $(1+i)^5 + (1-i)^5$ b $(1+i)^5(1-i)^5$

a We first convert each term into its polar form:

$$1+i = \sqrt{2}cis\left(\frac{\pi}{4}\right)$$

$$(1+i)^5 = (\sqrt{2})^5 cis\left(\frac{5\pi}{4}\right)$$

$$= 4\sqrt{2}cis\left(\frac{5\pi}{4}\right)$$

$$1-i = \sqrt{2}cis\left(-\frac{\pi}{4}\right)$$

$$(1-i)^5 = (\sqrt{2})^5 cis\left(-\frac{5\pi}{4}\right)$$

$$= 4\sqrt{2}cis\left(-\frac{5\pi}{4}\right)$$

It follows that:

$$= 4\sqrt{2}\left[\left(-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}i\right)+\left(-\frac{1}{\sqrt{2}}-\frac{1}{\sqrt{2}}i\right)\right]$$

$$= 4\sqrt{2}\left[-\frac{2}{\sqrt{2}}\right]$$

$$= -8$$

b Using the previous results we have,

$$(1+i)^5(1-i)^5 = 4\sqrt{2}cis\left(\frac{5\pi}{4}\right)\times 4\sqrt{2}cis\left(-\frac{5\pi}{4}\right)$$

$$= 32cis\left(\frac{5\pi}{4}-\frac{5\pi}{4}\right)$$

$$= 32cis(0)$$

$$= 32$$

Notice that whenever we add or multiply the complex numbers $rcis(\theta)$ and $rcis(-\theta)$, a purely real complex number will always result. This can seen as follows:

1. Adding

$$\begin{aligned} rcis(\theta) + rcis(-\theta) &= r[cis(\theta) + cis(-\theta)] \\ &= r[(\cos\theta + i\sin\theta) + (\cos(-\theta) + i\sin(-\theta))] \\ &= r[(\cos\theta + i\sin\theta) + (\cos\theta - i\sin\theta)] \\ &= r[2\cos\theta] \\ &= 2r\cos\theta \end{aligned}$$

2. Multiplying

$$\begin{aligned} rcis(\theta)\times rcis(-\theta) &= r^2[cis(\theta)\times cis(-\theta)] \\ &= r^2[cis(\theta-\theta)] \\ &= r^2 cis(0) \\ &= r^2 \end{aligned}$$

Exercise 1.7.1

1. Express each of the following in the form $x+iy$.

a $(1+i)^5$ b $(-1+i)^4$

c $(2+2i)^3$ d $(-\sqrt{3}+i)^4$

e $(\sqrt{3}-i)^5$ f $(3-4i)^3$

2. Express each of the following in the form $x+iy$.

a $(1+i)^{-5}$ b $(-1+i)^{-4}$

c $(2+2i)^{-3}$ d $(-\sqrt{3}+i)^{-4}$

e $(\sqrt{3}-i)^{-5}$ f $(3-4i)^{-3}$

3. Express each of the following in the form $x+iy$.

a $\left(2cis\left(\frac{\pi}{2}\right)\right)^3$ b $\left(3cis\left(\frac{\pi}{6}\right)\right)^4$

c $\left(\sqrt{2}cis\left(-\frac{\pi}{4}\right)\right)^{-2}$ d $\left(5cis\left(\frac{3\pi}{2}\right)\right)^{-3}$

e $\left(-8cis\left(-\frac{\pi}{3}\right)\right)^{-1}$ f $\left(\frac{\sqrt{2}}{3}cis\left(\frac{7\pi}{3}\right)\right)^4$

4. Find each of the following, expressing your answer in the form $x+iy$.

a $(1+i)^3(2-2i)^4$ b $(\sqrt{3}+i)^2(1-i)^2$

c $\frac{(2+2\sqrt{3}i)^3}{(i-1)^2}$ d $(\sqrt{3}+i)^4(1+\sqrt{3}i)^4$

e $\frac{(3+4i)^4}{(3-4i)^2}$ f $\frac{(1+i)^4}{(1-i)^2}$

5. a Prove that $cis(\theta + 2k\pi) = cis(\theta)$, for all integer values of k.

Using part a, evaluate the following.

i $cis(37\pi)$ ii $cis(-43\pi)$

iii $cis\left(\frac{29}{2}\pi\right)$.

6. Simplify the following.

a $cis(\pi)cis\left(-\frac{3\pi}{2}\right)$

b $2cis\left(\frac{\pi}{12}\right) \times 6cis\left(\frac{\pi}{6}\right)$

c $\dfrac{\sqrt{8}cis\left(\frac{\pi}{8}\right)}{\sqrt{2}cis\left(-\frac{\pi}{2}\right)}$

7. a Express $cis\left(\frac{\pi}{4}\right)$ and $cis\left(\frac{\pi}{3}\right)$ in the form $x + iy$.

Hence, express $cis\left(\frac{7\pi}{12}\right)$ in the form $x + iy$.

b Use part a to find the exact value of:

i $\sin\left(\frac{7\pi}{12}\right)$ ii $\cos\left(\frac{7\pi}{12}\right)$

8. Use De Moivre's theorem to prove that:

if $z = rcis(\theta)$ then $(\bar{z})^n = \overline{(z^n)}$.

Extra questions

The *n*th roots of a Complex Number

Definition The nth roots of the complex number $x + iy$ are the solutions of the equation $z^n = x + iy$.

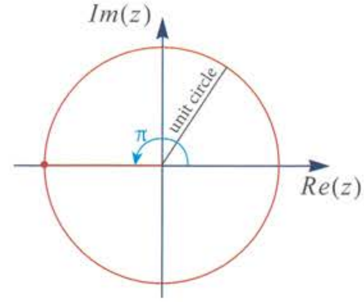
de Moivre's Theorem suggests a geometric approach.

Example 1.7.5
Find the cube roots of -1.

This question amounts to asking for all the solutions to:

$z^3 = -1$ or $z^3 = \text{cis}\pi$.

First locate -1 on the Argand Diagram:



As a consequence of de Moivre's Theorem, any solution to this question must have a modulus of 1 ($1^3 = 1$).

Also, any solution must have an argument which, when multiplied by 3, will give π.

The most obvious answer is an argument of $\frac{\pi}{3}$.

Is $\text{cis}\frac{\pi}{3} = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ a cube root of -1?

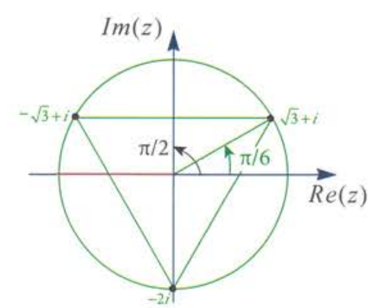
Check:

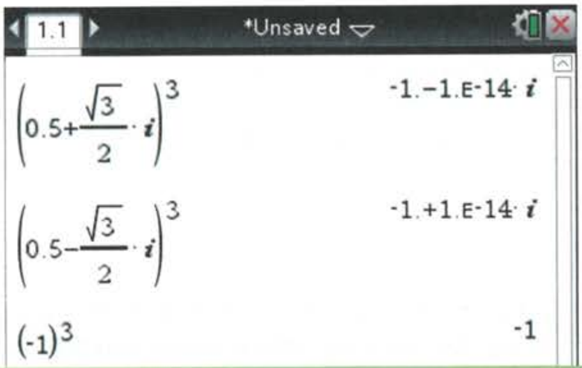
$$\begin{aligned}\left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)^3 &= \left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\\ &= \left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\left(\frac{1}{4}+\frac{\sqrt{3}}{4}i+\frac{\sqrt{3}}{4}i+\frac{3}{4}i^2\right)\\ &= \left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\left(\frac{1}{4}+\frac{\sqrt{3}}{2}i-\frac{3}{4}\right)\\ &= \left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\left(-\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)\\ &= -\frac{1}{4}+\frac{\sqrt{3}}{4}i-\frac{\sqrt{3}}{4}i+\frac{3}{4}i^2\\ &= -\frac{1}{4}-\frac{3}{4}\\ &= -1\end{aligned}$$

However, there are two other answers:

$\text{cis}-\pi = -1$ and $\text{cis}-\frac{\pi}{3} = \frac{1}{2} - \frac{\sqrt{3}}{2}i$

The three solutions lie at the vertices of an equilateral triangle:





Geometrically, we have that the nth roots of a complex number are represented in an Argand diagram as the vertices of a regular polygon of n sides, inscribed in a circle of radius $\sqrt[n]{r}$, and spaced at intervals of $\frac{2\pi}{n}$ from each other.

The steps involved in solving equations of the form $z^n = x+iy$ (even for the case that $y = 0$) are:

Step 1. Express $x+iy$ in polar form, $rcis(\theta)$

Step 2. Realise that $rcis(\theta) = rcis(\theta+2k\pi)$, where k is an integer, because every time you add another 2π, you return to the same position.

Step 3. Use de Moivre's theorem:

$$z^n = rcis(\theta+2k\pi) \therefore z = [rcis(\theta+2k\pi)]^{\frac{1}{n}} = r^{\frac{1}{n}} cis\left(\frac{\theta+2k\pi}{n}\right)$$

Step 4. Use n values of k, usually start at $k = 0, 1, \ldots$ and end at $k = n-1$. This will produce the n required solutions

Example 1.7.6

Find the 6th roots of 64, leaving your answer in polar form.

Setting $z^6 = 64$ we have,

$$z^6 = 64+0i = 64[cis(0)+i\sin(0)]$$

$$= 64cis(0)$$

$$= 64cis(0+2k\pi)$$

$$\therefore z = 64^{1/6}cis\left(\frac{2k\pi}{6}\right), k = 0, 1, 2, 3, 4, 5$$

Therefore, we have $z = 2cis\left(\frac{\pi k}{3}\right), k = 0, 1, 2, 3, 4, 5$.

So that,

$$z = 2cis(0), 2cis\left(\frac{\pi}{3}\right), 2cis\left(\frac{2\pi}{3}\right), 2cis(\pi), 2cis\left(\frac{4\pi}{3}\right), 2cis\left(\frac{5\pi}{3}\right)$$

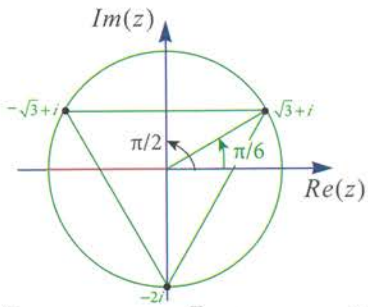
Example 1.7.7

Find the cube roots of $8i$.

8i has a modulus of 8 and an argument of $\frac{\pi}{2}$.

By de Moivre's Theorem, one these cube roots will have an argument of one third of the argument of $8i$. The moduli of all the roots will be the cube root of 8 (= 2).

The other two roots will be at the vertices of an equilateral triangle (triangle because we are looking for a cube root).



That is, $z = \sqrt{3}+i$ or $z = -\sqrt{3}+i$ or $z = -2i$.

Example 1.7.8

Find the four fourth roots of $1+i\sqrt{3}$. Give your answer in polar form.

We start by expressing $1+i\sqrt{3}$ in its polar form:

$$1+i\sqrt{3} = 2cis\left(\frac{\pi}{3}\right).$$

Then, set $z^4 = 2cis\left(\frac{\pi}{3}\right) = 2cis\left(\frac{\pi}{3}+2k\pi\right) = 2cis\left(\frac{\pi+6k\pi}{3}\right)$

So that, $z = \sqrt[4]{2}cis\left(\frac{\pi+6k\pi}{12}\right), k = 0, 1, 2, 3$.

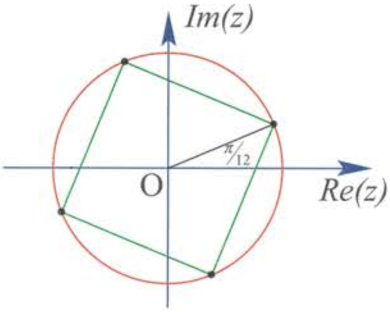
For $k = 0$, $z = \sqrt[4]{2}cis\left(\frac{\pi}{12}\right)$;

$k = 1$, $z = \sqrt[4]{2}cis\left(\frac{\pi+6\pi}{12}\right) = \sqrt[4]{2}cis\left(\frac{7\pi}{12}\right)$;

$k = 2$, $z = \sqrt[4]{2}cis\left(\frac{\pi + 12\pi}{12}\right) = \sqrt[4]{2}cis\left(\frac{13\pi}{12}\right)$;

$k = 3$, $z = \sqrt[4]{2}cis\left(\frac{\pi + 18\pi}{12}\right) = \sqrt[4]{2}cis\left(\frac{19\pi}{12}\right)$

Therefore, the four roots of $1 + i\sqrt{3}$ lie on the circumference of a circle of radius $\sqrt[4]{2}$ units and are evenly separated by an angle of $\frac{\pi}{2}$.



Again notice that the roots in this instance do not occur in conjugate pairs.

Exercise 1.7.2

1. Use the nth root method to solve the following:

   a $z^3 = 27$ b $z^3 = 27i$

   c $z^3 = -8i$ d $z^4 = -16$

2. Find the fourth roots of -4 in the form $x + iy$ and hence factorise $z^4 + 4$ into linear factors.

3. Find the square roots of:

   a i b $3 + 4i$ c $-1 + \sqrt{3}i$.

   Represent these roots on an Argand diagram.

4. Find the cube roots of:

   a $1 - i$ b $-1 + \sqrt{3}i$ c i

   Represent these roots on an Argand diagram.

5. Solve the following equations.

   a $z^4 = 1 + i$

   b $z^4 = i$

   c $z^3 + i = 0$

   d $z^4 = 8 - 8\sqrt{3}i$

   e $z^3 = 64i$

   f $z^2 = \sqrt{3} + i$

   Represent these roots on an Argand diagram.

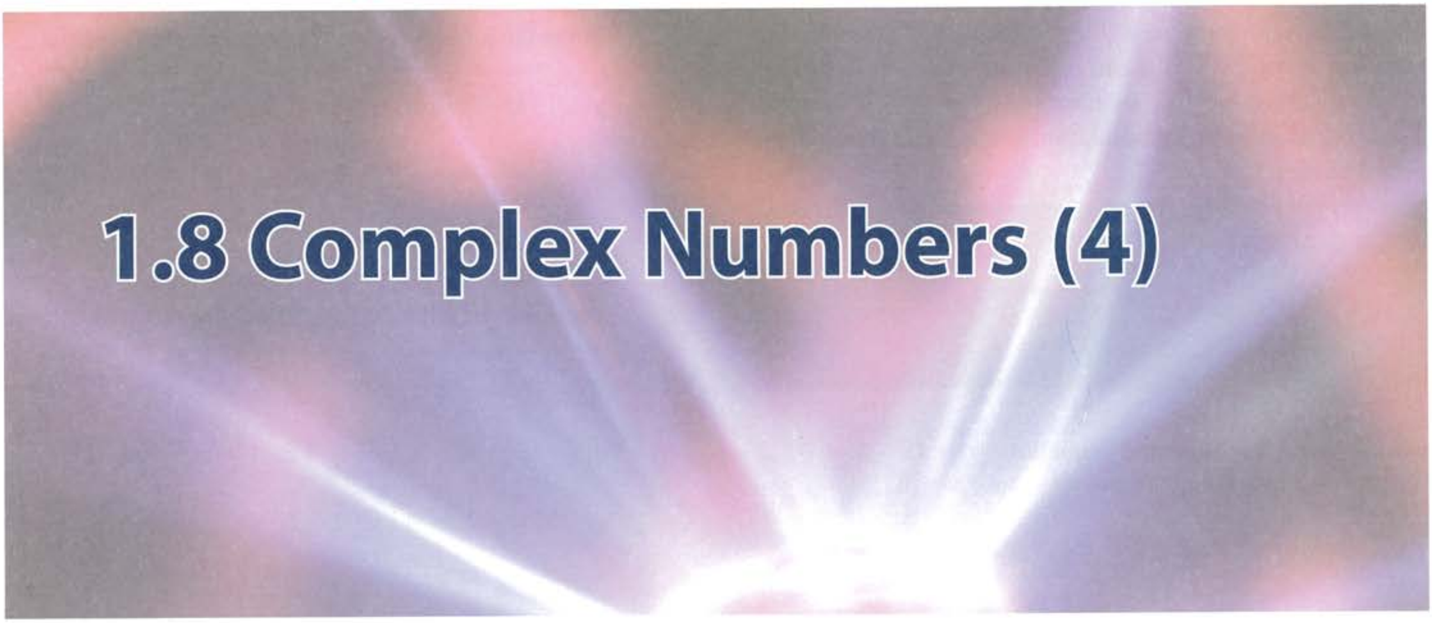
6 a Find the cube root of unity.

   b Hence, show that if $w^3 = 1$, then $1 + w + w^2 = 0$.

7. Three points, of which $1 + i\sqrt{3}$ is one point, lie on the circumference of a circle of radius 2 units and centre at the origin. If these three points form the vertices of an equilateral triangle, find the other two points.

Extra questions

Answers



1.8 Complex Numbers (4)

Polynomials

This section will look at polynomials with real coefficients in which the variable may take complex values. To emphasise this, the variable is generally labelled z (rather than x).

$P(z) = 2z^2 + 3z - 4$ is an example of a complex polynomial with real coefficients.

$P(z) = 2z^2 + 3iz - 4$ is an example of a complex polynomial with a complex coefficient (shown in green). Such polynomials are not included in this course.

Arising from such polynomials are equations with complex solutions. Our cover shows a plasma discharge. Solving many problems in the natural sciences involves complex numbers. Even though they may be 'imaginary' (a term many dispute), complex numbers figure in the solution of 'real' problems.

Quadratic equations

We start this section by looking at equations of the form $ax^2 + bx + c = 0$ where the discriminant, $\Delta = b^2 - 4ac < 0$. Such equations will produce complex solutions.

Example 1.8.1

Factorise, over the complex number field, z^2+2z+2. Hence, solve the equation $z^2+2z+2=0$.

We start the same way we would when dealing with any quadratic expression:

$$
\begin{aligned}
z^2+2z+2 &= \left(z^2+2z+1\right)+1 \quad \text{(complete the square)}\\
&= (z+1)^2+1\\
&= (z+1)^2-i^2 \quad \text{(difference of two squares)}\\
&= (z+1+i)(z+1-i)
\end{aligned}
$$

To solve $z^2+2z+2=0$, we have:

$$(z+1+i)(z+1-i) = 0 \Leftrightarrow z = -1-i \text{ or } z = -1+i.$$

Therefore, the two complex solutions are $z = -1 - i$ and $z = -1 + i$. Notice that the solutions are a conjugate pair.

Example 1.8.2

Solve the equation $z^2+3z+5=0$ over the complex field.

Rather than factorizing the equation, we will use the quadratic formula.

$$
\begin{aligned}
z^2+3z+5 = 0 \Leftrightarrow z &= \frac{-3 \pm \sqrt{3^2-4\times1\times5}}{2\times1}\\
&= \frac{-3\pm\sqrt{-11}}{2}\\
&= \frac{-3\pm\sqrt{11}\,i}{2}
\end{aligned}
$$

Therefore, the two complex solutions are

$$z = -\frac{3}{2}+\frac{\sqrt{11}}{2}i,\ z = -\frac{3}{2}-\frac{\sqrt{11}}{2}i.$$

Again, notice the conjugate pair that make up the solution.

Quadratics also come in a 'hidden form'. For example, the equation $z^6 + 4z^3 - 5 = 0$ can be considered to be a 'hidden form'. i.e. letting $w = z^3$ we have $w^2 + 4w - 5 = 0$. And so we can then solve the quadratic in w. We could then obtain solutions for z.

Example 1.8.3

Solve the equation $z^4 + 4z^2 - 5 = 0$ over the complex field.

Let $w = z^2$ so that the equation $z^4 + 4z^2 - 5 = 0$ is transformed into the quadratic $w^2 + 4w - 5 = 0$.

Then, we have $w^2 + 4w - 5 = 0 \Leftrightarrow (w+5)(w-1) = 0$

$$\Leftrightarrow (z^2+5)(z^2-1) = 0$$

$$\Leftrightarrow (z-\sqrt{5}i)(z+\sqrt{5}i)(z-1)(z+1) = 0$$

Therefore, we have that $z = \sqrt{5}i$ or $z = -\sqrt{5}i$ or $z = 1$ or $z = -1$.

That is, we have four solutions, two real and two complex (again, the complex solutions are conjugate pairs).

Exercise 1.8.1

1. Factorise the following over the complex number field.

a	$x^2 - 6x + 10$	b	$x^2 + 4x + 13$
c	$x^2 - 2x + 2$	d	$z^2 + 4z + 5$
e	$z^2 - 3z + 4$	f	$z^2 + 10z + 30$
g	$4w^2 + 4w + 17$	h	$3w^2 - 6w + 6$
i	$-2w^2 + 8w - 11$		

2. Solve the following over the complex number field.

   a $z^2 + 4z + 8 = 0$

   b $z^2 - z + 3 = 0$

   c $3z^2 - 3z + 1 = 0$

   d $2w^2 + 5w + 4 = 0$

   e $w^2 + 10w + 29 = 0$

3. Solve the following over the complex number field.

   a $z^4 - 3z^2 - 4 = 0$

   b $w^4 - 8w^2 - 9 = 0$

   c $z^4 - 5z^2 - 36 = 0$

4. Factorise the following over the complex number field.

a	$z^2 + 25$	b	$z^2 + 49$
c	$z^2 + 4z + 5$	d	$z^2 + 6z + 11$
e	$z^4 + 2z^2 - 8$	f	$z^4 - z^2 - 6$

Polynomial equations (of order ≥ 3)

We now look at some of the more general polynomial equations that provide a combination of real and imaginary roots and factors. The important thing to remember is that the laws for real polynomials hold equally well for complex polynomials.

A polynomial, $P(z)$ of degree n in one variable is an expression of the form

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$$

If the coefficients, $a_n, a_{n-1}, \ldots, a_1, a_0$ are real, the polynomial is a **polynomial over the real number field**, while if they are complex numbers, the polynomial is a **polynomial over the complex field**. We shall, however, concentrate on polynomials over the real field.

We state some standard results:

Remainder Theorem

If a polynomial $P(x)$ is divided by a linear polynomial $(x-a)$, the remainder is $P(a)$.

Factor Theorem

If, when a polynomial $P(x)$ is divided by a linear polynomial $(x-a)$, the remainder $P(a)$ is zero, then $(x-a)$ is a factor of $P(x)$.

Fundamental Theorem of Algebra

Every polynomial equation of the form $P(z) = 0$, $z \in C$, of degree $n \in \mathbb{Q}^+$ has at least one complex root.

This theorem is the basis for the next important result:

A polynomial $P_n(z)$, $z \in C$, of degree $n \in \mathbb{Q}^+$, can be expressed as the product of n linear factors and hence, produce exactly n solutions to the equation $P_n(z) = 0$.

We have already observed, in previous examples, the occurrence of conjugate pairs when solving quadratics with real coefficients. We now state another result.

Conjugate Root Theorem (C.R.T)

The complex roots of a polynomial equation with real coefficients occur in conjugate pairs.

Example 1.8.4

Factorise the polynomial $z^3 - 3z^2 + 4z - 12$, hence solve $z^3 - 3z^2 + 4z - 12 = 0$.

Grouping like terms, we have:

$z^3 - 3z^2 + 4z - 12 = z^2(z-3) + 4(z-3) = (z^2+4)(z-3)$

i.e. $z^3 - 3z^2 + 4z - 12 = (z-2i)(z+2i)(z-3)$

And so, $z^3 - 3z^2 + 4z - 12 = 0 \Leftrightarrow (z-2i)(z+2i)(z-3) = 0$

Therefore, we have that $z = 2i$ or $z = -2i$ or $z = 3$.

We observe that two of the roots are conjugate pairs, and when we look at the polynomial, we see that all of the coefficients are real (as expected from the C.R.T).

Example 1.8.5

Given that $z = 1 - i$ is a root of the equation $2z^3 - 7z^2 + 10z - 6 = 0$, find the other roots.

As all of the coefficients of the polynomial are real, it means that the C.R.T applies. That is, given that $z = 1 - i$ is a root, so too then, is $z = 1 + i$.

Therefore, we have two factors, namely, $z - 1 + i$ and $z - 1 - i$.

This means that $(z-1+i)(z-1-i) = z^2 - 2z + 2$ is also a factor.

As in the last example, we can factorise by inspection: $2z^3 - 7z^2 + 10z - 6 = (az+b)(z^2-2z+2)$

That is, knowing that we are looking for a cubic, and given that we already have a quadratic factor, we are left with a linear factor, which is $(az+b)$. Then, comparing the coefficients of the z^3 term and the constant term we have that:

$a = 2$ and $2b = -6 \Leftrightarrow b = -3$.

That is, $2z^3 - 7z^2 + 10z - 6 = (2z-3)(z^2-2z+2)$

Therefore, the roots are $1-i, 1+i, \frac{3}{2}$.

Example 1.8.6

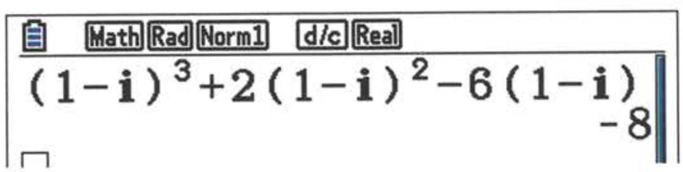
If $z - 1 + i$ is a factor of $P(z) = z^3 + 2z^2 - 6z + k$, find the value of k.

Given that $z - 1 + i$ is a factor of $P(z) = z^3 + 2z^2 - 6z + k$, then, by the factor theorem we must have that $P(1-i) = 0$.

So, $(1-i)^3 + 2(1-i)^2 - 6(1-i) + k = 0 \Leftrightarrow -8 + k = 0$

$\Leftrightarrow k = 8$

Calculators are useful in situations that involve simple evaluation of complex numbers.



Example 1.8.7

Solve the equation $z^3 - 4z^2 + 9z - 10 = 0$ where z is a complex number.

Let $P(z) = z^3 - 4z^2 + 9z - 10$. Using trial and error (or at least factors of 10), we have:

$P(1) = 1 - 4 + 9 - 10 = -4 \therefore (z-1)$ is not a factor.

$P(2) = 8 - 16 + 18 - 10 = 0 \Rightarrow (z-2)$ is a factor.

Therefore, $P(z) = (z-2)(az^2 + bz + c)$.

Comparing coefficients of the leading term and constant term we have:

$a = 1$ and $-2c = -10 \Leftrightarrow c = 5$

Therefore, $P(z) = (z-2)(z^2+bz+5)$.

Then, comparing the coefficient of the z^2 term, we have that $b-2 = -4 \therefore b = -2$.

So, $P(z) = (z-2)(z^2-2z+5) = (z-2)[(z^2-2z+1)+4]$ (completing the square)

$$= (z-2)[(z-1)^2+4]$$

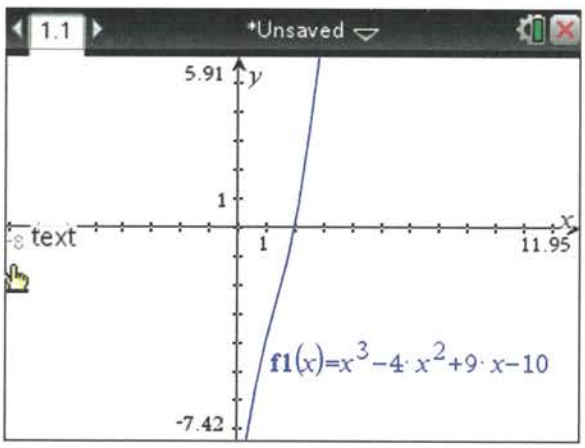
$$= (z-2)(z-1+2i)(z-1-2i)$$

Therefore, $P(z) = 0 \Leftrightarrow (z-2)(z-1+2i)(z-1-2i) = 0$

And so, $z = 2$ or $z = 1-2i$ or $z = 1+2i$.

We could have used long or synthetic division to factorize $P(z)$ to the stage $P(z) = (z-1)(z^2-3z+10)$. Both methods are equally valid.

Also, rather than using trial and error you could use your graphics calculator to help find a first real factor.



Exercise 1.8.2

1. Factorize the following over the complex number field.

   a z^3+2z^2+z+2 b z^3-9z^2+z-9

   c z^3-2z^2+2z-4

2. Factorise the following over the complex number field.

   a $w^3+2w-12$ b z^3-5z^2+9z-5

   c z^3+z^2-2 d x^4-3x^2-4

   e w^3-2w+4 f z^4-625

3. Solve each of the following over the complex number field.

   a $z^3-7z^2+31z-25 = 0$

   b $z^3-8z^2+25z-26 = 0$

   c $z^4-3z^3-2z^2+10z-12 = 0$

   d $2w^3+3w^2+2w-2 = 0$

   e $6z^4-11z^3+z^2+33z-45 = 0$

   f $z^3+7z^2+16z+10 = 0$

4. Given that $\frac{1}{2}(-1+\sqrt{3}i)$ is a root of:

   $3z^3+2z^2+2z-1 = 0$, find all other roots.

5. Given that $(z-1-2i)$ is a factor of $2z^3-3z^2+8z+5$ solve the equation $2z^3-3z^2+8z+5 = 0$ over the complex number field.

6. Given that $P(2-3i) = 0$, find all three linear factors of $z^3-7z^2+25z-39$.

7. Find all complex numbers, z, such that $z^4-z^3+6z^2-z+15 = 0$ and $z = 1+2i$ is a solution to the equation.

8. Factorise the following.

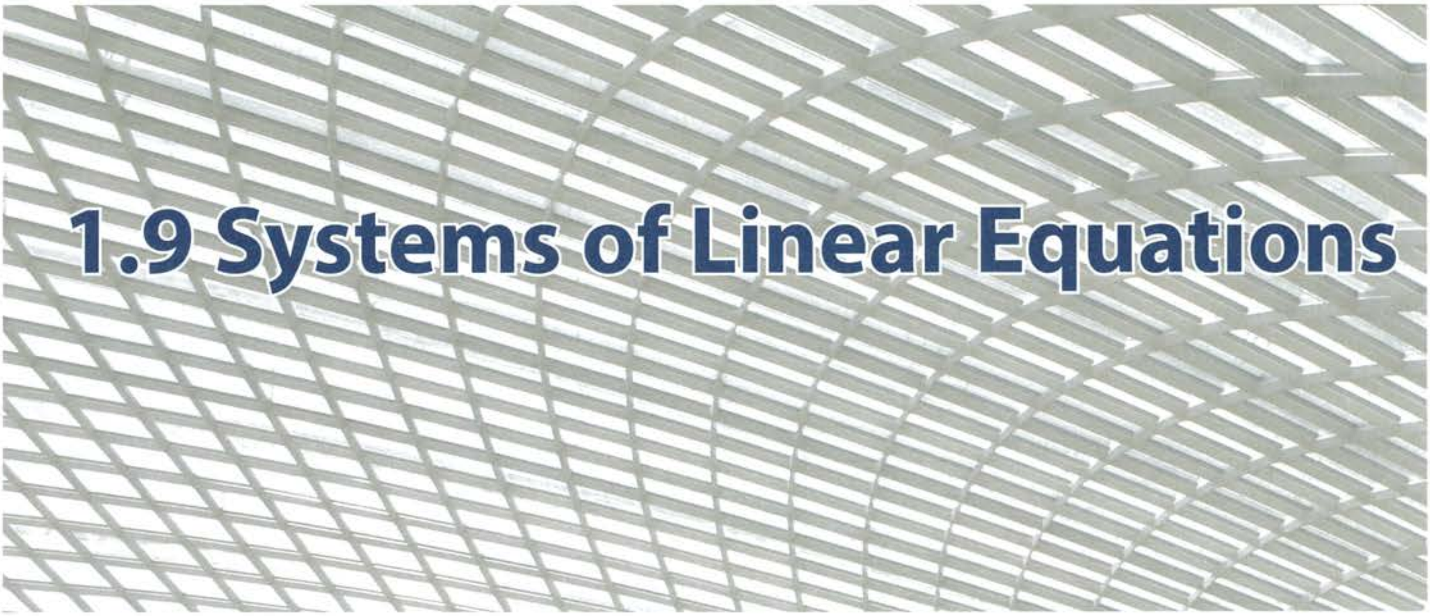
   a $2z^3-z^2+2z-1$ b z^4+z^2-12

9. Given that $2-i$ is a root of $z^3+az^2+z+5 = 0$ where a is a real number, find all the roots to this equation.

10. Given that $2+3i$ is a root of $z^3+az^2+b = 0$, where a and b are real numbers, find all the roots of this equation.

Extra questions

Answers



1.9 Systems of Linear Equations

Simultaneous linear equations in two unknowns

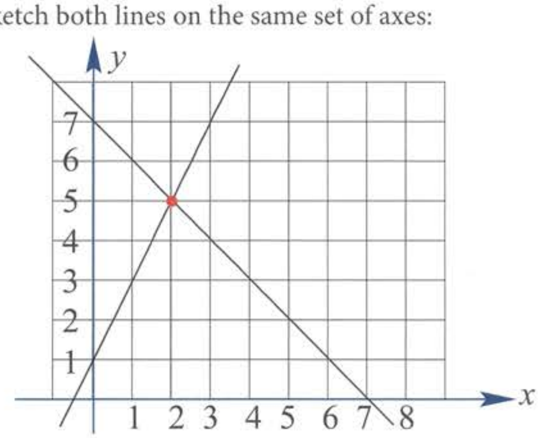
Pairs of simultaneous equations in two unknowns may be solved in two ways, either algebraically or graphically. To solve means to find where the two straight lines intersect once they have been sketched. So, we are looking for the point of intersection.

Method 1: Graphical

Example 1.9.1

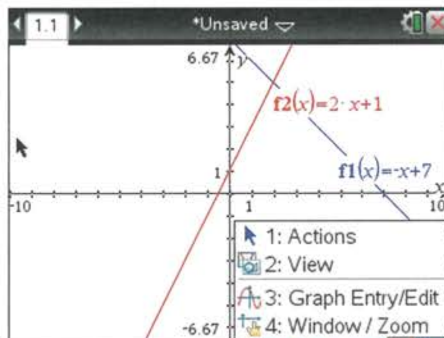
Solve the system of linear equations $y = -x + 7$ and $y = 2x + 1$.

We sketch both lines on the same set of axes:

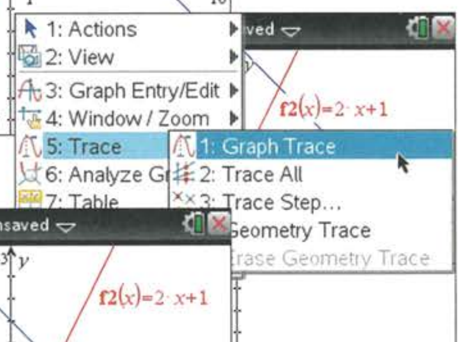


Reading off the grid we can see that the straight lines meet at the point with coordinates (2, 5). So, the solution to the given system of equations is $\boldsymbol{x} = 2$ and $\boldsymbol{y} = 5$.

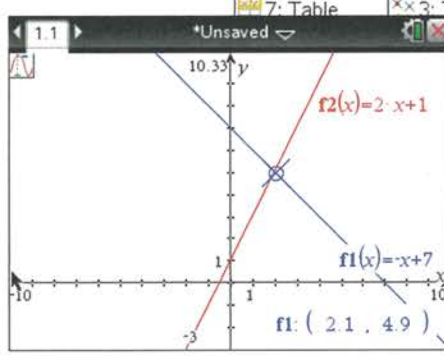
There are a number of ways that the graphics calculator can be used.



1. Plot the graphs

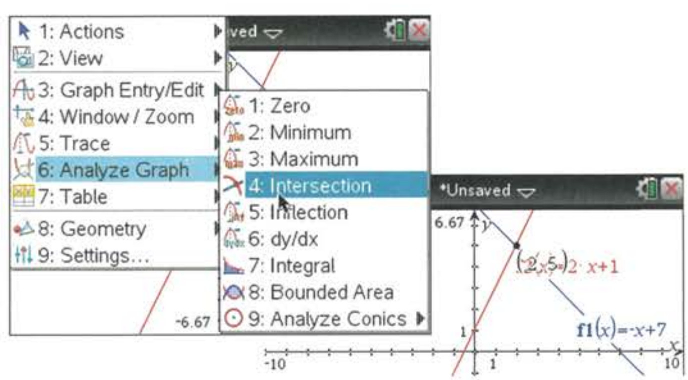


2. Use graph/trace

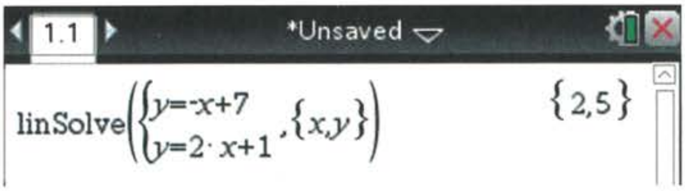


3. Solution
Note that the pixels on the screen produce an approximate solution

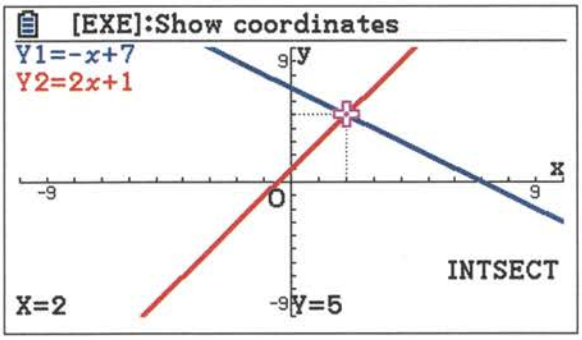
A more satisfactory way is to use the calculator to find the intersection.



Thirdly, you can use the solve facility:



Similar calculations can be performed on Casio models. This screen uses Graph mode (5) followed by F6-draw, Shift F5-G-Solv and F5-INTSCT to find the intersection.



Method 2: Algebraic

There are two possible approaches when dealing with simultaneous equations algebraically. They are the process of:

1. Elimination

2. Substitution

The choice of method often depends on the way the equations are presented.

Elimination method

The **key step** in using the elimination method is to obtain, for one of the variables (in both equations), coefficients that are the same (or only differ in sign). Then:

1. if the coefficients are the same, you subtract one equation from the other – this will eliminate one of the variables – leaving you with only one unknown.

2. if the coefficients only differ in sign, you add the two equations – this will eliminate one of the variables – leaving you with only one unknown.

Example 1.9.2

Use the elimination method to solve: $\begin{aligned} x-2y&=-7 \\ 2x+3y&=0 \end{aligned}$

As it is easier to add than subtract, we try to eliminate the variable which differs in sign. In this case the variable '*y*' is appropriate. However, the coefficients still need to be manipulated. We label the equations as follows:

$$x-2y = -7 \quad -(1)$$

$$2x+3y = 0 \quad -(2)$$

$3\times(1)$: $\quad 3x-6y = -21 \quad -(3)$

$2\times(2)$: $\quad 4x+6y = 0 \quad -(4)$

Adding $(3)+(4)$: $\quad 7x+0 = -21$

$$\Leftrightarrow x = -3$$

Substituting into (1) we can now obtain the y-value:

$-3-2y = -7 \Leftrightarrow -2y = -4 \Leftrightarrow y = 2$.

Therefore, the solution is $x = -3$, $y = 2$.

Once you have found the solution, always check with one of the original equations.

Using equation (2) we have: L.H.S $= 2\times-3+3\times2 = 0 =$ R.H.S.

Note that we could also have multiplied equation (1) by 2 and then subtracted the result from equation (2). Either way, we have the same answer.

Substitution method

The substitution method relies on making one of the variables the subject of one of the equations. Then we substitute this equation for its counterpart in the other equation. This will then produce a new equation that involves only one unknown. We can solve for this unknown and then substitute its value back into the first equation. This will then provide a solution pair.

Example 1.9.3

Use the substitution method to solve: $\begin{aligned} 5x-y&=4 \\ x+3y&=4 \end{aligned}$

Label the equations as follows: $\quad 5x-y = 4 \quad -(1)$

$$x+3y = 4 \quad -(2)$$

From equation (1) we have that $y = 5x - 4$ - (3)

Substituting (3) into (2) we have: $x + 3(5x - 4) = 4$

$$\Leftrightarrow 16x - 12 = 4$$
$$\Leftrightarrow 16x = 16$$
$$\Leftrightarrow x = 1$$

Substituting $x = 1$ into equation (3) we have:

$$y = 5 \times 1 - 4 = 1$$

Therefore, the solution is given by $x = 1$ and $y = 1$.

Check: Using equation (2) we have: L.H.S $= 1 + 3 \times 1 = 4$ = R.H.S

Not all simultaneous equations have unique solutions. Some pairs of equations have no solutions while others have infinite solution sets. You will need to be able to recognise the 'problem' in the processes of both algebraic and graphical solutions when dealing with such equations.

The following examples illustrate these possibilities.

Example 1.9.4

Solve: a $\begin{array}{l} 2x+6y=8 \\ 3x+9y=12 \end{array}$ b $\begin{array}{l} 2x+6y=8 \\ 3x+9y=15 \end{array}$

a **Algebraic solution:**

Label the equations as follows:

$$2x + 6y = 8 \text{ - (1)}$$

$$3x + 9y = 12 \text{ - (2)}$$

$3 \times (1)$: $6x + 18y = 24$ - (3)

$2 \times (2)$: $6x + 18y = 24$ - (4)

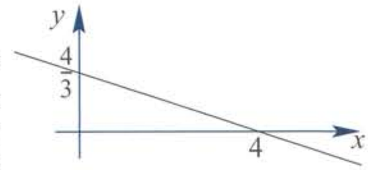
In this case, we have the same equation. That is, the straight lines are **coincident**.

If we were to 'blindly' continue with the solution process, we would have:

$3 \times (1) - 2 \times (2)$: $0 = 0$ ✓

The algebraic method produces an equation that is always true, i.e. zero will always equal zero. This means that any pair of numbers that satisfy either equation will satisfy both and are, therefore, solutions to the problem. Examples of solutions are: $x = 4, y = 0, x = 1, y = 1, x = 7, y = -1$. In this case we say that there is an **infinite** number of solutions.

Graphical solution:



Graphically, the two equations produce the same line. The coordinates of any point on this line will be solutions to both equations.

b **Algebraic solution:**

Label the equations as follows:

$$2x + 6y = 8 \text{ - (1)}$$

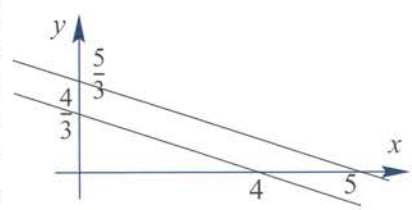
$$3x + 9y = 15 \text{ - (2)}$$

$3 \times (1)$: $6x + 18y = 24$ - (3)

$2 \times (2)$: $6x + 18y = 30$ - (4)

$(4) - (3)$: $0 = 6$ ×

Graphical solution:



The algebraic method produces an equation that is never true. This means that there are no solutions to the equations. Graphically, the two lines are parallel and produce no points of intersection.

Exercise 1.9.1

1. Solve these simultaneous equations, giving exact answers.

a $\begin{array}{l} 3x - 2y = -1 \\ 5x + 2y = 9 \end{array}$ b $\begin{array}{l} 3x + 5y = 34 \\ 3x + 7y = 44 \end{array}$

c $\begin{array}{l} 2x + 4y = 6 \\ 4x - 3y = -10 \end{array}$ d $\begin{array}{l} 3x + 2y = 2 \\ 2x - 6y = -6 \end{array}$

e $\begin{aligned} 5x+4y &= -22 \\ 3x-y &= -3 \end{aligned}$ f $\begin{aligned} 5x-9y &= -34 \\ 2x+3y &= -7 \end{aligned}$

2. Solve these simultaneous equations, giving fractional answers where appropriate.

a $\begin{aligned} 3x-y &= 2 \\ 5x+2y &= 9 \end{aligned}$ b $\begin{aligned} 4x+2y &= 3 \\ x-3y &= 0 \end{aligned}$

c $\begin{aligned} -3x+y &= 0 \\ 2x-4y &= 0 \end{aligned}$ d $\begin{aligned} \frac{x}{2}-3y &= 4 \\ 4x+\frac{3y}{2} &= -1 \end{aligned}$

e $\begin{aligned} 5x+\frac{2y}{3} &= -4 \\ 4x+y &= 2 \end{aligned}$ f $\begin{aligned} \frac{3x}{5}-4y &= \frac{1}{2} \\ x-2y &= \frac{1}{3} \end{aligned}$

3. Find the values of m such that these equations have no solutions.

a $\begin{aligned} 3x-my &= 4 \\ x+y &= 12 \end{aligned}$ b $\begin{aligned} 5x+y &= 12 \\ mx-y &= -2 \end{aligned}$

c $\begin{aligned} 4x-2y &= 12 \\ 3x+my &= 2 \end{aligned}$

4. Find the values of m and a such that these equations have infinite solution sets.

a $\begin{aligned} 4x+my &= a \\ 2x+y &= 4 \end{aligned}$ b $\begin{aligned} 5x+2y &= 12 \\ mx+4y &= a \end{aligned}$

c $\begin{aligned} 3x+my &= a \\ 2x-4y &= 6 \end{aligned}$

Extra questions

Simultaneous linear equations in three unknowns

So far we have looked at linear equation in two unknowns. However, this can be extended to linear equations in three unknowns. Equations such as these, involving the variables x, y and z take on the general form $ax + by + cz = k$ where a, b, c and k are real constants.

Just as for the case with two unknowns, where we required two equations to (hopefully) obtain a unique solution to the system of simultaneous equations, when dealing with three unknowns we will require a minimum of three equations to (hopefully) obtain a unique solution.

The solution process for a system of linear equations in three unknowns will require, primarily, the use of the elimination method. The method usually involves the reduction of a system of three equations in three unknowns to one of two equations in two unknowns. This will then enable the use of the methods already discussed to solve the "reduced" system. Once two of the unknowns have been determined from this "reduced" system, we substitute back into one of the original three equations to solve for the third unknown.

Example 1.9.5

Solve the simultaneous equations: $\begin{aligned} x+3y-z&=13 \\ 3x+y-z&=11 \\ x+y-3z&=11 \end{aligned}$

We label the equations as follows:

$$x+3y-z = 13 \quad -(1)$$

$$3x+y-z = 11 \quad -(2)$$

$$x+y-3z = 11 \quad -(3)$$

Reduce the system to one involving two equations and two unknowns.

We first eliminate the variable z:

(2) – (1): $2x-2y = -2 \quad -(4)$

$3 \times (2)$ – (3): $8x+2y = 22 \quad -(5)$

Solve the reduced system of equations.

(4) + (5): $10x = 20 \quad \Leftrightarrow x = 2$

Substitute into (4):
$2\times 2-2y = -2 \Leftrightarrow -2y = -6 \Leftrightarrow y = 3$.

Solve for the third unknown.

Substituting $x = 2$ and $y = 3$ into (1):

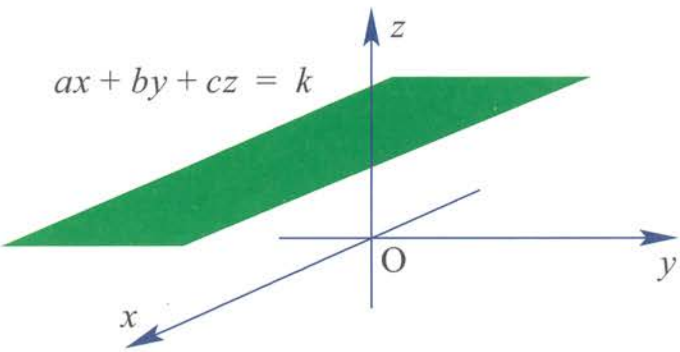
$2+3\times 3-z = 13 \Leftrightarrow z = -2$

Therefore the solution is given by $x = 2$, $y = 3$ and $z = -2$.

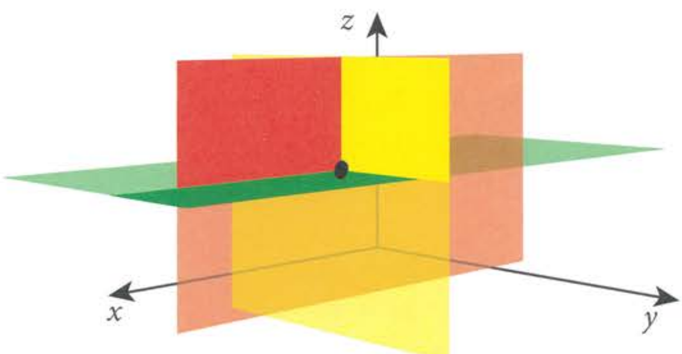
Check: Using equation (2):
L.H.S. $= 2 + 3 - 3 \times -2 = 11 =$ R.H.S

We have already seen that linear equations in two unknowns are represented by straight lines on the Cartesian axes. The question then becomes, "What do linear equations in three unknowns look like?"

Equations of the form $ax + by + cz = k$ represent a plane in space. To draw such a plane we need to set up three mutually perpendicular axes that coincide at some origin O. This is commonly drawn with a horizontal x–y plane and the z-axis in the vertical direction:

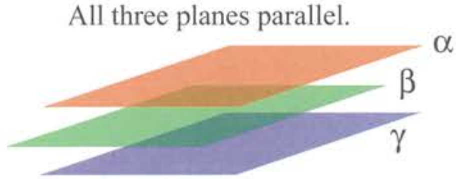


In Example 1.9.5 we obtained a unique solution. This (2, 3, –2 means that the three planes must have intersected at a unique point. We can represent such a solution as shown in the diagram below:



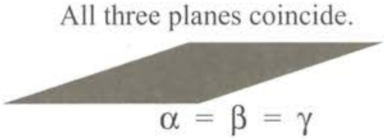
There are a number of possible combinations for how three planes in space can intersect (or not). Labelling the planes as α, β and γ the possible outcomes are shown below.

All three planes parallel.

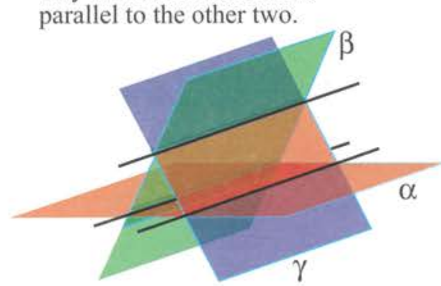


Two planes coincide.

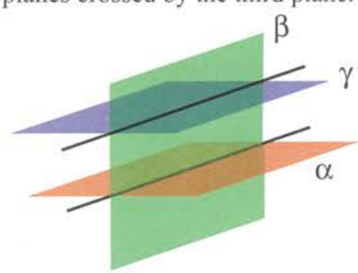
All three planes coincide.



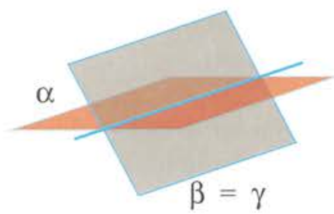
Any line of intersection is parallel to the other two.



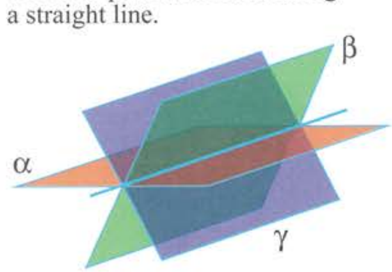
Two parallel non-coincident planes crossed by the third plane.



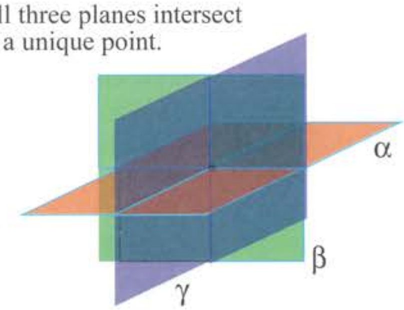
Two parallel coincident planes crossed by the third plane.



All three planes intersect along a straight line.



All three planes intersect at a unique point.



Example 1.9.6

Solve the simultaneous equations:

$$x+2y=10$$
$$3x+2y-4z=18$$
$$y+z=3$$

We label the equations as follows;

$x+2y = 10 \qquad -(1)$

$3x+2y-4z = 18 \qquad -(2)$

$y+z = 3 \qquad -(3)$

We eliminate $\boldsymbol{x}$ using equations (1) and (2):

(2) – 3 × (1): $\quad -4y-4z = -12$

$\Leftrightarrow y+z = 3 \quad -(4)$

We are now left with equations (3) and (4). However, these two equations are identical.

To obtain the solution set to this problem we introduce a **parameter,** we let z be any arbitrary value, say $z = k$ where k is some real number.

Then, substituting into equation (4), we have:

$y+k = 3 \Rightarrow y = 3-k$.

Next, we substitute into (1) so that

$x+2(3-k) = 10 \Rightarrow x = 4+2k$.

Therefore, the solution is given by

$x = 4+2k, y = 3-k, z = k$.

Notice the nature of the solution. Each of the variables is expressed as a linear function of $\boldsymbol{k}$. This means that we have a situation where the three original planes meet along a straight line.

Supplementary example - matrices.

Exercise 1.9.2

Solve the simultaneous equations:

a
$$\begin{aligned} 6x+4y-z &= 3 \\ x+2y+4z &= -2 \\ 5x+4y &= 0 \end{aligned}$$

b
$$\begin{aligned} x+y+z &= 2 \\ 4x+y &= 4 \\ -x+3y+2z &= 8 \end{aligned}$$

c
$$\begin{aligned} 4x+9y+13z &= 3 \\ -x+3y+24z &= 17 \\ 2x+6y+14z &= 6 \end{aligned}$$

d
$$\begin{aligned} x-2y-3z &= 3 \\ x+y-2z &= 7 \\ 2x-3y-2z &= 0 \end{aligned}$$

e
$$\begin{aligned} x-y-z &= 2 \\ 3x+3y-7z &= 7 \\ x+2y-3z &= 3 \end{aligned}$$

f
$$\begin{aligned} x-2y &= -1 \\ -x-y+3z &= 1 \\ y-z &= 0 \end{aligned}$$

g
$$\begin{aligned} x+y+z &= 1 \\ x-y+z &= 3 \\ 4x+2y+z &= 6 \end{aligned}$$

h
$$\begin{aligned} -2x+y-2z &= 5 \\ x+4z &= 1 \\ x+y+10z &= 10 \end{aligned}$$

Answers

Theory of Knowledge

The Need for New Concepts and Notations

Throughout history, various notations and operations were introduced by mathematicians when they discovered their current set of notations was inadequate to address certain new mathematical concepts. In this chapter, we have studied the logarithm as an inverse operation to exponentiation.

The concept of exponentiation was used by Euclid as early as 300BC in ancient Greek. In other parts of the world, mathematicians continued to explore this concept and discovered new rules governing the proper use of exponents. However, it was not until the 17th century that the logarithm was first introduced by John Napier in a book titled *Mirifici Logarithmorum Canonis Descriptio.*

During the years when the idea of the logarithm was not formalized in the field of mathematics, were people not able to find the inverse of exponentiation? Taking this mathematical operation as an example, have you ever wondered what people used before a certain concept or notation was introduced?

In mathematics, as well as in other disciplines, before a concept was introduced and accepted to address a knowledge gap, does it mean that particular concept did not exist or was it irrelevant at that moment in time? What drives the discovery of a new concept in mathematics? Does intuition play a pivotal role in recognizing a gap in the current set of knowledge and notations before a mathematician can formalize and present a new concept, a new symbol, or a new notation? In other words, before a new mathematical concept is introduced did the empirical evidence or the rational thinking come first? Is it necessary before a new concept is formalized and accepted that it must have both empirical evidence and rational thinking?

With respect to the written notations in mathematics, have you ever wondered why certain symbols and notations are reused in different contexts and have very different meaning? For example, if you are presented with (3,5), does it suggest a coordinate pair on the Cartesian plane or does it suggest an open interval between 3 and 5 exclusively? Similarly, have you ever doubted your understanding of the difference between $f^{-1}(x)$ and $f(x)^{-1}$? These are just some examples to illustrate how a precise language like mathematics can also be ambiguous to a certain extent. If mathematics is a language, what grammatical rules are you following? Is this language evolving with time? Or is this language static and unchanging over time, thus limiting its ability to communicate newer concepts in mathematics?

Assumptions and Conventions

By definition, an *assumption* is a claim for a concept, a thing, or a situation, that is accepted as true without evidence, justification, or proof. Conversely, by definition, *convention* is a way in which an action is usually taken or a way in which something is usually done.

In mathematics, when one attempts to provide an answer to a question, it is necessary to show the logical deductive reasoning to ensure there is no error in applying the algebraic rules. However, does it merely mean that assumptions and conventions are not to be used and considered when one studies mathematics? If one only provides an answer solely based on assumptions, does it warrant the answer to be inaccurate?

It is understood by mathematicians that x is x^1, when the exponent 1 is already assumed in the written notation of x. Similarly, the expression $\log x$ is assumed to be written in base 10 (i.e. $\log_{10}x$) or the radical term $\sqrt{x}$ is already understood to be the same as $\sqrt[2]{x}$. Do these assumptions consequently affect the validity of the answer? Likewise, if these assumptions are generally accepted as a convention in written mathematics, then who decides which conventions are to be adopted or rejected? How do cultural and historical factors influence these assumptions and conventions? Are these written mathematical conventions infallible?

If mathematics is constructed from deductions, in which one must assume certain things before inferring conclusions, then how does it affect the validity of the conclusions if the original assumptions are not entirely true? If the assumptions were not entirely true, then would it imply the conclusion to be false? Or would that be considered as an exception to the general rule?

If not true, then false?

Finding an answer for every question in mathematics may be an impossible task despite utilizing the finite set of axioms and the abstract language of the subject. Most mathematics questions in pre-tertiary school contexts often present themselves into the polarity of right or wrong, correct or incorrect, true or false, *et cetera*. However, in the absence of correctness in the answer of a given mathematics question, does it immediately imply that it is incorrect? In other words, is incorrectness in mathematics the same as being wrong in mathematics? Similarly, how does inaccuracy in mathematics fit into the discussion of incorrectness and wrongfulness?

Topic 1 in the Higher Level programme introduces the notion of proofs, and in particular, proof by mathematical induction. Indeed, mathematical proofs are essential for new conjectures to be proven and their validity accepted. However, is it right to claim that we gain new knowledge in mathematics if a given proof is mathematically valid? If so, then is it necessary for everything in mathematics to be proven true first before one can use it? If not, then how do we distinguish those concepts which are infallible without proofs and those which are proven true, subject to the validity of the proof? More importantly, the foremost critical question is to ask what is considered to be the validity of a proof? If a proof is shown to be true, then does it automatically infer its validity?

One important aspect of mathematical proofs is to provide generalizations of a result. The process of moving from specific results to a generalization is definitely an art. However, have you ever wondered about the potential risk in this process of generalization? How rigourous does one need to be in order to ensure the generalization is not over simplifying the result? If a phenomenon exists with absolute certainty, then what parameters must be established before it could be generalized with symbols and axioms?

Similar to other subject areas, even when a certain knowledge claim has been proven to be true today, no one can guarantee its validity will withstand challenges through future times. Even though it only takes one example to disprove a certain conjecture or theorem, it involves more effort and time than one could ever imagine. Take geometry as an example, what is the shortest distance between two distinct points? In most primary and secondary mathematics classes, the shortest distance between two distinct points is a straight path connecting them; and this is certainly true according to Euclidean geometry. This is Euclid's fifth postulate which dates back to 330 BC.

If a straight line crossing two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if extended indefinitely, meet on that side on which are the angles less than the two right angles.

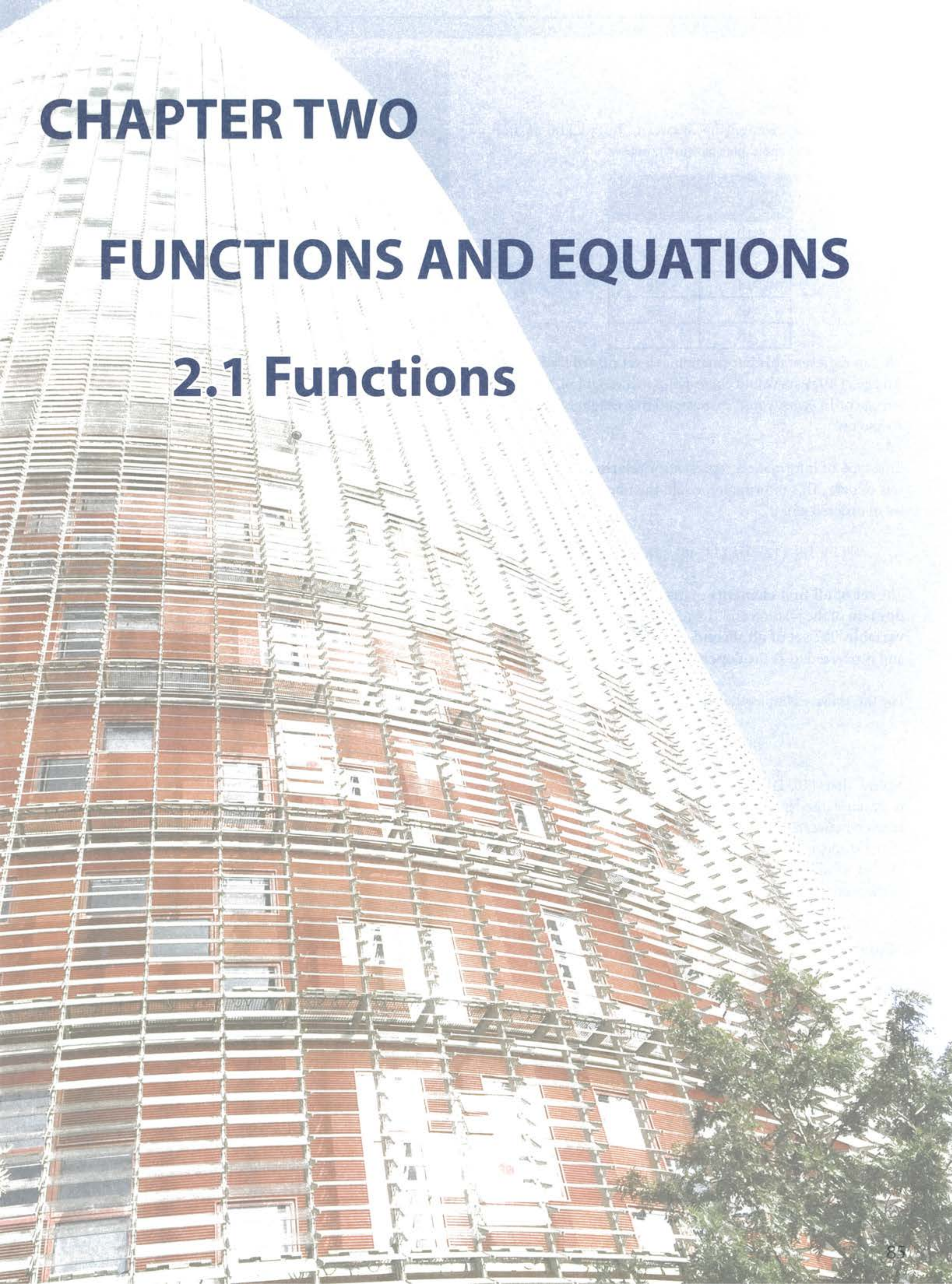
Euclidean geometry remained unchallenged until the early 19th century when non-Euclidean geometric concepts started to emerge. In Riemannian Geometry and in hyperbolic geometry, the shortest distance between two points is no longer necessarily a straight path. It is often not until the final year of secondary education or even in tertiary education when these challenges are presented and begin to question the first set of knowledge. When situations like this emerge in mathematics (or even just within secondary mathematics education), does it suggest that the content of primary and secondary mathematics is inadequate to enable students to appreciate the fullness of the discipline? Similarly, how incorrect was the first introduction of a straight path being the shortest distance in the classroom? Have you ever questioned whether or not it is acceptable to present some not entirely true statement for the sake of simplifying a complex discussion in mathematics?

Many students and the general public often see mathematics as a subject which has its strength in the provision of absolute certainty. However, in statistics the results are also presented with a tolerance level of uncertainty. How does it affect the validity of the result when it may not necessarily be certain? Conversely, if a given result is presented with 100% certainty, does it mean it is less valid without a certain level of deviation and significance level?

Imaginary Numbers

An imaginary number is a complex number with a real number and an imaginary unit. However, does the term imaginary number suggest that it is merely just an invention to satisfy the desire of mathematicians? In other words, when the given limits for real numbers do not address the additional mathematical concepts, have the mathematicians created this new concept to fill in the gap? The concept of imaginary numbers was not widely used and adopted until the 18th century by Leonhard Euler and Carl Gauss. When a mathematician cannot find known concepts and existing knowledge to address a new mathematical phenomenon, does it give him/her an automatic pass to create new sets of rules to govern his/her findings?

When there is a need to define new number systems, new mathematical rules and theorems, or new methods in approaching emerging topics in mathematics, does it imply that the existing set of axioms is obsolete and inadequate to meet the new demands? If it is necessary for newer rules, is it better to simply create new ones as extensions of the current system, or is it better to start from scratch and disregard all existing rules? If the mathematical field continues to build extensions from the existing set of axioms and rules to facilitate new findings, will there become a time when it becomes impossible to extend any further?



CHAPTER TWO

FUNCTIONS AND EQUATIONS

2.1 Functions

Relations

Consider the relationship between the weight of five students and their ages as shown below.

Age (years)	Weight (kg)
10	31
12	36
14	48
16	53
18	65

We can represent this information as a **set of ordered pairs**. An age of 10 years would correspond to a weight of 31 kg. An age of 16 years would correspond to a weight of 53 kg and so on.

This type of information represents a **relation** between two sets of data. This information could then be represented as a set of ordered pairs,

$$\{(10, 31), (12, 36), (14, 48), (16, 53), (18, 65)\}$$

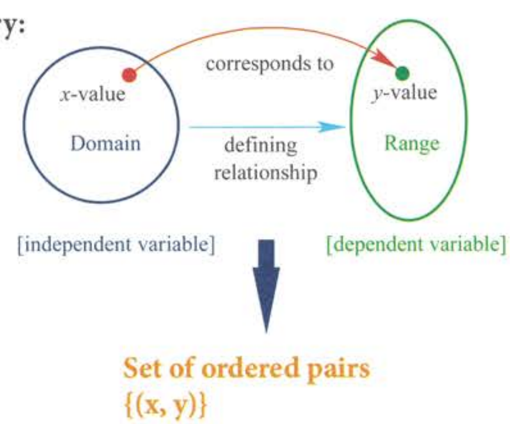
The **set of all first elements** of the ordered pair is called the **domain** of the relation and is referred to as the **independent variable**. The **set of all second elements** is called the **range** and is referred to as the **dependent variable**.

For the above example, the domain= $\{10, 12, 14, 16, 18\}$

and the range = $\{31, 36, 48, 53, 65\}$.

Notice that $(10, 31)$ and $(31,10)$ are not the same! This is because the ordered pair $(10, 31)$ provides the correct relation between age and weight, i.e. at age 10 years the weight of the student is 31 kg. On the other hand, the ordered pair $(31,10)$ would be informing us that at age 31 years the weight of the student is 10 kg!

Summary:



Example 2.1.1

Determine the domain and range for each of the following relations:

a $\{(0, 0) (1, 1), (2, 4), (3, 9), 4, 16), (5, 25)\}$

b $\{(-3, 4), (-1, 0), (2, -2), (-2, 2)\}$.

a The domain is the set of all first elements, i.e. $\{0, 1, 2, 3, 4, 5\}$.

The range is the set of all second elements, i.e. $\{0, 1, 4, 9, 16, 25\}$.

b The domain is the set of all first elements, i.e. $\{-3, -1, 2, -2\}$.

The range is the set of all second elements, i.e. $\{4, 0, -2, 2\}$

The letter "X" is often used to denote the domain and the letter "Y" to denote the range. For part a this means that we could write X = $\{0, 1, 2, 3, 4, 5\}$ and Y = $\{0, 1, 4, 9, 16, 25\}$ and for part b we could write X = $\{-3, -1, 2, -2\}$ and Y = $\{4, 0, -2, 2\}$.

This is a convention, nothing more.

Rather than giving a verbal description of how the independent variable and the dependent variable are related, it is much clearer to provide a **mathematical rule** that shows how the elements in the range relate to the elements in the domain.

Example 2.1.2

A relation is defined by the rule .

a Determine the range of this relation.

b Express this relation as a set of ordered pairs.

a The domain of this relation is given by the x-values, i.e. $\{0, 1, 2, 3, 4\}$. We can therefore substitute these values into the equation $y = x + 2$ and determine their corresponding y-values. This will provide the range of the relation.

Substituting we have, $x = 0 \Rightarrow y = 0 + 2 = 2$

$x = 1 \Rightarrow y = 1 + 2 = 3$

$x = 2 \Rightarrow y = 2 + 2 = 4$, and so on.

This produces a set of y-values $\{2, 3, 4, 5, 6\}$ that defines the range.

b The set of ordered pairs would be $\{(0, 2), (1, 3), (2, 4), (3, 5), (4, 6)\}$.

Notice that we can describe the set of ordered pairs more formally as:

$\{(x,y): y = x + 2, x \in \{0, 1, 2, 3, 4\}\}$.

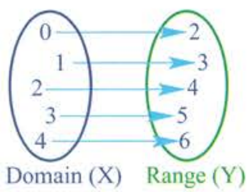
which is read as:

"The set of ordered pairs x and y, such that $y = x + 2$, where x is an element of the set of values $\{0, 1, 2, 3, 4\}$."

The information in Example 2.1.2 can be displayed in different ways. Both those shown are *visual* displays – they show the mappings in different ways.

Mapping diagram

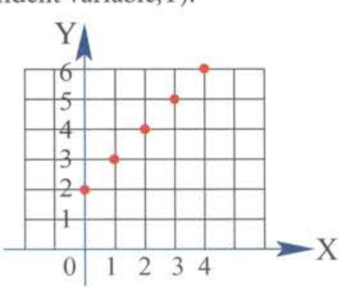
The mapping diagram below displays which y-value corresponds to a given x-value.



However it is often not easy to see the 'pattern' between the variables with this style of diagram.

Cartesian plane

The Cartesian plane is made up of a horizontal axis (independent variable, X) and a vertical axis (dependent variable, Y).



We plot the points on the grid, so that (3, 5) is 3 units to the right and 5 units up.

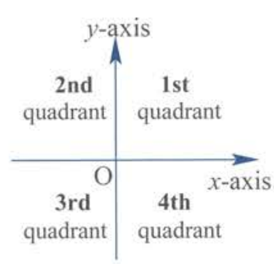
Notice that in the mapping diagram that uses the Cartesian plane, we have not joined the points together in a straight line. This is because the domain specifies that the only values of x that can be used must be from the set $\{0, 1, 2, 3, 4\}$, and so a value such as $x = 2.4$ cannot be used.

Both these visual representations are useful in displaying which values in the domain generate a given value in the range. However, the Cartesian plane more readily gives a quick overview of what the underlying relationship between the two variables is. It is very easy (and quick) to see that as the x-values increase, so too do the y-values. We can do this by simply looking at the points on the graph and observing the 'trend' without really concerning ourselves with what the actual values are.

We now provide a formal definition of the Cartesian plane and a relation.

The Cartesian Plane

The Cartesian plane (named after Rene Descartes - see picture) is formed by constructing two real lines that intersect at a right-angle where the point of intersection of these two lines becomes the origin. The horizontal real line is usually referred to as the x-axis and the vertical real line is usually called the y-axis. This also implies that the plane has been divided into four quadrants. Each point on this plane is represented by an ordered pair (x,y) where x and y are real numbers and are the coordinates of the point.



The set of all ordered pairs (x, y), where $x \in X$ and $y \in Y$ can also be defined by making use of the Cartesian product,

$$X \times Y = \{(x,y): x \in X, y \in Y\}$$

Implied domain

So far we have looked at examples for which a domain has been specified. Suppose we were asked to find the range of the relation $y = 1 + x^2, x \geq 3$? After sketching its graph, we would determine its range to be $[10, \infty)$. However, what if we wanted to know the range of the relation $y = 1 + x^2$? In this case, because we have not been provided with any restriction on the x-values, we will need to assume that we can **use the largest possible set of x-values for which the relation is defined** – this domain is known as the **implied domain** (or **maximal domain**) – in this case that would be the real number set, $\mathbb{R}$. Then, after sketching the graph of $y = 1 + x^2$ for all real values of x we would have a range defined by $[1,\infty)$.

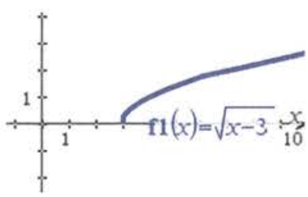
Example 2.1.3

Determine the domain and range of the following relations:

a $y = \sqrt{x-3}$ b $y = \dfrac{2}{\sqrt{x-3}}$

c $y = \dfrac{3}{2-x}$

a Using a calculator to sketch the graph of $y = \sqrt{x-3}$ (i.e. the square root relation) we observe that its domain is $[3,\infty)$.



Now, let's take a closer look at why that is the case.

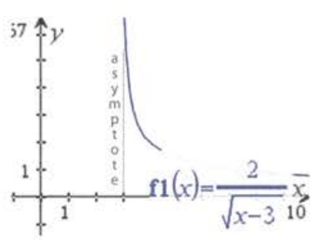
Because we are dealing with an expression that involves a square root, then, the term 'inside' the square root (radicand) must be greater than or equal to zero (as we cannot take the square root of a negative number).

So, we must have that $x - 3 \geq 0 \Leftrightarrow x \geq 3$. Therefore, the implied domain is $\{x: x \geq 3\}$.

From the graph, the range can be seen to be $[0,\infty)$.

It should be noted that the TI–83 uses the implied domain when graphing. Also realise that from the sketch, we could be misled into thinking that there is a 'gap' at the point (3, 0). Be careful with this – use the graphics calculator as an aid, then, double check to make sure.

b The equation: $y = \dfrac{2}{\sqrt{x-3}}$ represents the reciprocal of a square root relation.

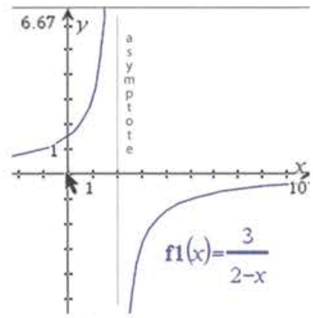


As in part a, we must have that $x - 3 \geq 0 \Leftrightarrow x \geq 3$.

However, this time we have another restriction – we cannot divide by zero and so we cannot include $x = 3$ in our domain. So, at $x = 3$, we draw an **asymptote**.

We then have $x - 3 > 0 \Leftrightarrow x > 3$. This leads to a range of $(0,\infty)$ (or $]0,\infty[$).

c The only restriction that can be readily seen for the relation $y = \dfrac{3}{2-x}$ is that we cannot divide by zero and so, we must have that $2 - x \neq 0$. That is, $x \neq 2$.



As it is a reciprocal relation, we have an asymptote at $x = 2$. So, the domain is given by $]-\infty,2[\cup]2,\infty$ or simply, $\mathbb{R}\backslash\{2\}$.

The range can then be seen to be $\mathbb{R}\backslash\{0\}$.

Exercise 2.1.1

1 State the domain and range of the following relations.

a $\{(2,4), (3,-9), (-2,4), (3,9)\}$

b $\{(1,2), (2,3), (3,4), (5,6), (7,8), (9,10)\}$

c $\{(0,1), (0, 2), (1,1), (1,2)\}$

2 . Find the range for each of the following.

a $\{(x,y): y = x + 1, x \in \mathbb{R}^+\}$

b $\{(x,y): y \geq x, x \geq 0\}$

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

d $y = 2x - x^2, x \in \mathbb{R}$

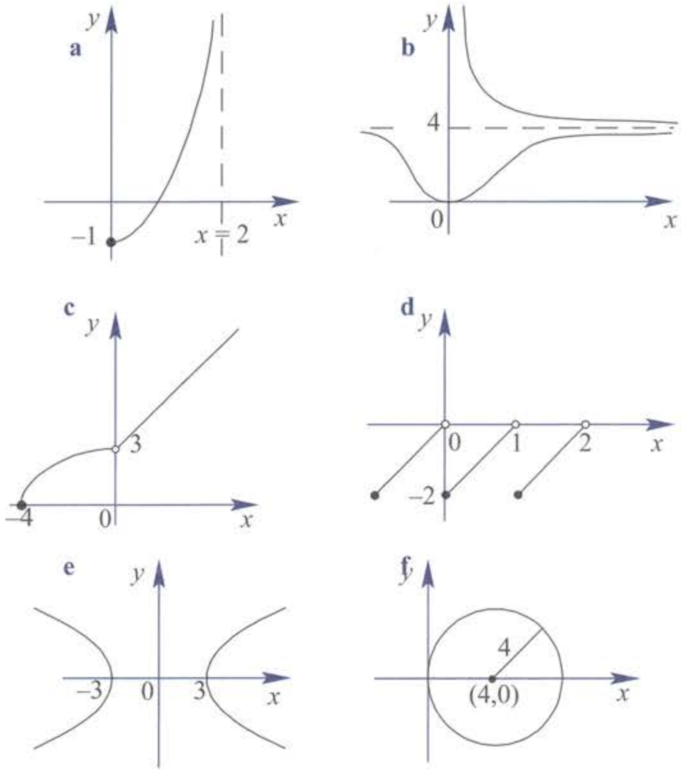
e $x^2 + y^2 = 9, -3 \leq x \leq 3$

f $x^2 - y^2 = 9, x \geq 3$

g $y = x - 1, 0 < x \leq 1$

h $y = 4 - x^2, -2 \leq x < 1$

3. State the range and domain for each of the following relations.



4. Determine the implied domain for each of the following relations.

a $y = \dfrac{2x}{x+2}$ b $y = \dfrac{3}{\sqrt{9-x}}$

c $y = \sqrt{16-x^2}$ d $y = \sqrt{x^2-4}$

e $xy - x = 3$ f $y = \dfrac{2}{x^2+1}$

g $y = \dfrac{2}{x^3+1}$

5. Find the range of the following relations.

a $y = x - a, x < 0, a > 0$

b $y = \dfrac{ab}{x+1}, x \geq 0, ab > 0$

c $y = a^2x - ax^2, x \geq \dfrac{1}{2}a, a > 0$

d $y = a^2x - ax^2, x \geq \dfrac{1}{2}a, a < 0$

e $y = \dfrac{a}{x} + a, a > 0$

Extra questions

Functions

There is a special group of relations which are known as **functions**. This means that every set of ordered pairs is a relation, but **every relation is not a function**. Functions then make up a subset of all relations.

A function is defined as a relation such that each domain element has a unique image in the codomain. That is a function is a relation for which no ordered pairs have the same first element.

For example, the function 'Take a real number, double it and add one' is commonly expressed in mathematica not ation as:

$$f(x) = 2x+1, x \in \mathbb{R} \text{ or } f : \mathbb{R} \mapsto \mathbb{R}, f(x) = 2x+1$$

When you use a function on a calculator (such as x^2) you get a single answer - not a choice. This is why mathematicians like functions - they remove doubt!

There are two ways to determine if a relation is a function.

Method 1: Algebraic approach

For Method 1 we use the given equation and determine the number of y-values that can be generated from one x-value.

Example 2.1.4

Determine which (if any) of the following are functions.

a $y^3 - x = 2$ b $y^2 + x = 2$

a From $y^3 - x = 2$, we have $y = \sqrt[3]{2+x}$, then for any given value of x, say $x = a$, we have that $y = \sqrt[3]{2+a}$ which will only ever produce one unique y-value.

Therefore, the relation $y^3 - x = 2$ is a function. In fact, it is a one-to-one function.

b From $y^2 + x = 2$, we have:
$y^2 = 2 - x \Leftrightarrow y = \pm\sqrt{2-x}$.

Then, for any given value of x, say $x = a$ (where $a \leq 2$), we have that $y = \pm\sqrt{2-a}$, meaning that we have two different y-values; $y_1 = \sqrt{2-a}$ and $y_2 = -\sqrt{2-a}$, for the same x-value.

Therefore, this relation is not a function.

Method 2: Vertical line test

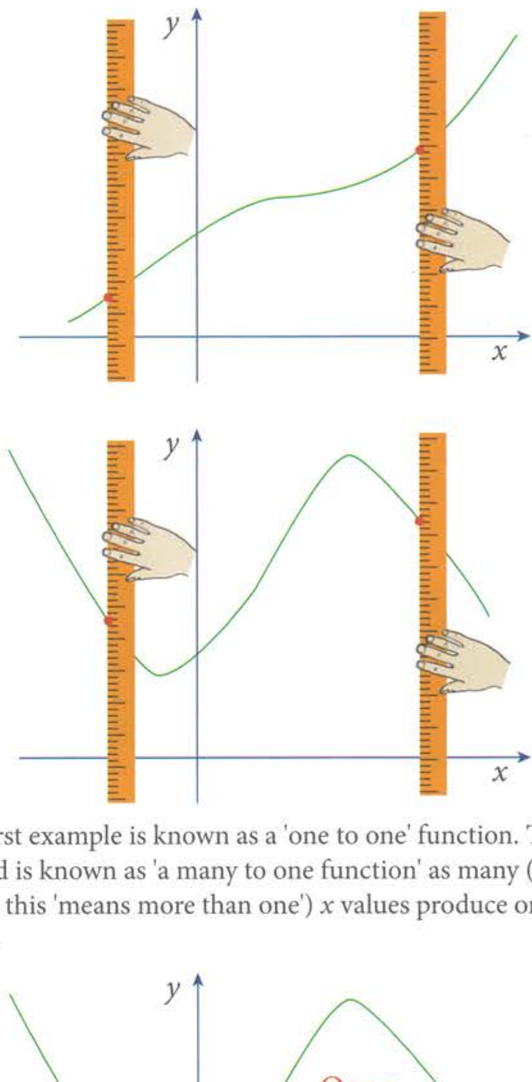
Step 1: Sketch the graph of the relation.

Step 2: Make a visual check of the number of times a vertical line would cut the graph.

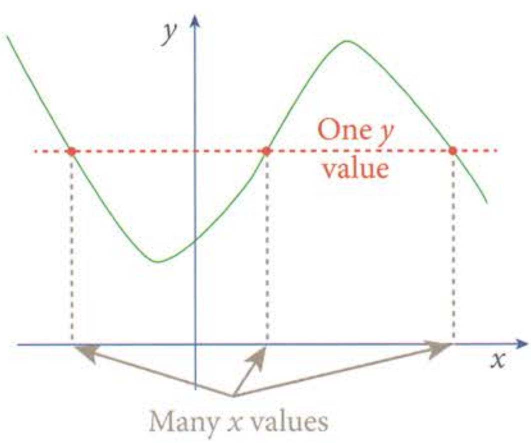
Step 3: If the vertical line only ever cuts at one place for every value in the domain the relation is a function.

In both these examples, any vertical line (ruler) cuts the graph in at most one place.

Vertical Line Test



The first example is known as a 'one to one' function. The second is known as 'a many to one function' as many (in maths this 'means more than one') x values produce one y value.



Example 2.1.5

Which of the following defines a function?

a $\{(0,2), (1,2), (2,1)\}$

b $\{(x,y): y=x^3+1, x\in\mathbb{R}\}$

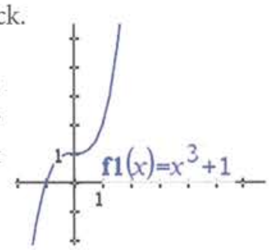
c $y^2=x, x\geq 0$

d $\{(x,y): x^2+y^2=16\}$

a Clearly, we have every first element of the ordered pairs different.

This means that this relation is also a function:

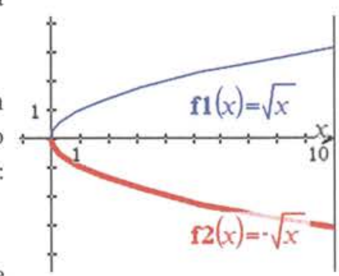
b A graph provides a visual check.



From the graph shown, a vertical line drawn anywhere on the domain for which the relation is defined, will cut the graph at only one place.

This relation is therefore a function.

c Again we make use of a visual approach to determine if the relation is a function.



First we write the relation in a form that will enable us to enter it into the calculator: $y^2 = x \Rightarrow y = \pm\sqrt{x}$

We can therefore define the relation $Y_1 = \sqrt{X}$ and $Y_2 = -\sqrt{X}$ and sketch both on the same set of axes.

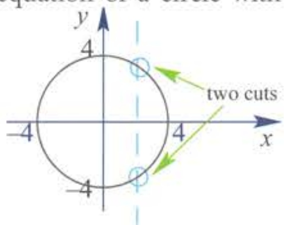
Placing a vertical line over sections of the domain shows that the line cuts the graph in two places (except at the origin). Therefore this relation is not a function.

Algebraic proof

We can also determine if a relation is a function by using algebraic means. Begin by choosing a value of x that lies in the domain. For example $x = 4$. This gives the following equation: $y^2 = 4 \Rightarrow y = \pm\sqrt{4}$.

From which we can say that when $x = 4$, $y = 2$ **and** $y = -2$, so that there are two ordered pairs, (4, 2) and (4, −2). As we have two different y-values for one x-value this relation is not a function.

d This relation describes the equation of a circle with radius 4 units and centre at the origin. The graph of this relation is shown alongside. The graph fails the vertical line test, and so is not a function.



Exercise 2.1.2

1. A function is defined as follows, $f{:}x \mapsto 2x + 3, x \geq 0$.

a Find the value of $f(0), f(1)$.

b Evaluate the expressions: i $f(x+a)$
ii $f(x+a) - f(x)$

c Find $\{x{:}f(x) = 9\}$.

2. If $f(x) = \dfrac{x}{x+1}, x \in [0, 10]$, find
:

a $f(0), f(10)$ b $\{x{:}f(x) = 5\}$

c the range of $f(x) = \dfrac{x}{x+1}, x \in [0, 10]$.

3. For the mapping $x \mapsto 2 - \frac{1}{2}x^2, x \in \mathbb{R}$, find:

a $f(x+1)$, $f(x-1)$ b a, given that $f(a) = 1$

c b, given that $f(b) = 10$.

4. A function is defined as, $y = x^3 - x^2, x \in [-2, 2]$

a Find the value(s) of x such that $y = 0$.

b Sketch the graph of $y = x^3 - x^2, x \in [-2, 2]$ and determine its range.

5. The function f is defined as $f{:}]-\infty, \infty[\mapsto \mathbb{R}$,where $f(x) = x^2 - 4$.

a Sketch the graph of:

i f ii $y = x + 2, x \in]-\infty, \infty[$

b Find: i $\{x{:}f(x) = 4\}$

ii $\{x{:}f(x) = x + 2\}$

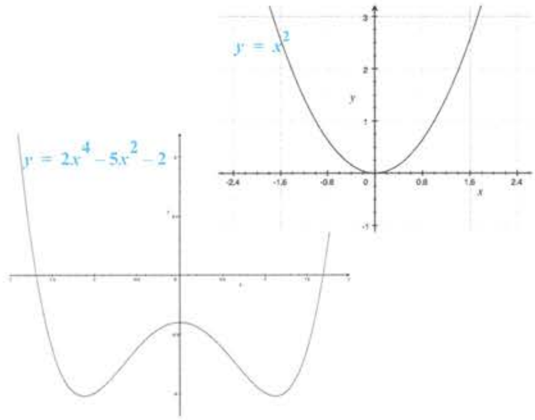
Extra questions

Odd and Even Functions

Functions can be classified into three categories based on the symmetries of their graphs.

Even functions

If a function has line symmetry about the y-axis, it is said to be even.



Formally, a function is even if $f(x) = f(-x)$ for all x in the domain.

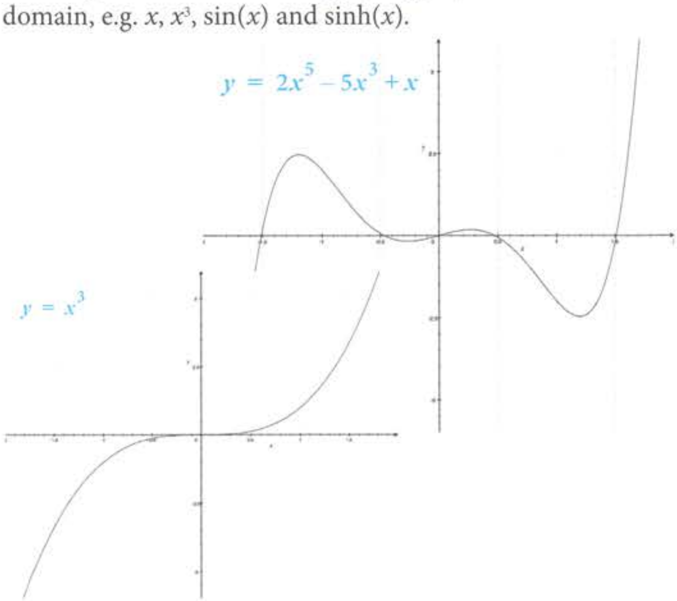
The most obvious examples of even functions are the even polynomials

, e.g. $|x|$, x^2, x^4, $\cos(x)$ and $\cosh(x)$.

Odd functions

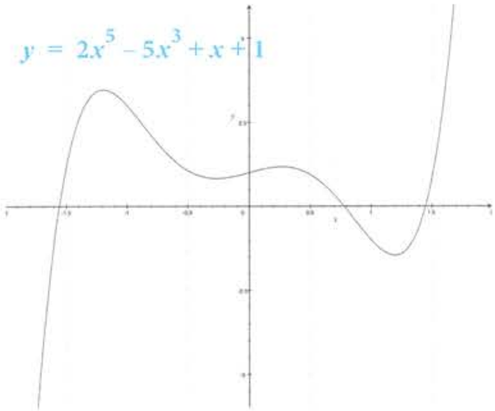
A function is odd if it has two-fold rotational symmetry about the origin. If the graph is 'pinned' at the origin and turned through 180°, it will fit back over the original graph.

Formally, a function is odd if $f(x) = -f(-x)$ for all x in the domain, e.g. x, x^3, $\sin(x)$ and $\sinh(x)$.



The concept applies to other, non-polynomial functions. The cosine function is even, the sine function is odd and the logarithm function is neither.

Not every function is either odd or even. Most functions are neither.

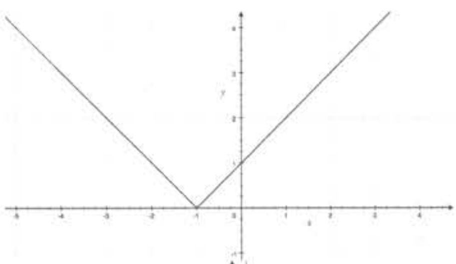


Example 2.1.6

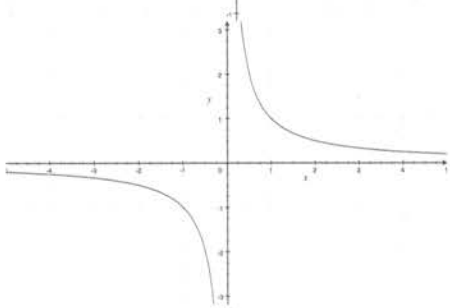
Classify the following functions as even, odd or neither.

a $f(x) = |x+1|$ b $f(x) = \dfrac{1}{x}$ c $f(x) = \left|\dfrac{1}{x}\right|$

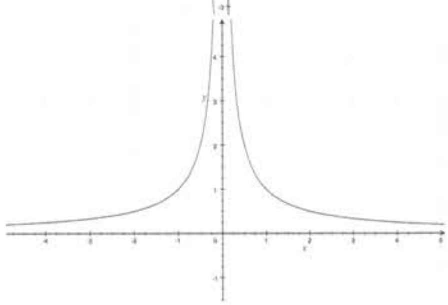
a Neither odd nor even.



b Odd



c Even



Exercise 2.1.3

1. Classify the following functions as even, odd or neither.

   a $y = 4$ b $y = |x^3|$

   c $y = (x-1)^3$ d $y = \dfrac{x}{x-2}$

   e $y = \ln(x^2)$ f $y = \dfrac{1}{x}$

   g $y = \dfrac{x^2+2}{x^3+x}$ h $y = \dfrac{x}{x^3-x}$

   i $y = \dfrac{\sin x}{\cos x} + x^3$

2. Prove that the product of two even functions is even.

3. Is it necessary for $f(0)=0$ for a function to be odd?

4. Explain why the composite of two odd functions is odd.

5. Prove that the quotient of two even functions is even.

6. A function has the full real line as its domain and is both odd and even. What is the rule for the function?

Basic operations and composite functions

For any two real functions f: $d_f \mapsto \mathbb{R}, y = f(x)$ and g: $d_g \mapsto \mathbb{R}, y = g(x)$, defined over domains d_f and d_g respectively, the following rules of algebra apply:

Addition

$(f+g)(x) = f(x) + g(x), d_{f+g} = d_f \cap d_g$ for example:

$f(x) = \sqrt{x}, x \geq 0$ and $g(x) = x, x \in \mathbb{R}$

$(f+g)(x) = \sqrt{x} + x, x \geq 0$

The domain of the sum is the intersection of the two functions that make up the sum - both functions have to 'work' for the sum to exist.

Subtraction

$(f-g)(x) = f(x) - g(x), d_{f-g} = d_f \cap d_g$

$f(x) = \sqrt{x}, x \geq 0$ and $g(x) = x, x \in \mathbb{R}$

$(f-g)(x) = \sqrt{x} - x, x \geq 0$

Multiplication

$(f \times g)(x) = f(x) \times g(x), d_{f \times g} = d_f \cap d_g$

$f(x) = \sqrt{x}, x \geq 0$ and $g(x) = x, x \in \mathbb{R}$

$(f \times g)(x) = \sqrt{x} \times x, x \geq 0$

Division

$(f \div g)(x) = f(x) \div g(x)$

The basic domain is, as before: $d_{f \div g} = d_f \cap d_g$. However, we must also exclude cases in which we would be dividing by zero. Thus, values in the domain for which $g(x) = 0$ must also be excluded.

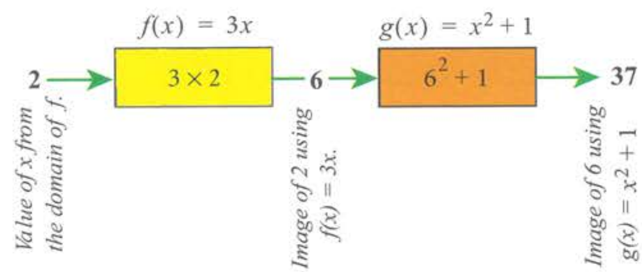
For example: $h(x) = \sqrt{x+1}, x \geq -1$ and $i(x) = x, x \in \mathbb{R}$

$$(g \div i)(x) = \frac{\sqrt{x+1}}{x} \times x, x \geq -1, x \neq 0$$

Composite functions

We now investigate another way in which we can combine functions, namely **composition**.

Consider the two functions $f(x) = 3x$ and $g(x) = x^2 + 1$. Observe what happens to the value $x = 2$ as we first apply the function $f(x)$ and then the function $g(x)$ to the image of the first mapping, i.e.



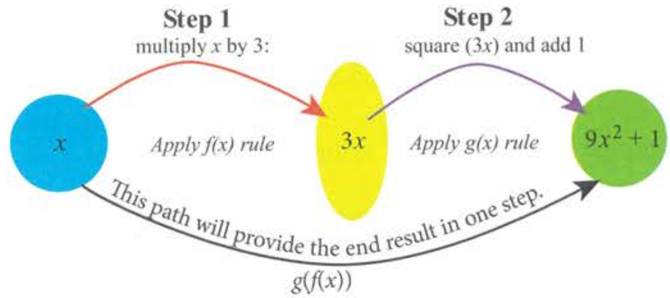
Such a combination of functions leads to the question "Is there a third function that will enable us to produce the same result in one step?"

We consider any value x that belongs to the domain of f and follow 'its path':

1 This value of x, has as its image the value $f(x) = 3x$.

2 The resulting number, $3x$, now represents an element of the domain of g.

3 The image of $3x$ under the mapping g is given by $g(3x) = (3x)^2 + 1 = 9x^2 + 1$.



We can now test our result by using the value of $x = 2$ with the mapping $x \mapsto 9x^2 + 1$.

For $x = 2$, we have $9(2)^2 + 1 = 9 \times 4 + 1 = 37$, which agrees with our previous result.

The two **critical steps** in this process are:

1 That the image under the first mapping must belong to the domain of the second mapping.

2 The expression $g(f(x))$ exists.

Notation

The expression $g(f(x))$ is called the composite function of f and g and is denoted by gof.

Although f is applied first, it is placed second in $g \circ f$.

Example 2.1.7

Given the functions $f(x) = x^2 + 1$ and $g(x) = \ln(x-1)$, find the composite function $(gof)(x)$.

The composite function $(gof)(x) = g(f(x))$

$= \ln(f(x) - 1) = \ln(x^2 + 1 - 1) = \ln x^2$

Example 2.1.8

Given the functions $f(x) = 2 - x$ and $g(x) = \sqrt{x-1}$, find the composite function $(gof)(x)$.

The composite function $(gof)(x) = g(f(x))$

$= \sqrt{f(x) - 1} = \sqrt{(2-x) - 1} = \sqrt{1-x}$

In the previous example,. we have:

$(g \circ f)(-1) = \sqrt{1-(-1)} = \sqrt{2}$ but

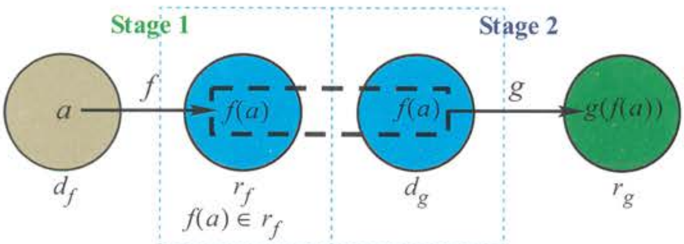
$(g \circ f)(2) = \sqrt{1-2} = \sqrt{-1}$ is undefined!

So what went wrong? To answer this question let's take another look at the composition process.

The process is made up of two stages:

Stage 1: An element from the domain of the first function, $f(x)$ is used to produce an image. That is, using $x = a$ we produce the image $f(a)$.

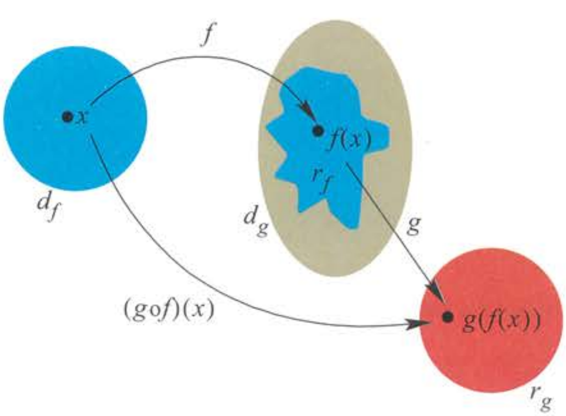
Stage 2: Using the second function, $g(x)$, the image, $f(a)$, is used to produce a second image $g(f(a))$.



From the diagram, the result of stage 1 is $f(a)$ (which belongs to the range of f) we also observe that at stage 2, when using the value $f(a)$ (produced from stage 1) we have assumed that $f(a)$ belongs to the domain of $g(x)$. This is where problems can arise – as seen in Example 2.1.8.

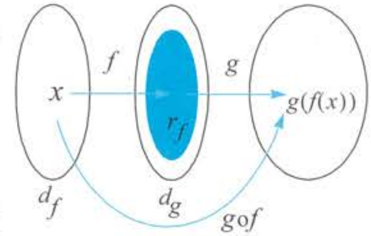
To overcome this difficulty we need to strengthen our definition of composition of functions as well as ensuring the existence of composite functions.

What we need to prove is that all values produced from stage 1, i.e. $f(a)$, are values in the domain of the function in stage 2, i.e. $f(a) \in d_g$. Making use of a mapping diagram, we show the inter-relation between the range of f, r_f, and the domain of g, d_g.



For $(g \circ f)(x) = g(f(x))$ to exist, then $r_f \subseteq d_g$.

What is the domain of $g \circ f$?



If we refer to the diagram alongside, we see that

if $r_f \subseteq d_g$, then $d_{g \circ f} = d_f$.

This means that we can substitute values of x that belong to the domain of f directly into the expression $g(f(x))$ (once we have established that it exists).

Example 2.1.9

If $f(x) = \sqrt{x+1}, x \in (0,\infty)$ and $g(x) = x^3, x \in \mathbb{R}$, determine if $g \circ f$ exists, and find an expression for $g \circ f$ if it does.

For $g \circ f$ to exist we must have that $r_f \subseteq d_g$.

Using the TI–83 we obtain the range of f from its sketch, in this case, $r_f = (1,\infty)$.

The domain of g is $(-\infty,\infty)$ (i.e. the real field).

Then, given that $(1,\infty) \subseteq (-\infty,\infty)$, $g \circ f$ does exist.

We are now able to determine $g \circ f$.

First we determine the equation $g(f(x))$:

$g(f(x)) = g(\sqrt{x+1}) = (\sqrt{x+1})^3 = (x+1)^{3/2}$.

Next we need the domain of $g \circ f$.

As we have seen, $d_{g \circ f} = d_f$, $\therefore d_{g \circ f} = (0,\infty)$.

Therefore, $g \circ f : (0,\infty) \mapsto \mathbb{R}$, $(g \circ f)(x) = (x+1)^{3/2}$.

Hint on setting out

When solving problems that involve the use of composition, it is useful to set up a **domain–range table** in order to help us determine the existence of the composition. Such a table includes information about the domain and range of both the functions under consideration:

	domain	range
f	d_f	r_f
g	d_g	r_g

The existence of $g \circ f$ can then be established by looking at r_f and d_g. Similarly, the existence of $f \circ g$ can be established by comparing r_g and d_f.

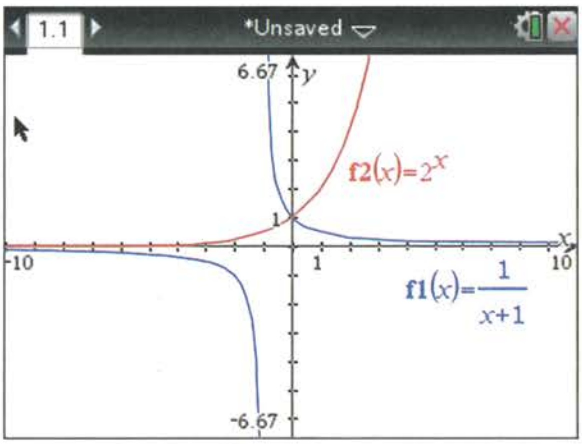
Example 2.1.10

Find $g \circ f$ and its range, given that:

$$g(x) = \frac{1}{x+1}, x \in \mathbb{R} \setminus \{-1\} \text{ and } f(x) = 2^x, x \in \mathbb{R} .$$

We first sketch the graphs of both functions to help us complete the domain–range table:

$$g(x) = \frac{1}{x+1}, x \in \mathbb{R} \setminus \{-1\}, \quad f(x) = 2^x, x \in \mathbb{R}.$$



We now complete the table:

	domain	range
f	$\mathbb{R}$	$]0, \infty[$
g	$\mathbb{R} \setminus \{-1\}$	$\mathbb{R} \setminus \{0\}$

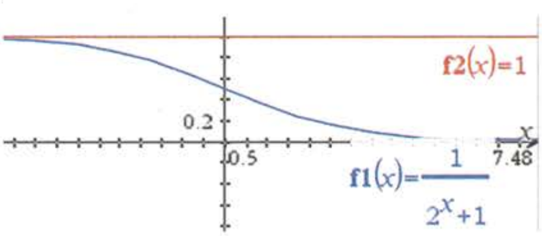
Using the table we see that $r_f \subseteq d_g \Rightarrow g \circ f$ exists.

We can now determine $g \circ f$: $g(f(x)) = \dfrac{1}{f(x)+1} = \dfrac{1}{2^x+1}$.

We also have that $d_{g \circ f} = d_f$, $\therefore d_{g \circ f} = \mathbb{R}$.

Therefore, $g \circ f : \mathbb{R} \mapsto \mathbb{R}$, where $(g \circ f)(x) = \dfrac{1}{2^x+1}$.

Making use of a calculator we see that the range of $g \circ f$ is $]0, 1[$.



Example 2.1.11

Given that $f(x) = \sqrt{x-1}$ and $g(x) = \ln(x)$. Does $f \circ g$ exist? If so, define fully the function $f \circ g$. If not, find a suitable restriction on the domain of g so that $f \circ g$ exists.

For $f \circ g$ to exist it is necessary that $r_g \subseteq d_f$. To determine the range of g we need to know the domain of g. Using the implied domain we have that $d_g =]0,\infty[$ and so, $r_g = \mathbb{R}$.

However, the implied domain of f is $[1, \infty[$. Then, as $r_g \not\subset d_f$, $f \circ g$ does not exist.

In order that $f \circ g$ exists we need to have $rg \subseteq [1,\infty[$, i.e. we must have that $g(x) \geq 1$. What remains then, is to find those values of x such that $g(x) \geq 1$.

Now, $g(x) = \ln x$ therefore, $g(x) \geq 1 \Leftrightarrow \ln x \geq 1 \Leftrightarrow x \geq e$.

So, if the domain of g is restricted to $[e, \infty[$ or any subset of $[e, \infty[$, then $f \circ g$ will exist.

Does *gof* = *fog*?

In general the answer is **no**! However, there exist situations when $(f \circ g)(x) = (g \circ f)(x)$ – we will look at such cases in the next section.

Consider Example 2.1.10, where $g(x) = \dfrac{1}{x+1}, x \in \mathbb{R} \setminus \{-1\}$ and $f(x) = 2^x, x \in \mathbb{R}$.

From our previous working, we have that $(g \circ f)(x) = \dfrac{1}{2^x+1}$.

To determine if $(f \circ g)(x)$ exists, we will need to determine if $r_g \subseteq d_f$. Using the domain–range table we have that $r_g = \mathbb{R} \setminus \{0\}$ and $d_f = \mathbb{R}$. Therefore as $r_g \subseteq d_f \Rightarrow f \circ g$ does exist.

We then have, $(f \circ g)(x) = f(g(x)) = 2^{g(x)} = 2^{\frac{1}{x+1}}$.

To determine the domain of $f \circ g$, we use the fact that, $d_{f \circ g} = d_g$ so that $d_g = \backslash\{\mathbb{R}1\}$.

Then, $f \circ g : \mathbb{R} \setminus \{-1\} \mapsto \mathbb{R}$, where $(f \circ g)(x) = 2^{\frac{1}{x+1}}$.

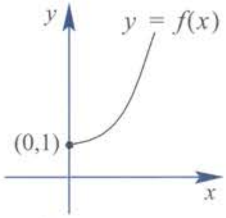
In this case, $(f \circ g)(x) \neq (g \circ f)(x)$.

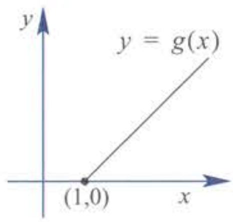
Example 2.1.12

Given $f(x)=x^2+1$ where $x\geq 0$ and $g(x)=x-1$ where $x\geq 0$, determine the functions $(f\circ g)(x)$ and $(g\circ f)(x)$ (if they exist). For the composite functions that exist, find the image of $x = 3$.

We first set up the domain–range table:

	domain	range
f	$[0, \infty[$	$[1, \infty[$
g	$[1, \infty[$	$[0, \infty[$





From the table we see that $r_g \subseteq d_f$ and $r_f \subseteq d_g$, and so, both $(f o g)(x)$ and $(g o f)(x)$ exist. We can now determine both composite functions.

We start with $(f o g)(x)$:

$(f o g)(x) = f(g(x)) = (g(x))^2 + 1$
$= (x-1)^2 + 1$

As $d_{fog} = d_g = [1, \infty[$, we have that:

$(f o g)(x) = (x-1)^2 + 1,$ where $x \geq 1$

Next, we find $(g o f)(x)$:

$(g o f)(x) = g(f(x)) = f(x) - 1 = (x^2+1) - 1 = x^2$

Also, $d_{gof} = d_f = [0, \infty[$, and so, $(g o f)(x) = x^2$, where $x \geq 0$

To find the image of 3, we substitute $x = 3$ into the final equations.

$(f o g)(3) = (3-1)^2 + 1 = 2^2 + 1 = 5$,

whereas, $(g o f)(3) = (3)^2 = 9$.

Exercise 2.1.4

1. Fully define the functions: a $f+g$ b fg given that :

   i $f(x) = x^2$ and $g(x) = \sqrt{x}$

   ii $f(x) = \ln x$ and $g(x) = \dfrac{1}{x}$

   iii $f(x) = \sqrt{9-x^2}$ and $g(x) = \sqrt{x^2-4}$

   Find the range for case a.

2. Fully define the functions: a $f-g$ b f/g (and find the range for case a) given that :

   i $f(x) = e^x$ and $g(x) = 1 - e^x$

   ii $f(x) = x+1$ and $g(x) = \sqrt{x+1}$

   iii $f(x) = |x-2|$ and $g(x) = |x+2|$

3. All of the following functions are mappings of $\mathbb{R} \mapsto \mathbb{R}$ unless otherwise stated.

   a Determine the composite functions $(f o g)(x)$ and $(g o f)(x)$, if they exist.

   b For the composite functions in part a that do exist, find their range.

   i $f(x) = x+1, g(x) = x^3$

   ii $f(x) = x^2+1, g(x) = \sqrt{x}, x \geq 0$

   iii $f(x) = (x+2)^2, g(x) = x-2$

   iv $f(x) = \dfrac{1}{x}, x \neq 0, g(x) = \dfrac{1}{x}, x \neq 0$

   v $f(x) = x^2, g(x) = \sqrt{x}, x \geq 0$

   vi $f(x) = x^2-1, g(x) = \dfrac{1}{x}, x \neq 0$

   vii $f(x) = \dfrac{1}{x}, x \neq 0, g(x) = \dfrac{1}{x^2}, x \neq 0$

4. Given the functions $f: x \mapsto 2x+1, x \in]-\infty,\infty[$ and $g: x \mapsto x+1, x \in]-\infty, \infty[$. Find the functions:

   a $(f o g)$ b $(g o f)$ c $(f o f)$

5. Determine the function g, given that $f: x \mapsto x+1, x \in \mathbb{R}$ and $g o f: x \mapsto x^2 + 2x + 2, x \in \mathbb{R}$.

6. The functions f and g are defined by $f: x \mapsto x+1, x \in \mathbb{R}$ and $g: x \mapsto x + \dfrac{1}{x}, x \in \mathbb{R}\backslash\{0\}$.

   Find the composite functions (where they exist) of the following, stating the range in each case.

   a $f o g$ b $g o f$ c $g o g$

7. If $g: x \mapsto x^3 + 1, x \in \mathbb{R}$ and $f: x \mapsto \sqrt{x}, x \in [0, \infty[$, find:

   a $(g o f)(4)$ b $(f o g)(2)$

8. Given that $f: x \mapsto x+5, x \in \mathbb{R}$ and $h: x \mapsto x-7, x \in \mathbb{R}$, show that $(f o h)(x)$ is equal to $(h o f)(x)$ for all $x \in \mathbb{R}$.

9. If $f(x)=4x+1$ and $g(f(x))=16x^2+8x+4$, find $g(x)$.

10. If $f(x)=2x+1$ and $g(f(x))=\frac{1}{2}\sqrt{4x^2+4x}+2$, find $g(x)$.

11. Solve the equation $(f\circ g)(x) = 0$, where:

    a $f{:}x \mapsto x+5, x \in \mathbb{R}$ and $g{:}x \mapsto x^2-6, x \in \mathbb{R}$.

    b $f{:}x \mapsto x^2-4, x \in \mathbb{R}$ and $g{:}x \mapsto x+1, x \in \mathbb{R}$.

12. Given that $f{:}x \mapsto 2x+1, x \in \mathbb{R}$, determine the two functions g, given that:

    a $(g\circ f)(x) = \frac{1}{2x+1}$

    b $(f\circ g)(x) = \frac{1}{2x+1}$

Extra questions

Identity and inverse functions

Before we start our discussion of inverse functions, it is worthwhile looking back at a fundamental area of algebra – algebraic operations. The relevant algebraic properties for real numbers $a \in \mathbb{R}$, $b \in \mathbb{R}$ and $c \in \mathbb{R}$ are:

	Under addition
Closure	$a+b \in \mathbb{R}$
Commutativity	$a+b = b+a$
Associativity	$(a+b)+c = a+(b+c)$
Existence of the identity	**0** : $a+0 = 0+a = a$
Inverse element	$-a$: $a+(-a) = 0 = (-a)+a$

	Under multiplication
Closure	$a \times b \in \mathbb{R}$
Commutativity	$a \times b = b \times a$
Associativity	$(a \times b) \times c = a \times (b \times c)$
Existence of the identity	$1 : a \times 1 = 1 \times a = a$
Inverse element	$\frac{1}{a} : \left(\frac{1}{a}\right) \times a = 1 = a \times \left(\frac{1}{a}\right)$

Just as there exists an **identity element** for addition, i.e. 0 and for multiplication, i.e. 1 under the real number system, it seems reasonable to assume that an identity element exists when dealing with functions.

It should be noted that without an identity element, equations such as $x+2 = 7$ and $2x = 10$ could not be solved under the real number system. Because we take the process of solving these equations for granted, sometimes we lose sight of the underpinning algebraic process that led to their solution.

For example, to solve $x+2 = 7$ we would write $x = 5$ as the next step. However, if we break the process down we have the following:

$x+2 = 7 \Leftrightarrow x+2+(-2) = 7+(-2)$ (Inverse element)

$\Leftrightarrow x+0 = 5$ (Identity)

$\Leftrightarrow x = 5$

So that without the identity element, we would not be able to make the last statement!

Consider the two functions $f : x \mapsto \mathbb{R}, \sqrt[3]{x}, x \in \mathbb{R}$ and $g : x \mapsto \mathbb{R}, x^3, x \in \mathbb{R}$. The composite functions $f\circ g$ and $g\circ f$ exist. The composite functions are then given by:

$$(f\circ g)(x) = f(g(x)) = \sqrt[3]{g(x)} = \sqrt[3]{x^3} = x$$

and

$$(g\circ f)(x) = g(f(x)) = (f(x))^3 = \left(\sqrt[3]{x}\right)^3 = x.$$

For this particular example we have the result that:

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

Using an analogy to the algebraic properties for the real number system, we introduce the identity function.

Identity function

We define the identity function, I, as $I(x) = x$

If an identity function, I, exists then it must have the property that for any given function f

$$f \circ I = I \circ f = f \text{ or } f(I(x)) = I(f(x)) = f(x)$$

As we have already seen, it is found that a function with the rule $I(x) = x$ has the required properties. However, unlike its counterpart in the real number system, where the identities are unique, the domain of the identity function is chosen to match that of the function f.

For example, if $f(x) = \sqrt{x}$, $x \geq 0$ then $I(x) = x$ where $x \geq 0$.

Whereas if $f(x) = 2x$, $x \in]-\infty,\infty[$ then $I(x) = x$ where $x \in]-\infty,\infty[$.

The existence of the identity function leads us to investigate the existence of an **inverse function**.

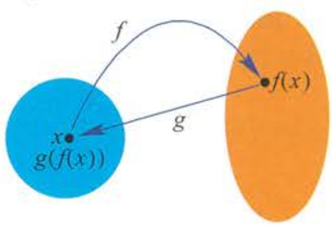
Inverse function

Using an analogy to the real number system, the concept of an inverse requires that given some function f, there exists an inverse function, f^{-1}, such that

$$f \circ f^{-1} = I = f^{-1} \circ f \text{ or } f(f^{-1}(x)) = x = f^{-1}(f(x))$$

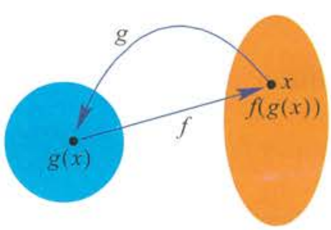
The '−1' used in f^{-1} should not be mistaken for an exponent, i.e. $f^{-1}(x) \neq \frac{1}{f(x)}$!

Looking back at the two functions $f : x \mapsto \mathbb{R}, \sqrt[3]{x}, x \in \mathbb{R}$ and $g : x \mapsto \mathbb{R}, x^3, x \in \mathbb{R}$ we notice that the function g is the reverse operation of function f, and function f is the reverse operation of function g. Making use of a mapping diagram we can 'visualise' the process:



1. Use an element (x) from the domain of the function f and obtain its image $f(x)$.

2. Using this image, which must be an element of the domain of g, we then apply g to $f(x)$ and obtain its image, $g(f(x))$ resulting in the value x that we started with.



1. Use an element (x) from the domain of the function g and obtain its image $g(x)$.

2. Using this image, which must be an element of the domain of f, we then apply f to $g(x)$ and obtain its image, $f(g(x))$ resulting in the value x that we started with.

We can now make some observations:

1. The domain of f, d_f, must equal the range of g, r_g and the domain of g, d_g must equal the range of f, r_f.

2. For the uniqueness of 'x' to be guaranteed both ***f* and *g* must be one-to-one functions.**

We therefore have the result:

If f and g are **one-to-one functions**, such that $f(g(x)) = x = g(f(x))$, then g is known as the inverse of f and f as the inverse of g.

In our case we write, $g(x) = f^{-1}(x)$ and $f(x) = g^{-1}(x)$. This then brings us back to the notation we first introduced for the inverse function. We summarise our findings:

For the inverse function of f, f^{-1} (read as f inverse) to exists, then:

1. f must be a one to one function.

2. i the domain of f is equal to the range of f^{-1}, i.e. $d_f = r_{f^{-1}}$

   ii the range of f is equal to the domain of f^{-1}, i.e. $d_{f^{-1}} = r_f$

3. $f(f^{-1}(x)) = x = f^{-1}(f(x))$.

How do we find the inverse function?

A guideline for determining the inverse function can be summarized as follows:

Step 1: Check that the function under investigation is a one-to-one function. This is best done by using a sketch of the function.

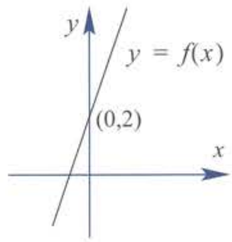
Step 2: Use the expression $f(f^{-1}(x)) = x$ to solve for $f^{-1}(x)$.

Step 3: Use the fact that $d_{f^{-1}} = r_f$, and $d_f = r_{f^{-1}}$ to complete the problem.

Example 2.1.13

Find the inverse of the function:
$f: x \mapsto 5x+2$, where $x \in \mathbb{R}$

We start by checking if the function is a one-to-one function.



From the graph, it is clearly the case that $f(x)$ is a one-to-one function.

Making use of the result that $f(f^{-1}(x)) = x$ to solve for $f^{-1}(x)$:

$f(f^{-1}(x)) = 5f^{-1}(x) + 2$

Then, $f(f^{-1}(x)) = x \Rightarrow 5f^{-1}(x) + 2 = x$

$$\Leftrightarrow \quad 5f^{-1}(x) = x - 2$$

$$\Leftrightarrow \quad f^{-1}(x) = \frac{1}{5}(x-2)$$

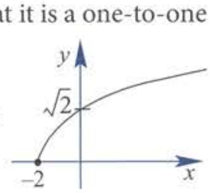
To complete the question we need the domain of f^{-1}. We already know that $d_{f^{-1}} = r_f$, therefore all we now need is the range of f, so $d_{f^{-1}} = \mathbb{R}$ (using the graph of $f(x)$).

That is, $f^{-1}: \mathbb{R} \mapsto \mathbb{R}$ where $f^{-1}(x) = \frac{1}{5}(x-2)$.

Example 2.1.14

Find the inverse function of $f(x) = \sqrt{x+2}, x \geq -2$ and sketch its graph.

A quick sketch of the graph of f verifies that it is a one-to-one function.



We can now determine the inverse function, $f^{-1}(x)$:

Now, $f(f^{-1}(x)) = \sqrt{f^{-1}(x)+2}$, so, using $f(f^{-1}(x)) = x$ we have:

$\sqrt{f^{-1}(x)+2} = x \Leftrightarrow f^{-1}(x) + 2 = x^2$ (squaring both sides)

$$\Leftrightarrow \quad f^{-1}(x) = x^2 - 2$$

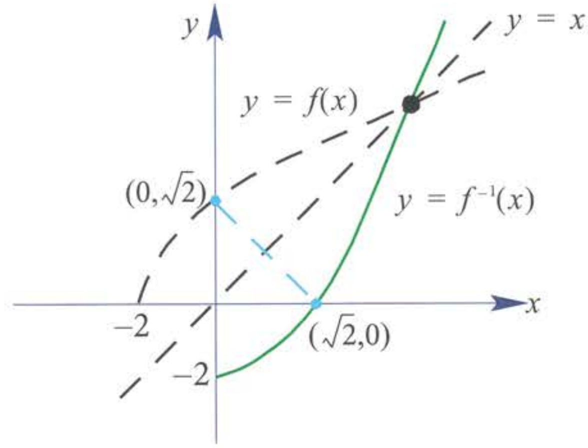
To fully define f^{-1} we need to determine the domain of f^{-1}.

We do this by using the result that $d_{f^{-1}} = r_f$.

Using the graph shown above, we have that $r_f = [0, \infty[$ $\therefore d_{f^{-1}} = [0, \infty[$. We are now in a position to fully define the inverse function.

We have, $f^{-1}: [0, \infty) \mapsto \mathbb{R}, f^{-1}(x) = x^2 - 2$.

We can now sketch the graph of the inverse function:

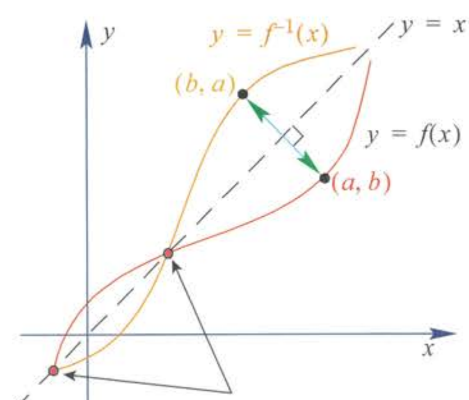


We make two important observations:

1. The graph of $y = f^{-1}(x)$ is the graph of $y = f(x)$ reflected about the line $y = x$.
2. Points of intersection of the graphs $y = f^{-1}(x)$ and $y = f(x)$ will always occur where both curves meet the line $y = x$.

The relationship between the graph of a function f and the graph of its inverse function f^{-1} can be explained rather neatly, because all that has actually happened is that we have interchanged the x- and y-values, i.e.$(a,b) \leftrightarrow (b,a)$. In doing so, we find that $d_{f^{-1}} = r_f$, and $d_f = r_{f^{-1}}$.

This then has the result that the graph of the inverse function $f^{-1}(x)$ is a reflection of the graph of the original function $f(x)$ about the line with equation $y = x$.



If the curve $y = f(x)$ intersects with the line $y = x$ then $y = f(x)$ and $y = f^{-1}(x)$ will meet on the line.

Graphing the inverse function

Because of the nature of the functions $f^{-1}(x)$ and $f(x)$, when finding the points of intersection of the two graphs, rather than solving $f^{-1}(x) = f(x)$, it might be easier to solve the equations $f^{-1}(x) = x$ or $f(x) = x$. The only caution when solving the latter two, is to always keep in mind the domain of the original function.

One interesting function is $f(x) = \frac{1}{x}, x \neq 0$.

It can be established that its inverse function is given by:

$$f^{-1}(x) = \frac{1}{x}, x \neq 0$$

i.e. $f(f^{-1}(x)) = x \Rightarrow \frac{1}{f^{-1}(x)} = x \Leftrightarrow f^{-1}(x) = \frac{1}{x}$.

Then, as the two functions are identical, it is its **self-inverse**.

Sketching the graph of $f(x) = \frac{1}{x}, x \neq 0$ and reflecting it about the line $y = x$ will show that the two graphs overlap each other. Note then, that self-inverse functions also intersect at points other than just those on the line $y = x$ – basically because they are the same functions!

There is a second method that we can use to find the inverse of a function.The steps required are:

Step 1: Let y denote the expression $f(x)$.

Step 2: Interchange the variables x and y.

Step 3: Solve for y.

Step 4: The expression in step 3 gives the inverse function, $f^{-1}(x)$.

We work through an example using this method.

Example 2.1.15

Find the inverse function, f^{-1} of $f:]1,\infty[\mapsto \mathbb{R}$ where $f(x) = \frac{2}{x-1} + 1$ and sketch its graph.

Let $y = f(x)$, giving $y = \frac{2}{x-1} + 1$.

Next, we interchange the variables x and y: $x = \frac{2}{y-1} + 1$.

We now solve for y: $x = \frac{2}{y-1} + 1 \Leftrightarrow x - 1 = \frac{2}{y-1}$

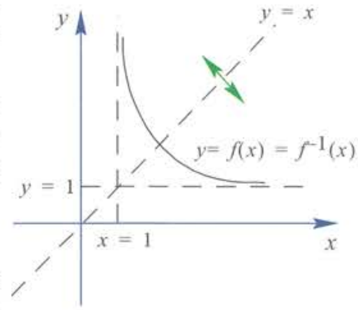
$\Leftrightarrow \frac{1}{x-1} = \frac{y-1}{2}$ (inverting both sides)

$\Leftrightarrow y - 1 = \frac{2}{x-1}$

Therefore, we have that $y = \frac{2}{x-1} + 1$

That is, $f^{-1}(x) = \frac{2}{x-1} + 1, x > 1$.

We notice that the original function $y = f(x)$ and its inverse function $y = f^{-1}(x)$ are the same. This becomes obvious when we sketch both graphs on the same set of axes.



When reflected about the line $y = x$, they are identical!

Again, we have an example of a self-inverse function.

Example 2.1.16

Find the inverse of the function $g: x \mapsto 4^x - 2, x \in \mathbb{R}$.

The function $g: x \mapsto 4^x - 2, x \in \mathbb{R}$ is a one-to-one function and so its inverse function exists.

Using the result that $g(g^{-1}(x)) = x$ we have:

$4^{g^{-1}(x)} - 2 = x$

$\Leftrightarrow 4^{g^{-1}(x)} = x + 2$

$\therefore g^{-1}(x) = \log_4(x+2)$ (using $N = b^x \Leftrightarrow x = \log_b N$)

Now, $d_{g^{-1}} = r_g =]-2, \infty[$ (we have obtained the range by using a sketch of $g(x)$)

Therefore the inverse, g^{-1}, is given by

$g^{-1}: x \mapsto \log_4(x+2), x > -2$

Example 2.1.17

Find the inverse of $g: \left(\frac{1}{2}, \infty\right) \mapsto \mathbb{R}, x \mapsto \log_{10}(2x-1) + 2$.

This time we make use of the second method of finding the inverse function.

Let $y = \log_{10}(2x-1)+2$, interchanging x and y, we have:

$x = \log_{10}(2y-1)+2$

$\Leftrightarrow \ x-2 = \log_{10}(2y-1)$

$\Leftrightarrow \ 2y-1 = 10^{x-2}$ (using $N = b^x \Leftrightarrow x = \log_b N$)

$\Leftrightarrow \ y = \frac{1}{2}(10^{x-2}+1)$.

Therefore, we have that $f^{-1}(x) = \frac{1}{2}(10^{x-2}+1)$.

Next, $d_{f^{-1}} = r_f = (-\infty, \infty)$, so the inverse function is: $f^{-1}: x \mapsto \frac{1}{2}(10^{x-2}+1), x \in (-\infty, \infty)$

Example 2.1.18

Find the inverse function of $f(x) = 4x^2+8x-1, x \geq -1$

Let $y = f(x)$

$$\begin{aligned} y &= 4x^2+8x-1 \\ &= 4(x^2+2x)-1 \\ &= 4(x^2+2x+1)-1-4 \\ &= 4(x+1)^2-5 \end{aligned}$$

Interchanging x and y.

$$x = 4(y+1)^2-5$$

$$(y+1)^2 = \frac{x+5}{4}$$

$$y+1 = \pm\frac{\sqrt{x+5}}{2}$$

$$y = \pm\frac{\sqrt{x+5}}{2}-1$$

Since the original domain is restricted to $x \geq -1$ and the range is $y \geq -5$, the inverse of the original quadratic equation must be restricted to the domain $x \geq -5$ with the range of $y \geq -1$.

Hence $f^{-1}(x) = \frac{\sqrt{x+5}}{2}-1, x \geq -5$

Exercise 2.1.5

1. Find the inverse function for each of the following.

   a $f(x) = 2x+1, x \in \mathbb{R}$

   b $f(x) = x^3, x \in \mathbb{R}$

   c $g(x) = \frac{1}{3}x-3, x \in \mathbb{R}$

   d $g(x) = \frac{2}{5}x+2, x \in \mathbb{R}$

   e $h(x) = \sqrt{x+1}, x > -1$

   f $f(x) = \sqrt{x}+1, x \geq 0$

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

   h $h(x) = \frac{1}{\sqrt{x}}-1, x > 0$

2. Using the graph of the original function, sketch the graph of the corresponding inverse function for each part in Question 1.

3. Find and sketch the inverse function of:

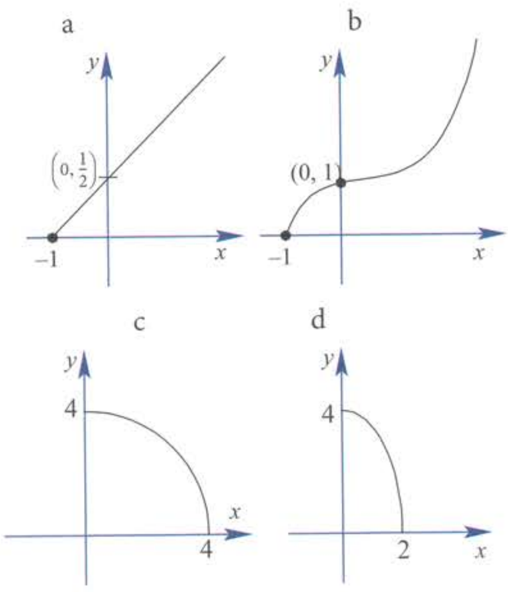
   a $f(x) = x^2-3, x \geq 0$.

   b $f(x) = x^2-3, x \leq 0$.

4. Show that $f(x) = \frac{x}{\sqrt{x^2+1}}, x \in \mathbb{R}$ is a one-to-one function.

   Hence find its inverse.

5. Sketch the inverse of the following functions.



6. Find the inverse function (if it exists) of the following.

a $f(x) = 3^x + 1, x \in \,]-\infty, \infty[$

b $f(x) = 2^x - 5, x \in \,]-\infty, \infty, [$

c $f(x) = 3^{2x+1}, x \in \,]-\infty, \infty[$

d $g(x) = 3 - 10^{x-1}, x \in \,]-\infty, \infty[$

e $h(x) = \dfrac{2}{3^x - 1}, x \neq 0$

f $g(x) = \dfrac{1}{2^x} - 1, x \in \,]-\infty, \infty[$

7. Using the graph of the original functions in Question 6, sketch the graph of their inverses.

8. Find the inverse function (if it exists) of the following functions:

a $f(x) = \log_2(x+1), x > -1$

b $f(x) = \log_{10}(2x), x > 0$

c $h(x) = 1 - \log_2 x, x > 0$

d $g(x) = \log_3(x-1) - 1, x > 1$

e $h(x) = 2\log_5(x-5), x > 5$

f $f(x) = 2 - \frac{1}{3}\log_{10}(1-x), x < 1$

9. Find the inverse function of $f(x) = x^2 + 2x, x \geq -1$, stating both its domain and range. Sketch the graph of f^{-1}.

10. Find the inverse function of:

a $f(x) = -x + a, x \in \,]-\infty, \infty[$, where a is real.

b $h(x) = \dfrac{2}{x-a} + a, x > a$, where a is real.

c $f(x) = \sqrt{a^2 - x^2}, 0 \leq x \leq a$, where a is real.

11. Find the inverse of $h(x) = -x^3 + 2, x \in \mathbb{R}$. Sketch both $h(x)$ and $h^{-1}(x)$ on the same set of axes.

12. Find the largest possible set of positive real numbers **S**, that will enable the inverse function h^{-1} to exist, given that $h(x) = (x-2)^2, x \in \mathbf{S}$.

13. Determine the largest possible positive valued domain, X, so that the inverse function, $f^{-1}(x)$, exists,

given that $f(x) = \dfrac{3x+2}{2x-3}, x \in \mathrm{X}$.

14. a Sketch the graph of $f(x) = x - \dfrac{1}{x}, x > 0$.

Does the inverse function, f^{-1} exist? Give a reason for your answer.

b Consider the function $g : \mathrm{S} \to \mathbb{R}$ where, $g(x) = x - \dfrac{1}{x}$

Find the two largest sets S so the inverse function, g^{-1}, exists. Find both inverses and on separate axes, sketch their graphs.

15. Find f^{-1} given that $f(x) = \sqrt{\dfrac{x}{a} - 1}$ where $a > 1$.

On the same set of axes sketch both the graphs of $y = f(x)$ and $y = f^{-1}(x)$. Find $\{x : f(x) = f^{-1}(x)\}$.

16. Find and sketch the inverse, f^{-1}, of the functions:

a $f(x) = \begin{cases} -\frac{1}{2}(x+1), & x > 1 \\ -x^3, & x \leq 1 \end{cases}$

b $f(x) = \begin{cases} e^{x+1}, & x \leq 0 \\ x + e, & x > 0 \end{cases}$

c $f(x) = \begin{cases} -\ln(x-1), & x > 2 \\ 2 - x, & x \leq 2 \end{cases}$

d $f(x) = \begin{cases} \sqrt{x} + 4, & x > 0 \\ x + 4, & -4 < x < 0 \end{cases}$

17. a On the same set of axes sketch the graph of $f(x) = \dfrac{1}{a}\ln(x-a), x > a > 0$ and its reciprocal.

b Find and sketch the graph of f^{-1}.

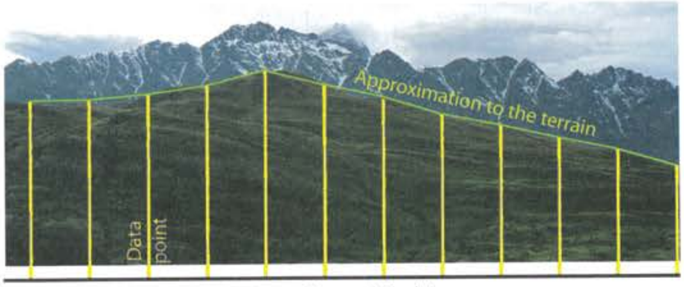
Extra questions

Answers

2.2 Graphs

Engineers designing a road through mountanous terrain will want to minimise the earth moving necessary to provide a reasonably level road. Bridges and tunnels are even more expensive. The choice of the 'right route' is a task requiring considerable skill.

The terrain will almost certainly not follow a simple mathematical formula. Instead, it will be a set of data points and a 'graph'.



Horizontal 'zero altitude'

The set of data points can now be treated as a graph. This can be used to estimate the amount of earth that will need to be moved if a particular route is chosen.

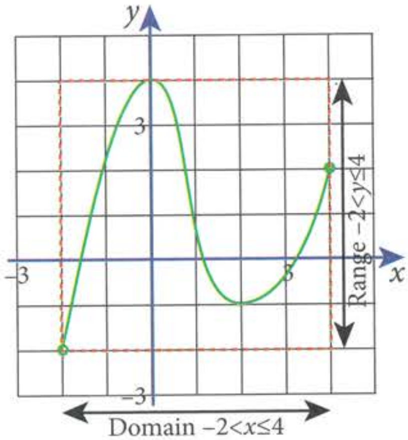
This section will deal with the visual aspects of graphs. We begin by reviewing some of the main features of graphs.

Domain and Range

The domain is the horizontal extent of a graph.

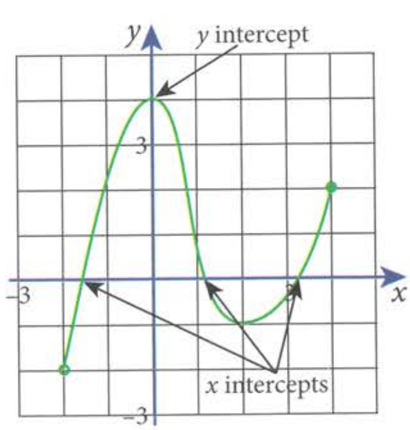
The range is its vertical extent.

By convention, points represented by solid dots are included in the graph and points with open dots are excluded.



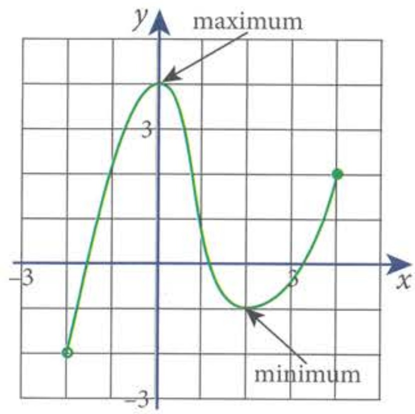
Intercepts

The points at which a graph cuts the axes are known as intercepts. Note that a function can have at most one intercept on the vertical axis.



Maxima and Minima

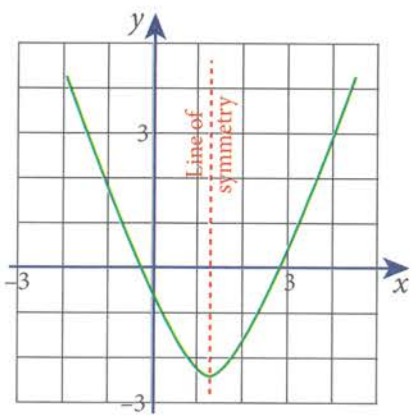
These are the peaks and valleys on a graph and are not necessarily the highest and lowest overall points.



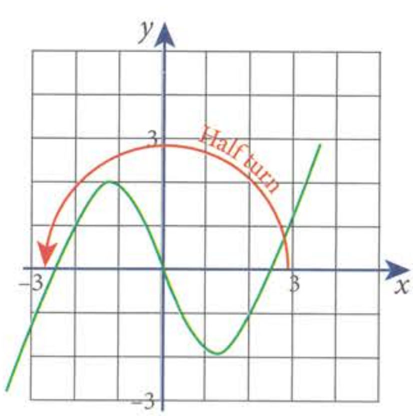
Symmetry

The two main types of symmetry encountered are:

1. Reflection (a line of symmetry):

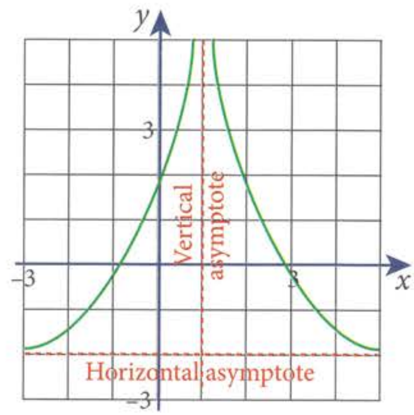


2. Rotational symmetry (often, though not always, about the origin).

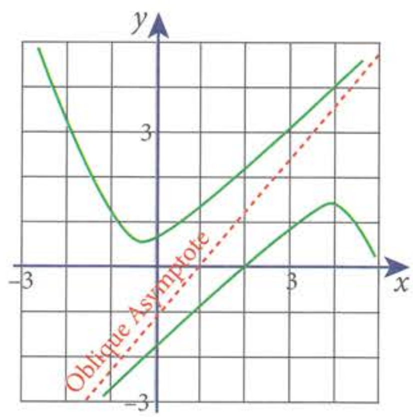


Asymptotes

Some graphs approach, without ever reaching, other lines on a graph. This is known as asymptotic behaviour.



OR:



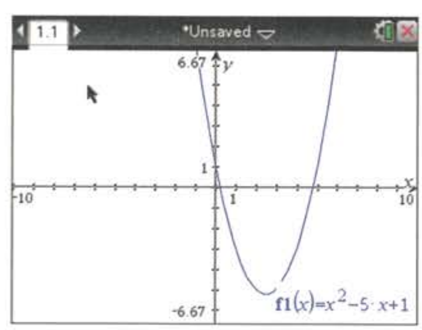
Note that this last example fails the vertical line test and is not a function.

Graphs modified by the absolute value function

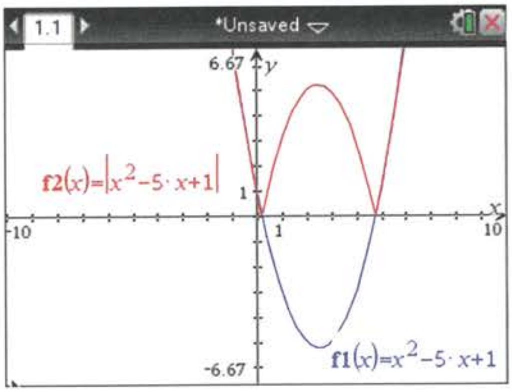
The absolute value function has no effect on a positive number, but makes negative numbers positive. A graphing calculator is a good way to explore this as the examples will show.

1. Comparing the graphs of $y = f(x)$ and $y = |f(x)|$.

The skill here is to take a graph and convert it to the absolute value graph. If using a calculator, begin with a simple function such as

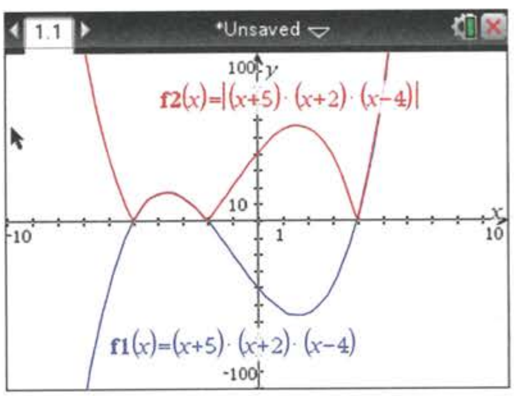


Next, add the absolute value of the function:



The parts of the graph above the x-axis are unchanged as the absolute value function leaves positive numbers unchanged. The parts of the graph below the x-axis are reflected in the x-axis as the absolute value function makes negative numbers positive.

Here is a second example:



Again, the negative parts of the blue graph have become positive. Note that the viewing window has been adjusted using ZOOM.

A useful way of seeing this is as a composite function with the absolute value function happening second - $\text{abs}(f(x))$.

2. Comparing the graphs of $y=f(x)$ and $y=f(|x|)$.

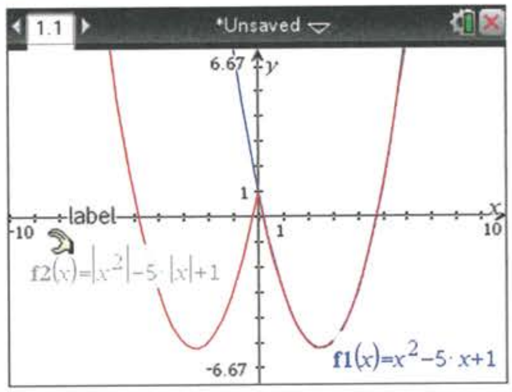
This is the composite in the reverse order - and we have already seen that we should not expect the same result. The first thing to happen is that the x values are made positive. They are then fed throgh f. The result is a left-right symmetric graph based in the right hand half.

Looking again at the first example.

The base function is: $y=x^2-5x+1$.

The composite is $y=|x|^2-5|x|+1$.

Now look at the two graphs:



The graph to the left of the y-axis is a mirror image of the graph to the right.

When answering questions on this topic, remember that you will not be given an algebraic rule and will not be able to use a calculator. You will need to 'see' the symmetries of the shapes.

Some functions you might like to try using your calculator are: $y=ax+b, y=ax(x-b), y=x^a, y=a\log(bx)$ where you can choose values for a & b.

Reciprocal Functions

How are the graphs of the two possible composites of a function defined by $y=f(x)$ and the reciprocal function related?

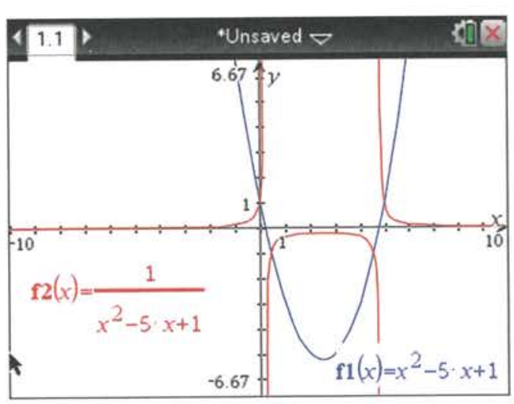
To begin with, recall that the reciprocal of a big number is small and the reciprocal of a small number is big. The sign is not changed.

Comparing $y=f(x)$ with $y=\dfrac{1}{f(x)}$.

Beginning with the parabola used in previous examples:

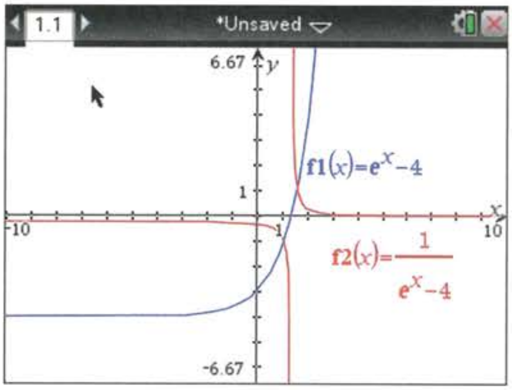
$$y=x^2-5x+1$$

When this is big and positive, we expect the reciprocal to be small and positive and vice-versa. This means that we get vertical asymptotes to correspond with x-intercepts.



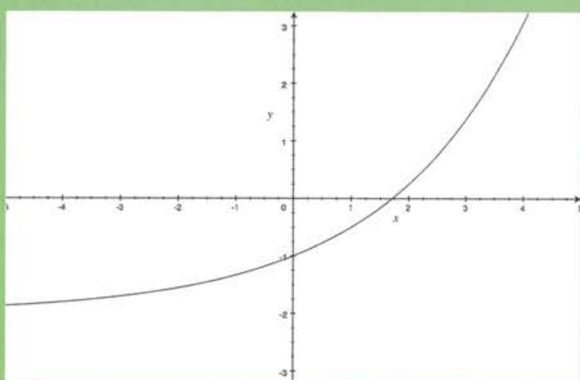
Negative parts of the blue graph correspond to negative parts of the red graph. Note also how the intercpts of the blue graph correspond to the vertical asymptotes of the red graph.

Here is a second example:



Example 2.2.1

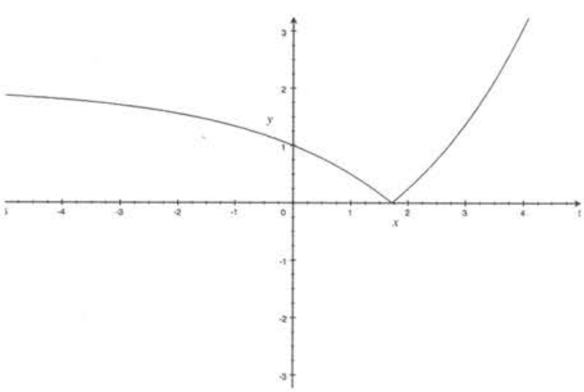
A function has the following graph:



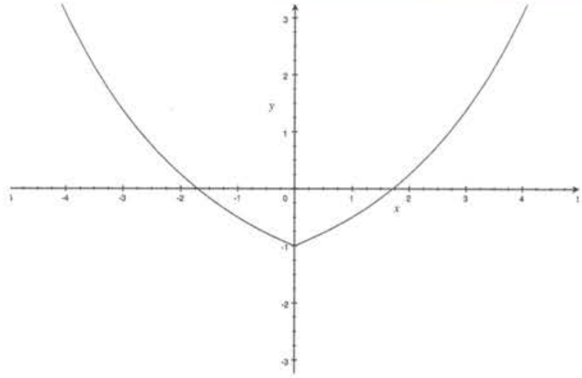
Sketch the graphs of: i $y=|f(x)|$

ii $y=f(|x|)$ iii $y=\dfrac{1}{f(x)}$

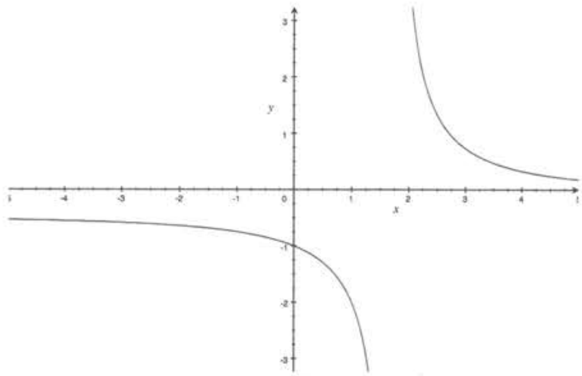
i



ii



iii

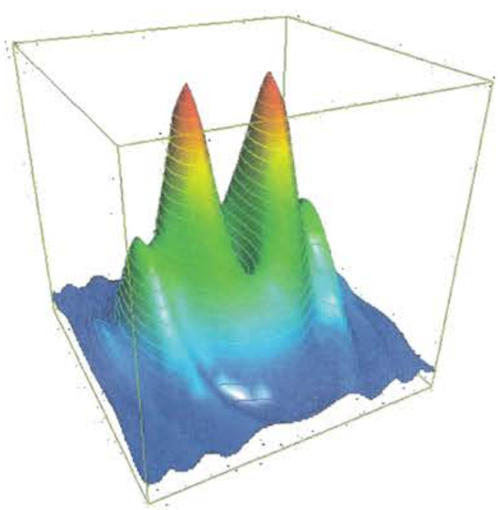


This section has been about the mathematics of 'shape'. There is a tendency to think of 'real mathematics' as being an activity that has to include a hefty dose of calculation.

This has only ever been partly true. The ancient Greek mathematicians were particularly interested in the mathematics of shape and their thoughts form the bedrock of the modern discipline.

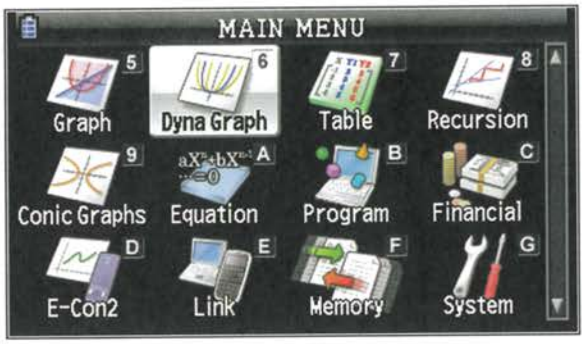
Dr Jacob Bronowski in his excellent *The Ascent of Man* (both a book and a TV series) explains an example from his own work on the evolution of the skull. The example comes in the chapter on mathematics - *The Music of the Spheres*. We also refer to this excellent resource in the Theory of Knowledge section.

Before working through the exercise, take a look at the 'graph'. To what extent could it be thought of as the same as a real mountain range?



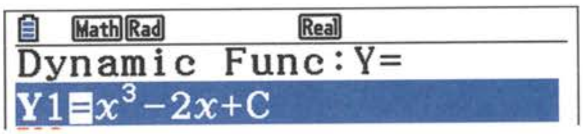
Dynamic Graphing

A particularly effective way of learning about transformations of graphs is to use the Dynamic Graphing feature of Casio graphic calculators.

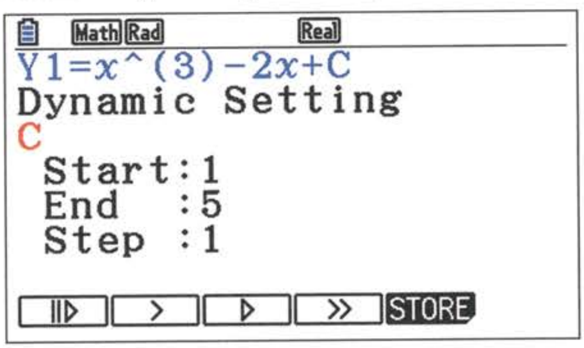


For example, if we look at the function $y = x^3 - 2x + C$ where C is a parameter that we can vary.

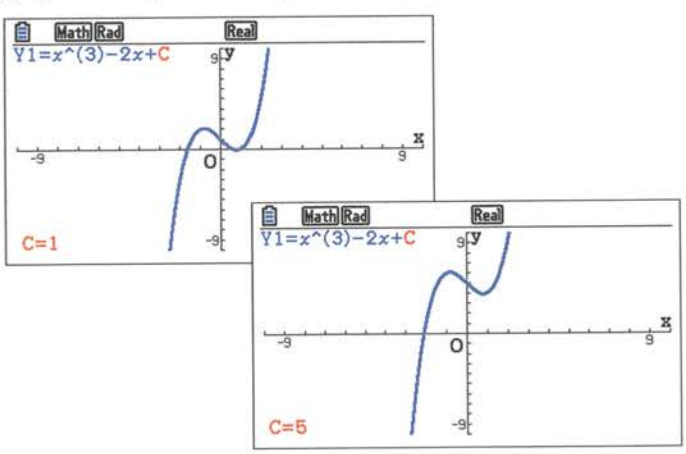
Open the Dyna Graph module (6) and enter the function.



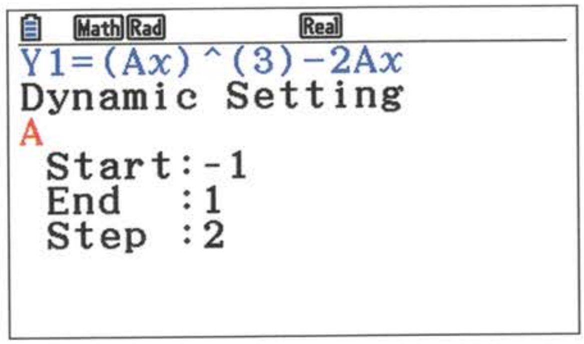
Then enter a range for the parameter (F4-VAR,F2-SET).



Press EXE, F6-DYNA and the calculator will animate the graphs as they run through the values of C.

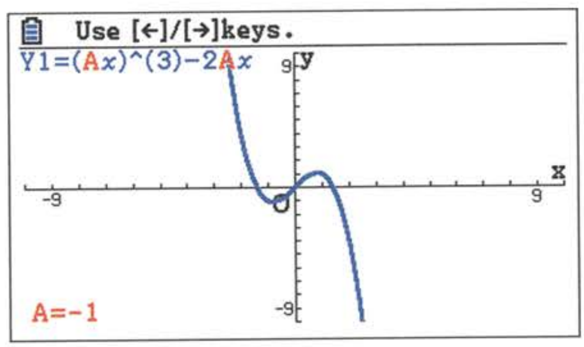


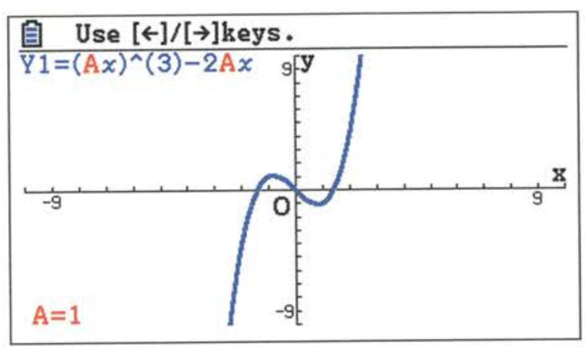
A second example illustrates reflection in the vertical axis.



Note that the two relevant values of the parameter A are –1 and 1 and the necessary step between the two values is 2.

The 'dyna' function will now allow you to animate the reflection between the two graphs.





Video

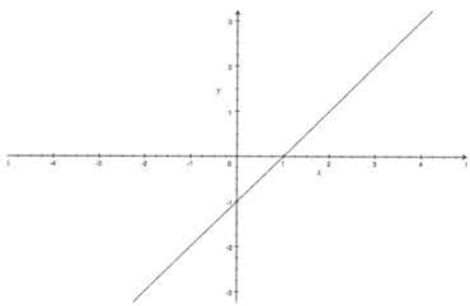
Exercise 2.2.1

1. For each of the following graphs, sketch:

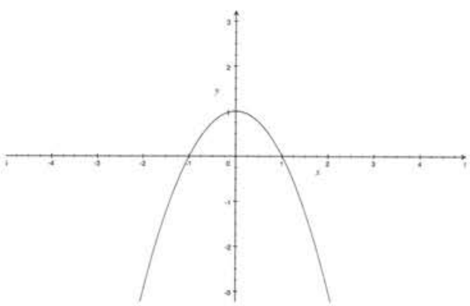
i $y=|f(x)|$ ii $y=f(|x|)$

iii $y=\dfrac{1}{f(x)}$

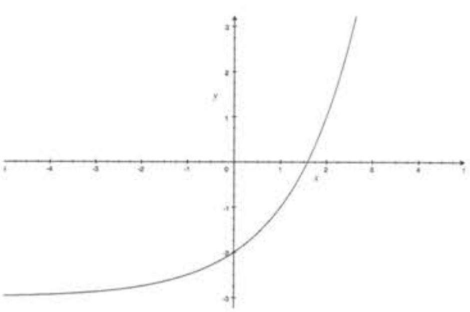
1.



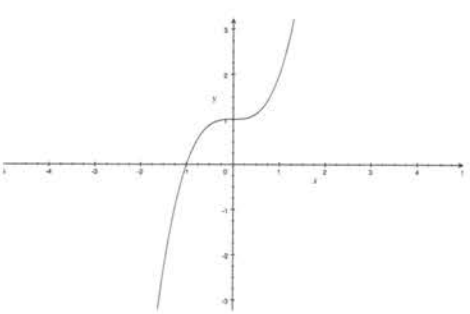
2.



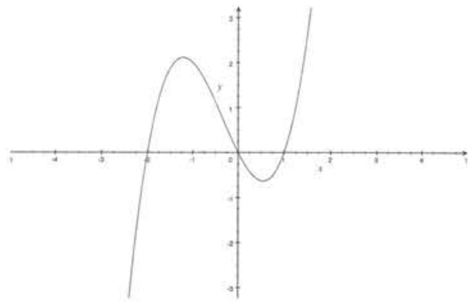
3.



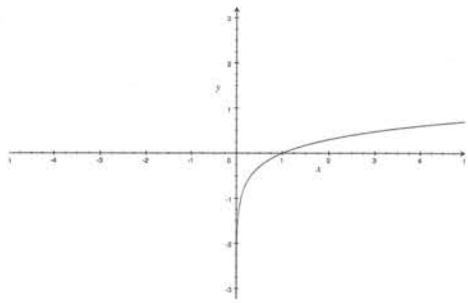
4.



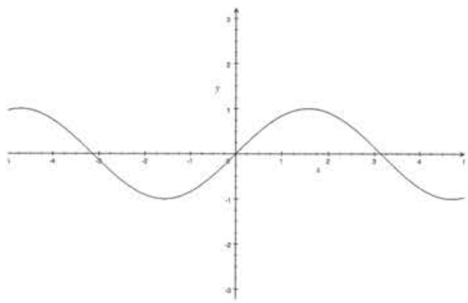
5.



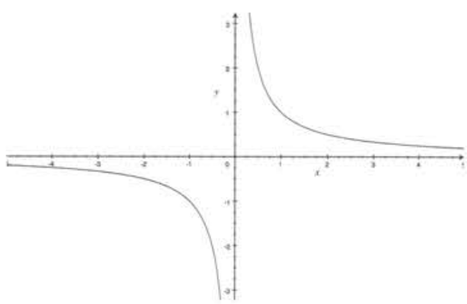
6.



7.



8.



2. For the following functions, graph $y=f(x)$, $y=|f(x)|$, $y=f(|x|)$, and $y=|f(|x|)|$ on the same set of axes.

i $f(x)=3x-5$

ii $f(x)=x^2+4x+8$

iii $f(x)=\sqrt{2x+5}-2$

Answers

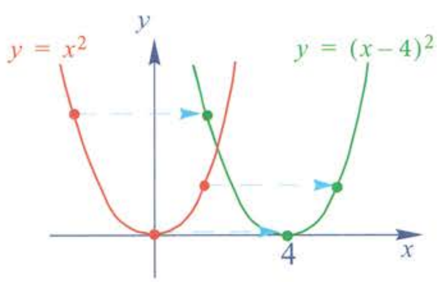
2.3 Transformations of Graphs

Horizontal translation

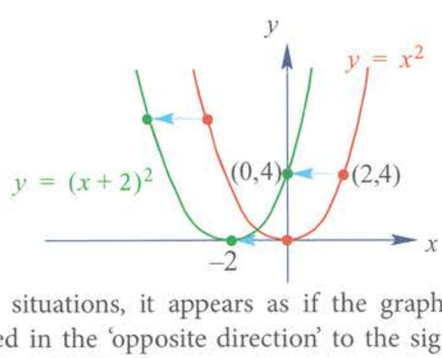
A horizontal translation takes the form: $f(x) \to f(x-a)$.

We start by looking at the transformation of the basic parabolic graph with equation $y = x^2$. The horizontal translation of $y = x^2$ is given by $y = (x-a)^2$. This transformation represents a **translation along the *x*-axis**.

For example, the graph of $y = (x-4)^2$ represents the parabola $y = x^2$ translated 4 units to the right ($a = 4$):



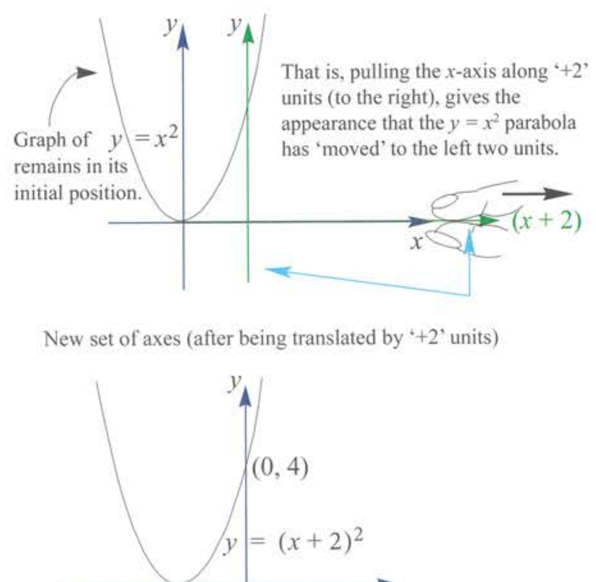
Similarly, the graph of $y = (x+2)^2$ represents the parabola $y = x^2$ translated 2 units to the left ($a = -2$):



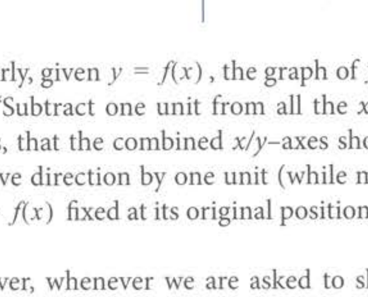
In both situations, it appears as if the graphs have been translated in the 'opposite direction' to the sign of '*a*'. That is, $y = (x+2)^2$ has been translated 2 units back (i.e. in the negative direction), while $y = (x-2)^2$ has been translated 2 units forward (i.e. in the positive direction).

The reason for this is **the transformation is applied to the *x*-values, not the graph.**

That is, given $y = f(x)$, the graph of $y = f(x+2)$ is telling us to 'Add two units to all the *x*-values'. In turn, this means, that the combined *x*/*y*–axes should be moved in the positive direction by two units (whilst the graph of $y = f(x)$ remains exactly where it is):

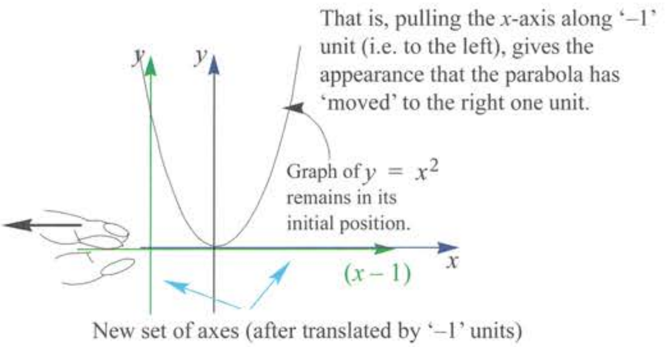


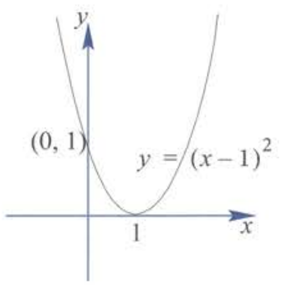
New set of axes (after being translated by '+2' units)



Similarly, given $y = f(x)$, the graph of $y = f(x-1)$ is telling us to 'Subtract one unit from all the *x*-values'. In turn, this means, that the combined *x*/*y*–axes should be moved in the negative direction by one unit (while maintaining the graph of $y = f(x)$ fixed at its original position):

However, whenever we are asked to sketch a graph, rather than drawing the axes after the graph has been sketched, the first thing we do is draw the set of axes and then sketch the graph. This is why there appears to be a sense in which we seem to do the 'opposite' when sketching graphs that involve transformations.



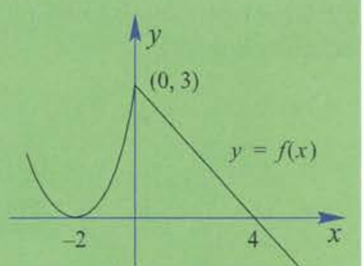


Example 2.3.1

Given the graph of $y = f(x)$ sketch the graph of:

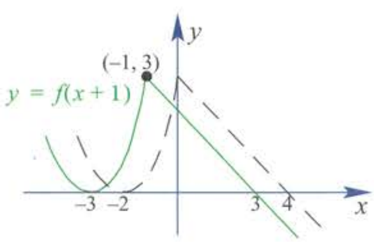
a $y = f(x+1)$

b $y = f\left(x - \frac{1}{2}\right)$



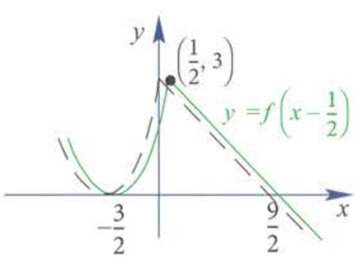
a

The graph of $y = f(x+1)$ represents a translation along the x-axis of the graph of $y = f(x)$ by 1 unit to the left.



b

The graph of $y = f\left(x - \frac{1}{2}\right)$ represents a translation along the x-axis of the graph of $y = f(x)$ by $\frac{1}{2}$ unit to the right.

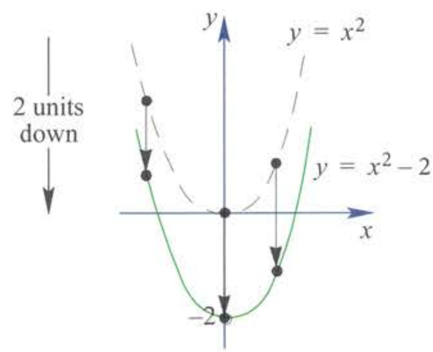


Vertical translation

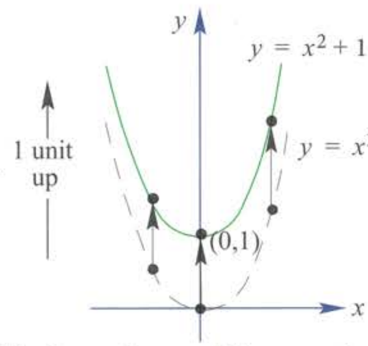
A vertical translation takes on the form $f(x) \rightarrow f(x)+b$.

Again we consider the transformation of the basic parabolic graph with equation $y = x^2$. The vertical translation of $y = x^2$ is given by $y = x^2 + b$. This transformation represents a translation along the y-axis.

For example, the graph of $y = x^2 - 2$ represents the parabola $y = x^2$ translated 2 units down:



Similarly, the graph of $y = x^2 + 1$ represents the parabola $y = x^2$ translated 1 unit up:

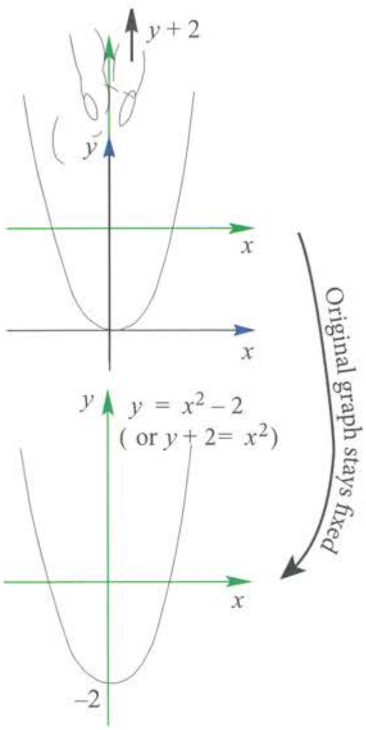


Note that this time, when applying a vertical translation, we are consistent with the sign of b!

The reason is that, although we are sketching the graph of $y = f(x) + b$, we are in fact sketching the graph of $y - b = f(x)$. So that this time, the transformation is applied to the y-values, and not the graph!

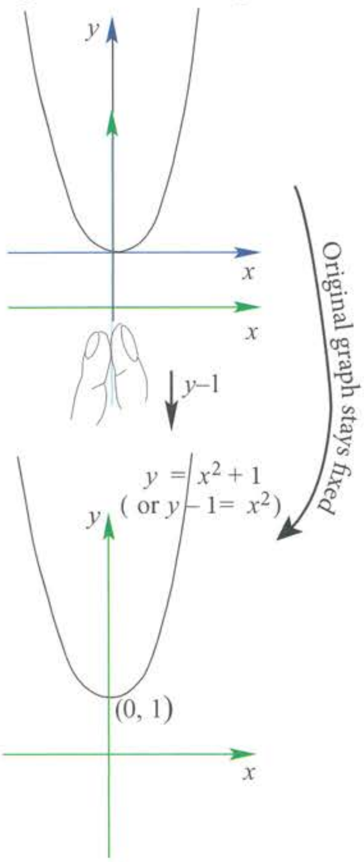
So, if we consider the two previous examples, we have $y = x^2 - 2 \Leftrightarrow y + 2 = x^2$, and so, in this case we would be pulling the y-axis **UP** 2 units, which gives the appearance that the parabola has been moved down 2 units.

That is, pulling the y-axis along '+2' units (i.e. upwards), gives the appearance that the parabola has 'moved' down 2 units.



Similarly, $y = x^2 + 1 \Leftrightarrow y - 1 = x^2$, so that in this case we would be pulling the y-axis **DOWN** 1 unit, which gives the appearance that the parabola has been moved up 1 unit.

That is, pulling the y-axis along '–1' units (i.e. downwards), gives the appearance that the parabola has 'moved' up 1 unit.

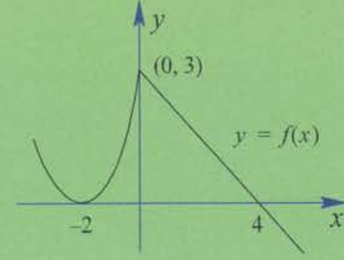


Example 2.3.2

Given the graph of $y = f(x)$ sketch the graph of:

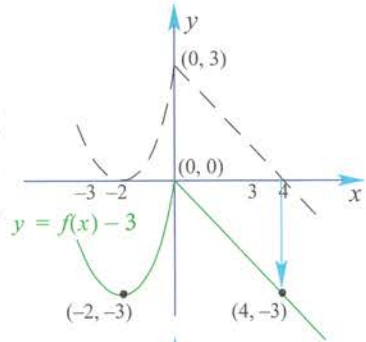
a $\quad y = f(x) - 3$

b $\quad y = f(x) + 2$



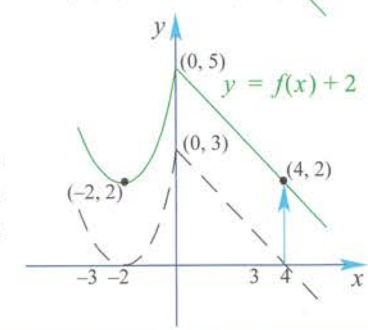
a

The graph of $y = f(x) - 3$ represents a translation along the y-axis of the graph of $y = f(x)$ by 3 units in the downward direction.



b

The graph of $y = f(x) + 2$ represents a translation along the y-axis of the graph of $y = f(x)$ by 2 units in the upward direction.



Summary

$y=f(x-a)$, $a > 0$: translation of $f(x)$ along the x-axis of a units to the right.

$y=f(x+a)$ $a > 0$: translation of $f(x)$ along the x-axis of a units to the left.

$y=f(x)+b$, $b > 0$: translation of $f(x)$ along the y-axis of b units up.

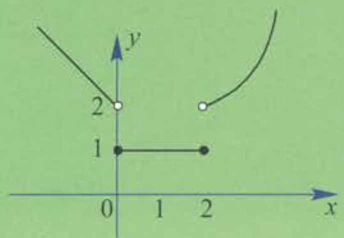
$y=f(x)-b$, $b > 0$: translation of $f(x)$ along the y-axis of b units down.

Of course it is also possible to apply both a vertical and horizontal translation to the one graph at the same time. That is, the graph of $y = (x-1)^2 + 2$ would represent the graph of $y = x^2$ after it had been translated one unit to the right and two units up.

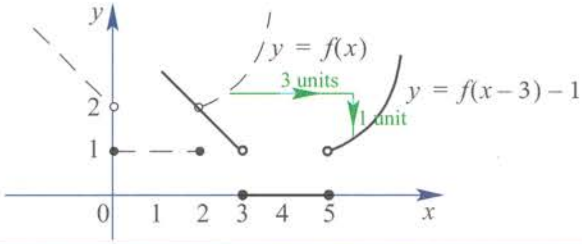
Such a combination takes the form $f(x) \rightarrow f(x-a)+b$, representing a horizontal translation of 'a' along the x-axis and a vertical translation of 'b' along the y-axis.

Example 2.3.3

Sketch the graph of $y = f(x-3)-1$ for the graph shown.

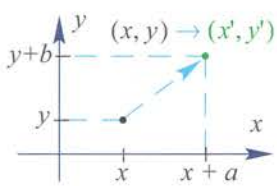


The graph of $y = f(x-3)-1$ represents the graph of $y = f(x)$ after a translation along the x-axis of 3 units to the right followed by a translation along the y-axis of 1 unit down.



Vector Notation

The mapping from the original coordinates (x, y) to the new coordinates (x', y') can also be presented in vector form. That is, if the point (x, y) is translated 'a' units along the x-axis and 'b' units along the y-axis, the new coordinates would be given by $(x+a, y+b)$.



That is, $x' = x+a$ and $y' = y+b$.

The vector notation for such a translation is given by:

$$\boxed{\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}}$$

From $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$ we then have:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \Leftrightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x'-a \\ y'-b \end{pmatrix} \therefore x = x'-a \;\&\; y = y'-b$$

Substituting these results into the equation $y = f(x)$ we obtain the transformed equation:

$$y'-b = f(x'-a) \Rightarrow y' = f(x'-a)+b$$

As x' and y' are only dummy variables, we can rewrite this last equation as $y = f(x-a)+b$.

Under the vector translation $\begin{pmatrix} a \\ b \end{pmatrix}$, we have the mapping $f(x) \rightarrow f(x-a)+b$

Example 2.3.4

Find the equation of $y = x^3 - 1$ under the translation $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$.

Let the new set of axes be u and v so that:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 3 \\ 5 \end{pmatrix} \Rightarrow \left.\begin{array}{l} u = x+3 \\ v = y+5 \end{array}\right\} \text{ or } \left.\begin{array}{l} x = u-3 \\ y = v-5 \end{array}\right\}.$$

Substituting these into $y = x^3 - 1$, we have:

$$v-5 = (u-3)^3 - 1 \Rightarrow v = (u-3)^3 + 4.$$

However, u and v are dummy variables, and so we can rewrite the last equation in terms of x and y, i.e. $y = (x-3)^3 + 4$. This represents the graph of $y = x^3$ translated 3 units to the right (along the x-axis) and 4 units up (along the y-axis).

Notice that the original relation is in fact $y = x^3 - 1$, so that relative to this graph, the graph of $y = (x-3)^3 + 4$ has been translated 3 units to the right (along the x-axis) and 5 units up (along the y-axis).

Exercise 2.3.1

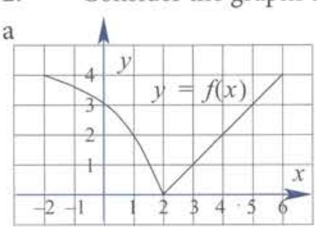
1. Find the equation of the given relation under the translation indicated.

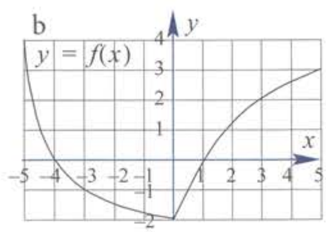
   a $y = x^2$; $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$ b $y = x^2$; $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$

   c $y = x^2$; $\begin{pmatrix} 0 \\ 5 \end{pmatrix}$ d $x^2 + y = 2$; $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$

   e $x^2 + y = 2$; $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$

2. Consider the graphs shown below.

a


b


In each case, sketch the graph of:

i $y = f(x+2)$ ii $y = f(x-1)$

iii $y = f(x)-3$ iv $y = f(x)+1$

3. Using translations on the graph of $f(x) = \sqrt{x}$, sketch the graphs of the following.

   a $y = f(x-4)$ b $y = f(x) - 2$

   c $y = f(x-2) + 3$

4. Using translations on the graph of $f(x) = \frac{1}{x}$, sketch the graphs of the following.

   a $y = f(x+1)$ b $y = f(x) - 4$

   c $y = f(x+2) - 3$

5. Using translations on the graph of $f(x) = \frac{1}{x^2}$, sketch the graphs of the following.

   a $y = f(x) - 1$ b $y = f(x-1)$

   c $y = 2 + f\left(x - \frac{3}{2}\right)$

6. On the same set of axes sketch the graphs of:

   a $y = x^2 - 4$ and $y = (x-4)^2$

   b $y = x^2 + 5$ and $y = (x+5)^2$

   c $y = x^2 + 2$ and $y = x^2 - 2$

   d $y = \left(x + \frac{3}{2}\right)^2$ and $y = \left(x - \frac{3}{2}\right)^2$

7. Sketch the graphs of the following functions, making sure to include all axial intercepts and labelling the equations of asymptotes (where they exist).

   a $y = (x-2)^2 + 3$

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

   c $y = (x-1)^3 - 1$

   d $y = \frac{1}{x-2} + \frac{1}{2}$

   e $y = (x+2)^3 - 8$

   f $y = (x+3)^2 - 9$

   g $y = \sqrt{x-2} + 2$

   h $y = \sqrt{4+x} + 2$

Extra questions

Dilations

Dilation from the *x*-axis

Before we start our discussion it should be pointed out that other commonly used expressions for *dilations from the x-axis* are: *dilation along the y-axis* and *dilation parallel to the y-axis* – any one of these three expressions can be used when describing this dilation.

The equation $y = p.f(x)$ can be written as $\frac{y}{p} = f(x)$.

We have rearranged the expression so that we can more clearly see the effects that p has on the y-values. That is, the term $^y/_p$ represents a transformation on the y-axis as opposed to a transformation on the graph of $f(x)$. The effect of 'p' in the term $^y/_p$ is that of a **dilation from the x-axis.**

If $|p| > 1$, we *shrink* the y-axis (seeing as we are dividing the y-values by a number larger than one). Whereas if $0 < |p| < 1$ we *stretch* the y-axis.

However, we still need to describe the effect this transformation has on the final appearance of the graph of $f(x)$.

We summarise these results, stating the effects on the graph of $f(x)$:

For the curve with equation $y = p.f(x)$,

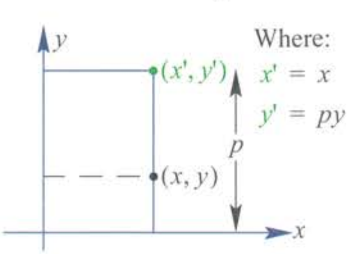
$|p| > 1$, represents '**stretching**' $f(x)$ by a factor p from the x-axis.

$0 < |p| < 1$, represents '**shrinking**' $f(x)$ by a factor $\frac{1}{p}$ from the x-axis.

A 'stretch' of factor $^1/_3$ is the same as a 'shrink' of factor 3. However, it is more common that when referring to a dilation from the x-axis, we refer to it as a stretch. So a dilation from the x-axis of factor 3 would imply a stretching effect whereas a dilation from the x-axis of factor $^1/_3$ would imply a shrinking effect.

The relationship between the original coordinates (x, y) and the new coordinates (x', y') can be seen in the diagram below.

Unlike the translation vector discussed earlier, we have no dilation vector to describe this transformation [although there does exist a *dilation matrix*].



Example 2.3.5

Describe the effects on the graph of $y=f(x)$ when the following graphs are sketched.

a $y=2.f(x)$ b $y=\frac{1}{4}f(x)$

c $2y=f(x)$

a $y = 2f(x)$ represents a dilation of factor 2 from the x-axis. This means that the graph of $y = f(x)$ would be stretched by a factor of two along the y-axis.

b $y=\frac{1}{4}f(x)$ shows a dilation of factor $^1/_4$ from the x-axis.

This means that the graph of $y = f(x)$ would shrink by a factor of four along the y-axis.

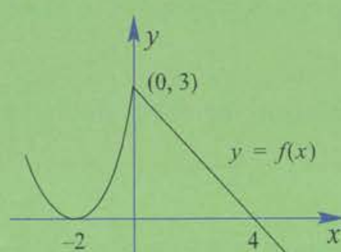
c $2y = f(x)$ needs to first be written as $y=\frac{1}{2}f(x)$.

This represents a dilation of factor $^1/_2$ from the x-axis. This means that the graph of $y = f(x)$ would shrink by a factor of two along the y-axis.

Example 2.3.6

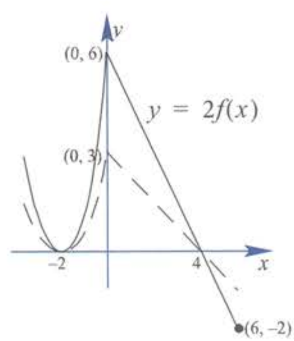
Given the graph of $y=f(x)$, sketch the graph of:

a $y = 2f(x)$ b $3y = f(x)$



a The graph of $y = 2f(x)$ represents a dilation of factor 2, i.e. the graph of $y = f(x)$ will be stretched by a factor of 2.

Notice how the x-intercepts are invariant, i.e. they have not altered after the transformation. This is because the y-value at these points is zero, and multiplying zero by any number will still be zero.



b $3y = f(x) \Leftrightarrow y = \frac{1}{3}f(x)$

This represents a dilation of factor $^1/_3$, i.e. the graph of $y = f(x)$ will shrink by a factor of 3.

Dilation from the y-axis

The equation: $y=f\left(\frac{x}{q}\right)$

represents a transformation applied to the x-values. We now need to consider how this factor 'q' affects the graph of $f(x)$. The term $^x/_q$ represents a transformation on the x-values as opposed to a transformation on the graph of $f(x)$. The effect of 'q' in the term $^x/_q$ is that of a **dilation from the y-axis**.

If $|q| > 1$, we *shrink* the x-axis (seeing as we are dividing the x-values by a number larger than one). Whereas if $0 < |q| < 1$ we *stretch* the x-axis (because we are dividing by a number less than one but greater than zero). However, we still need to describe the effect this transformation has on the final appearance of the graph of $f(x)$.

We summarise these results, stating the effects on the graph of $f(x)$:

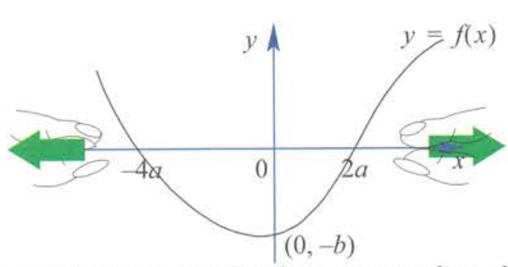
For the curve with equation:

$$y=f\left(\frac{x}{q}\right)$$

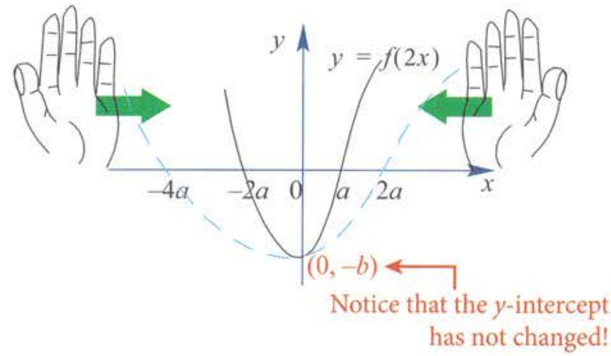
$|q| > 1$, represents 'stretching' $f(x)$ by a factor q from the y-axis.

$0 < |q| < 1$, represents 'shrinking' $f(x)$ by a factor $\frac{1}{q}$ from the y-axis.

So a dilation from the y-axis of factor 2 (e.g. $y = f(x/2)$) would imply a stretching effect whereas a dilation from the y-axis of factor $^1/_2$ (e.g. $y = f(2x)$) would imply a shrinking effect. We show this in the following diagram.

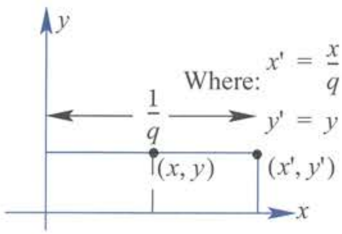


Multiplying the x-values by 2, i.e. stretching the x-axis by a factor of 2 has the same effect as squashing the graph of $y = f(x)$ by a factor of 2.



Again, it must be remembered that the graph of $y = f(x)$ has not changed, it is simply the illusion that the graph has been squashed. That is, relative to the new (stretched) x-axis, the graph of $y = f(x)$ appears to have been squashed.

The relationship between the original coordinates (x, y) and the new coordinates (x', y') can be seen in the diagram below.

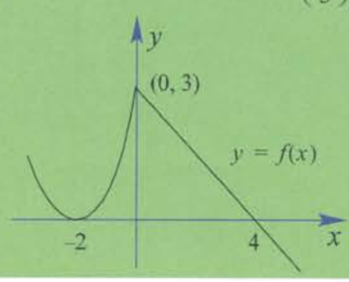


Again, we have no dilation vector to describe this transformation (although there does exist a *dilation matrix*).

Example 2.3.7

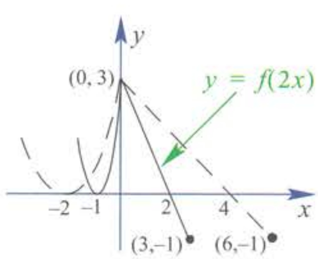
Given the graph of $y = f(x)$, sketch the graph of:

a $\quad y = f(2x) \quad$ b $\quad y = f\left(\frac{x}{3}\right)$



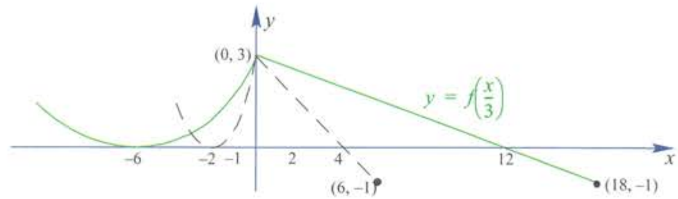
a First express the dilation in the form $y = f(x/q)$:

$y = f(2x) \Rightarrow y = f\left(x/\left(\frac{1}{2}\right)\right)$.



This means that we have a dilation from the y-axis of factor $^1/_2$. That is, the graph of $y = f(x)$ will 'shrink' (or rather be squashed) by a factor of 2. Because the x-values are doubled (from the $2x$ term in the expression $y = f(2x)$) it seems reasonable to deduce that on the new set of axes the graph will be squashed by a factor of 2.

The term $\frac{x}{3}$ in the expression $y = f\left(\frac{x}{3}\right)$ implies that the new x-values will be a third of the original x-values. This means that the new x-axis will be compressed by a factor of 3. This in turn will have a stretching effect on $y = f(x)$ of factor 3 (along the x-axis).



Exercise 2.3.2

1. On the same set of axes, sketch the graphs of:

   a $\quad f(x) = x^2,\ y = f(2x)$

   b $\quad f(x) = \sqrt{x},\ y = f(4x)$

   c $\quad f(x) = \frac{1}{x},\ y = f\left(\frac{x}{3}\right)$

   d $\quad f(x) = x^3,\ y = f\left(\frac{x}{2}\right)$

2. On the same set of axes, sketch the graphs of:

   a $\quad f(x) = x^2,\ y = 2f(x)$

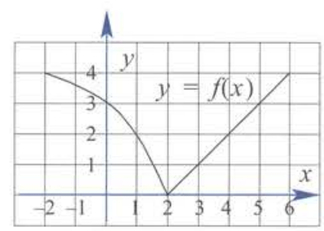
   b $\quad f(x) = \sqrt{x},\ y = 4f(x)$

   c $\quad f(x) = \frac{1}{x},\ y = \frac{1}{3}f(x)$

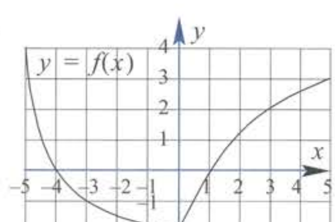
   d $\quad f(x) = x^3,\ y = \frac{1}{2}f(x)$

3. Consider the graphs shown below.

   a



   b



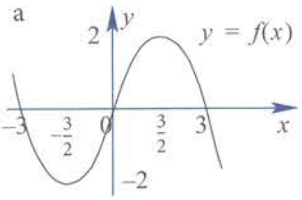
In each case, sketch the graphs of:

i $y = f(0.5x)$ ii $y = f(2x)$

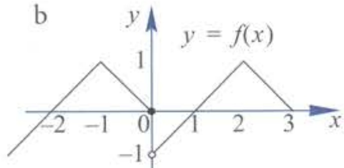
iii $y = 0.5f(x)$ iv $y = 2f(x)$

4. Consider the relations shown below.

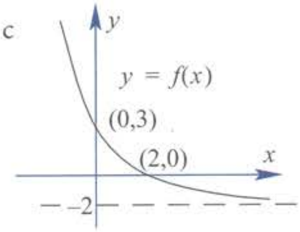
a



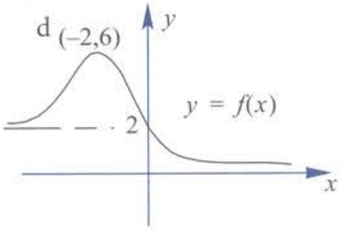
b



c



d



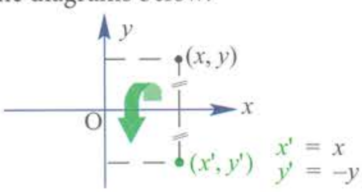
Sketch the following.

i $y = f\left(\frac{2}{3}x\right)$ ii $y = 4f(x)$

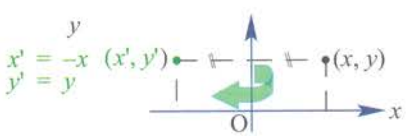
Extra questions

Reflections

We first consider reflections about the x-axis and about the y-axis. The effects of reflecting a curve about these axes can be seen in the diagrams below:



When reflecting about the x-axis we observe that the coordinates (x, y) are mapped to the coordinates $(x, -y)$ meaning that the x-values remain the same but the y-values change sign.



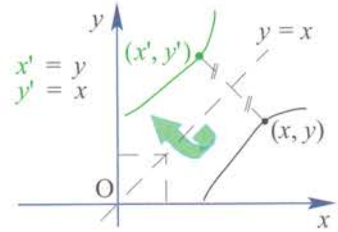
When reflecting about the y-axis we observe that the coordinates (x, y) are mapped to the coordinates $(-x, y)$, meaning that the y-values remain the same but the x-values change sign.

We summarise the effects of these two transformations of the graph of $y = f(x)$

Another type of reflection is the reflection about the line $y = x$, when sketching the inverse of a function. Inverse functions are dealt with in detail in Chapter 2,1, so here we give a summary of that result.

Reflection about the line $y = x$

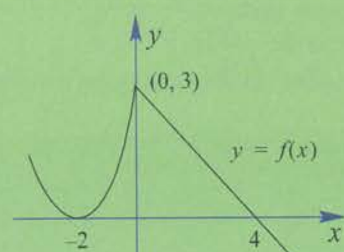
When reflecting about the line $y = x$ we observe that the coordinates (x,y) are mapped to the coordinates (y,x), meaning that the x-values and the y-values are interchanged. If a one-one function undergoes such a reflection, we call its transformed graph the inverse function and denote it by $y = f^{-1}(x)$.



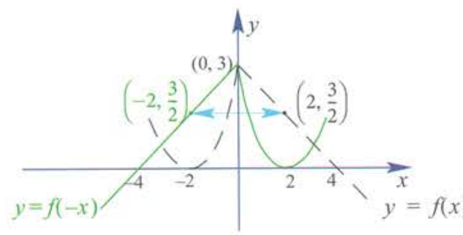
Example 2.3.8

Given the graph of $y = f(x)$, sketch the graph of:

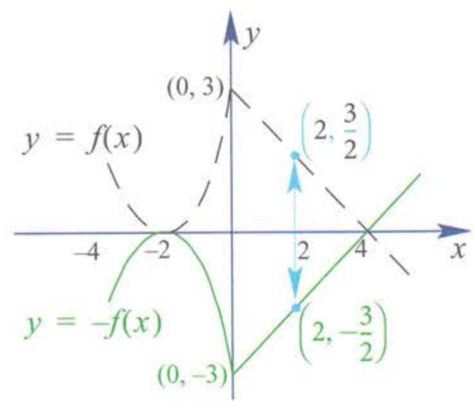
a $y = f(-x)$ b $y = -f(x)$



a The graph of $y = f(-x)$ represents a reflection of the graph of $y = f(x)$ about the y-axis:



b The graph of $y = -f(x)$ represents a reflection of the graph of $y = f(x)$ about the x-axis:



We now consider a combination of the transformations we have looked at so far.

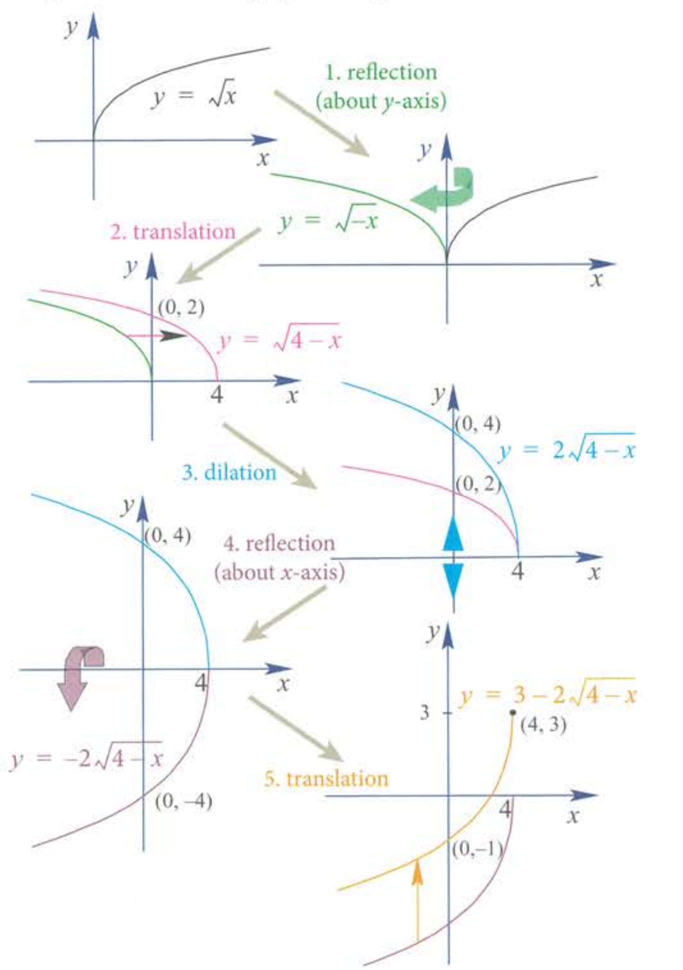
Example 2.3.9

Sketch the graph of $y = 3 - 2\sqrt{4-x}$.

We start by considering the function $f(x) = \sqrt{x}$, then, the expression $y = 3 - 2\sqrt{4-x}$ can be written in terms of $f(x)$ as follows: $y = 3 - 2f(4-x)$ or $y = -2f(4-x) + 3$. This represents:

1. reflection about the y-axis (due to the '$-x$' term)
2. translation of 4 units to the right (due to the '$4 - x$' term). Note: $4 - x = -(x-4)$
3. dilation of factor 2 along the y-axis (due to the $2f(x)$ term)
4. reflection about the x-axis (due to the '$-$' in front of the $2f(x)$ term)
5. translation of 3 units up (due to the '+3' term)

We produce the final graph in stages:

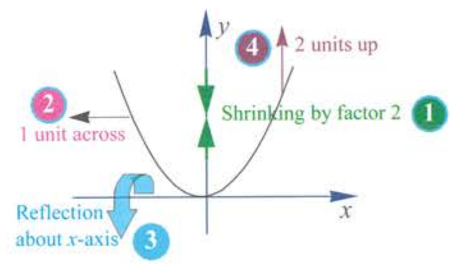


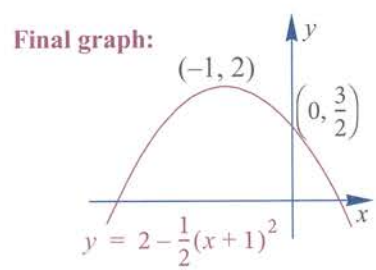
The previous example gave a step-by-step account of how to produce the final graph, however, there is no need to draw that many graphs to produce the final outcome. We can reduce the amount of work involved by including all the transformations on one set of axes and then produce the final graph on a new set of axes.

Example 2.3.10

Sketch the graph of $y = 2 - \frac{1}{2}(x+1)^2$.

If we consider the function $f(x) = x^2$, the graph of $y = 2 - \frac{1}{2}(x+1)^2$ can be written as $y = 2 - \frac{1}{2}f(x+1)$. This represents a 'shrinking' effect of factor 2 from the x-axis followed by a reflection about the x-axis then a translation of 1 to the left and finally a translation of 2 units up.





At this stage we have not looked at the x- or y-intercepts, although these should always be determined.

Important note!

Note the order in which we have carried out the transformations - although there is some freedom in this - there are some transformations that must be carried out before others.

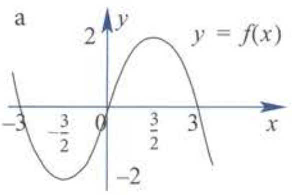
You should try to alter the order in which the transformations in Example 2.3.10 have been carried out. For example, does it matter if we apply Step 2 before Step 1? Can Step 2 be carried out after Step 4?

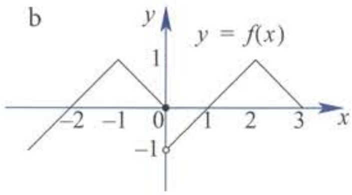
Exercise 2.3.3

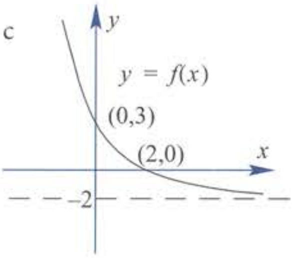
1. Sketch the graphs of:

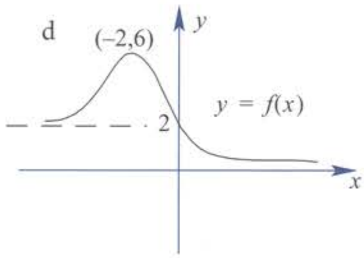
   i $y = f(-x)$ ii $y = -f(x)$

   for each of the following.

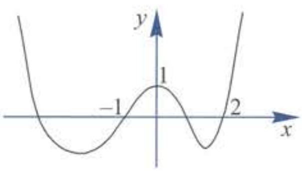
a


b


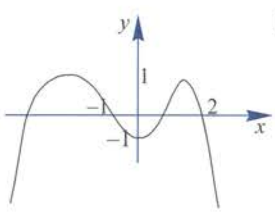
c


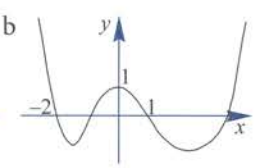
d


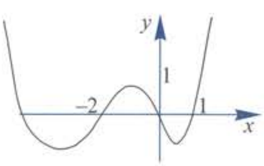
2. The diagram below shows the graph of the function $y = f(x)$.

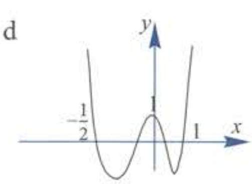


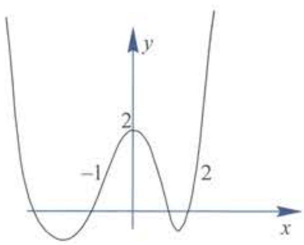
Find the equation in terms of $f(x)$ for each of the following graphs.

a


b


c


d


e


3. Sketch the graphs of the following.

   a $f(x) = -3x^2 + 9$

   b $f(x) = 4 - \frac{1}{2}x^2$

   c $f(x) = 1 - \frac{1}{8}x^3$

   d $f(x) = -(x+2)^2 + 3$

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

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

   g $f(x) = -(x-2)^3 - 2$

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

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

   j $f(x) = 3 - \frac{2}{1-x}$

   k $f(x) = -(1-x)^2$

   l $f(x) = 4\sqrt{9-x}$

   m $f(x) = -\frac{2}{(2x-1)^2}$

4. The graph of $y = f(x)$ is shown opposite. Use it to sketch the graphs of:

   a $y = f(x-1)$

   b $y = f(x) - 1$

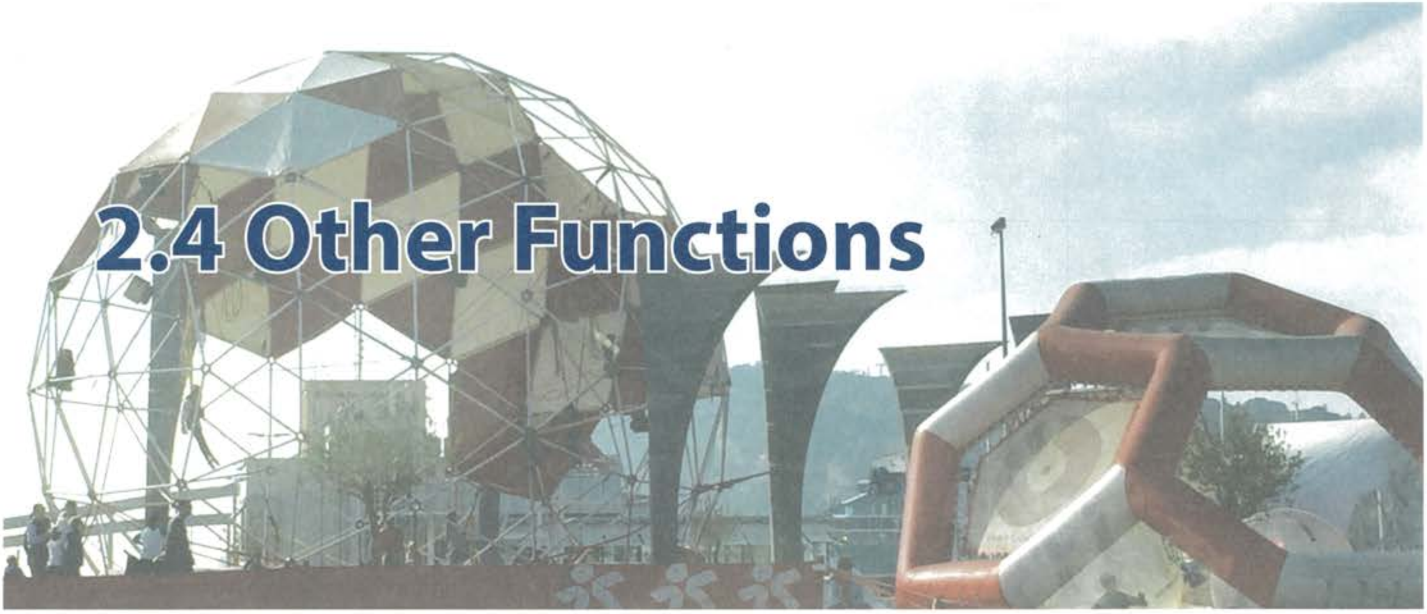
   c $y = f(x+1)$ d $y = 1 - f(x)$

   e $y = 1 + f(-x)$ f $y = 2 - f(-x)$

   g $y = -\frac{1}{2}f(x)$ h $y = f(-2x)$

Extra questions

Answers



2.4 Other Functions

The Rational Function

Rational fuctions take the form: $f(x)=\dfrac{ax+b}{cx+d}$

Graphs of this nature possess two types of asymptotes, one vertical and the other horizontal.

1. The vertical asymptote

A vertical asymptote occurs when the denominator is zero, that is, where $cx+d=0$. Where this occurs, we place a vertical line (usually dashed), indicating that the curve cannot cross this line under any circumstances. This must be the case, because the function is undefined for that value of x.

For example, the function

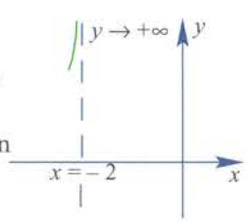
$$x \mapsto \frac{3x+1}{2x+4}$$

is undefined for that value of x where $2x+4=0$. That is, the function is undefined for $x=-2$. This means that we would need to draw a vertical asymptote at $x=-2$. In this case, we say that the asymptote is defined by the equation $x=-2$.

Using limiting arguments provides a more formal approach to 'deriving' the equation of the vertical asymptote. The argument is based along the following lines:

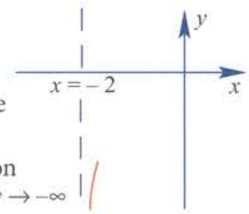
as $x \to -2^-, \dfrac{3x+1}{2x+4} \to +\infty$

That is, as x tends to -2 from the left or 'below', (hence the minus sign next to the two) the function tends to positive infinity.



as $x \to -2^-, \dfrac{3x+1}{2x+4} \to +\infty$

That is, as x tends to -2 from the right or 'above', (hence the plus sign next to the two) the function tends to negative infinity.



Therefore we write:

$$\left.\begin{array}{l} \text{As } x \to -2^+ \;\; f(x) \to -\infty \\ \text{As } x \to -2^- \;\; f(x) \to +\infty \end{array}\right\} \begin{array}{l} \therefore x=-2 \\ \text{is a vertical} \\ \text{asymptote} \end{array}$$

$$\text{of } f(x)=\frac{3x+1}{2x+4}, x \neq -2.$$

2. The horizontal asymptote

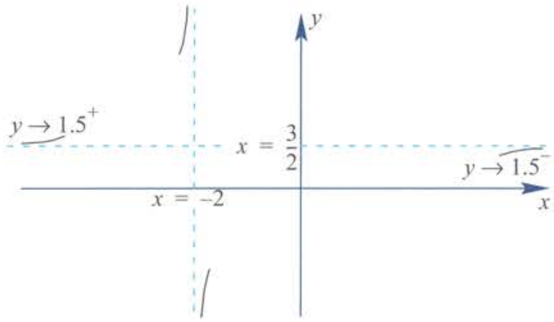
To determine the equation of the horizontal asymptote, we use a limiting argument, however, this time we observe the behaviour of the function as $x \to \pm\infty$.

It will be easier to determine the behaviour of the function (as $x \to \pm\infty$) if we first 'simplify' the rational function (using long division):

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

Next we determine the behaviour for extreme values of x.

$$\left.\begin{array}{l} \text{As } x \to +\infty \;\; f(x) \to \left(\frac{3}{2}\right)^- \\ \text{As } x \to -\infty \;\; f(x) \to \left(\frac{3}{2}\right)^+ \end{array}\right\} \begin{array}{l} \text{Therefore, } y=\frac{3}{2} \\ \text{is the horizontal} \\ \text{asymptote} \end{array}$$



We can now add a few more features of the function:

3. Axial intercepts

x–intercept

To determine the x-intercept(s) we need to solve for $f(x) = 0$.

In this case we have:

$$f(x) = \frac{3x+1}{2x+4} = 0 \Leftrightarrow 3x+1 = 0 \Leftrightarrow x = -\frac{1}{3}.$$

That is, the curve passes through the point $\left(-\frac{1}{3}, 0\right)$.

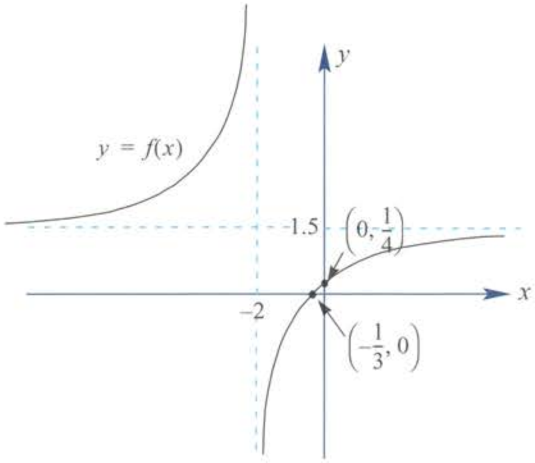
y-intercept

To determine the y-intercept we find the value of $f(0)$ (if it exists, for it could be that the line $x = 0$ is a vertical asymptote).

In this case we have $f(0) = \dfrac{3 \times 0 + 1}{2 \times 0 + 4} = \dfrac{1}{4}$.

Therefore the curve passes through the point $(0, {}^1/_4)$.

Having determined the behaviour of the curve near its asymptotes (i.e. if the curve approaches the asymptotes from above or below) and the axial–intercept, all that remains is to find the stationary points (if any).



Example 2.4.1

Sketch the graph of: $y = \dfrac{3x+1}{1-x}$

The vertical asymptote is at $x = 1$. Just to the left of the asymptote (e.g. at $x = -1.1$) the numerator is negative and the denominator small and negative. The y value is large and positive. Just to the right of the asymptote (e.g. at $x = -0.9$) the numerator is negative and the denominator small and positive. The y value is large and negative.

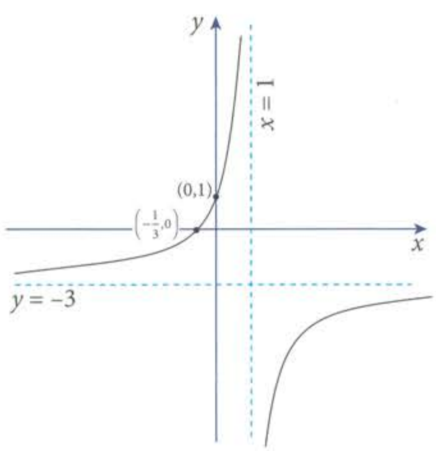
The horizontal asymptote occurs when x is large. When this happens, y tends to −3. For a big negative x (e.g. $x = -100$), $y = {}^{-299}/_{101}$ i.e. a bit smaller than −3 (above the asymptote). For a big positive x (e.g. $x = 100$), $y = {}^{301}/_{-99}$ i.e. a bit bigger than −3 (below the asymptote).

Intercepts:

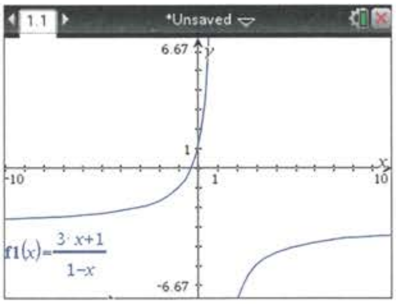
x intercept at $3x+1=0 \Rightarrow x = -\dfrac{1}{3}$

y intercept at $y = \dfrac{3\times 0+1}{1-0} = 1$

The sketch should show all the important features:



Modern graphic calculators will produce quite good plots of rational functions. They do not, however, show the key features which a good sketch must include.



Exercise 2.4.1

1. Use a limiting argument to determine the equations of the vertical and horizontal asymptotes for the following.

   a $f(x)=\dfrac{2x+1}{x+1}$ b $f(x)=\dfrac{3x+2}{3x+1}$

   c $f(x)=\dfrac{2x-1}{4x+1}$ d $f(x)=\dfrac{4-x}{x+3}$

   e $f(x)=3-\dfrac{1}{x}$ f $f(x)=5-\dfrac{1}{2-x}$

2, Make use of a graphics calculator to verify your results from Question 1 by sketching the graph of the given functions.

3. Sketch the following curves, clearly labelling all intercepts, stating the equations of all asymptotes, and, in each case, showing that there are no stationary points.

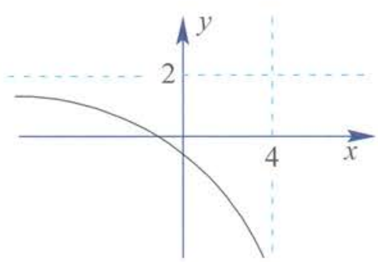
   a $x\mapsto\dfrac{3}{2x+1}$ b $x\mapsto\dfrac{x+1}{x+2}$

   c $x\mapsto\dfrac{5-x}{2x-1}$ d $x\mapsto 3+\dfrac{1}{x}$

   e $x\mapsto\dfrac{1}{x-3}-2$ f $x\mapsto 1-\dfrac{2}{2x-3}$

4. The figure at below shows part of the graph of the function whose equation is:

   $x\mapsto\dfrac{ax+2}{x-c}$.

   Find the values of a and c.

5. Given that $f{:}x\mapsto x+2$ and that $g{:}x\mapsto\dfrac{1}{x-1}$, sketch the graphs of:

   a $f\circ g$ b $g\circ f$.

Extra questions

The Exponential Function

The **exponential function** takes the form:

$$f(x)=a^x, x\in\mathbb{R}, a>0, a\neq 1$$

where the independent variable is the exponent.

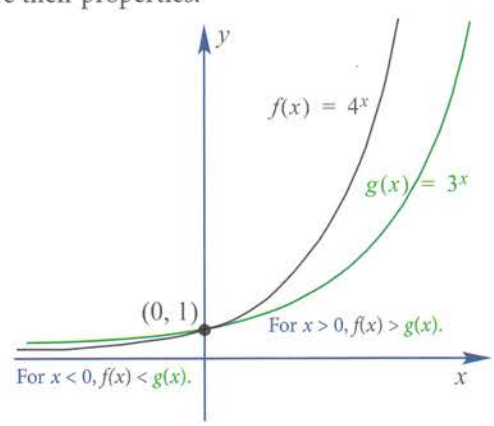
Graphs with a > 1

An example of an exponential function is $f(x)=2^x, x\in\mathbb{R}$ So, how does the graph of $f(x)=2^x$ compare with that of $f(x)=x^2$?

We know that the graph of $f(x)=x^2$ represents a parabola with its vertex at the origin, and is symmetrical about the y-axis. To determine the properties of the exponential function we set up a table of values and use these values to sketch a graph of $f(x)=2^x$.

x	...	−2	−1	0	1	2	...
$f(x)=x^2$	...	4	1	0	1	4	...
$f(x)=2^x$	...	$^1/_4$	$^1/_2$	1	2	4	...

We can now plot both graphs on the same set of axes and compare their properties:



Notice how different the graphs of the two functions are, even though their rules appear similar. The difference being that for the quadratic function, the variable x is the base, whereas for the exponential, the variable x is the power.

Properties of $f(x) = 2^x$

The function increases for all values of x (i.e. as x increases so too do the values of y).

The function is always positive (i.e. it lies above the x-axis).

As $x\to\infty$ then $y\to\infty$

$x \to -\infty$ then $y \to 0$. i.e. the x-axis is an asymptote.

When $x > 0$ then $y > 1$,

$x = 0$ then $y = 1$

$x < 0$ then $0 < y < 1$.

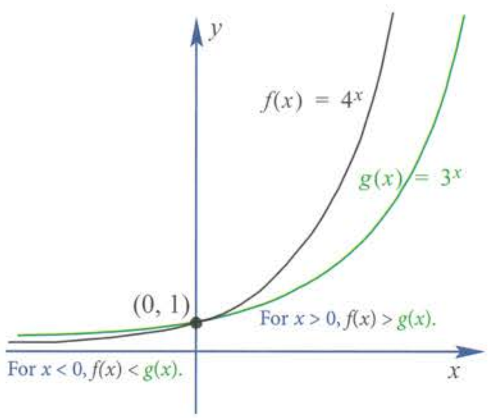
We can now investigate the exponential function for different bases. Consider the exponential functions $f(x) = 4^x$ and $g(x) = 3^x$: From the graphs we can see that $f(x) = 4^x$ increases much faster than $g(x) = 3^x$ for $x > 0$.

For example, at $x = 1$, $f(1) = 4$, $g(1) = 3$ and then, at $x = 2$, $f(2) = 16$, $g(2) = 9$. However, for $x < 0$ we have, $f(x) = 4^x$ decreases faster than $g(x) = 3^x$.

Notice then that at $x = 0$, both graphs pass through the point (0, 1).

From the graphs we can see that for values of x less than zero, the graph of $f(x) = 4^x$ lies below that of $g(x) = 3^x$. Whereas for values of x greater than zero, then the graph of $f(x) = 4^x$ lies above that of $g(x) = 3^x$.

Exponential functions that display these properties are referred to as exponential growth functions.



What happens when $0 < a < 1$?

We make use of a calculator to investigate such cases. Consider the case where $a = ½$.

Rather than using a table of values we provide a sketch of the curve. The graph shows that the function is decreasing – such exponential functions are referred to as exponential decay.

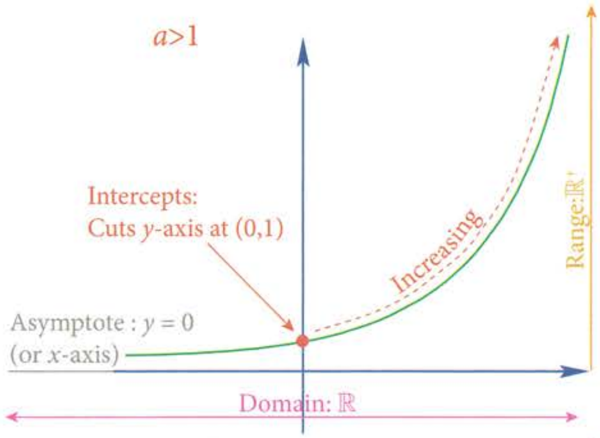
In fact, from the second screen we can see that the graph of $y = (½)^x$ is a reflection of $y = 2^x$ about the y-axis.

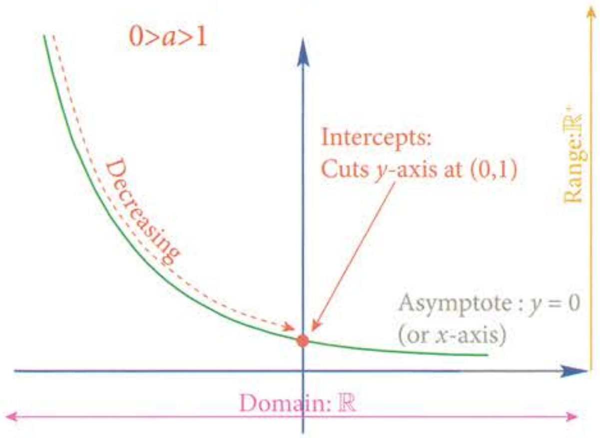
We note that the function $y = (½)^x$ can also be written as $y = (2^{-1})^x = 2^{-x}$.

There are two ways to represent an exponential decay function, either as $f(x) = a^x$, $0<a<1$ or $f(x) = a^{-x}$, $a>1$.

For example, the functions $f(x) = \left(\frac{1}{4}\right)^x$ and $g(x) = 4^{-x}$ are identical.

We can summarize the exponential function as follows:





There also exists an important exponential function known as the natural exponential function.

This function is such that the base has the special number 'e'. The number 'e', which we also consider in Chapter 1.2 has a value that is given by the expression:

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \approx 2.71828...$$

However, at this stage it suffices to realise that the number 'e' is greater than one. This means that a function of the form $f(x) = e^x$ will have the same properties as that of $f(x) = a^x$ for $a > 1$.

That is, it will depict an exponential growth. Whereas the function $f(x) = e^{-x}$ will depict an exponential decay.

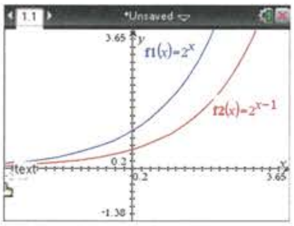
Example 2.4.2

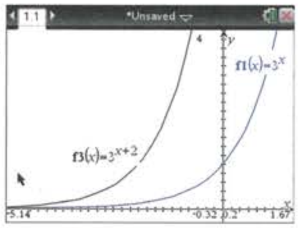
On the same set of axes sketch the following.

a $\quad f(x) = 2^x, g(x) = 2^{x-1}$

b $\quad f(x) = 3^x, g(x) = 3^{x+2}$

Using a calculator:

a 

b 

Notice that the principles of transformations discussed in the previous section (2.3) apply to exponential graphs. In Example 2.4.2 a we have a 'right one unit' translation and in b, a 'left two units' translation.

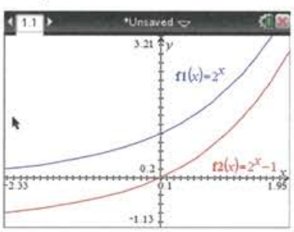
Example 2.4.3

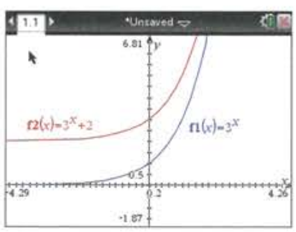
On the same set of axes sketch the following.

a $\quad f(x) = 2^x, g(x) = 2^x - 1$

b $\quad f(x) = 3^x, g(x) = 3^x + 2$

Using a calculator:

a 

b 

'down 1' 'up 2'

Example 2.4.4

On the same set of axes sketch the following.

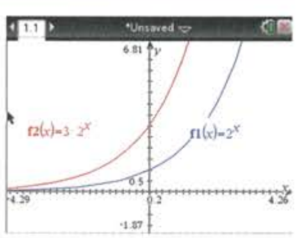
a $\quad f(x) = 2^x, g(x) = 3 \times 2^x$

b $\quad f(x) = 3^x, g(x) = -\left(\frac{1}{2}\right) \times 3^x$

Using a calculator:

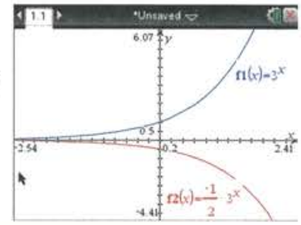
a

Vertical stretch factor 3.



b

Reflection in x-axis and dilation of ½ in the y direction.

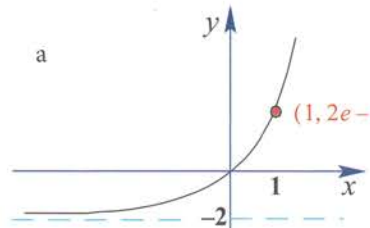


Example 2.4.5

Sketch the following.

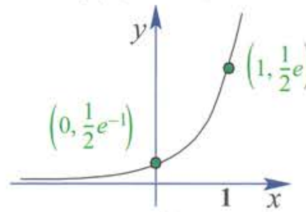
a $\quad f(x) = 2 \times e^x - 2$ $\qquad$ b $\quad f(x) = \frac{1}{2}e^{2x-1}$

a



When $x = 0$, $f(0) = 2 \times e^0 - 2 = 0$.
When $x = 1$, $f(1) = 2 \times e^1 - 2$.

b



When $x = 0$, $f(0) = \frac{1}{2}e^{0-1} = \frac{1}{2}e^{-1}$.

When $x = 1$, $f(1) = \frac{1}{2}e^{2-1} = \frac{1}{2}e^{1}$.

Notice that in both cases the general shape of the exponential growth remains unaltered. Only the main features of the graph are of interest when sketching is involved.

Exercise 2.4.2

1. On separate sets of axes, sketch the graphs of the following functions and determine the range of each function.

a	$f(x) = 4^x$	b	$f(x) = 3^x$
c	$f(x) = 5^x$	d	$f(x) = (2.5)^x$
e	$f(x) = (3.2)^x$	f	$f(x) = (1.8)^x$
g	$f(x) = \left(\frac{1}{2}\right)^x$	h	$f(x) = \left(\frac{1}{3}\right)^x$
i	$f(x) = \left(\frac{1}{5}\right)^x$	j	$f(x) = \left(\frac{3}{4}\right)^x$
k	$f(x) = \left(\frac{5}{8}\right)^x$	l	$f(x) = (0.7)^x$

2. Sketch the following on the same set of axes, clearly labelling the y-intercept.

$f(x) = 3^x + c$ where i c = 1 ii c = –2

$f(x) = 2^{-x} + c$ where i c = 0.5 ii c = –0.5

3. Sketch the following on the same set of axes, clearly labelling the y-intercept.

$f(x) = b \times 3^x$ where i b = 2 ii b = –2

$f(x) = b \times \left(\frac{1}{2}\right)^x$ where i b = 3 ii b = – 2

4. On the same set of axes, sketch the following graphs.

a $f(x) = 3^x$ and $f(x) = 3^{-x}$

b $f(x) = 5^x$ and $f(x) = 5^{-x}$

c $f(x) = 10^x$ and $f(x) = 10^{-x}$

d $f(x) = \left(\frac{1}{3}\right)^x$ and $f(x) = \left(\frac{1}{3}\right)^{-x}$

5. Find the range of the following functions.

a $f:[0,4] \mapsto \mathbb{R}, y = 2^x$

b $f:[1,3] \mapsto \mathbb{R}, y = 3^x$

c $f:[-1,2] \mapsto \mathbb{R}, y = 4^x$

d $f:[-1,2] \mapsto \mathbb{R}, y = 2^x$

e $f:[2,3] \mapsto \mathbb{R}, y = 2^{-x}$

f $f:[-1,1] \mapsto \mathbb{R}, y = 10^{-x}$

6. Sketch the graphs of the following functions, stating their range.

a $f: \mathbb{R} \mapsto \mathbb{R}$ where $f(x) = 2e^x + 1$

b $f: \mathbb{R} \mapsto \mathbb{R}$ where $f(x) = 3 - e^{x-1}$

c $f: \mathbb{R} \mapsto \mathbb{R}$ where $f(x) = e - e^{-x}$

d $f: \mathbb{R} \mapsto \mathbb{R}$ where $f(x) = 2 + \frac{1}{2}e^{-x}$

Extra questions

The Logarithmic Function

The logarithmic function, with base 'a' is represented by the expression $g(x) = \log_a x$, $x>0$.

The logarithmic function is the inverse of the exponential function:

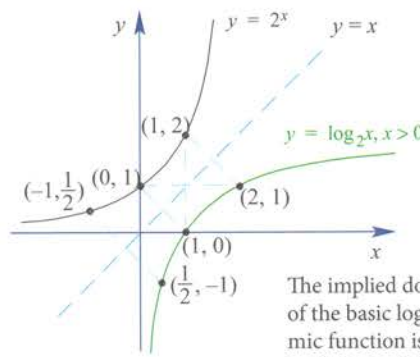
$$\text{If } f(x) = 2^x \text{ then } f^{-1}(x) = \log_2 x$$

To determine the shape of its graph we start by constructing a table of values for the function $f^{-1}(x) = \log_2 x$ and comparing it with the table of values for $f(x) = 2^x$:

x	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	0	1	2
$y = \log_2 x$	–3	–2	–1	–	0	1

x	–3	–2	–1	0	1	2
$y = 2^x$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4

From the table of values we observe that the x- and y-values have interchanged! Plotting these results on the same set of axes, we observe that the graph of the logarithmic function is a reflection of the exponential function about the line $y = x$. So, whereas for the exponential function, the asymptote is the x-axis (i.e. $y = 0$), for the logarithmic function, the asymptote is the y-axis (i.e. $x = 0$).

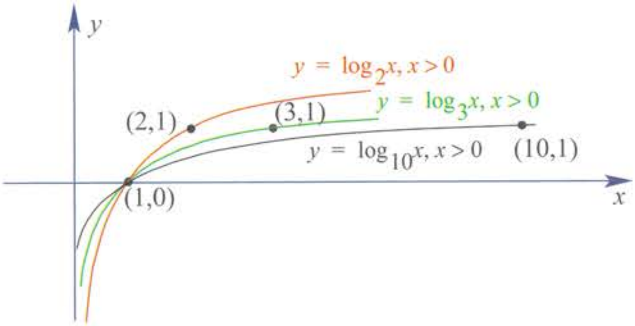


The implied domain of the basic logarithmic function is]0, ∞[and the y-axis is its asymptote.

So, how do the graphs of $y = \log_e x$, $y = \log_{10} x$, $y = \log_3 x$ compare to $y = \log_2 x$? The best way to see this is to sketch the graphs on the same set of axes as the diagram on the following page shows:

The implied domain of the basic logarithmic function with any positive base is]0, ∞[and has the y-axis as its asymptote.

Observe that each of the logarithmic functions is a reflection about the line $y = x$ of its corresponding exponential function (of the same base).



Notice that for all $a > 0$ $\log_a 1 = 0$ and $\log_a a = 1$.

As is the case for the exponential functions, the base 'e' also plays an important role when dealing with logarithmic functions. When using the number 'e' as the base for the logarithmic function, we refer to it as the **natural logarithmic function** and can write it in one of two ways:

$$f(x) = \log_e x,\ x > 0 \text{ or } f(x) = \ln x,\ x > 0$$

Example 2.4.6

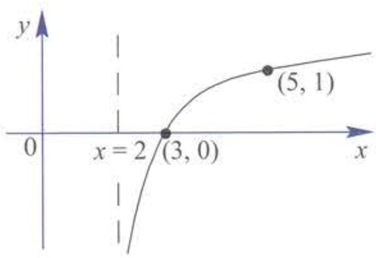
Sketch the following, specifying the implied domain in each case.

a $\quad f(x) = \log_3(x-2)$ $\quad$ b $\quad g(x) = \log_2(2x+3)$

a

A '2 right' translation.

Domain of $f = (2, \infty)$.



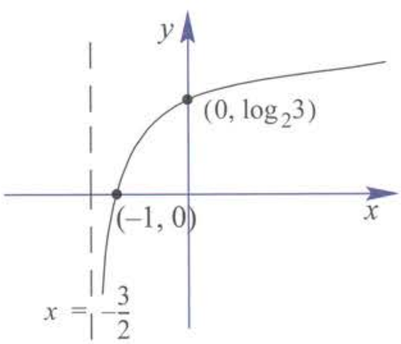
b

Looking at the domain:

$2x + 3 > 0 \Leftrightarrow x > -\frac{3}{2}$.

Therefore, the implied domain of $g = \left(-\frac{3}{2}, \infty\right)$.

The vertical asymptote is: $x = -\frac{3}{2}$.



Video

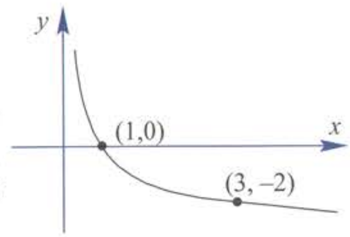
Example 2.4.7

Sketch the following, specifying the implied domain in each case.

a $\quad f(x) = -2\log_3 x$ $\quad$ b $\quad f(x) = \frac{1}{3}\log_e x$

a The implied domain in this case is $x > 0$. So, the vertical asymptote has the equation $x = 0$.

We note that the negative sign in front of the $\log_3 x$ will have the effect of reversing the sign of the $\log_3 x$ values.



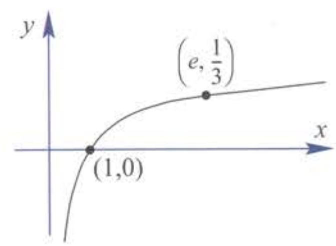
That is, the graph of $f(x) = -\log_3 x$ is a reflection about the x-axis of the graph of $y = \log_3 x$.

The factor of 2 will have the effect of 'stretching' the graph of $y = \log_3 x$ by a factor of 2 along the y-axis.

Also, we have that $f(1) = -2\log_3 1 = 0$ and $f(3) = -2\log_3 3 = -2$.

b This time the implied domain is $]0, \infty[$.

Therefore, the equation of the asymptote is $x = 0$. The one third factor in front of $\log_e x$ will have the effect of 'shrinking' the graph of $y = \log_e x$ by a factor of 3.



Then, $f(1) = {}^1/_3 \log_e 1 = 0$ and $f(e) = {}^1/_3 \log_e e = {}^1/_3$.

Again, we have the following observations:

The graph of $y = k \times \log_a x,\ k > 0$ is identical to $y = \log_a x$ but

i $\quad$ Stretched along the y-axis if $k > 1$.

ii $\quad$ Shrunk along the y-axis if $0 < k < 1$.

The graph of $y = k \times \log_a x,\ k < 0$ is identical to $y = \log_a x$ but

i $\quad$ Reflected about the x-axis and stretched along the y-axis if $k < -1$.

ii $\quad$ Reflected about the x-axis and shrunk along the y-axis if $-1 < k < 0$.

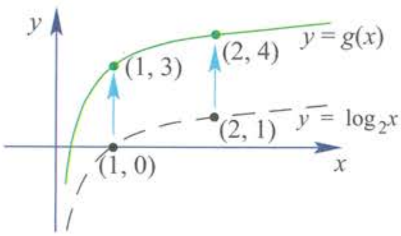
Example 2.4.8

Sketch the following, specifying the implied domain in each case.

a $\quad g(x) = \log_2 x + 3 \qquad$ b $\quad g(x) = \log_2 x - 2$

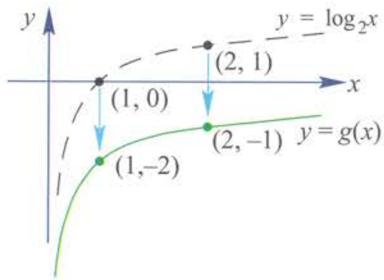
c $\quad g(x) = \log_2(3 - x)$

a The effect of adding 3 to the graph of $y = \log_2 x$ will result in $g(x) = \log_2 x + 3$ being moved up 3 units.

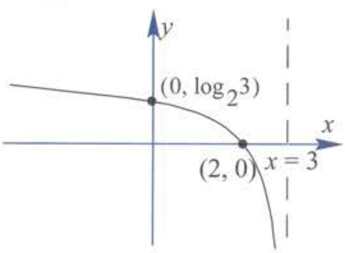


Its implied domain is $(0, \infty)$ and its asymptote has equation $x = 0$.

b The effect of subtracting 2 from the graph of $y = \log_2 x$ will result in $g(x) = \log_2 x - 2$ being moved down 2 units. Its implied domain is $(0, \infty)$ and its asymptote has equation $x = 0$.



c As we cannot have the logarithm of a negative number we must have that $3 - x > 0 \Leftrightarrow x < 3$.



This means that the vertical asymptote is given by $x = 3$ and the graph must be drawn to the left of the asymptote.

Exercise 2.4.3

1. Sketch the graph of the following functions, clearly stating domains and labelling asymptotes.

   a $\quad f(x) = \log_4(x - 2) \qquad$ b $\quad f(x) = \log_2(x + 3)$

   c $\quad h(x) = \log_{10} x + 2 \qquad$ d $\quad g(x) = -3 + \log_3 x$

   e $\quad f(x) = \log_5(2x - 1) \qquad$ f $\quad h(x) = \log_2(2 - x)$

   g $\quad g(x) = 2\log_{10} x \qquad$ h $\quad f(x) = -\log_{10} x + 1$

2. Sketch the graph of the following functions, clearly stating domains and labelling asymptotes.

   a $\quad f(x) = 2\log_2 x + 3 \qquad$ b $\quad f(x) = 10 - 2\log_{10} x$

   c $\quad h(x) = 2\log_2 2(x - 1)$

   d $\quad g(x) = -\frac{1}{2}\log_{10}(1 - x)$

   e $\quad f(x) = \log_2(3x + 2) - 1$

   f $\quad h(x) = 3\log_2\left(\frac{1}{2}x - 1\right) + 1$

3. Sketch the graph of the following functions, clearly stating domains and labelling asymptotes.

   a $\quad f(x) = 2\ln x \qquad$ b $\quad g(x) = -5\ln x$

   c $\quad f(x) = \ln(x - e) \qquad$ d $\quad f(x) = \ln(1 - ex)$

   e $\quad f(x) = 5 - \ln x \qquad$ f $\quad h(x) = \ln x - e$

4. Sketch the graph of the following functions, clearly stating domains and labelling asymptotes.

   a $\quad f(x) = \log_2 \sqrt{x} \qquad$ b $\quad f(x) = \log_{10} x^2$

   c $\quad h(x) = \ln|x| \qquad$ d $\quad g(x) = \ln\left(\frac{1}{x}\right)$

   e $\quad h(x) = \ln(1 - x^2) \qquad$ f $\quad f(x) = \log_2(x^2 - 4)$

5. Sketch the graph of the following functions, clearly stating domains and labelling asymptotes.

   a $\quad f(x) = |\log_{10} x|$

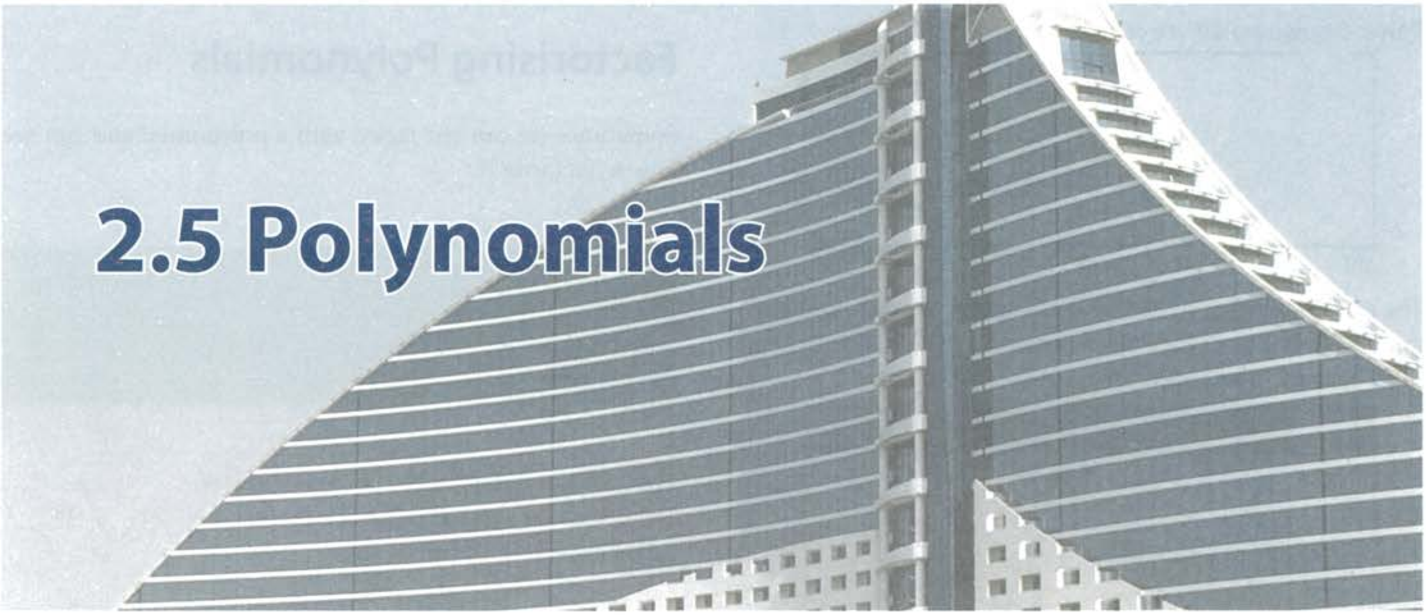
   b $\quad g(x) = |\log_2(x - 1)|$

   c $\quad h(x) = |\ln x - 1| \qquad$ d $\quad h(x) = 2 - |\ln x|$

   e $\quad f(x) = \log_2|x + 2| \qquad$ f $\quad f(x) = \log_5\left|\frac{1}{x}\right|$

Extra questions

Answers



2.5 Polynomials

This chapter will deal with three important results and the ways in which they can help us sketch the graphs of polynomials.

Polynomials are functions of the form:

$$P(x)=a_0+a_1x+a_2x^2+...+a_{n-1}x^{n-1}+a_nx^n$$

Remainder Theorem

If a polynomial $P(x)$ is divided by a linear polynomial $(x - a)$, the remainder is $P(a)$.

In general: dividend = divisor ×quotient + remainder

Factor Theorem

If, when a polynomial $P(x)$ is divided by a linear polynomial $(x - a)$, the remainder $P(a)$ is zero, then $(x - a)$ is a factor of $P(x)$.

Fundamental Theorem of Algebra

Every polynomial equation of the form $P(z) = 0$, $z \in \mathbb{C}$, of degree $n \in \mathbb{Q}^+$ has at least one complex root.

This has the important result that:

A polynomial $P_n(z) = 0$, $z \in \mathbb{C}$, of degree $n \in \mathbb{Q}^+$, can be expressed as the product of n linear factors and, hence, produce exactly n solutions to the equation $P_n(z) = 0$.

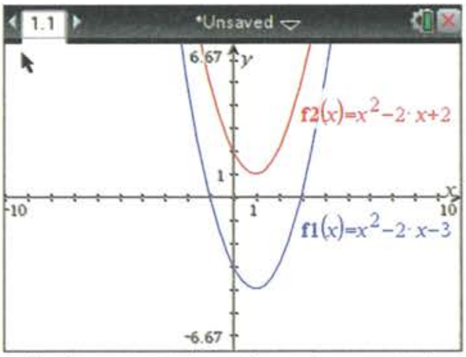
This does not, however, mean that, for example, all cubic equations have three **real** solutions.

The truth of these matters can best be understood by looking at the graphs of polynomials. A good way to do this is with a graphic calculator.

Polynomial Graphs

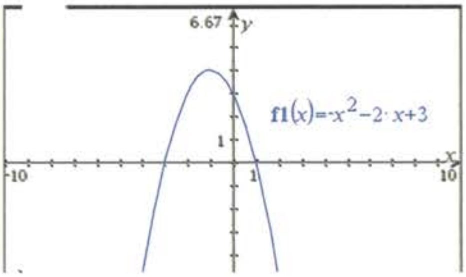
The basic shapes can be investigated using a graphic calculator. Linear functions in which the highest power is 1 are straight lines.

Power 2 polynomials are called quadratics. Their graphs are parabolas. These may or may not intersect the x-axis.

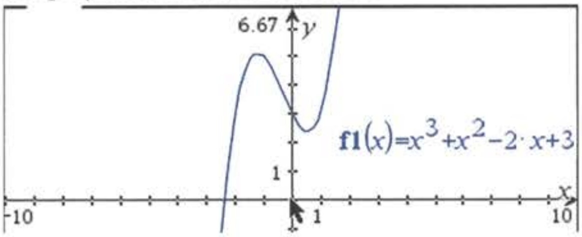


Note that all the principles of graph translation apply to polynomials.

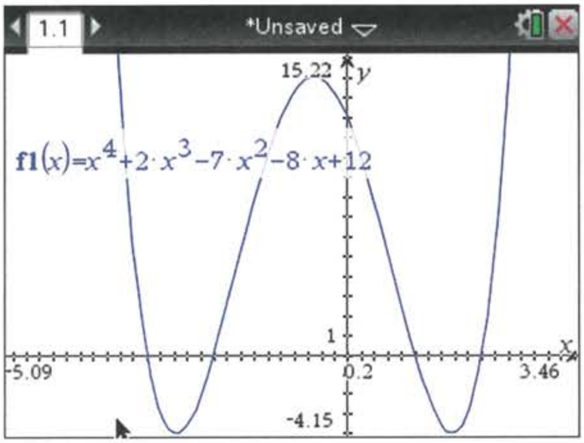
If the squared term is negative, the parabola will be 'inverted'.



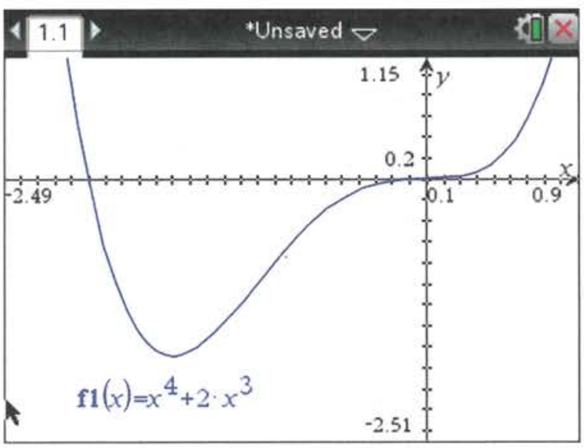
Power 3 polynomials are called cubics.



The extra power has added an extra 'hump' - a maximum or minimum. It seems likely that a power 4 polynomial will have three 'humps' - and this is often true:



There are, however, cases in which a maximum and minimum can coalesce to form a 'point of inflection'. At such points, the graph flattens out for a moment and then carries on in the same direction:



In this case, a maximum and minimum have become an inflection point at the origin.

With that qualification, it is possible to infer the general shape of the graph of a polynomial from the highest power (or degree) of the polynomial.

The other key features are the axes intercepts. The y intercept can be found by evaluating $P(0)$.

The x-intercepts are harder as we need to solve $P(x) = 0$. This is usually approached by factorisation using the Factor Theorem.

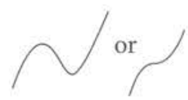
Factorising Polynomials

Sometimes we can 'get lucky' with a polynomial and can see how to factorise it.

Example 2.5.1

Sketch the graph of: $y=x^3-x^2-2x$

The polynomial is cubic. If $x\to\infty$, $y\to+\infty$. Thus, with the qualifications mentioned, we might expect the shape at right.



$x = 0$, $y = 0$ implies the graph passes through the origin.

The other x-intercepts must be found by solving $y = 0$.

or $x^3-x^2-2x=0$.

Since x is a common factor, an immediate factorisation is possible: $x\left(x^2-x-2\right)=0$

The quadratic will factorise using the inspection method:

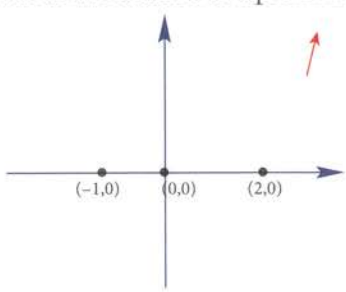
$x(x+1)(x-2)=0$

Now we use the 'null factor rule' to solve:

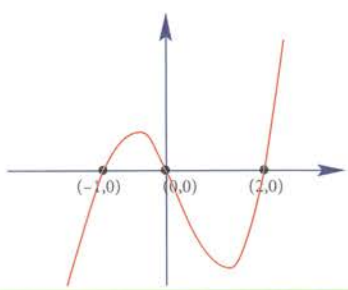
We already have $x = 0$

$x + 1 = 0 \Rightarrow x = -1$ and $x - 2 = 0 \Rightarrow x = 2$.

These features can now be added to a preliminary sketch:



which gives:



If the factorisation is not immediately obvious we use the Factor Theorem.

This works very like the factorisation of numbers. Suppose we were asked to factorise 126? We might begin by noticing that the number is even and so 2 is a factor. Once that is discovered there is a second factor that can be found by the division 126÷2. Many people will do this in their head. However, we will review the process of division as its algebraic version follows an identical pattern.

The common layout for a division of numbers is:

$$2\,\overline{)\,1\,2\,6}$$

Unlike the other three arithmetic processes (which move from right to left or from small numbers to large), division works from the large numbers to the small, or left to right.

This is because division is sharing. If this problem was to share \$126 between 2 people, we would probably begin by sharing the hundred dollar notes out first - which we cannot do as there is not enough money. Next we view it as 12 \$10 notes - so each person gets \$60. This leaves \$6 to share giving the answer as \$63.

This is usually written as:

1 $$\begin{array}{ll} & 0 \\ 2 & \overline{)\,1\,2\,6} \end{array}$$ Divide 1 by 2 (=0).

2 $$\begin{array}{ll} & 0 \\ 2 & \overline{)\,1\,2\,6} \\ & 0 \end{array}$$ Multiply the dividend (0) by the divisor (2) to get 0.

3 $$\begin{array}{ll} & 0 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0} \\ & 1 \end{array}$$ Subtract (1–0=1) to get the remainder 1.

4 $$\begin{array}{ll} & 0 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \end{array}$$ Include the next column to the right - "bring down".

REPEAT the 4 processes.

1 $$\begin{array}{ll} & 0\,6 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \end{array}$$ Divide 12 by 2 (=6).

2 $$\begin{array}{ll} & 0\,6 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & 1\,2 \end{array}$$ Multiply the dividend (6) by the divisor (2) to get 12.

3 $$\begin{array}{ll} & 0\,6 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & \underline{1\,2} \\ & \;\;0 \end{array}$$ Subtract (12–12=0) to get the remainder 0.

4 $$\begin{array}{ll} & 0\,6 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & \underline{1\,2}\downarrow \\ & \;\;0\,6 \end{array}$$ Include the next column to the right - "bring down".

REPEAT these processes.

1 $$\begin{array}{ll} & 0\,6\,3 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & \underline{1\,2}\downarrow \\ & \;\;0\,6 \end{array}$$ Divide 6 by 2 (=3).

2 $$\begin{array}{ll} & 0\,6\,3 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & \underline{1\,2}\downarrow \\ & \;\;0\,6 \\ & \;\;\;\;6 \end{array}$$ Multiply the dividend (3) by the divisor (2) to get 6.

3 $$\begin{array}{ll} & 0\,6\,3 \\ 2 & \overline{)\,1\,2\,6} \\ & \underline{0}\downarrow \\ & 1\,2 \\ & \underline{1\,2}\downarrow \\ & \;\;0\,6 \\ & \;\;\;\;\underline{6} \\ & \;\;\;\;0 \end{array}$$ Subtract (6–6=0) to get the remainder 0.

There are no numbers left.
The process is complete.
The answer is 63 remainder 0

We have shown this process in detail because keeping it in mind can help when working through a polynomial division. The technique is the same except it is performed with algebra instead of arithmetic.

Example 2.5.2

Factorise: $x^3-3x^2-10x+24$.

Hence sketch the graph of $y=x^3-3x^2-10x+24$.

There is no obvious common factor, so we use the Factor Theorem.

Let: $P(x)=x^3-3x^2-10x+24$

First, we look for a zero:

$$P(1)=1^3-3\times1^2-10\times1+24\neq0$$
$$P(-1)=(-1)^3-3\times(-1)^2-10\times(-1)+24\neq0$$
$$\begin{aligned}P(2)&=2^3-3\times2^2-10\times2+24\neq0\\&=8-12-20+24\\&=0\end{aligned}$$

By the Factor Theorem, $x-2$ is a factor of $P(x)$.

The other factor can be found by division: $P(x)\div(x-2)$.

1
$$\begin{array}{r|l} & x^2 \\ \hline x-2 & x^3-3x^2-10x+24 \end{array}$$
Divide x^3 by x $(=x^2)$. Only look at the highest powers.

2
$$\begin{array}{r|l} & x^2 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & x^3-2x^2 \end{array}$$
Multiply the dividend (x^2) by the divisor $(x-2)$ to get x^3-2x^2.

3
$$\begin{array}{r|l} & x^2 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2} \\ & \quad\;\, -x^2 \end{array}$$
Subtract $(x^3-3x^2-(x^3-2x^2)=-x^2)$ to get the remainder $-x^2$. Take care with signs!

4
$$\begin{array}{r|l} & x^2 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \end{array}$$
Include the next column to the right - "bring down".

1
$$\begin{array}{r|l} & x^2-x \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \end{array}$$
REPEAT the 4 processes.

Divide $-x^2$ by x $(=-x)$.

2
$$\begin{array}{r|l} & x^2-x \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, -x^2+2x \end{array}$$
Multiply the dividend $(-x)$ by the divisor $(x-2)$ to get $-x^2+2x$.

3
$$\begin{array}{r|l} & x^2-x \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, \underline{-x^2+2x} \\ & \qquad\quad -12x \end{array}$$
Subtract $(-x^2-10x-(-x^2+2x)=-12x)$ to get the remainder $-12x$.

4
$$\begin{array}{r|l} & x^2-x \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, \underline{-x^2+2x}\;\;\downarrow \\ & \qquad\quad -12x+24 \end{array}$$
Include the next column to the right - "bring down".

1
$$\begin{array}{r|l} & x^2-x-12 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, \underline{-x^2+2x}\;\;\downarrow \\ & \qquad\quad -12x+24 \end{array}$$
REPEAT these processes.

Divide $-12x$ by x $(=-12)$.

2
$$\begin{array}{r|l} & x^2-x-12 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, \underline{-x^2+2x}\;\;\downarrow \\ & \qquad\quad -12x+24 \\ & \qquad\quad -12x+24 \end{array}$$
Multiply the dividend (-12) by the divisor $(x-2)$ to get $-12x+24$.

3
$$\begin{array}{r|l} & x^2-x-12 \\ \hline x-2 & x^3-3x^2-10x+24 \\ & \underline{x^3-2x^2}\;\downarrow \\ & \quad\;\, -x^2-10x \\ & \quad\;\, \underline{-x^2+2x}\;\;\downarrow \\ & \qquad\quad -12x+24 \\ & \qquad\quad \underline{-12x+24} \\ & \qquad\qquad\qquad 0 \end{array}$$
Subtract $(-12x+24-(-12x+24)=0)$ to get the remainder 0.

This means that: $P(x)=(x-2)(x^2-x-12)$

The quadratic factor can now be factorised as a trinomial:

$$P(x)=(x-2)(x+3)(x-4)$$

We can now set about sketching the graph of:

$$y=x^3-3x^2-10x+24$$

If $x=0$, $y=24$.

If $y=0$, $\quad x^3-3x^2-10x+24=0$

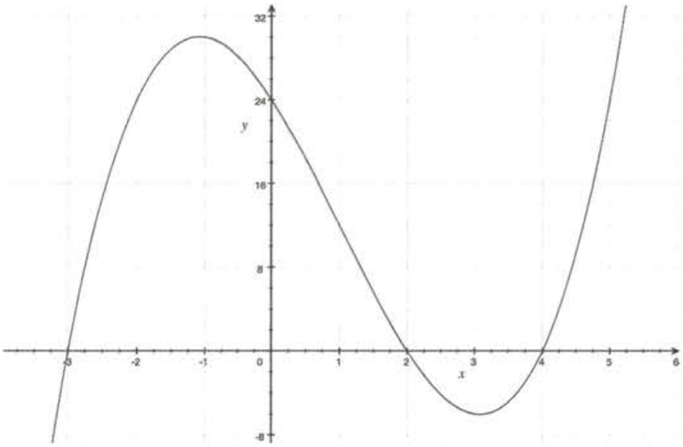
$$(x-2)(x+3)(x-4)=0$$
$$x-2=0\Rightarrow x=2$$
$$x+3=0\Rightarrow x=-3$$
$$x-4=0\Rightarrow x=4$$

We now have all the intercepts: (0,24), (−3,0), (2,0) & (4,0)

Once these four points are on the graph, and with the general shape of the cubic in mind, the sketch can be completed.

An alternative method (synthetic division) is discussed at:

What follows is a computer derived plot. Note that the 'humps' are not symmetric.



Example 2.5.3

Factorise: $P(x)=x^3-2x^2-5x+6$.

Hence sketch the graph of $y=x^3-2x^2-5x+6$.

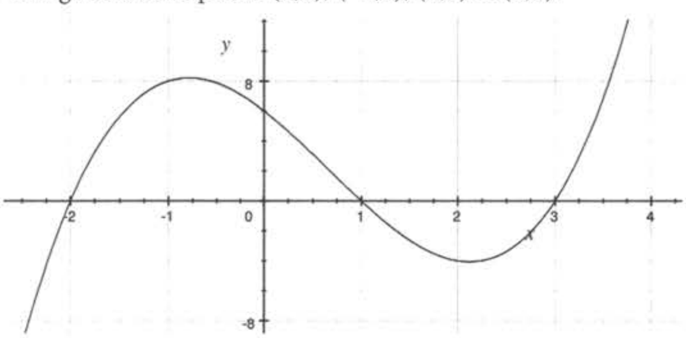
Using the Factor Theorem: $P(1)=1^3-2\times1^2-5\times1+6$
$$=1-2-5+6$$
$$=0$$

By the Factor Theorem, $x-1$ is a factor of $P(x)$. The other factor can be found by division: $P(x)\div(x-1)$. The required division should look like this:

$$\begin{array}{r|l} & x^2 - x \quad - 6 \\ \hline x-1 & x^3-2x^2-5x+6 \\ & x^3 - x^2 \\ \hline & \quad -x^2 - 5x \\ & \quad -x^2 + x \\ \hline & \qquad -6x + 6 \\ & \qquad -6x + 6 \\ \hline & \qquad\qquad 0 \end{array}$$

So: $P(x)=x^3-2x^2-5x+6$
$$=(x-1)(x^2-x-6)$$
$$=(x-1)(x-3)(x+2)$$

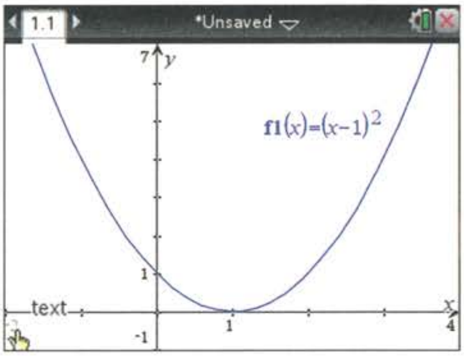
This gives intercepts of: (0,6), (−2,0), (1,0) & (3,0).



Repeated Factors

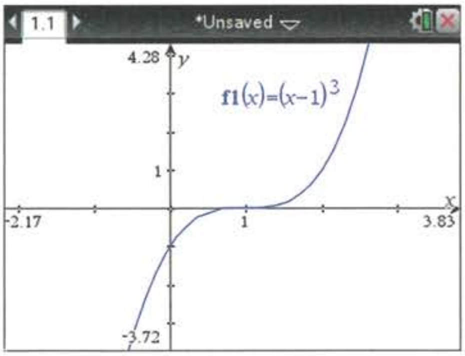
We have a further complication when trying to link the degree of a polynomial with the number of zeros (solutions of $P(x)=0$ and hence the number of intercepts on the graph of $y=P(x)$.

This is illustrated by the graph of $y=(x-1)^2$. This should have two intercepts as it is of degree 2. However, this is not the case. As using a graphic calculator is a good way of understanding this issue, we will use screen grabs in this section.



There is an intercept at (1,0), but it is a 'toucher'. The graph touches, but does not cut, the x-axis.

What if the power of the repeated factor is 3?



This time, the intercept is an inflection point and an intercept all in one. It is a good idea to use a calculator to see what happens if the factor is repeated even more times.

The even powers give **touching intercepts** with the graph getting flatter the higher the power.

The odd powers give **inflection intercepts** with the graph getting flatter the higher the power.

With a Casio model, use the Dyna Graph Module (6).

E.g. plot $y=(x-1)^A$.

Example 2.5.4

Sketch the graph of: $y=(x+1)^2(x-1)^3(x-3)$

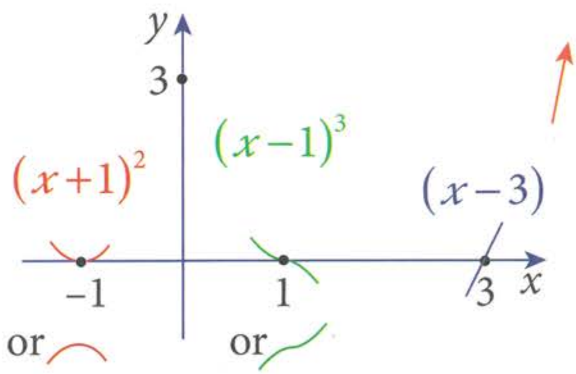
$$y=(x+1)^2(x-1)^3(x-3)$$

The red term (repeated twice) gives a touching intercept at (-1,0).

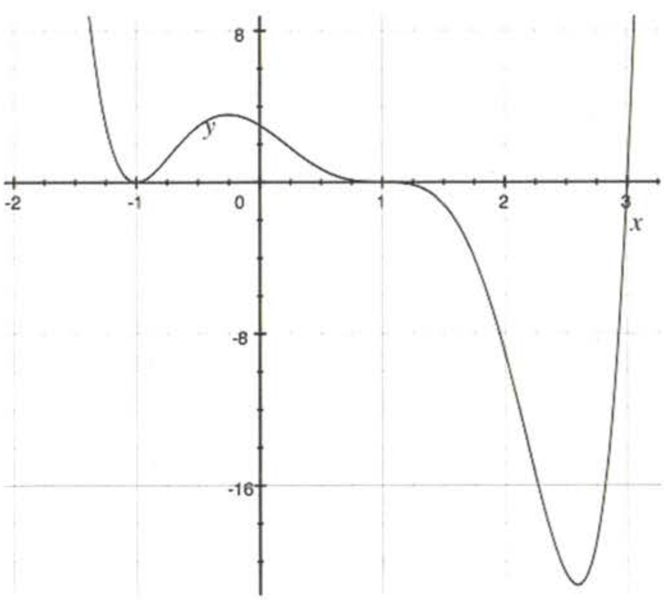
The green term (repeated three times) gives an inflection intercept at (1,0).

The blue term (one only) gives a cutting intercept at (3,0).

The y-intercept ($x = 0$) is (0,3). If $x \to \infty$, $y \to +\infty$.



The actual graph is:



Rational Root Theorem

When given a higher degree polynomial, it is not always easy to find the first zero to begin factorisation. Instead of randomly selecting a value from the real number system, you can consider using the Rational Root Theorem to identify a subset of potential zeros for the given polynomial.

The **Rational Root Theorem** states that if

$$f(x)=a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2 x^2 + a_1 x^1 + a_0$$

is a polynomial of degree n, then the subset of potential zeros of this given polynomial is $\frac{p}{q}$ where p is the list of the integer factors of a_0 and q is the list of factors of a_n.

Example 2.5.5

State all the possible rational roots of:

$f(x)=x^4-8x^3+3x^2+40x-12$

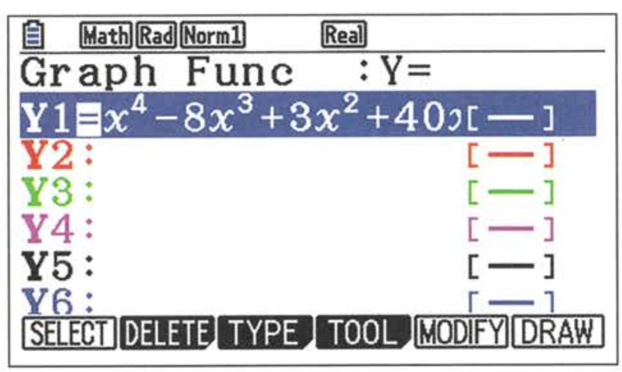
$a_0=-12$ so the list of factors p is: $\{\pm1,\pm2,\pm3,\pm4,\pm6,\pm12\}$.

$a_n=1$ so the list of factors q is: $\{\pm1\}$.

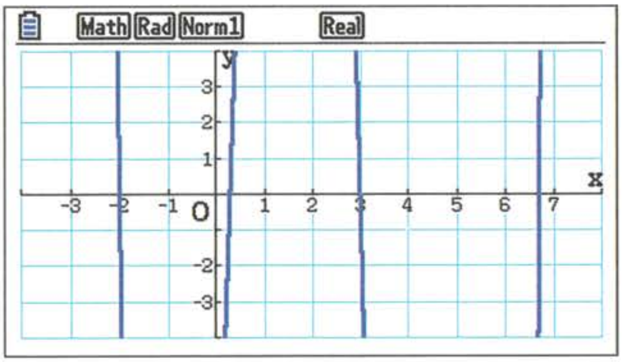
Hence, the list of potential rational roots is $\pm\frac{1,2,3,4,6,12}{1}$.

This will cut down the amount of 'trial and error' involved in using the Factor Theorem.

The other way of doing that is to use a calculatore to draw the graph:

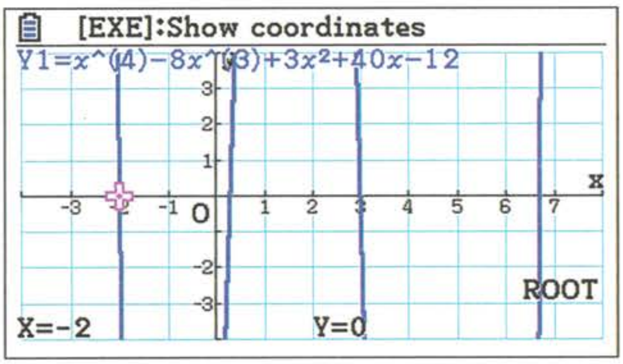


Always remember that you may need to adjust the viewing window. Don't sit looking at your calculator wondering why it "isn't working" when all that has happened is that the graph is off the screen!

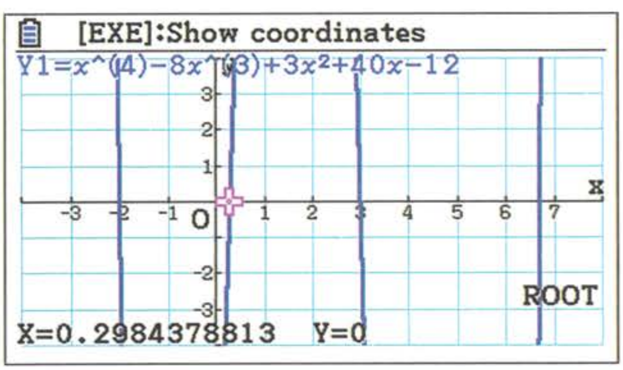


This suggests that –2 is an intercept and hence $(x + 2)$ is a factor of the polynomial.

If necessary, use G-Solve, or Analyse Graph etc. to identify the roots more precisely:



In this case, two of the roots appear to be non-integer.



Theory of Knowledge

It is not very often that mathematics texts get to recount a tale of passion and revenge.

However, the search for the solutions of polynomial equations is such an opportunity and we are not going to pass it up.

The solution of polynomial equations started with linear and quadratic equations which were 'cracked' quite early on in the History of Mathematics.

However, cubics and higher orders presented a much tougher set of problems.

Some progress had been made in China by Wang Xiaotong in the 7th century and by the Persian poet and scholar Omar Khayyám (1048–1131), both of whom solved a few cubic equations.

The solution to the general cubic, however, remained elusive.

In Bologna at the beginning of the 16th Century, it had become fashionable for the Universtity to stage problem solving competitions. These were a popular 'spectator sport' and drew large crowds.

Two mathematicians, Antonio Fiore and Niccolò Tartaglia (pictured) claimed success with the cubic.

The gauntlet was thrown down and a showdown was arranged in which cubic equations were to be solved against the clock. Tartaglia won.

At this point, a third player, Gerolamo Cardano (pictured) , entered the fray.

Cardano succeeded in persuading Tartaglia to tell him his secret. Tartaglia agreed on the condition that Cardano was not to reveal the method to anyone.

However, Cardano shared the secret with a student Lodovico Ferrari with the result that they extended the method to the general solution of the quartic.

What followed was one of the bitterest disputes in the History of Mathematics.

You can read more about these two colourful individuals at:

and

Exercise 2.5.1

1. Sketch the graphs of the following polynomials:

   a $P(x) = x(x-2)(x+2)$

   b $P(x) = (x-1)(x-3)(x+2)$

   c $T(x) = (2x-1)(x-2)(x+1)$

   d $P(x) = \left(\frac{x}{3}-1\right)(x+3)(x-1)$

   e $P(x) = (x-2)(3-x)(3x+1)$

   f $T(x) = (1-3x)(2-x)(2x+1)$

2. Sketch the graph of the following polynomials:

   a $P(x) = x^3-4x^2-x+4$

   b $P(x) = x^3-6x^2+8x$

   c $P(x) = 6x^3+19x^2+x-6$

   d $P(x) = -x^3+12x+16$

   e $P(x) = x^4-5x^2+4$

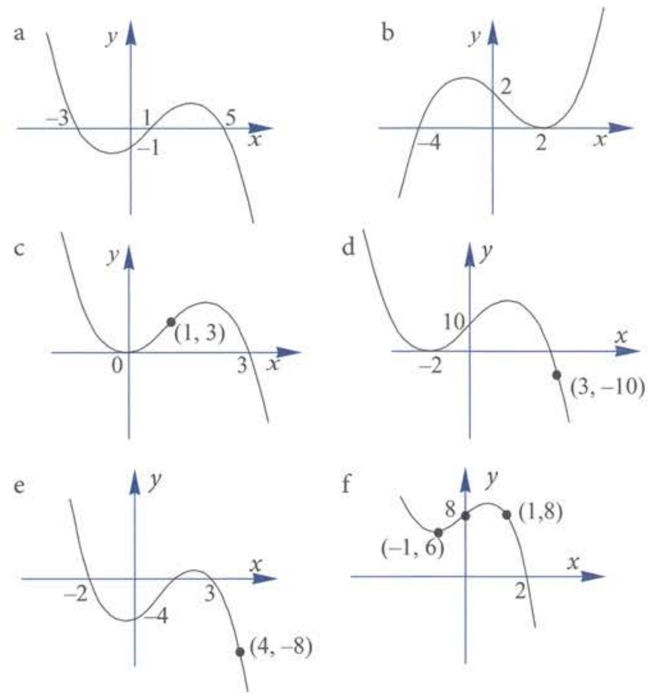
   f $P(x) = 3x^3-6x^2+6x-12$

3. Sketch the graphs of:

   a) $P(x) = x^3-kx$ where i $k = b^2$ ii $k = -b^2$.

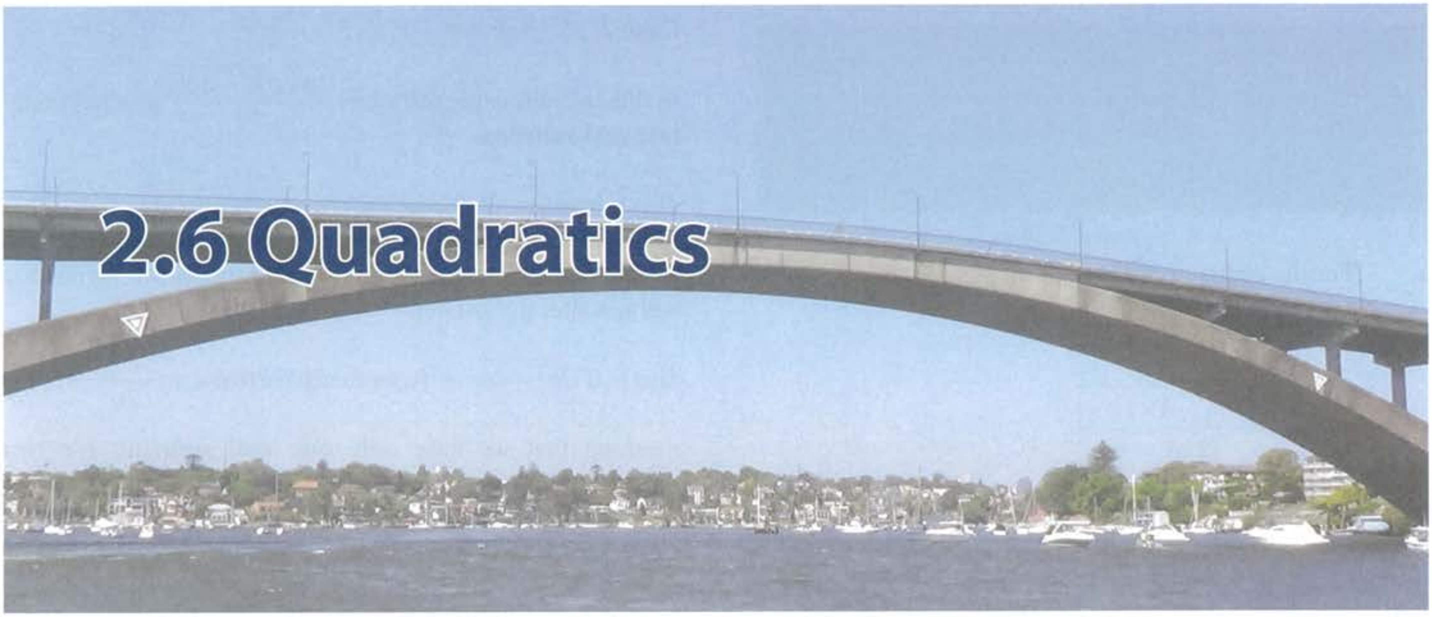
   b $P(x) = x^3-kx^2$ where i $k = b^2$ ii $k = -b^2$.

4. Determine the equations of the following cubic functions:



Extra questions

Answers



2.6 Quadratics

Quadratic Equation

A **quadratic equation** in the **variable** x (say) takes on the form $ax^2+bx+c=0$ where a, b and c are real constants. The equation is a quadratic because x is raised to the power of two.

The solution(s) to such equations can be obtained in one of two ways.

Method 1:Factorise the quadratic and use the Null Factor Law.

Method 2:Use the quadratic formula.

Method 1: Factorisation and the Null Factor Law

First of all we must have one side of the equation as 0, otherwise the Null Factor Law cannot be used. Next, when factorizing the quadratic, you will need to rely on your ability to recognise the form of the quadratic and hence which approach to use. A summary of the factorisation process for quadratics is shown below:

1. Trial and error:$(x+\alpha)(x+\beta)$
   e.g. $x^2+12x+32=(x+4)(x+8)$

2. Perfect square:$(x+\alpha)^2$ or $(x-\alpha)^2$
   e.g. $x^2+6x+9=(x+3)^2$

3. Difference of two squares: $(x+\alpha)(x-\alpha)$
   e.g. $x^2-16x=(x+4)(x-4)$

Note that sometimes you might need to use a perfect square approach on part of the quadratic and then complete the factorisation process by using the difference of two squares.

Example 2.6.1

Solve the quadratic $x^2+6x+7=0$.

In this instance it is not obvious what the factors are and so trial and error is not appropriate. However, we notice that x^2+6x+7 can be broken up into $x^2+6x+9-2$.

That is, part of the quadratic has been expressed as a perfect square, so that $x^2+6x+9-2=(x+3)^2-2$.

Then, we are left with a difference of perfect squares:

$(x+3)^2-2=\left(x+3+\sqrt{2}\right)\left(x+3-\sqrt{2}\right)$.

Therefore:

$$x^2+6x+7=0$$
$$\left(x+3+\sqrt{2}\right)\left(x+3-\sqrt{2}\right)=0$$
$$x=-3\pm\sqrt{2}$$

Method 2 Quadratic Formula and the Discriminant

A formula that allows us to solve any quadratic equation $ax^2+bx+c=0$ (if real solutions exist), is given by

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

Obtaining solutions requires that we make the appropriate substitution for a, b and c.

Example 2.6.2

Use the formula to solve the quadratic equations:

a $x^2 - x - 4 = 0$ b $2x^2 = 4 - x$

a For the equation, we have $a = 1$, $b = -1$ and $c = -4$,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$= \frac{1 \pm \sqrt{(-1)^2 - 4 \times 1 \times -4}}{2 \times 1}$$
$$= \frac{1 \pm \sqrt{17}}{2}$$

b Similarly, if $2x^2 = 4 - x$, then, $2x^2 + x - 4 = 0$ so that $a = 2$, $b = 1$ and $c = -4$, so that

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$= \frac{-1 \pm \sqrt{(1)^2 - 4 \times 2 \times -4}}{2 \times 2}$$
$$= \frac{-1 \pm \sqrt{33}}{4}$$

The Discriminant

Closer inspection of this formula indicates that much can be deduced from the term under the square root sign, i.e. $b^2 - 4ac$. The expression $b^2 - 4ac$ is known as the **discriminant** and is often represented by the delta symbol $\Delta = b^2 - 4ac$.

In particular, there are three cases to address:

Case 1. $b^2 - 4ac > 0$

In this case, the expression $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ produces **two real solutions**

Taking the square root of a positive number will produce another positive real number. This implies that there will be one solution corresponding to the '+' term and one solution corresponding to the '–' term.

That is, say that $\sqrt{b^2 - 4ac} = K$, where K is a real number.

We then have that $x = \frac{-b \pm K}{2a}$, i.e. $x_1 = \frac{-b + K}{2a}$, $x_2 = \frac{-b - K}{2a}$

giving **two distinct real solutions**.

Case 2. $b^2 - 4ac = 0$

In this case, the expression $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ produces only **one real solution**.

This is because taking the square root of zero gives zero. This in turn implies that there will be only one solution because adding and subtracting '0' to the '$-b$' term in the numerator will not alter the answer.

That is, if $\sqrt{b^2 - 4ac} = 0$, we then have that $x = \frac{-b \pm 0}{2a} = -\frac{b}{2a}$

meaning that we have only **one real solution** (or two repeated solutions).

Case 3. $b^2 - 4ac < 0$

In this case, the expression $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ produces **no real solution**.

This is because the square root of a negative number will not produce a **real** number. This in turn implies that the formula cannot be utilized (if we are dealing with quadratic equations under the real numbers).

There are, however, two complex solutions.

Example 2.6.3

a Find the value(s) of m for which the equation $2x^2 + mx + 1 = 0$ has one real solution.

b Find the value(s) of k for which the equation $x^2 + 4x + k = 0$ has two real solutions.

a. For one real solution to exist, we must have that $\Delta = b^2 - 4ac = 0$.

For this quadratic we have that $\boldsymbol{a} = 2$, $\boldsymbol{b} = \boldsymbol{m}$ and $\boldsymbol{c} = 1$. Therefore, we need that $m^2 - 4 \times 2 \times 1 = 0$

$$\Leftrightarrow m^2 - 8 = 0$$
$$\Leftrightarrow (m - \sqrt{8})(m + \sqrt{8}) = 0$$
$$\therefore m = 2\sqrt{2} \text{ or } m = -2\sqrt{2}$$

b For 2 real solutions, we must have $\Delta = b^2 - 4ac > 0$.

For this quadratic we have that $\boldsymbol{a} = 1$, $\boldsymbol{b} = 4$ and $\boldsymbol{c} = \boldsymbol{k}$. Therefore, we need that $4^2 - 4 \times 1 \times k > 0$

$$\Leftrightarrow 16 - 4k > 0 \quad \Leftrightarrow 16 > 4k \quad \Leftrightarrow 4 > k.$$

i.e. the quadratic $x^2 + 4x + k = 0$ will have two real solutions as long as $k < 4$.

Example 2.6.4

Find k if the equation $x^2 - (k + 3)x + (k + 6) = 0$ has:

a 1 real root. b 2 real roots. c no real root.

a First we find an expression for the discriminant in terms of k. Using the values $a = 1$, $b = k + 3$ and $c = k + 6$, we have:

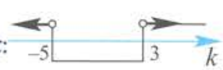
$$\Delta = b^2 - 4ac = (k+3)^2 - 4 \times 1 \times (k+6)$$
$$= k^2 + 2k - 15$$
$$= (k+5)(k-3)$$

For the equation to have 1 solution, the discriminant, $\Delta = 0$ thus,

$(k+5)(k-3) = 0 \Leftrightarrow k = -5$, or $k = 3$

That is, the solution set is $\{k: k = -5, 3\}$.

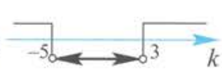
b For the equation to have 2 solutions, the discriminant, $\Delta > 0$, thus, $(k+5)(k-3) > 0 \Leftrightarrow k < -5$ or $k > 3$.

Using a sign diagram for k: 

That is, the solution set is $\{k : k < -5\} \cup \{k : k > 3\}$.

c For the equation to have no real solutions, the discriminant, $\Delta < 0$, thus,

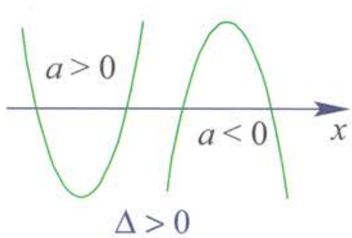
$$(k+5)(k-3) < 0 \Leftrightarrow -5 < k < 3$$

Using a sign diagram for k: 

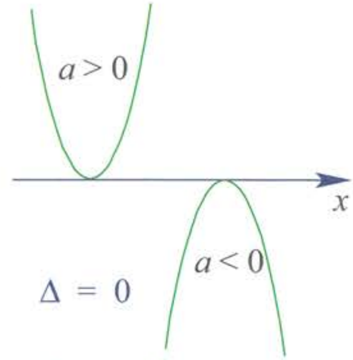
That is, the solution set is $\{k : -5 < k < 3\}$.

Graphical Interpretation

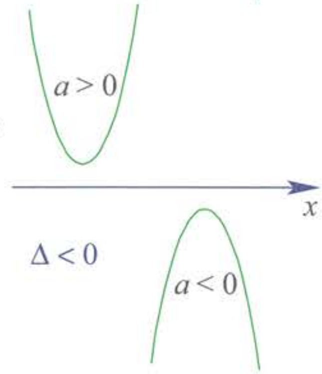
If $\Delta = b^2 - 4ac > 0$, then there are **two x–intercepts**.



If $\Delta = b^2 - 4ac = 0$, then there is **one x–intercept**.



If $\Delta = b^2 - 4ac < 0$, then there are **no x–intercepts**.



Exercise 2.6.1

1. By using a factorisation process, solve for x:

a $x^2 + 10x + 25 = 0$

b $x^2 - 10x + 24 = 0$

c $3x^2 + 9x = 0$

d $x^2 - 4x + 3 = 0$

2. Without using the quadratic formula, solve for the given variable.

a $u + \frac{1}{u} = -2$ b $x + 2 = \frac{35}{x}$

c $5x - 13 = \frac{6}{x}$ d $\frac{x}{2} - \frac{1}{x+1} = 0$

e $y + 1 = \frac{4}{y+1}$ f $v + \frac{20}{v} = 9$

3. By completing the square, solve for the given variable.

a $x^2 + 2x = 5$ b $x^2 + 4 = 6x$

c $x^2 - 2x = 4$ d $4x^2 + x = 2$

e $2y^2 = 9y - 1$ f $3a^2 - a = 7$

4. Use the quadratic formula to solve these equations.

a $x^2 - 3x - 7 = 0$ b $x^2 - 5x = 2$

c $x^2 - 3x - 6 = 0$ d $x^2 = 7x + 2$

e $x(x+7) = 4$ f $x^2+2x-8 = 0$

g $x^2+2x-7 = 0$ h $x^2+5x-7 = 0$

5. For what value(s) of p does the equation $x^2+px+1 = 0$ have:

   a no real solutions?

   b one real solution?

   c two real solutions?

6. Find the values of m for which the quadratic $x^2+2x+m = 0$ has:

   a one real solution.

   b two real solutions.

   c no real solutions.

7. Find the values of m for which the quadratic $x^2+mx+2 = 0$ has:

   a one real solution.

   b two real solutions.

   c no real solutions.

8. Find the values of k for which the quadratic $2x^2+kx+9 = 0$ has:

   a one real solution.

   b two real solutions.

   c no real solutions.

9. Consider the equation $x^2+2x = 7$. Prove that this equation has two real roots.

10. Find the value(s) of p such that the equation $px^2-px+1 = 0$ has exactly one real root.

11. Prove that the equation $kx^2+3x = k$ has two real solutions for all non-zero real values of k.

Extra questions

Sum and Product of the Roots of a Quadratic Equation

The quadratic equation $ax^2 + bx + c = 0$ can be rearranged to

$$x^2+\frac{bx}{a}+\frac{c}{a} = 0.$$

Note that the case $a = 0$ can be discounted as it refers to a linear equation.

Assume that the equation factorises to $(x-\alpha)(x-\beta) = 0$. We can then use the null factor rule to identify the roots of the equation as α and β.

Expanding $(x-\alpha)(x-\beta) = 0$ gives $x^2-(\alpha+\beta)x+\alpha\beta = 0$.

If this is to be identical to the original equation, then the coefficients of each term must be equal.

Original equation: $x^2+\frac{b}{a}+\frac{c}{a} = 0$

New version: $x^2-(\alpha+\beta)x+\alpha\beta = 0$

The coefficients of x^2 are both 1 and so are equal.

The coefficients of $\boldsymbol{x}$: $-(\alpha+\beta) = \frac{b}{a}$

which implies: $(\alpha+\beta)$ – 'the sum of the roots' $= -\frac{b}{a}$

The constants: $\alpha\beta$ – 'the product of the roots' $= -\frac{b}{a}$

It is thus possible to make statements about the roots of quadratic equations without actually solving them. This includes situations in which the equation does not have real roots ($b^2 < 4ac$).

Example 2.6.5

a Without solving the equation, find the sum and products of $2x^2 - 4x - 12 = 0$.

b Given that one of the roots of the equation $5x^2 + 7x - 3 = 0$ is $\frac{-7-\sqrt{109}}{10}$, find the other root without using the quadratic formula.

a Using the sum of the roots $= -\frac{b}{a}$ and the product of the roots $= \frac{c}{a}$:

Sum of the roots $= -\frac{-4}{2} = 2$

Product of the roots $= \frac{-12}{2} = -6$

b One of the roots is $\frac{-7-\sqrt{109}}{10}$.

If the second root is β, then, using the sum of the roots:

$$\beta + \frac{-7-\sqrt{109}}{10} = \frac{-7}{5}$$

$$\beta = -\frac{-7-\sqrt{109}}{10} + \frac{-7}{5} \quad = \frac{7+\sqrt{109}}{10} - \frac{14}{10} = \frac{-7+\sqrt{109}}{10}$$

Note that this confirms what might have been expected from the quadratic formula.

Exercise 2.6.2

1. Find the sum and product of the roots of each of these equations.

   a $x^2 + 2x + 4 = 0$

   b $x^2 - 3x - 7 = 0$

   c $x^2 - 3x - 3 = 0$

   d $5x^2 - 7x + 3 = 0$

   e $2x^2 + 5x - 3 = 0$

   f $-9x^2 + 4x + 2 = 0$

   g $3x^2 = 7x - 4$

   h $5x^2 + 8x = 13$

   i $\frac{4x+1}{4x-1} = 2x$

2. Given one of the roots of each of these equations, find the other.

   a $x^2 - 5x + 6 = 0,\ \alpha = 2$

   b $x^2 - 1 = 0,\ \alpha = 1$

   c $2x^2 - 7x + 3 = 0,\ \alpha = 3$

   d $6x^2 + x - 1 = 0,\ \alpha = \frac{1}{3}$

   e $9x^2 + 12x = 5,\ \alpha = \frac{1}{3}$

   f $10x^2 + 27 = 33x,\ \alpha = 1.5$

Extra questions

Indicial Equations

Solving equations of the form $x^{\frac{1}{2}} = 3$, where the **variable is the base**, requires that we square both sides of the equation so that:

$$\left(x^{\frac{1}{2}}\right)^2 = 3^2 \Rightarrow x = 9$$

However, when the **variable is the power** and not the base we need to take a different approach.

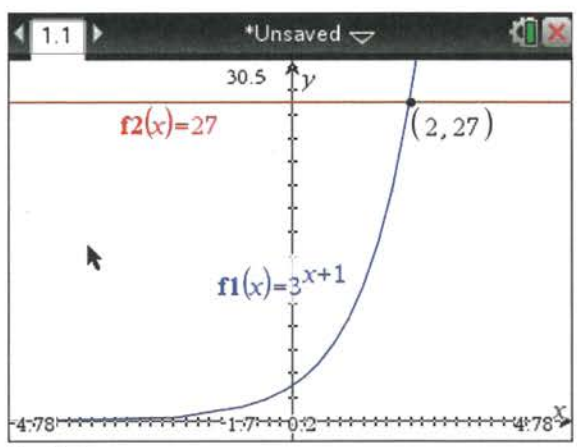
Consider the case where we wish to solve for x given that $2^x = 8$. In this case we need to think of a value of x so that when 2 is raised to the power of x the answer is 8. Using trial and error, it is not too difficult to arrive at $x = 3$ $(2^3 = 2 \times 2 \times 2 = 8)$.

Next consider the equation $3^{x+1} = 27$. Again, we need to find a number such that when 3 is raised to that number, the answer is 27. Here we have that $27 = 3^3$. Therefore we can rewrite the equation as $3^{x+1} = 3^3$.

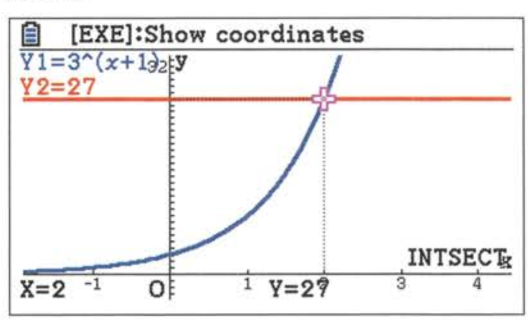
As the base on both sides of the equality is the same we can then equate the powers, that is:

$$3^{x+1} = 27 \Leftrightarrow 3^{x+1} = 3^3$$
$$\Leftrightarrow x + 1 = 3$$
$$\Leftrightarrow x = 2$$

This can be solved graphically by plotting $y=3^{(x+1)}$ and $y=27$. Then use Analyse Graph / Intersection.



or with Casio:



Example 2.6.6

Solve the following.

a $3^x = 81$ b $2 \times 5^u = 250$ c $2^x = \dfrac{1}{32}$

a $3^x = 81 \Leftrightarrow 3^x = 3^4$

b $2 \times 5^u = 250 \Leftrightarrow 5^u = 125$

$\Leftrightarrow 5^u = 5^3$

$\Leftrightarrow u = 3$

c $2^x = \dfrac{1}{32} \Leftrightarrow 2^x = \dfrac{1}{2^5}$

$\Leftrightarrow 2^x = 2^{-5}$

$\Leftrightarrow x = -5$

Example 2.6.7

Find:

a $\left\{x \mid \left(\dfrac{1}{2}\right)^x = 16\right\}$

b $\left\{x \mid 3^{x+1} = 3\sqrt{3}\right\}$

c $\left\{x \mid 4^{x-1} = 64\right\}$

a $\left(\dfrac{1}{2}\right)^x = 16 \Leftrightarrow (2^{-1})^x = 16$

$\Leftrightarrow 2^{-x} = 2^4$

$\Leftrightarrow -x = 4$

$\Leftrightarrow x = -4$

i.e. solution set is $\{-4\}$.

b $3^{x+1} = 3\sqrt{3} \Leftrightarrow 3^{x+1} = 3 \times 3^{1/2}$

$\Leftrightarrow 3^{x+1} = 3^{3/2}$

$\Leftrightarrow x + 1 = \dfrac{3}{2}$

$\Leftrightarrow x = \dfrac{1}{2}$

i.e. solution set is $\{0.5\}$

c $4^{x-1} = 64 \Leftrightarrow (2^2)^{x-1} = 2^6$

$\Leftrightarrow 2^{2x-2} = 2^6$

$\Leftrightarrow 2x - 2 = 6$

$\Leftrightarrow 2x = 8$

$\Leftrightarrow x = 4$

i.e. solution set is $\{4\}$.

Exercise 2.6.3

1. Solve the following equations:

   a $\{x \mid 4^x = 16\}$

   b $\left\{x \mid 7^x = \dfrac{1}{49}\right\}$

   c $\{x \mid 8^x = 4\}$

   d $\{x \mid 3^x = 243\}$

   e $\{x \mid 3^{x-2} = 81\}$

   f $\left\{x \mid 4^x = \dfrac{1}{32}\right\}$

   g $\{x \mid 3^{2x-4} = 1\}$

   h $\{x \mid 4^{2x+1} = 128\}$

   i $\{x \mid 27^x = 3\}$

2. Solve the following equations.

   a $\{x \mid 7^{x+6} = 1\}$

   b $\left\{x \mid 8^x = \dfrac{1}{4}\right\}$

   c $\{x \mid 10^x = 0.001\}$

   d $\{x \mid 9^x = 27\}$

   e $\{x \mid 2^{4x-1} = 1\}$

   f $\{x \mid 25^x = \sqrt{5}\}$

   g $\left\{x \mid 16^x = \dfrac{1}{\sqrt{2}}\right\}$

   h $\{x \mid 4^{-x} = 32\sqrt{2}\}$

   i $\{x \mid 9^{-2x} = 243\}$

Answers

2.7 Inequations

Linear Inequations

Inequations are solved in the same way as equations, with the exception that, when both sides are multiplied or divided by a negative number, the direction of the inequality sign reverses.

Example 2.7.1

Find:

a $\{x: x + 1 < 4\}$ b $\{x: 2x - 5 < 1\}$

a $x + 1 < 4 \Leftrightarrow x < 3$.

Therefore, the solution set (s.s.) is $\{x : x < 3\}$.

b $2x - 5 < 1 \Leftrightarrow 2x < 6$

$\Leftrightarrow x < 3$

Therefore, the solution set (s.s.) is $\{x : x < 3\}$.

Example 2.7.2

Find: a $\{x: x + 2 < 3 - 2x\}$ b $\left\{x: \frac{3-2x}{7} \leq \frac{4x-3}{2}\right\}$

a $x + 2 > 3 - 2x \Leftrightarrow 3x + 2 > 3$

$\Leftrightarrow 3x > 1 \quad \Leftrightarrow x > \frac{1}{3}$

Therefore, s.s. $= \left\{x : x > \frac{1}{3}\right\}$

b $\frac{3-2x}{7} \leq \frac{4x-3}{2} \Leftrightarrow 14(\frac{3-2x}{7}) \leq 14(\frac{4x-3}{2})$
(multiply both sides by 14)

$\Leftrightarrow 2(3 - 2x) \leq 7(4x - 3)$

$\Leftrightarrow 6 - 4x \leq 28x - 21$

$\Leftrightarrow -32x \leq -27$

$\Leftrightarrow x \geq \frac{27}{32}$ (notice the reversal of the inequality – as we divided by a negative number)

Therefore, s.s. is $\left\{x : x \geq \frac{27}{32}\right\}$.

Exercise 2.7.1

1. Solve the following inequalities.

   a $2x + 1 < x - 3$ b $\frac{x-4}{3} \geq 2x - 1$

   c $x + 1 > \frac{x+3}{2}$ d $x \geq 3(x + 4)$

   e $\frac{x-4}{5} > \frac{2-x}{2}$ f $1 - 3x < 5x - 2$

2. Solve the following inequalities.

   a $\frac{2x+1}{5} - \frac{2-x}{3} > 3$ b $\frac{1+x}{2} + \frac{1-x}{4} \leq 1$

   c $\frac{x}{5} + \frac{2-3x}{3} \geq -2$

3. Solve the following inequalities.

a $a(x+1) > 2a, a < 0$

b $\frac{a-x}{2} + 1 > a, a > 0$

c $\frac{x}{a} + \frac{b}{a^2} < \frac{4x}{a} - \frac{b}{a^2}, b > a > 0$

d $x + \frac{x-1}{a+1} \geq \frac{x+1}{a+1} - ax$, a > 0

Extra questions

Quadratic Inequalities

Quadratic inequations arise from replacing the '=' sign in a quadratic by an inequality sign. Solving inequations can be carried out in two ways, either algebraically or graphically.

Method 1: Algebraic Method

This method relies on factorizing the quadratic and then using the fact that, when two terms, a and b are multiplied, the following rules apply:

$ab > 0 \Leftrightarrow a > 0$ and $b > 0$ or $a < 0$ and $b < 0$ OR

$ab < 0 \Leftrightarrow a > 0$ and $b < 0$ or $a < 0$ and $b < 0$

Example 2.7.3
Find:

a $\{x \mid x^2 - 6x + 8 > 0\}$

b $\{x \mid 2x^2 + 5x - 3 \leq 0\}$

c $\{x \mid x^2 - 3 < 2x\}$

a We start by factorizing the quadratic:

$x^2 - 6x + 8 = (x-2)(x-4)$

Then, $x^2 - 6x + 8 > 0 \Leftrightarrow (x-2)(x-4) > 0$

Which means either: $x - 2 > 0$ **and** $x - 4 > 0$, i.e. $x > 2$ and $x > 4$ $\Rightarrow x > 4$ – (1)

or $x - 2 < 0$ **and** $x - 4 < 0$, i.e. $x < 2$ and $x < 4 \Rightarrow x < 2$ – (2)

Then, combining (1) and (2) we have $\{x \mid x^2 - 6x + 8 > 0\} = \{x \mid x < 2\} \cup \{x \mid x > 4\}$.

b Now, $2x^2 + 5x - 3 \leq 0 \Leftrightarrow (2x-1)(x+3) \leq 0$.

Meaning that either $2x - 1 \leq 0$ and $x + 3 \leq 0$, i.e. $x \leq \frac{1}{2}$ and $x \leq -3$ – (1)

or $2x - 1 \geq 0$ and $x + 3 \geq 0$, i.e. $x \geq \frac{1}{2}$ and $x \geq -3$ – (2)

From result (1) we have that $-3 \leq x \leq \frac{1}{2}$.

However, the inequalities in result (2) are inconsistent, i.e. we cannot have that x is both greater than or equal to ½ and less than or equal to –3 simultaneously. Therefore we discard this inequality.

Therefore, $\{x \mid 2x^2 + 5x - 3 \leq 0\} = \left\{x \mid -3 \leq x \leq \frac{1}{2}\right\}$

c This time we need some rearranging:

$x^2 - 3 < 2x \Leftrightarrow x^2 - 2x - 3 < 0 \Leftrightarrow (x+1)(x-3) < 0$

Then, we must have that $x + 1 < 0$ and $x - 3 > 0$, i.e. $x < -1$ and $x > 3$ – (1)

$x + 1 > 0$ and $x - 3 < 0$, i.e. $x > -1$ and $x < 3$ – (2)

This time (1) is inconsistent, so we discard it and from (2) we have $-1 < x < 3$.

Therefore, $\{x \mid x^2 - 3 < 2x\} = \{x \mid -1 < x < 3\}$

Method 2: Graphical Method

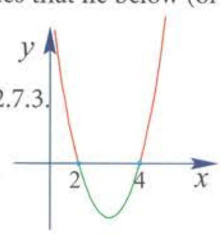
This method relies on examining the graph of the corresponding quadratic function and then:

1. quoting the x-values that produce y-values that lie above (or on) the x-axis (i.e. $y > 0$ or $y \geq 0$)

or

2. quoting x-values that produce y-values that lie below (or on) the x-axis (i.e. $y < 0$ or $y \leq 0$)

We consider inequations from Example 2.7.3.



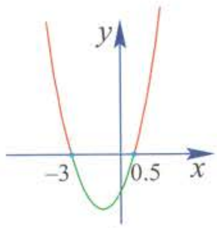
a The corresponding function in this case is $f(x) = x^2 - 6x + 8$.

That part of the graph corresponding to $f(x) > 0$ is highlighted in red. The values of x that correspond to these parts are $x < 2$ as well as $x > 4$.

Therefore, s.s. $= \{x | x < 2\} \cup \{x | x > 4\}$

b The corresponding function in this case is $f(x) = 2x^2 + 5x - 3$.

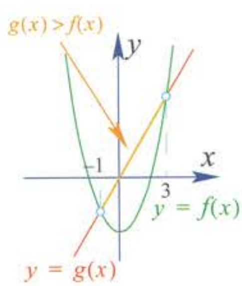
That part of the graph corresponding to $f(x) \leq 0$ is highlighted in green.



The values of x that correspond to these parts are $-3 \leq x \leq ½$.

Therefore, s.s. $= \left\{x | -3 \leq x \leq \frac{1}{2}\right\}$

c This time we have two functions, $f(x) = x^2 - 3$ and $g(x) = 2x$, and we want to find those values of x where $f(x) < g(x)$.



We do this by sketching both graphs on the same set of axes and then finding those values of x for which $f(x) < g(x)$ i.e. where the graph of $y = g(x)$ lies above that of $y = f(x)$.

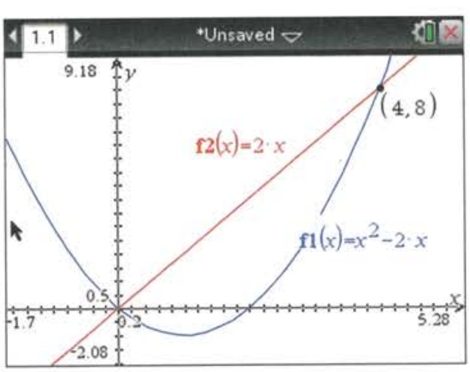
Once we have found the point of intersection, i.e. once we have solved $f(x) = g(x)$, we refer to the graph.

$$f(x) = g(x) \Leftrightarrow x^2 - 3 = 2x \Leftrightarrow x^2 - 2x - 3 = 0$$
$$\Leftrightarrow (x-3)(x+1) = 0$$
$$\Leftrightarrow x = 3 \text{ or } x = -1$$

Then, $f(x) < g(x)$ for $-1 < x < 3$.
i.e. $\{x | x^2 - 3 < 2x\} = \{x | -1 < x < 3\}$.

The graphical method, particularly when allied to the use of technology can be used to solve a range of inequations.

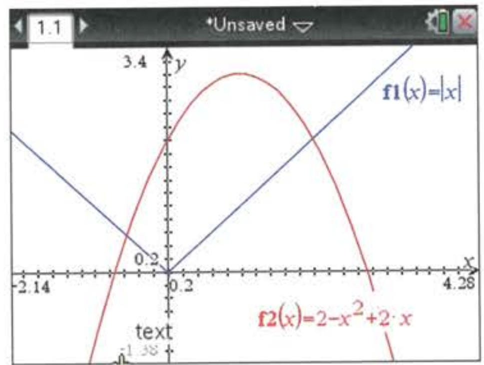
The screen that would solve part c above is:



Example 2.7.4

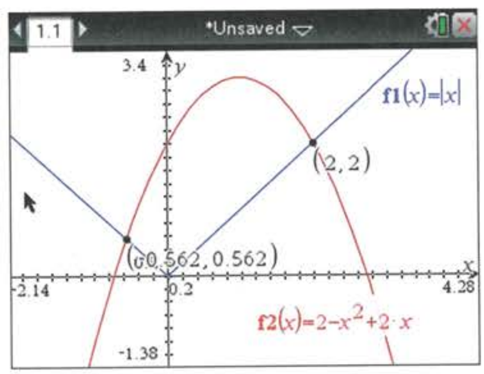
Use a graphical method to find: $\{x:|x| < 2 - x^2 + 2x\}$

Let $g(x) = |x|$ and $f(x) = 2 - x^2 + 2x$, we sketch these graphs on the same set of axes.



Next we need to find where the two graphs intersect. That is, we need to solve the equation $g(x) = f(x)$. So, we have: $|x| = 2 - x^2 + 2x$.

Using the intersect option:

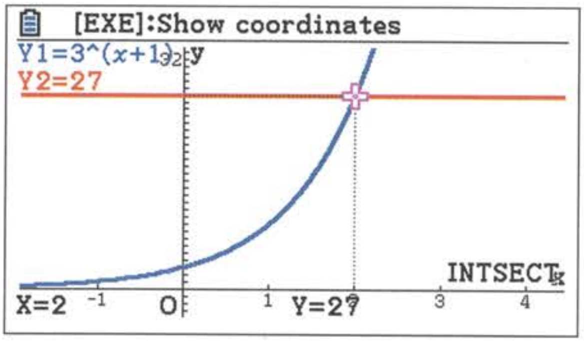


We are looking for the interval(s) for which the red graph is above the blue graph.

Therefore, from our results we have that:
$\{x | |x| < 2 - x^2 + 2x\} = \{x | -0.562 < x < 2\}$.

Notice once again, that the graphics calculator could only provide an approximate answer for one of the points of intersection.

Casio models use the Graph module followed by G-Solv.



If $(x-a)(x-b) \geq 0$, then the solution is $x \leq a$ and $x \geq b$.

If $(x-a)(x-b) \leq 0$, then the solution is $a \leq x \leq b$.

Exercise 2.7.2

1. Find the solution set for each of the following inequalities.

   a $(x-1)(x+2) > 0$

   b $(x+3)(x-2) \leq 0$

   c $x(4-x) \leq 0$

   d $(1-3x)(x-3) > 0$

   e $(3+2x)(x+1) \geq 0$

   f $(5-2x)(3-4x) < 0$

2. Find the solution set for each of the following inequalities.

   a $x^2+3x+2 > 0$

   b $x^2-x-6 < 0$

   c $2x^2-5x-3 \geq 0$

   d $x^2-4 \leq 0$

   e $x^2+x-5 < 0$

   f $-x^2+x+6 \leq 0$

   g $-x^2+x+1 \geq 0$

   h $-2x^2-3x+5 \geq 0$

   i $2x^2+5x-3 > 0$

   j $x^2-4x+3 < 0$

   k $2x^2+x-1 < 0$ l $x^2+3 < 0$

   m $-x^2-2 > 0$ n $2x^2-7x \leq 15$

   o $3x^2+5x > 2$

3. a For what value(s) of k is the inequation $x^2+2kx-k > 0$ true for all values of x?

   b For what value(s) of k is the inequation $x^2-kx+2 \geq 0$ true for all values of x?

   c For what value(s) of n is the inequation $x^2+2x \geq 2n$ true for all values of x?

4. By sketching on the same set of axes, the graphs of the functions $f(x)$ and $g(x)$, solve the following inequalities:

   i $f(x) < g(x)$ ii $f(x) \geq g(x)$

   a $f(x) = x+2$, $g(x) = x^2$

   b $f(x) = x-1$, $g(x) = x^2-4x+5$

   c $f(x) = x^2+2$, $g(x) = 4x-1$

   d $f(x) = 3x^2-1$, $g(x) = x+1$

   e $f(x) = 5-x^2$, $g(x) = x^2-3$

   f $f(x) = x^2-3x-3$, $g(x) = x-4$

5. On the same set of axes sketch the graphs of $f(x) = |x-1|$ and $g(x) = 1-x^2$. Hence find $\{x : |x-1| < 1-x^2\}$.

6. Given that $f(x) = x^2+3x+2$ and $g(x) = 4-x^2$, find $\{x \mid f(x) \leq g(x)\}$.

7. Find $\{x : |x^2-4x| < k\}$ for:

   i $k=2$ ii $k=4$ iii $k=8$

8. Find:

   i $\{x : |2x-3| \leq 3x-x^2\}$

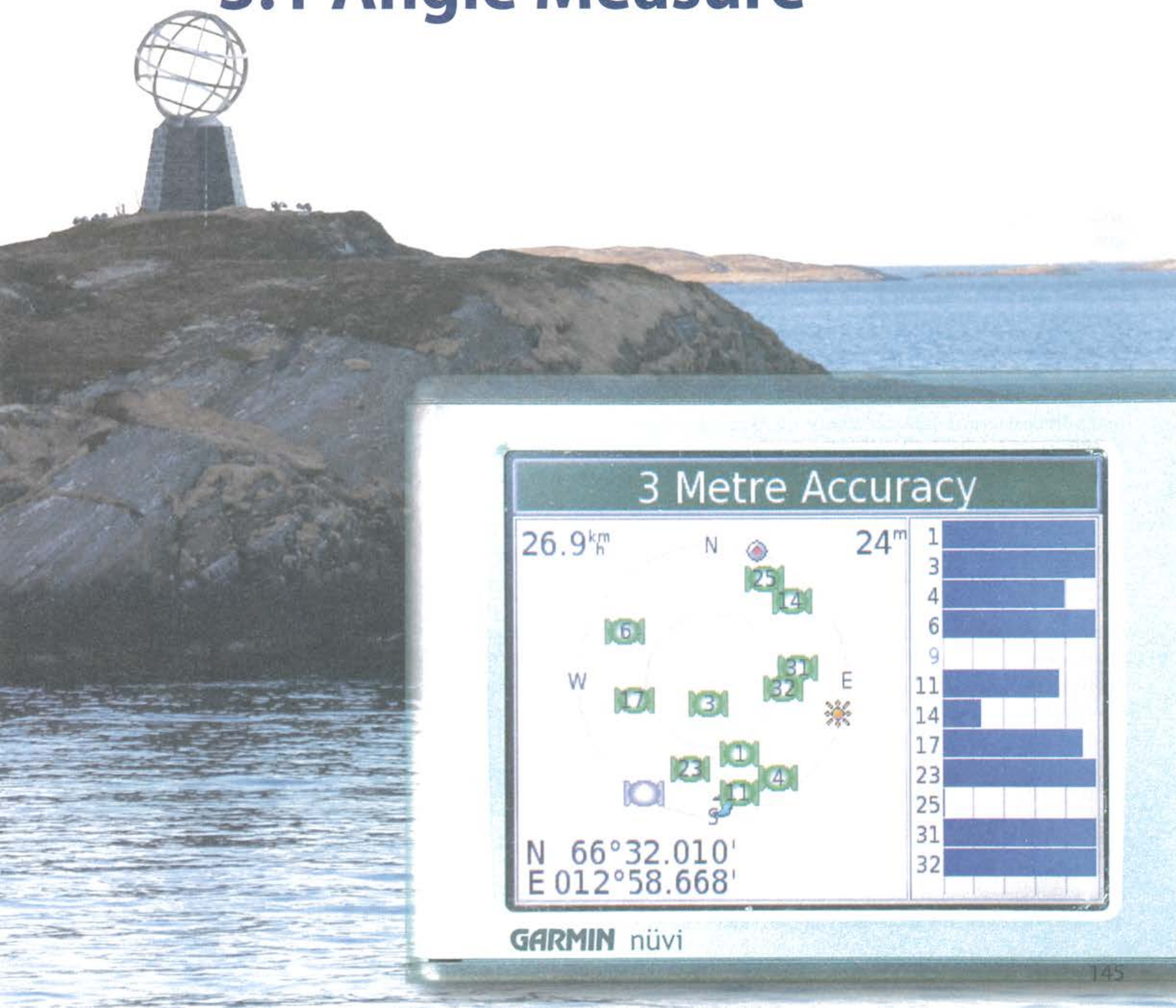
   ii $\left\{x : |3-|x|| \leq \left|3-\frac{1}{3}x^2\right|\right\}$

Answers

CHAPTER THREE

CIRCULAR FUNCTIONS AND TRIGONOMETRY

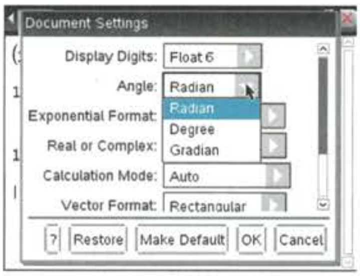
3.1 Angle Measure



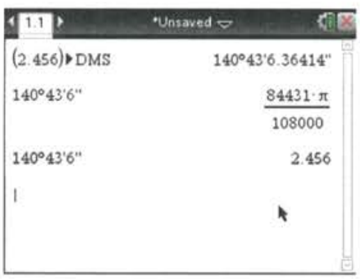
Our cover picture shows a modern GPS unit as it crosses the Arctic Circle - and the geographic marker that confirms it. It all depends on angles!

Radian Measure of an Angle

In Middle School Mathematics angles are measured in degrees. However, while this has been very useful, such measurements are not suitable for many topics in mathematics. Instead, we introduce a new measure, called the radian measure.



The degree measure of angle is based on dividing the complete circle into 360 equal parts known as degrees. Each degree is divided into sixty smaller parts known as minutes, and each minute is divided into sixty seconds. If using a calculator, you need to know how to set the calculator to radians or degrees.



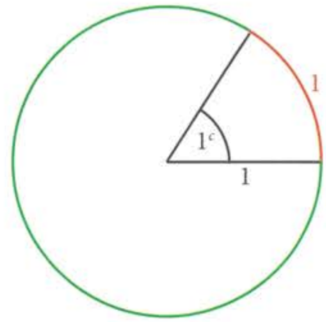
Decimal parts of a degree can be converted into degrees, minutes and seconds using the book key (to the right of the 9, pressing D to scroll quickly to DMS and then selecting the function. The calculator should be in degree mode.

This can be useful as calculators generally produce answers in the decimal format. It should also be noted that the degree, minute, second angle system is the same as the hours minutes seconds system that we use to measure time. The above screen could be interpreted as 2.456° and is equal to 2 degrees 27 minutes and 21.6 seconds or as 2.456 hours which is the same as 2 hours 27 minutes and 21.6 seconds. Note also that there is a difference between 'ENTER' and 'CTRL/ENTER'.

The degree system is arbitrary in the sense that the decision was made (in the past and due to astronomical measurements) to divide the complete circle into 360 parts. The radian system is an example of a natural measurement system.

One radian is defined as the ratio between the arc and the radius of the circle.

Two radians is the angle that gives an arc length of twice the radius etc. giving a natural linear conversion between the measure of a radian, the arc length and the radius of a circle.



A complete circle has an arc length of $2\pi r$.

It follows that a complete circle corresponds to $\frac{2\pi r}{r} = 2\pi$ radians.

This leads to the conversion factor between these two systems:

$360° = 2\pi$ radians or $180° = \pi$ radians (often written as π^c).

So, exactly how large is a radian?

Using the conversion above, if $360° = 2\pi^c$, then $1^c = \frac{360}{2\pi} \approx 57.2957°$

That is, the angle which subtends an **arc of length 1 unit** in a **circle of radius 1** unit, is **1 radian**.

More generally we have:

To convert from degrees to radians, multiply angle by $\frac{\pi}{180}$

To convert from radians to degrees, multiply angle by $\frac{180}{\pi}$

All conversions between the two systems follow this ratio. It is not generally necessary to convert between the systems as problems are usually worked either entirely in the degree system (as in the previous sections) or in radians (as in the functions and calculus chapters). In the case of arc length and sector areas, it is generally better to work in the radian system.

Example 3.1.1

Convert: a $70°$ into radians

b 2.34^c into degrees

c $\frac{\pi^c}{6}$ into degrees

Using the above conversion factors we have:

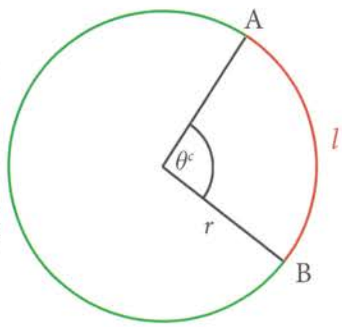
a $70° = 70 \times \frac{\pi^c}{180} = \frac{7\pi^c}{18}$ or 1.2217^c.

b $\quad 2.34^c = 2.34 \times \frac{180°}{\pi} = 134.0721° = 134°4'20''$

c $\quad \frac{\pi^c}{6} = \frac{\pi}{6} \times \frac{180°}{\pi} = 30°$

Arc length

As the arc length AB of a circle is directly proportional to the angle which AB subtends at its centre, then, the arc length AB is a fraction of the circumference of the circle of radius r.



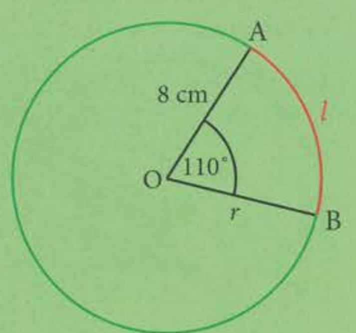
So, if the angle is θ^c, then the arc length is $\frac{\theta}{2\pi}$ of the circumference.

Then, the (minor) arc length, AB, denoted by l, is given by $l = \frac{\theta}{2\pi} \times 2\pi r = r\theta$.

The longer arc AB, called the major arc, has a length of $2\pi r - l$.

Example 3.1.2

Using the circle shown, find the arc length AB.



First we need to convert 110° into radian measure.

$$110° = 110 \times \frac{\pi^c}{180} = \frac{11\pi^c}{18}$$

Then, the arc length, l, is given by,

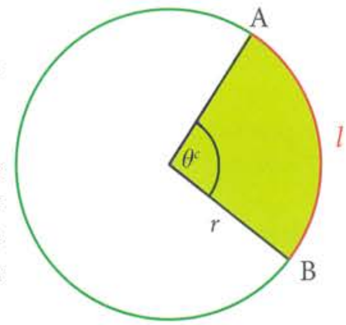
$$l = r\theta = 8 \times \frac{11\pi}{18} = \frac{44\pi}{9} = 15.3588\ldots$$

Therefore, the arc length is 15.36 cm.

Area of a Sector

The formula for the area of a sector is derived as follows:

If a sector is cut from a circle of radius r using an angle at the centre of θ radians, the area of the complete circle is πr^2.

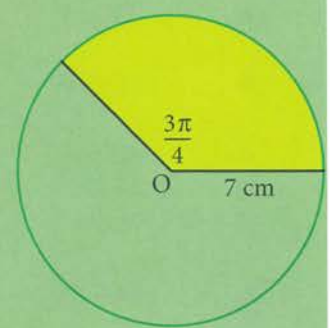


The fraction of the circle that forms the sector is $\frac{\theta}{2\pi}$ of the complete circle, so the area of the sector is:

$$\pi r^2 \times \frac{\theta}{2\pi} = \frac{1}{2}r^2\theta.$$

Example 3.1.3

Find the area and perimeter of the sector shown.



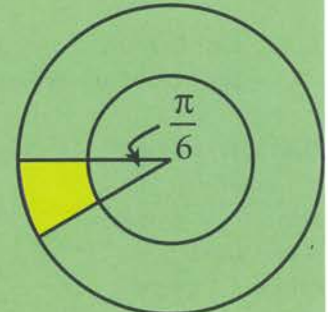
Area of sector $= \frac{1}{2} \times 7^2 \times \frac{3\pi}{4} = \frac{147\pi}{8}$ cm^2.

The perimeter is made up from two radii (14 cm) and the arc $l = r\theta = 7 \times \frac{3\pi}{4} = \frac{21\pi}{4}$.

The perimeter is $14 + \frac{21\pi}{4}$ cm.

Example 3.1.4

Find the area and perimeter of the shaded part of the diagram. The radius of the inner circle is 4 cm and the radius of the outer circle is 9 cm.



The angle of the shaded segment $= 2\pi - \frac{\pi}{6} = \frac{11\pi^c}{6}$

The shaded area can be found by subtracting the area of the sector in the smaller circle from that in the larger circle.

Shaded area =

$$\frac{1}{2}\times 9^2\times\frac{11\pi}{6}-\frac{1}{2}\times 4^2\times\frac{11\pi}{6} = \frac{11\pi}{12}(9^2-4^2) = 59\frac{7}{12}\pi\ \text{cm}^2$$

The perimeter is made up from two straight lines (each 9 – 4 = 5 cm long) and two arcs.

$$\text{Perimeter} = 10+4\times\frac{11\pi}{6}+9\times\frac{11\pi}{6} = 10+\frac{143\pi}{6}\ \text{cm}.$$

Exercise 3.1.1

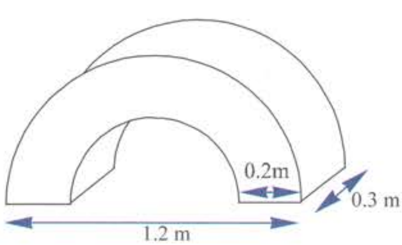
1. Find the areas and perimeters of the following sectors.

	Radius	Angle
a	2.6 cm	$\frac{\pi}{3}$
b	11.5 cm	$\frac{\pi}{4}$
c	44 cm	$\frac{\pi}{4}$
d	6.8 m	$\frac{2\pi}{3}$
e	0.64 cm	$\frac{3\pi}{4}$
f	7.6 cm	$\frac{5\pi}{6}$
g	324 m	$\frac{\pi}{10}$

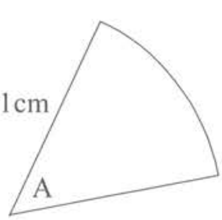
2. A cake has a circumference of 30cm and a uniform height of 7cm. A slice is to be cut from the cake with two straight cuts meeting at the centre. If the slice is to contain $50\,\text{cm}^3$ of cake, find the angle between the two cuts, giving the answer in radians to 2 significant figures and in degrees correct to the nearest degree.

3. The diagram shows a part of a Norman arch. The dimensions are shown in metres.

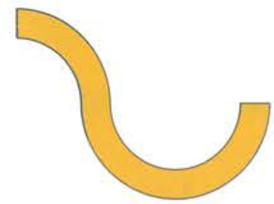
Find the volume of stone in the arch, giving your answer in cubic metres, correct to three significant figures.



4. In the diagram, find the value of the angle A in radians, correct to three significant figures, if the perimeter is equal to 40 cm.

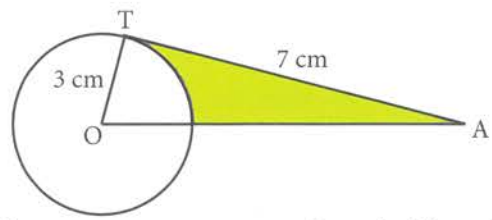


5. The diagram shows a design for a shop sign. The arcs are each one quarter of a complete circle. The radius of the smaller circle is 7 cm and the radius of the larger circle is 9 cm.



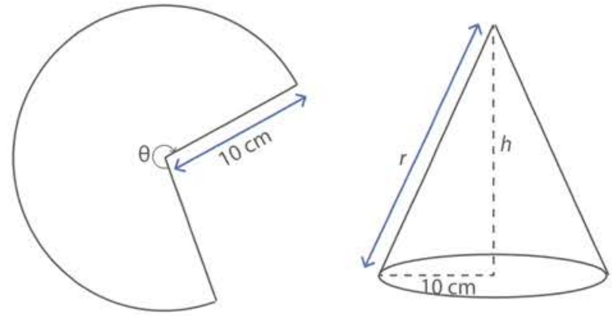
Find the perimeter of the shape, correct to the nearest centimetre.

6. Find the shaded area in the diagram. The dimensions are given in centimetres. O is the centre of the circle and AT is a tangent.



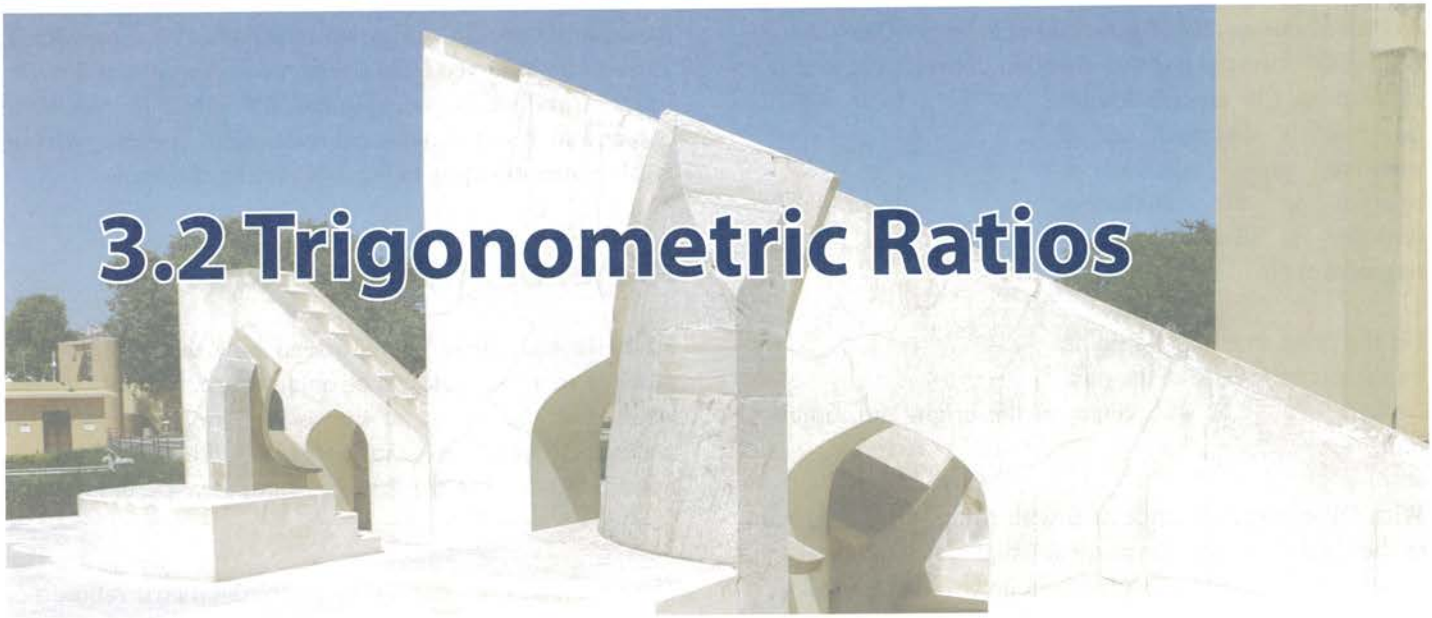
Give your answer correct to three significant figures.

7. The diagrams show a circular sector of radius 10cm and angle θ radians which is formed into a cone of slant height 10 cm. The vertical height h of the cone is equal to the radius r of its base. Find the angle θ radians.



Extra questions

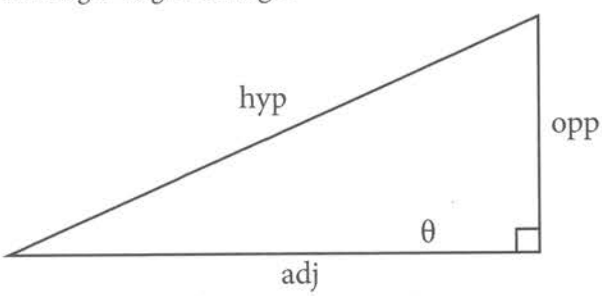
Answers



3.2 Trigonometric Ratios

The Unit Circle

Middle School courses usually include how to find the sine, cosine and tangent of acute angles contained within a right-angled triangle.



$$\sin\theta = \frac{\text{opp}}{\text{hyp}},\ \cos\theta = \frac{\text{adj}}{\text{hyp}},\ \tan\theta = \frac{\text{opp}}{\text{adj}}$$

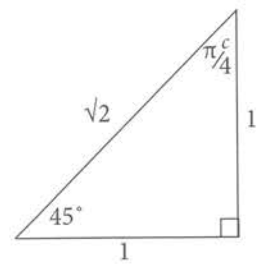
We can extend this to enable us to find the sine and cosine ratio of obtuse angles. To see why this works, or indeed why it would work for an angle of any magnitude, we need to reconsider how angles are measured. To do this we start by making use of the unit circle and introduce some definitions.

From this point on we define the angle θ as a real number that is measured in either degrees or radians. So that, an expression such as $\sin(180^\circ - \theta)$ will imply that θ is measured in degrees as opposed to the expression $\sin(\pi^c - \theta)$ which would imply that θ is measured in radians. In both cases, it should be clear from the context of the question which one it is.

Exact Values

There are several 'exact values' of acute angles that are important. They result from two special triangles. It is probably easier to remember the triangles rater than a table of values.

Isosceles Right Triangle

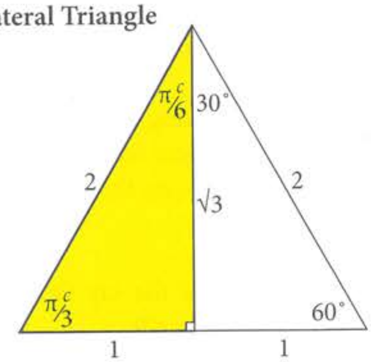


$$\sin 45^\circ = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

$$\cos 45^\circ = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

$$\tan 45^\circ = \tan\frac{\pi}{4} = 1$$

Right Triangle in an Equilateral Triangle



$$\sin 30^\circ = \sin\frac{\pi}{6} = \frac{1}{2}$$

$$\cos 30^\circ = \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}$$

$$\tan 30^\circ = \tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}$$

$$\sin 60^\circ = \sin\frac{\pi}{3} = \frac{1}{2} = \frac{\sqrt{3}}{2}$$

$$\cos 60^\circ = \cos\frac{\pi}{3} = \frac{1}{2}$$

$$\tan 60^\circ = \tan\frac{\pi}{3} = \sqrt{3}$$

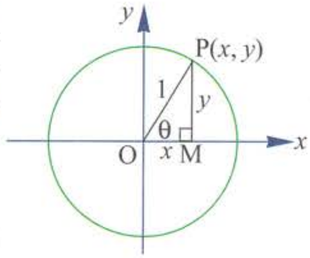
Finally, we have the extreme values of 0° and 90° which are not really representable in a right triangle.

$$\sin 0^\circ = 0, \cos 0^\circ = 1, \tan 0^\circ = 0$$

$$\sin 90^\circ = 1, \cos 90^\circ = 0, \tan 90^\circ \text{ is undefined}$$

Note that $\tan 90^\circ$ is undefined. We will shortly see why this is the case.

By convention, an angle θ is measured in terms of the rotation of a ray OP from the positive direction of the x-axis, so that a rotation in the **anticlockwise** direction is described as a **positive** angle, whereas a rotation in the **clockwise** direction is described as a **negative** angle.



Let the point $P(x, y)$ be a point on the circumference of the unit circle, $x^2 + y^2 = 1$, with centre at the origin and radius 1 unit.

With OP making an angle of θ with the positive direction of the x-axis, we draw the perpendicular from P to meet the x-axis at M. This then provides the following definitions:

$$\sin\theta = \frac{MP}{OP} = \frac{y\text{-coordinate of P}}{OP} = \frac{y}{1} = y$$

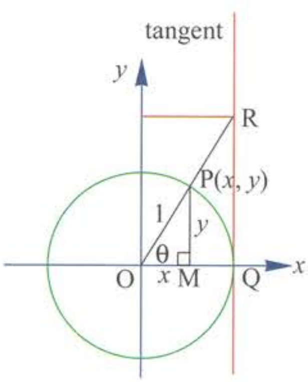
$$\cos\theta = \frac{OM}{OP} = \frac{x\text{-coordinate of P}}{OP} = \frac{x}{1} = x$$

$$\tan\theta = \frac{MP}{OM} = \frac{y\text{-coordinate of P}}{x\text{-coordinate of P}} = \frac{y}{x}$$

Note that this means that the y-coordinate corresponds to the sine of the angle θ, that the x-coordinate corresponds to the cosine of the angle θ and that the tangent,. . . , well, for the tangent, let's revisit the unit circle, but this time we will make an addition to the diagram.

Using the existing unit circle, we draw a tangent at the point where the circle cuts the positive x-axis, Q.

Next, we extend the ray OP to meet the tangent at R.



Using similar triangles, we have that $\frac{PM}{OM} = \frac{RQ}{OQ} = \frac{RQ}{1}$.

That is, $\tan\theta = RQ$ – which means that the value of the tangent of the angle θ corresponds to the y-coordinates of point R cut off on the tangent at Q by the extended ray OP.

Also, it is worth noting that $\tan\theta = \frac{PM}{OM} = \frac{y}{x} = \frac{\sin\theta}{\cos\theta}$ (as long as $\cos\theta \neq 0$).

That is, $\tan\theta = \frac{\sin\theta}{\cos\theta}, \cos\theta \neq 0$.

From our list of exact values, we note that $\tan 90^\circ$ is undefined. This can be observed from the previous diagram. If $\theta = 90^\circ$, P lies on the y-axis, meaning that OP would be parallel to QR, and so, P would never cut the tangent, meaning that no y-value corresponding to R could ever be obtained.

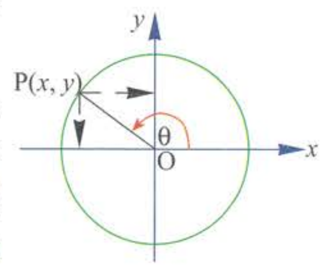
Angle of any Magnitude

From the unit circle we have seen how the trigonometric ratios of an acute angle can be obtained, i.e. for the sine ratio we read off the y-axis, for the cosine ratio, we read off the x-axis and for the tangent ratio we read off the tangent. As the point P is located in the first quadrant, then $x \geq 0$, $y \geq 0$ and $\frac{y}{x} \geq 0, x \neq 0$.

This means that we obtain positive trigonometric ratios.

So, what if P lies in the second quadrant?

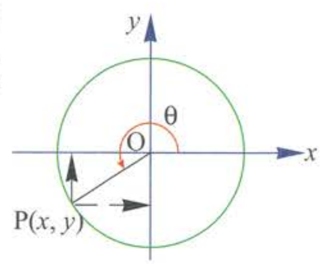
We start by drawing a diagram for such a situation:



From our diagram we see that if P lies in the second quadrant, the y-value is still positive, the x-value is negative and therefore the ratio, y/x is negative.

This means that, $\sin\theta > 0$, $\cos\theta < 0$ and $\tan\theta < 0$.

In a similar way, we can conclude that if $180^\circ < \theta < 270^\circ$, i.e. the point P is in the third quadrant, then,

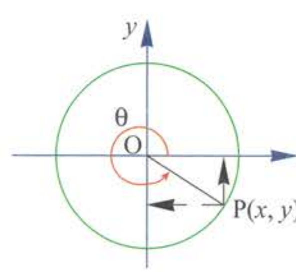


y-value is negative $\sin\theta < 0$

x-value is negative $\cos\theta < 0$

the quotient is positive $\tan\theta > 0$

For the ***fourth quadrant*** we have, $270^\circ < \theta < 360^\circ$, so that:

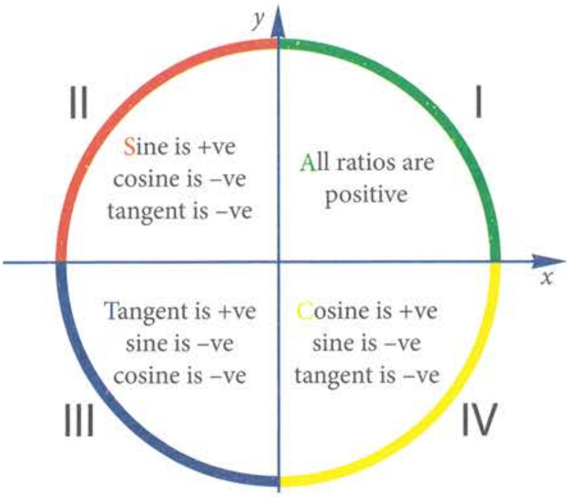


y-value is negative $\sin\theta < 0$

x-value is positive $\cos\theta > 0$

the quotient is negative $\tan\theta < 0$

We now know that, depending on which quadrant an angle lies in, the sign of the trigonometric ratio will be either positive or negative. In fact, we can summarize this as follows:

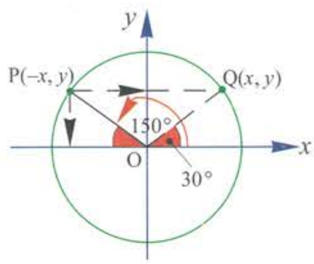


Mnemonic:

All Stations To City

However, knowing the sign of a trigonometric ratio reflects only half the information. We still need to determine the numerical value. We start by considering a few examples.

Consider the value of $\sin 150^\circ$. Using the unit circle we have:



By symmetry we see that the y-coordinate of Q and the y-coordinate of P are the same and so, $\sin 150^\circ = \sin 30^\circ$.

Therefore, $\sin 150^\circ = ½$

Note that $150^\circ = 150 \times \frac{\pi}{180} = \frac{5\pi}{6}$ and $30^\circ = \frac{\pi}{6}$,

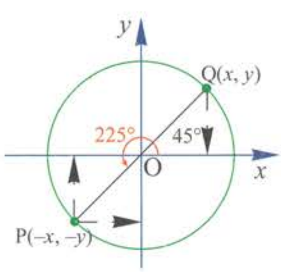
so that in radian form we have, $\sin\frac{5\pi}{6} = \sin\frac{\pi}{6} = \frac{1}{2}$.

In other words, we were able to express the sine of an angle in the second quadrant in terms of the sine of an angle in the first quadrant. In particular, we have that

If $0^\circ < \theta < 90^\circ$, $\sin(180^\circ - \theta) = \sin\theta$

If $0^c < \theta < \frac{\pi^c}{2}$, $\sin(\pi^c - \theta) = \sin\theta$

Next, consider the value of $\cos 225^\circ$. Using the unit circle we have:



By symmetry we see that the x-coordinate of P has the same magnitude as the x-coordinate of Q but is of the opposite sign.

So, we have that $\cos 225^\circ = -\cos 45^\circ$.

Therefore, $\cos 225^\circ = -\frac{1}{\sqrt{2}}$.

Similarly, as $225^\circ = \frac{3\pi^c}{4}$ and $45^\circ = \frac{\pi^c}{4}$,

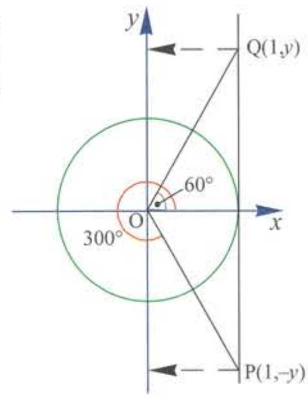
$\cos\frac{3\pi^c}{4} = -\cos\frac{\pi^c}{4} = -\frac{1}{\sqrt{2}}$.

In other words, we were able to express the cosine of an angle in the third quadrant in terms of the cosine of an angle in the first quadrant. In particular, we have that:

If $0^\circ < \theta < 90^\circ$, $\cos(180^\circ + \theta) = -\cos\theta$

If $0^c < \theta < \frac{\pi^c}{2}$, $\cos(\pi^c + \theta) = -\cos\theta$

As a last example we consider the value of $\tan 300^\circ$. This time we need to add a tangent to the unit circle cutting the positive x-axis:



By symmetry we see that the y-coordinate of P has the same magnitude as the y-coordinate of Q but is of the opposite sign.

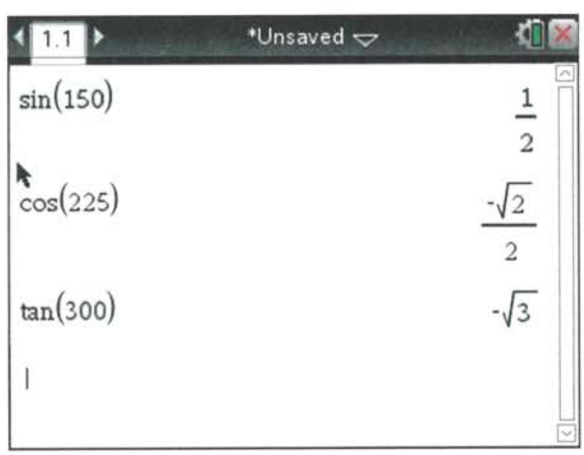
So: $\tan 300^\circ = -\tan 60^\circ$.

Therefore, $\tan 300^\circ = -\sqrt{3}$.

As, $300^\circ = \frac{5\pi^c}{3}$, $60^\circ = \frac{\pi^c}{3}$

$\tan\frac{5\pi^c}{3} = -\tan\frac{\pi^c}{3} = -\sqrt{3}$.

In other words, we were able to express the tangent of an angle in the fourth quadrant in terms of the tangent of an angle in the first quadrant. In particular, we have that:

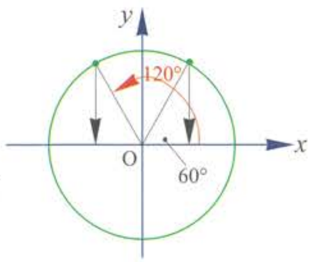


Example 3.2.1

Find the exact values of:

a	$\cos 120°$	b	$\sin 210°$
c	$\cos\frac{7\pi}{4}$	d	$\tan\frac{5\pi}{4}$

a



Step 1: Start by drawing the unit circle:

Step 2: Trace out an angle of 120°

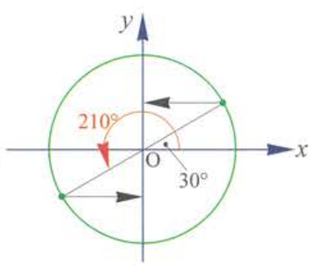
Step 3: Trace out the **reference angle** in the first quadrant. In this case it is 60°.

Step 4: Use the symmetry between the reference angle and the given angle.

Step 5: State relationship and give answer:

$$\cos 120° = -\cos 60° = -\frac{1}{2}$$

b



Step 1: Start by drawing the unit circle:

Step 2: Trace out an angle of 210°

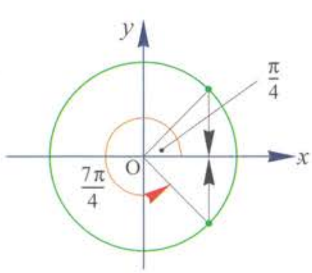
Step 3: Trace out the **reference angle** in the first quadrant. In this case it is 30°.

Step 4: Use the symmetry between the reference angle and the given angle.

Step 5: State relationship and give answer:

$$\sin 210° = -\sin 30° = -\frac{1}{2}$$

c



Step 1: Start by drawing the unit circle:

Step 2: Trace out an angle of $\frac{7\pi}{4}$ $(= 315°)$

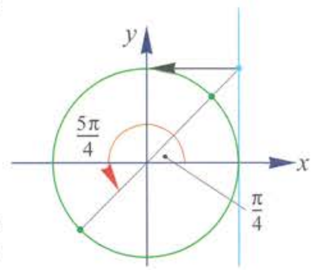
Step 3: Trace out the **reference angle** in the first quadrant. In this case it is $\frac{\pi}{4}$.

Step 4: Use the symmetry between the reference angle and the given angle.

Step 5: State relationship and give answer:

$$\cos\frac{7\pi}{4} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

d



Step 1: Start by drawing the unit circle:

Step 2: Trace out an angle of $\frac{5\pi}{4}$ $(= 225°)$

Step 3: Trace out the **reference angle** in the first quadrant. In this case it is $\frac{\pi}{4}$.

Step 4: Use the symmetry between the reference angle and the given angle.

Step 5: State relationship and give answer:

$$\tan\frac{5\pi}{4} = \tan\frac{\pi}{4} = 1$$

The results we have obtained, that is, expressing trigonometric ratios of any angle in terms of trigonometric ratios of acute angles in the first quadrant (i.e. reference angles) are known as **trigonometric reduction formulae**. There are too many formulae to commit to memory, and so it is advisable to draw a unit circle and then use symmetry properties as was done in Example 3.2.1. We list a number of these formulae in the table below, where $0 < \theta < \frac{\pi}{2}$ $(= 90°)$.

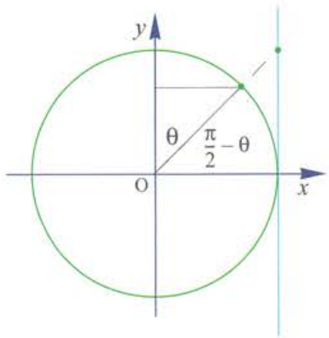
Note: From this point on, angles without the degree symbol or radian symbol will mean an angle measured in radian mode.

Q	θ in degrees	θ in radians
2	$\sin(180° - \theta) = \sin\theta$ $\cos(180° - \theta) = -\cos\theta$ $\tan(180° - \theta) = -\tan\theta$	$\sin(\pi - \theta) = \sin\theta$ $\cos(\pi - \theta) = -\cos\theta$ $\tan(\pi - \theta) = -\tan\theta$
3	$\sin(180° + \theta) = -\sin\theta$ $\cos(180° + \theta) = -\cos\theta$ $\tan(180° + \theta) = \tan\theta$	$\sin(\pi + \theta) = -\sin\theta$ $\cos(\pi + \theta) = -\cos\theta$ $\tan(\pi + \theta) = \tan\theta$
4	$\sin(360° - \theta) = -\sin\theta$ $\cos(360° - \theta) = \cos\theta$ $\tan(360° - \theta) = -\tan\theta$	$\sin(2\pi - \theta) = -\sin\theta$ $\cos(2\pi - \theta) = \cos\theta$ $\tan(2\pi - \theta) = -\tan\theta$

There is another set of results that is suggested by symmetry through the fourth quadrant:

Quadrant	θ in degrees or θ in radians
4	$\sin(-\theta) = -\sin\theta$ $\cos(-\theta) = \cos\theta$ $\tan(-\theta) = -\tan\theta$

The symmetry of the unit circle leads to a number of further relationships



Quadrant	θ in radians or θ in degrees
1	$\sin\left(\frac{\pi}{2} - \theta\right) = \cos\theta \quad \sin(90° - \theta) = \cos\theta$ $\cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta \quad \cos(90° - \theta) = \sin\theta$ $\tan\left(\frac{\pi}{2} - \theta\right) = \cot\theta \quad \tan(90° - \theta) = \cot\theta$

Rather than trying to memorise these results, reference to the unit circle and its symmetries will reveal them when they are needed.

Reciprocals

Note the introduction of a new trigonometric ratio, $\cot\theta$. This is one of a set of three other trigonometric ratios known as the **reciprocal trigonometric ratios**, namely **cosecant**, **secant** and **cotangent** ratios. These are defined as:

$$\text{cosecant ratio: } \operatorname{cosec}\theta = \frac{1}{\sin\theta}, \sin\theta \neq 0$$

$$\text{secant ratio: } \sec\theta = \frac{1}{\cos\theta}, \cos\theta \neq 0$$

$$\text{cotangent ratio: } \cot\theta = \frac{1}{\tan\theta}, \tan\theta \neq 0$$

Note then, that $\cot\theta = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta}$, $\sin\theta \neq 0$ and cosec is often written 'csc'.

Example 3.2.2

Find the exact values of:

a $\sec 45°$ b $\operatorname{cosec} 150°$

c $\cot\frac{11\pi}{6}$ d $\sec 0$

a $\sec 45° = \frac{1}{\cos 45°} = \frac{1}{\left(\frac{1}{\sqrt{2}}\right)} = \sqrt{2}$

b $\operatorname{cosec} 150° = \frac{1}{\sin 150°} = \frac{1}{\sin 30°} = \frac{1}{\left(\frac{1}{2}\right)} = 2$

c $\cot\frac{11\pi}{6} = \frac{1}{\tan\left(\frac{11\pi}{6}\right)} = \frac{1}{\tan\left(-\frac{\pi}{6}\right)} = \frac{1}{-\tan\frac{\pi}{6}}$

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

d $\sec 0 = \frac{1}{\cos 0} = \frac{1}{1} = 1$

Exercise 3.2.1

1. Convert the following angles to degrees.

   a $\frac{2\pi}{3}$ b $\frac{3\pi}{5}$

   c $\frac{12\pi}{10}$ d $\frac{5\pi}{18}$

2. Convert the following angles to radians.

   a 180° b 270°

   c 140° d 320°

3. Find the exact value of:

   a $\sin 120°$ b $\cos 120°$

   c $\tan 120°$ d $\sec 120°$

   e $\sin 210°$ f $\cos 210°$

   g $\tan 210°$ h $\cot 210°$

   i $\sin 225°$ j $\cos 225°$

k	$\tan 225°$	l	$\operatorname{cosec} 225°$
m	$\sin 315°$	n	$\cos 315°$
o	$\tan 315°$	p	$\sec 315°$
q	$\sin 360°$	r	$\cos 360°$
s	$\tan 360°$	t	$\operatorname{cosec} 360°$

4. Find the exact value of:

a	$\sin\pi$	b	$\cos\pi$
c	$\tan\pi$	d	$\sec\pi$
e	$\sin\frac{3\pi}{4}$	f	$\cos\frac{3\pi}{4}$
g	$\tan\frac{3\pi}{4}$	h	$\operatorname{cosec}\frac{3\pi}{4}$
i	$\sin\frac{7\pi}{6}$	j	$\cos\frac{7\pi}{6}$
k	$\tan\frac{7\pi}{6}$	l	$\cot\frac{7\pi}{6}$
m	$\sin\frac{5\pi}{3}$	n	$\cos\frac{5\pi}{3}$

5. Find the exact value of:

a	$\sin(-210°)$	b	$\cos(-30°)$
c	$\tan(-135°)$	d	$\cos(-420°)$
e	$\cot(-60°)$	f	$\sin(-150°)$
g	$\sec(-135°)$	h	$\operatorname{cosec}(-120°)$

6. Find the exact value of:

a	$\sin\left(-\frac{\pi}{6}\right)$	b	$\cos\left(-\frac{3\pi}{4}\right)$
c	$\tan\left(-\frac{2\pi}{3}\right)$	d	$\sec\left(-\frac{4\pi}{3}\right)$
e	$\cot\left(-\frac{3\pi}{4}\right)$	f	$\sin\left(-\frac{7\pi}{6}\right)$
g	$\cot\left(-\frac{\pi}{3}\right)$	h	$\cos\left(-\frac{7\pi}{6}\right)$
i	$\operatorname{cosec}\left(-\frac{2\pi}{3}\right)$	j	$\tan\left(-\frac{11\pi}{6}\right)$
k	$\sec\left(-\frac{13\pi}{6}\right)$	l	$\sin\left(-\frac{7\pi}{3}\right)$

Extra questions

The Pythagorean Identity

We have seen a number of important relationships between trigonometric ratios. Relationships that are true for all values of θ are known as **identities**. To signal an identity (as opposed to an equation) the **equivalence** symbol is used, i.e. '≡'.

For example, we can write $(x+1)^2 \equiv x^2+2x+1$ as this statement is true for all values of x. However, we would have to write $(x+1)^2 = x^2+1$, as this relationship is only true for some values of x (which need to be determined).

One trigonometric identity is based on the unit circle.

Consider the point $P(x, y)$ on the unit circle,
$x^2 + y^2 = 1$ – (1)

From the previous section, we know that

$$x = \cos\theta \quad – (2)$$

$$y = \sin\theta \quad – (3)$$

Substituting (2) and (3) into (1) we have: $(\cos\theta)^2 + (\sin\theta)^2 = 1$
or

$$\boxed{\sin^2\theta+\cos^2\theta=1} \quad – (4)$$

This is known as the fundamental trigonometric identity. Note that we have not used the identity symbol, i.e. we have not written $\sin^2\theta+\cos^2\theta\equiv 1$. This is because more often than not, it will be 'obvious' from the setting as to whether a relationship is an identity or an equation. And so, there is a tendency to forgo the formal use of the identity statement and restrict ourselves to the equality statement.

By rearranging the identity we have that $\sin^2\theta=1-\cos^2\theta$ and $\cos^2\theta\equiv 1-\sin^2\theta$. Similarly we obtain the following two new identities:

Divide both sides of (4) by $\cos^2\theta$:

$$\frac{\sin^2\theta+\cos^2\theta}{\cos^2\theta} = \frac{1}{\cos^2\theta} \Leftrightarrow \frac{\sin^2\theta}{\cos^2\theta}+\frac{\cos^2\theta}{\cos^2\theta} = \frac{1}{\cos^2\theta}$$

$$\Leftrightarrow \tan^2\theta + 1 = \sec^2\theta \quad – (5)$$

Divide both sides of (4) by $\sin^2\theta$:

$$\frac{\sin^2\theta+\cos^2\theta}{\sin^2\theta} = \frac{1}{\sin^2\theta} \Leftrightarrow \frac{\sin^2\theta}{\sin^2\theta}+\frac{\cos^2\theta}{\sin^2\theta} = \frac{1}{\sin^2\theta}$$

$$\Leftrightarrow 1 + \cot^2\theta = \operatorname{cosec}^2\theta \quad – (6)$$

Example 3.2.3

If $\cos\theta = -\frac{3}{5}$, where $\pi \le \theta \le \frac{3\pi}{2}$, find:

a $\sin\theta$ b $\tan\theta$

Although we solved problems like this in section 10.1 by making use of a right-angled triangle, we now solve this question by making use of the trigonometric identities we have just developed.

a From $\sin^2\theta + \cos^2\theta = 1$ we have

$$\sin^2\theta + \left(-\frac{3}{5}\right)^2 = 1 \Leftrightarrow \sin^2\theta + \frac{9}{25} = 1$$
$$\Leftrightarrow \sin^2\theta = \frac{16}{25}$$
$$\therefore \sin\theta = \pm\frac{4}{5}$$

Now, as $\pi \le \theta \le \frac{3\pi}{2}$,

this means the angle is in the third quadrant, where the sine value is negative.

Therefore, we have that $\sin\theta = -\frac{4}{5}$.

b Using the identity $\tan\theta = \frac{\sin\theta}{\cos\theta}$,

we have $\tan\theta = \frac{(-4/5)}{(-3/5)} = \frac{4}{3}$.

Example 3.2.4

If $\tan\theta = \frac{5}{12}$, where $\pi \le \theta \le \frac{3\pi}{2}$, find:

a $\cos\theta$ b $\operatorname{cosec}\theta$

a From the identity $\tan^2\theta + 1 = \sec^2\theta$ we have:

$$\left(\frac{5}{12}\right)^2 + 1 = \sec^2\theta \Leftrightarrow \sec^2\theta = \frac{25}{144} + 1$$
$$\therefore \sec^2\theta = \frac{169}{144}$$
$$\therefore \sec\theta = \pm\frac{13}{12}$$

Therefore, as $\cos\theta = \frac{1}{\sec\theta} \Rightarrow \cos\theta = \pm\frac{12}{13}$.

However, $\pi \le \theta \le \frac{3\pi}{2}$, meaning that θ is in the third quadrant.

And so, the cosine is negative. That is, $\cos\theta = -\frac{12}{13}$.

Now, $\operatorname{cosec}\theta = \frac{1}{\sin\theta}$, but,

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \Leftrightarrow \sin\theta = \tan\theta\cos\theta \therefore \sin\theta = \frac{5}{12} \times -\frac{12}{13}$$
$$= -\frac{5}{13}$$

Therefore, $\operatorname{cosec}\theta = \frac{1}{(-5/13)} = -\frac{13}{5}$.

Example 3.2.5

Simplify the following expressions.

a $\cos\theta + \tan\theta\sin\theta$ b $\frac{\cos\theta}{1+\sin\theta} - \frac{1-\sin\theta}{\cos\theta}$

a
$$\cos\theta + \tan\theta\sin\theta = \cos\theta + \frac{\sin\theta}{\cos\theta}\sin\theta$$
$$= \cos\theta + \frac{\sin^2\theta}{\cos\theta}$$
$$= \frac{\cos^2\theta + \sin^2\theta}{\cos\theta}$$
$$= \frac{1}{\cos\theta}$$
$$= \sec\theta$$

b
$$\frac{\cos\theta}{1+\sin\theta} - \frac{1-\sin\theta}{\cos\theta}$$
$$= \frac{\cos^2\theta}{(1+\sin\theta)\cos\theta} - \frac{(1-\sin\theta)(1+\sin\theta)}{(1+\sin\theta)\cos\theta}$$
$$= \frac{\cos^2\theta}{(1+\sin\theta)\cos\theta} - \frac{1-\sin^2\theta}{(1+\sin\theta)\cos\theta}$$
$$= \frac{\cos^2\theta - 1 + \sin^2\theta}{(1+\sin\theta)\cos\theta}$$
$$= \frac{(\cos^2\theta + \sin^2\theta) - 1}{(1+\sin\theta)\cos\theta}$$
$$= \frac{1-1}{(1+\sin\theta)\cos\theta}$$
$$= 0$$

Example 3.2.6

Show that $\frac{1-2\cos^2\theta}{\sin\theta\cos\theta} = \tan\theta - \cot\theta$.

$$\begin{aligned}\text{R.H.S} &= \tan\theta - \cot\theta \\ &= \frac{\sin\theta}{\cos\theta} - \frac{\cos\theta}{\sin\theta} \\ &= \frac{\sin^2\theta - \cos^2\theta}{\sin\theta\cos\theta} \\ &= \frac{(1-\cos^2\theta) - \cos^2\theta}{\sin\theta\cos\theta} \\ &= \frac{1 - 2\cos^2\theta}{\sin\theta\cos\theta} \\ &= \text{L.H.S}\end{aligned}$$

Exercise 3.2.2

1. Prove the identity.

a $\sin\theta + \cot\theta\cos\theta = \operatorname{cosec}\theta$

b $\dfrac{\sin\theta}{1+\cos\theta} + \dfrac{1+\cos\theta}{\sin\theta} = 2\operatorname{cosec}\theta$

c $\dfrac{\sin^2\theta}{1-\cos\theta} = 1 + \cos\theta$

d $3\cos^2 x - 2 = 1 - 3\sin^2 x$

e $\tan^2 x\cos^2 x + \cot^2 x\sin^2 x = 1$

f $\sec\theta - \sec\theta\sin^2\theta = \cos\theta$

g $\sin^2\theta(1+\cot^2\theta) - 1 = 0$

h $\dfrac{1}{1-\sin\phi} + \dfrac{1}{1+\sin\phi} = 2\sec^2\phi$

i $\dfrac{\cos\theta}{1+\sin\theta} + \tan\theta = \sec\theta$

j $\dfrac{1-\sin\theta}{\cos\theta} = \dfrac{\cos\theta}{1+\sin\theta}$

k $\dfrac{1}{\sec x + \tan x} = \sec x - \tan x$

l $\sin x + \dfrac{\cos^2 x}{1+\sin x} = 1$

m $\dfrac{\sec\phi + \operatorname{cosec}\phi}{\tan\phi + \cot\phi} = \sin\phi + \cos\phi$

n $\dfrac{\sin x + 1}{\cos x} = \dfrac{\sin x - \cos x + 1}{\sin x + \cos x - 1}$

o $\tan x + \sec x = \dfrac{\tan x + \sec x - 1}{\tan x - \sec x + 1}$

2. Prove the following.

a $(\sin x + \cos x)^2 + (\sin x - \cos x)^2 = 2$

b $\sec^2\theta\operatorname{cosec}^2\theta = \sec^2\theta + \operatorname{cosec}^2\theta$

c $\sin^4 x - \cos^4 x = (\sin x + \cos x)(\sin x - \cos x)$

d $\sec^4 x - \sec^2 x = \tan^4 x + \tan^2 x$

e $\dfrac{\sin^3 x + \cos^3 x}{\sin x + \cos x} = 1 - \sin x\cos x$

f $(\cot x - \operatorname{cosec} x)^2 = \dfrac{\sec x - 1}{\sec x + 1}$

g $(2b\sin x\cos x)^2 + b^2(\cos^2 x - \sin^2 x)^2 = b^2$

3. Eliminate θ from each of the following pairs.

a $x = k\sin\theta, y = k\cos\theta$

b $x = b\sin\theta, y = a\cos\theta$

c $x = 1 + \sin\theta, y = 2 - \cos\theta$

d $x = 1 - b\sin\theta, y = 2 + a\cos\theta$

e $x = \sin\theta + 2\cos\theta, y = \sin\theta - 2\cos\theta$

4. a If $\tan\theta = \frac{3}{4}, \pi \le \theta \le \frac{3\pi}{2}$,

find: i $\cos\theta$ ii $\operatorname{cosec}\theta$

b If $\sin\theta = -\frac{3}{4}, \frac{3\pi}{2} \le \theta \le 2\pi$,

find: i $\sec\theta$ ii $\cot\theta$

5. Solve the following, where $0 \le \theta \le 2\pi$:

a $4\sin\theta = 3\operatorname{cosec}\theta$

b $2\cos^2\theta + \sin\theta - 1 = 0$

c $2 - \sin\theta = 2\cos^2\theta$

d $2\sin^2\theta = 2 + 3\cos\theta$

Extra questions

Answers

3.3 Identities

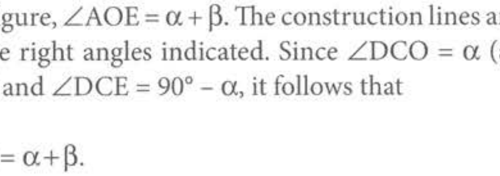
Compound Angle Identities

As we have seen in the previous section, there are numerous trigonometric identities. However, they were all derived from the fundamental identities. In this section we develop some more fundamental identities (which will lead us to more identities). These fundamental identities are known as compound angle identities. That is, they are identities that involve the sine, cosine and tangent of the sum and difference of two angles.

We start with the sine of the sum of two angles, $\sin(\alpha+\beta)$:

The Diploma Course does not expect students to prove this result. It is included for the sake of completeness.

A commonly given proof of these identities is again only valid for acute angles:



In the figure, $\angle AOE = \alpha+\beta$. The construction lines are drawn with the right angles indicated. Since $\angle DCO = \alpha$ (alternate angles) and $\angle DCE = 90° - \alpha$, it follows that

$\angle AOE = \alpha+\beta$.

Therefore, we have, $\sin(\alpha+\beta) = \sin AOE = \dfrac{AE}{OE}$

$$
\begin{aligned}
&= \frac{AD+DE}{OE}\\
&= \frac{AD}{OE}+\frac{DE}{OE}\\
&= \frac{BC}{OE}+\frac{DE}{OE}\\
&= \frac{BC}{OC}\times\frac{OC}{OE}+\frac{DE}{EC}\times\frac{EC}{OE}\\
&= \sin\alpha\times\cos\beta+\cos\alpha\times\sin\beta
\end{aligned}
$$

It is now possible to prove the difference formula, replacing β by $-\beta$ we have:

$$
\begin{aligned}
\sin(\alpha-\beta) &= \sin(\alpha+(-\beta))\\
&= \sin\alpha\cos(-\beta)+\cos\alpha\sin(-\beta)\\
&= \sin\alpha\cos\beta-\cos\alpha\sin\beta
\end{aligned}
$$

($\cos(-\beta) = \cos\beta$ and $\sin(-\beta) = -\sin\beta$)

And so we have the addition and difference identities for sine:

$$
\begin{aligned}
\sin(\alpha+\beta) &= \sin\alpha\cos\beta+\cos\alpha\sin\beta\\
\sin(\alpha-\beta) &= \sin\alpha\cos\beta-\cos\alpha\sin\beta
\end{aligned}
$$

A similar identity can be derived for the cosine function (using the same diagram):

$$
\begin{aligned}
\cos(\alpha+\beta) = \frac{OA}{OE} &= \frac{OB-AB}{OE}\\
&= \frac{OB}{OE}-\frac{AB}{OE}\\
&= \frac{OB}{OE}-\frac{CD}{OE}\\
&= \frac{OB}{OC}\times\frac{OC}{OE}-\frac{CD}{EC}\times\frac{EC}{OE}\\
&= \cos\alpha\cos\beta-\sin\alpha\sin\beta
\end{aligned}
$$

Similarly, from this and replacing β by $-\beta$ we have that $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$.

And so we have the addition and difference identities for cosine:

$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta$$
$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta$$

Also, the tangent addition identity can be proved as follows:

Using $\frac{\sin\theta}{\cos\theta} = \tan\theta$

$$\begin{aligned}\tan(\alpha+\beta) &= \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)}\\ &= \frac{\sin\alpha\cos\beta+\cos\alpha\sin\beta}{\cos\alpha\cos\beta-\sin\alpha\sin\beta}\\ &= \frac{\dfrac{\sin\alpha\cos\beta+\cos\alpha\sin\beta}{\cos\alpha\cos\beta}}{\dfrac{\cos\alpha\cos\beta-\sin\alpha\sin\beta}{\cos\alpha\cos\beta}}\\ &= \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}\end{aligned}$$

Again, if we replace β by $-\beta$ we have:

$$\tan(\alpha-\beta) = \frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta}.$$

And so we have the addition and difference identities for tangent:

$$\tan(\alpha+\beta)=\frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}$$
$$\tan(\alpha-\beta)=\frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta}$$

As a special case of the compound identities we have obtained so far, we have a set of identities known as the **double-angle identities**.

Using the substitution $\theta = \alpha = \beta$ we obtain the identities:

$$\sin 2\theta=2\sin\theta\cos\theta$$
$$\cos 2\theta=\cos^2\theta-\sin^2\theta$$

i.e. substituting $\theta = \alpha = \beta$ into:

$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$ we obtain

$$\begin{aligned}\sin(\theta+\theta) &= \sin\theta\cos\theta + \cos\theta\sin\theta\\ \therefore \sin 2\theta &= 2\sin\theta\cos\theta\end{aligned}$$

Similarly, substituting $\theta = \alpha = \beta$ into:

$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ we obtain

$$\begin{aligned}\cos(\theta+\theta) &= \cos\theta\cos\theta - \sin\theta\sin\theta\\ \therefore \cos 2\theta &= \cos^2\theta - \sin^2\theta\end{aligned}$$

The second of these can be further developed to give:

$$\cos 2\theta = \cos^2\theta - \sin^2\theta = \cos^2\theta - (1-\cos^2\theta) = 2\cos^2\theta - 1$$

and

$$\cos 2\theta = \cos^2\theta - \sin^2\theta = (1-\sin^2\theta) - \sin^2\theta = 1 - 2\sin^2\theta$$

Finally, we have a double-angle identity for the tangent:

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta}$$

Summary of double-angle identities

$$\begin{aligned}\sin 2\theta &= 2\sin\theta\cos\theta\\ \cos 2\theta &= \cos^2\theta-\sin^2\theta\\ &= 2\cos^2\theta-1\\ &= 1-2\sin^2\theta\\ \tan 2\theta &= \frac{2\tan\theta}{1-\tan^2\theta}\end{aligned}$$

We have seen how trigonometric identities can be used to solve equations, simplify expressions and to prove further identities. We now illustrate this using the new set of identities.

Example 3.3.1

Simplify the expression: $\frac{\sin 3\alpha}{\sin\alpha} - \frac{\cos 3\alpha}{\cos\alpha}$.

$$\begin{aligned}\frac{\sin 3\alpha}{\sin\alpha} - \frac{\cos 3\alpha}{\cos\alpha} &= \frac{\sin 3\alpha\cos\alpha - \cos 3\alpha\sin\alpha}{\sin\alpha\cos\alpha}\\ &= \frac{\sin(3\alpha-\alpha)}{\sin\alpha\cos\alpha}\\ &= \frac{\sin 2\alpha}{\frac{1}{2}\sin 2\alpha}\\ &= 2\end{aligned}$$

Example 3.3.2

Prove the identity $\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha$.

$$\cos 3\alpha = \cos(2\alpha + \alpha)$$

$$= \cos 2\alpha \cos\alpha - \sin 2\alpha \sin\alpha$$

$$= (2\cos^2\alpha - 1)\cos\alpha - 2\sin\alpha\cos\alpha\sin\alpha$$

$$= 2\cos^3\alpha - \cos\alpha - 2\sin^2\alpha\cos\alpha$$

$$= 2\cos^3\alpha - \cos\alpha - 2(1 - \cos^2\alpha)\cos\alpha$$

$$= 2\cos^3\alpha - \cos\alpha - 2\cos\alpha + 2\cos^3\alpha$$

$$= 4\cos^3\alpha - 3\cos\alpha$$

Example 3.3.3

Using a compound identity, show that $\cos\left(\frac{3\pi}{2} - \theta\right) = -\sin\theta$

L.H.S $= \cos\left(\frac{3\pi}{2} - \theta\right)$

$$= \cos\left(\frac{3\pi}{2}\right)\cos\theta + \sin\left(\frac{3\pi}{2}\right)\sin\theta$$

$$= 0 \times \cos\theta + (-1) \times \sin\theta$$

$$= -\sin\theta$$

$= \text{R.H.S}$

Example 3.3.4

Find the exact value of:

a $\cos 15°$ b $\tan\frac{5\pi}{12}$

a

$$\cos 15° = \cos(45° - 30°) = \cos 45°\cos 30° + \sin 45°\sin 30°$$

$$= \frac{1}{\sqrt{2}} \times \frac{\sqrt{3}}{2} + \frac{1}{\sqrt{2}} \times \frac{1}{2}$$

$$= \frac{\sqrt{3}+1}{2\sqrt{2}}$$

b

$$\tan\frac{5\pi}{12} = \tan\left(\frac{3\pi}{12} + \frac{2\pi}{12}\right) = \frac{\tan\frac{3\pi}{12} + \tan\frac{2\pi}{12}}{1 - \tan\frac{3\pi}{12}\tan\frac{2\pi}{12}} = \frac{\tan\frac{\pi}{4} + \tan\frac{\pi}{6}}{1 - \tan\frac{\pi}{4}\tan\frac{\pi}{6}}$$

$$= \frac{1 + \frac{1}{\sqrt{3}}}{1 - 1 \times \frac{1}{\sqrt{3}}} = \frac{\sqrt{3}+1}{\sqrt{3}-1}$$

Example 3.3.5

Prove that $\frac{\sin 2\phi + \sin\phi}{\cos 2\phi + \cos\phi + 1} = \tan\phi$.

L.H.S $= \frac{\sin 2\phi + \sin\phi}{\cos 2\phi + \cos\phi + 1} = \frac{2\sin\phi\cos\phi + \sin\phi}{2\cos^2\phi - 1 + \cos\phi + 1}$

$$= \frac{\sin\phi(2\cos\phi + 1)}{\cos\phi(2\cos\phi + 1)}$$

$$= \frac{\sin\phi}{\cos\phi}$$

$$= \tan\phi$$

$= \text{R.H.S}$

Exercise 3.3.1

1. Expand the following.

a	$\sin(\alpha + \phi)$	b	$\cos(3\alpha + 2\beta)$
c	$\sin(2x - y)$	d	$\cos(\phi - 2\alpha)$
e	$\tan(2\theta - \alpha)$	f	$\tan(\phi - 3\omega)$

2. Simplify the following.

a $\sin 2\alpha\cos 3\beta - \sin 3\beta\cos 2\alpha$

b $\cos 2\alpha\cos 5\beta - \sin 2\alpha\sin 5\beta$

c $\sin x\cos 2y + \sin 2y\cos x$

d $\cos x\cos 3y + \sin x\sin 3y$

e $\frac{\tan 2\alpha - \tan\beta}{1 + \tan 2\alpha\tan\beta}$

f $\frac{\tan(x-y) + \tan y}{1 - \tan(x-y)\tan y}$

g $\frac{1 - \tan\phi}{1 + \tan\phi}$

h $\frac{1}{\sqrt{2}}\sin(\alpha + \beta) + \frac{1}{\sqrt{2}}\cos(\alpha + \beta)$

3. Given that $\sin\theta = \frac{4}{5}, 0 \le \theta \le \frac{\pi}{2}$ and $\cos\phi = -\frac{5}{13}, \pi \le \phi \le \frac{3\pi}{2}$, evaluate:

a $\sin(\theta+\phi)$

b $\cos(\theta+\phi)$

c $\tan(\theta-\phi)$

4. Given that $\sin\theta = -\frac{3}{5}, \pi \le \theta \le \frac{3\pi}{2}$ and $\cos\phi = -\frac{12}{13}, \pi \le \phi \le \frac{3\pi}{2}$, evaluate:

a $\sin(\theta-\phi)$

b $\cos(\theta-\phi)$

c $\tan(\theta+\phi)$

5. Given that $\sin\theta = -\frac{5}{6}, \frac{3\pi}{2} \le \theta \le 2\pi$, evaluate:

a $\sin 2\theta$

b $\cos 2\theta$

c $\tan 2\theta$

d $\sin 4\theta$

6. Given that $\tan x = -3, \frac{\pi}{2} \le x \le \pi$, evaluate:

a $\sin 2x$

b $\cos 2x$

c $\tan 2x$

d $\tan 4x$

7. Find the exact value of:

a $\sin\frac{5\pi}{12}$ b $\sin 105°$

c $\cos\frac{11\pi}{12}$ d $\tan 165°$

8. Given that $\tan x = \frac{a}{b}, \pi \le x \le \frac{3\pi}{2}$, evaluate

a $\sin 2x$ b $\mathrm{cosec}\,2x$

c $\cos 4x$ d $\tan 2x$

9. Prove the following identities.

a $\cot x - \cot 2x = \mathrm{cosec}\,2x$

b $\sin(x+y)\sin(x-y) = \sin^2 x - \sin^2 y$

c $\sec^2 x = 1 + \tan^2 x$

d $\tan(\theta+\phi) + \tan(\theta-\phi) = \dfrac{2\sin 2\theta}{\cos 2\theta + \cos 2\phi}$

e $\cos^4\alpha - \sin^4\alpha = 1 - 2\sin^2\alpha$

f $\dfrac{1}{\sin y\cos y} - \dfrac{\cos y}{\sin y} = \tan y$

g $\dfrac{1+\cos 2y}{\sin 2y} = \dfrac{\sin 2y}{1-\cos 2y}$

h $\csc\left(\theta + \frac{\pi}{2}\right) = \sec\theta$

i $\cos 3x = \cos x - 4\sin^2 x\cos x$

j $\dfrac{1+\sin 2\theta}{\cos 2\theta} = \dfrac{\cos\theta + \sin\theta}{\cos\theta - \sin\theta}$

k $(\cot x + \csc x)^2 = \dfrac{1+\cos x}{1-\cos x}$

l $\sin 3\alpha = 3\sin\alpha - 4\sin^3\alpha$

m $\cos 2x = \dfrac{1-\tan^2 x}{1+\tan^2 x}$

n $2\cot\theta\sin^2\theta = \sin 2\theta$

o $\tan\left(\frac{\phi}{2}\right) = \csc\phi - \cot\phi$

p $2\csc x = \tan\left(\frac{x}{2}\right) + \cot\left(\frac{x}{2}\right)$

q $\cos\beta + \sin\beta = \dfrac{\cos 2\beta}{\cos\beta - \sin\beta}$

r $\tan\alpha + \tan\beta = \dfrac{\sin(\alpha+\beta)}{\cos\alpha\cos\beta}$

s $\sin^2\frac{\theta}{2} = \dfrac{1-\cos\theta}{2}$

t $\dfrac{\sin^3 x + \cos^3 x}{\sin x + \cos x} = 1 - \frac{1}{2}\sin 2x$

Extra examples and questions

Answers

3.4 Trigonometric Functions

The sine, cosine and tangent functions

As we saw at the beginning of this chapter, there is an infinite set of angles all of which give values (when they exist) for the main trigonometric ratios. We also noticed that the trigonometric ratios behave in such a way that values are repeated over and over. This is known as periodic behaviour. Many real world phenomena are periodic in the sense that the same patterns repeat over time. The trigonometric functions are often used to model such phenomena which include sound waves, light waves, alternating current electricity and other more approximately periodic events such as tides and even our moods (i.e. biorhythms).

Notice how we have introduced the term 'trigonometric function', replacing the term 'trigonometric ratio'. By doing this we can extend the use of the trigonometric ratios to a new field of problems.

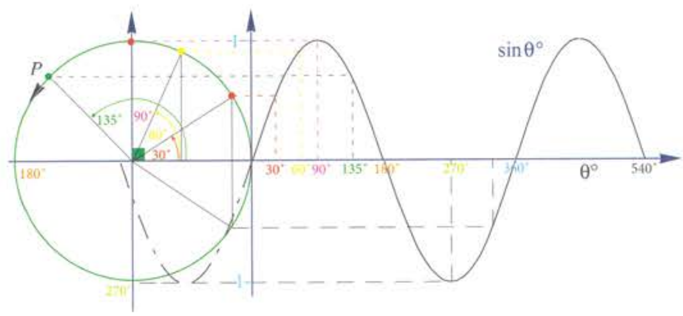
When the trigonometric functions are used for these purposes, the angles are almost always measured in radians. However, there is no reason why we cannot use degrees. It will always be obvious from the equation as to which mode of angle we are using. An expression such as $\sin x$ will imply (by default) that the angle is measured in radians, otherwise it will be written as $\sin x^{\circ}$, implying that the angle is measured in degrees.

What do trigonometric functions look like?

The sine and cosine values have displayed a periodic nature. This means that, if we were to plot a graph of the sine values versus their angles or the cosine values versus their angles, we could expect their graphs to demonstrate periodic behaviour. We start by plotting points.

The Sine Function

θ	0	30	45	60	90	120	135	150	180	...	330	360
$\sin\theta^{\circ}$	0.0	0.5	0.71	0.87	1.0	0.86	0.71	0.5	0.0	...	–0.5	0.0



Notice that, as the sine of angle θ corresponds to the y-value of point P on the unit circle, as P moves around the circle in an anticlockwise direction, we can measure the y-value and plot it on a graph as a function of θ (as above).

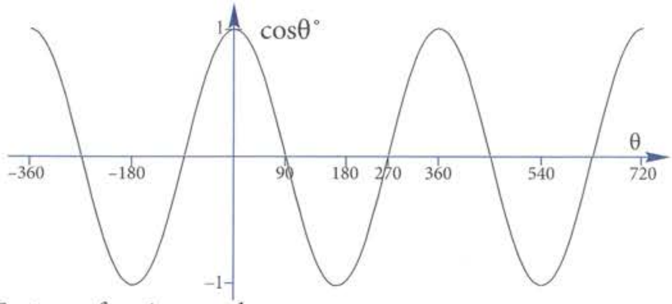
Feature of sine graph:

1. Maximum value = 1, Minimum value = –1
2. Period = 360° (i.e. graph repeats itself every 360°)
3. If P moves in a clockwise direction, y-values continue in their periodic nature (see dashed part of graph).

The Cosine Function

θ	0	30	45	60	90	120	135	150	180	...	330	360
$\cos\theta^{\circ}$	1.0	0.87	0.71	0.5	0.0	–0.5	–0.7	–0.8	–1.0	...	0.8	1.0

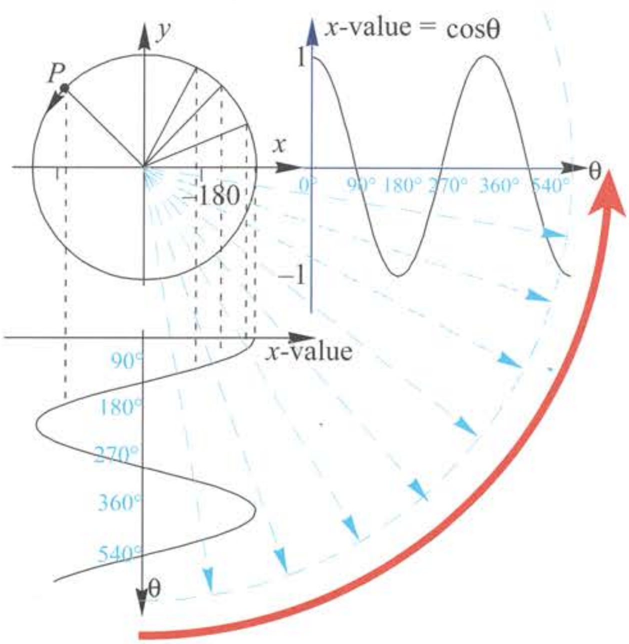
Plotting these points on a cosθ° versus θ-axis, we have:



Feature of cosine graph:

1. Maximum value = 1, Minimum value = −1
2. Period = 360° (i.e. graph repeats itself every 360°)
3. If P moves in a clockwise direction, x-values continue in their periodic nature.

There is a note to be made about using the second method (the one used to obtain the sine graph) when dealing with the cosine graph. As the cosine values correspond to the x-values on the unit circle, the actual cosine graph should have been plotted as shown below. However, for the sake of consistency, we convert the 'vertical graph' to the more standard 'horizontal graph':



There are some common observations to be made from these two graphs:

1. We have that the period of each of these functions is 360°. This is the length that it takes for the curve to start repeating itself.
2. The amplitude of the function is the distance between the centre line (in this case the θ -axis) and one of the maximum points. In this case, the amplitude is 1.

The sine and cosine functions are useful for modelling wave phenomena such as sound, light, water waves etc.

The Tangent Function

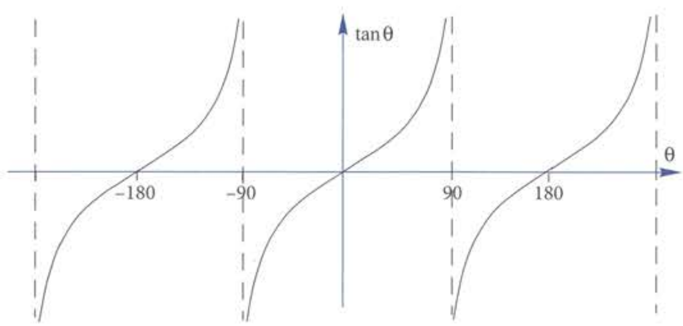
The third trigonometric function (tangent) is defined as:

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

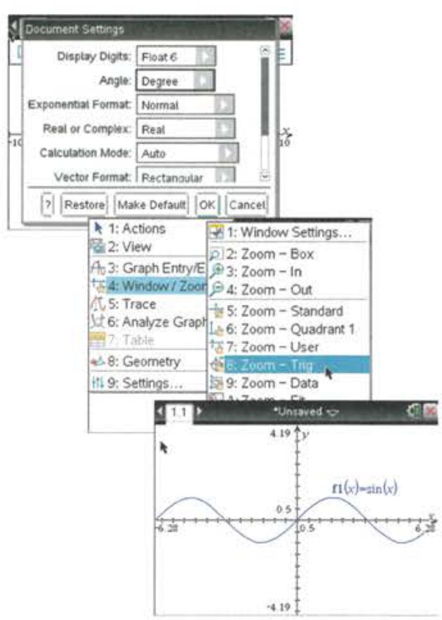
and so is defined for all angles for which the cosine function is non-zero.

The angles for which the tangent function is not defined correspond to the x-axis intercepts of the cosine function which are ±90°, ±270°, ±450°, … . At these points the graph of the tangent function has vertical asymptotes.

The period of the tangent function is 180°, which is half that of the sine and cosine functions. Since the tangent function has a vertical asymptote, it cannot be said to have an amplitude. It is also generally true that the tangent function is less useful than the sine and cosine functions for modelling applications. The graph of the basic tangent function is:



When sketching these graphs using a calculator, be sure that the **WINDOW** settings are appropriate for the **MODE** setting. In the case of degrees we have:



Transformations of Trigonometric Functions

We now consider some of the possible transformations that can be applied to the standard sine and cosine function and look at how these transformations affect the basic properties of both these graphs. These effects are the same as those discussed in §2.3.

1 Vertical translations

Functions of the type $f(x) = \sin(x) + c$, $f(x) = \cos(x) + c$ and $f(x) = \tan(x) + c$. represent vertical translations of the curves of $\sin(x)$, $\cos(x)$ and $\tan(x)$ respectively. If $c > 0$ the graph is moved vertically up and if $c < 0$ the graph is moved vertically down.

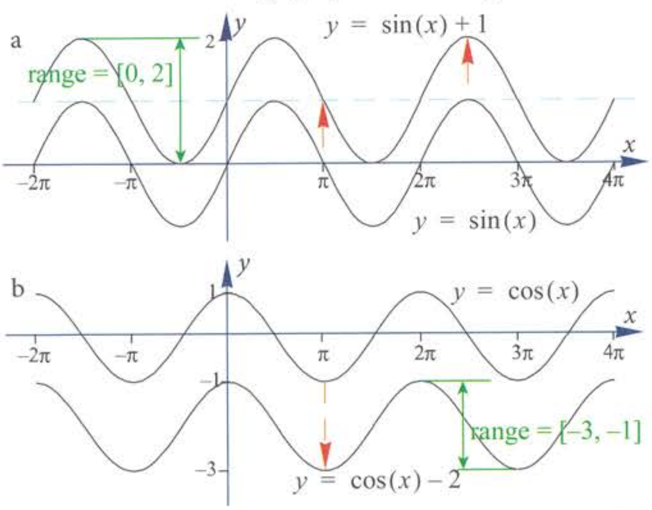
Example 3.4.1

Sketch the graphs of the functions for x-values in the range -2π to 4π.

a $y = \sin(x) + 1$ b $y = \cos(x) - 2$

A graph sketch should show all the important features of a graph. In this case, the axes scales are important and should show the correct period (2π) and range $[-3,-1]$.

That is, adding or subtracting a fixed amount to a trigonometric function translates the graph parallel to the y-axis.



2 Horizontal translations

Functions of the type $f(x) = \sin(x \pm \alpha)$, $f(x) = \cos(x \pm \alpha)$ and $f(x) = \tan(x \pm \alpha)$ where $\alpha > 0$ are horizontal translations of the curves $\sin(x)$, $\cos(x)$ and $\tan(x)$ respectively. In the context of trigonometric functions, horizontal transformations are often referred to as **phase shifts**.

So that:

$f(x) = \sin(x - \alpha)$, $f(x) = \cos(x - \alpha)$ and $f(x) = \tan(x - \alpha)$

are translations to the right.

while

$f(x) = \sin(x + \alpha)$, $f(x) = \cos(x + \alpha)$ and $f(x) = \tan(x + \alpha)$

are translations to the left.

Example 3.4.2

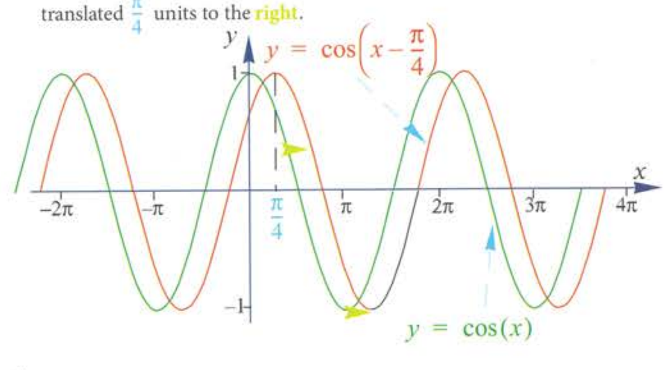
For $-2\pi \le x \le 4\pi$, sketch the graphs of the curves with equations:

a $y = \cos\left(x - \frac{\pi}{4}\right)$ b $y = \cos\left(x + \frac{\pi}{3}\right)$

c $y = \tan\left(x - \frac{\pi}{6}\right)$

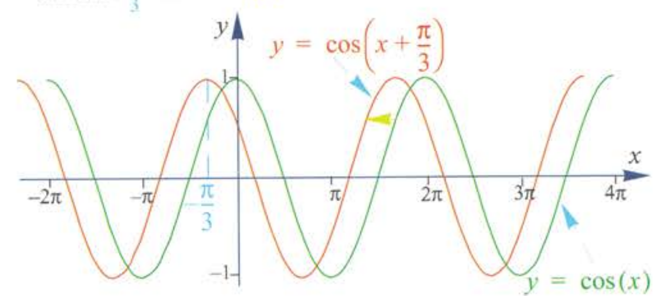
a

This is the basic cosine graph translated $\frac{\pi}{4}$ units to the right.

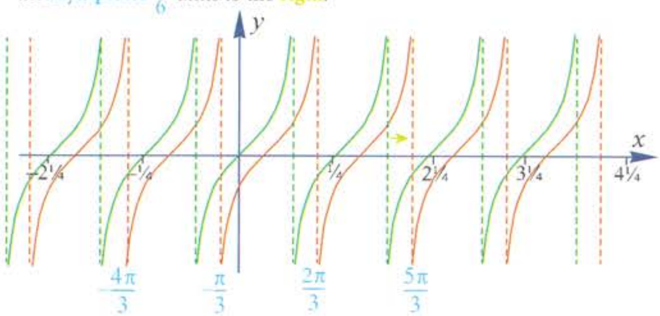


b

This graph is the basic cosine graph translated $\frac{\pi}{3}$ units to the left.



This is the tangent graph translated $\frac{\pi}{6}$ units to the right. This also translates the asymptotes $\frac{\pi}{6}$ units to the right.

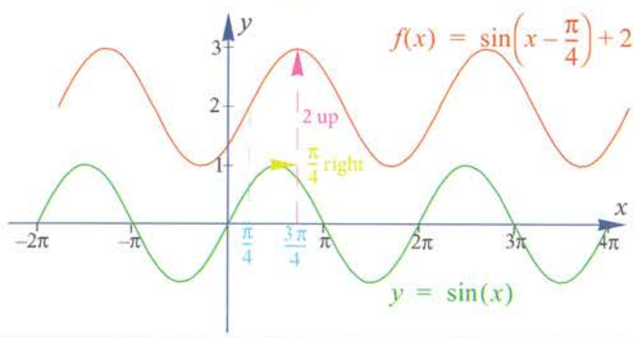


Of course, it is also possible to combine vertical and horizontal translations, as the next example shows.

Example 3.4.3

Sketch the graphs of the function $f(x) = \sin\left(x - \frac{\pi}{4}\right) + 2$, $-2\pi \le x \le 4\pi$.

Basic sine graph translated $\frac{\pi}{4}$ units right and 2 up.



3 Dilations

Functions of the form $f(x) = a\sin(x)$, $f(x) = a\cos(x)$ and $f(x) = a\tan(x)$ are dilations of the curves $\sin(x)$, $\cos(x)$ and $\tan(x)$ respectively, parallel to the ***y***-axis.

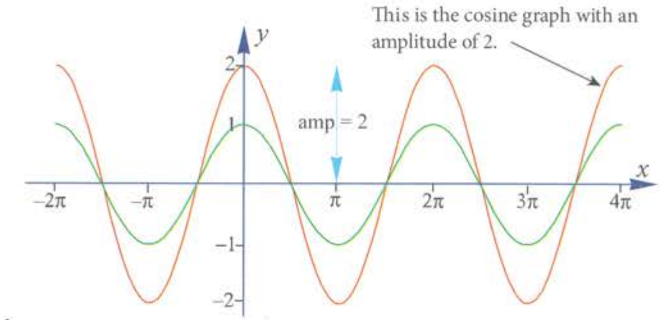
In the case of the sine and cosine functions, the amplitude becomes $|a|$ and not 1. This dilation does not affect the shape of the graph. Also, as the tangent function extends indefinitely, the term amplitude has no relevance.

Example 3.4.4

Sketch the graphs of the following functions for $-2\pi \le x \le 4\pi$.

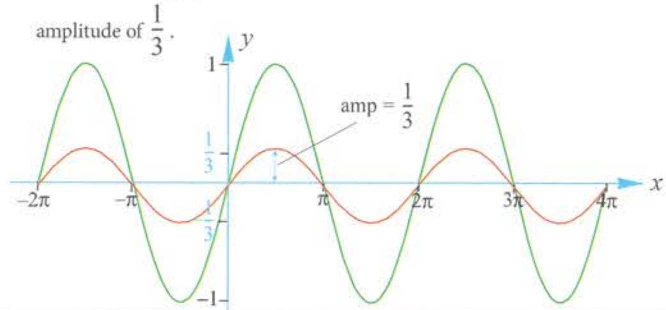
a $f(x) = 2\cos x$ b $f(x) = \frac{1}{3}\sin x$

a



b

This is the sine graph with an amplitude of $\frac{1}{3}$.



Functions of the form $f(x) = \sin(bx)$, $f(x) = \cos(bx)$ and $f(x) = \tan(bx)$ are dilations of the curves $\sin(x)$, $\cos(x)$ and $\tan(x)$ respectively, parallel to the ***x***-axis.

This means that the period of the graph is altered. It can be valuable to remember and use the formula that relates the value of b to the period τ of the dilated function:

1. The graph of $f(x) = \cos(bx)$ will show b cycles in 2π radians, meaning that its period will be given by $\tau = 2\pi/b$.

2. The graph of $f(x) = \sin(bx)$ will show b cycles in 2π radians, meaning that its period will be given by $\tau = 2\pi/b$.

3. The graph of $f(x) = \tan(bx)$ will show b cycles in π radians, meaning that its period will be given by $\tau = \pi/b$.

Note: In the case of the tangent function, whose original period is π, the new period is $\tau = \pi/b$.

Example 3.4.5

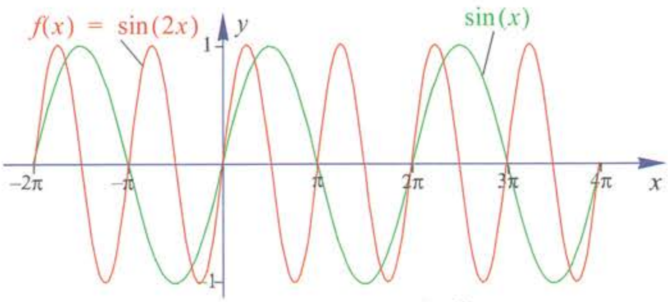
Sketch graphs of the following functions for x-values in the range -2π to 4π.

a $\quad f(x) = \sin(2x)$ $\qquad$ b $\quad f(x) = \cos\left(\frac{x}{2}\right)$

c $\quad f(x) = \tan\left(\frac{x}{4}\right)$

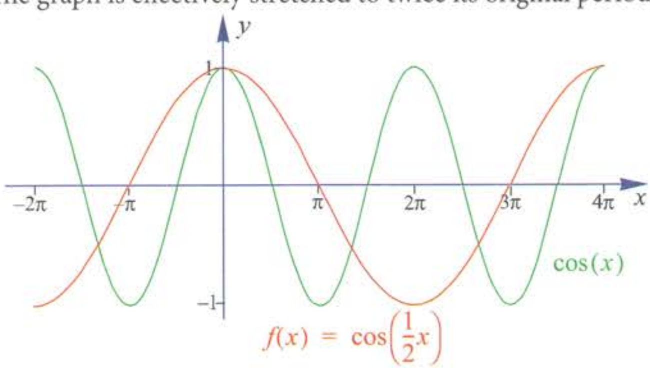
a $\quad$ The value of n is 2 so the period is $\tau = \frac{2\pi}{n} = \frac{2\pi}{2} = \pi$.

Note that this means that the period is **half** that of the basic sine function.



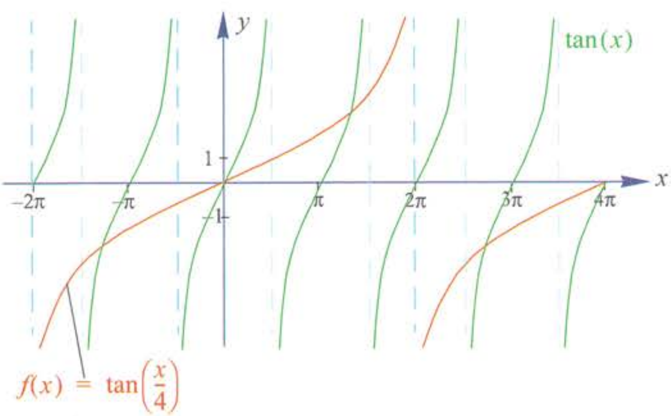
b $\quad$ In this case the value of $n = \frac{1}{2}$ and the period $\tau = \frac{2\pi}{n} = \frac{2\pi}{1/2} = 4\pi$.

The graph is effectively stretched to twice its original period.



c $\quad$ In this case, with $f(x) = \tan\left(\frac{x}{4}\right)$, the value of $n = \frac{1}{4}$

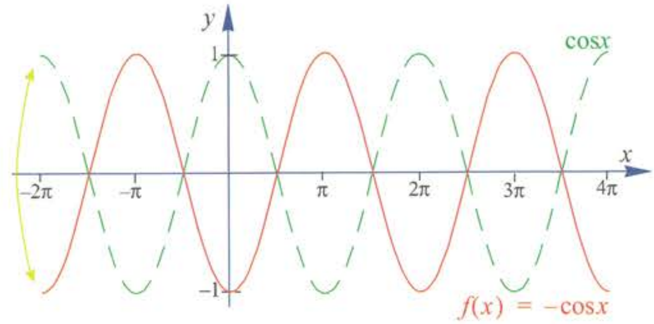
and the period $\tau = \frac{\pi}{n} = 4\pi$.



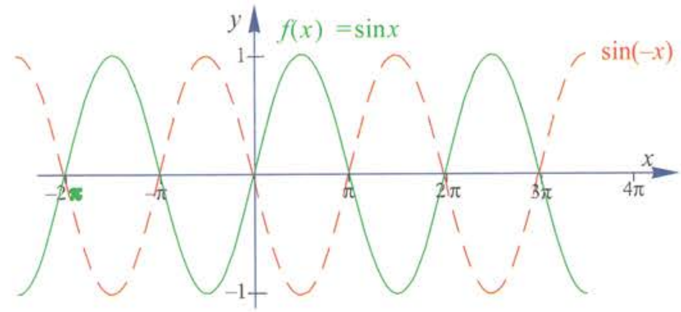
4 Reflections

Recall that the graph of $y = -f(x)$ is the graph of $y = f(x)$ reflected about the x-axis, while that of $y = f(-x)$ is the graph of $y = -f(x)$ reflected about the y-axis.

$f(x) = -\cos x$ is the basic cosine graph (green broken line) reflected in the $\boldsymbol{x}$-axis.



$f(x) = \sin(-x$ is the basic sine graph reflected in the $\boldsymbol{y}$-axis.



5 Combined transformations

You may be required to combine some or all of these transformations in a single function.

The functions of the type:

$$f(x) = a\sin[b(x+c)]+d \text{ and } f(x) = a\cos[b(x+c)]+d$$

have:

1. an amplitude of $|a|$ (i.e. the absolute value of a).

2. a period of $\frac{2\pi}{b}$

3. a horizontal translation of c units, $c > 0 \Rightarrow$ to the left $c < 0 \Rightarrow$ to the right.

4. a vertical translation of d units, $d > 0 \Rightarrow$ up $d < 0 \Rightarrow$ down.

Care must be taken with the horizontal translation.

For example, the function $f(x) = 2\cos\left(3x + \frac{\pi}{2}\right) - 1$

has a horizontal translation of $\frac{\pi}{6}$ to the left, not $\frac{\pi}{2}$!

This is because:

$$f(x) = 2\cos\left(3x + \frac{\pi}{2}\right) - 1 = 2\cos\left[3\left(x + \frac{\pi}{6}\right)\right] - 1 .$$

i.e. if the coefficient of x is not one, we must first express the function in the form $a\cos[b(x + c)] + d$.

Similarly we have: $f(x) = a\tan[b(x+c)]+d$

1. no amplitude (as it is not appropriate for the tan function).

2. a period of $\frac{\pi}{b}$

3. a horizontal translation of c units, $c > 0 \Rightarrow$ to the left $c < 0 \Rightarrow$ to the right.

4. a vertical translation of d units, $d > 0 \Rightarrow$ up $d < 0 \Rightarrow$ down.

Example 3.4.6

Sketch graphs of the following functions for x-values in the range -2π to 4π.

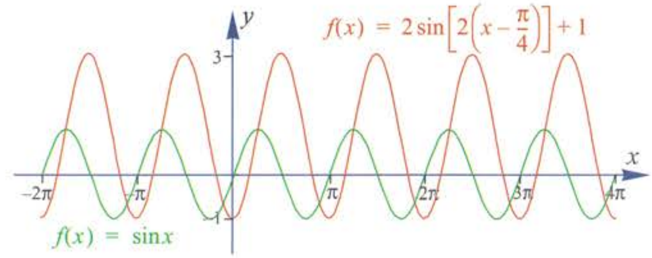
a $f(x) = 2\sin\left[2\left(x - \frac{\pi}{4}\right)\right] + 1$

b $f(x) = -\cos\frac{1}{2}\left(x - \frac{\pi}{3}\right) + 2$

c $f(x) = -\frac{1}{2}\tan\frac{1}{2}\left(x - \frac{\pi}{2}\right)$

a $f(x) = 2\sin\left[2\left(x - \frac{\pi}{4}\right)\right] + 1 .$

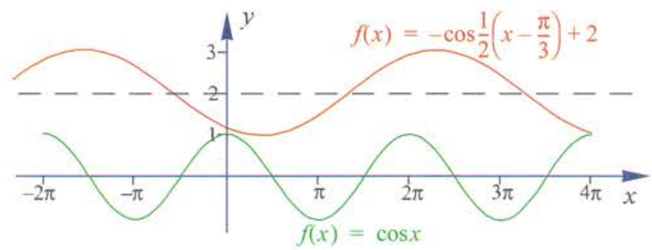
This graph has an amplitude of 2, a period of π, a horizontal translation of $\frac{\pi}{4}$ units to the right and a vertical translation of 1 unit up.



b

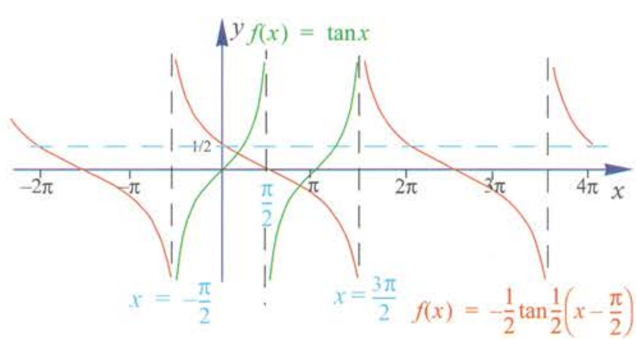
$$f(x) = -\cos\frac{1}{2}\left(x - \frac{\pi}{3}\right) + 2$$

The transformations are a reflection in the x-axis, a dilation of factor 2 parallel to the x-axis and a translation of $\frac{\pi}{3}$ right and 2 up.

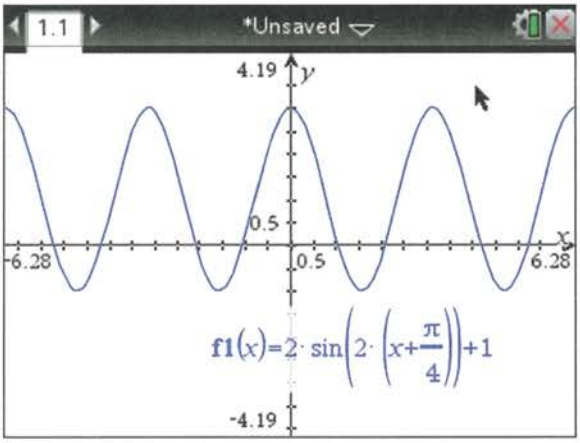


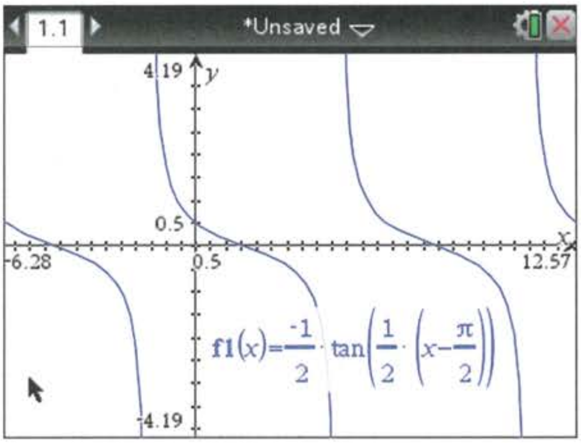
c $f(x) = -\frac{1}{2}\tan\frac{1}{2}\left(x - \frac{\pi}{2}\right) .$

The transformations are a reflection in the x-axis, a vertical dilation with factor ½, a horizontal dilation with factor 2 and a translation of $\frac{\pi}{2}$ to the right.



Again we see that a graphics calculator is very useful in such situations – in particular it allows for a checking process.





Exercise 3.4.1

1. State the period of the following functions.

   a $f(x) = \sin\frac{1}{2}x$

   b $f(x) = \cos 3x$

   c $f(x) = \tan\frac{x}{3}$

   d $g(x) = \cos\left(\frac{x}{2} - \pi\right)$

   e $g(x) = 4\sin(\pi x + 2)$

   f $g(x) = 3\tan\left(\frac{\pi}{2} - 2x\right)$

2. State the amplitude of the following functions.

   a $f(x) = 5\sin 2x$

   b $g(x) = -3\cos\frac{x}{2}$

   c $g(x) = 4 - 5\cos(2x)$

   d $f(x) = \frac{1}{2}\sin(3x)$

3. Find the period and, where appropriate, the amplitude of the following functions.

   a $y = 2\sin x$

   b $y = 3\cos\frac{x}{3}$

   c $y = 3\tan x$

   d $2\tan(x - 2\pi)$

   e $y = -4\sin\left[2\left(x + \frac{\pi}{6}\right)\right] + 1$

   f $y = 2 - 3\cos(2x - \pi)$

   g $y = -2\tan\frac{x}{6}$

   h $y = \frac{1}{4}\cos\left[3\left(x - \frac{3\pi}{4}\right)\right] + 5$

   i $y = 4\tan\left(\frac{x-4}{3}\right) - 3$

   j $y = -\frac{2}{3}\cos\left(\frac{3}{4}\left(x + \frac{3\pi}{5}\right)\right) + 5$

4. Sketch the graphs of the curves with equations:

   a $y = 3\cos x, 0 \le x \le 2\pi$

   b $y = \sin\frac{x}{2}, -\pi \le x \le \pi$

   c $y = 2\cos\left(\frac{x}{3}\right), 0 \le x \le 3\pi$

d $y = -\frac{1}{2}\sin 3x,\ 0 \le x \le \pi$

e $y = 4\tan\left(\frac{x}{2}\right),\ 0 \le x \le 2\pi$

f $y = \tan(-2x),\ -\pi \le x \le \frac{\pi}{4}$

g $y = \frac{1}{3}\cos(-3x),\ -\frac{\pi}{3} \le x \le \frac{\pi}{3}$

h $y = 3\sin(-2x),\ -\pi \le 0 \le \pi$

5. Sketch the graphs of the curves with equations:

a $y = 3\cos x + 3,\ 0 \le x \le 2\pi$

b $y = \sin\frac{x}{2} - 1,\ -\pi \le x \le \pi$

c $y = 2\cos\left(\frac{x}{3}\right) - 2,\ 0 \le x \le 3\pi$

d $y = -\frac{1}{2}\sin 3x + 2,\ 0 \le x \le \pi$

e $y = 4\tan\left(\frac{x}{2}\right) - 1,\ 0 \le x \le 2\pi$

f $y = \tan(-2x) + 2,\ -\pi \le x \le \frac{\pi}{4}$

g $y = \frac{1}{3}\cos(-3x) + \frac{1}{3},\ -\frac{\pi}{3} \le x \le \frac{\pi}{3}$

h $y = 3\sin(-2x) - 2,\ -\pi \le 0 \le \pi$

6. Sketch the graphs of the curves with equations:

a $y = 3\cos\left(x + \frac{\pi}{2}\right),\ 0 \le x \le 2\pi$

b $y = \sin\left(\frac{x}{2} - \pi\right),\ -\pi \le x \le \pi$

c $y = 2\cos\left(\frac{x}{3} + \frac{\pi}{6}\right),\ 0 \le x \le 3\pi$

d $y = -\frac{1}{2}\sin(3x + 3\pi),\ 0 \le x \le \pi$

e $y = 4\tan\left(\frac{x}{2} - \frac{\pi}{4}\right),\ 0 \le x \le 2\pi$

f $y = \tan(-2x + \pi),\ -\pi \le x \le \frac{\pi}{4}$

g $y = \frac{1}{3}\cos(-3x - \pi),\ -\frac{\pi}{3} \le x \le \frac{\pi}{3}$

h $y = 3\sin\left(-2x - \frac{\pi}{2}\right),\ -\pi \le 0 \le \pi$

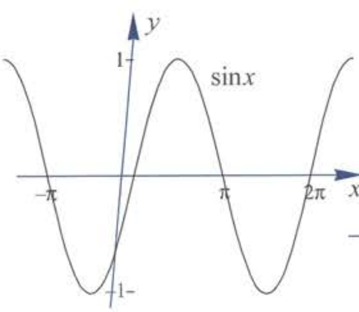
Extra questions

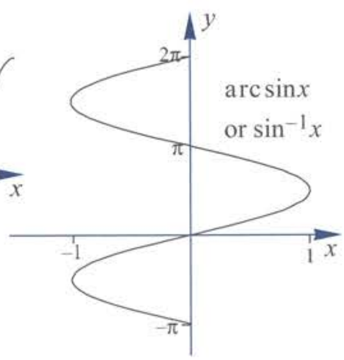
Answers

3.5 Inverse Trigonometric Functions

The Inverse Sine Function

The trigonometric functions are many-to-one which means that, unless we are careful about defining domains, their inverses are not properly defined. The basic graphs of the sine function and its inverse (after reflection about the line $y = x$ for the arcsin$x$ function) are:



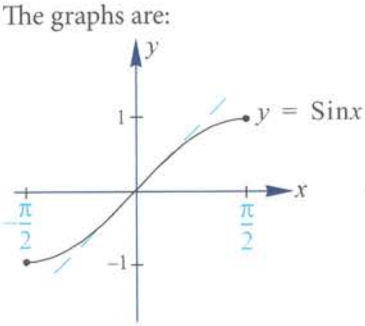


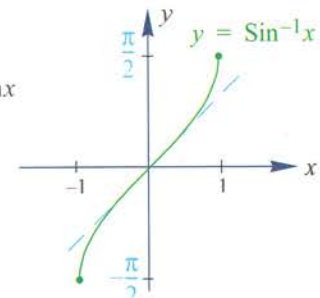
The inverse as depicted here is not a function (as it is one : many). This is inconvenient as the inverse trigonometric functions are useful. The most useful solution to this problem is to restrict the domain of the function to an interval over which it is one-to-one.

In the case of the sine function, this is usually taken as $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Though this is not the only possible choice, it is one that allows for consistency to be maintained in literature and among mathematicians. The function thus defined is written with a capital letter: $f(x) = \text{Sin}(x), x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

The graphs are:





Notice then that the domain of $\text{Sin}^{-1}x$ = range of $\text{Sin}x = [-1, 1]$

and the range of $\text{Sin}^{-1}x$ = domain of $\text{Sin}x = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

With these restrictions, we refer to $\text{Sin}^{-1}x$ (which is sometimes denoted by Arcsin$x$) as the principal value of arcsin$x$.

For example, $\arcsin\left(\frac{1}{2}\right) = \frac{\pi}{6}$ or $\frac{5\pi}{6}$ or $-\frac{7\pi}{6}$ or ....

However, $\text{Arcsin}\left(\frac{1}{2}\right)$ has only one value (the principal value),

so that $\text{Arcsin}\left(\frac{1}{2}\right) = \frac{\pi}{6}$.

From our fundamental identity property of inverse functions, i.e. $f o f^{-1}(x) = f^{-1} o f(x) = x$, we have that:

$$\text{Sin}(\text{Sin}^{-1}x) = x, \ -1 \le x \le 1 \text{ and } \text{Sin}^{-1}(\text{Sin}x) = x, \ -\frac{\pi}{2} \le x \le \frac{\pi}{2}$$

Therefore, $\text{Sin}(\text{Sin}^{-1}x) = x = \text{Sin}^{-1}(\text{Sin}x)$ only if $-1 \le x \le 1$.

This then means that sometimes we can provide a meaningful interpretation to expressions such as $\sin(\text{Sin}^{-1}x)$ & $\text{Sin}^{-1}(\sin x)$ – as long as we adhere to the relevant restrictions.

Example 3.5.1

Give the exact value of:

a $\quad \text{Sin}^{-1}\frac{1}{2}$ $\qquad$ b $\quad \text{Arcsin}\left(-\frac{\sqrt{3}}{2}\right)$

c $\quad \text{Sin}^{-1}(1.3)$ $\qquad$ d $\quad \text{Sin}^{-1}(\sin\pi)$

a $\quad$ As $\frac{1}{2} \in [-1, 1] \Rightarrow \text{Sin}^{-1}\frac{1}{2}$ exists.

Therefore, $\text{Sin}^{-1}\frac{1}{2} = \frac{\pi}{6}$.

b $\quad$ As $-\frac{\sqrt{3}}{2} \in [-1, 1] \Rightarrow \text{Arcsin}\left(-\frac{\sqrt{3}}{2}\right)$ exists.

Now, $\text{Arcsin}\left(-\frac{\sqrt{3}}{2}\right) = -\text{Arcsin}\left(\frac{\sqrt{3}}{2}\right) = -\frac{\pi}{3}$.

c $\quad$ As $1.3 \notin [-1, 1] \Rightarrow \text{Sin}^{-1}(1.3)$ does **not** exist.

d $\quad$ As $\sin\pi \in [-1, 1] \Rightarrow \text{Sin}^{-1}(\sin\pi)$ exists.

So, $\text{Sin}^{-1}(\sin\pi) = \text{Sin}^{-1}(0) = 0$.

Note that $\text{Sin}^{-1}(\sin\pi) \neq \pi$! Why?

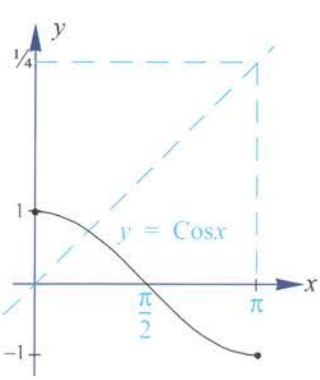
The Inverse Cosine Function

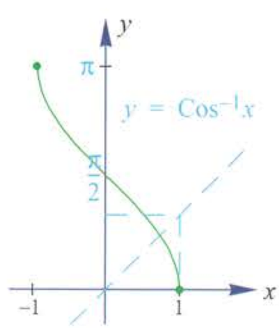
For similar reasons as those for the sine function, the cosine function, $\cos x$, $x \in]-\infty,\infty[$ being a many-to-one function, with its inverse, $\arccos x$, $-1 \leq x \leq 1$ (or $\cos^{-1}x$, $-1 \leq x \leq 1$) needs to be restricted to the domain $[0,\pi]$, to produce a function that is one-to-one.

The function $y = \text{Cos}x$, $x \in [0,\pi]$, $-1 \leq y \leq 1$ (with a capital 'C') will have the inverse function defined as:

$$f(x) = \text{Cos}^{-1}x,\ -1 \leq x \leq 1,\ -1 \leq y \leq \pi$$

The graphs of these functions are:





Notice that the domain of $\text{Cos}^{-1}x$ = range of $\text{Cos}x = [-1, 1]$ and the range of $\text{Cos}^{-1}x$ = domain of $\text{Cos}x = [0,\pi]$.

When these restrictions are adhered to, we refer to $\text{Cos}^{-1}x$ (which is sometimes denoted by $\text{Arccos}x$) as the principal value of $\arccos x$.

From our fundamental identity property of inverse functions,

i.e. $f\circ f^{-1}(x) = f^{-1}\circ f(x) = x$, we have that:

$\text{Cos}(\text{Cos}^{-1}x) = x,\ -1 \leq x \leq 1$ and $\text{Cos}^{-1}(\text{Cos}x) = x,\ -0 \leq x \leq \pi$

Therefore, $\text{Cos}(\text{Cos}^{-1}x) = x = \text{Cos}^{-1}(\text{Cos}x)$ only if $0 \leq x \leq 1$.

This means that we can provide a meaningful interpretation of expressions such as $\cos(\text{Cos}^{-1}x)$ and $\text{Cos}^{-1}(\text{Cos}x)$– as long as we adhere to the relevant restrictions.

Note also that in this case, $\text{Cos}^{-1}(-x) \neq -\text{Cos}^{-1}(x)$.

Example 3.5.2

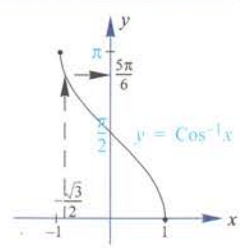
Given the exact value of:

a $\quad \text{Cos}^{-1}\frac{1}{2}$ $\qquad$ b $\quad \text{Arccos}\left(-\frac{\sqrt{3}}{2}\right)$

c $\quad \text{Cos}^{-1}\left(\cos\left(\frac{3\pi}{2}\right)\right)$

a $\quad$ As $\frac{1}{2} \in [-1, 1] \Rightarrow \text{Cos}^{-1}\frac{1}{2}$ exists.

Therefore, $\text{Cos}^{-1}\frac{1}{2} = \frac{\pi}{3}$.



b $\quad$ As $-\frac{\sqrt{3}}{2} \in [-1, 1] \Rightarrow \text{Cos}^{-1}\left(-\frac{\sqrt{3}}{2}\right)$ exists.

Let $y = \text{Cos}^{-1}\left(-\frac{\sqrt{3}}{2}\right)$, then, $\text{Cos}y = -\frac{\sqrt{3}}{2}$, $0 \leq y \leq \pi$.

$$\Leftrightarrow y = \pi - \frac{\pi}{6}$$

$$= \frac{5\pi}{6}$$

c $\quad$ As $\cos\left(\frac{3\pi}{2}\right) \in [-1, 1] \Rightarrow \text{Cos}^{-1}\left(\cos\left(\frac{3\pi}{2}\right)\right)$ exists.

$\text{Cos}^{-1}\left(\cos\left(\frac{3\pi}{2}\right)\right) = \text{Cos}^{-1}(0) = \frac{\pi}{2}$.

Notice that $\text{Cos}^{-1}\left(\cos\left(\frac{3\pi}{2}\right)\right) \neq \frac{3\pi}{2}$.

Example 3.5.3

Give the exact value of:

a $\sin\left(\text{Arccos}\left(\frac{1}{\sqrt{2}}\right)\right)$ b $\cos\left(\text{Sin}^{-1}\left(\frac{1}{4}\right)\right)$

c $\sin\left(\frac{\pi}{2} - \text{Cos}^{-1}\left(\frac{3}{4}\right)\right)$

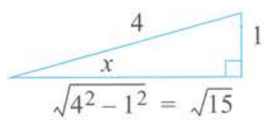
a

Let $\text{Arccos}\left(\frac{1}{\sqrt{2}}\right) = x \therefore$ as $\frac{1}{\sqrt{2}} \in [0, 1] \Rightarrow \text{Arccos}\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}$

Then, $\sin\left(\text{Arccos}\left(\frac{1}{\sqrt{2}}\right)\right) = \sin(x) = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}$.

Let $\text{Sin}^{-1}\left(\frac{1}{4}\right) = x \therefore$ as $\frac{1}{4} \in [-1, 1] \Rightarrow \text{Sin}^{-1}\left(\frac{1}{4}\right)$ exists.

b However, this time we cannot obtain an exact value for x, so we make use of a right-angled triangle:



Therefore, from the triangle we have that $\cos x = \frac{\sqrt{15}}{4}$.

i.e. $\cos\left(\text{Sin}^{-1}\left(\frac{1}{4}\right)\right) = \cos x = \frac{\sqrt{15}}{4}$.

c Let $\text{Cos}^{-1}\left(\frac{3}{4}\right) = \theta \therefore$ as $\frac{3}{4} \in [-1, 1] \Rightarrow \text{Cos}^{-1}\left(\frac{3}{4}\right)$

Then, $\sin\left(\frac{\pi}{2} - \text{Cos}^{-1}\left(\frac{3}{4}\right)\right) = \sin\left(\frac{\pi}{2} - \theta\right) = \cos\theta$.

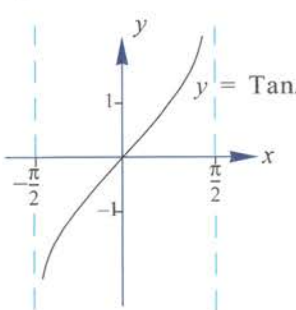
Therefore, $\sin\left(\frac{\pi}{2} - \text{Cos}^{-1}\left(\frac{3}{4}\right)\right) = \cos\left(\text{Cos}^{-1}\left(\frac{3}{4}\right)\right) = \frac{3}{4}$

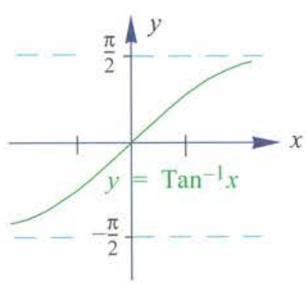
The Inverse Tangent Function

The tangent function can be made one : one by restricting its domain to the open interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

$f(x) = \text{Tan}(x), x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

The function $y = \text{Tan}(x), x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), -\infty < y < \infty$, (with a capital 'T') will have the inverse function defined as $f(x) = \text{Tan}^{-1}(x), , -\infty < x < \infty$. The graphs of these functions are:





Notice then that the domain of $\text{Tan}^{-1}x$ = range of $\text{Tan}x = (-\infty, \infty)$ and the range of $\text{Tan}^{-1}x$ = domain of $\text{Tan}x = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

When these restrictions are adhered to, we refer to $\text{Tan}^{-1}x$ (which is sometimes denoted by Arctan$x$) as the principal value of arctan$x$.

$\text{Tan}(\text{Tan}^{-1}x) = x, -\infty \le x \le \infty$ and $\text{Tan}^{-1}(\text{Tan}x) = x, -\frac{\pi}{2} \le x \le \frac{\pi}{2}$

From our fundamental identity property of inverse functions, i.e. $f o f^{-1}(x) = f^{-1} o f(x) = x$, we have that

Therefore, $\text{Tan}(\text{Tan}^{-1}x) = x = \text{Tan}^{-1}(\text{Tan}x)$ only if $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$.

As we saw with the sine and cosine functions, it **may** also be possible to evaluate expressions such as $\tan(\text{Tan}^{-1}x)$ and = $\text{Tan}^{-1}(\tan x)$.

For example, $\tan(\text{Tan}^{-1}1) = \tan\left(\frac{\pi}{4}\right) = 1$,

however, $\text{Tan}^{-1}\left(\tan\frac{2\pi}{3}\right) = \text{Tan}^{-1}(-\sqrt{3}) = -\frac{\pi}{3}$.

Note also that $\text{Tan}^{-1}(-x) = -\text{Tan}(x)$

Example 3.5.4

Give the exact value of:

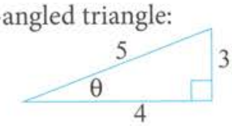
a $\tan\left(\text{Sin}^{-1}\left(-\frac{3}{5}\right)\right)$

b $\sin\left(2\text{Tan}^{-1}\left(\frac{1}{3}\right)\right)$

a As $-\frac{3}{5} \in [-1, 1] \Rightarrow \text{Sin}^{-1}\left(-\frac{3}{5}\right)$ exists.

Then, we let $\theta = \text{Sin}^{-1}\left(\frac{3}{5}\right)$, so that $\text{Sin}\theta = \frac{3}{5}$.

Next we construct an appropriate right-angled triangle:



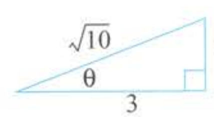
So, $\tan\left(\text{Sin}^{-1}\left(-\frac{3}{5}\right)\right) = \tan\left(-\text{Sin}^{-1}\left(\frac{3}{5}\right)\right)$

$= \tan(-\theta) = -\tan\theta = -\frac{3}{4}$

b As $\frac{1}{3} \in (-\infty, \infty) \Rightarrow \text{Tan}^{-1}\left(\frac{1}{3}\right)$ exists.

Let $\text{Tan}^{-1}\left(\frac{1}{3}\right) = \theta \therefore \text{Tan}\theta = \frac{1}{3}$.

Next we construct an appropriate right-angled triangle:



Then, $\sin\left(2\text{Tan}^{-1}\left(\frac{1}{3}\right)\right) = \sin 2\theta = 2\sin\theta\cos\theta$

$$= 2 \times \frac{1}{\sqrt{10}} \times \frac{3}{\sqrt{10}}$$

$$= \frac{3}{5}$$

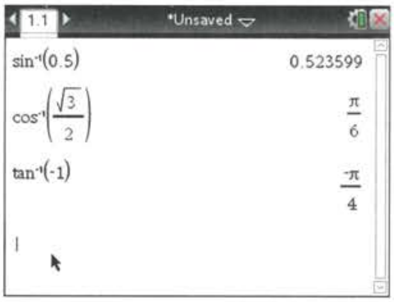
It is these restricted functions that are programmed into most calculators, spreadsheets etc.

If the calculator is set in radian mode, some sample calculations are:

$$\text{Sin}^{-1}\frac{1}{2} = \frac{\pi}{6}$$

$$\text{Cos}^{-1}\frac{\sqrt{3}}{2} = \frac{\pi}{6}$$

$$\text{Tan}^{-1}-1 = -\frac{\pi}{4}$$

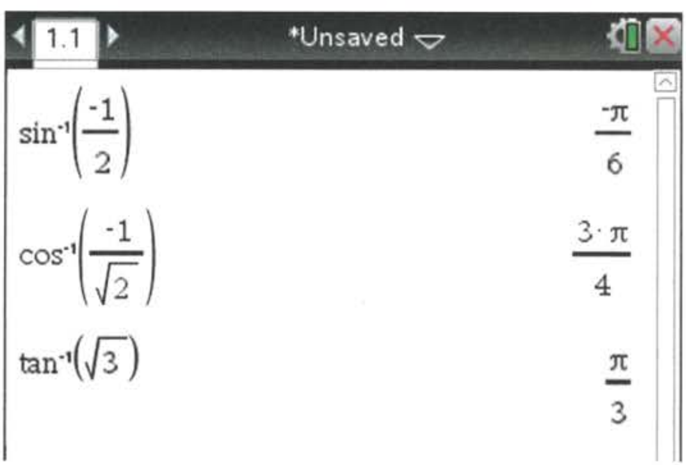


Example 3.5.5

Find the principal values of the following.

a $\text{Sin}^{-1}\left(-\frac{1}{2}\right)$ b $\text{Cos}^{-1}\left(-\frac{1}{\sqrt{2}}\right)$

c $\text{Tan}^{-1}\sqrt{3}$



Exercise 3.5.1

1. Find the principal values of the following, giving answers in radians.

a $\text{Tan}^{-1}1$ b $\text{Arc}\sin 1$

c $\text{Arc}\cos -1$

d $\text{Sin}^{-1}\frac{\sqrt{3}}{2}$ e $\text{Cos}^{-1}\frac{1}{\sqrt{2}}$

f $\text{Tan}^{-1}-\sqrt{3}$ g $\text{Tan}^{-1}2$

h $\text{Sin}^{-1}-0.7$ i $\text{Arc}\tan 0.1$

j $\text{Arc}\cos 0.3$ k $\text{Sin}^{-1}-0.6$

l $\text{Tan}^{-1}5$ m $\text{Cos}^{-1}3$

n $\text{Tan}^{-1}-30$ o $\text{Sin}^{-1}\left(\frac{7}{8}\right)$

2. Solve the following equations, giving exact answers.

a $\text{Arctan}x = \frac{3\pi}{4}$

b $\text{Arcsin}(2x) = \frac{\pi}{3}$

c $\text{Arccos}(3x) = \frac{5\pi}{4}$

3. Prove:

a $\text{Arctan}(4) - \text{Arctan}\left(\frac{3}{5}\right) = \frac{\pi}{4}$

b $\text{Sin}^{-1}\left(\frac{4}{5}\right) + \text{Sin}^{-1}\left(-\frac{4}{5}\right) = 0$

4. Solve for x, where:

$\text{Arctan}(3x) - \text{Arctan}(2x) = \text{Arctan}\left(\frac{1}{5}\right)$

5. Find the exact value of:

a $\sin\left[\frac{\pi}{2} - \text{Cos}^{-1}\left(\frac{2}{3}\right)\right]$

b $\cos\left[\frac{\pi}{2} + \text{Sin}^{-1}\left(-\frac{1}{3}\right)\right]$

c $\cos[\text{Tan}^{-1}(-\sqrt{3})]$

d $\tan\left(\text{Cos}^{-1}\left(\frac{4}{5}\right)\right)$

e $\sec\left(\text{Sin}^{-1}\left(-\frac{1}{3}\right)\right)$

f $\cot(\text{Tan}^{-1}(-1))$

Extra example and questions

Answers

3.6 Trigonometric Equations

We have already encountered solutions to trigonometric equations as part of a general observation in this chapter. There are two basic methods that can be used when solving trigonometric equations:

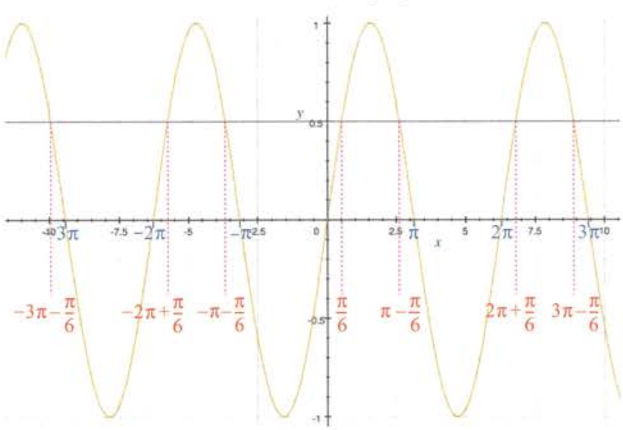
Method 1. Use the unit circle as a visual aid.

Method 2. Use the graph of the function as a visual aid.

The method you choose depends entirely on what you feel comfortable with. However, it is recommended that you become familiar with both methods.

Solution of $\sin x = a$, $\cos x = a$ and $\tan x = a$

The equation $\sin x = ½$ produces an infinite number of solutions. This can be seen from the graph of the sine function.



Using the principal angle $\left(\frac{\pi}{6}\right)$ and symmetry, the solutions generated are:

For $x \geq 0$ $x = \frac{\pi}{6}, \pi - \frac{\pi}{6}, 2\pi + \frac{\pi}{6}, 3\pi - \frac{\pi}{6}, \ldots$

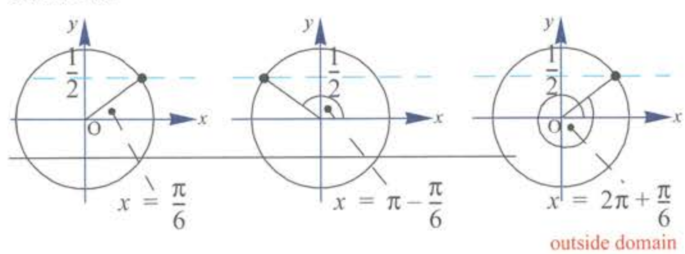
$$= \frac{\pi}{6}, \frac{5\pi}{6}, \frac{13\pi}{6}, \frac{17\pi}{6}, \ldots$$

For $x < 0$ $x = -\pi - \frac{\pi}{6}, -2\pi + \frac{\pi}{6}, -3\pi - \frac{\pi}{6}, \ldots$

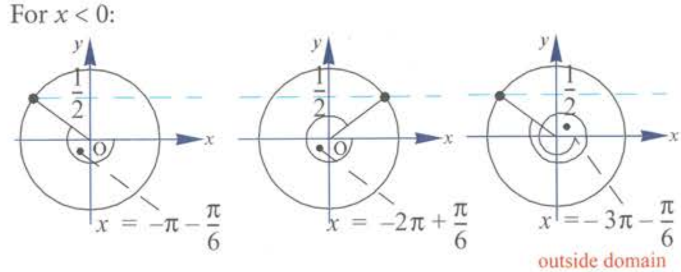
$$= -\frac{7\pi}{6}, -\frac{11\pi}{6}, -\frac{19\pi}{6}, \ldots$$

The same problem could have been solved using Method 1. We start by drawing a unit circle and we continue to move around the circle until we have all the required solutions within the domain restriction. Again, the use of symmetry plays an important role in solving these equations.

For $x \geq 0$:



For $x < 0$:



Again, for the restricted domain $-2\pi < x < \pi$,

we have $\sin x = \frac{1}{2}$ if $x = -\frac{7\pi}{6}, -\frac{11\pi}{6}, \frac{\pi}{6}, \frac{5\pi}{6}$.

The process is identical for the cosine and tangent functions.

Example 3.6.1

Solve the following, for $0 \le x \le 4\pi$, giving answers to 4 decimal places if no exact answers are available.

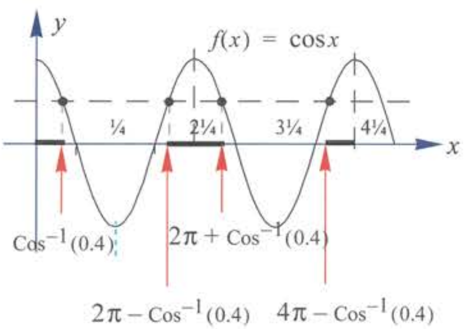
a $\cos x = 0.4$ b $\tan x = -1$

c $5\cos x - 2 = 0$

a Step 1: Find the reference angle:

$$x = \text{Cos}^{-1}(0.4) = 1.1593.$$

Step 2: Sketch the cosine graph (or use the unit circle):



Step 3: Use the reference angle and symmetry to obtain solutions.

Therefore, solutions are,

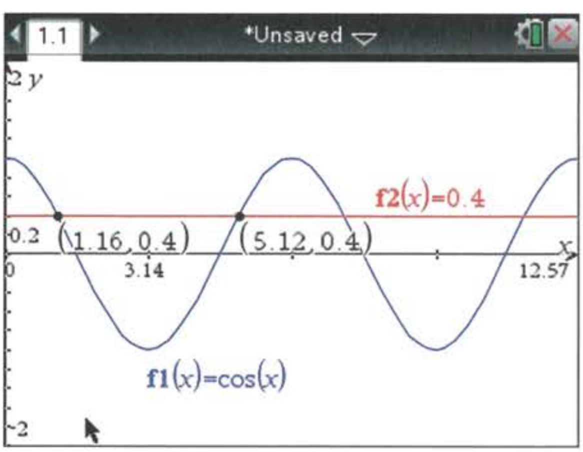
$$\text{Cos}^{-1}(0.4), 2\pi - \text{Cos}^{-1}(0.4), 2\pi + \text{Cos}^{-1}(0.4), 4\pi - \text{Cos}^{-1}(0.4)$$

$$= 1.1593, 5.1239, 7.4425, 11.4071$$

Step 4: Check that:

i all solutions are within the domain

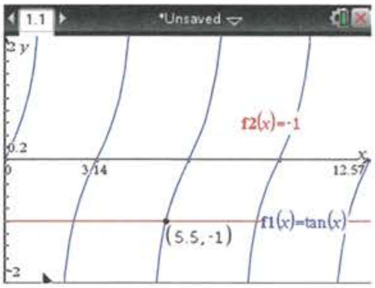
ii you have obtained all the solutions in the domain.



Step1: Find reference angle (in first quadrant):

$$\text{Tan}^{-1}(1) = \frac{\pi}{4} \qquad = \frac{3\pi}{4}, \frac{7\pi}{4}, \frac{11\pi}{4}, \frac{15\pi}{4}$$

From a calculator, we see that there are four solutions.



Then, as the four solutions obtained all lie in the interval $[0, 4\pi]$, Step 4 is satisfied.

c $$5\cos x - 2 = 0 \Leftrightarrow 5\cos x = 2 \Leftrightarrow \cos x = \frac{2}{5}$$

i.e. $\cos x = 0.4$

This is, in fact, identical to the equation in part a and so, we have that $x = 1.1593, 5.1239, 7.4425, 11.4071$

Part c Example 3.6.1 highlights the fact that it is possible to transpose a trigonometric equation into a simpler form, which can then be solved. Rather than remembering (or trying to commit to memory) the different possible forms of trigonometric equations and their specific solution processes, the four steps used (with possibly some algebraic manipulation) will always transform a (seemingly) difficult equation into one having a simpler form, as in Example 3.6.1.

Some forms of trigonometric equations are:

$\sin(kx) = a$, $\cos(x+c) = a$, $\tan(kx+c) = a$,

$b\cos(kx+c) = a$, $b\sin(kx+c)+d = a$

And, of course, then there are equations involving the secant, cosecant and cotangent functions.

However, even the most involved of these equations, e.g. $b\sin(kx+c)+d = a$, can be reduced to a simpler form:

1. Transpose:

$$b\sin(kx+c)+d = a \Leftrightarrow b\sin(kx+c) = a-d$$

$$\Leftrightarrow \sin(kx+c) = \frac{a-d}{b}$$

2. Substitute:

Then, setting $kx+c = \theta$ and $\frac{a-d}{b} = m$, we have $\sin\theta = m$ which can be readily solved.

3. Solve for new variable:

   So that the solutions to $\sin\theta = m$ are $\theta = \theta_1, \theta_2, \theta_3, \ldots$

4. Solve for original variable:

   We substitute back for θ and solve for x:

$$\begin{aligned} kx + c &= \theta_1, \theta_2, \theta_3, \ldots \\ \Leftrightarrow kx &= \theta_1 - c, \theta_2 - c, \theta_3 - c, \ldots \\ \Leftrightarrow x &= \frac{\theta_1 - c}{k}, \frac{\theta_2 - c}{k}, \frac{\theta_3 - c}{k}, \ldots \end{aligned}$$

   All that remains is to check that all the solutions have been obtained and that they all lie in the restricted domain.

The best way to see how this works is through a number of examples.

Example 3.6.2

Solve the following, for $0 \le x \le 2\pi$.

a $\cos(2x) = 0.4$ b $\tan\left(\frac{1}{2}x\right) = -1$

a Let $\theta = 2x$, so that we now solve the equation $\cos\theta = 0.4$.

From Example 3.6.1 a we already have the solutions, namely;

$$\text{Cos}^{-1}(0.4), 2\pi - \text{Cos}^{-1}(0.4), 2\pi + \text{Cos}^{-1}(0.4), 4\pi - \text{Cos}^{-1}(0.4)$$

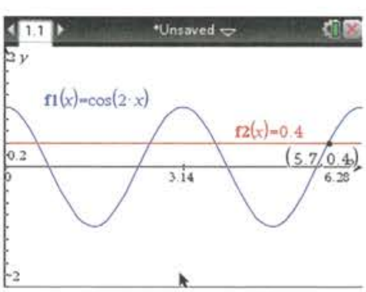
$$= 1.1593, 5.1239, 7.4425, 11.4071$$

However, we want to solve for x not θ. So, we substitute back for x:

i.e. $2x = 1.1593, 5.1239, 7.4425, 11.4071$

$$\therefore x = 0.5796, 2.5620, 3.7212, 5.7045$$

To check that we have all the solutions, we sketch the graphs of $y = \cos(2x)$ and $y = 0.4$ over the domain $0 \le x \le 2\pi$.



The diagram shows that there should be four solutions.

This time, to solve $\tan\left(\frac{1}{2}x\right) = -1$ we first let $\theta = \frac{1}{2}x$ so that we now need to solve the simpler equation $\tan\theta = -1$.

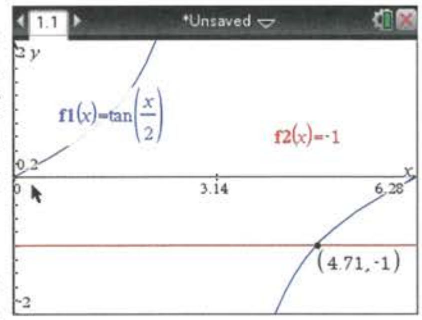
b From Example 3.6.1 b we have that

$$\begin{aligned} \theta &= \pi - \frac{\pi}{4}, 2\pi - \frac{\pi}{4}, 3\pi - \frac{\pi}{4}, 4\pi - \frac{\pi}{4} \\ &= \frac{3\pi}{4}, \frac{7\pi}{4}, \frac{11\pi}{4}, \frac{15\pi}{4} \end{aligned}$$

However, we want to solve for x, not θ . So, we substitute for x:

$$\begin{aligned} \frac{1}{2}x &= \frac{3\pi}{4}, \frac{7\pi}{4}, \frac{11\pi}{4}, \frac{15\pi}{4} \\ \therefore x &= \frac{3\pi}{2}, \frac{7\pi}{2}, \frac{11\pi}{2}, \frac{15\pi}{2} \end{aligned}$$

To check that we have all the solutions, we sketch the graphs of $y = \tan\left(\frac{1}{2}x\right)$ and $y = -1$ over the domain $0 \le x \le 2\pi$.



The diagram shows that there should be only one solution.

Therefore, the only solution is $x = \frac{3\pi}{2}$.

There is of course another step that could be used to help us predetermine which solutions are valid. This requires that we make a substitution not only into the equation, but also into the restricted domain statement.

In Example 3.6.2 b, after setting $\theta = \frac{1}{2}x$ to give $\tan\theta = -1$,

we next adjust the restricted domain: $\theta = \frac{1}{2}x \Leftrightarrow x = 2\theta$

So, from $0 \le x \le 2\pi$ we now have $0 \le 2\theta \le 2\pi \Leftrightarrow 0 \le \theta \le \pi$

That is, we have the equivalent equations:

$$\tan\left(\frac{1}{2}x\right) = -1, 0 \le x \le 2\pi : \tan\theta = -1, 0 \le \theta \le \pi$$

Example 3.6.3

Solve: $4.\sin(3x) = 2, 0^\circ \le x \le 360^\circ$.

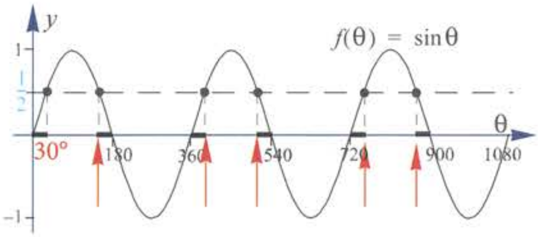
$4\sin(3x) = 2 \Leftrightarrow \sin(3x) = 0.5, 0° \le x \le 360°$.

Let $\theta = 3x \Rightarrow \sin\theta = 0.5$.

New domain:
$0° \le x \le 360° \Leftrightarrow 0° \le \frac{\theta}{3} \le 360° \Leftrightarrow 0° \le \theta \le 1080°$

Therefore, we have, $\sin\theta = 0.5, 0° \le \theta \le 1080°$.

The reference angle is $30°$. Then, by symmetry, we have:



$$\therefore \theta = 30°, 180° - 30°, 360° + 30°, 540° - 30°, 720° + 30°, 900° - 30°$$

$$\therefore 3x = 30°, 150°, 390°, 510°, 750°, 870°$$
$$\therefore x = 10°, 50°, 130°, 170°, 250°, 290°$$

All solutions lie within the **original** specified domain, $0° \le x \le 360°$.

Example 3.6.4

Solve $2\cos\left(\frac{x}{2} + \frac{\pi}{2}\right) - \sqrt{3} = 0$, for $-\pi \le x \le 4\pi$.

$2\cos\left(\frac{x}{2} + \frac{\pi}{2}\right) - \sqrt{3} = 0 \Leftrightarrow \cos\left(\frac{x}{2} + \frac{\pi}{2}\right) = \frac{\sqrt{3}}{2}, -\pi \le x \le 4\pi$.

Let $\frac{x}{2} + \frac{\pi}{2} = \theta \Rightarrow \cos\theta = \frac{\sqrt{3}}{2}$

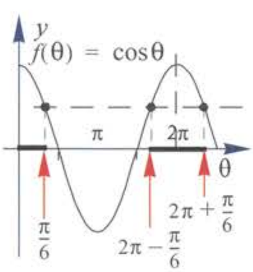
Next, $\frac{x}{2} + \frac{\pi}{2} = \theta \Leftrightarrow \frac{x}{2} = \theta - \frac{\pi}{2} \Leftrightarrow x = 2\theta - \pi$.

(Obtain x in terms of q)

New domain:
$-\pi \le x \le 4\pi \Leftrightarrow -\pi \le 2\theta - \pi \le 4\pi \Leftrightarrow 0 \le 2\theta \le 5\pi$

$$\Leftrightarrow 0 \le \theta \le \frac{5\pi}{2}$$

Therefore, our equivalent statement is.
$\cos\theta = \frac{\sqrt{3}}{2}, 0 \le \theta \le \frac{5\pi}{2}$



$$\therefore \theta = \frac{\pi}{6}, 2\pi - \frac{\pi}{6}, 2\pi + \frac{\pi}{6}$$

$$= \frac{\pi}{6}, \frac{11\pi}{6}, \frac{13\pi}{6}$$

But we still need to find the x-values:

Therefore, substituting $\theta = \frac{x}{2} + \frac{\pi}{2}$ back into the solution set, we have:

$$\frac{x}{2} + \frac{\pi}{2} = \frac{\pi}{6}, \frac{11\pi}{6}, \frac{13\pi}{6}$$
$$\Leftrightarrow x + \pi = \frac{\pi}{3}, \frac{11\pi}{3}, \frac{13\pi}{3}$$
$$\Leftrightarrow x = -\frac{2\pi}{3}, \frac{8\pi}{3}, \frac{10\pi}{3}$$

And, we notice that all solutions lie within the **original** specified domain, $-\pi \le x \le 4\pi$.

Example 3.6.5

Solve the following, for $0 \le x \le 2\pi$.

a $2\sin x = 3\cos x$ b $2\sin 2x = 3\cos x$
c $\sin x = \cos 2x$

a

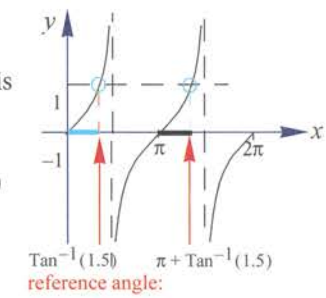
$2\sin x = 3\cos x \Leftrightarrow \frac{\sin x}{\cos x} = \frac{3}{2}, 0 \le x \le 2\pi$

$\Leftrightarrow \tan x = 1.5, 0 \le x \le 2\pi$

The reference angle is $\text{Tan}^{-1}(1.5) \approx 0.9828$

$x = \text{Tan}^{-1}(1.5), \pi + \text{Tan}^{-1}(1.5)$

$\approx 0.9828, 4.1244$



b In this case we make use of the double-angle identity, $\sin 2x = 2\sin x\cos x$.

$$\therefore 2\sin 2x = 3\cos x \Leftrightarrow 2(2\sin x\cos x) = 3\cos x$$
$$\Leftrightarrow 4\sin x\cos x - 3\cos x = 0$$
$$\Leftrightarrow \cos x(4\sin x - 3) = 0$$
$$\cos x = 0, \sin x = \frac{3}{4}$$

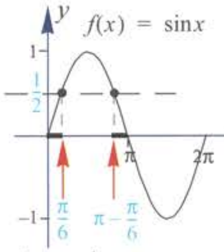
Solving for $\cos x = 0 : x = \frac{\pi}{2}, \frac{3\pi}{2}$.

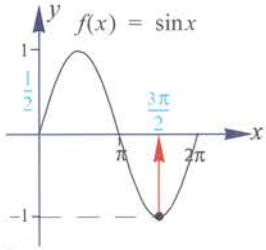
Solving for $\sin x = \frac{3}{4} : x = \text{Sin}^{-1}\left(\frac{3}{4}\right), \pi - \text{Sin}^{-1}\left(\frac{3}{4}\right)$

We solved two separate equations, giving the solution, $x = 0, \text{Sin}^{-1}\left(\frac{3}{4}\right), \frac{\pi}{2}, \pi - \text{Sin}^{-1}\left(\frac{3}{4}\right), \frac{3\pi}{2}$

c This time we make use of the cosine double-angle formula, $\cos 2x = 1 - 2\sin^2 x$.

$$\begin{aligned} \therefore\ \sin x &= \cos 2x \\ \Leftrightarrow \sin x &= 1 - 2\sin^2 x \\ \Leftrightarrow 2\sin^2 x + \sin x - 1 &= 0 \\ \Leftrightarrow (2\sin x - 1)(\sin x + 1) &= 0 \\ \Leftrightarrow \sin x &= \frac{1}{2} \text{ or } \sin x = -1 \end{aligned}$$





Again, we have two equations to solve.

Solving for $\sin x = \frac{1}{2}$: Solving for $\sin x = -1$:

Therefore, solution set is $x = \frac{\pi}{6}, \frac{5\pi}{6}, \frac{3\pi}{2}$.

As the next example shows, the working required to solve trigonometric equations can be significantly reduced, especially if you know the exact values for the basic trigonometric angles as well as the symmetry properties (without making use of a graph). However, we encourage you to use a visual aid when solving such equations.

Example 3.6.6

Solve the equation $4\sin x\cos x = \sqrt{3}, -2\pi \le x \le 2\pi$.

$$\begin{aligned} 4\sin x\cos x &= \sqrt{3}, -2\pi \le x \le 2\pi \\ 2\sin 2x &= \sqrt{3} \quad \text{using } \sin 2\theta \equiv 2\sin\theta\cos\theta \\ \sin 2x &= \frac{\sqrt{3}}{2} \\ 2x &= \ldots \frac{-11\pi}{3}, \frac{-10\pi}{3}, \frac{-5\pi}{3}, \frac{-4\pi}{3}, \frac{\pi}{3}, \frac{2\pi}{3}, \frac{7\pi}{3}, \frac{8\pi}{3}, \ldots \\ x &= \frac{-11\pi}{6}, \frac{-5\pi}{3}, \frac{-5\pi}{6}, \frac{-2\pi}{3}, \frac{\pi}{6}, \frac{\pi}{3}, \frac{7\pi}{6}, \frac{4\pi}{3} \end{aligned}$$

Exercise 3.6.1

1. If $0 \le x \le 2\pi$, find:

   a $\sin x = \frac{1}{\sqrt{2}}$ b $\sin x = -\frac{1}{2}$

   c $\sin x = \frac{\sqrt{3}}{2}$ d $\sin 3x = \frac{1}{2}$

   e $\sin\left(\frac{x}{2}\right) = \frac{1}{2}$ f $\sin(\pi x) = -\frac{\sqrt{2}}{2}$

2. If $0 \le x \le 2\pi$, find:

   a $\cos x = \frac{1}{\sqrt{2}}$ b $\cos x = -\frac{1}{2}$

   c $\cos x = \frac{\sqrt{3}}{2}$ d $\cos\left(\frac{x}{3}\right) = \frac{1}{2}$

   e $\cos(2x) = \frac{1}{2}$ f $\cos\left(\frac{\pi}{2}x\right) = -\frac{\sqrt{2}}{2}$

3. If $0 \le x \le 2\pi$, find:

   a $\tan x = \frac{1}{\sqrt{3}}$ b $\tan x = -1$

   c $\tan x = \sqrt{3}$ d $\tan\left(\frac{x}{4}\right) = 2$

   e $\tan(2x) = -\sqrt{3}$ f $\tan\left(\frac{\pi}{4}x\right) = -1$

4. If $0 \le x \le 2\pi$ or $0 \le x \le 360$, find:

   a $\sin(x° + 60°) = \frac{1}{2}$

   b $\cos(x° - 30°) = -\frac{\sqrt{3}}{2}$

   c $\tan(x° + 45°) = -1$

   d $\sin(x° - 20°) = \frac{1}{\sqrt{2}}$

   e $\cos\left(2x - \frac{\pi}{2}\right) = \frac{1}{2}$

   f $\tan\left(\frac{\pi}{4} - x\right) = 1$

   g $\sec(2x + \pi) = 2$

   h $\cot\left(2x + \frac{\pi}{2}\right) = 1$

5. If $0 \le x \le 2\pi$ or $0 \le x \le 360$, find:

   a $\cos x° = \frac{1}{2}$

   b $2\sin x + \sqrt{3} = 0$

   c $\sqrt{3}\tan x = 1$

   d $5\sin x° = 2$

e $4\sin^2 x - 3 = 0$

f $\frac{1}{\sqrt{3}}\tan x + 1 = 0$

g $2\sin\left(x + \frac{\pi}{3}\right) = -1$

h $5\cos(x+2) - 3 = 0$

i $\tan\left(x - \frac{\pi}{6}\right) = \frac{1}{\sqrt{3}}$

j $2\cos 2x + 1 = 0$

k $\tan 2x - \sqrt{3} = 0$

l $2\sin x^\circ = 5\cos x^\circ$

m $2\operatorname{cosec}\left(\frac{x}{2}\right) = 4$

n $\frac{1}{2}\cot(2x) = 0$

o $\sec\left(\frac{x}{3}\right) = -\sqrt{2}$

6. Solve the following equations for the intervals indicated, giving exact answers:

a $\sin\theta\cos\theta = \frac{1}{2}, -\pi \le \theta \le \pi$

b $\cos^2\theta - \sin^2\theta = -\frac{1}{2}, -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$

c $\tan A = \frac{1 - \tan^2 A}{2}, -\pi \le A \le \pi$

d $\frac{\sin\theta}{1 + \cos\theta} = -1, -\pi \le \theta \le \pi$

e $\cos^2 x = 2\cos x, -\pi \le x \le \pi$

f $\sec 2x = \sqrt{2}, 0 \le x \le 2\pi$

7. Solve each of the following equations in the specified domain.

a $2\sin^2 x + \sin x - 1 = 0, 0 \le x \le 4\pi$

b $\cos^2 x + 2\cos x - 3 = 0, -2\pi \le x \le 2\pi$

c $2\sin^2 x - \sin x = 3, -\pi \le x \le 5\pi$

d $2\sin^2 x - 5\sin x + 2 = 0, 0 \le x \le 6\pi$

e $6\cos^2 x + 5\cos x + 1 = 0, -2\pi \le x \le 4\pi$

Extra questions

Applications

Functions of the type considered in the previous section are useful for modelling periodic phenomena.

The pitching and rolling of a ship in a heavy sea.

The flexing of large structures such as suspension bridges.

The vibration of aircraft wings.

These sorts of applications usually start with data that has been measured in an experiment. The next task is to find a function that 'models' the data in the sense that it produces function values that are similar to the experimental data. Once this has been done, the function can be used to predict values that are missing from the measured data (interpolation) or values that lie outside the experimental data set (extrapolation).

Example 3.6.7

The table shows the depth of water at the end of a pier at various times (measured, in hours after midnight on the first day of the month.)

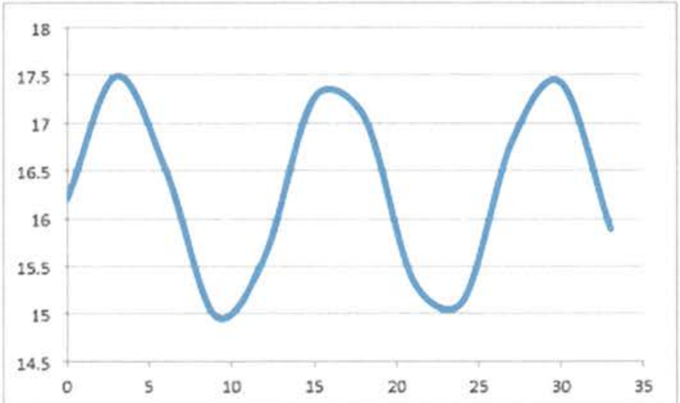
t (hr)	0	3	6	9	12	15	18
d m	16.20	17.49	16.51	14.98	15.60	17.27	17.06

t (hr)	21	24	27	30	33
d m	15.34	15.13	16.80	17.42	15.89

Use your model to predict the time of the next high tide.

Plot the data as a graph. Use your results to find a rule that models the depth data.

We start by entering the data as lists and then plotting them using graph paper or a graph plotter. This graph was produced using a spreadsheet with 'scatter/smooth line' selected.

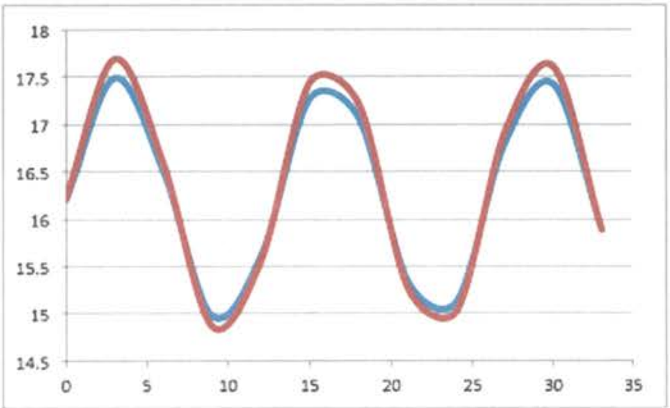


This does suggest that the depth is varying periodically. It appears that the period is approximately 13 hours. This is found by looking at the time between successive high tides. This is not as easy as it sounds as the measurements do not appear to have been made exactly at the high tides. This means that an estimate will need to be made based upon the observation that successive high tides appear to have happened after 3, 16 and 32 hours. Next, we look at the amplitude and vertical translation. Again, because we do not have exact readings at high and low tides, these will need to be estimated. The lowest tide recorded is 14.98 and the highest is 17.49.

A first estimate of the vertical translation is $\frac{17.49 + 14.98}{2} = 16.235$ and the amplitude is

$17.7 - 16.235 = 1.465$. Since the graph starts near the mean depth and moves up it seems likely that the first model to try might be:

$y = 1.465 \times \sin\left(\frac{2\pi t}{13}\right) + 16.235$



The modelling function is shown in red.

Notice that the dilation factor (along the x-axis) is found by using the result that if

$$\tau = 13 \Rightarrow \frac{2\pi}{n} = 13 \therefore n = \frac{2\pi}{13}.$$

The model should now be 'evaluated' which means testing how well it fits the data. This can be done by making tables of values of the data and the values predicted by the model and working to make the differences between these as small as possible. This can be done using a scientific or graphics calculator.

The model shown is quite good as errors are small with some being positive and some being negative. The function used is $y = 1.465 \times \sin\left(\frac{2\pi t}{13}\right) + 16.235$ and this can now be used to predict the depth for times that measurements were not made. Also, the graph of the modelling function can be added to the graph of the data (as shown).

The modelling function can also be used to predict depths into the future (extrapolation).

The next high tide, for example can be expected to be 13 hours after the previous high tide at about 29.3 hours. This is after 42.3 hours.

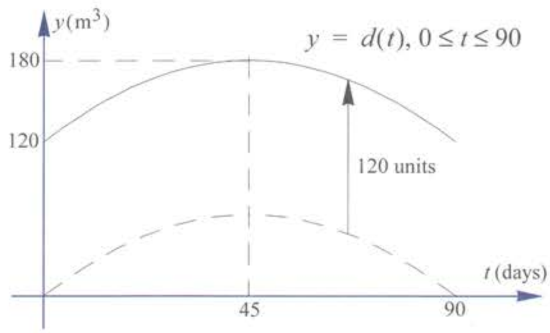
Example 3.6.8

A reservoir supplies water to an outer suburb based on the water demand, $D(t) = 120 + 60\sin\left(\frac{\pi}{90}t\right), 0 \le t \le 90$, where t measures the number of days from the start of summer (which lasts for 90 days).

Sketch the graph of $D(t)$.

What are the maximum and minimum demands made by the community over this period?

a



The features of this function are:

Period $= \dfrac{2\pi}{\left(\frac{\pi}{90}\right)} = 180$ days

Amplitude $= 60$

Translation $= 120$ units up.

We 'pencil in' the graph of $y = 60\sin\left(\frac{\pi}{90}t\right)$ and then move it up 120 units:

b The minimum is 120m³ and the maximum is 180m³.

Example 3.6.9

When a person is at rest, the blood pressure, P millimetres of mercury at any time t seconds, can be approximately modelled by the equation:

$$P(t) = -20\cos\left(\frac{5\pi}{3}t\right) + 100, t \ge 0$$

a Determine the amplitude and period of P.

b What is the maximum blood pressure reading that can be recorded for this person?

c Sketch the graph of $P(t)$, showing one full cycle.

d Find the first two times when the pressure reaches a reading of 110 mmHg.

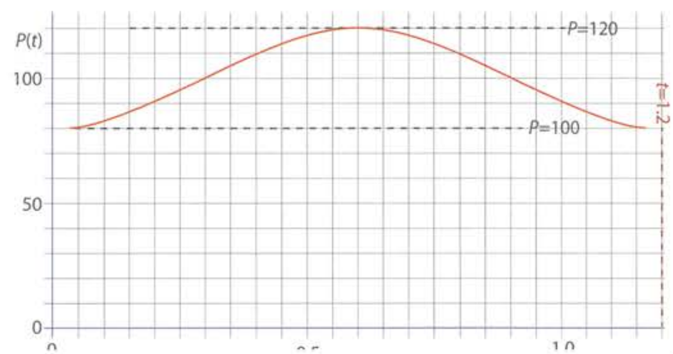
a The amplitude is 20 mmHg and the period is given by:

$$\frac{2\pi}{\left(\frac{5\pi}{3}\right)} = \frac{6}{5} = 1.2 \text{ seconds.}$$

b The maximum is given by (100 + amplitude) = 100 + 20 =120.

c One full cycle is 1.2 seconds long:

d



Note that the graph has been drawn as opposed to sketched. That is, it has been accurately sketched, meaning that the scales and the curve are accurate. Because of this we can read directly from the graph.

In this case, $P = 110$ when $t = 0.4$ and 0.8.

Even though we have drawn the graph, we will now solve the relevant equation:

$$P(t) = 110 \Leftrightarrow 110 = -20\cos\left(\frac{5\pi}{3}t\right) + 100$$

$$\Leftrightarrow 10 = 20\cos\left(\frac{5\pi}{3}t\right)$$

$$\Leftrightarrow \cos\left(\frac{5\pi}{3}t\right) = -\frac{1}{2}$$

$$\therefore \frac{5\pi}{3}t = \pi - \text{Cos}^{-1}\left(\frac{1}{2}\right), \pi + \text{Cos}^{-1}\left(\frac{1}{2}\right)$$

$$\Leftrightarrow \frac{5\pi}{3}t = \frac{2\pi}{3}, \frac{4\pi}{3}$$

$$\Leftrightarrow t = \frac{2}{5}, \frac{4}{5}$$

Exercise 3.6.2

1. The table shows the temperature in an office block over a 36-hour period.

t (hr)	0	3	6	9	12	15	18
T °C	18.3	15.0	14.1	16.0	19.7	23.0	23.9

t (hr)	21	24	27	30	33	36
T °C	22.0	18.3	15.0	14.1	16.0	19.7

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

   c. Use your rule to predict the temperature after 40 hours.

2. The table shows the light level L during an experiment on dye fading.

t (hr)	0	1	2	3	4	5
L	6.6	4.0	7.0	10.0	7.5	4.1

t (hr)	6	7	8	9	10
L	6.1	9.8	8.3	4.4	5.3

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

3. The table shows the value in $s of an industrial share over a 20-month period.

Month	0	2	4	6	8	10
Value	7.0	11.5	10.8	5.6	2.1	4.3

Month	12	14	16	18	20
Value	9.7	11.9	8.4	3.2	2.5

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

4. The table shows the population (in thousands) of a species of fish in a lake over a 22-year period.

Year	0	2	4	6	8	10
Pop	11.2	12.1	13.0	12.7	11.6	11.0

Year	12	14	16	18	20	22
Pop	11.6	12.7	13.0	12.1	11.2	11.2

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

5. The table shows the average weekly sales (in thousands of $) of a small company over a 15-year period.

Time	0	1.5	3	4.5	6	7.5
Sales	3.5	4.4	7.7	8.4	5.3	3.3

Time	9	10.5	12	13.5	15
Sales	5.5	8.5	7.6	4.3	3.6

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

6. The table shows the average annual rice production, P, (in thousands of tonnes) of a province over a 10-year period.

t (yr)	0	1	2	3	4	5
P	11.0	11.6	10.7	10.5	11.5	11.3

t (yr)	6	7	8	9	10
P	10.4	11.0	11.6	10.7	10.5

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

7. The table shows the depth of water (D metres) over a 5-second period as waves pass the end of a pier.

t (sec)	0	0.5	1	1.5	2
D	11.3	10.8	10.3	10.2	10.4

t (sec)	2.5	3	3.5	4	4.5	5
D	10.9	11.4	11.7	11.8	11.5	11.0

   a. Estimate the amplitude, period, horizontal and vertical translations.

   b. Find a rule that models the data.

8. The population (in thousands) of a species of butterfly in a nature sanctuary is modelled by the function:

$$P = 3 + 2\sin\left(\frac{3\pi t}{8}\right),\ 0 \le t \le 12$$

where t is the time in weeks after scientists first started making population estimates.

a What is the initial population?

b What are the largest and smallest populations?

c When does the population first reach 4 thousand butterflies?

9. A water wave passes a fixed point. As the wave passes, the depth of the water (D metres) at time t seconds is modelled by the function:

$$D = 7 + \frac{1}{2}\cos\left(\frac{2\pi t}{5}\right), t > 0$$

a What are the greatest and smallest depths?

b Find the first two times at which the depth is 6.8 metres.

10. The weekly sales (S) (in hundreds of cans) of a soft drink outlet is modelled by the function:

$$S = 13 + 5.5\cos\left(\frac{\pi t}{6} - 3\right), t > 0$$

t is the time in months with $t = 0$ corresponding to 1st January 1990.

a Find the minimum and maximum sales during 1990.

b Find the value of t for which the sales first exceed 1500 ($S = 15$).

c During which months do the weekly sales exceed 1500 cans?

11. The rabbit population, $R(t)$ thousands, in a northern region of South Australia is modelled by the equation:

$$R(t) = 12 + 3\cos\left(\frac{\pi}{6}t\right), 0 \le t \le 24,$$

where t is measured in months after the first of January.

a What is the largest rabbit population predicted by this model?

b How long is it between the times when the population reaches consecutive peaks?

c Sketch the graph of $R(t)$ for $0 \le t \le 24$.

d Find the longest time span for which $R(t) \ge 13.5$.

e Give a possible explanation for the behaviour of this model.

12. Samantha is sitting in a rocking chair. The distance $d(t)$, in centimetres, between the wall and the rear of the chair varies sinusoidally with time t, in seconds. At time $t = 1$, the chair is closest to the wall and $d(1) = 18$. At $t = 2$, the chair is farthest from the wall and $d(2) = 34$.

a What is the period of the function?

b How far is the chair from the wall when no one is rocking in it?

c What is the domain of the function, if Samantha rocks the chair back and forth 20 times?

d What is the range of the function?

e What is the amplitude of the function?

f What is the equation of the sinusoidal function?

g What is the distance between the wall and the chair at t = 8s?

13. A weight attached to the end of a long spring is bouncing up and down. As it bounces, its distance from the floor varies sinusoidally with time. You start a stopwatch. When the stopwatch reads 0.4s, the weight first reaches a high point 72cm above the floor. The next low point, 32cm above the floor, occurs at 1.6s.

a Find a particular equation for distance from the floor as a function of time.

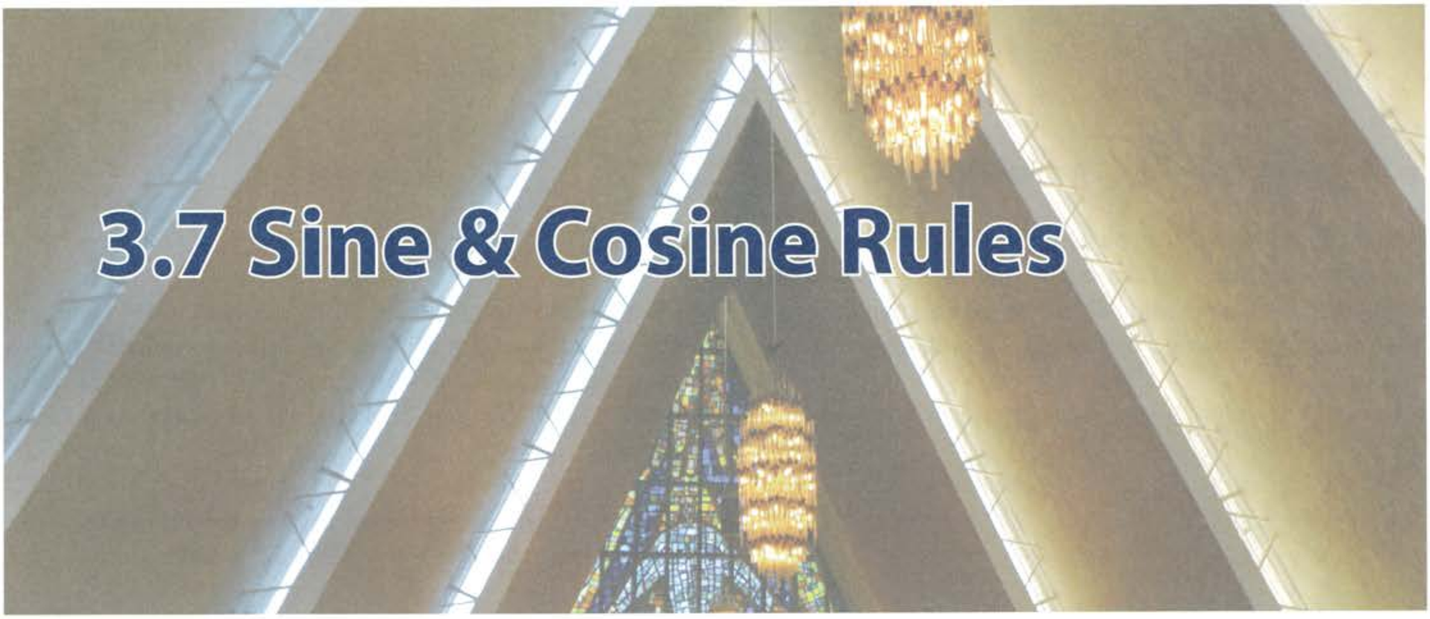
b What is the distance from the floor when the stopwatch reads 17.2s?

c What was the distance from the floor when you started the stopwatch?

d What is the first positive value of time when the weight is 59cm above the floor?

Extra questions

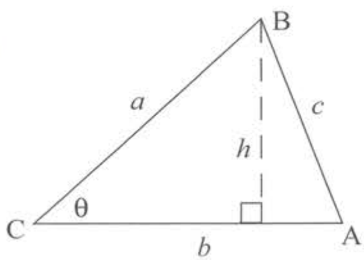
Answers

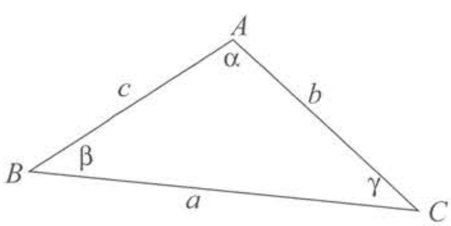


3.7 Sine & Cosine Rules

The Sine Rule

Previous sections dealt with the trigonometry of right-angled triangles. The trigonometric ratios can be used to solve non-right-angled triangles. There are two main methods for solving non-right-angled triangles, the **sine rule** and the **cosine rule** (which we look at later in this section). Both are usually stated using a standard labelling of the triangle. This uses capital letters to label the vertices and the corresponding small letters to label the sides opposite these vertices.

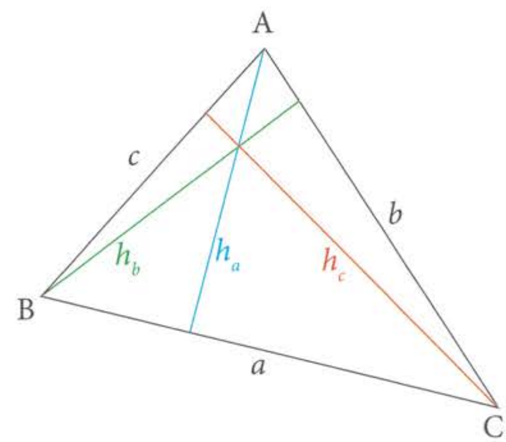




$$\sin A = \sin\alpha$$
$$\sin B = \sin\beta$$
$$\sin C = \sin\gamma$$

If we add the three altitudes to the triangle, we can calculate the area of the triangle in three ways.

The altitude from vertex A to side a is labelled h_a. These will meet in a single point, though this is not crucial to the following argument.



Area of triangle = ½ base × height. There are three versions of this:

$$\text{Area} = \frac{1}{2}\times a\times h_a = \frac{1}{2}\times a\times c\times \sin B$$
$$\text{Area} = \frac{1}{2}\times b\times h_b = \frac{1}{2}\times b\times a\times \sin C$$
$$\text{Area} = \frac{1}{2}\times c\times h_c = \frac{1}{2}\times c\times b\times \sin A$$

These must be equal so (with the usual provisos about non-zero denominators):

$$\frac{1}{2}\times a\times c\times \sin B = \frac{1}{2}\times b\times a\times \sin C = \frac{1}{2}\times c\times b\times \sin A$$
$$a\times c\times \sin B = b\times a\times \sin C = c\times b\times \sin A$$
$$\frac{a\times c\times \sin B}{a\times b\times c} = \frac{b\times a\times \sin C}{a\times b\times c} = \frac{c\times b\times \sin A}{a\times b\times c}$$
$$\frac{\sin B}{b} = \frac{\sin C}{c} = \frac{\sin A}{a}$$

The reciprocal version of this is the more usual version of the **sine rule**:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \text{ or } \frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

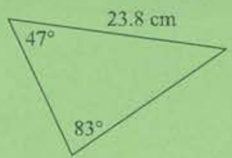
So, when should/can we make use of the sine rule?

Although the sine rule can be used for right-angled triangles, it is more often used for situations when we do not have a right-angled triangle, and when the given triangle has a known side and the angle opposite it is also known:

Example 3.7.1

Solve the following triangles giving the lengths of the sides in centimetres, correct to one decimal place and angles correct to the nearest degree.

a



b

a Firstly, label the triangle using the standard method of lettering. 'Solve the triangle' means find all the angles and the lengths of all the sides. Since two of the angles are known, the third is C = 180˚ – 47˚ – 83˚ = 50˚.

The lengths of the remaining sides can be found using the known pairing of side and angle, b and B.

$$\frac{a}{\sin A} = \frac{b}{\sin B} \Leftrightarrow \frac{a}{\sin 47°} = \frac{23.8}{\sin 83°}$$

$$a = \frac{23.8 \times \sin 47°}{\sin 83°}$$

$$= 17.5369\ldots$$

That is, BC is 17.5 cm (correct to one d.p.).

Similarly, the remaining side can be calculated:

$$\frac{c}{\sin C} = \frac{b}{\sin B} \Leftrightarrow \frac{c}{\sin 50°} = \frac{23.8}{\sin 83°}$$

$$\therefore c = \frac{23.8 \times \sin 50°}{\sin 83°}$$

$$= 18.3687\ldots$$

That is, AB is 18.4 cm (correct to one d.p.).

b This triangle is different from the previous example in that only one angle is known. It remains the case that a pair of angles and an opposite side are known and that the sine rule can be used. The angle A must be found first.

$$\frac{\sin A}{a} = \frac{\sin B}{b} \Leftrightarrow \frac{\sin A}{28.7} = \frac{\sin 42°}{92.4}$$

$$\Leftrightarrow \sin A = \frac{28.7 \times \sin 42°}{92.4}$$

$$= 0.207836$$

$$\therefore A = \sin^{-1} 0.207836$$

$$= 11.9956°$$

$$= 11°59'44''$$

The answer to the first part of the question is 12° correct to the nearest degree. It is important, however, to carry a much more accurate version of this angle through to subsequent parts of the calculation. This is best done using the calculator memory.

The third angle can be found because the sum of the three angles is 180°. So, C = 180˚ – 12˚ – 42˚ = 126˚.

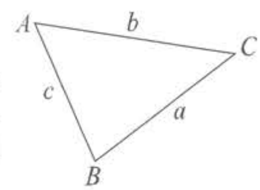
An accurate version of this angle must also be carried to the next part of the calculation. Graphics calculators have multiple memories labelled A, B, C etc. and students are advised to use these in such calculations.

The remaining side is: $\frac{c}{\sin 126°} = \frac{28.7}{\sin 12°} \Leftrightarrow c = \frac{28.7 \sin 126°}{\sin 12°}$

$$\therefore c = 111.6762\ldots$$

That is, AB is 111.7 cm (correct to one d.p.)

Exercise 3.7.1



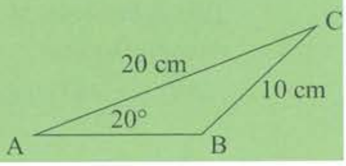
1. Use the sine rule to complete the following table, which refers to the standard labelling of a triangle.

	a cm	b cm	c cm	A	B	C
1			48.2		29°	141°
2		1.2		74°	25°	
3			11.3	60°		117°
4			51.7	38°		93°
5	18.5	11.4		68°		
6	14.6	15.0			84°	
7		7.3			16°	85°
8			28.5	39°		124°
9	0.8		0.8	82°		
10			33.3	36°		135°
11	16.4			52°	84°	
12			64.3		24°	145°
13	30.9	27.7		75°		
14			59.1	29°		102°

	a cm	b cm	c cm	A	B	C
15		9.8	7.9		67°	
16			54.2	16°		136°
17	14.8		27.2			67°
18			10.9		3°	125°
19			17.0		15°	140°
20			40.1	30°		129°

Example 3.7.2

For the triangle shown, find the angle ABC.



Making use of the sine rule we have:

$$\frac{\sin A}{a} = \frac{\sin B}{b} \Leftrightarrow \frac{\sin 20°}{10} = \frac{\sin B}{20}$$

$$\Leftrightarrow \sin B = \frac{20\sin 20°}{10}$$

$$\therefore B = \sin^{-1}(2\sin 20°)$$

$$= 43.1601\ldots$$

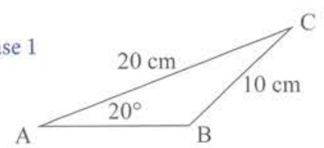
That is, $B = 43°10'$

However, from our diagram, the angle ABC should have been greater than 90°! That is, we should have obtained an **obtuse angle** (90° < B < 180°) rather than an **acute angle** (0° < B < 90°).

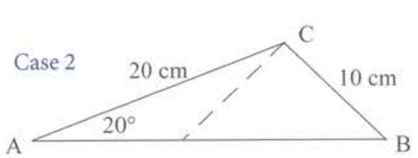
So, what went wrong?

This example is a classic case of what is known as the **ambiguous case**, in that, from the given information it is possible to draw two different diagrams, both having the same data. We show both these triangles:

Case 1



Case 2



Notice that the side BC can be pivoted about the point C and therefore two different triangles can be formed with BC = 10. This is why there are two possible triangles based on the same information. In the solution above, B = 43° 10' – representing Case 2. However, our diagram is represented by Case 1! Therefore, the correct answer is 180° – 43° 10' = 136° 50'.

The Ambiguous Case

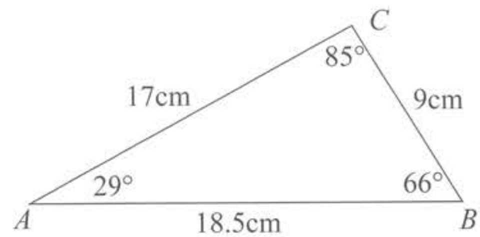
From Example 3.7.2, it can be seen that an ambiguous (having a 'double meaning') case can arise when using the sine rule. In the given situation we see that the side CB can be pivoted about its vertex, forming two possible triangles. We consider another such triangle in the next example.

Example 3.7.3

Draw diagrams showing the triangles in which AC = 17 cm, BC = 9 cm and A = 29° and solve these triangles.

Applying the sine rule to the triangle gives:

$$\frac{\sin B}{17} = \frac{\sin 29°}{9} \Leftrightarrow \sin B = \frac{17 \times \sin 29°}{9}$$



$$= 0.91575$$

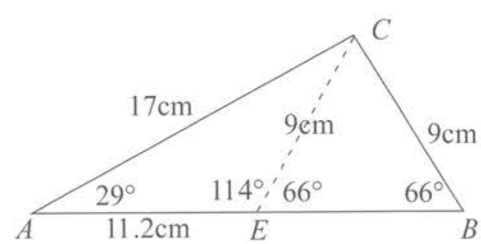
$$\therefore B = 66°$$

Next, we have,

$$C = 180° - 29° - 66° = 85°$$

$$\frac{c}{\sin 85°} = \frac{9}{\sin 29°} \Leftrightarrow c = 18.5$$

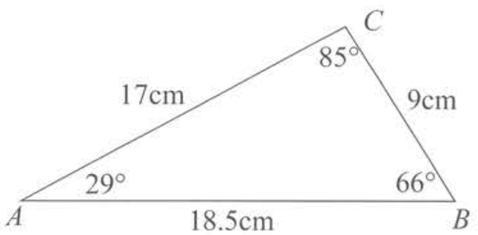
There is, however, a second solution that results from drawing an isosceles triangle BCE.

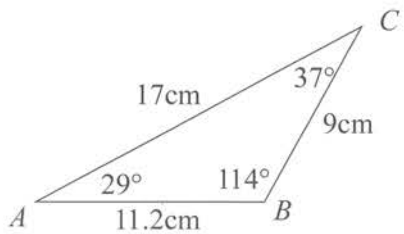


This creates the triangle AEC which also fits the data. The third angle of this triangle is 37° and the third side is:

$$\frac{\text{AE}}{\sin 37°} = \frac{9}{\sin 29°} \Leftrightarrow \text{AE} = 11.2$$

The original data is ambiguous in the sense that there are two triangles that are consistent with it.





You should also notice that the two angles in the solution are 66° and 114° and that sin66° = sin114°. (That is, sin 66° = sin(180° – 66°) = sin 114°.

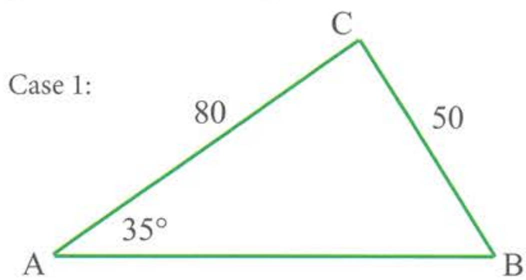
Example 3.7.4

Find ∠ABC for the triangle ABC given that $a = 50$, $b = 80$ and A = 35°.

We first determine the value of $b\sin\alpha$ and compare it with the value a:

Now, $b\sin\alpha = 80\sin 35° = 45.89$

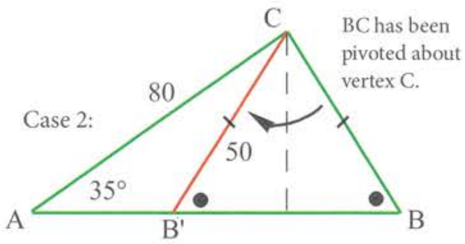
Therefore we have that $b\sin\alpha$ (= 45.89) $< a$ (= 50) $< b$ (= 80) meaning that we have an ambiguous case.



Using the sine rule, $\frac{\sin A}{a} = \frac{\sin B}{b}$, we have

$$\frac{\sin 35°}{50} = \frac{\sin B}{80} \Leftrightarrow \sin B = \frac{80\sin 35°}{50}$$

$$\therefore B = 66°35'$$



From case 1, the obtuse angle B' is given by 180° – 66°35' = 113°25'.

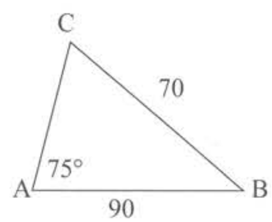
This is because $\Delta B'CB$ is an **isosceles** triangle, so that

$\angle AB'C = 180° - \angle CB'B$

Example 3.7.5

Find $\angle ACB$ for the triangle ABC given that $a = 70$, $c = 90$ and $A = 75°$.

We start by drawing the triangle with the given information:



Using the sine rule we have:

$$\frac{\sin C}{90} = \frac{\sin 75}{70} \Leftrightarrow \sin C = \frac{90\sin 75}{70}$$

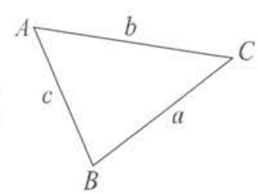
$$\therefore \sin C = 1.241\ldots$$

Which is impossible to solve for as the sine of an angle can never be greater than one.

Therefore no such triangle exists.

Exercise 3.7.2

Find the two solutions to these triangles which are defined using the standard labelling:



	a cm	b cm	A
1	7.4	18.1	20°
2	13.3	19.5	14°
3	13.5	17	28°
4	10.2	17	15°
5	7.4	15.2	20°

	a cm	b cm	A
6	10.7	14.1	26°
7	11.5	12.6	17°
8	8.3	13.7	24°
9	13.7	17.8	14°
10	13.4	17.8	28°
11	12.1	16.8	23°
12	12	14.5	21°
13	12.1	19.2	16°
14	7.2	13.1	15°
15	12.2	17.7	30°
16	9.2	20.9	14°
17	10.5	13.3	20°
18	9.2	19.2	15°
19	7.2	13.3	19°
20	13.5	20.4	31°

2. Solve the following triangles.

a $\alpha = 75°, a = 35, c = 45$

b $\alpha = 35°, a = 30, b = 80$

c $\beta = 40°, a = 22, b = 8$

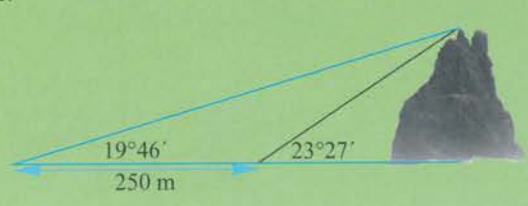
d $\gamma = 50°, a = 112, c = 80$

Applications of the sine rule

Just as in the case of right-angled triangles, the sine rule becomes very useful. In particular, it means that previous problems that required the partitioning of a non-right-angled triangle into two (or more) right-angled triangles can be solved using the sine rule.

Example 3.7.6

A surveying team are trying to find the height of a hill. They take a 'sight' on the top of the hill and find that the angle of elevation is 23°27′. They move a distance of 250 metres on level ground directly away from the hill and take a second 'sight'. From this point, the angle of elevation is 19°46′. Find the height of the hill, correct to the nearest metre.



Labelling the given diagram using the standard notation we have:

With $\beta = 180 - 23°27' = 156°33'$

and $\gamma = 180 - 19°46' - 156°33' = 3°41'$

Then, using the sine rule,

$$\frac{b}{\sin 156°33'} = \frac{250}{\sin 3°41'}$$

$$\Leftrightarrow b = \frac{250 \sin 156°33'}{\sin 3°41'}$$

$$= 1548.63\ldots$$

Then, using ΔACP we have,

$$\sin 19°46' = \frac{h}{b} \Leftrightarrow h = b \sin 19°46'$$

$$= 523.73$$

So, the hill is 524 m high (to nearest metre).

Exercise 3.7.3

1. A short course biathlon meet requires the competitors to run in the direction S60°W to their bikes and then ride S40°E to the finish line, situated 20 km due south of the starting point. What is the distance of this course?

2. A pole is slanting towards the sun and is making an angle of 10° to the vertical. It casts a shadow 7 metres long along the horizontal ground. The angle of elevation of the top of the pole to the tip of its shadow is 30°. Find the length of the pole, giving your answer to 2 d.p.

3. A statue A, is observed from two other statues B and C which are 330 m apart. The angle between the lines of sight AB and BC is 63° and the angle between the lines of sight AC and CB is 75°. How far is statue A from statue B?

4. Town A is 12 km from town B and its bearing is 132°T from B. Town C is 17 km from A and its bearing is 063°T from B. Find the bearing of A from C.

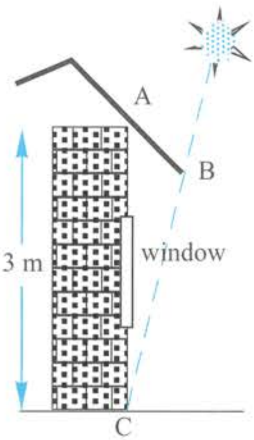
5. The angle of elevation of the top of a building from a park bench on level ground is 18°. The angle of elevation from a second park bench, 300 m closer to the base of the building is 30°. Assuming that the two benches and the building all lie on the same vertical plane, find the height of the building.

a A man standing 6 m away from a lamp post casts a shadow 10 m long on a horizontal ground. The angle of elevation from the tip of the shadow to the lamp light is 12°. How high is the lamp light?

b If the shadow is cast onto a road sloping at 30° upwards, how long would the shadow be if the man is standing at the foot of the sloping road and 6 metres from the lamp post?

6. At noon the angle of elevation of the sun is 72° and is such that a three metre wall AC, facing the sun, is just in the shadow due to the overhang AB. The angle that the overhang makes with the vertical wall is 50°.

Copy and illustrate this information on the diagram shown.



a Find the length of the overhang.

b At 4 p.m. the angle of elevation of the sun is 40° and the shadow due to the overhang just reaches the base of the window.

c How far from the ground is the window?

7. The lookout on a ship sailing due east at 25 km/h observes a reef N62°E at a distance of 30 km.

a How long will it be before the ship is 15 km from the reef, assuming that it continues on its easterly course.

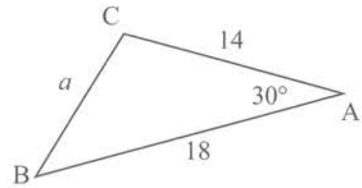
b How long is it before it is again 15 km from the reef?

c What is the closest that the ship will get to the reef?

Extra questions

The Cosine Rule

Sometimes the sine rule is not enough to help us solve for a non right-angled triangle. For example, in the triangle shown, we do not have enough information to use the sine rule. That is, the sine rule only provided the following:



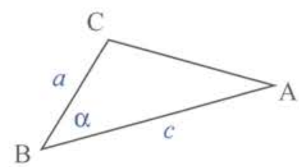
$$\frac{a}{\sin 30°} = \frac{14}{\sin B} = \frac{18}{\sin C}$$

where there are too many unknowns.

For this reason we derive another useful result, known as the cosine rule. The cosine rule may be used when

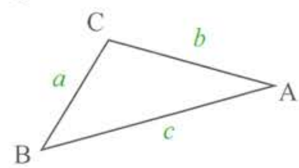
1. **two sides and an included angle are given:**

   This means that the third side can be determined and then we can make use of the sine rule (or the cosine rule again).



2. **three sides are given:**

   This means we could then determine any of the angles.

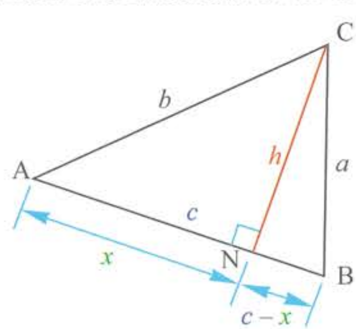


The cosine rule, with the standard labelling of the triangle has three versions:

$$a^2 = b^2 + c^2 - 2bc\cos A$$
$$b^2 = a^2 + c^2 - 2ac\cos B$$
$$c^2 = a^2 + b^2 - 2ab\cos C$$

The cosine rule can be remembered as a version of Pythagoras's Theorem with a correction factor. We now show why this works.

Consider the case where there is an acute angle at A. Draw a perpendicular from C to N as shown in the diagram.



In ΔANC we have $\quad b^2 = h^2 + x^2$

$$\Leftrightarrow h^2 = b^2 - x^2 \quad -(1)$$

In ΔBNC we have $\quad a^2 = h^2 + (c-x)^2$

$$\Leftrightarrow h^2 = a^2 - (c-x)^2 \quad -(2)$$

Equating (1) and (2) we have,

$$a^2 - (c-x)^2 = b^2 - x^2$$
$$\Leftrightarrow a^2 - c^2 + 2cx - x^2 = b^2 - x^2$$
$$\Leftrightarrow a^2 = b^2 + c^2 - 2cx$$

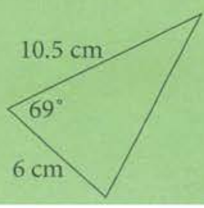
However, from ΔANC: $\cos A = \frac{x}{b} \Leftrightarrow x = b\cos A$

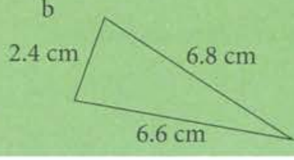
Substituting this result for x: $\boxed{a^2 = b^2 + c^2 - 2bc\cos A}$

Although we have shown the result for an acute angle at A, the same rule applies if A is obtuse.

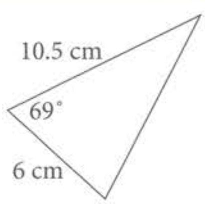
Example 3.7.7

Solve the following triangles giving the lengths of the sides in centimetres, correct to one decimal place and angles correct to the nearest degree.

a


b


a



The data does not include an angle and the opposite side so the sine rule cannot be used. The first step, as with the sine rule, is to label the sides of the triangle. Once the triangle has been labelled, the correct version of the cosine rule must be chosen. In this case, the solution is:

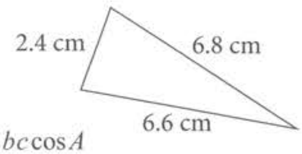
$$c^2 = a^2 + b^2 - 2ab\cos C$$
$$c^2 = 10.5^2 + 6^2 - 2 \times 10.5 \times 6 \times \cos 69°$$
$$= 101.0956$$
$$a = 10.1$$

The remaining angles can be calculated using the sine rule. Again, it is important to carry a high accuracy for the value of c to the remaining problem:

$$\frac{\sin B}{b} = \frac{\sin C}{c} \Leftrightarrow \sin B = \frac{6 \times \sin 69°}{10.0546} \quad \therefore B = 34°$$

Finally, $A = 180° - 34° - 69° = 77°$

b In this case, there are no angles given. The cosine rule can be used to solve this problem as follows:



$$a^2 = b^2 + c^2 - 2bc\cos A$$
$$6.6^2 = 2.4^2 + 6.8^2 - 2 \times 2.4 \times 6.8 \times \cos A$$
$$2 \times 2.4 \times 6.8 \times \cos A = 2.4^2 + 6.8^2 - 6.6^2$$
$$\cos A = \frac{2.4^2 + 6.8^2 - 6.6^2}{2 \times 2.4 \times 6.8}$$
$$= 0.25858$$
$$A = 75.014°$$
$$= 75°1'$$

Next, use the sine rule:

$$\frac{\sin B}{b} = \frac{\sin A}{a} \Leftrightarrow \sin B = \frac{2.4 \times \sin 75}{6.6} \quad \therefore B = 20°34'$$

So that $C = 180° - 75° - 21° = 84°$

The three angles, correct to the nearest degree are $A = 75°$, $B = 21°$ & $C = 84°$.

Exercise 3.7.4

Solve the following triangles.

	a cm	*b* cm	*c* cm	*A*	*B*	*C*
1	13.5		16.7		36°	
2	8.9	10.8				101°
3	22.8		12.8		87°	
4	21.1	4.4				83°
5		10.6	15.1	74°		
6		13.6	20.3	20°		
7	9.2		13.2		46°	
8	23.4	62.5				69°
9		9.6	15.7	41°		
10	21.7	36.0	36.2			
11	7.6	3.4	9.4			
12	7.2	15.2	14.3			
13	9.1		15.8		52°	
14	14.9	11.2	16.2	63°	42°	75°
15	2.0	0.7	2.5			
16	7.6	3.7	9.0			
17	18.5	9.8	24.1			
18	20.7	16.3	13.6			

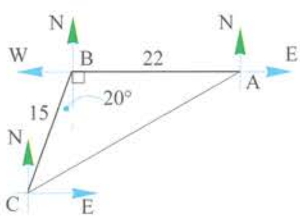
	a cm	b cm	c cm	A	B	C
19		22.4	29.9	28°		
20	7.0		9.9		42°	
21	21.8	20.8	23.8			
22	1.1		1.3		89°	
23		1.2	0.4	85°		
24	23.7	27.2				71°
25	3.4	4.6	5.2			

Applications of the Cosine Rule

Example 3.7.8

A cyclist rode her bike for 22 km on a straight road heading in a westerly direction towards a junction. Upon reaching the junction, she headed down another straight road bearing 200°T for a distance of 15 km. How far is the cyclist from her starting position?

We start with a diagram:



Note that:

$\angle ABC = 90° + 20° = 110°$

Using the cosine rule we have,

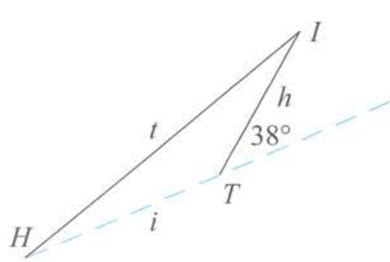
$$AC^2 = 15^2 + 22^2 - 2 \times 15 \times 22 \cos 110°$$
$$\Rightarrow AC = \sqrt{225 + 484 - 660 \times (-0.3420\ldots)}$$
$$\therefore AC = 30.5734\ldots$$

That is, she is (approximately) 30.57 km from her starting point.

Example 3.7.9

A yacht starts from a harbour and sails for a distance of 11 km in a straight line. The yacht then makes a turn to port (left) of 38° and sails for 7 km in a straight line in this new direction until it arrives at a small island. Draw a diagram that shows the path taken by the yacht and calculate the distance from the harbour to the island.

The question does not give the bearing of the first leg of the trip so the diagram can show this in any direction. H is the harbour, I the island and T the point where the yacht makes its turn.



The angle in the triangle at T is $180° - 38° = 142°$.

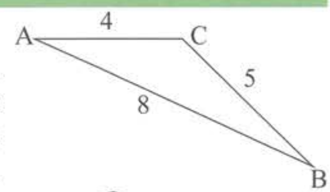
The problem does not contain an angle and the opposite side and so must be solved using the cosine rule.

$$t^2 = h^2 + i^2 - 2hi\cos T$$
$$= 7^2 + 11^2 - 2 \times 7 \times 11 \times \cos 142°$$
$$= 291.354$$
$$\therefore t = 17.1$$

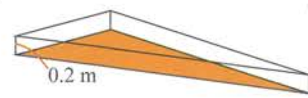
That is, distance from the harbour to the port is 17.1 km (to one d.p)

Example 3.7.10

A triangular sandpit having side lengths 5 m, 4 m and 8 m is to be constructed to a depth of 20 cm. Find the volume of sand required to fill this sandpit.



We will need to find an angle. In this case we determine the largest angle, which will be the angle opposite the longest side.



From our diagram we have

$$8^2 = 4^2 + 5^2 - 2 \times 4 \times 5 \cos C$$
$$\therefore 64 = 16 + 25 - 40 \cos C$$
$$\Leftrightarrow \cos C = \frac{16 + 25 - 64}{40}$$
$$= -\frac{23}{40}$$
$$\therefore C = 125°6'$$

To find the volume of sand we first need to find the surface area of the sandpit.

Area $= \frac{1}{2}ab\sin C = \frac{1}{2} \times 4 \times 5 \times \sin(125°6') = 8.1815 \text{ m}^2$.

The volume of sand required is $0.2 \times 8.1815 = 1.64 \text{ m}^3$.

Exercise 3.7.5

1. Thomas has just walked 5 km in a direction N70°E when he realises that he needs to walk a further 8 km in a direction E60°S.

   a How far from the starting point will Thomas have travelled?

b What is his final bearing from his starting point?

2. Two poles, 8 m apart, are facing a rugby player who is 45 m from the left pole and 50 m from the right one. Find the angle that the player makes with the goal mouth.

3. The lengths of the adjacent sides of a parallelogram are 4.80 cm and 6.40 cm. If these sides have an inclusive angle of 40˚, find the length of the shorter diagonal.

4. During an orienteering venture, Patricia notices two rabbit holes and estimates them to be 50 m and 70 m away from her. She measures the angle between the line of sight of the two holes as 54°. How far apart are the two rabbit holes?

5. To measure the length of a lake, a surveyor chooses three points. Starting at one end of the lake she walks in a straight line for 223.25 m to some point X, away from the lake. She then heads towards the other end of the lake in a straight line and measures the distance covered to be 254.35 m. If the angle between the paths she takes is 82°25', find the length of the lake.

6. A light aeroplane flying N87°W for a distance of 155 km, suddenly needs to alter its course and heads S34°E for 82 km to land on an empty field.

   a How far from its starting point did the plane land.

   b What was the plane's final bearing from its starting point?

Area of a Triangle

Given **any** triangle with sides a and b, height h and included angle θ, the area, A, is given by:

$$A = \frac{1}{2}bh$$

However, $\sin\theta = \frac{h}{a} \Leftrightarrow h = a \times \sin\theta$ and so, we have that:

Area = ½ $a \times b$ sinθ

where θ is the angle between sides a and b.

Note that the triangle need not be a right-angled triangle.

Because of the standard labelling system for triangles, the term sinθ is often replaced by sin$C$, giving the expression Area = ½ $a \times b$ sin$C$.

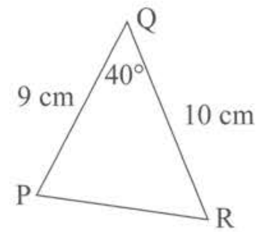
A similar argument can be used to generate the formulae:

Area = ½ $b \times c$ sin$A$ = ½ $a \times c$ sin$B$

Example 3.7.11

Find the area of the triangle PQR given that PQ = 9 cm, QR = 10 cm and ∠PQR = 40˚.

Based on the given information we can construct the following triangle:



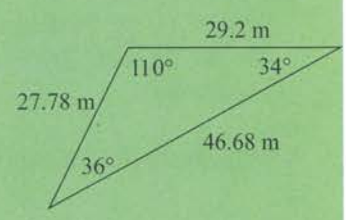
The required area, A, is given by:

$$A = \frac{1}{2}ab\sin\theta = \frac{1}{2}\times 9\times 10\times \sin 40^\circ$$
$$= 28.9$$

That is, the area is 28.9 cm².

Example 3.7.12

The diagram shows a triangular children's playground.



Find the area of the playground.

Since all the measurements of the triangle are known, any one of the three formulae could be used. Many people remember the formula as 'Area equals half the product of the lengths of two sides times the sine of the angle between them'.

$$\text{Area} = \frac{1}{2}\times 27.78\times 46.68\times \sin 36^\circ = 381\,\text{m}^2$$

$$\text{Area} = \frac{1}{2}\times 27.78\times 29.2\times \sin 110^\circ = 381\,\text{m}^2$$

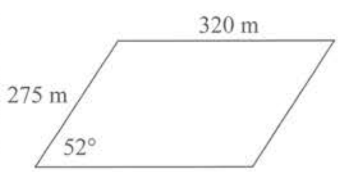
$$\text{Area} = \frac{1}{2}\times 29.2\times 46.68\times \sin 34^\circ = 381\,\text{m}^2$$

Exercise 3.7.6

1. Find the areas of these triangles that are labelled using standard notation.

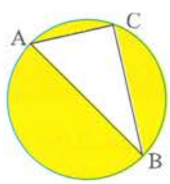
	a cm	b cm	c cm	A	B	C
a	35.94	128.46	149.70	12°	48°	120°
b	35.21	54.55	81.12	20°	32°	128°
c	46.35	170.71	186.68	14°	63°	103°
d	33.91	159.53	163.10	12°	78°	90°
e	42.98	25.07	48.61	62°	31°	87°
f	39.88	24.69	34.01	84°	38°	58°
g	43.30	30.26	64.94	34°	23°	123°
h	12.44	2.33	13.12	68°	10°	102°
i	43.17	46.44	24.15	67°	82°	31°
j	23.16	32.71	24.34	45°	87°	48°
k	50.00	52.91	38.64	64°	72°	44°
l	44.31	17.52	48.77	65°	21°	94°
m	12.68	23.49	22.34	32°	79°	69°
n	42.37	42.37	68.56	36°	36°	108°
o	40.70	15.65	41.26	77°	22°	81°

2. A car park is in the shape of a parallelogram. The lengths of the sides of the car park are given in metres.



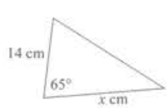
What is the area of the car park?

3. The diagram shows a circle of radius 10 cm. AB is a diameter of the circle. AC = 6 cm.



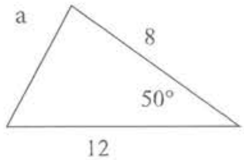
Find the area of the shaded region, giving an exact answer.

4. The triangle shown has an area of 110 cm^2. Find x.

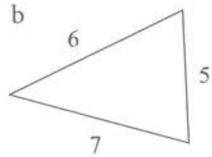


5. Find the area of the following.

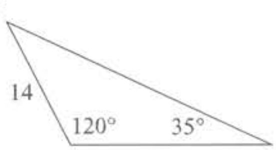
a



b

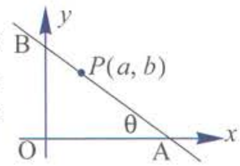


c



6. A napkin is in the shape of a quadrilateral with diagonals 9 cm and 12 cm long. The angle between the diagonals is 75°. What area does the napkin cover when laid out flat?

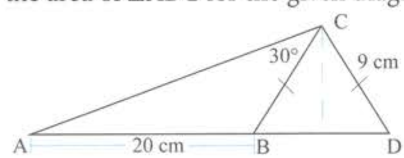
7. A triangle of area 50 cm^2 has side lengths 10 cm and 22 cm. What is the magnitude of the included angle?



8. A variable triangle OAB is formed by a straight line passing through the point $P(a,b)$ on the Cartesian plane and cutting the x-axis and y-axis at A and B respectively.

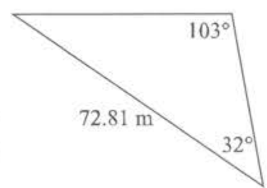
If $\angle OAB = \theta$, find the area of ΔOAB in terms of a, b and θ.

9. Find the area of ΔABC for the given diagram.



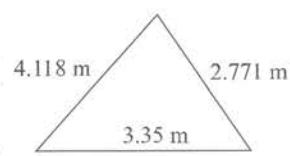
Exercise 3.7.7

1. The diagram shows a triangular building plot. The distances are given in metres. Find the length of the two remaining sides of the plot giving your answers correct to the nearest hundredth of a metre.



2. Xiang is standing on level ground. Directly in front of him and 32 metres away is a flagpole. If Xiang turns 61° to his right, he sees a post box 26.8 metres in front of him. Find the distance between the flagpole and the post box.

3. A triangular metal brace is part of the structure of a bridge. The lengths of the three parts are shown in metres. Find the angles of the brace.



4. Find the smallest angle in the triangle whose sides have length 35.6 cm, 58.43 cm and 52.23 cm.

5. Ayton is directly north of Byford. A third town, Canfield, is 9.93km from Ayton on a bearing of 128°T. The distance from Byford to Canfield is 16.49km. Find the bearing of Canfield from Byford.

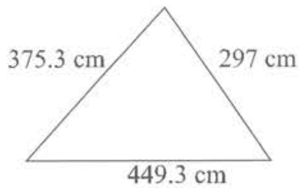
6. A parallelogram has sides of length 21.90 cm and 95.18 cm. The angle between these sides is 121°. Find the length of the long diagonal of the parallelogram.

7. A town clock has 'hands' that are of length 62cm and 85cm.

   a Find the angle between the hands at half past ten.

   b Find the distance between the tips of the hands at half past ten.

8. A shop sign is to be made in the shape of a triangle. The lengths of the edges are shown. Find the angles at the vertices of the sign.

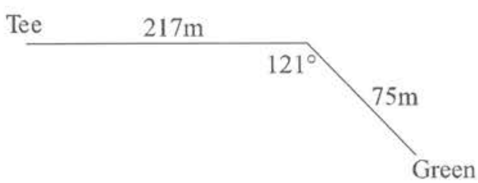


9. An aircraft takes off from an airstrip and then flies for 16.2 km on a bearing of 066°T. The pilot then makes a left turn of 88° and flies for a further 39.51 km on this course before deciding to return to the airstrip.

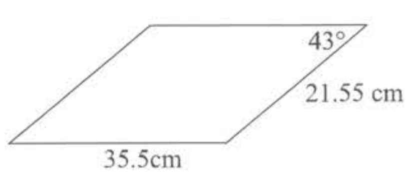
   a Through what angle must the pilot turn to return to the airstrip?

   b How far will the pilot have to fly to return to the airstrip?

10. A golfer hits two shots from the tee to the green. How far is the tee from the green?



11. The diagram shows a parallelogram. Find the length of the longer of the two diagonals.



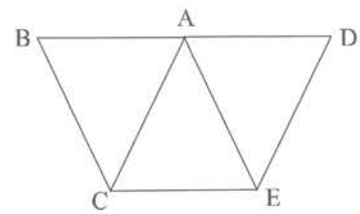
12. A triangle has angles 64°, 15° and 101°. The shortest side is 49 metres long. What is the length of the longest side?

13. The diagram shows a part of the support structure for a tower. The main parts are two identical triangles, ABC and ADE.

    AC = DE = 27.4cm and BC = AE = 23.91cm

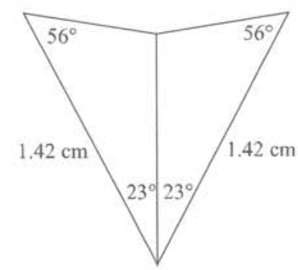
    The angles ACB and AED are 58°.

    Find the distance BD.



14. The diagram shows a design for the frame of a piece of jewellery. The frame is made of wire.

    Find the length of wire needed to make the frame.



15. A triangular cross-country running track begins with the runners running North for 2050 metres. The runners then turn right and run for 5341 metres on a bearing of 083°T. Finally, the runners make a turn to the right and run directly back to the starting point.

    a Find the length of the final leg of the run.

    b Find the total distance of the run.

    c What is the angle through which the runners must turn to start the final leg of the race?

    d Find the bearing that the runners must take on the final leg of the race.

16 Show that for any standard triangle ABC,

$$\frac{\cos A}{a} + \frac{\cos B}{b} + \frac{\cos C}{c} = \frac{a^2 + b^2 + c^2}{2abc}$$

Extra questions

Theory of Knowledge

The Rightness of Right Triangles

When you began studying trigonometry, perhaps the first lesson was an introduction to the primary trigonometric ratios of sine, cosine, and tangent. These particular primary trigonometric ratios are all derived from the ratios between two sides within a right triangle. The two special triangles (i.e. 45°-45°-90° and 30°-60°-90°) are also right triangles and they are the building blocks for the unit circle. It appears that trigonometry depends critically on right triangles. However, as you continue your study in trigonometry and circular functions, you will soon come across questions when you need to consider trigonometric ratios of angles greater than a right angle. On what basis are you confident in extending your knowledge based on right triangles to angles greater than 90°? What other knowledge claims must you establish before you can generalize your understanding of trigonometry?

Trigonometry was first introduced around the 3rd century BC. It stems from the Greek origin of *trigono* and *metron*. This branch of mathematics mainly involves right triangles, even when non-right triangles are presented. However, the basic understanding of trigonometry becomes less absolute when you start studying spherical trigonometry when the figure is no longer a planar triangle. For non-planar triangles, the interior angles of such figures may have a sum greater or less than 180°. As you have already learned in your pre-DP math classes, primarily in your Euclidean Geometry lesson, the Angle Sum Triangle Theorem (ASTT) dictates that the sum of the interior angles of a triangle will always be 180°. However, when you began to move away from Euclidean Geometry, did you ever wonder whether it is necessary to have a new set of theorems and conjectures?

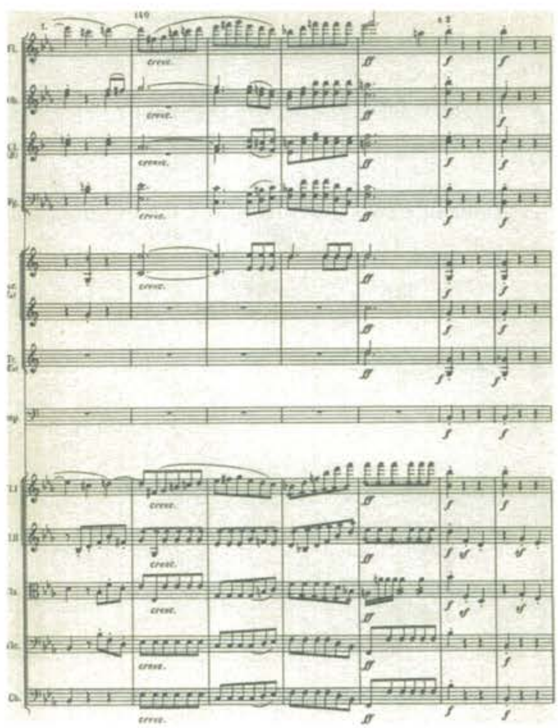
In the process of the development of new theorems, mathematicians need to go through a very rigorous process to ensure it is true for all cases. However, it merely takes one counterexample to disprove the theorem's validity. If you follow this argument then, when you extend your knowledge beyond Euclidean geometry, certain geometrical facts are no longer true. Hence, can you safely suggest that the original theorems and facts are invalid? If not, then when is it appropriate to set limitations to ensure the validity of theorems and how wide should those limitations and restrictions be?

Mathematics and Music

Music, especially tones and sound waves, are often expressed using mathematics. Sound waves allow the listener to visualize tones, pitches, volumes, and all other related properties which are uniquely understood by one's audio perception. Does it then merely imply that music is mathematical or does it suggest that mathematics is musical? Mathematics is in Group 5 while Music is part of Group 6, and yet, it appears that mathematics and music are highly associated with each other. How do you see this connection? Does music help you to see the truth in mathematics? Does mathematical truth help you to understand music?

If you are a musician or if you study music, can you merely use mathematics alone to create a piece of music? If a random set of numbers has no particular mathematical pattern, can you conclude that it is also the same as a collection of random musical notes with no harmony? Conversely, do all mathematical patterns translate into a corresponding set of musical tones?

Music and Mathematics - the only 'universal languages'?

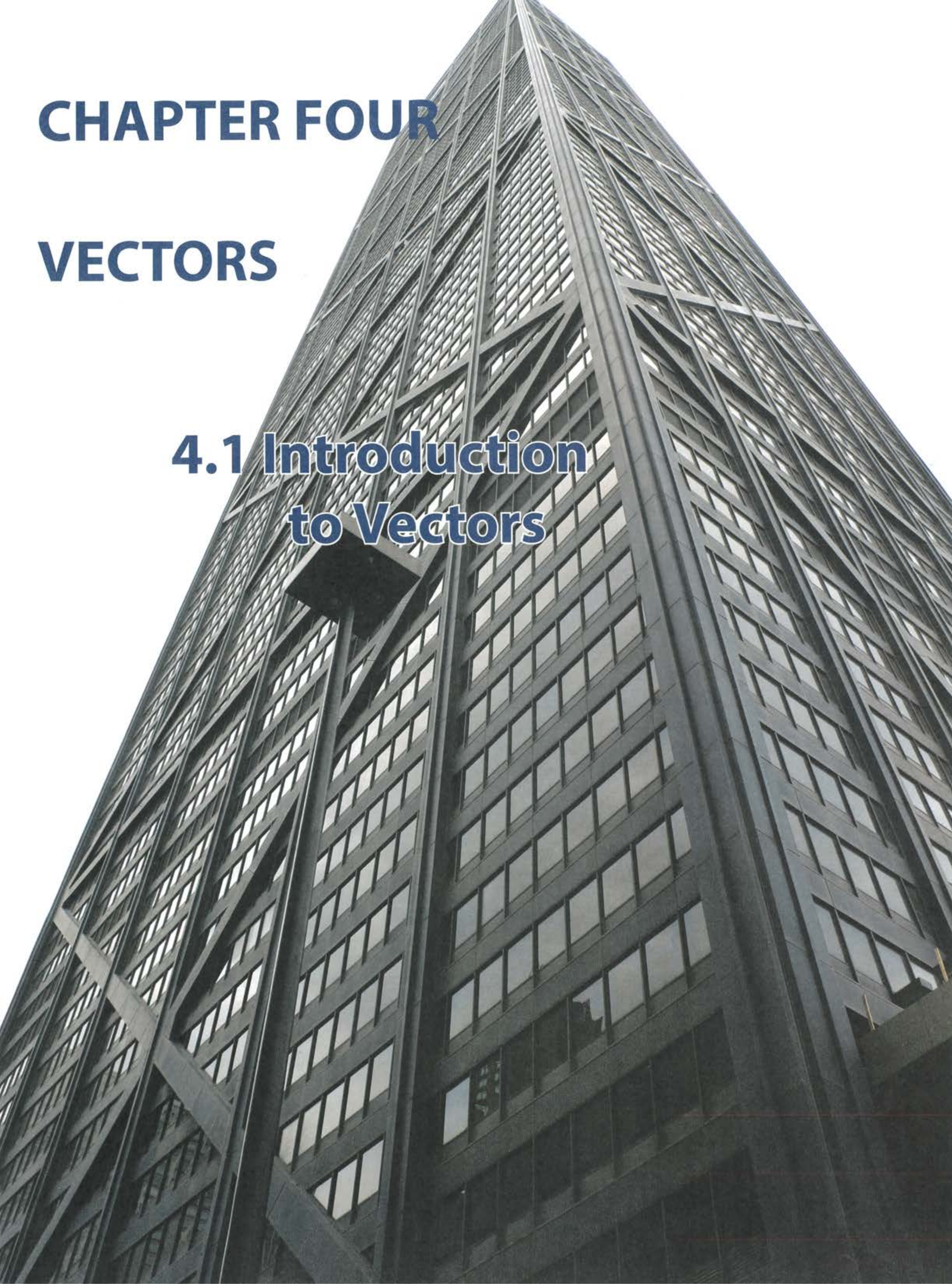


Answers

CHAPTER FOUR

VECTORS

4.1 Introduction to Vectors



Modern structures such as our main picture are often made from supporting structures of steel wires and beams. These are frequently visible.

It is both the tension forces in these components and their direction that gives the building its strength. This dual feature (force and direction) defines a vector.

Scalar and vector quantities

Numerical measurement scales are in widespread use. It is important to be able to distinguish between two distinct types of measurement scales, scalars and vectors.

Scalar quantities

A scalar is a quantity that has magnitude (size) but no direction. For example, we measure the mass of objects using a variety of scales such as 'kilograms' and 'pounds'. These measures have magnitude in that more massive objects (such as the sun) have a larger numerical mass than small objects (such as this book). Giving the mass of this book does not, however, imply that this mass has a direction. This does **not** mean that scalar quantities must be positive. Signed scalar quantities, such as temperature as measured by the Celsius or Fahrenheit scales (which are commonly used) also exist.

Vector quantities

Some measurements have both magnitude and direction. When we pull on a door handle, we exert what is known as a force. The force that we exert has both magnitude (we either pull hard or we pull gently) and direction (we open or close the door). Both the size of the pull and its direction are important in determining its effect. Such quantities are said to be vectors. Other examples of vectors are velocity, acceleration and displacement. The mathematics that will be developed in this section can be applied to problems involving any type of vector quantity.

Exercise 4.1.1

The following situations need to be described using an appropriate measure. Classify the measure as a scalar (s) or a vector (v).

1. A classroom chair is moved from the front of the room to the back.
2. The balance in a bank account.
3. The electric current passing through an electric light tube.
4. A dog, out for a walk, is being restrained by a lead.
5. An aircraft starts its take-off run.
6. The wind conditions before a yacht race.
7. The amount of liquid in a jug.
8. The length of a car.

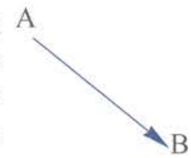
Representing Vectors

Directed line segment

There are a number of commonly used notations for vectors:

Notation 1:

This vector runs from A to B and is depicted as $\overrightarrow{AB}$ or **AB** with the arrow giving the direction of the vector. Point A is known as the tail of the vector $\overrightarrow{AB}$ and point B is known as the head of vector $\overrightarrow{AB}$.

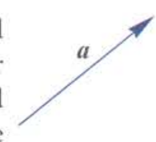


We also say the $\overrightarrow{AB}$ is the position vector of B relative to (from) A.

In the case where a vector starts at the origin (O), the vector running from O to another point C is simply called the position vector of C, $\overrightarrow{OC}$ or **OC**.

Notation 2:

Rather than using two reference points, A and B, as in notation 1, we can also refer to a vector by making reference to a single letter attached to an arrow. In essence we are 'naming' the vector.



The vector $\boldsymbol{a}$ can be expressed in several ways. In text books they are often displayed in bold type, however, in written work, the following notations are generally used:

$$\boldsymbol{a} = \underset{\sim}{a} = \vec{a}$$

We will consider another vector notation later in this chapter.

Magnitude of a vector

The magnitude or modulus of a vector is its length, which is the distance between its tail and head. We denote the magnitude of AB by |AB| (or more simply by AB).

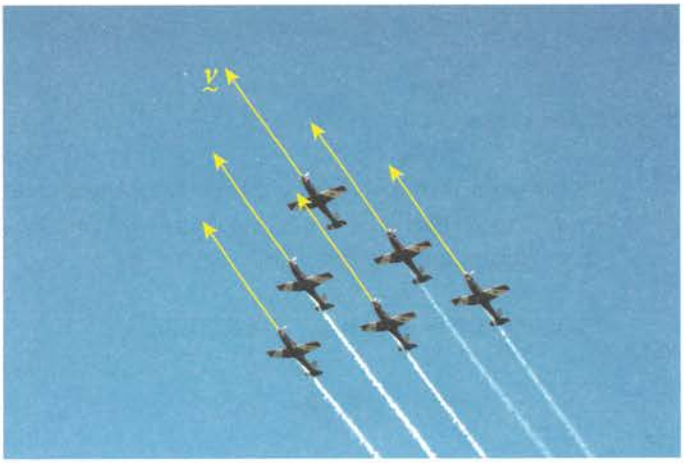
Similarly, if we are using vector notation 2, we may denote the magnitude of $\boldsymbol{a}$ by $|\boldsymbol{a}| = a$.

Note then that $|\boldsymbol{a}| \geq 0$.

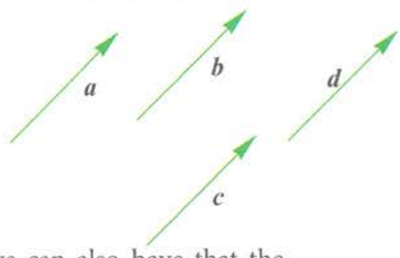
Equal vectors

Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are said to be equal if they have the same direction and the same magnitude, i.e. if they point in the same direction and $|\boldsymbol{a}| = |\boldsymbol{b}|$.

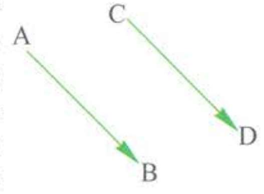
These aircraft must have equal velocity vectors if they are to maintain their formation.



Notice that if $\boldsymbol{a} = \boldsymbol{b}$, then vector $\boldsymbol{b}$ is a translation of vector $\boldsymbol{a}$. Using this notation, where there is no reference to a fixed point in space, we often use the term *free* vectors. That is, free vectors are vectors that have no specific position associated with them. In the diagram below, although the four vectors occupy a different space, they are all equal.

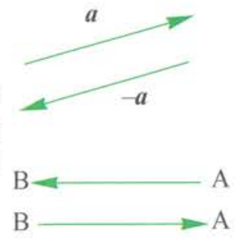


Note that we can also have that the vectors $\mathbf{AB} = \mathbf{CD}$, so that although they do not have the same starting point (or ending point) they are still equal because their magnitudes are equal and they have the same direction.



Negative vectors

The negative of a vector $\boldsymbol{a}$, denoted by '$-\boldsymbol{a}$' is the vector $\boldsymbol{a}$ but pointing in the opposite direction to $\boldsymbol{a}$.



Similarly, the negative of $\mathbf{AB}$ is $-\mathbf{AB}$ or $\mathbf{BA}$, because rather than starting at A and ending at B the negative of $\mathbf{AB}$ starts at B and ends at A.

Note that $|\boldsymbol{a}| = |-\boldsymbol{a}|$ and $|\mathbf{AB}| = |-\mathbf{AB}| = |\mathbf{BA}|$.

Zero vector

The zero vector has zero magnitude, $|\mathbf{0}| = 0$ and has no definite direction. It is represented geometrically by joining a point onto itself. Note then that for any non-zero vector $\boldsymbol{a}$, $|\boldsymbol{a}| > 0$.

Orientation and vectors

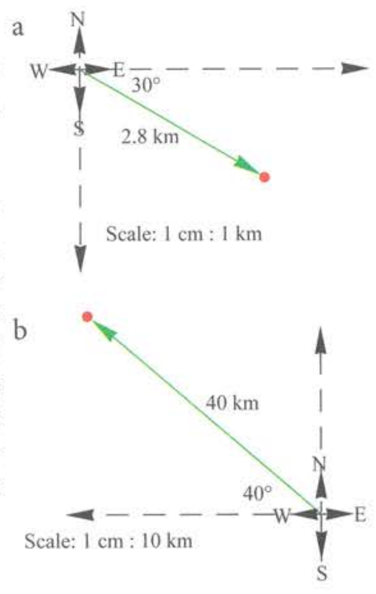
Vectors are useful when representing positions relative to some starting point. Consider:

the position of a man who has walked 2.8 km across a field in a direction East 30° South or

a car moving at 20 km/h in a direction W 40° N for 2 hours.

Each of these descriptions can be represented by a vector.

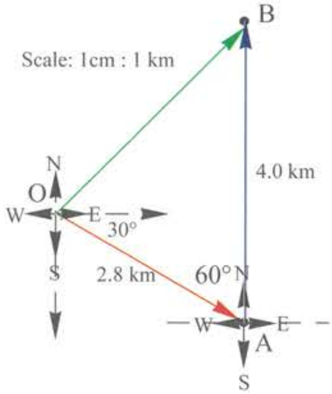
We start by setting up a set of axes and then we represent the above vectors showing the appropriate direction and magnitude. Representing the magnitude can be done using a scale drawing or labelling the length of the vector.



Example 4.1.1

Find the position of a bushwalker if, on the first part of her journey, she walks 2.8 km across a field in a direction East 30° South, and then continues for a further 4 km in a northerly direction.

We start by representing her journey using a vector diagram. The first part of her journey is represented by vector OA and the second partby AB. Note then that because her final position is at point B, her final position, relative to O, is given by the vector OB.



All that remains is to find the direction of OB and its magnitude. To do this we make use of trigonometry.

Finding |**OB**|:Using the cosine rule we have:

$$OB^2 = OA^2 + AB^2 - 2(AB)(OA)\cos(60°)$$

$$= 2.8^2 + 4.0^2 - 2 \times 2.8 \times 4.0 \times 0.5$$

$$= 12.64$$

$$\therefore OB = 3.56$$

Next, we find the angle BOA:

$$AB^2 = OA^2 + OB^2 - 2(OA)(OB)\cos(\angle BOA)$$

$$4.0^2 = 2.8^2 + 12.64 - 2(2.8)(\sqrt{12.64})\cos(\angle BOA)$$

$$\therefore \cos(\angle BOA) = \frac{2.8^2 + 12.64 - 4.0^2}{2(2.8)(\sqrt{12.64})}$$

$$\angle BOA = \cos^{-1}(0.2250)$$

$$= 76°59'45''$$

$$\approx 77°$$

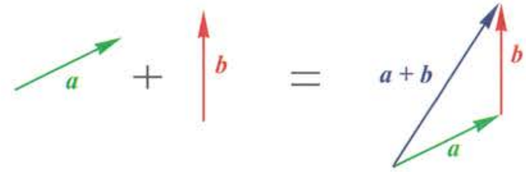
That is, the bushwalker is 3.56 km E 47° N from her starting point.

Although we will investigate the algebra of vectors in the next section, in Example 4.1.1 we have already looked at adding two vectors informally. That is, the final vector **OB** was found by *joining* the vectors **OA** and **AB**. Writing this in vector form we have, $\mathbf{OB} = \mathbf{OA} + \mathbf{AB}$.

To add two vectors, ***a*** and ***b***, geometrically we

1. first draw ***a***,
2. draw vector ***b*** so that its tail meets the arrow end of vector ***a***,
3. draw a line segment from the tail of vector ***a*** to the arrow end of vector ***b***.

This vector then represents the result ***a*** + ***b***.

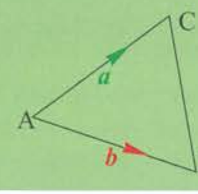


Example 4.1.2

For the equilateral triangle shown, express in terms of *a* and *b*.

a CA b BC

c |**AB** + **BC**|



a $\mathbf{CA} = -\mathbf{AC} = -\boldsymbol{a}$.

b To get from B to C we first get from B to A and then from A to C. That is, we 'join' the vectors **BA** and **AC**. In vector notation we have: $\mathbf{BC} = \mathbf{BA} + \mathbf{AC}$

However, $\mathbf{AB} = \boldsymbol{b} \Rightarrow \mathbf{BA} = -\mathbf{AB} = -\boldsymbol{b}$

$$\therefore \mathbf{BC} = -\boldsymbol{b} + \boldsymbol{a}$$

c $\mathbf{AB} + \mathbf{BC} = \mathbf{AC} = \boldsymbol{a} \therefore |\mathbf{AB} + \mathbf{BC}| = |\boldsymbol{a}|$

Exercise 4.1.1

1. Using a scale of 1 cm representing 10 units sketch the vectors that represent:

   a 30 km in a westerly direction.

   b 20 newtons applied in a NS direction.

   c 15 m/s N 60° E.

   d 45 km/h W 30° S.

2. The vector $\longleftarrow$ represents a velocity of 20 m/s due west. Represent the following vectors:

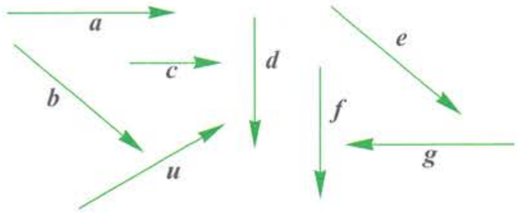
   a 20 m/s due east

   b 40 m/s due west

   c 60 m/s due east

   d 40 m/s due NE

3. State which of the vectors shown:



a have the same magnitude.

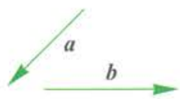
b are in the same direction.

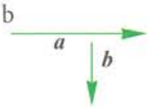
c are in opposite directions.

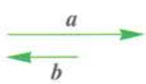
d are equal.

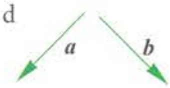
e are parallel.

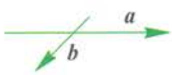
4. For each of the following pairs of vectors, find $\boldsymbol{a} + \boldsymbol{b}$.

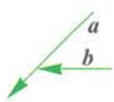
a


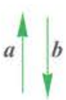
b


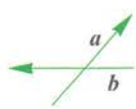
c


d


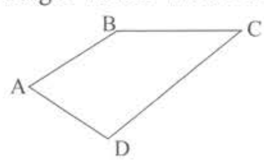
e


f


g


h


5. For the shape shown, find a single vector which is equal to:



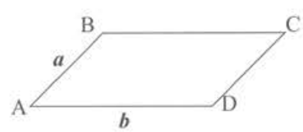
a **AB** + **BC**

b **AD** + **DB**

c **AC** + **CD**

d **BC** + **CD** + **DA** e **CD** + **DA** + **AB** + **BC**

6. Consider the parallelogram shown alongside. Which of the following statements are true?



a **AB** = **DC** b $|\boldsymbol{a}| = |\boldsymbol{b}|$

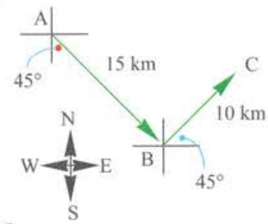
c **BC** = $\boldsymbol{b}$ d $|\mathbf{AC} + \mathbf{CD}| = |\boldsymbol{b}|$

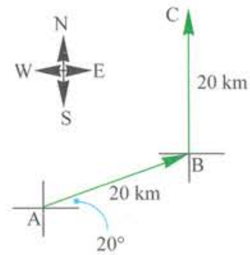
e **AD** = **CB**

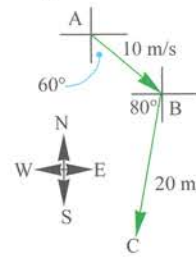
7. For each of the following:

i complete the diagram by drawing the vector **AB**+**BC**.

ii find $|\mathbf{AB} + \mathbf{BC}|$.

a


b


c


8. Two forces, one of 40 newtons acting in a northerly direction and one of 60 newtons acting in an easterly direction, are applied at a point A. Draw a vector diagram representing the forces. What is the resulting force at A?

9. Two trucks, on opposite sides of a river, are used to pull a barge along a straight river. They are connected to the barge at one point by ropes of equal length. The angle between the two ropes is 50°. Each truck is pulling with a force of 1500 newtons.

a Draw a vector diagram representing this situation.

b Find the magnitude and direction of the force acting on the barge.

10. An aircraft is flying at 240 km/h in a northerly direction when it encounters a 40 km/h wind from:

i the north. ii the north-east.

a Draw a vector diagram representing these situations.

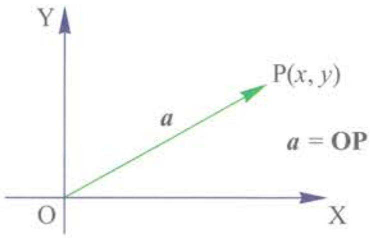
b In each case, find the actual speed and direction of the aircraft.

Extra question

Cartesian Representation of Vectors

Representation in two dimensions

When describing vectors in two-dimensional space it is often helpful to make use of a rectangular Cartesian coordinate system.



As such, the position vector of the point P, **OP**, has the coordinates (x, y).

The vector $\boldsymbol{a}$ can be expressed as a column vector $\begin{pmatrix} x \\ y \end{pmatrix}$.

That is:

$\boldsymbol{a} = \begin{pmatrix} x \\ y \end{pmatrix}$ is the position vector OP where P has the coordinates (x,y).

Unit vector and base vector notation

We define the unit vector $\boldsymbol{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

as the position vector of the point having coordinates $(1, 0)$,

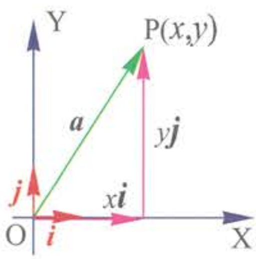
and the unit vector $\boldsymbol{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

as the position vector of the point having coordinates $(0, 1)$.

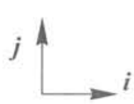
The term unit vector refers to the fact that the vector has a magnitude of one.

$$\boldsymbol{a} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix} = x\boldsymbol{i} + y\boldsymbol{j}$$

i.e. the position vector of any point can be expressed as the sum of two vectors, one parallel to the x-axis and one parallel to the y-axis.



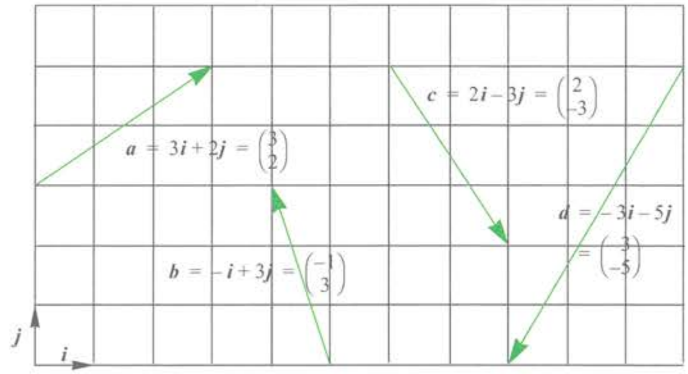
The unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ are also known as the base vectors. If we confine ourselves to vectors that exist in the plane of this page, the most commonly used basis is:



where $|\boldsymbol{i}| = |\boldsymbol{j}| = 1$

Notice the definite direction of the base vectors, i.e. $\boldsymbol{i}$ points in the positive x-axis direction while $\boldsymbol{j}$ points in the positive y-axis direction.

Vectors can now be expressed in terms of these base vectors.



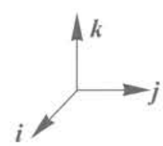
The vector $\boldsymbol{a}$ is 'three steps to the right and two steps up' and can be written in terms of the standard basis as $\boldsymbol{a} = 3\boldsymbol{i} + 2\boldsymbol{j}$.

The vector $\boldsymbol{b}$ is 'one step to the left and three steps up'. 'One step to the left' is in the opposite direction of the basis element $\boldsymbol{i}$ and is written $-\boldsymbol{i}$, giving the definition of the vector $\boldsymbol{b} = -\boldsymbol{i} + 3\boldsymbol{j}$. The vectors $-\boldsymbol{i}$ and $3\boldsymbol{j}$ are known as components of the vector $\boldsymbol{b}$.

The other definitions follow in a similar way.

Representation in three dimensions

When vectors are represented in three-dimensional space, a third vector must be added to the basis, in this case it is a unit vector $\boldsymbol{k}$ and is such that the three unit vectors are mutually perpendicular as shown.



In addition, extra basis vectors can be added to generate higher dimensional vector spaces. These may not seem relevant to us, inhabiting as we do, a three dimensional space. However, it remains the case that it is possible to do calculations in higher dimensional spaces and these have produced many valuable results for applied mathematicians.

As was the case for vectors in two dimensions, we can represent vectors in three dimensions using column vectors as follows:

The position vector $\boldsymbol{a}$ = **OP** where P has coordinates (x, y, z) is given by

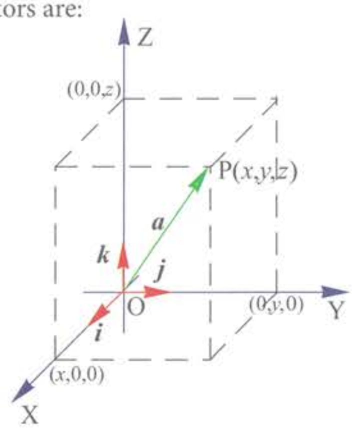
$$\boldsymbol{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$= x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$$

Where this time the base vectors are:

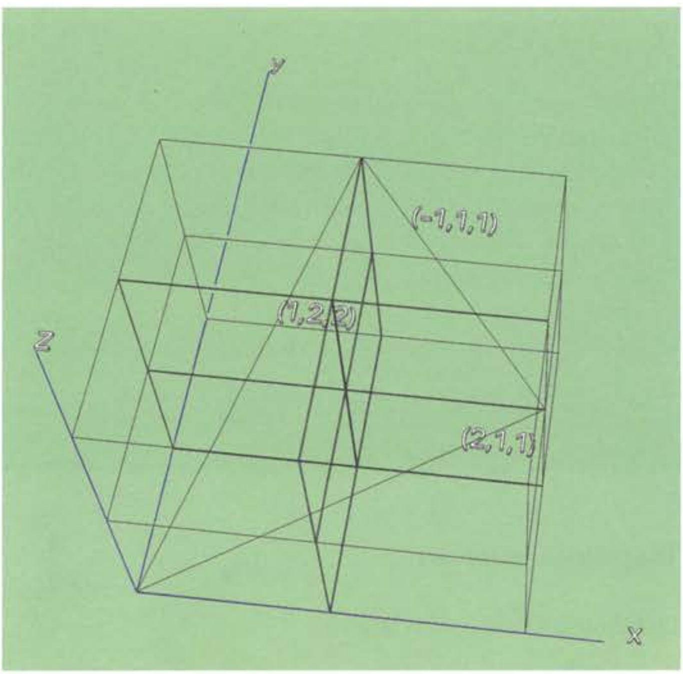
$$i = \begin{pmatrix}1\\0\\0\end{pmatrix},\ j = \begin{pmatrix}0\\1\\0\end{pmatrix} \text{ and}$$

$$k = \begin{pmatrix}0\\0\\1\end{pmatrix}$$



Vectors in three dimensions can be difficult to visualise.

This diagram is a representation of the sum of vectors in three dimensions:



The diagram shows:

$$(2,1,1) + (-1,1,1) = (1,2,2)$$

The following QR code links to a 3 dimensional image of this calculation that you will be able to 'tumble' in order to get a better idea of the geometry of the situation.

3-d image

Vector operations

Addition and subtraction

If $a = \begin{pmatrix}x_1\\y_1\end{pmatrix} = x_1 i + y_1 j$ and $b = \begin{pmatrix}x_2\\y_2\end{pmatrix} = x_2 i + y_2 j$ then:

$$a \pm b = \begin{pmatrix}x_1\\y_1\end{pmatrix} \pm \begin{pmatrix}x_2\\y_2\end{pmatrix} = \begin{pmatrix}x_1 \pm x_2\\y_1 \pm y_2\end{pmatrix} = (x_1 \pm x_2)i + (y_1 \pm y_2)j$$

If $a = \begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix} = x_1 i + y_1 j + z_1 k$ and $b = \begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix} = x_2 i + y_2 j + z_2 k$

then

$$a \pm b = \begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix} \pm \begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix} = \begin{pmatrix}x_1 \pm x_2\\y_1 \pm y_2\\z_1 \pm z_2\end{pmatrix}$$

$$= (x_1 \pm x_2)i + (y_1 \pm y_2)j + (z_1 \pm z_2)k$$

Scalar multiplication

If $a = \begin{pmatrix}x\\y\end{pmatrix} = xi + yj$

then $ka = k\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}kx\\ky\end{pmatrix} = kxi + kyj \quad , k \in \mathbb{R}$.

If $a = \begin{pmatrix}x\\y\\z\end{pmatrix} = xi + yj + zk$ then:

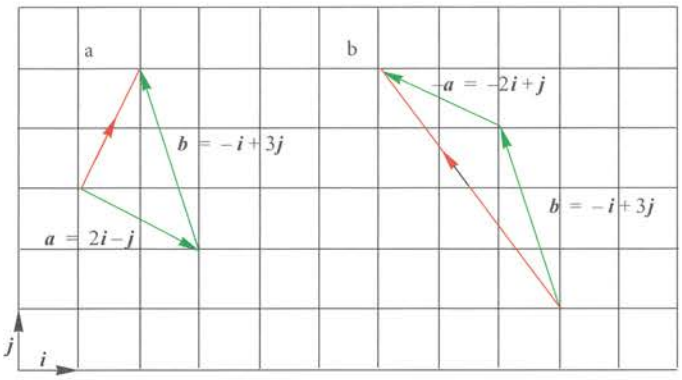
$$ka = k\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}kx\\ky\\kz\end{pmatrix} = kxi + kyj + kzk \quad , k \in \mathbb{R}$$

Example 4.1.3

If $a = 2i - j$ and $b = -i + 3j$, find:

a $a + b$ b $b - a$ c $3b - 2a$

Vectors are added 'nose to tail':



a Vectors are added in much the same way as are algebraic terms. Only like terms can be added or subtracted, so that $a + b = (2i - j) + (-i + 3j)$

$$= (2-1)i + (-1+3)j$$
$$= i + 2j$$

b This problem is solved in a similar way:

$$b - a = (-i + 3j) - (2i - j)$$
$$= (-1-2)i + (3-(-1))j$$
$$= -3i + 4j$$

Note that we could also have expressed the sum as:

$$b - a = b + (-a) = (-i + 3j) + (-2i + j).$$

$$= -3i + 4j$$

(i.e. the negative of a vector is the same length as the original vector but points in the opposite direction.)

c Combining the properties of scalar multiplication with those of addition and subtraction we have:

$$3b - 2a = 3(-i + 3j) - 2(2i - j)$$
$$= -3i + 9j - 4i + 2j$$
$$= -7i + 11j$$

Example 4.1.4

If $p = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix}$ and $q = \begin{pmatrix} -2 \\ 0 \\ 3 \end{pmatrix}$, find:

a $p + q$ b $p - \frac{q}{2}$ c $\frac{3}{2}q - p$.

a $p + q = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} + \begin{pmatrix} -2 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3-2 \\ -1+0 \\ 4+3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 7 \end{pmatrix}$

b $p - \frac{q}{2} = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -2 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3+1 \\ -1-0 \\ 4-1.5 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 2.5 \end{pmatrix}$

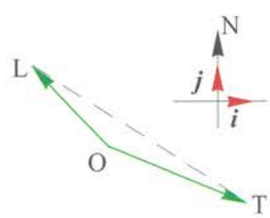
c $\frac{3}{2}q - p = 1.5\begin{pmatrix} -2 \\ 0 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix} = \begin{pmatrix} -3-3 \\ 0+1 \\ 4.5-4 \end{pmatrix} = \begin{pmatrix} -6 \\ 1 \\ 0.5 \end{pmatrix}$

Example 4.1.5

A surveyor is standing at the top of a hill. Call this point 'the origin' (O). A lighthouse (L) is visible 4 km to the west and 3 km to the north. A town (T) is visible 5 km to the south and 2 km to the east. Using a vector basis in which i is a 1 km vector running east and j is a 1 km vector running north,

find the position vectors of the lighthouse, $\overrightarrow{OL}$ and the town $\overrightarrow{OT}$. Hence find the vector $\overrightarrow{LT}$ and the position of the town relative to the lighthouse.

The position vectors are:



Lighthouse $\overrightarrow{OL} = -4i + 3j$ and

Town $\overrightarrow{OT} = 2i - 5j$.

Then, to get from L to T we have $\overrightarrow{LT} = \overrightarrow{LO} + \overrightarrow{OT}$.

$$= -\overrightarrow{OL} + \overrightarrow{OT}$$
$$= -(-4i + 3j) + (2i - 5j)$$
$$= 4i - 3j + 2i - 5j$$
$$= 6i - 8j$$

This means that the town is 6 km east of the lighthouse and 8 km south.

Exercise 4.1.2

1. If $\boldsymbol{a} = \boldsymbol{i} + 7\boldsymbol{j} - \boldsymbol{k}$ and $\boldsymbol{b} = 4\boldsymbol{i} + 7\boldsymbol{j} + 5\boldsymbol{k}$, find:

   a $4\boldsymbol{a}$ b $3\boldsymbol{b}$

   c $2\boldsymbol{a} - \boldsymbol{b}$ d $2(\boldsymbol{a} - \boldsymbol{b})$

2. The position vectors of A and B are $\overrightarrow{OA} = -3\boldsymbol{i} + 4\boldsymbol{j} - 2\boldsymbol{k}$ and $\overrightarrow{OB} = \boldsymbol{i} - 4\boldsymbol{j} - 3\boldsymbol{k}$. Find:

   a $\overrightarrow{AO}$ b $\overrightarrow{OA} - 5\overrightarrow{OB}$

   c $-5\overrightarrow{OA} + 3\overrightarrow{OB}$ d $3\overrightarrow{OA} + 6\overrightarrow{BO}$

3. If: $\boldsymbol{p} = \begin{pmatrix} -1 \\ -2 \\ 4 \end{pmatrix}$ and $\boldsymbol{q} = \begin{pmatrix} 6 \\ 1 \\ 2 \end{pmatrix}$. Find:

   a $\boldsymbol{p} + 2\boldsymbol{q}$ b $-3\boldsymbol{p} - 5\boldsymbol{q}$

   c $3\boldsymbol{p}$ d $2\boldsymbol{p} + 3\boldsymbol{q}$

4. Find the position vectors that join the origin to the points with coordinates A $(2, -1)$ and B $(-3, 2)$. Express your answers as column vectors. Hence find $\overrightarrow{AB}$.

5. A point on the Cartesian plane starts at the origin. The point then moves 4 units to the right, 5 units up, 6 units to the left and, finally 2 units down. Express these translations as a sum of four column vectors. Hence find the coordinates of the final position of the point.

6. Two vectors are defined as $\boldsymbol{a} = \boldsymbol{i} + \boldsymbol{j} + 4\boldsymbol{k}$ and $\boldsymbol{b} = -7\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$. Find:

   a $-6\boldsymbol{a} - 2\boldsymbol{b}$ b $-5\boldsymbol{a} + 2\boldsymbol{b}$

   c $4\boldsymbol{a} + 3\boldsymbol{b}$ d $-2(\boldsymbol{a} + 3\boldsymbol{b})$

7. If $\boldsymbol{x} = \begin{pmatrix} 4 \\ -4 \\ 2 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} 4 \\ 3 \\ 7 \end{pmatrix}$, find as column vectors.

   a $2\boldsymbol{x} + 3\boldsymbol{y}$ b $\boldsymbol{x} + 2\boldsymbol{y}$

   c $5\boldsymbol{x} - 6\boldsymbol{y}$ d $\boldsymbol{x} - 6\boldsymbol{y}$

8. Find the values of A and B if:

   $$A(7\boldsymbol{i} + 7\boldsymbol{j} + 4\boldsymbol{k}) - 3(3\boldsymbol{i} - \boldsymbol{j} + B\boldsymbol{k}) = -37\boldsymbol{i} - 25\boldsymbol{j} + 5\boldsymbol{k}$$

8. Find the values of A and B if:

   $$A(7\boldsymbol{i} + 7\boldsymbol{j} + 4\boldsymbol{k}) - 3(3\boldsymbol{i} - \boldsymbol{j} + B\boldsymbol{k}) = -37\boldsymbol{i} - 25\boldsymbol{j} + 5\boldsymbol{k}$$

9. Two vectors are defined as:

   $$\boldsymbol{a} = \begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} 6 \\ -6 \\ -5 \end{pmatrix}.$$

   Find values of the scalars X and Y if $X\boldsymbol{a} + Y\boldsymbol{b}$ is equal to:

   a $\begin{pmatrix} -36 \\ 32 \\ 33 \end{pmatrix}$

   b $\begin{pmatrix} 30 \\ -22 \\ -31 \end{pmatrix}$

   c $\begin{pmatrix} -12 \\ 24 \\ 1 \end{pmatrix}$.

10. A submarine (which is considered the origin of the vector system) is 60 metres below the surface of the sea when it detects two surface ships. A destroyer (D) is 600 metres to the east and 800 metres to the south of the submarine. An aircraft carrier (A) is 1200 metres to the west and 300 metres to the south.

    a Define a suitable vector basis for this problem.

    b Using the submarine as the origin, state the position vectors of the destroyer and the aircraft carrier.

    c A helicopter pilot, based on the aircraft carrier, wants to make a supplies delivery to the destroyer. Find, in vector terms, the course along which the pilot should fly.

Answers

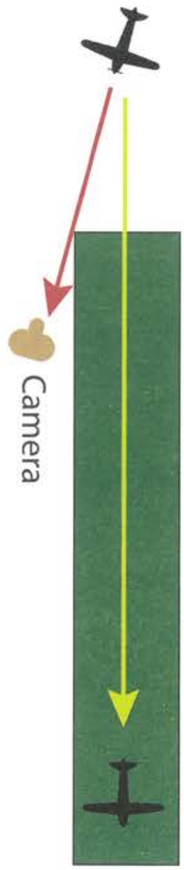
Application

Crosswind Landing

This small aeroplane is landing at a short grass landing strip on a coral atoll.

It appears that the aeroplane is heading almost straight for the camera.

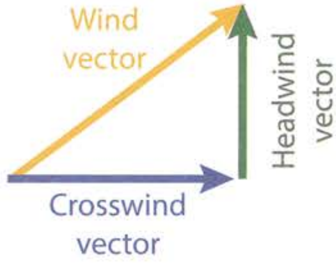
Fortunately for the photographer, this is not so. The aircraft is appraching the landing strip 'crabwise' in order to offset the drift created by a cross wind.



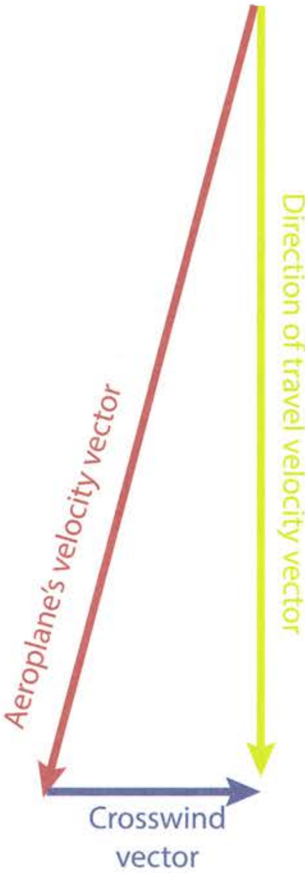
The vectors concerned are:

1. Wind

The crosswind is coming from the pilot's right and resolves into a component straight down the runway (green) and a component across the runway (blue).



2. The aeroplane



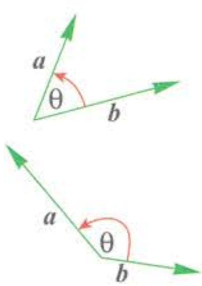
The crosswind component of the red vector balances the blue vector and the aeroplane travels straight down the runway. Without this, the aeroplane would drift off the runway line to the pilot's left during the approach. Just before touchdown, the pilot will straighten the aeroplane by using left rudder to 'yaw' it to the left. Some right aileron is also necessary to counteract the roll that happens during this 'de-crabbing'. Can you see why?

4.2 Scalar Product

Definition of the Scalar Product

The scalar product (or **dot product**) of two vectors is defined by:

$$\boldsymbol{a} \bullet \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}| \cos\theta$$



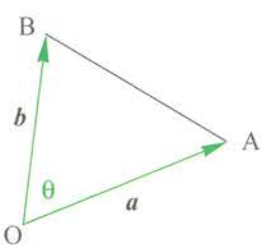
where θ is the angle between the two vectors and may be an obtuse angle. The angle must be measured between the directions of the vectors. That is, the angle between the two vectors once they are joined tail to tail.

The three quantities on the right-hand side of the equation are all scalars and it is important to realise that, when the **scalar** product of two **vectors** is calculated, the result is a **scalar**.

Example 4.2.1

Find the scalar product of the vectors $2i - 3j + k$ and $i + j - k$.

Let $\boldsymbol{a} = 2\boldsymbol{i} - 3\boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}$, then to determine the scalar product, $\boldsymbol{a} \bullet \boldsymbol{b}$, we need to find:



$|\boldsymbol{a}|$, $|\boldsymbol{b}|$ and $\cos\theta$, where θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

Finding:

$$|\boldsymbol{a}| = \sqrt{2^2 + (-3)^2 + 1^2} = \sqrt{14}.$$

$$|\boldsymbol{b}| = \sqrt{1^2 + 1^2 + (-1)^2} = \sqrt{3}.$$

Finding $\cos\theta$ requires a little work. Relative to a common origin O, the points A(2, –3, 1) and B(1, 1, –1) have position vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

Before making use of the cosine rule we need to determine the length of AB. Using the distance formula between two points in space, we have:

$$\begin{aligned} AB &= \sqrt{(1-2)^2 + (1-(-3))^2 + (-1-1)^2} \\ &= \sqrt{1+16+4} \\ &= \sqrt{21} \end{aligned}$$

Cosine rule:

$$AB^2 = OA^2 + OB^2 - 2 \cdot OA \cdot OB \cdot \cos\theta$$

$$(\sqrt{21})^2 = (\sqrt{14})^2 + (\sqrt{3})^2 - 2 \cdot \sqrt{14} \cdot \sqrt{3} \cdot \cos\theta$$

$$21 = 14 + 3 - 2\sqrt{42}\cos\theta$$

$$\therefore \cos\theta = -\frac{2}{\sqrt{42}}$$

Next, from the definition of the scalar product:

$\boldsymbol{a} \bullet \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos\theta$, we have

$$\boldsymbol{a} \bullet \boldsymbol{b} = \sqrt{14} \times \sqrt{3} \times -\frac{2}{\sqrt{42}} = -2$$

The solution to Example 4.2.1 was rather lengthy. However, we now look at the scalar product from a slightly different viewpoint.

First consider the dot product $\boldsymbol{i} \bullet \boldsymbol{i}$:

Using the definition, we have that

$$\boldsymbol{i} \bullet \boldsymbol{i} = |\boldsymbol{i}||\boldsymbol{i}|\cos 0 = 1 \times 1 \times 1 = 1$$

(the angle between the vectors $\boldsymbol{i}$ and $\boldsymbol{i}$ is 0 and so $\cos\theta = \cos 0 = 1$).

Next consider the product $\boldsymbol{i} \bullet \boldsymbol{j}$:

Using the definition, we have that:

$$\boldsymbol{i} \bullet \boldsymbol{j} = |\boldsymbol{i}||\boldsymbol{j}|\cos 90 = 1 \times 1 \times 0 = 0$$

(the angle between the vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ is 90° and so $\cos\theta = \cos 90° = 0$).

Similarly, we end up with the following results for all possible combinations of the $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ vectors:

$$\boldsymbol{i} \bullet \boldsymbol{i} = \boldsymbol{j} \bullet \boldsymbol{j} = \boldsymbol{k} \bullet \boldsymbol{k} = 1$$

and

$$\boldsymbol{i} \cdot \boldsymbol{j} = \boldsymbol{i} \cdot \boldsymbol{k} = \boldsymbol{j} \cdot \boldsymbol{k} = \boldsymbol{j} \cdot \boldsymbol{i} = \boldsymbol{k} \cdot \boldsymbol{i} = \boldsymbol{k} \cdot \boldsymbol{j} = 0$$

Armed with these results we can now work out the scalar product of the vectors $\boldsymbol{a} = x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}$ and $\boldsymbol{b} = x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}$ as follows:

$$\begin{aligned} a \bullet b &= (x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}) \bullet (x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}) \\ &= x_1x_2(\boldsymbol{i} \bullet \boldsymbol{i}) + x_1y_2(\boldsymbol{i} \cdot \boldsymbol{j}) + x_1z_2(\boldsymbol{i} \cdot \boldsymbol{k}) \\ &\quad + y_1x_2(\boldsymbol{j} \cdot \boldsymbol{i}) + y_1y_2(\boldsymbol{j} \bullet \boldsymbol{j}) + y_1z_2(\boldsymbol{j} \cdot \boldsymbol{k}) \\ &\quad + z_1x_2(\boldsymbol{k} \cdot \boldsymbol{i}) + z_1y_2(\boldsymbol{k} \cdot \boldsymbol{j}) + z_1z_2(\boldsymbol{k} \bullet \boldsymbol{k}) \end{aligned}$$

$$a \bullet b = x_1x_2 + y_1y_2 + z_1z_2$$

That is, if:

$\boldsymbol{a} = x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}$ and $\boldsymbol{b} = x_2\boldsymbol{i} + y_2\boldsymbol{j} + z_2\boldsymbol{k}$

then $a \bullet b = x_1x_2 + y_1y_2 + z_1z_2$

Using this result with the vectors of Example 4.2.1, $2\boldsymbol{i} - 3\boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}$ we have:

$$\begin{aligned} (2\boldsymbol{i} - 3\boldsymbol{j} + \boldsymbol{k}) \bullet (\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}) &= 2 \times 1 + (-3) \times 1 + 1 \times (-1) \\ &= 2 - 3 - 1 \\ &= -2 \end{aligned}$$

This is a much faster process!

However, the most usual use of scalar product is to calculate the angle between vectors using a rearrangement of the definition of scalar product:

$$\cos\theta = \frac{\boldsymbol{a} \bullet \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}$$

Example 4.2.2

For the following pairs of vectors, find their magnitudes and scalar products. Hence find the angles between the vectors, correct to the nearest degree.

a $\quad -\boldsymbol{i} + 3\boldsymbol{j}$ and $-\boldsymbol{i} + 2\boldsymbol{j}$

b $\quad \begin{pmatrix} 0 \\ -5 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} -5 \\ -1 \\ -3 \end{pmatrix}$.

a In using the scalar product, it is necessary to calculate the magnitudes of the vectors.

$$|-\boldsymbol{i} + 3\boldsymbol{j}| = \sqrt{(-1)^2 + 3^2} = \sqrt{10} \text{ and}$$

$$|-\boldsymbol{i} + 2\boldsymbol{j}| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$

Next, calculate the scalar product: $-\boldsymbol{i} + 3\boldsymbol{j} \qquad -\boldsymbol{i} + 2\boldsymbol{j}$

$$(-\boldsymbol{i} + 3\boldsymbol{j}) \bullet (-\boldsymbol{i} + 2\boldsymbol{j}) = -1 \times -1 + 3 \times 2 = 7$$

Finally, the angle is: $\cos\theta = \dfrac{\boldsymbol{a} \bullet \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \dfrac{7}{\sqrt{10} \times \sqrt{5}} \Rightarrow \theta \approx 8°$

b $\quad \left|\begin{pmatrix} 0 \\ -5 \\ 4 \end{pmatrix}\right| = \sqrt{0^2 + (-5)^2 + 4^2} = \sqrt{41}$

and $\quad \left|\begin{pmatrix} -5 \\ -1 \\ -3 \end{pmatrix}\right| = \sqrt{(-5)^2 + (-1)^2 + (-3)^2} = \sqrt{35}$

Next, the scalar product:

$$\begin{pmatrix} 0 \\ -5 \\ 4 \end{pmatrix} \bullet \begin{pmatrix} -5 \\ -1 \\ -3 \end{pmatrix} = 0 \times (-5) + (-5) \times (-1) + 4 \times (-3) = -7$$

Finally, the angle can be calculated:

$$\cos\theta = \frac{\boldsymbol{a} \bullet \boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \frac{-7}{\sqrt{41} \times \sqrt{35}} \Rightarrow \theta \approx 101°$$

The use of cosine means that obtuse angles between vectors (which occur when the scalar product is negative) are calculated correctly when using the inverse cosine function on a calculator.

Properties of the Scalar Product

Closure The scalar product of two vectors is a scalar (i.e. the result is not a vector). The operation is not closed and so closure does not apply.

Commutative

Now, $\boldsymbol{a} \bullet \boldsymbol{b} = |a||b|\cos\theta = |b||a|\cos\theta = \boldsymbol{b} \bullet \boldsymbol{a}$

That is, $\boldsymbol{a} \bullet \boldsymbol{b} = \boldsymbol{b} \bullet \boldsymbol{a}$.

Therefore the operation of scalar product is commutative.

Associative If the associative property were to hold it would take on the form $(\boldsymbol{a} \bullet \boldsymbol{b}) \bullet \boldsymbol{c} = \boldsymbol{a} \bullet (\boldsymbol{b} \bullet \boldsymbol{c})$. However, $\boldsymbol{a} \bullet \boldsymbol{b}$ is a real number and therefore the operation $(\boldsymbol{a} \bullet \boldsymbol{b}) \bullet \boldsymbol{c}$ has no meaning (you cannot 'dot' a scalar with a vector).

Distributive The **scalar product is distributive** (over addition).

We leave the proof of this result as an exercise - it was assumed in the discussion on the previous page.

Identity As the operation of scalar product is not closed, an identity cannot exist.

Inverse As the operation of scalar product is not closed, an inverse cannot exist.

Note that although the scalar product is non-associative, the following 'associative rule' holds for the scalar product:

If $k \in \mathbb{R}$, then, $\boldsymbol{a} \bullet (k\boldsymbol{b}) = k(\boldsymbol{a} \bullet \boldsymbol{b})$

Special cases of the scalar product

Perpendicular vectors

If the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are perpendicular then:

$$\boldsymbol{a} \bullet \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos\frac{\pi}{2} = 0.$$

(Note: We are assuming that $\boldsymbol{a}$ and $\boldsymbol{b}$ are non-zero vectors.)

Zero vector

For any vector $\boldsymbol{a}$, $\boldsymbol{a} \bullet \boldsymbol{0}$: $\boldsymbol{a} \bullet \boldsymbol{0} = |\boldsymbol{a}||\boldsymbol{0}|\cos\theta = 0$

Parallel vectors

If vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are parallel then, $\boldsymbol{a} \bullet \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos 0 = |\boldsymbol{a}||\boldsymbol{b}|$

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are antiparallel then, $\boldsymbol{a} \bullet \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos\pi = -|\boldsymbol{a}||\boldsymbol{b}|$.

(Note: We are assuming that $\boldsymbol{a}$ and $\boldsymbol{b}$ are non-zero vectors.)

Combining the results of 1 and 2 above, we have the important observation:

If $\boldsymbol{a} \bullet \boldsymbol{b} = 0$ then either:

1. $\boldsymbol{a}$ and/or $\boldsymbol{b}$ are both the zero vector, **0**.

Or

2. $\boldsymbol{a}$ and $\boldsymbol{b}$ are perpendicular with neither $\boldsymbol{a}$ nor $\boldsymbol{b}$ being the zero vector.

Notice how this result differs from the standard Null Factor Law when dealing with real numbers, where given $ab = 0$ then a or b or both are zero! That is, the cancellation property that holds for real numbers does not hold for vectors.

A nice application using the perpendicular property above can be seen in the next example.

Example 4.2.3

Three towns are joined by straight roads. Oakham is the state capital and is considered as the 'origin'. Axthorp is 3 km east and 9 km north of Oakham and Bostock is 5 km east and 5 km south of Axthorp.

Considering $\boldsymbol{i}$ as a 1 km vector pointing east and $\boldsymbol{j}$ a 1 km vector pointing north:

a Find the position vector of Axthorp relative to Oakham.

b Find the position vector of Bostock relative to Oakham.

A bus stop (S) is situated two thirds of the way along the road from Oakham to Axthorp.

c Find the vectors $\overline{\text{OS}}$ and $\overline{\text{BS}}$.

d Prove that the bus stop is the closest point to Bostock on the Oakham to Axthorp road.

a Axthorp is 3 km east and 9 km north of Oakham so $\overrightarrow{\text{OA}} = 3\boldsymbol{i} + 9\boldsymbol{j}$

b $\overrightarrow{\text{OB}} = \overrightarrow{\text{OA}} + \overrightarrow{\text{AB}} = 3\boldsymbol{i} + 9\boldsymbol{j} + 5\boldsymbol{i} - 5\boldsymbol{j} = 8\boldsymbol{i} + 4\boldsymbol{j}$

c $\quad \overrightarrow{OS} = \frac{2}{3}(\overrightarrow{OA}) = \frac{2}{3}(3\boldsymbol{i}+9\boldsymbol{j}) = 2\boldsymbol{i}+6\boldsymbol{j}$

$$\overrightarrow{BS} = \overrightarrow{BO}+\overrightarrow{OS} = -(8\boldsymbol{i}+4\boldsymbol{j})+2\boldsymbol{i}+6\boldsymbol{j} = -6\boldsymbol{i}+2\boldsymbol{j}$$

d $\quad$ The next step is to calculate the angle between $\overrightarrow{OS}$ and $\overrightarrow{BS}$ by calculating the scalar product of the two vectors:

$$\overrightarrow{OS}\bullet\overrightarrow{BS} = (2\boldsymbol{i}+6\boldsymbol{j})\bullet(-6\boldsymbol{i}+2\boldsymbol{j}) = 2\times(-6)+6\times 2 = 0$$

This means that $\overrightarrow{OS}$ and $\overrightarrow{BS}$ are at right angles to each other. It follows that the bus stop is the closest point to Bostock on the Oakham to Axthorp road.

Example 4.2.4

Find the value(s) of m for which the vectors $2m\boldsymbol{i}+m\boldsymbol{j}+8\boldsymbol{k}$ and $\boldsymbol{i}+3m\boldsymbol{j}-\boldsymbol{k}$ are perpendicular.

As the two vectors are perpendicular, then:

$$(2m\boldsymbol{i}+m\boldsymbol{j}+8\boldsymbol{k})\bullet(\boldsymbol{i}+3m\boldsymbol{j}-\boldsymbol{k}) = 0$$
$$\Rightarrow 2m+3m^2-8 = 0$$
$$\Leftrightarrow 3m^2+2m-8 = 0$$
$$\Leftrightarrow (3m-4)(m+2) = 0$$
$$\Leftrightarrow m = \frac{4}{3} \text{ or } m = -2$$

Example 4.2.5

Find a vector perpendicular to $\boldsymbol{u} = 4\boldsymbol{i}-3\boldsymbol{j}$.

Let the vector perpendicular to $\boldsymbol{u} = 4\boldsymbol{i}-3\boldsymbol{j}$ be $\boldsymbol{v} = x\boldsymbol{i}+y\boldsymbol{j}$.

Then, as $\boldsymbol{u}\perp\boldsymbol{v}\Rightarrow\boldsymbol{u}\bullet\boldsymbol{v} = 0$ so that $(4\boldsymbol{i}-3\boldsymbol{j})\bullet(x\boldsymbol{i}+y\boldsymbol{j}) = 0$

$$\therefore 4x-3y = 0 \quad -(1)$$

Unfortunately, at this stage we only have one equation for two unknowns! We need to obtain a second equation from somewhere. To do this we recognise the fact that if $\boldsymbol{v}$ is perpendicular to $\boldsymbol{u}$, then so too will the unit vector, $\hat{\boldsymbol{v}}$, be perpendicular to $\boldsymbol{u}$.

Then, as $|\hat{\boldsymbol{v}}| = 1 \Rightarrow \sqrt{x^2+y^2} = 1 \therefore x^2+y^2 = 1 \quad -(2)$

From (1) we have that $y = \frac{4}{3}x \quad -(3)$

Substituting (3) into (2) we have:

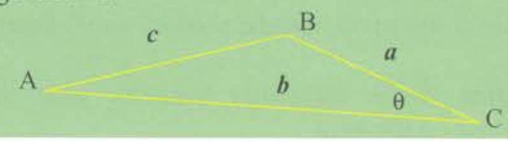
$$x^2+\left(\frac{4}{3}x\right)^2 = 1 \Leftrightarrow 25x^2 = 9 \Leftrightarrow x = \pm\frac{3}{5}$$

Substituting into (3) we have: $y = \pm\frac{4}{5}$

Therefore, both $\boldsymbol{v} = \frac{3}{5}\boldsymbol{i}+\frac{4}{5}\boldsymbol{j}$ and $\boldsymbol{v} = -\left(\frac{3}{5}\boldsymbol{i}+\frac{4}{5}\boldsymbol{j}\right)$ are perpendicular to $\boldsymbol{u}$.

Example 4.2.6

Use a vector method to derive the cosine rule for the triangle shown.



From the triangle rule for vector addition we have $\boldsymbol{a}+\boldsymbol{c} = \boldsymbol{b}\Leftrightarrow\boldsymbol{c} = \boldsymbol{b}-\boldsymbol{a}$.

Now, using the scalar product we have:

$$\begin{aligned}\boldsymbol{c}\bullet\boldsymbol{c} &= (\boldsymbol{b}-\boldsymbol{a})\bullet(\boldsymbol{b}-\boldsymbol{a})\\ &= \boldsymbol{b}\bullet\boldsymbol{b}-\boldsymbol{b}\bullet\boldsymbol{a}-\boldsymbol{a}\bullet\boldsymbol{b}+\boldsymbol{a}\bullet\boldsymbol{a}\\ &= |\boldsymbol{b}|^2-2\boldsymbol{a}\bullet\boldsymbol{b}+|\boldsymbol{a}|^2\\ \therefore|\boldsymbol{c}|^2 &= |\boldsymbol{b}|^2+|\boldsymbol{a}|^2-2|\boldsymbol{a}||\boldsymbol{b}|\cos\theta\end{aligned}$$

Example 4.2.7

Find a vector perpendicular to both $\boldsymbol{a} = 2\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{i}+3\boldsymbol{j}+\boldsymbol{k}$.

Let the vector $\boldsymbol{c} = x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}$ be perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$.

Then, we have that $\boldsymbol{a}\bullet\boldsymbol{c} = 0$ and $\boldsymbol{b}\bullet\boldsymbol{c} = 0$.

From $\boldsymbol{a}\bullet\boldsymbol{c} = 0$ we obtain:

$$(2\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k})\bullet(x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}) = 2x+y-z = 0 \quad -(1)$$

From $\boldsymbol{b}\bullet\boldsymbol{c} = 0$ we obtain:

$$(\boldsymbol{i}+3\boldsymbol{j}+\boldsymbol{k})\bullet(x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}) = x+3y+z = 0 \quad -(2)$$

In order to solve for the three unknowns we need one more equation. We note that if $\boldsymbol{c}$ is perpendicular to $\boldsymbol{a}$ and $\boldsymbol{b}$ then so too will the unit vector, $\hat{\boldsymbol{c}}$. So, without any loss in generality, we can assume that $\boldsymbol{c}$ is a unit vector. This will provide a third equation.

As we are assuming that $\boldsymbol{c}$ is a unit vector, we have:

$|\boldsymbol{c}| = 1 \therefore x^2 + y^2 + x^2 = 1$ – (3)

We can now solve for x, y and z:

(1) + (2): $3x + 4y = 0$ – (4)

2×(1) – (2): $5y + 3z = 0$ – (5)

Substituting (4) and (5) into (3): $\left(-\frac{4}{3}y\right)^2 + y^2 + \left(-\frac{5}{3}y\right)^2 = 1$

$$\Leftrightarrow 16y^2 + 9y^2 + 25y^2 = 9$$

$$\Leftrightarrow 50y^2 = 9$$

$$\Leftrightarrow y = \pm\frac{3}{5\sqrt{2}}$$

$$\therefore y = \pm\frac{3\sqrt{2}}{10}$$

Substituting into (4) and (5) we have $x = -\frac{4}{3} \times \pm\frac{3\sqrt{2}}{10} = \pm\frac{2\sqrt{2}}{5}$

and $z = -\frac{5}{3} \times \pm\frac{3\sqrt{2}}{10} = \pm\frac{\sqrt{2}}{2}$.

Therefore, $\pm\frac{2\sqrt{2}}{5}\boldsymbol{i} \pm \frac{3\sqrt{2}}{10}\boldsymbol{j} \pm \frac{\sqrt{2}}{2}\boldsymbol{k}$ or $\pm\left(\frac{2\sqrt{2}}{5}\boldsymbol{i} - \frac{3\sqrt{2}}{10}\boldsymbol{j} + \frac{\sqrt{2}}{2}\boldsymbol{k}\right)$

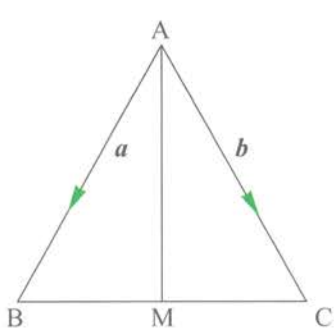
are two vectors perpendicular to $\boldsymbol{a}$ and $\boldsymbol{b}$. Of course, any multiple of this vector will also be perpendicular to $\boldsymbol{a}$ and $\boldsymbol{b}$.

As we have already seen in Example 4.2.6, the scalar product is a very powerful tool when proving theorems in geometry. We now look at another theorem that is otherwise lengthy to prove by standard means.

Example 4.2.8

Prove that the median to the base of an isosceles triangle is perpendicular to the base.

Consider the triangle ABC as shown, where M is the midpoint of the base $\overline{\mathrm{BC}}$. Next, let $\boldsymbol{a} = \mathbf{AB}$ and $\boldsymbol{b} = \mathbf{AC}$. We then wish to show that $\mathbf{AM} \perp \mathbf{BC}$ (or $\mathbf{AM} \bullet \mathbf{BC} = 0$).



Now, $\mathbf{AM} = \mathbf{AB} + \mathbf{BM} = \mathbf{AB} + \frac{1}{2}\mathbf{BC}$

$$= \boldsymbol{a} + \frac{1}{2}(\boldsymbol{b} - \boldsymbol{a})$$

$$= \frac{1}{2}(\boldsymbol{a} + \boldsymbol{b})$$

Therefore, $\mathbf{AM} \bullet \mathbf{BC} = \frac{1}{2}(\boldsymbol{a} + \boldsymbol{b}) \bullet (\boldsymbol{b} - \boldsymbol{a})$

$$= \frac{1}{2}(\boldsymbol{a} \bullet \boldsymbol{b} - \boldsymbol{a} \bullet \boldsymbol{a} + \boldsymbol{b} \bullet \boldsymbol{b} - \boldsymbol{b} \bullet \boldsymbol{a})$$

$$= \frac{1}{2}(-|\boldsymbol{a}|^2 + |\boldsymbol{b}|^2) \text{ (because } \boldsymbol{a} \bullet \boldsymbol{b} = \boldsymbol{b} \bullet \boldsymbol{a}\text{)}$$

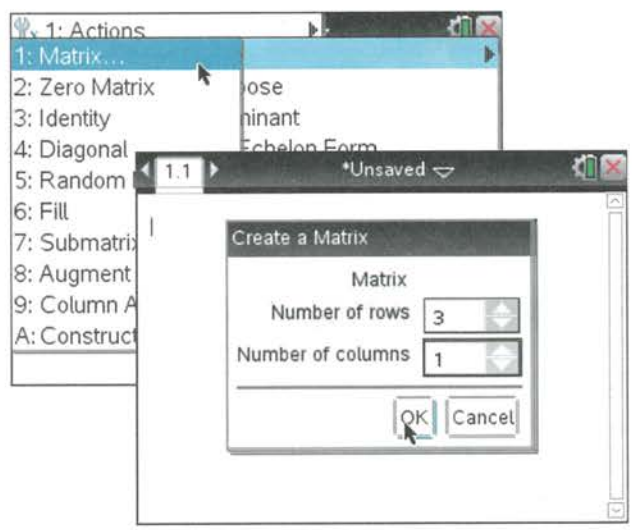
$$= 0 \text{ (because } |\boldsymbol{a}| = |\boldsymbol{b}|\text{)}$$

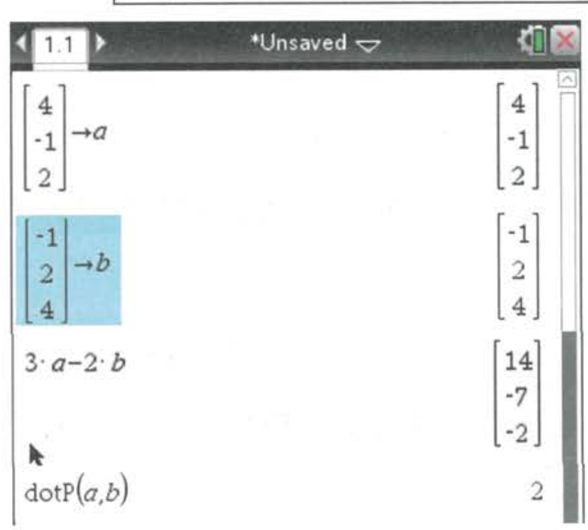
Therefore:

As $\mathbf{AM} \neq \mathbf{0}$ and $\mathbf{BC} \neq \mathbf{0}$, then $\mathbf{AM} \bullet \mathbf{BC} = 0 \Rightarrow \mathbf{AM} \perp \mathbf{BC}$

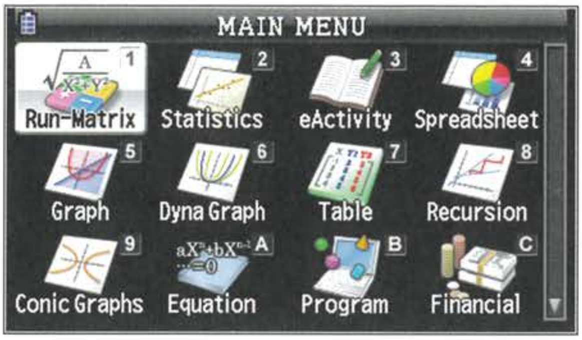
i.e. the median is perpendicular to the base.

Most graphic calculators can perform vector calculations. You should know how to do the basic procedures such as entering and saving vectors.





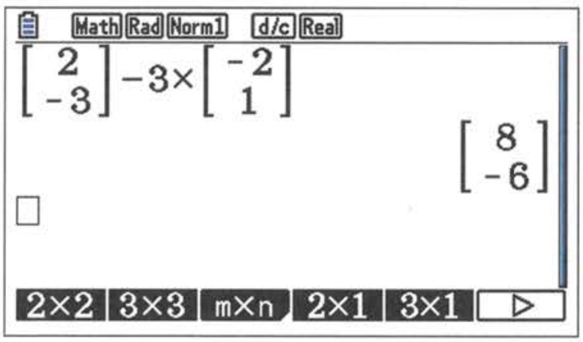
If using Casio, select Module 1.



If using the Run mode (1):

Vectors can be entered as needed and arithmetic performed on them by pressing F4-MATH and F1-MAT/VCT. This provides a screen from which common vector (and matrix) layouts can be accessed and basic operations performed.

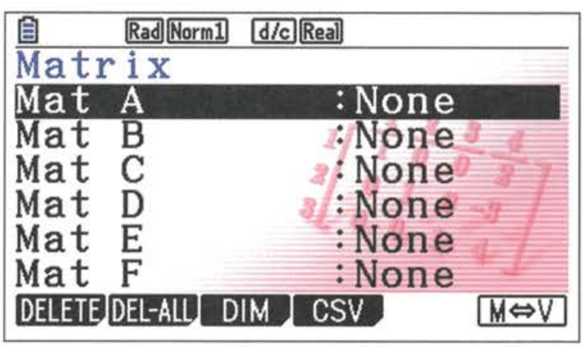
If using 2 by 1 vectors, a blank vector of the right size can be found by pressing F4. The values can now be entered from the keyboard.



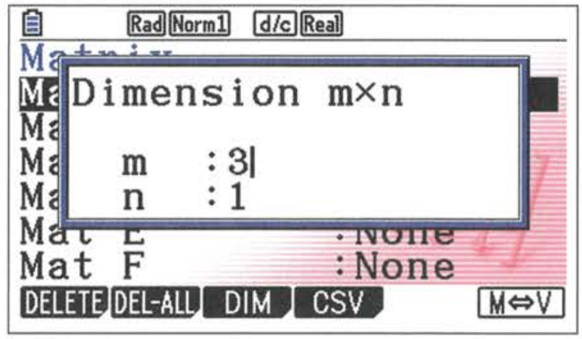
Many applications will make use of the same vectors.

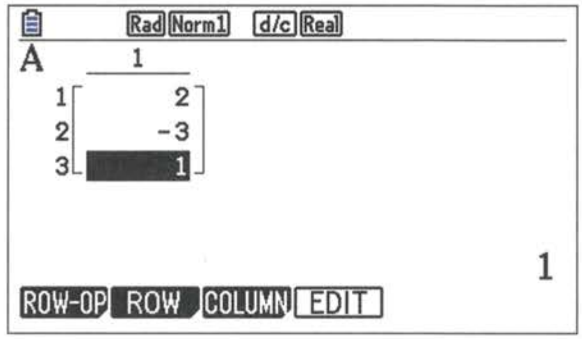
These can be entered (still in Run mode) by using F3-MATH/VCT.

This opens a screen for defining matrices and vectors.



A 3 by 1 vector can now be entered as Mat A

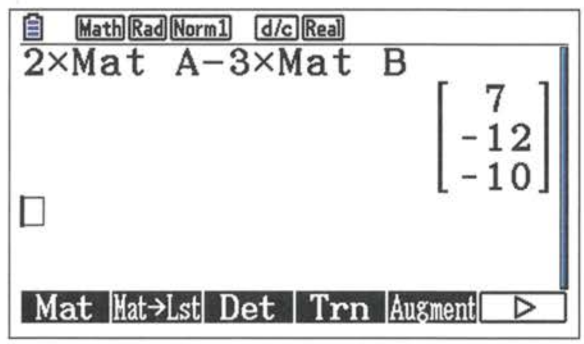




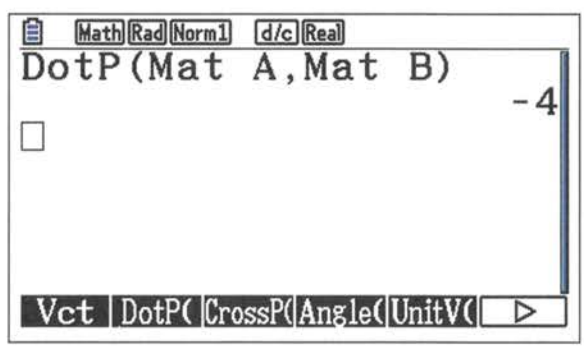
If a second vector is stored $\begin{bmatrix}-1\\2\\4\end{bmatrix}$,

both can be accessed repeatedly to perform calculations.

The vector A is accessed by pressing OPTN, F2, F1 followed by ALPHA A to name the vector.

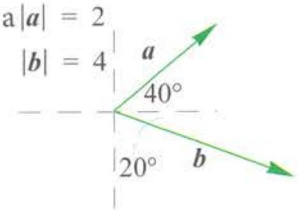


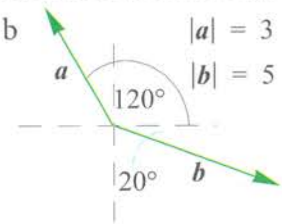
Scalar product calculations can be found by scrolling twice (using F6) to the right and pressing F2.

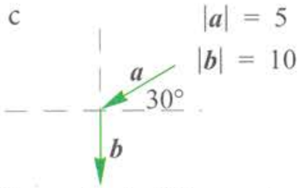


Exercise 4.2.1

1. Find the scalar product, $\boldsymbol{a} \bullet \boldsymbol{b}$, for each of the following:

a


b


c


2. Find the scalar products of these pairs of vectors.

a $3\boldsymbol{i} + 2\boldsymbol{j}$ and $2\boldsymbol{i} + 3\boldsymbol{j}$

b $3\boldsymbol{i} + 7\boldsymbol{j}$ and $2\boldsymbol{i} + 3\boldsymbol{j}$

c $3\boldsymbol{i} - \boldsymbol{j}$ and $-2\boldsymbol{i} + 2\boldsymbol{j}$

d $6\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}$ and $-7\boldsymbol{i} - 4\boldsymbol{j} + 3\boldsymbol{k}$

e $-\boldsymbol{j} + 5\boldsymbol{k}$ and $-4\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}$

f $-\boldsymbol{i} + 5\boldsymbol{j} + 4\boldsymbol{k}$ and $5\boldsymbol{i} - 4\boldsymbol{k}$

g $\begin{pmatrix} 0 \\ 6 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 7 \\ 2 \\ -6 \end{pmatrix}$

h $\begin{pmatrix} -3 \\ -1 \\ 7 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$

i $\begin{pmatrix} -6 \\ -1 \\ 7 \end{pmatrix}$ and $\begin{pmatrix} 7 \\ 3 \\ 5 \end{pmatrix}$

3. Find the angles between these pairs of vectors, giving the answers in degrees, correct to the nearest degree.

a $-4\boldsymbol{i} - 4\boldsymbol{j}$ and $-3\boldsymbol{i} + 2\boldsymbol{j}$

b $\boldsymbol{i} - \boldsymbol{j}$ and $3\boldsymbol{i} + 6\boldsymbol{j}$

c $-4\boldsymbol{i} - 2\boldsymbol{j}$ and $-\boldsymbol{i} - 7\boldsymbol{j}$

d $-7\boldsymbol{i} + 3\boldsymbol{j}$ and $-2\boldsymbol{i} - \boldsymbol{j}$

e $\boldsymbol{i} + 3\boldsymbol{j} + 7\boldsymbol{k}$ and $6\boldsymbol{i} + 7\boldsymbol{j} - \boldsymbol{k}$

f $\boldsymbol{j} + 3\boldsymbol{k}$ and $-\boldsymbol{j} - 2\boldsymbol{k}$

g $\begin{pmatrix} -3 \\ -1 \\ -5 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ 5 \\ -5 \end{pmatrix}$

h $\begin{pmatrix} -2 \\ 7 \\ -7 \end{pmatrix}$ and $\begin{pmatrix} 5 \\ 2 \\ -5 \end{pmatrix}$

4. Two vectors are defined as $\boldsymbol{a} = 2\boldsymbol{i} + x\boldsymbol{j}$ and $\boldsymbol{b} = \boldsymbol{i} - 4\boldsymbol{j}$ Find the value of x if:

a the vectors are parallel.

b the vectors are perpendicular.

5. If $\boldsymbol{a} = 2\boldsymbol{i} - 3\boldsymbol{j} + \boldsymbol{k}$, $\boldsymbol{b} = -\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}$ and $\boldsymbol{c} = \boldsymbol{i} + \boldsymbol{k}$, find, where possible,

a $\boldsymbol{a} \bullet \boldsymbol{b}$

b $(\boldsymbol{a} - \boldsymbol{b}) \bullet \boldsymbol{c}$

c $\boldsymbol{a} \bullet \boldsymbol{b} \bullet \boldsymbol{c}$

d $(\boldsymbol{a} - \boldsymbol{b}) \bullet (\boldsymbol{a} + \boldsymbol{b})$

e $\dfrac{\boldsymbol{a}}{\boldsymbol{c}}$

f $\boldsymbol{b} \bullet \boldsymbol{0}$

6. If $\boldsymbol{a} = 2\boldsymbol{i} - \sqrt{3}\boldsymbol{j}$, $\boldsymbol{b} = \sqrt{3}\boldsymbol{i} - \boldsymbol{j}$ and $\boldsymbol{c} = \boldsymbol{i} + \boldsymbol{j}$, find, where possible:

a $\boldsymbol{a} \bullet (\boldsymbol{b} + \boldsymbol{c}) + \boldsymbol{b} \bullet (\boldsymbol{c} - \boldsymbol{a}) + \boldsymbol{c} \bullet (\boldsymbol{a} - \boldsymbol{b})$

b $(\boldsymbol{b} - \boldsymbol{c}) \bullet (\boldsymbol{c} - \boldsymbol{b}) + |\boldsymbol{b}|^2$

c $2|\boldsymbol{a}|^2 - \sqrt{3}\boldsymbol{c} \bullet \boldsymbol{c}$

d $\sqrt{\dfrac{\boldsymbol{a}}{|\boldsymbol{a}|} + \dfrac{\boldsymbol{b}}{|\boldsymbol{b}|} + \dfrac{\boldsymbol{c}}{|\boldsymbol{c}|}}$

7. Find the value(s) of x for which the vectors $x\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}$ and $x\boldsymbol{i} - 2x\boldsymbol{j} - \boldsymbol{k}$ are perpendicular.

8. P, Q and R are three points in space with coordinates (2, –1, 4), (3, 1, 2) and (–1, 2, 5) respectively. Find angle Q in the triangle PQR.

9. Find the values of x and y if $\boldsymbol{u} = x\boldsymbol{i} + 2y\boldsymbol{j} - 8\boldsymbol{k}$ is perpendicular to both $\boldsymbol{v} = 2\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{w} = 3\boldsymbol{i} + 2\boldsymbol{j} - 4\boldsymbol{k}$.

10. Find the unit vector that is perpendicular to both $\boldsymbol{a} = 3\boldsymbol{i} + 6\boldsymbol{j} - \boldsymbol{k}$ and $\boldsymbol{b} = 4\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}$.

11. Show that, if $\boldsymbol{u}$ is a vector in three dimensions, then $\boldsymbol{u} = (\boldsymbol{u} \bullet \boldsymbol{i})\boldsymbol{i} + (\boldsymbol{u} \bullet \boldsymbol{j})\boldsymbol{j} + (\boldsymbol{u} \bullet \boldsymbol{k})\boldsymbol{k}$.

12.

a Find a vector perpendicular to both $\boldsymbol{a} = -\boldsymbol{i} + 2\boldsymbol{j} + 4\boldsymbol{k}$ and $\boldsymbol{b} = 2\boldsymbol{i} - 3\boldsymbol{j} + 2\boldsymbol{k}$.

b Find a vector perpendicular to $2\boldsymbol{i} + \boldsymbol{j} - 7\boldsymbol{k}$.

13. Show that if $|\boldsymbol{a} - \boldsymbol{b}| = |\boldsymbol{a} + \boldsymbol{b}|$, where $\boldsymbol{a} \neq \boldsymbol{0}$ and $\boldsymbol{b} \neq \boldsymbol{0}$, then $\boldsymbol{a}$ and $\boldsymbol{b}$ are perpendicular.

14. If $\boldsymbol{a} \bullet \boldsymbol{b} = \boldsymbol{a} \bullet \boldsymbol{c}$ where $\boldsymbol{a}$ ¦ $\boldsymbol{0}$ ¦ $\boldsymbol{b}$, what conclusion(s) can be made?

15. Using the scalar product for vectors prove that the cosine of the angle between two lines with direction cosines l_1, m_1, n_1 and l_2, m_2, n_2 is given by $\cos\theta = l_1 l_2 + m_1 m_2 + n_1 n_2$.

16. Find the cosine of the acute angle between:

a two diagonals of a cube.

b the diagonal of a cube and one of its edges.

17.a On the same set of axes sketch the graphs of:

$x + 3y - 6 = 0$ and $2x - y + 6 = 0$,

clearly labelling all intercepts with the axes.

b Find a unit vector along the line:

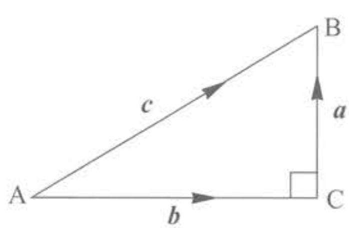
i $x + 3y - 6 = 0$.

ii $2x - y + 6 = 0$.

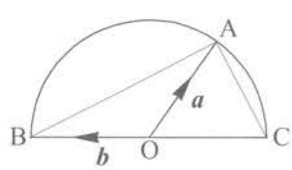
c Hence find the acute angle between the two lines $x + 3y - 6 = 0$ and $2x - y + 6 = 0$.

18. Find a unit vector $\boldsymbol{a}$ such that $\boldsymbol{a}$ makes an angle of 45° with the z-axis and is such that the vector $\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{a}$ is a unit vector.

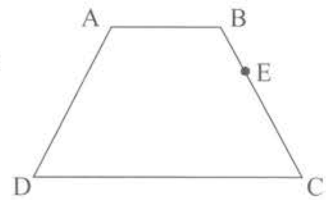
19. Using the scalar product for vectors prove Pythagoras's Theorem for the triangle ABC shown.



20 Prove that an angle inscribed in a semicircle is a right angle.



21. In the trapezium shown, BE:BC = 1:3.



Show that $3\mathbf{AC} \bullet \mathbf{DE} = 2(4m^2 - n^2)$

where $|\mathbf{AB}| = m$, $|\mathbf{DC}| = 2|\mathbf{AB}|$ and $|\mathbf{DA}| = n$

22. Prove that the altitudes of any triangle are concurrent.

23. An oil pipeline runs from a well (W) to a distribution point (D) which is 4 km east and 8 km north of the well. A second well (S) is drilled at a point 9 km east and 7 km south of the distribution point. It is desired to lay a new pipeline from the second well to a point (X) on the original pipeline where the two pipes will be joined. This new pipeline must be as short as possible.

a Set up a suitable vector basis using the first well as the origin.

b Express $\overrightarrow{WD}$, $\overrightarrow{WS}$, $\overrightarrow{DS}$ in terms of your basis.

c Write a unit vector in the direction of $\overrightarrow{WD}$.

d If the point X is d km along the pipeline from the first well, write a vector equal to $\overrightarrow{WX}$.

e Hence find the vector $\overrightarrow{WX}$ such that the new pipeline is as short as possible.

Link to a 3-d visualisation of two vectors, the plane in which they exist and a vector perpendicular to this plane.

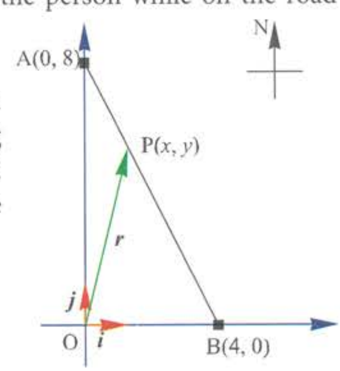
Answers

4.3 Vector Equations

Vector equation of a line in two dimensions

We start this section by considering the following problem:

Relative to an origin O, a house, situated 8 km north of O, stands next to a straight road. The road runs past a second house, located 4 km east of O. If a person is walking along the road from the house north of O to the house east of O, determine the position of the person while on the road relative to O.



We start by drawing a diagram and place the person along the road at some point P. We need to determine the position vector of point P.

We have:

$$\boldsymbol{r} = \mathbf{OP} = \mathbf{OA} + \mathbf{AP}$$

Now, as P lies somewhere along $\overline{\text{AB}}$, we can write:

$\mathbf{AP} = l\,\mathbf{AB}$, where $0 \le \lambda \le 1$, so that when $\lambda = 0$ the person is at A and when $\lambda = 1$ the person is at B.

Next, $\mathbf{AB} = \mathbf{AO} + \mathbf{OB} = -8\boldsymbol{j} + 4\boldsymbol{i}$, and so we have:

$$\boldsymbol{r} = 8\boldsymbol{j} + \lambda(-8\boldsymbol{j} + 4\boldsymbol{i})$$

This provides us with the position vector of the person while walking on the road.

We take this equation a little further. The position vector of P can be written as $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j}$ and so we have that

$$x\boldsymbol{i} + y\boldsymbol{j} = 8\boldsymbol{j} + \lambda(-8\boldsymbol{j} + 4\boldsymbol{i})$$

That is, we have $x\boldsymbol{i} + y\boldsymbol{j} = 4\lambda\boldsymbol{i} + (8 - 8\lambda)\boldsymbol{j}$ meaning that

$$x = 4\lambda \text{ and } y = 8 - 8\lambda$$

The equations $x = 4\lambda$ – (1) and $y = 8 - 8\lambda$ – (2) are known as the parametric form of the equations of a straight line

Next, from these parametric equations, we have $\lambda = \frac{x}{4}$ – (3) and $\lambda = \frac{y-8}{-8}$ – (4)

Then, equating (3) and (4) we have $\frac{x}{4} = \frac{y-8}{-8}$. This equation is known as the Cartesian form of the equation of a straight line. We can go one step further and simplify this last equation.

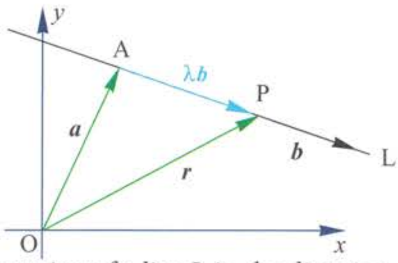
$$\frac{x}{4} = \frac{y-8}{-8} \Leftrightarrow -2x = y - 8 \Leftrightarrow y = -2x + 8$$

which corresponds to the straight line passing through A and B.

This approach to describe the position of an object (or person) is of great value when dealing with objects travelling in a straight line. When planes are coming in for landing, it is crucial that their positions along their flight paths are known, otherwise one plane could be heading for a collision with another plane in the air.

We now formalise the definition of the vector equation of a line in a plane:

The vector equation of a line L in the direction of the vector b, passing through the point A with position vector a is given by r = a + λb where λ is a scalar parameter.



The vector equation of a line L in the direction of the vector $\boldsymbol{b}$, passing through the point A with position vector $\mathbf{a}$ is given by:

$$\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$$

where λ is a scalar parameter.

Proof:

Let the point P(x, y) be any point on the line L, then the vector $\mathbf{AP}$ is parallel to the vector $\boldsymbol{b}$.

$$\begin{aligned} \boldsymbol{r} &= \mathbf{OP} \\ &= \mathbf{OA} + \mathbf{AP} \end{aligned}$$

$$\therefore \boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$$

So the equation of L is given by $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$ as required.

We can now derive two other forms for equations of a line. We start by letting the coordinates of A be (a_1, a_2), the coordinates of P be (x, y) and the vector $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$.

From $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$ we have:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \lambda\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \Leftrightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_1 + \lambda b_1 \\ a_2 + \lambda b_2 \end{pmatrix}$$

This provides us with the:

Parametric form for the equation of a straight line:

$$x = a_1 + \lambda b_1 \quad y = a_2 + \lambda b_2$$

Next, from the parametric form we have:

$$x = a_1 + \lambda b_1 \Leftrightarrow x - a_1 = \lambda b_1 \Leftrightarrow \lambda = \frac{x - a_1}{b_1} \qquad -(1)$$

and $$y = a_2 + \lambda b_2 \Leftrightarrow y - a_2 = \lambda b_2 \Leftrightarrow \lambda = \frac{y - a_2}{b_2} \qquad -(2)$$

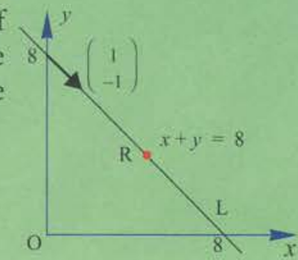
Equating (1) and (2) provides us with the:

Cartesian form for the equation of a straight line:

$$\frac{x - a_1}{b_1} = \frac{y - a_2}{b_2}$$

Example 4.3.1

Find the vector equation of the line L, as shown in the diagram. Comment on the uniqueness of this equation.



The vector equation of the line L is based on finding (or using) *any* point on the line, such as (0,8) , and *any* vector in the direction of the line L, such as $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

The position vector of any point R on the line can then be written as $\boldsymbol{r} = \begin{pmatrix} 0 \\ 8 \end{pmatrix} + \lambda\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

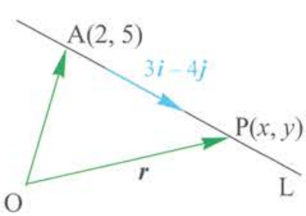
As λ varies, different points on the line are generated, and conversely any point on the line has a corresponding value of λ. For example, substituting $\lambda = 3$ gives the point (3,5) and the point (8,0) corresponds to $\lambda = 8$.

NB: the vector equation (in parametric form) is not unique. The equation $\boldsymbol{r} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} + \lambda\begin{pmatrix} -2 \\ 2 \end{pmatrix}$ is an equally valid description of the line, and in this case substituting $\lambda = 0.5$ generates the point (3,5).

Example 4.3.2

Find the vector equation of the line L, passing through the point A(2, 5) and parallel to the vector $3\boldsymbol{i} - 4\boldsymbol{j}$.

Rather than depend on a standard formula, it is always helpful to visualise problems such as these, in particular, when we move onto straight lines in space. We draw a general representation of this situation and work from there.



Let the point P be any point on the line L with position vector **r**, then $\mathbf{OP} = \mathbf{OA} + \mathbf{AP}$

However, as A and P lie on the line L, then $\mathbf{AP} = \lambda(3\boldsymbol{i} - 4\boldsymbol{j})$.

Therefore, $\boldsymbol{r} = (2\boldsymbol{i} + 5\boldsymbol{j}) + \lambda(3\boldsymbol{i} - 4\boldsymbol{j})$

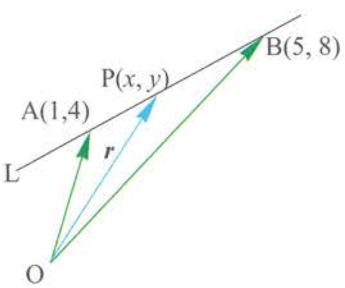
This represents the vector equation of the line L in terms of the parameter λ, where $\lambda \in \mathbb{R}$.

The equation could also be written as, $\boldsymbol{r} = (2 + 3\lambda)\boldsymbol{i} + (5 - 4\lambda)\boldsymbol{j}$

Example 4.3.3

Find the vector equation of the line L, passing through the points A(1, 4) and B(5, 8). Give both the parametric form and Cartesian form of L.

We start with a sketch of the situation described:



Let the point P be any point on the line L with position vector $\boldsymbol{r}$, then

$\mathbf{OP} = \mathbf{OA} + \mathbf{AP}$

Then, as :

$\mathbf{AP} \parallel \mathbf{AB} \Rightarrow \mathbf{AP} = l\,\mathbf{AB}$

where $\lambda \in \mathbb{R}$.

This means that we need to find the vector **AB** which will be the vector parallel to the line L. So, we have

$$\mathbf{AB} = \mathbf{AO} + \mathbf{OB} = -\begin{pmatrix}1\\4\end{pmatrix} + \begin{pmatrix}5\\8\end{pmatrix} = \begin{pmatrix}4\\4\end{pmatrix} = 4\begin{pmatrix}1\\1\end{pmatrix}$$

Therefore, from $\mathbf{OP} = \mathbf{OA} + \mathbf{AP}$ we have

$$\mathbf{OP} = \begin{pmatrix}1\\4\end{pmatrix} + \lambda \times 4\begin{pmatrix}1\\1\end{pmatrix}$$

That is, $\boldsymbol{r} = \begin{pmatrix}1\\4\end{pmatrix} + t\begin{pmatrix}1\\1\end{pmatrix}$ where $t = 4\lambda$

This represents the vector equation of the straight line L.

To find the **parametric form** of L we make use of the equation:

$$\boldsymbol{r} = \begin{pmatrix}1\\4\end{pmatrix} + t\begin{pmatrix}1\\1\end{pmatrix}.$$

As P(x, y) is any point on the line L, we write the vector equation as:

$$\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1\\4\end{pmatrix} + t\begin{pmatrix}1\\1\end{pmatrix}.$$

From where we obtain the parametric equations, $x = 1 + t$ and $y = 4 + t$.

To find the **Cartesian form** of L we now make use of the parametric equations.

From $x = 1 + t$ we have $t = x - 1$ – (1) and from $y = 4 + t$ we have $t = y - 4$ – (2)

Then, equating (1) and (2) we have $x - 1 = y - 4$ (or $y = x + 3$).

Example 4.3.4

The vector equation of the line L, is given by $\boldsymbol{r} = \begin{pmatrix}3 + 2\lambda\\5 - 5\lambda\end{pmatrix}$.

Express the vector equation in the standard form $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$.

Find a unit vector in the direction of L.

Find the Cartesian form of the line L.

$\boldsymbol{r} = \begin{pmatrix}3 + 2\lambda\\5 - 5\lambda\end{pmatrix} = \begin{pmatrix}3\\5\end{pmatrix} + \begin{pmatrix}2\lambda\\-5\lambda\end{pmatrix} = \begin{pmatrix}3\\5\end{pmatrix} + \lambda\begin{pmatrix}2\\-5\end{pmatrix}$ (which is in the form $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$).

The direction of the line L is provided by the vector $\boldsymbol{b}$, i.e. $\begin{pmatrix}2\\-5\end{pmatrix}$.

To find the unit vector we need $\left|\begin{pmatrix}2\\-5\end{pmatrix}\right| = \sqrt{4 + 25} = \sqrt{29}$.

$$\therefore \hat{\boldsymbol{b}} = \frac{1}{\sqrt{29}}\begin{pmatrix}2\\-5\end{pmatrix}.$$

Using the point P(x, y) as representing any point on the line L, we have that $\boldsymbol{r} = \begin{pmatrix}x\\y\end{pmatrix}$.

Therefore, we can write the vector equation as $\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}3 + 2\lambda\\5 - 5\lambda\end{pmatrix}$

From this equation we then have:

$$x = 3 + 2\lambda \text{ – (1) and } \quad y = 5 - 5\lambda \text{ – (2)}$$

We can now find the Cartesian equation by eliminating the parameter λ using (1) and (2).

From (1): $\lambda = \dfrac{x - 3}{2}$.

From (2): $\lambda = \dfrac{y - 5}{-5}$

Therefore, $\dfrac{x - 3}{2} = \dfrac{y - 5}{-5}$.

Example 4.3.5

Find the angle between the lines $\frac{x-2}{4} = \frac{y+1}{3}$ and $\frac{x+2}{-1} = \frac{y-4}{2}$.

1. If finding the angles between two vectors, then the answer can either be acute or obtuse depending on the original arrangement between the two vectors.

2. If finding the angles between two lines, the answer should be stated as an acute angle, since we will "create" two vectors from the lines and hence, depending on how we have created the vectors, the angle may be obtuse or acute.

We must first express the lines in their vector form. To do this we need to introduce a parameter for each line.

Let $\frac{x-2}{4} = \frac{y+1}{3} = \lambda$ giving the parametric equations:

$x = 2+4\lambda$ and $y = -1+3\lambda$.

We can now express these two parametric equations in the vector form:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2+4\lambda \\ -1+3\lambda \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \lambda\begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

This vector equation informs us that the line $\frac{x-2}{4} = \frac{y+1}{3}$ is parallel to the vector $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$.

In the same way we can obtain the vector equation of the line:

$\frac{x+2}{-1} = \frac{y-4}{2}$.

Let $\frac{x+2}{-1} = \frac{y-4}{2} = t$ giving the parametric equations:

$x = -2-t$ and $y = 4+2t$.

From here we obtain the vector equation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2-t \\ 4+2t \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix} + t\begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

This vector equation informs us that the line $\frac{x+2}{-1} = \frac{y-4}{2}$ is parallel to the vector $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

To find the angle between the two lines we use their direction vectors, $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ along with their scalar product:

$$\begin{pmatrix} 4 \\ 3 \end{pmatrix} \bullet \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \left|\begin{pmatrix} 4 \\ 3 \end{pmatrix}\right| \times \left|\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right| \cos\theta$$

$$\Rightarrow -4+6 = \sqrt{16+9} \times \sqrt{1+4}\cos\theta$$

$$\Leftrightarrow \cos\theta = \frac{2}{5\sqrt{5}}$$

$$\therefore \theta \approx 79°42'$$

Exercise 4.3.1

1. For the straight line with equation $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$ where $\boldsymbol{a} = \boldsymbol{i}+2\boldsymbol{j}$ and $\boldsymbol{b} = -2\boldsymbol{i}+3\boldsymbol{j}$, find the coordinates of the points on the line for which:

   i $\lambda = 0$ ii $\lambda = 3$ iii $\lambda = -2$

   Sketch the graph of $\boldsymbol{r} = \boldsymbol{i}+2\boldsymbol{j}+\lambda(-2\boldsymbol{i}+3\boldsymbol{j})$.

2. Find the vector equation of the line passing through the point A and parallel to the vector $\boldsymbol{b}$, where:

   a $A \equiv (2, 5)$, $\boldsymbol{b} = 3\boldsymbol{i}-4\boldsymbol{j}$

   b $A \equiv (-3, 4)$, $\boldsymbol{b} = -\boldsymbol{i}+5\boldsymbol{j}$

   c $A \equiv (0, 1)$, $\boldsymbol{b} = 7\boldsymbol{i}+8\boldsymbol{j}$

   d $A \equiv (1, -6)$, $\boldsymbol{b} = 2\boldsymbol{i}+3\boldsymbol{j}$

   e $A \equiv (-1, -1)$, $\boldsymbol{b} = \begin{pmatrix} -2 \\ 10 \end{pmatrix}$

   f $A \equiv (1, 2)$, $\boldsymbol{b} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$

3. Find a vector equation of the line passing through the points A and B where:

   a A(2, 3), B(4, 8)

   b A(1, 5), B(−2, 1)

   c A(4, −3), B(−1, −2)

4. Find the vector equation of the straight line defined by the parametric equations:

   a $x = 9+\lambda, y = 5-3\lambda$

   b $x = 6-4t, y = -6-2t$

   c $x = -1-4\lambda, y = 3+8\lambda$

   d $x = 1+\frac{1}{2}\mu, y = 2-\frac{1}{3}\mu$

5. Find the parametric form of the straight line having the vector equation:

a $\quad r = \begin{pmatrix}-8\\10\end{pmatrix} + \mu\begin{pmatrix}2\\1\end{pmatrix}$ $\qquad$ b $\quad r = \begin{pmatrix}7\\4\end{pmatrix} + \mu\begin{pmatrix}-3\\-2\end{pmatrix}$

c $\quad r = \begin{pmatrix}5\\3\end{pmatrix} + \frac{\mu}{2}\begin{pmatrix}5\\1\end{pmatrix}$ $\qquad$ d $\quad r = \begin{pmatrix}0.5-0.1t\\0.4+0.2t\end{pmatrix}$

6. Find the Cartesian form of the straight line having the vector equation:

a $\quad r = \begin{pmatrix}1\\3\end{pmatrix} + \mu\begin{pmatrix}3\\1\end{pmatrix}$

b $\quad r = \begin{pmatrix}2\\4\end{pmatrix} - \lambda\begin{pmatrix}7\\5\end{pmatrix}$

c $\quad r = -\begin{pmatrix}2\\4\end{pmatrix} + \lambda\begin{pmatrix}1\\8\end{pmatrix}$

d $\quad r = \begin{pmatrix}0.5\\0.2\end{pmatrix} - t\begin{pmatrix}-1\\11\end{pmatrix}$

e $\quad r = \begin{pmatrix}7\\5\end{pmatrix} + \lambda\begin{pmatrix}0\\1\end{pmatrix}$

7. Write the following lines in vector form:

a $\quad y = \frac{1}{3}x + 2$ $\qquad$ b $\quad y = x - 5$

c $\quad 2y - x = 6$

8. Find the position vector of the point of intersection of each pair of lines:

$r_1 = \begin{pmatrix}2\\1\end{pmatrix} + \begin{pmatrix}\lambda\\3\lambda\end{pmatrix}$ and $r_2 = \begin{pmatrix}1\\3\end{pmatrix} + \mu\begin{pmatrix}1\\2\end{pmatrix}$.

$r_1 = \begin{pmatrix}0\\4\end{pmatrix} + \lambda\begin{pmatrix}2\\5\end{pmatrix}$ and $r_2 = \begin{pmatrix}2\\-2\end{pmatrix} + \mu\begin{pmatrix}1\\1\end{pmatrix}$.

9. Find the equation of the line that passes through the point A (2, 7) and is perpendicular to the line with equation $\boldsymbol{r} = -\boldsymbol{i} - 3\boldsymbol{j} + \lambda(3\boldsymbol{i} - 4\boldsymbol{j})$.

10. Let the position vectors of the points $P(x_1, y_1)$ and $Q(x_2, y_2)$ be $\boldsymbol{p}$ and $\boldsymbol{q}$ respectively.

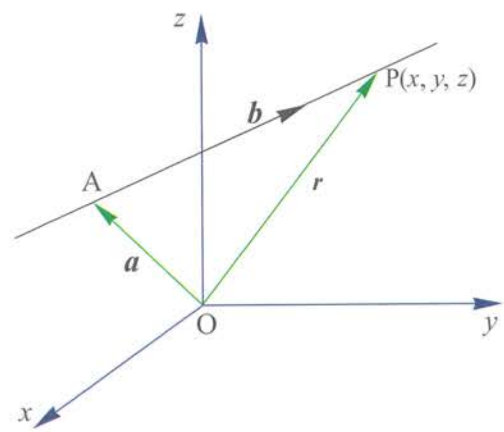
Show that the equation $\boldsymbol{r} = (1-\lambda)\boldsymbol{p} + \lambda\boldsymbol{q}$ represents a vector equation of the line through P and Q, where $\lambda \in \mathbb{R}$.

Extra questions

Lines in three dimensions

In three-dimensional work always try to visualise situations very clearly. Because diagrams are never very satisfactory, it is useful to use the corner of a table with an imagined vertical line for axes; then pencils become lines and books or sheets of paper become planes.

It is tempting to generalise from a two-dimensional line like $x + y = 8$ and think that the Cartesian equation of a three dimensional line will have the form $x + y + z = 8$. This is not correct – as we will see later **this represents a plane, not a line**.



We approach lines in three dimensions in exactly the same way we did for lines in two dimensions. For any point P(x, y, z) on the line having the position vector $\boldsymbol{r}$, passing through the point A and parallel to a vector in the direction of the line, $\boldsymbol{b}$ say, we can write the equation of the line as $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b}$.

So, for example, the line passing through the point (4, 2, 5) and having the direction vector $\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$ can be written as:

$$\boldsymbol{r} = \begin{pmatrix}4\\2\\5\end{pmatrix} + \lambda\begin{pmatrix}1\\-1\\2\end{pmatrix}$$

Or, it could also have been written in $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ form as

$$\boldsymbol{r} = 4\boldsymbol{i} + 2\boldsymbol{j} + 5\boldsymbol{k} + \lambda(\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k})$$

As for the case in 2–D, the parametric form or Cartesian form of the equation is obtained by using a point P(x, y, z) on the line with position vector:

$$\boldsymbol{r} = \begin{pmatrix}x\\y\\z\end{pmatrix} \text{ so that } \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}4\\2\\5\end{pmatrix} + \lambda\begin{pmatrix}1\\-1\\2\end{pmatrix}$$

From here we first get the **parametric equations**:

$x = 4 + \lambda$, $y = 2 - \lambda$ and $z = 5 + 2\lambda$.

Solving each of these for λ, we get:

$$\lambda = x - 4 = 2 - y = \frac{z-5}{4}$$

The parameter λ plays no part in the Cartesian equation, so we drop it and write the **Cartesian equation** as:

$$x - 4 = 2 - y = \frac{z-5}{4}.$$

It is important to be clear what this means: if we choose x, y and z satisfying the Cartesian equation, then the point P(x, y, z) will be on the line.

For example $x = 10$, $y = -4$ and $z = 17$ satisfies the Cartesian equation, and if we think back to our original parametric equation we can see that:

$$\begin{pmatrix} 10 \\ -4 \\ 17 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} + 6\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

To convert a Cartesian equation into parametric form we reverse the process and introduce a parameter λ. For example if the Cartesian equation is:

$\frac{x-1}{3} = \frac{y+2}{2} = \frac{z-6}{4}$ we write:

$$\frac{x-1}{3} = \frac{y+2}{2} = \frac{z-6}{4} = \lambda$$

$$x = 1 + 3\lambda$$

$$y = -2 + 2\lambda$$

$$z = 6 + 4\lambda$$

$$\boldsymbol{r} = \begin{pmatrix} 1 \\ -2 \\ 6 \end{pmatrix} + \lambda\begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}$$

You will probably have noticed the strong connection between the numbers in the fractions in the Cartesian form and the numbers in the vectors in the parametric form.

Consider the **Cartesian form** of any straight L passing through the point $P(x_1, y_1, z_1)$:

$$\frac{x - x_1}{a} = \frac{y - y_1}{b} = \frac{z - z_1}{c}$$

From this equation we obtain the **parametric form** of the straight line:

$$\frac{x - x_1}{a} = \lambda \Rightarrow x = x_1 + \lambda a$$

$$\frac{y - y_1}{b} = \lambda \Rightarrow y = y_1 + \lambda b$$

$$\frac{z - z_1}{c} = \lambda \Rightarrow z = z_1 + \lambda c$$

which then leads to the **vector form** of the straight line:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \lambda\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

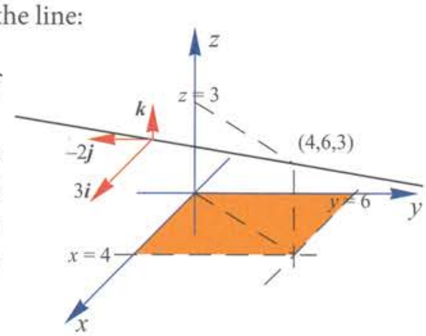
That is, the denominators of the Cartesian form of a straight line provide the coefficients of the directional vector of the line. This is an important observation, especially when finding the angle between two lines when the equation of the line is provided in Cartesian form. However, rather than simply committing this observation to memory, it is always a good idea to go through the (very short) working involved.

Example 4.3.6

Find the Cartesian form of the straight line passing through the point (4, 6, 3) and having direction vector $3\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k}$. Draw a sketch of this line on a set of axes.

We start by sketching the line:

The direction vector of the line is $3\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k}$ and as the line passes through the point (4, 6, 3), the vector equation of the line is given by



$$\boldsymbol{r} = (4\boldsymbol{i} + 6\boldsymbol{j} + 3\boldsymbol{k}) + \lambda(3\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k}) \quad .$$

From the vector equation we obtain the parametric form of the line: $x = 4 + 3\lambda, y = 6 - 2\lambda$ and $z = 3 + \lambda$.

From these equations we have, $\lambda = \frac{x-4}{3}$, $\lambda = \frac{y-6}{-2}$ and $\lambda = \frac{z-3}{1}$

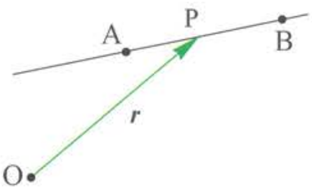
Then, eliminating λ we have $\frac{x-4}{3} = \frac{y-6}{-2} = \frac{z-3}{1}$ or $\frac{x-4}{3} = \frac{y-6}{-2} = z - 3$

which represents the Cartesian form of the line.

Example 4.3.7

Find the vector form of the equation of the line through the point A(2, 1, 1) and the point B(4, 0, 3).

We make a very rough sketch –



there is no point in trying to plot A and B accurately. Let the position vector of any point P on the line be $\boldsymbol{r}$.

Then the vector form of the line is $\boldsymbol{r} = \mathbf{OA} + \lambda\mathbf{AB}$.

Now, $\mathbf{OP} = \boldsymbol{r} = \mathbf{OA} + \mathbf{AP}$.

But $\quad \mathbf{AP} = l\,\mathbf{AB} \therefore \boldsymbol{r} = \mathbf{OA} + l\,\mathbf{AB}$ and

$$\mathbf{AB} = \mathbf{AO} + \mathbf{OB} = -\mathbf{OA} + \mathbf{OB}$$

$$\therefore \mathbf{AB} = -\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} \text{ and so,}$$

$$\boldsymbol{r} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}$$

Example 4.3.8

Find the acute angle between the straight lines

$$L_1: \ \frac{x-3}{2} = \frac{y+2}{-1} = \frac{z}{\sqrt{3}} \text{ and } L_2: \ \frac{x+1}{1} = \frac{y-2}{1} = \frac{z-1}{\sqrt{3}}.$$

Because the lines are given in their standard Cartesian form, we know that the denominators represent the coefficients of the direction vectors of these lines. As the angle between the lines is the same as the angle between their direction vectors we need only use the direction vectors of each line and then apply the dot product.

For L_1 the direction vector is $\boldsymbol{b}_1 = 2\boldsymbol{i} - \boldsymbol{j} + \sqrt{3}\boldsymbol{k}$ and for L_2 it is $\boldsymbol{b}_2 = \boldsymbol{i} + \boldsymbol{j} + \sqrt{3}\boldsymbol{k}$.

Using the dot product we have:

$$\boldsymbol{b}_1 \bullet \boldsymbol{b}_2 = |\boldsymbol{b}_1||\boldsymbol{b}_2|\cos\theta$$

$$\therefore (2\boldsymbol{i} - \boldsymbol{j} + \sqrt{3}\boldsymbol{k}) \bullet (\boldsymbol{i} + \boldsymbol{j} + \sqrt{3}\boldsymbol{k}) = \sqrt{8} \times \sqrt{5}\cos\theta$$

$$2 - 1 + 3 = \sqrt{40}\cos\theta$$

$$\cos\theta = \frac{4}{\sqrt{40}}$$

$$\therefore \theta = 50°46'$$

Example 4.3.9

Write the equation of the line $\dfrac{x+1}{3} = \dfrac{4-y}{2} = z$

in parametric form, and show that it is parallel to

$-\boldsymbol{i} + 5\boldsymbol{j} + \boldsymbol{k} + \mu(-6\boldsymbol{i} + 4\boldsymbol{j} - 2\boldsymbol{k})$.

From the Cartesian form of the line $\dfrac{x+1}{3} = \dfrac{4-y}{2} = z = \lambda$ (say) we obtain the parametric form:

$x = -1 + 3\lambda$, $y = 4 - 2\lambda$ and $z = \lambda$.

We can then write this in the vector form

$\boldsymbol{r} = -\boldsymbol{i} + 4\boldsymbol{j} + \lambda(3\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k})$.

Comparing the direction vectors of the two lines we see that:

$$-6\boldsymbol{i} + 4\boldsymbol{j} - 2\boldsymbol{k} = -2(3\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k})$$

and so the direction vectors (and hence the lines) are parallel.

It is worth emphasising, that lines will be parallel or perpendicular if their direction vectors are parallel or perpendicular.

1. If the two lines are perpendicular we have $\boldsymbol{b}_1 \bullet \boldsymbol{b}_2 = 0 \Rightarrow x_1x_2 + y_1y_2 + z_1z_2 = 0$.

2. If the two lines are parallel we have $\boldsymbol{b}_1 = m\boldsymbol{b}_2, m \neq 0$

Example 4.3.10

Line L passes through the points (4, 3, 9) and (7, 8, 5), while line M passes through the points (12, 16, 4) and (k, 26, –4), where $k \in \mathbb{R}$. Find the value(s) of k, if:

a L is parallel to M.

b L is perpendicular to M.

We first need to determine direction vectors for both L and M.

For L: Let the points be A(4, 3, 9) and B(7, 8, 5), then a direction vector for L,

$\boldsymbol{b}_1$ (for example), is given by $\boldsymbol{b}_1 = \begin{pmatrix} 7-4 \\ 8-3 \\ 5-9 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ -4 \end{pmatrix}$.

For M: Let the points be X(12, 16, 4) and Y(k, 26, –4), then a direction vector for M,

$\boldsymbol{b}_2$ (for example), is given by $\boldsymbol{b}_2 = \begin{pmatrix} k-12 \\ 26-16 \\ -4-4 \end{pmatrix} = \begin{pmatrix} k-12 \\ 10 \\ -8 \end{pmatrix}$.

a If $\mathrm{L} \parallel \mathrm{M}$ we must have that $\boldsymbol{b}_1 = c\boldsymbol{b}_2, c \in \mathbb{R}$.

i.e. $\begin{pmatrix} 3 \\ 5 \\ -4 \end{pmatrix} = c\begin{pmatrix} k-12 \\ 10 \\ -8 \end{pmatrix} \Rightarrow \frac{3}{k-12} = \frac{5}{10} = -\frac{4}{-8}$

So that $\frac{3}{k-12} = \frac{1}{2} \Leftrightarrow k-12 = 6 \Leftrightarrow k = 18$.

b If $\mathrm{L} \perp \mathrm{M}$ we must have that $\boldsymbol{b}_1 \bullet \boldsymbol{b}_2 = 0$.

i.e. $\begin{pmatrix} 3 \\ 5 \\ -4 \end{pmatrix} \bullet \begin{pmatrix} k-12 \\ 10 \\ -8 \end{pmatrix} = 0 \Rightarrow 3(k-12) + 50 + 32 = 0$

$\Leftrightarrow 3k = -46 \Leftrightarrow k = -\frac{46}{3}$

Exercise 4.3.2

1. Find the vector form of the line passing through the point:

   a A(2, 1, 3) which is also parallel to the vector $\boldsymbol{i} - 2\boldsymbol{j} + 3\boldsymbol{k}$.

   b A(2, –3, –1) which is also parallel to the vector $-2\boldsymbol{i} + \boldsymbol{k}$.

2. Find the vector form of the line passing through the points:

   a $A(2, 0, 5)$ and $B(3, 4, 8)$.

   b $A(3, -4, 7)$ and $B(7, 5, 2)$.

   c $A(-3, 4, -3)$ and $B(4, 4, 4)$.

3. Find the Cartesian form of the line having the vector form:

   a $\boldsymbol{r} = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} + s\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$

   b $\boldsymbol{r} = \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix} + t\begin{pmatrix} 5 \\ 0 \\ -2 \end{pmatrix}$ c $\boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

4. Find the Cartesian equation of the line passing through the points A(5, 2, 6) and B(–2, 4, 2). Also, provide the parametric form of this line.

5. For the line defined by the parametric equations $x = 3 + 2t, y = 4 - 3t$ and $z = 1 + 5t$, find the coordinates of where the line crosses the xy-plane.

6. Convert these lines to their parametric form:

   a $\frac{x-2}{3} = y-5 = 2(z-4)$

   b $\frac{2x-1}{3} = y = \frac{4-z}{2}$

   c $\frac{x-3}{-1} = \frac{2-y}{3} = \frac{z-4}{2}$

   d $\frac{2x-2}{4} = \frac{3-y}{-2} = \frac{2z-4}{1}$

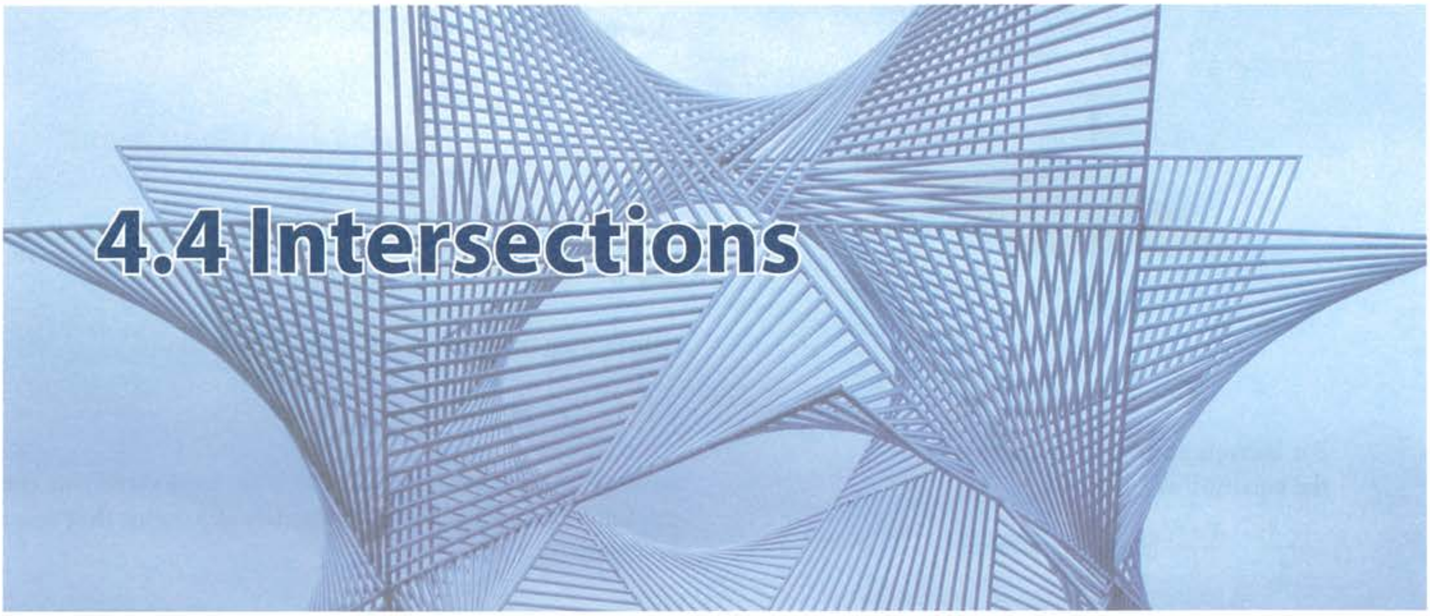
7. Convert these lines to their Cartesian form:

   a $\boldsymbol{r} = \begin{pmatrix} 4 \\ 1 \\ -2 \end{pmatrix} + t\begin{pmatrix} 3 \\ -4 \\ -2 \end{pmatrix}$

   b $\boldsymbol{r} = 2\boldsymbol{i} + \boldsymbol{k} + \mu(\boldsymbol{j} - 3\boldsymbol{k})$

Extra questions

Answers



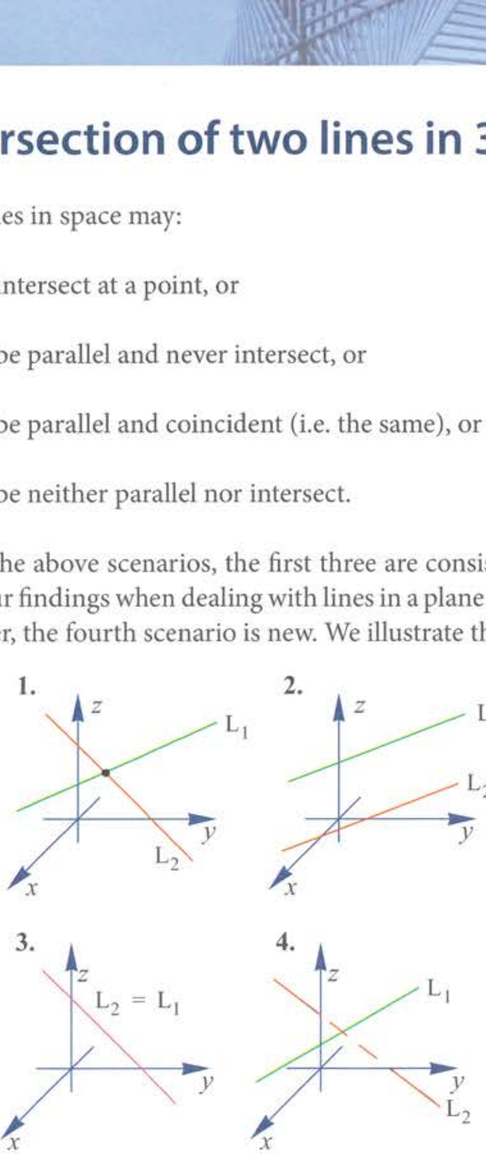
4.4 Intersections

Intersection of two lines in 3–D

Two lines in space may:

1. intersect at a point, or
2. be parallel and never intersect, or
3. be parallel and coincident (i.e. the same), or
4. be neither parallel nor intersect.

Of the above scenarios, the first three are consistent with our findings when dealing with lines in a plane (i.e. 2-D), however, the fourth scenario is new. We illustrate these now.



Two lines that meet at (at least) one point must lie in the same plane (cases **1** and **3**). Two intersecting lines or two parallel lines are said to be coplanar (cases **1**, **2** and **3**). Two lines which are not parallel and which do not intersect are said to be skew – skew lines do not lie on the same plane, i.e. they are not coplanar (case **4**).

Lines lying on the *xy*–, *xz*– and *yz*– planes

From the Cartesian form of the straight line,

L: $\dfrac{x-x_1}{a} = \dfrac{y-y_1}{b} = \dfrac{z-z_1}{c}$ we can write:

$$\frac{x-x_1}{a} = \frac{y-y_1}{b} \Leftrightarrow b(x-x_1) = a(y-y_1) \quad -(1)$$

$$\frac{x-x_1}{a} = \frac{z-z_1}{c} \Leftrightarrow c(x-x_1) = a(z-z_1) \quad -(2)$$

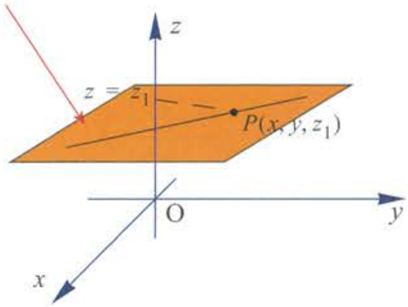
$$\frac{y-y_1}{b} = \frac{z-z_1}{c} \Leftrightarrow c(y-y_1) = b(z-z_1) \quad -(3)$$

Equations (1), (2) and (3) represent the planes perpendicular to the *xy*-, *xz*- and *yz* planes respectively. Each of these equations is an equation of a plane containing L. The simultaneous solution of any pair of these planes will produce the same line. In fact, the three equations are not independent because any one of them can be derived from the other two. If any one of the numbers *a*, *b* or *c* is zero we obtain a line lying in one of the *xy*-, *xz*- or *yz* planes. For example, consider the case that $c = 0$ and neither *a* nor *b* is zero.

In such a case we have, $\dfrac{x-x_1}{a} = \dfrac{y-y_1}{b}$ and $z = z_1$ meaning that the line lies on the plane containing the point $z = z_1$ and parallel to the *xy*-plane.

3-d image showing that skew lines may appear to intersect from some viewpoints.

$$\frac{x-x_1}{a} = \frac{y-y_1}{b} \text{ and } z = z_1$$



For 'convenience' we sometimes write the equation as

$\frac{x-x_1}{a} = \frac{y-y_1}{b} = \frac{z-z_1}{0}$, although

clearly, $\frac{z-z_1}{0}$ has no meaning.

Example 4.4.1

Line L passes through the points A(1, 2, −1) and B(11, −2, −7) while line M passes though the points C(2, −1, −3) and D(9, −10, 3). Show that L and M are skew lines.

We start by finding the vector equations of both lines. For L we have a direction vector given by $\boldsymbol{b}_1 = (11-1)\boldsymbol{i} + (-2-2)\boldsymbol{j} + (-7-(-1))\boldsymbol{k} = 10\boldsymbol{i} - 4\boldsymbol{j} - 6\boldsymbol{k}$.

Then, as L passes through A(1, 2, −1), it has a vector equation given by $\boldsymbol{r} = \boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k} + \lambda(10\boldsymbol{i} - 4\boldsymbol{j} - 6\boldsymbol{k})$

This gives the parametric form as, $x = 1 + 10\lambda, y = 2 - 4\lambda$ and $z = -1 - 6\lambda$ – (1)

Similarly, we can find the parametric form for M.

The vector form of M is given by:

$$\boldsymbol{r} = 2\boldsymbol{i} - \boldsymbol{j} - 3\boldsymbol{k} + \mu(7\boldsymbol{i} - 9\boldsymbol{j} + 6\boldsymbol{k}),$$

so the parametric form is given by $x = 2 + 7\mu, y = -1 - 9\mu$ and $z = -3 + 6\mu$ – (2)

Now, as the set of coefficients of the direction vector of M and L are not proportional, i.e. as $\frac{10}{7} \neq \frac{-4}{-9} \neq \frac{-6}{6}$, the lines L and M are not parallel.

Then, for the lines to intersect, there must be a value of λ and μ that will provide the same point (x_0, y_0, z_0) lying on both L and M. Using (1) and (2) we equate the coordinates and try to determine this point (x_0, y_0, z_0) :

$1 + 10\lambda = 2 + 7\mu$ – (3)

$2 - 4\lambda = -1 - 9\mu$ – (4)

$-1 - 6\lambda = -3 + 6\mu$ – (5)

Solving for λ and μ using (4) and (5) we obtain: $\mu = -\frac{5}{39}$ and $\lambda = \frac{18}{39}$.

Substituting these values into (1), we have

$$\text{L.H.S} = 1 + 10 \times \frac{18}{39} \neq 2 + 7 \times -\frac{5}{39} = \text{R.H.S.}$$

As the first equation is not consistent with the other two, the lines do not intersect and, as they are not parallel, they must be skew.

The techniques we have been discussing can be used to solve problems in particle motion (kinematics).

Example 4.4.2

Particle A moves at a constant velocity from the point (1,2,3) to the point (21,32,23) over a period of 10 seconds. Particle B moves from (5,18,7) to (15,8,17) over the same period. Will the particles collide?

Particle A is translated (over 10 seconds):

$$\begin{pmatrix} 21-1 \\ 32-2 \\ 23-3 \end{pmatrix} = \begin{pmatrix} 20 \\ 30 \\ 20 \end{pmatrix}$$

This represents a velocity vector of $\begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$ (per sec).

The position of particle A is: $\boldsymbol{r}_A = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t\begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$

Particle B is translated (over 10 seconds):

$$\begin{pmatrix} 15-5 \\ 8-18 \\ 17-7 \end{pmatrix} = \begin{pmatrix} 10 \\ -10 \\ 10 \end{pmatrix}$$

This represents a velocity vector of $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ (per sec).

The position of particle B is: $\boldsymbol{r}_B = \begin{pmatrix} 5 \\ 18 \\ 7 \end{pmatrix} + t\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$

If the particles collide, there is a time at which they are in the same position. This means that there is a value of t such that:

$$\begin{pmatrix}1\\2\\3\end{pmatrix}+t\begin{pmatrix}2\\3\\2\end{pmatrix}=\begin{pmatrix}5\\18\\7\end{pmatrix}+t\begin{pmatrix}1\\-1\\1\end{pmatrix}$$

This means that:

$$1+2t=5+t\Rightarrow t=4$$
$$2+3t=18-t\Rightarrow t=4$$
$$3+2t=7+t\Rightarrow t=4$$

and the particles collide after 4 seconds.

Example 4.4.3

Two aircraft are flying in the vicinity of an aerodrome. The positions of the two aircraft relative to the ground radar are given by:

$$r_A=\begin{pmatrix}1\\12\\5\end{pmatrix}+t\begin{pmatrix}2\\-3\\-1\end{pmatrix}\quad r_B=\begin{pmatrix}4\\5\\3\end{pmatrix}+t\begin{pmatrix}-1\\1\\0\end{pmatrix}$$

Distance units are nautical miles and time is in minutes.

Find the vector that represents the separation of the two aircraft. Find the distance of closest approach and when this occurs.

Vector from A to B is:

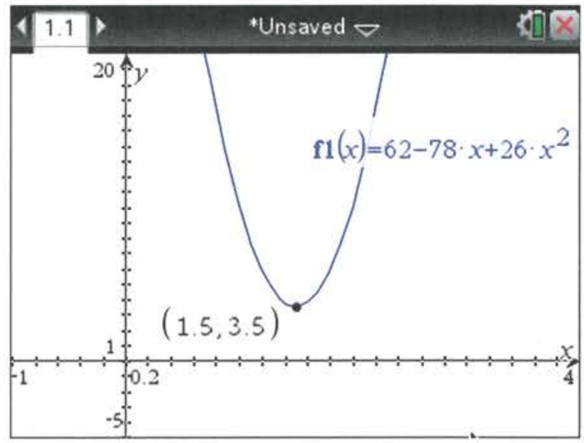
$$\begin{aligned}r_B-r_A&=\begin{pmatrix}4\\5\\3\end{pmatrix}+t\begin{pmatrix}-1\\1\\0\end{pmatrix}-\begin{pmatrix}1\\12\\5\end{pmatrix}-t\begin{pmatrix}2\\-3\\-1\end{pmatrix}\\&=\begin{pmatrix}3\\-7\\-2\end{pmatrix}+t\begin{pmatrix}-3\\4\\1\end{pmatrix}\end{aligned}$$

The distance between the aircraft is the absolute value of this function. We will work with the square of this absolute value:

$$\begin{aligned}|r_B-r_A|^2&=(3-3t)^2+(-7+4t)^2+(-2+t)^2\\&=9-18t+9t^2+49-56t+16t^2+4-4t+t^2\\&=62-78t+26t^2\end{aligned}$$

We can look for the time at which this expression is a minimum. This is because the square root function is one to one and increasing.

We are after the minimum and can use a graph to find it. As with many 'applications' questions, it is necessary to adjust the graph window. We have used Analyze Graph to locate the minimum.



The closest approach occurs at t = 1.5 and is $\sqrt{3.5}$ or about 1.9nm. Note also that one of the pilots will need to pay attention to avoid hitting the ground!

Exercise 4.4.1

1. Find the Cartesian equation of the lines joining the points

   a (−1, 3, 5) to (1, 4, 4)

   b (2, 1, 1) to (4, 1, −1)

2. Find the coordinates of the point where the line:

   a $r=\begin{pmatrix}-2\\5\\3\end{pmatrix}+t\begin{pmatrix}-1\\2\\1\end{pmatrix}$ intersects the x-y plane.

   b The line $\frac{x-3}{4}=y+2=\frac{4-z}{5}$ passes through the point (a, 1, b). Find the values of a and b.

3. Find the Cartesian equation of the line having the vector form:

   a $r=\begin{pmatrix}1\\4\\-2\end{pmatrix}+t\begin{pmatrix}1\\-1\\0\end{pmatrix}$ b $r=\begin{pmatrix}2\\1\\3\end{pmatrix}+t\begin{pmatrix}2\\0\\0\end{pmatrix}$.

   In each case, provide a diagram showing the lines.

4. Find the vector equation of the line represented by the Cartesian form $\frac{x-1}{2}=\frac{1-2y}{3}=z-2$.

   Clearly describe this line.

5. Find the acute angle between the following lines.

   a $r=\begin{pmatrix}0\\2\\3\end{pmatrix}+s\begin{pmatrix}3\\4\\5\end{pmatrix}$ and $r=\begin{pmatrix}-2\\5\\3\end{pmatrix}+t\begin{pmatrix}-1\\2\\1\end{pmatrix}$.

b $\quad \boldsymbol{r} = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix} + s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$ and $\boldsymbol{r} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$

c $\quad \frac{x-3}{-1} = \frac{2-y}{3} = \frac{z-4}{2}$ and

$\frac{x-1}{2} = \frac{y-2}{-2} = z-2$

6. Find the point of intersection of the lines:

a $\quad \frac{x-5}{-2} = y-10 = \frac{z-9}{12}$ - $x = 4, \frac{y-9}{-2} = \frac{z+9}{6}$

b $\quad \frac{2x-1}{3} = \frac{y+5}{3} = \frac{z-1}{-2}$ - $\frac{2-x}{4} = \frac{y+3}{2} = \frac{4-2z}{1}$

7. Find the Cartesian form of the lines with parametric equation given by:

L : $x = \lambda, y = 2\lambda + 2, z = 5\lambda$ and

M : $x = 2\mu - 1, y = -1 + 3\mu, z = 1 - 2\mu$

a $\quad$ Find the point of intersection of these two lines.

b $\quad$ Find the acute angle between these two lines.

Find the coordinates of the point where:

i $\quad$ L cuts the x-y plane.

ii $\quad$ M cuts the y-z plane.

8. Show that the lines $\frac{x-2}{3} = \frac{y-3}{-2} = \frac{z+1}{5}$ and $\frac{x-5}{-3} = \frac{y-1}{2} = \frac{z-4}{-5}$ are coincident.

9. Show that the lines $\frac{x-1}{-3} = y-2 = \frac{7-z}{11}$ and $\frac{x-2}{3} = \frac{y+1}{8} = \frac{z-4}{-7}$ are skew.

10. Find the equation of the line passing through the origin and the point of intersection of the lines with equations

$x-2 = \frac{y-1}{4}, z = 3$ and $\frac{x-6}{2} = y-10 = z-4$.

11. The lines $\frac{x}{3} = \frac{y-2}{4} = 3+z$ and $x = y = \frac{z-1}{2k}$,

$k \in \mathbb{R} \setminus \{0\}$ meet at right angles. Find k.

12. Consider the lines L : $x = 0, \frac{y-3}{2} = z+1$ and M : $\frac{x}{4} = \frac{y}{3} = \frac{z-10}{-1}$.

Find, correct to the nearest degree, the angle between the lines L and M.

13. Find the value(s) of k, such that the lines:

$\frac{x-2}{k} = \frac{y}{2} = \frac{3-z}{3}$ and $\frac{x}{k-1} = \frac{y+2}{3} = \frac{z}{4}$ are perpendicular.

14. Find a direction vector of the line that is perpendicular to both:

$\frac{x+1}{3} = \frac{y+1}{8} = \frac{z+1}{12}$ and $\frac{1-2x}{-4} = \frac{3y+1}{9} = \frac{z}{6}$.

15. Are the lines:

$\frac{x-1}{5} = \frac{y+2}{4} = \frac{4-z}{3}$ and $\frac{x+2}{3} = \frac{y+7}{2} = \frac{2-z}{3}$

parallel? Find the point of intersection of these lines. What do you conclude?

16. Two particles have position vectors:

$$\boldsymbol{r}_A = \begin{pmatrix} 2 \\ 2 \\ 10 \end{pmatrix} + t\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

$$\boldsymbol{r}_B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}$$

Find when the particles collide.

17. Find the point of coincidence of:

$$\boldsymbol{r}_A = \begin{pmatrix} -11 \\ 17 \\ -7 \end{pmatrix} + t\begin{pmatrix} 3 \\ -4 \\ 2 \end{pmatrix}$$

$$\boldsymbol{r}_B = \begin{pmatrix} -1 \\ -5 \\ -3 \end{pmatrix} + t\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

Will the particles collide?

18. Find the closest approach of these two particles:

$$\boldsymbol{r}_A = \begin{pmatrix} -4 \\ -2 \\ -3 \end{pmatrix} + t\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

$$\boldsymbol{r}_B = \begin{pmatrix} 9 \\ 5 \\ 8 \end{pmatrix} + t\begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix}$$

Answers

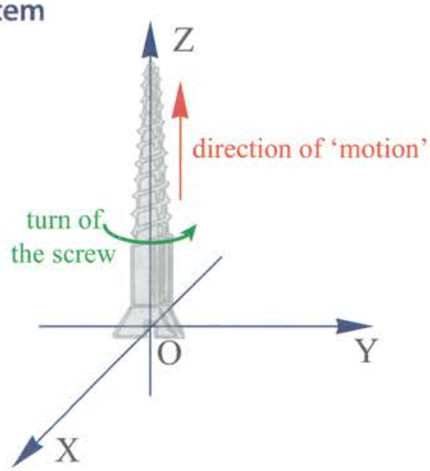
4.5 Vector Product

Three-dimensional Geometry

We start this section by establishing a definition:

Right-handed system

When dealing with three-dimensional space, three base vectors (not coplanar) must be defined. We also conveniently use base vectors that are mutually orthogonal (at right-angles) and which are **right-handed**.



So, what do we mean by right-handed?

If we place a screw at some origin O and rotate it from OX to OY, then the screw would move in the direction OZ. This defines what is known as a **right-handed system**. This definition becomes important when we look at the operation of vector product.

Vector Product

Unlike the scalar product of two vectors, which results in a scalar value, the vector product or as it is often called, the **cross product**, produces a vector.

We define the vector product as follows:

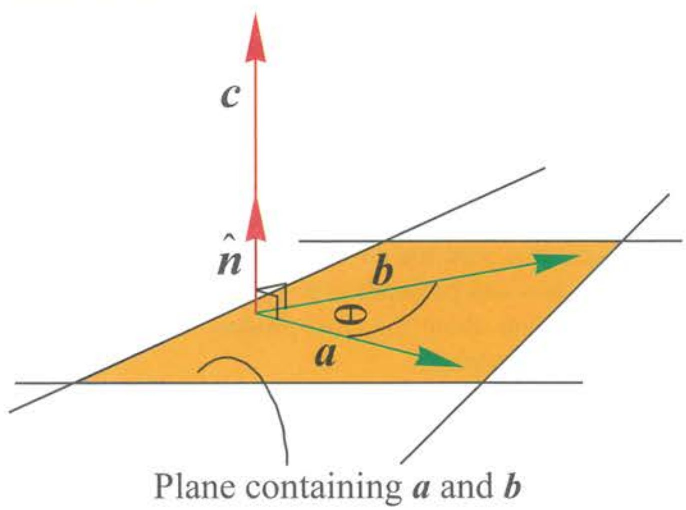
The vector product (or cross product) of two vectors, $\boldsymbol{a}$ and $\boldsymbol{b}$ produces a third vector, $\boldsymbol{c}$, where

$$\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta\hat{\boldsymbol{n}}$$

and θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ and $\hat{\boldsymbol{n}}$ is a unit vector perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e. to the plane of $\boldsymbol{a} \times \boldsymbol{b}$. This means that the vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\hat{\boldsymbol{n}}$ (**in that order**) form a right-handed system.

We now consider some properties of the vector product.

Direction of $\boldsymbol{a} \times \boldsymbol{b}$



The resulting vector, $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$ is a vector that is parallel to the unit vector $\hat{\boldsymbol{n}}$ (unless $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0}$).

The direction of $\hat{\boldsymbol{n}}$ (and hence $\boldsymbol{c}$) is always either:

1. perpendicular to the plane containing $\boldsymbol{a}$ and $\boldsymbol{b}$ which is **determined by the right-hand rule** (as shown in the diagram).

or

2. is the zero vector, $\mathbf{0}$.

Magnitude of $\boldsymbol{a} \times \boldsymbol{b}$

The magnitude of $\boldsymbol{a} \times \boldsymbol{b}$ is given by $|\boldsymbol{a} \times \boldsymbol{b}| = ||\boldsymbol{a}||\boldsymbol{b}|\sin\theta\,\hat{\boldsymbol{n}}|$

$= |\boldsymbol{a}||\boldsymbol{b}||\sin\theta||\hat{\boldsymbol{n}}|$

But, $|\hat{\boldsymbol{n}}| = 1$ and $0 \le \theta \le \pi \Rightarrow \sin\theta \ge 0$, therefore, we have that:

$$|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta$$

Notice that from **1** and **2**, we can also conclude that:

If $\boldsymbol{a} \times \boldsymbol{b} = \mathbf{0}$, then either:

1. $\boldsymbol{a} = \mathbf{0}$ or $\boldsymbol{b} = \mathbf{0}$ or both $\boldsymbol{a}$ and $\boldsymbol{b}$ are $\mathbf{0}$ or

2. $\sin\theta = 0 \Rightarrow \theta = 0$ or π (as $0 \le \theta \le \pi$).

Observation 2, i.e. $\sin\theta = 0 \Rightarrow \theta = 0$ or π, implies that $\boldsymbol{a}$ and $\boldsymbol{b}$ would be either parallel or antiparallel, which would not define a plane and so, the unit vector would not be defined.

This means that for any vector, $\boldsymbol{a}$, $\boldsymbol{a} \times \boldsymbol{a} = \mathbf{0}$, which brings up a very interesting result for our $\boldsymbol{i}$-$\boldsymbol{j}$-$\boldsymbol{k}$ – vector system:

$$\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{k} \times \boldsymbol{k} = \mathbf{0}$$

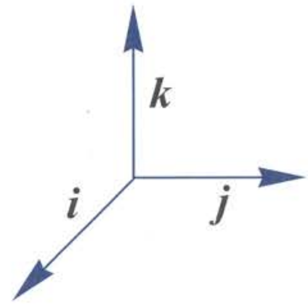
So, unlike the scalar product, where $\boldsymbol{a} \bullet \boldsymbol{a} = |\mathbf{a}|^2 > 0$ for a non-zero vector $\boldsymbol{a}$, with the cross product we have $\boldsymbol{a} \times \boldsymbol{a} = \mathbf{0}$. Also, recall that with the dot product, if the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are non-zero and perpendicular, then $\boldsymbol{a} \bullet \boldsymbol{b} = 0$. So, what can we conclude about the cross product of two non-zero perpendicular vectors?

If the non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are perpendicular then

$\theta = \frac{\pi}{2} \Rightarrow \sin\theta = 1 \therefore \boldsymbol{a} \times \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\hat{\boldsymbol{n}}$.

This means that the magnitude of $|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}||\hat{\boldsymbol{n}}| = |\boldsymbol{a}||\boldsymbol{b}|$.

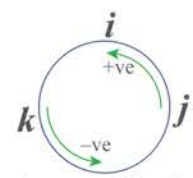
As a result of this property, we have for our $\boldsymbol{i}$-$\boldsymbol{j}$-$\boldsymbol{k}$ – vector system the following results:



$\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}, \boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}, \boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$
and
$\boldsymbol{i} \times \boldsymbol{k} = -\boldsymbol{j}, \boldsymbol{j} \times \boldsymbol{i} = -\boldsymbol{k}, \boldsymbol{k} \times \boldsymbol{j} = -\boldsymbol{i}$

The reason for the negative signs in the above is to ensure consistency within the right-hand system.

So that for example, the vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ (in that order) form a right-hand system as do the vectors $\boldsymbol{i}$, $\boldsymbol{k}$ and $-\boldsymbol{j}$ (in that order). A useful way of remembering which sign applies is to use the cyclic diagram shown:



1. Going **clockwise**, we take the **positive sign**, e.g. $\boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$
2. Going **anticlockwise**, we take the **negative sign**, e.g. $\boldsymbol{j} \times \boldsymbol{i} = -\boldsymbol{k}$

Operational properties

Closure

As $\boldsymbol{a} \times \boldsymbol{b}$ produces a unique vector, then the operation of vector product is closed.

Commutative

As $\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}$ (to conform with the right-hand system) the operation of vector product is **not** commutative.

In fact, because of the change in sign, we say that the vector product is anti-commutative.

Notice also that $|\boldsymbol{a} \times \boldsymbol{b}| = |-\boldsymbol{b} \times \boldsymbol{a}| = |\boldsymbol{b} \times \boldsymbol{a}|$, i.e. the vector $\boldsymbol{a} \times \boldsymbol{b}$ has the same magnitude as $\boldsymbol{b} \times \boldsymbol{a}$ but is in the opposite direction.

Associative

You should try to verify that $(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} \neq \boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c})$ (e.g. use $\boldsymbol{a} = \boldsymbol{i}$, $\boldsymbol{b} = \boldsymbol{j}$ and $\boldsymbol{c} = \boldsymbol{k}$) and so the vector product is **non-associative**.

Distributive

Also, try to verify that $\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c}$ and as such, the vector product **is distributive over addition**.

Identity

No identity element exists for the operation of vector product.

Inverse

No inverse element exists for the operation of vector product.

Exercise 4.5.1

1. For each pair of coplanar vectors, find the magnitude of their cross product.

   a $|\boldsymbol{a}| = 5, |\boldsymbol{b}| = 2$ and the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ is $30°$.

   b $|\boldsymbol{u}| = 1, |\boldsymbol{v}| = 8$ and the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $60°$.

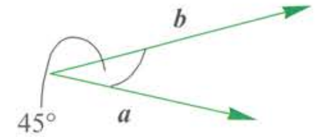
   c $|\boldsymbol{a}| = 3, |\boldsymbol{b}| = 4$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are parallel.

   d $|\boldsymbol{u}| = 0.5, |\boldsymbol{v}| = 12$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular.

   e $|\boldsymbol{a}| = 7, |\boldsymbol{b}| = 3$ and $\boldsymbol{a}$ and $\boldsymbol{b}$ are anti-parallel.

2. Sketch the following cross products for each pair of coplanar vectors:

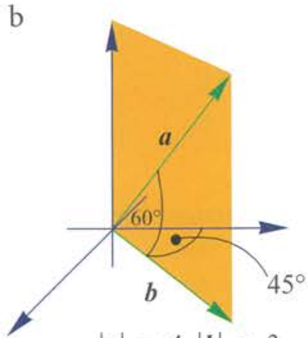
   a



$|\boldsymbol{a}| = 2, |\boldsymbol{b}| = 3$

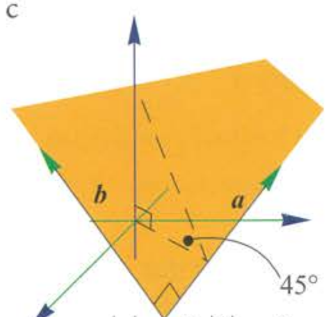
Where $\boldsymbol{a}$ and $\boldsymbol{b}$ can be considered as lying on the surface of an upright table.

   b



$|\boldsymbol{a}| = 4, |\boldsymbol{b}| = 3$

   c



$|\boldsymbol{a}| = 2, |\boldsymbol{b}| = 2$

   i $\boldsymbol{a} \times \boldsymbol{b}$ ii $\boldsymbol{b} \times \boldsymbol{a}$ iii $\boldsymbol{a} \times \boldsymbol{a}$

3.

   a If $|\boldsymbol{a}| = 5$, $|\boldsymbol{b}| = 4$ and $\boldsymbol{a} \bullet \boldsymbol{b} = 6$, find $|\boldsymbol{a} \times \boldsymbol{b}|$.

   b If $|\boldsymbol{a}| = 5$, $|\boldsymbol{b}| = 4$ and $\boldsymbol{a} \bullet \boldsymbol{b} = 12$, find $|\boldsymbol{a} \times \boldsymbol{b}|$

4. If $|\boldsymbol{a}| = 2$, $|\boldsymbol{b}| = 9$ and $|\boldsymbol{a} \times \boldsymbol{b}| = 15$, find the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

5. If $|\boldsymbol{a}| = 3$, $|\boldsymbol{b}| = 3$ and $|\boldsymbol{a} \times \boldsymbol{b}| = 6$, find $\boldsymbol{a} \bullet \boldsymbol{b}$.

6. If $|\boldsymbol{a}| = 1$, $|\boldsymbol{b}| = \sqrt{3}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are mutually perpendicular, find:

   a $|(\boldsymbol{a} + \boldsymbol{b}) \times (\boldsymbol{a} - \boldsymbol{b})|$.

   b $|(2\boldsymbol{a} + \boldsymbol{b}) \times (\boldsymbol{a} - 2\boldsymbol{b})|$.

Vector form of the Vector Product

1. Component form

The vector product is only defined when both vectors are three dimensional.

The vector product of $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ is given by:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}$$

This is known as the **component form** of the cross product. The result is a third vector that is at right angles to the two original vectors. This can be verified by making use of the dot product. Using the 'product' $\boldsymbol{a} \bullet (\boldsymbol{a} \times \boldsymbol{b})$ we have:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \bullet \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}$$

$$= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1)$$
$$= a_1a_2b_3 - a_1a_3b_2 + a_2a_3b_1 - a_2a_1b_3 + a_3a_1b_2 - a_3a_2b_1$$
$$= 0$$

You should check for yourself that the vector product is also perpendicular to the second vector.

Also, notice that in the above diagram, the resulting vector $\boldsymbol{c}$, points in the direction that is consistent with the right-hand rule.

Example 4.5.1

Find the vector product $\begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} \times \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix}$.

$$\begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} \times \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 \times -2 - 1 \times 4 \\ 1 \times -1 - (-2) \times 2 \\ 2 \times 4 - (-1) \times 4 \end{pmatrix} = \begin{pmatrix} -12 \\ 3 \\ 12 \end{pmatrix}$$

Check:

$$\begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} -12 \\ 3 \\ 12 \end{pmatrix} = -24 + 12 + 12 = 0,$$

$$\begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix} \bullet \begin{pmatrix} -12 \\ 3 \\ 12 \end{pmatrix} = 12 + 12 - 24 = 0$$

2. The Determinant form

When vectors are given in base vector notation, a more convenient method of finding the **Vector Cross Product** relies on a determinant representation. Given two vectors $\boldsymbol{a} = a_1\boldsymbol{i} + a_2\boldsymbol{j} + a_3\boldsymbol{k}$ and $\boldsymbol{b} = b_1\boldsymbol{i} + b_2\boldsymbol{j} + b_3\boldsymbol{k}$, the vector product $\boldsymbol{a} \times \boldsymbol{b}$ is defined as:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}\boldsymbol{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}\boldsymbol{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\boldsymbol{k}$$

Applying this to the vectors in Example 4.5.1, where $\boldsymbol{a} = 2\boldsymbol{i} + 4\boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{b} = -\boldsymbol{i} + 4\boldsymbol{j} - 2\boldsymbol{k}$ we have:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & 4 & 1 \\ -1 & 4 & -2 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} 4 & 1 \\ 4 & -2 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 2 & 1 \\ -1 & -2 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 2 & 4 \\ -1 & 4 \end{vmatrix}$$

$$= -12\boldsymbol{i} + 3\boldsymbol{j} + 12\boldsymbol{k}$$

which agrees with our previous answer.

Example 4.5.2

Find $\boldsymbol{a} \times \boldsymbol{b}$ if $\boldsymbol{a} = 2\boldsymbol{i} + \boldsymbol{k}$ and $\boldsymbol{b} = 3\boldsymbol{i} - 4\boldsymbol{j} + 2\boldsymbol{k}$. Hence, find $|\boldsymbol{a} \times \boldsymbol{b}|$.

Using the determinant form of the cross product we have:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & 0 & 1 \\ 3 & -4 & 2 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} 0 & 1 \\ -4 & 2 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 2 & 0 \\ 3 & -4 \end{vmatrix}$$

$$= (0 - (-4))\boldsymbol{i} - (4 - 3)\boldsymbol{j} + (-8 - 0)\boldsymbol{k}$$

$$= 4\boldsymbol{i} - \boldsymbol{j} - 8\boldsymbol{k}$$

Therefore, $|\boldsymbol{a} \times \boldsymbol{b}| = \sqrt{16 + 1 + 64} = \sqrt{81} = 9$

Example 4.5.3

Find the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ if $\boldsymbol{a} = 2\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{b} = 3\boldsymbol{i} - 4\boldsymbol{j} + 2\boldsymbol{k}$.

We first need to determine $\boldsymbol{a} \times \boldsymbol{b}$:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & -1 & 1 \\ 3 & -4 & 2 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} -1 & 1 \\ -4 & 2 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 2 & -1 \\ 3 & -4 \end{vmatrix}$$
$$= 2\boldsymbol{i} - \boldsymbol{j} - 5\boldsymbol{k}$$

Next, $|\boldsymbol{a} \times \boldsymbol{b}| = \sqrt{4+1+25} = \sqrt{30}$.

From $\boldsymbol{a} \times \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta\hat{\boldsymbol{n}}$ we have that $|\boldsymbol{a} \times \boldsymbol{b}| = ||\boldsymbol{a}||\boldsymbol{b}|\sin\theta\hat{\boldsymbol{n}}| = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta$, where θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

$|\boldsymbol{a}| = \sqrt{4+1+1} = \sqrt{6}$ and $|\boldsymbol{b}| = \sqrt{9+16+4} = \sqrt{29}$, so:

$$\sqrt{30} = \sqrt{6} \times \sqrt{29}\sin\theta \Leftrightarrow \sin\theta = \frac{\sqrt{30}}{\sqrt{6} \times \sqrt{29}}$$

$$\therefore \theta \approx 24°32'$$

Of course, it would have been much easier to do Example 4.5.3 using the scalar product!

Example 4.5.4

Find a vector of magnitude 5 units perpendicular to both $\boldsymbol{a} = -2\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{i} - 3\boldsymbol{j} - \boldsymbol{k}$.

The cross product, $\boldsymbol{a} \times \boldsymbol{b}$, will provide a vector that is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$. In fact, it is important to realise that the vector $\boldsymbol{a} \times \boldsymbol{b}$ is perpendicular to the plane that contains the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. This information will be very useful in the next sections, when the equation of a plane must be determined.

Let $\boldsymbol{c}$ be the vector perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$.

$\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b} =$

$$\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -2 & 1 & 1 \\ 1 & -3 & -1 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} 1 & 1 \\ -3 & -1 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} -2 & 1 \\ 1 & -3 \end{vmatrix}$$
$$= (-1+3)\boldsymbol{i} - (2-1)\boldsymbol{j} + (6-1)\boldsymbol{k}$$
$$= 2\boldsymbol{i} - \boldsymbol{j} + 5\boldsymbol{k}$$

However, we want a vector of magnitude 5 units, that is, we want the vector $5\hat{\boldsymbol{c}}$.

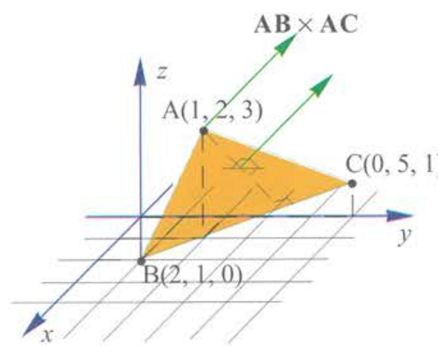
$$\hat{\boldsymbol{c}} = \frac{1}{|\boldsymbol{c}|}\boldsymbol{c} = \frac{1}{\sqrt{4+1+25}}(2\boldsymbol{i} - \boldsymbol{j} + 5\boldsymbol{k}) = \frac{1}{\sqrt{30}}(2\boldsymbol{i} - \boldsymbol{j} + 5\boldsymbol{k}).$$

So, $5\hat{\boldsymbol{c}} = \frac{5}{\sqrt{30}}(2\boldsymbol{i} - \boldsymbol{j} + 5\boldsymbol{k})$.

Example 4.5.5

Find a unit vector that is perpendicular to the plane containing the points A(1, 2, 3), B(2, 1, 0) and C(0, 5, 1).

We start by drawing a diagram of the situation described so that the triangle ABC lies on the planes containing the points A, B and C.



Then, the vector, perpendicular to the plane containing the points A, B and C will be parallel to the vector produced by the cross product $\mathbf{AB} \times \mathbf{AC}$.

Now, $\mathbf{AB} = \mathbf{AO} + \mathbf{OB} = -\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix}$

and $\mathbf{AC} = \mathbf{AO} + \mathbf{OC} = -\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$.

Then,

$$\mathbf{AB} \times \mathbf{AC} = \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix} \times \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \times -2 - 3 \times -3 \\ -3 \times -1 - 1 \times -2 \\ 1 \times 3 - (-1) \times -1 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ 2 \end{pmatrix}$$

Let $\boldsymbol{c} = \mathbf{AB} \times \mathbf{AC}$, $\therefore \hat{\boldsymbol{c}} = \frac{1}{\sqrt{150}}\begin{pmatrix} 11 \\ 5 \\ 2 \end{pmatrix}$.

3-d realisation

Exercise 4.5.2

1. A set of vectors is defined by:

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}, \boldsymbol{c} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix}, \boldsymbol{d} = \begin{pmatrix} -2 \\ -4 \\ 3 \end{pmatrix}$$

Find the vector products:

a $\boldsymbol{a} \times \boldsymbol{b}$ b $\boldsymbol{a} \times \boldsymbol{c}$ c $\boldsymbol{a} \times \boldsymbol{d}$

d $\boldsymbol{b} \times \boldsymbol{c}$ e $\boldsymbol{b} \times \boldsymbol{d}$ f $\boldsymbol{c} \times \boldsymbol{d}$

2. Find a vector that is perpendicular to both:
$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ 6 \\ -2 \end{pmatrix}$.

3. Verify that the vector $\boldsymbol{a} = \boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}$ is perpendicular to the cross product $\boldsymbol{a} \times \boldsymbol{b}$ where $\boldsymbol{b} = 2\boldsymbol{i} - 3\boldsymbol{j} + \boldsymbol{k}$.

4. Verify that if $\boldsymbol{a} = \boldsymbol{i} + 6\boldsymbol{j} - 3\boldsymbol{k}$, $\boldsymbol{b} = -\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{c} = 2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k}$ then

a $\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c}$.

b $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \bullet \boldsymbol{c})\boldsymbol{b} - (\boldsymbol{a} \bullet \boldsymbol{b})\boldsymbol{c}$.

5. If $\boldsymbol{a} = m\boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k}$ and $\boldsymbol{b} = 2\boldsymbol{i} + n\boldsymbol{j} - \boldsymbol{k}$,

a Find: i $\boldsymbol{a} \times \boldsymbol{a}$ ii $\boldsymbol{a} \times \boldsymbol{b}$

b Show that $mn - 4 = 0$ if $\boldsymbol{a} \parallel \boldsymbol{b}$.

6. Find a vector that is perpendicular to both the vectors $\boldsymbol{i} + 6\boldsymbol{j} + 3\boldsymbol{k}$ and $\boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k}$ and has a magnitude of 2.

7. Find a vector that is perpendicular to the plane containing the points:

a A(0, 0, 0), B(0, 5, 0) and C(2, 0, 0).

b A(2, 3, 1), B(2, 6, 2) and C(–1, 3, 4).

8. Using the cross product, find, to the nearest degree, the angle between the vectors:

a $\boldsymbol{u} = 2\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$ and $\boldsymbol{v} = -\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}$.

b $\boldsymbol{a} = 3\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{j} + \boldsymbol{k}$.

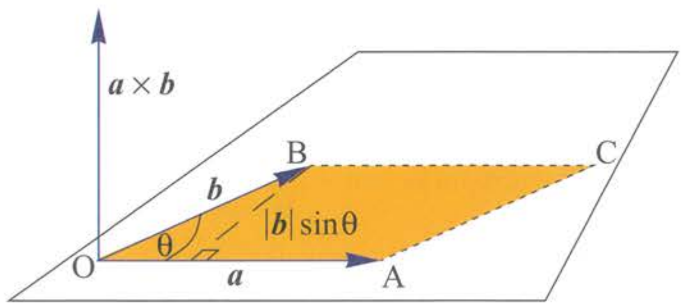
9. Prove that $(\boldsymbol{a} + \boldsymbol{b}) \times (\boldsymbol{a} - \boldsymbol{b}) = 2\boldsymbol{b} \times \boldsymbol{a}$.

Extra questions

Applications of the Vector Product

1. Area

Consider the parallelogram OACB lying on the plane, with the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as shown.



Then, the area of OACB is given by:

$$\begin{aligned} \text{OA} \times |\boldsymbol{b}|\sin\theta &= |\boldsymbol{a}|(|\boldsymbol{b}|\sin\theta) \\ &= |\boldsymbol{a} \times \boldsymbol{b}| \end{aligned}$$

i.e. the **area of the parallelogram** OACB is given by the **magnitude of** the cross product $\boldsymbol{a} \times \boldsymbol{b}$.

We can prove this by using the result $|\boldsymbol{a} \times \boldsymbol{b}|^2 = |\boldsymbol{a}|^2|\boldsymbol{b}|^2 - (\boldsymbol{a} \bullet \boldsymbol{b})^2$ where we replace $\boldsymbol{a} \bullet \boldsymbol{b}$ with $|\boldsymbol{a}||\boldsymbol{b}|\cos\theta$ and then carry through with some algebra. We leave this proof for the next set of exercises.

Example 4.5.6

Find the area of the parallelogram determined by the vectors $\boldsymbol{a} = 2\boldsymbol{i} + \boldsymbol{j} + 3\boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{i} + 4\boldsymbol{j} - \boldsymbol{k}$.

We first need to determine the cross product, $\boldsymbol{a} \times \boldsymbol{b}$:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & 1 & 3 \\ 1 & 4 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 3 \\ 4 & -1 \end{vmatrix}\boldsymbol{i} - \begin{vmatrix} 2 & 3 \\ 1 & -1 \end{vmatrix}\boldsymbol{j} + \begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix}\boldsymbol{k}$$

$$= -13\boldsymbol{i} + 5\boldsymbol{j} + 7\boldsymbol{k}$$

Now,

$|\boldsymbol{a} \times \boldsymbol{b}| = |-13\boldsymbol{i} + 5\boldsymbol{j} + 7\boldsymbol{k}| = \sqrt{169 + 25 + 49} = \sqrt{243}$

3-d realisation

Then, the area of the parallelogram is $\sqrt{243}$ unit^2.

Example 4.5.7

Find the area of the triangle with vertices (1, 6, 3), (0, 10, 1) and (5, 8, 3).

We construct the vectors from the vertex (1, 6, 3) to the vertex (0, 10, 1) and also the vector from (1, 6, 3) to (5, 8, 3).

These vectors are:

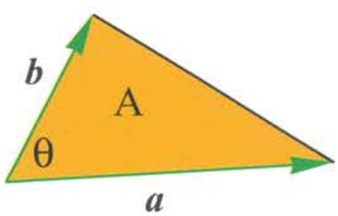
$$\boldsymbol{a} = \begin{pmatrix} 0 \\ 10 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix}$$

and $$\boldsymbol{b} = \begin{pmatrix} 5 \\ 8 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix}.$$

Next, we calculate the vector product:

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{pmatrix} -1 \\ 4 \\ -2 \end{pmatrix} \times \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ -8 \\ -18 \end{pmatrix}$$

Next, using the fact that $|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta$ is a measure of the area of the parallelogram containing the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we can deduce the **area**, A, **of the triangle** containing these vectors to be:



$$A = \frac{1}{2}|\boldsymbol{a}||\boldsymbol{b}|\sin\theta = \frac{1}{2}|\boldsymbol{a} \times \boldsymbol{b}|$$

In this case, the result is:

$$A = \frac{1}{2}\sqrt{4^2 + (-8)^2 + (-18)^2} = \frac{1}{2}\sqrt{404}\text{units}^2 = \sqrt{101}\text{units}^2$$

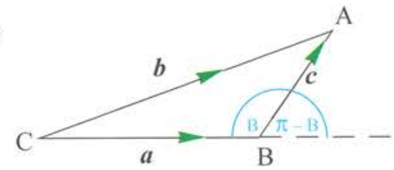
2. Geometric proofs

In the same way that we used the scalar product to neatly prove geometric theorems, for example, proving the cosine rule, we find that the vector product serves just as well for other geometric theorems. We now use the vector product to prove the sine rule.

Example 4.5.8

By making use of the vector product, derive the sine rule, $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$.

Consider the triangle ABC with associated vectors as shown:



From the diagram we have that $\boldsymbol{a} = \boldsymbol{b} - \boldsymbol{c}$.

Then:

$\boldsymbol{a} \times \boldsymbol{a} = \boldsymbol{a} \times (\boldsymbol{b} - \boldsymbol{c})$

But, as $\boldsymbol{a} \times \boldsymbol{a} = \boldsymbol{0}$,

then $\quad \boldsymbol{0} = \boldsymbol{a} \times (\boldsymbol{b} - \boldsymbol{c})$

i.e. $\quad \boldsymbol{0} = \boldsymbol{a} \times \boldsymbol{b} - \boldsymbol{a} \times \boldsymbol{c}$

$$\Rightarrow \boldsymbol{a} \times \boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$$

$$\Rightarrow |\boldsymbol{a} \times \boldsymbol{c}| = |\boldsymbol{a} \times \boldsymbol{b}|$$

$\therefore\ |\boldsymbol{a}||\boldsymbol{c}|\sin(\pi - B) = |\boldsymbol{a}||\boldsymbol{b}|\sin C$

$\therefore\ |\boldsymbol{a}||\boldsymbol{c}|\sin B = |\boldsymbol{a}||\boldsymbol{b}|\sin C$ (as $\sin(\pi - B) = \sin B$)

And so, $|\boldsymbol{c}|\sin B = |\boldsymbol{b}|\sin C$

$$\Leftrightarrow \frac{|\boldsymbol{c}|}{\sin C} = \frac{|\boldsymbol{b}|}{\sin B}$$

That is, $\frac{c}{\sin C} = \frac{b}{\sin B}$.

Similarly, we can prove that $\frac{|\boldsymbol{c}|}{\sin C} = \frac{|\boldsymbol{a}|}{\sin A}$, leading to the results:

$$\frac{a}{\sin A} = \frac{c}{\sin C} = \frac{b}{\sin B}.$$

Exercise 4.5.3

1. Find the area of the parallelogram with adjacent vectors:

   a $2\boldsymbol{i}+\boldsymbol{k}$ and $-\boldsymbol{i}-\boldsymbol{j}+3\boldsymbol{k}$

   b $3\boldsymbol{i}-\boldsymbol{j}+2\boldsymbol{k}$ and $5\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}$

2. A parallelogram has two adjacent sides formed by the vectors:
   $\begin{pmatrix}1\\2\\1\end{pmatrix}$ and $\frac{1}{2}\begin{pmatrix}-14\\1\\-1\end{pmatrix}$.

   a Find the cross product of these two vectors.

   b Find the area of this parallelogram.

   c Hence find the angle between the two vectors.

3. A triangle has vertices (–1, 2, 4), (3, 7, –5) and (4, 2, 3). Find the area of this triangle.

4. Find x, where $x > 0$, if the area of the triangle formed by the adjacent vectors $x\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}$ and $\boldsymbol{j}-\boldsymbol{k}$ is 12 unit^2.

5. Find the area of the triangle with adjacent sides formed by the vectors $2\boldsymbol{i}+3\boldsymbol{j}-4\boldsymbol{k}$ and $2\boldsymbol{i}-3\boldsymbol{j}+4\boldsymbol{k}$. Hence find the angle enclosed by these two vectors.

6. Show that the quadrilateral with vertices at O(4, 1, 0), A(7, 6, 2), B(5, 5, 4) and C(2, 0, 2) is a parallelogram. Hence find its area.

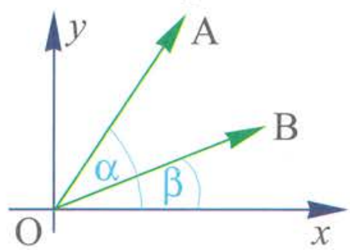
7. Find the area of the parallelogram having diagonals $\boldsymbol{u} = 3\boldsymbol{i}-\boldsymbol{j}+2\boldsymbol{k}$ and $\boldsymbol{v} = \boldsymbol{i}-2\boldsymbol{j}+\boldsymbol{k}$.

8. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are three-dimensional vectors and θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$, use the result that $|\boldsymbol{a}\times\boldsymbol{b}|^2 = |\boldsymbol{a}|^2|\boldsymbol{b}|^2-(\boldsymbol{a}\bullet\boldsymbol{b})^2$ to prove that $|\boldsymbol{a}\times\boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta$.

9. Find, in terms of α and β the vector expressions for:

   a i **OA**

   ii **OB**



   where both **OA** and **OB** are unit vectors.

   b Use the vector product to prove the trigonometric identity $\sin(\alpha-\beta) = \sin\alpha\cos\beta-\sin\beta\cos\alpha$.

10. Let ABCD be a quadrilateral such that its diagonals, [AC] and [BD], intersect at some point O. If triangle ABC has the same area as triangle CBD, show that O is the mid-point of the diagonal [AC].

11. Show that the condition for three points A, B and C to be collinear is that their respective position vectors, $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ satisfy the equation $(\boldsymbol{a}\times\boldsymbol{b})+(\boldsymbol{b}\times\boldsymbol{c})+(\boldsymbol{c}\times\boldsymbol{a}) = \boldsymbol{0}$.

Extra questions

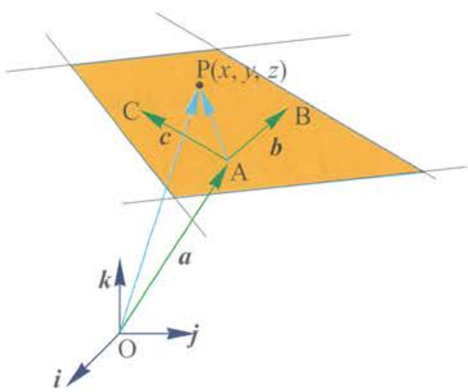
Answers

4.6 Planes

Vector Equation of a Plane

The approach to determine the vector equation of a plane requires only a small extension of the ideas of Section 4,3. In fact, apart from introducing the form that the equation of a plane has, this section has its foundations in Section 4.5.

We begin with the vector equation of a plane.



Let P(x, y, z), whose position vector is $\boldsymbol{r} = \mathbf{OP}$ be any point on the plane relative to some origin O.

Consider three points, A, B and C on this plane where $\mathbf{OA} = \boldsymbol{a}$, $\mathbf{AB} = \boldsymbol{b}$ and $\mathbf{AC} = \boldsymbol{c}$. That is, the plane contains the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$, where $\boldsymbol{b} \mid \mathbf{0} \mid \boldsymbol{c}$ and the vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are non-coplanar.

Now, as $\mathbf{AP}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are coplanar, then we can express $\mathbf{AP}$ in terms of $\boldsymbol{b}$ and $\boldsymbol{c}$: $\mathbf{AP} = \lambda\boldsymbol{b} + \mu\boldsymbol{c}$ for some real λ and μ.

Then, $\boldsymbol{r} = \mathbf{OP} = \mathbf{OA} + \mathbf{AP} = \boldsymbol{a} + \lambda\boldsymbol{b} + \mu\boldsymbol{c}$.

That is, every point on the plane has a position vector of this form.

As such, we say that the vector equation of a plane is given by $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b} + \mu\boldsymbol{c}$

This means that to find the vector form of the equation of a plane we need to know:

1. the position vector of a point A in the plane, and
2. two non-parallel vectors in the plane.

Example 4.6.1

Find the vector equation of the plane containing the vectors:

$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$ which also includes the point (1, 2, 0).

Let $\boldsymbol{b} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{c} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$ be two vectors on the plane.

Then, as the point (1, 2,0) lies on the plane we let $\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ be the position of this point.

Using the vector form of the equation of a plane,

i.e. $\boldsymbol{r} = \boldsymbol{a} + \lambda\boldsymbol{b} + \mu\boldsymbol{c}$, we have $\boldsymbol{r} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \mu\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$.-

Cartesian Equation of a Plane

In the same way that we were able to produce a Cartesian equation for a line in 2-D, we now derive the Cartesian equation of a plane.

Using Example 4.6.1 we obtain the parametric equations and use them to derive the Cartesian equation of the plane.

From the vector equation $\boldsymbol{r} = \begin{pmatrix}1\\2\\0\end{pmatrix} + \lambda\begin{pmatrix}2\\1\\1\end{pmatrix} + \mu\begin{pmatrix}3\\0\\-1\end{pmatrix}$ we obtain

the following parametric equations:

$x = 1 + 2\lambda + 3\mu$ – (1)

$y = 2 + \lambda$......– (2)

$z = \quad \lambda - \mu$ – (3)

Now we find expressions for λ and μ in terms of x, y and z, taking care to use all three equations while doing this:

From (1) and (3) we obtain: $\lambda = \dfrac{x+3z-1}{5}$ – (4)

From (2) and (3) we obtain: $\mu = y - z - 2$ – (5)

Finally we substitute these back into one of the equations. In this particular case it will be easiest to use (4) and (2) - and in fact we didn't need the expression for μ, though in most cases we will.

Substituting (4) into (2) we obtain: $y = 2 + \dfrac{x+3z-1}{5}$

and simplifying we get: $x - 5y + 3z = -9$.

This result tells us that the:

Cartesian form of a plane is given by the equation:

$$ax + by + cz = d$$

Example 4.6.2
Find the Cartesian equation of the plane defined by the vector equation:

$$\boldsymbol{r} = \begin{pmatrix}1\\3\\4\end{pmatrix} + \lambda\begin{pmatrix}-2\\1\\1\end{pmatrix} + \mu\begin{pmatrix}1\\1\\2\end{pmatrix}$$

From the vector equation of the plane, namely:

$$\boldsymbol{r} = \begin{pmatrix}1\\3\\4\end{pmatrix} + \lambda\begin{pmatrix}-2\\1\\1\end{pmatrix} + \mu\begin{pmatrix}1\\1\\2\end{pmatrix},$$

we produce the parametric equations:

$x = 1 - 2\lambda + \mu$ – (1)

$y = 3 + \lambda + \mu$ – (2)

$z = 4 + \lambda + 2\mu$ – (3)

Next, we eliminate λ and μ :(2) – (1): $y - x = 2 + 3\lambda$ – (4)

$2 \times (2) - (3)$: $\quad 2y - z = 2 + \lambda$ – (5)

$(4) - 3 \times (5)$: $\quad -5y - x + 3z = -4$

That is, the Cartesian equation of the plane is given by $-5y - x + 3z = -4$ or $x + 5y - 3z = 4$.

Exercise 4.6.1

1. Find the vector equation of the plane containing the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ and passing through the point A. In each case, draw a rough diagram depicting the situation.

   a $\boldsymbol{b} = 3\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}$, $\boldsymbol{c} = -2\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$, $A \equiv (1, 0, 1)$.

   b $\boldsymbol{b} = \boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$, $\boldsymbol{c} = -\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$, $A \equiv (-1, 2, 1)$.

   c $\boldsymbol{b} = 2\boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k}$, $\boldsymbol{c} = 2\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}$, $A \equiv (4, 1, 5)$.

   d $\boldsymbol{b} = -3\boldsymbol{i} + \boldsymbol{j} - 2\boldsymbol{k}$, $\boldsymbol{c} = \boldsymbol{i} - 2\boldsymbol{j} + \frac{1}{2}\boldsymbol{k}$, $A \equiv (2, -3, -1)$.

2. Find the Cartesian equation for each of the planes in Question **1**.

3. Find the:

   i vector equation.

   ii Cartesian equation of the plane containing the points:

   a A(2, 3, 4), B(–1, 2, 1) and C(0, 5, 6).

   b A(3, –1, 5), B(1, 4, –6) and C(2, 3, 4).

4. A plane contains the vectors $\boldsymbol{b} = 2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k}$ and $\boldsymbol{c} = 3\boldsymbol{i} + \boldsymbol{j} + 2\boldsymbol{k}$.

   a Find the vector equation of the plane, containing the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ and passing through the point:

   i (2, –2, 3).

   ii (0, 0, 0).

   b Find the Cartesian equation for each plane in part **a**.

   c Express $\boldsymbol{b} \times \boldsymbol{c}$ in the form $a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$.

   d What do you notice about the coefficient of x, y and z in part **b** and the values a, b and c from part **c**?

Normal Vector form of a Plane

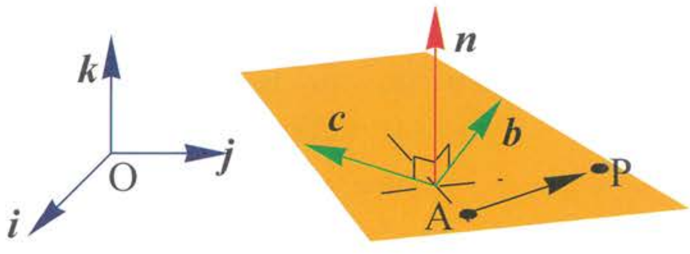
Before we formally derive the **normal vector form of a plane**, we consider an example that follows directly from the work covered so far. In particular, Question **4** from Exercise 4.6.1 – if you have not attempted this problem you should do so now, before proceeding further.

Consider a plane containing the vectors $\boldsymbol{b} = 3\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$ and $\boldsymbol{c} = 2\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}$ and passing through the point A(2, 1, 6). Now, the cross product $\boldsymbol{b} \times \boldsymbol{c}$ represents a vector that is perpendicular to the plane containing the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$.

Let $\boldsymbol{n} = \boldsymbol{b} \times \boldsymbol{c} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 3 & -1 & 2 \\ 2 & 2 & 1 \end{vmatrix} = \begin{vmatrix} -1 & 2 \\ 2 & 1 \end{vmatrix}\boldsymbol{i} - \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}\boldsymbol{j} + \begin{vmatrix} 3 & -1 \\ 2 & 2 \end{vmatrix}\boldsymbol{k}$

$$= -5\boldsymbol{i} + \boldsymbol{j} + 8\boldsymbol{k}$$

We now have a vector, $\boldsymbol{n} = -5\boldsymbol{i} + \boldsymbol{j} + 8\boldsymbol{k}$ that is perpendicular to the plane in question.



3-d realisation - plane and perpendicular vector

Next, consider any point P(x, y, z) on this plane. As P lies on the plane the vector **AP** must also be perpendicular to the vector $\boldsymbol{n}$. This means that $\boldsymbol{n} \bullet \mathbf{AP} = 0$.

To use the equation $\boldsymbol{n} \bullet \mathbf{AP} = 0$ we first need to find the vector **AP**. As **AP** = **AO** + **OP**, we have:

$$\begin{aligned} \mathbf{AP} &= -(2\boldsymbol{i} + \boldsymbol{j} + 6\boldsymbol{k}) + (x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}) \\ &= (x-2)\boldsymbol{i} + (y-1)\boldsymbol{j} + (z-6)\boldsymbol{k} \end{aligned}$$

Then, from $\boldsymbol{n} \bullet \mathbf{AP} = 0$ we have

$$\begin{aligned} (-5\boldsymbol{i} + \boldsymbol{j} + 8\boldsymbol{k}) \bullet ((x-2)\boldsymbol{i} + (y-1)\boldsymbol{j} + (z-6)\boldsymbol{k}) &= 0 \\ \Leftrightarrow -5(x-2) + (y-1) + 8(z-6) &= 0 \\ \Leftrightarrow -5x + y + 8z &= 39 \end{aligned}$$

That is, we have obtained the Cartesian equation of the plane containing the vectors $\boldsymbol{b} = 3\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$ and $\boldsymbol{c} = 2\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}$ and passing through the point A(2, 1, 6) without making use of the parametric form of the plane.

We check this result using the parametric form of the plane.

From the vector for, $\boldsymbol{r} = \begin{pmatrix} 2 \\ 1 \\ 6 \end{pmatrix} + \lambda\begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + \mu\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$

we obtain the parametric equations:

$$x = 2 + 3\lambda + 2\mu \quad -(1)$$

$$y = 1 - \lambda + 2\mu \quad -(2)$$

and $\quad z = 6 + 2\lambda + \mu \quad -(3)$

(1) – (2): $\quad x - y = 1 + 4\lambda \quad -(4)$

(2) – 2 × (3): $\quad y - 2z = -11 - 5\lambda \quad -(5)$

From (4) and (5) we obtain:

$$\frac{x-y-1}{4} = \frac{y-2z+11}{-5} \Leftrightarrow -5x + y + 8z = 39.$$

As expected, we produce the same equation.

To use this method, we require a vector that is perpendicular to the plane and a point that lies on the plane. We could use the vector, $\boldsymbol{n}$ (say) or the unit vector $\hat{\boldsymbol{n}}$, or even $-\boldsymbol{n}$, as they are all perpendicular to the plane.

We can summarise this process as follows:

To find the Cartesian equation of a plane through the point $P_0(x_0, y_0, z_0)$ having a non-zero normal vector $\boldsymbol{n}$ (or $\hat{\boldsymbol{n}}$) we

1. let $P(x, y, z)$ be any point on the plane, and

2. find the vector $\boldsymbol{n} = a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$.

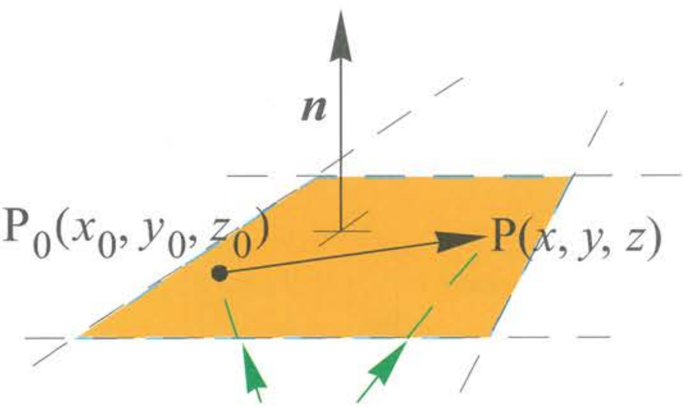
Then, as $\mathbf{P_0P} \perp \boldsymbol{n}$ for all points P on the plane, we have

$$\mathbf{P_0P} \bullet \boldsymbol{n} = 0$$

$$\Rightarrow [(x-x_0)\boldsymbol{i} + (y-y_0)\boldsymbol{j} + (z-z_0)\boldsymbol{k}] \bullet (a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}) = 0$$

$$\therefore a(x-x_0) + b(y-y_0) + c(z-z_0) = 0$$

Or, after some simplifying, $ax + by + cz = d$



Notice that if two planes, Π_1 and Π_2 have normal vectors, $\boldsymbol{n}_1 = a_1\boldsymbol{i} + b_1\boldsymbol{j} + c_1\boldsymbol{k}$ and $\boldsymbol{n}_2 = a_2\boldsymbol{i} + b_2\boldsymbol{j} + c_2\boldsymbol{k}$ respectively, then the two planes, Π_1 and Π_2 are

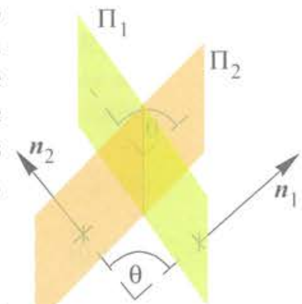
1. parallel iff their normal vectors are parallel, i.e. iff $\boldsymbol{n}_1 = m \times \boldsymbol{n}_2$, where $m \in \mathbb{R}$

   i.e. iff $\dfrac{a_1}{a_2} = \dfrac{b_1}{b_2} = \dfrac{c_1}{c_2} = m$

2. perpendicular iff their normal vectors are perpendicular. i.e. iff $\boldsymbol{n}_1 \bullet \boldsymbol{n}_2 = 0$

   i.e. iff $a_1a_2 + b_1b_2 + c_1c_2 = 0$

Taking this one step further, this result also means that we can use the normals to find the angle between two planes. The angle between two planes is defined as the angle between their normals.



If two planes, Π_1 and Π_2 have normal vectors $\boldsymbol{n}_1 = a_1\boldsymbol{i} + b_1\boldsymbol{j} + c_1\boldsymbol{k}$ and $\boldsymbol{n}_2 = a_2\boldsymbol{i} + b_2\boldsymbol{j} + c_2\boldsymbol{k}$ respectively, and intersect at an acute angle θ (or π – θ depending on their direction), the acute angle θ can be found from the product rule:

$$\boxed{\cos\theta = \frac{|\boldsymbol{n}_1 \bullet \boldsymbol{n}_2|}{|\boldsymbol{n}_1||\boldsymbol{n}_2|}}$$

Example 4.6.3

Find the Cartesian equation of the plane containing the point A(3, 1, 1) and with the normal vector given by $\boldsymbol{n} = 3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}$.

Using the normal vector, $\boldsymbol{n} = 3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}$ and a vector on the plane passing through the point A(3, 1, 1), i.e. the vector $\mathbf{AP} = (x-3)\boldsymbol{i} + (y-1)\boldsymbol{j} + (z-1)\boldsymbol{k}$, where $P(x, y, z)$ is an arbitrary point on the plane, we have

$\boldsymbol{n} \bullet \mathbf{AP} = 0$

$\Rightarrow (3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}) \bullet [(x-3)\boldsymbol{i} + (y-1)\boldsymbol{j} + (z-1)\boldsymbol{k}] = 0$

That is, $3(x-3) + (-2)(y-1) + 4(z-1) = 0$

Or, after some simplification, $3x - 2y + 4z = 11$

Example 4.6.4

Find the angle (to the nearest degree) between the planes with normal vectors $3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}$ and $\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}$.

The angle between the planes corresponds to the angle between their normals. So, using the dot product we have

$$(3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}) \bullet (\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}) = |3\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}||\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}|\cos\theta$$

$$3 + 2 + 12 = \sqrt{29} \times \sqrt{11}\cos\theta$$

$$\therefore \cos\theta = \frac{17}{\sqrt{29} \times \sqrt{11}}$$

And so, we have that $\theta \approx 17°52' = 18°$ (to the nearest degree).

Example 4.6.5

Find the angle (to the nearest degree) between the planes $2x + 3y - 8z = 9$ and $-x + y - 2z = 1$.

To find the angle between the planes we need the normal vectors to the planes. From our observations, we have that a normal vector can be directly obtained from the equation of a plane by using the coefficients of each variable.

For the plane $2x + 3y - 8z = 9$, a normal vector would be $2\boldsymbol{i} + 3\boldsymbol{j} - 8\boldsymbol{k}$ and for the plane $-x + y - 2z = 1$, a normal vector would be $-\boldsymbol{i} + \boldsymbol{j} - 2\boldsymbol{k}$.

Then, we proceed as in Example 4.6.4, using the cosine rule:

$$(2\boldsymbol{i} + 3\boldsymbol{j} - 8\boldsymbol{k}) \bullet (-\boldsymbol{i} + \boldsymbol{j} - 2\boldsymbol{k}) = |2\boldsymbol{i} + 3\boldsymbol{j} - 8\boldsymbol{k}||-\boldsymbol{i} + \boldsymbol{j} - 2\boldsymbol{k}|\cos\theta$$

$$\therefore -2 + 3 + 16 = \sqrt{77} \times \sqrt{6}\cos\theta$$

$$\therefore \cos\theta = \frac{17}{\sqrt{77} \times \sqrt{6}}$$

That is, $\theta \approx 37°44' = 38°$ (to the nearest degree).

Exercise 4.6.2

1. Find the Cartesian equation of the plane containing the point P and having a normal vector, $\boldsymbol{n}$.

   a $\boldsymbol{n} = 2\boldsymbol{i} - \boldsymbol{j} + 5\boldsymbol{k}$, $P \equiv (3, 4, 1)$

   b $\boldsymbol{n} = -4\boldsymbol{i} + 6\boldsymbol{j} - 8\boldsymbol{k}$, $P \equiv (-2, 3, -1)$

   c $\boldsymbol{n} = -\boldsymbol{i} + 3\boldsymbol{j} - 2\boldsymbol{k}$, $P \equiv (2, 4, 5)$

   d $\boldsymbol{n} = 5\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}$, $P \equiv (-1, 2, 1)$

2. Which of the planes in Question **1** pass through the origin?

3. Find the Cartesian equation of the plane containing the points:

   a A(2, 1, 5), B(3, 2, 7) and C(0, 1, 2)

   b A(0, 2, 4), B(1, 2, 3) and C(4, 2, 5)

   c A(1, 1, 7), B(2, –1, 5) and C(–1, 3, 7)

4. Find the angle (to the nearest degree) between the planes with normal vectors:

   a $\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$ and $\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}$.

   b $-3\boldsymbol{i} + 5\boldsymbol{j} - 2\boldsymbol{k}$ and $\boldsymbol{j} + \boldsymbol{k}$.

   c $4\boldsymbol{i} - 2\boldsymbol{j} + 7\boldsymbol{k}$ and $2\boldsymbol{i} + 11\boldsymbol{j} + 2\boldsymbol{k}$.

   d $-3\boldsymbol{i} + 2\boldsymbol{j} - 4\boldsymbol{k}$ and $9\boldsymbol{i} - 6\boldsymbol{j} + 8\boldsymbol{k}$.

5. Find the angle between the planes:

   a $\Pi_1 : -x + 3y - z = 9$ and $\Pi_2 : 6x + 2y + 3z = 4$

   b $\Pi_1 : 2x + 2y - 3 = z$ and $\Pi_2 : 2y - 3z + 2 = 0$

   c $\Pi_1 : 2x - y + 3z = 2$ and $\Pi_2 : 2x + y - 7z = 8$

6. Find the equation of the plane which passes through the point A(4, 2, 1) and:

   a contains the vector joining the points B(3, –2, 4) and C(5, 0, 1).

   b is perpendicular to the planes with equation $5x - 2y + 6z + 1 = 0$ and $2x - y - z = 4$.

7. Find the equation of the plane which passes through the point A(–1, 2, 1) and is parallel to the plane $x - 2y + 3z + 2 = 0$.

8. Find the equation of the plane which passes through the point A(–1, 2, 1) and is parallel to the plane $2y - 3 = 3x + 5z$.

9. The planes $4x - y + 6z = -5$ and $ax + by - z = 7$ are perpendicular. If both planes contain the point (1, 3, –1), find a and b.

10.

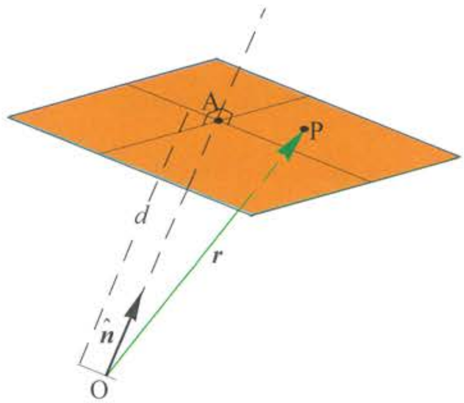
    a Find a vector equation of the line passing through the points (3, 2, 1) and (5, 7, 6).

    b Find the normal vector of the plane $3x + 2y + z = 6$.

    c Hence, find the inclination that the line $\frac{x-3}{2} = \frac{y-2}{5} = \frac{z-1}{5}$ makes with the plane $3x + 2y + z = 10$.

The Normal Form

We now formalise (or at least give a complete vectorial presentation for) the equation of a plane in three dimensions. The good news is that the normal form of the vector equation of a plane in three dimensions develops in almost the same way as the vector equation of a line in two and three dimensions.



Let $\hat{n}$ be a (unit) vector from O normal to the plane and d be the distance of the plane from the origin.

The condition for a point P to be on the plane is that **OA** is perpendicular to **AP**.

That is, $\mathbf{OA} \bullet \mathbf{AP} = 0$

Now, $\mathbf{AP} = \mathbf{AO} + \mathbf{OP} = -d\hat{\boldsymbol{n}} + \boldsymbol{r}$

So that $\quad d\hat{\boldsymbol{n}} \bullet (-d\hat{\boldsymbol{n}} + \boldsymbol{r}) = 0$

Now, dividing by d (assumed to be non-zero)

we have: $\hat{\boldsymbol{n}} \bullet (-d\hat{\boldsymbol{n}} + \boldsymbol{r}) = 0$

$$\therefore -d\hat{\boldsymbol{n}} \bullet \hat{\boldsymbol{n}} + \hat{\boldsymbol{n}} \bullet \boldsymbol{r} = 0$$

$$\Rightarrow \hat{\boldsymbol{n}} \bullet \boldsymbol{r} = d\hat{\boldsymbol{n}} \bullet \hat{\boldsymbol{n}}$$

$$\therefore \hat{\boldsymbol{n}} \bullet \boldsymbol{r} = d \text{ (as } \hat{n} \bullet \hat{n} = 1\text{)}$$

That is, the normal vector form of the equation of a plane is given by $\hat{\boldsymbol{n}} \bullet \boldsymbol{r} = d$.

If we are using $\boldsymbol{n}$ (not a unit vector) the equation becomes $\boldsymbol{n} \bullet \boldsymbol{r} = D$, where D is no longer the distance of the plane from the origin.

If we know the position vector $\boldsymbol{a}$ of a point on the plane we can write the equation as:

$$\boldsymbol{r} \bullet \boldsymbol{n} = \boldsymbol{a} \bullet \boldsymbol{n}$$

For example $\boldsymbol{r} \bullet \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 8$ is the equation of a plane.

We can get this into a Cartesian form by noting that $\boldsymbol{r}$ is the position vector of some arbitrary point P(x, y, z) on the plane and so we can write the vector expression as:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 8 \text{, or } x + y + z = 8.$$

Converting from Cartesian to vector form:

$2x - y + 4z = 2$ becomes $\boldsymbol{r} \bullet \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = 2$.

If we want to get the equation in $\hat{\boldsymbol{n}}$ form, i.e. in the form $\hat{\boldsymbol{n}} \bullet \boldsymbol{r} = d$ we can work out that the length of the vector

$\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}$ is $\sqrt{2^2 + (-1)^2 + 4^2} = \sqrt{21}$, and so, from the equation

$\boldsymbol{r} \bullet \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = 2$ we divide both sides by $\sqrt{21}$ to get:

$$\frac{1}{\sqrt{21}}\boldsymbol{r} \bullet \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = \frac{1}{\sqrt{21}} \times 2 \text{ or } \boldsymbol{r} \bullet \frac{1}{\sqrt{21}}\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = \frac{2}{\sqrt{21}}.$$

We then get the information that the distance of the plane from the origin is $\frac{2}{\sqrt{21}}$.

Example 4.6.6

Show that the line $\boldsymbol{r} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + k\begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}$ is parallel to the plane $x - 2y + 2z = 11$.

We need to prove that $\boldsymbol{n}$ is perpendicular to $\boldsymbol{v}$.

Rewriting $x - 2y + 2z = 11$ in the normal vector form, we have:

$$\boldsymbol{r} \bullet \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} = 11$$

From this equation, a suitable $\boldsymbol{n}$ is the vector $\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$.

From the vector equation of the line, the direction vector of $\boldsymbol{v}$ is:

$$\begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}$$

As $\begin{pmatrix}1\\-2\\2\end{pmatrix} \bullet \begin{pmatrix}4\\3\\1\end{pmatrix} = 4-6+2 = 0$, the vectors are perpendicular.

So the line and plane are parallel.

Example 4.6.7

Show that the line $r = \begin{pmatrix}2\\1\\0\end{pmatrix} + s\begin{pmatrix}5\\1\\-1\end{pmatrix}$ lies in the plane $x - 3y + 2z = -1$.

From the vector equation of the line we obtain the parametric equations:

$$x = 2 + 5s$$

$$y = 1 + s$$

and $$z = -s$$

If this line lies on the plane, then the parametric equations must satisfy the Cartesian equation of the plane. Substituting, into the equation $x - 3y + 2z = -1$, we get

$$\begin{aligned} \text{L.H.S} = x - 3y + 2z &= (2+5s) - 3(1+s) - 2s \\ &= 2 + 5s - 3 - 3s - 2s \\ &= -1 \end{aligned}$$

= R.H.S - Therefore, the line lies in the plane.

Exercise 4.6.3

For this set of exercises, where appropriate, make use of the normal vector form to solve the questions.

1. Convert these planes to Cartesian and vector form:

   a $r = \begin{pmatrix}1\\1\\-4\end{pmatrix} + \lambda\begin{pmatrix}2\\2\\3\end{pmatrix} + \mu\begin{pmatrix}-3\\0\\-1\end{pmatrix}$

   b $r = \begin{pmatrix}2\\1\\1\end{pmatrix} + \lambda\begin{pmatrix}2\\1\\0\end{pmatrix} + \mu\begin{pmatrix}0\\0\\1\end{pmatrix}$

2. Given A(1, 1, 0), B(2, 1, 3) and C(1, 2, –1), find the Cartesian equation of the plane containing A, B and C. (Find a parametric form first by taking A as the point in the plane and **AB** and **AC** as the two vectors in the plane.)

3. Re-solve Question **2** by taking the Cartesian form as $x + by + cz = d$, then calculating b, c and d (simultaneous equations in three unknowns).

4. Show that the line $x + 1 = \frac{y+2}{3} = \frac{4-z}{4}$ and the plane $5x + y + 2z = 20$ are parallel.

5. Find the distance of each of these planes from the origin (i.e. find d):

   a $2x - 3y + 6z = 21$

   b $2x - y + 2z = 5$

   c $x + y - 3z = 11$

   d $4x + 2y - z = 20$

6. Find the equation of the plane through (1, 2, 3) parallel to $3x + 4y - 5z = 0$.

7. Find the equation of the plane through the three points (1, 1, 0), (1, 2, 1) and (–2, 2, –1).

8. Show that the four points (0, –1, 0), (2, 1, 1), (1, 1, 1) and (3, 3, 2) are coplanar.

9. Find the equation of the plane through (2, –3, 1) normal to the line joining (3, 4, –1) and (2, –1, 5).

Applications

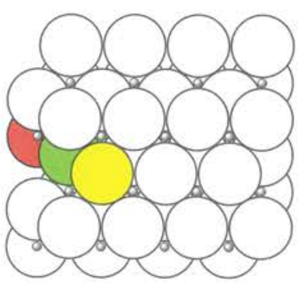
Crystals

The beautifully regular shapes of crystals arise naturally when molten minerals solidify or when solutions are concentrated by evaporation.

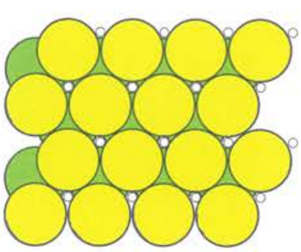
The regular shapes occur when the atoms (ions, molecules) 'close-pack' to form arrangements like a stack of tennis balls in a sports shop.

The techniques discussed in this section should enable you to investigate the shapes that arise when identical spheres form such crystals.

There are two types of stack:



ABC
3rd layer directly above
the hollows ○ in the 1st layer **ABC**



ABAB
3rd layer directly above
the hollows ● in the 2nd layer **ABA**

What two crystal forms result from these two arrangements?

Planes

The sections 4.1 to 4.7 deal with planes and lines in three dimensional space.

The following are three dimensional realisations of some of the situations discussed in these sections.

Two parallel planes

Two intersecting planes.

Note that the intersection is a straight edge.

Two intersecting planes

- and the angle between them

Choose a point on the edge where the planes intersect.

Draw two lines in each plane from this point and at right angles to the edge. Note that these lines may not be parallel to the edges of the planes.

The angle (we have shown the acute option) between these lines is the angle between the planes.

Answers

4.7 Intersections

Intersection of Two Lines

This topic was dealt with in detail in Section 4.4, however, we review it here.

In general, two lines (in three dimensions) will not intersect, but in certain circumstances they may. We can show, for example, that the lines:

$$r = \begin{pmatrix}-1\\4\\0\end{pmatrix} + \lambda\begin{pmatrix}3\\-2\\1\end{pmatrix} \text{ and } r = \begin{pmatrix}4\\4\\-1\end{pmatrix} + \mu\begin{pmatrix}2\\-3\\2\end{pmatrix}$$

do intersect, and we can find their point of intersection.

We show that there exist values of λ and μ which make the x-, y- and z- coordinates of the two lines identical. If we compare the x- and y-coordinates we get:

$$\begin{aligned} -1+3\lambda &= 4+2\mu \\ 4-2\lambda &= 4-3\mu \end{aligned}$$

We can solve these to get $\lambda = 3$ and $\mu = 2$. The point that will decide whether the two lines intersect is:

when $\lambda = 3$ and $\mu = 2$, are the z-coordinates also equal?

This can be tested: $\lambda = 3$ and $\mu = 2$, l has z-coordinate $= 0 + \lambda = 3$ and m has z-coordinate $= -1 + 2\mu = 3$. So the lines do intersect.

Substituting $\lambda = 3$ and $\mu = 2$ in the expressions for the x- and y-coordinates we find that the point of intersection is $(8, -2, 3)$. If the z-coordinates had been different, we would deduce that the lines do not intersect.

Recall that lines which do not intersect and are not parallel (a situation we looked at in section 4.4) are said to be ***skew***.

Exercise 4.7.1

1. a Show that the lines $r_A = 5i + j + k + \lambda(i + 2j - 2k)$ and $r_B = 11i + 4j - 2k + \mu(4i - j + k)$ intersect, and find their point of intersection.

   b By considering the scalar product $(i + 2j - 2k) \bullet (4i - j + k)$, show that the lines from part **a** intersect at right angles.

2. Given the lines:

   a $r = \begin{pmatrix}5\\2\\3\end{pmatrix} + \kappa\begin{pmatrix}-1\\3\\5\end{pmatrix}$

   b $r = \begin{pmatrix}1\\2\\1\end{pmatrix} + \lambda\begin{pmatrix}2\\-2\\3\end{pmatrix}$

   c $r = \begin{pmatrix}5\\-6\\2\end{pmatrix} + \mu\begin{pmatrix}1\\1\\4\end{pmatrix}$,

   find the two lines that intersect. Find also the coordinates of the point of intersection and the acute angle between the two lines.

3. Show that the line joining $(1, 4, 3)$ to $(7, -5, -6)$ intersects the line

   $$\frac{x-1}{2} = -y = \frac{3-z}{3},$$

   and find the point of intersection. (Find a parametric form for each line – remember to use a different parameter for each line.)

4. Show that the three lines:

L: $x = y+4 = \frac{z}{2}+1$ M: $\frac{x-1}{3} = 2y+1 = z-5$

N: $\frac{x}{4} = y+1 = \frac{z-3}{3}$

intersect at a single point, and give its coordinates.

Intersection of a Line and a Plane

In section 4.6 we considered the case of a line and a plane being parallel, and the case of a line lying in a plane. If neither of these happens then the line and plane must intersect in a point.

The angle between a line and a plane is defined as the angle between the line and its projection on the plane. To find the angle between a line and a plane we look at the vectors $\boldsymbol{n}$ (perpendicular to the plane) and $\boldsymbol{v}$ (in the direction of the line):

We can find angle ϕ from the formula $\cos\phi = \frac{\boldsymbol{v} \bullet \boldsymbol{n}}{|\boldsymbol{v}||\boldsymbol{n}|}$ then subtract from 90° to find θ.

Alternatively we can use the fact that $\cos\phi = \sin\theta$ to write directly $\sin\theta = \frac{\boldsymbol{v} \bullet \boldsymbol{n}}{|\boldsymbol{v}||\boldsymbol{n}|}$.

Example 4.7.1

Find the point of intersection of the line $\frac{x}{2} = \frac{y+6}{2} = 3z-1$ and the plane $3x+y-z=9$. Find also the angle between the line and the plane.

Introducing a parameter λ, we have the parametric equations:

$x = 2\lambda$, $y = 2\lambda-6$ and $z = \frac{\lambda+1}{3}$.

Substituting each of these values into the equation of the plane $3x+y-z = 9$ we obtain:

$$6\lambda+(2\lambda-6)-\frac{\lambda+1}{3} = 9$$

i.e. $18\lambda+6\lambda-18-(\lambda+1) = 27$

$$\therefore \lambda = 2$$

Substituting $\lambda = 2$, we get $x = 4$, $y = -2$ and $z = 1$, i.e. the point of intersection is $(4, -2, 1)$.

Writing the equation of the plane as $\boldsymbol{r} \bullet \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix} = 9$

and the equation of the line as $\boldsymbol{r} = \begin{pmatrix} 0 \\ -6 \\ \frac{1}{3} \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 2 \\ \frac{1}{3} \end{pmatrix}$,

we have that $\boldsymbol{n} = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} 2 \\ 2 \\ \frac{1}{3} \end{pmatrix}$.

Then $\boldsymbol{v} \bullet \boldsymbol{n} = 6+2-\frac{1}{3} = 7\frac{2}{3}$, $|\boldsymbol{v}| = \sqrt{8\frac{1}{9}}$ and $|\boldsymbol{n}| = \sqrt{11}$.

Hence $\cos\phi = 0.81165....$, $\phi = 35.7°$ and finally $\theta = 54.3°$.

Exercise 4.7.2

1. In each case find:

   i the point of intersection of the line and plane, and

   ii the angle between the line and plane:

	line	plane
a	$\boldsymbol{i}+2\boldsymbol{j}+\lambda(3\boldsymbol{i}+\boldsymbol{j}+\boldsymbol{k})$	$\boldsymbol{r} \bullet (2\boldsymbol{i}+4\boldsymbol{j}-\boldsymbol{k}) = 28$
b	$\frac{x-1}{2} = y = \frac{3-z}{4}$	$2x+3y+z = 11$
c	$\begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} + \kappa \begin{pmatrix} -1 \\ 3 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}$
d	$\frac{x-1}{2} = \frac{y-2}{-3} = \frac{z+3}{4}$	$2x+4y-z-1 = 0$

2.

   a A line joins the origin to (6, 10, 8). Find the coordinates of the point where the line cuts the plane $2x+2y+z = 10$.

   b Find the point where the line joining (2, 1, 3) to (4, −2, 5) cuts the plane $2x+y-z = 3$.

3. Try to describe with words and/or diagrams:

   a the plane $x + y = 6$.

   b the line $x = 4, y = 2z$.

   Now find their point of intersection.

4. Find the distance of the point (–1, –5, –10) from the point of intersection of the line $\frac{x-2}{3} = \frac{y+1}{4} = \frac{z-2}{12}$ and the plane $x - y + z = 5$.

Intersection of Two Planes

A full treatment of solving simultaneous equations in three unknowns is provided in Section 1.9. We revisit this area using the development of 3-D geometry that has evolved over this chapter and Chapter 24.

If two planes are parallel they will clearly not intersect (unless they coincide), and this case will be identifiable because their respective $\boldsymbol{n}$ vectors will be parallel. For example the planes $2x - y - z = 3$ and $-4x + 2y + 2z = 7$ are parallel because their respective $\boldsymbol{n}$ vectors are $2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k}$ and $-4\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}$ and $-4\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k} = -2(2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k})$. If two planes are not parallel they must intersect in a line.

Example 4.7.2

Find the equation of the line of intersection of the planes

$x + 3y + z = 5$ (1)

$2x - y - z = 1$ (2).

Find also the angle between the two planes.

Our strategy is to eliminate z and hence write x in terms of y.

Adding (1) and (2): $3x + 2y = 6$ and so $x = \frac{6-2y}{3}$.

Now we eliminate y and write x in terms of z.

Adding (1) to 3 × (2): $7x - 2z = 8$ and so $x = \frac{2z+8}{7}$.

Putting these together into a single equation we have the line $x = \frac{6-2y}{3} = \frac{2z+8}{7}$.

Note: having found the line it is worth choosing a simple-valued point on the line, such as (2, 0, 3), and checking that it lies on both planes – which in this case it does.

To find the angle between the planes we find the angle between their normal vectors.

Rewriting the equations as $\boldsymbol{r} \bullet (\boldsymbol{i} + 3\boldsymbol{j} + \boldsymbol{k}) = 5$ and $\boldsymbol{r} \bullet (2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k}) = 1$ we can calculate:

$$(\boldsymbol{i} + 3\boldsymbol{j} + \boldsymbol{k}) \bullet (2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k}) = -2$$

$$|(\boldsymbol{i} + 3\boldsymbol{j} + \boldsymbol{k})| = \sqrt{11}$$

$$|(2\boldsymbol{i} - \boldsymbol{j} - \boldsymbol{k})| = \sqrt{6}$$

Hence $\cos\theta = \frac{-2}{\sqrt{66}}$ and $\theta = 104.3°$.

If the acute angle was required it would be (180° – 104.3°) = 75.7°.

Exercise 4.7.3

1. Where possible, find a Cartesian equation of the line of intersection of the two planes and find the acute angle between them:

   a $x + y + z = 3$ and $2x + y + 3z = 0$

   b $2x + y + 4z = 7$ and $-x + 3y + z = -8$

   c $\boldsymbol{r} = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix} + p\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + q\begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix}$ and

   $\boldsymbol{r} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix} + \mu\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$

   d $\boldsymbol{r} \bullet (3\boldsymbol{i} + 2\boldsymbol{j} + \boldsymbol{k}) = 10$ and $\boldsymbol{r} \bullet (\boldsymbol{i} - 4\boldsymbol{j} - 2\boldsymbol{k}) = 8$

2.

   a Show that the point (5, 2, –1) lies on the line of intersection of the planes $x - 3y + z = -2$ and $2x + y + 3z = 9$.

   b Show that the line of intersection of the planes $x + y + z = 2$ and $2x - y + 3z = -4$ is perpendicular to $x = y = z$.

   c Show that the equation of the line of intersection of the planes $4x + 4y - 5z = 12$ and $8x + 12y - 13z = 32$ can be written as $\frac{x-1}{2} = \frac{y-2}{3} = \frac{z}{4}$.

3. Find the angle between the lines defined by the intersection of the planes:

$$\begin{cases} x-2y+z = 0 \\ x+y-z = 0 \end{cases} \text{ and } \begin{cases} x+2y+z = 0 \\ 8x+12y+5z = 0 \end{cases}.$$

Intersection of Three Planes

Case 1

When we write the equations of three planes such as:

$$x+y+2z = 0 \qquad (1)$$

$$2x-y+z = -6 \qquad (2)$$

$$3x+4y-z = -6 \qquad (3)$$

and consider their possible intersection, we are solving a system of equations in three unknowns, as already covered in Chapter 1. There are three possible outcomes:

1. a single solution
2. no solution
3. an infinity of solutions.

Before reading on it is worth playing with three planes (books, pieces of card) and trying to get a clear picture of the geometrical interpretation of each of these possibilities.

If M is the underlying 3×3 matrix of the system, in our case

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 3 & 4 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ -6 \\ -6 \end{bmatrix},$$

$\det \mathrm{M} \neq 0$ leads to outcome (i) and $\det \mathrm{M} = 0$ leads either to (ii) or to (iii).

$$\begin{aligned} \det \mathrm{M} &= 1(-1\times-1-4\times1)-1(2\times-1-3\times1)+2(2\times4-3\times-1) \\ &= 24 \end{aligned}$$

which means a unique solution, i.e. a single point of intersection.

To find this point we could eliminate z from (1) and (3), then from (2) and (3):

$$(1)+2(3) \qquad 7x+9y = -12$$

$$(2)+(3) \qquad 5x+3y = -12$$

and then solve. We get $x = -3$ and $y = 1$, and by going back to (1) we find $z = 1$. Hence the point of intersection is $(-3, 1, 1)$.

(There is considerable freedom as to which variable to eliminate and how to set about eliminating it.)

Case 2

Now we look at a case where $\det \mathrm{M} = 0$ but there is no solution - i.e. the planes have no common point. Such a system is:

$$3x+y+4z = 8 \qquad (1)$$

$$3x-y-z = 4 \qquad (2)$$

$$x+y+3z = 2 \qquad (3)$$

We set off in the same way as in Case (1): by eliminating one of the variables in two different ways. For this system the obvious variable to eliminate is y:

$$(1)+(2) \qquad 6x+3z = 12$$

$$(2)+(3) \qquad 4x+2z = 6$$

The first equation is equivalent to $2x+z = 4$ and the second is equivalent to $2x+z = 3$. The equations are inconsistent with each other and there is no solution to the system. The three dimensional picture is of three planes that have no point of intersection.

Case 3

In this system check that $\det \mathrm{M} = 0$:

$$3x-y-z = 1 \qquad (1)$$

$$x+2y+z = 4 \qquad (2)$$

$$x-5y-3z = -7 \qquad (3)$$

We could eliminate x in two ways:

$$3\times(2)-(1) \qquad 7y+4z = 11$$

$$(2)-(3) \qquad 7y+4z = 11.$$

It is important to be clear what this means: if we choose *any* y and z satisfying $7y+4z = 11$ we can find the value of x such that all three equations (1, 2 and 3) are satisfied. An example would be $y = z = 1$, leading to $x = 1$; check that all three equations are satisfied. But if we chose to satisfy $7y+4z = 11$ with $y = 5, z = -6$ we get $x = 0$, and again all three equations are satisfied.

Clearly we could find as many solutions as we wanted.

Solution is $\left(\frac{\lambda+6}{7}, \frac{11-4\lambda}{7}, \lambda\right)$.

To summarise: if det M = 0 there are two possibilities.

a When we eliminate one of the variables in two different ways and we get two inconsistent equations in the other two variables, then we have no solution. The three dimensional picture of this is three planes that fail to intersect.

b When we eliminate one of the variables in two different ways and we get two identical equations in the other two variables, then we have an infinity of solutions. The three dimensional picture of this is three planes intersecting in a line. (To find the equation of the line, find the equation of the line of intersection of any two of the planes.)

Exercise 4.7.4

1. Three planes can fail to have any point of intersection if two or more of them are parallel.

   Describe a situation where three planes fail to intersect but no pair of planes is parallel.

2. Analyse Case 2 in a little more detail:

   a Find a Cartesian equation of the line of intersection of $3x+y+4z = 8$ and $3x-y-z = 4$.

   b Show that this line is parallel to $x+y+3z = 2$.

3. Analyse Case 3 in a little more detail:

   a Find a Cartesian equation of the line of intersection of $3x-y-z = 1$ – (1) and $x+2y+z = 4$ – (2).

   b Show that this line lies in the plane $x-5y-3z = -7$ – (3).

   c Show that (1) = 2 × (2) + (3).

4. Classify each set of planes as:

   i intersecting in a single point, in which case give its coordinates, or

   ii no point of intersection, or

   iii intersecting in a line, in which case give a Cartesian equation.

   a $$\begin{aligned} x+y-z &= 10 \\ 2x-3y+z &= 5 \\ x-4y+2z &= 6 \end{aligned}$$

   b $$\begin{aligned} x+y+z &= 10 \\ 2x-y &= 9 \\ -x+3y+4z &= 14 \end{aligned}$$

   c $$\begin{aligned} x+2y-z &= 10 \\ 3x-y+z &= 11 \\ 2x+y+4z &= -1 \end{aligned}$$

   d $$\begin{aligned} 2x+y+3z &= -5 \\ x-2y+2z &= -9 \\ 3x+4y+4z &= -1 \end{aligned}$$

5. *This question involves concepts from the whole of this chapter.*

   OBCDEFGH is a cuboid with O(0, 0, 0); B(0, 0, 3); C(4, 0, 3); D(4, 0, 0); E(4, 2, 0); F(0, 2, 0); G(0, 2, 3); H(4, 2, 3).

   a Sketch the cuboid.

   b Find parametric forms for the equations of lines OH and BE. Show that the two lines intersect at the point (2, 1, 1.5).

   c Find the Cartesian equation of plane FHD. (A parametric form is $\boldsymbol{r} = \mathbf{OF} + s\mathbf{FH} + t\mathbf{FD}$. Now convert to Cartesian form.)

   d Find the coordinates of the point of intersection of line BE and plane FHD, and also the angle between the line and plane.

   e Find the angle between plane FHD and plane GHCB.

6. Show that the equations: $$\begin{aligned} x+y-z &= -1 \\ 5x+3y+z &= 3 \\ 2x+y+z &= a \end{aligned}$$
   are inconsistent for $a = 1$ and describe this situation geometrically in terms of intersecting planes.

7. Find the value of k for which the system of equations:

$$8x+3y+z=12$$
$$x+2z=3$$
$$2x+y-z=k$$

represents three planes that intersect in a common line and find the vector equation in parametric form of the line of intersection.

8. The planes $x-3y-z=0$ and $3x-5y-z=0$ intersect in a line, L, that passes through the origin.

a Find the vector product of the normals to both planes.

b Hence, find the vector equation of L.

c Find the value of k for which the system of equations:

$$x-3y-z=0$$
$$3x-5y-z=0$$
$$-x+ky+2z=k^2-4$$

has:

i no real solutions.

ii infinitely many solutions.

iii a unique solution.

9.

a On a set of axes, sketch the planes $x+y=2a, y+z=2b, z+x=2c$.

b Find where the planes meet, i.e. solve the system of equations:

$$x+y=2a$$
$$y+z=2b$$
$$z+x=2c$$

c Hence, deduce the solution to the system: $x+y=\frac{2}{a}, y+z=\frac{2}{b}, z+x=\frac{2}{c}$

10.

a Find the two values of k for which the planes with equations $-x+y+2z=3$, $kx+y-z=3k$ and $x+3y+kz=13$ have no unique solution.

b Show that for one value of k, there are in fact no solutions.

c Show that for the other value of k, the planes meet along a line. Find the Cartesian equation of this line.

11. Show that the equation for the plane passing through the point $M(x_0, y_0, z_0)$ and perpendicular to the planes $a_1x+b_1y+c_1z=d_1$ and $a_2x+b_2y+c_2z=d_2$ can be written in the form:

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix}=0.$$

12. Show that the equation for the plane passing through the points $M(x_0, y_0, z_0)$, $N(x_1, y_1, z_1)$ and perpendicular to the plane $ax+by+cz=d$ can be written in the form:

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ x_1-x_0 & y_1-y_0 & z_1-z_0 \\ a & b & c \end{vmatrix}=0.$$

13. Show that the equation for the plane passing through the point $M(x_0, y_0, z_0)$ and parallel to the straight lines:

$$L_1: \boldsymbol{r}=\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}+\lambda\begin{pmatrix} l_1 \\ m_1 \\ n_1 \end{pmatrix} \text{ and } L_2: \boldsymbol{r}=\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}+t\begin{pmatrix} l_2 \\ m_2 \\ n_2 \end{pmatrix}$$

may be written in the form $\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix}=0.$

14. Show that the equation for the plane which contains the lines

$$L_1: \boldsymbol{r}=\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}+\lambda\begin{pmatrix} l \\ m \\ n \end{pmatrix} \text{ and } L_2: \boldsymbol{r}=\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}+t\begin{pmatrix} l \\ m \\ n \end{pmatrix}$$

may be written in the form

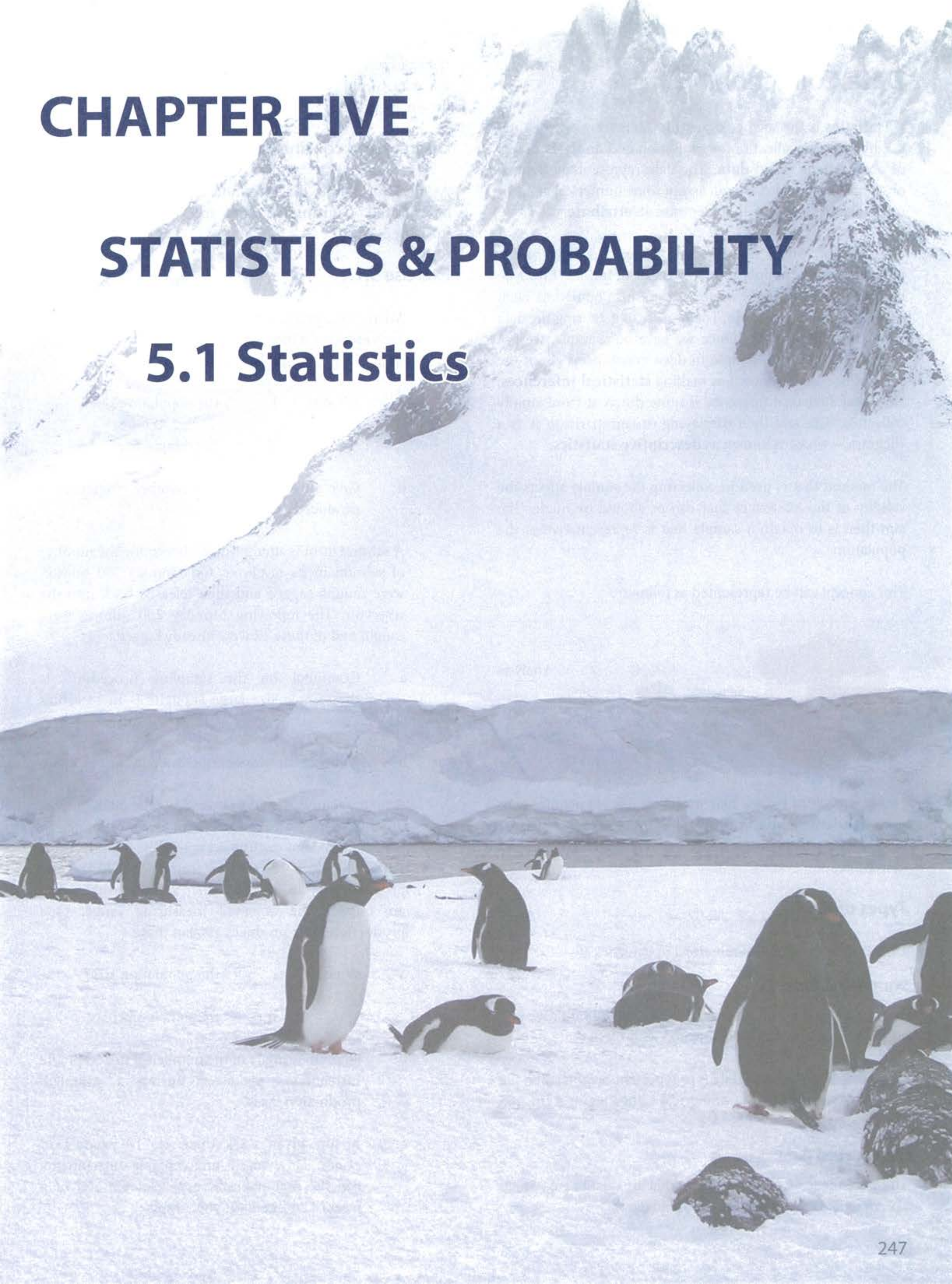
$$\begin{vmatrix} x-a_1 & y-b_1 & z-c_1 \\ a_2-a_1 & b_2-b_1 & c_2-c_1 \\ l & m & n \end{vmatrix}=0.$$

Answers

CHAPTER FIVE

STATISTICS & PROBABILITY

5.1 Statistics



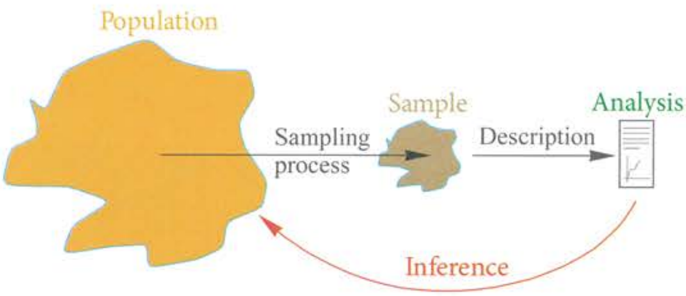
Data Collection

Statistics is the science of getting 'facts from figures' and involves the collection, organisation and **analysis** of sets of observations called **data**. The data represents individual observations to which we can assign some numerical value or categorise in some way that identifies an **attribute**.

The set from which the data is collected is known as the **population**. The population can be finite or infinite, however, the data collected will be a subset from this population. Such a subset is called a **sample**. The process of gathering the data is known as **sampling**. Once we have our sample, we use characteristics of the sample to draw conclusions about the population. This is known as making **statistical inferences**. Note that statistical inference is quite different from simply collecting data and then displaying or summarizing it as a 'diagram' – which is known as **descriptive statistics**.

The method that is used in collecting the sample affects the validity of the inferences that can or should be made. The aim then is to obtain a sample that is representative of the population.

This concept can be represented as follows:



One approach to reduce bias in the sample we acquire is to use a random sampling process. By doing this we stand a better chance of obtaining samples that reflect the population as a whole.

Types of data

Data can be classified as numerical or categorical –

Numerical data:

These are made up of observations that are quantitative and so have a numerical value associated with them.

For example, if the set of data is to represent heights, then the data would be collected as numerical values, e.g. 172 cm, 165 cm, etc.

Categorical data:

These are made up of observations that are qualitative (which are sometimes also known as nominal data).

For example, if the set of data is to represent hair colour, then the data would be collected as qualitative data e.g. black, brown, blue, etc.

Discrete and continuous data

As a rule of thumb, **discrete data** are sets of data that can be **counted**. **Continuous data** are sets of data that are measured.

Exercise 5.1.1

1. Micro Inc. produces 14 500 electrical components each month. Of these, 2,000 are randomly selected and tested. The test reveals 42 defective components.

   a What is: **i** the population size

   ii the sample size?

   b Give an estimate of the number of defectives produced during that month.

2. A salmon farm is attempting to determine the number of salmon in its reservoir. On Monday 300 salmon were caught, tagged and then released back into the reservoir. The following Monday 200 salmon were caught and of these 12 were already tagged.

   a Comment on the sampling procedure. Is the sample size large enough? Is there a bias involved?

   b Estimate the number of salmon in the reservoir.

3. A manufacturer wishes to investigate the quality of his product – a measuring instrument that is calibrated to within a 0.01 mm reading accuracy. The manufacturer randomly selects 120 of these instruments during one production cycle. She finds that 8 of the instruments are outside the accepted measuring range. One production cycle produces 1500 of these.

   a What is: **i** the population size?

   ii the sample size?

   b Give an estimate of the number of unacceptable instruments produced during a complete production cycle.

   c In any given week there are 10 production cycles. How many unacceptable instruments can the manufacturer expect at the end of a week? Comment on your result.

4. Classify the following as categorical or numerical data.

   a The winning margin in a soccer game.

   b The eye colour of a person.

   c The number of diagrams in a magazine.

   d The breed of a cat.

   e The fire-hazard levels during summer.

5. Classify the following as discrete or continuous data.

   a The number of cats in a town.

   b The length of a piece of string.

   c The time to run 100 metres.

   d The number of flaws in a piece of glass.

   e The volume of water in a one litre bottle.

Displaying Data

There are a number of ways that we may display sets of data that have been gathered. In fact, there are also a number of ways that we may gather the data. In our case, we shall confine our process to a method of random sampling and assume that, all things being equal, the data will be treated as the population for examination purposes.

Statistical Display, Ellis Island, New York.

Our first task is to deal with the types of data we collect, that is, discrete or continuous.

As mentioned earlier, if the sets of data are counted, they are considered discrete and if they are measured, they are considered continuous. So, for example, if we were counting the number of fish that were caught over a period of 365 days, this would be considered as a discrete measure (as we are carrying out a counting process). However, if we were looking to carry out some analysis about the length of these fish, then that would be considered as a continuous measure (as we are carrying out a measuring process).

Similarly, if we were looking at rainfall, we could measure the number of days on which it rained – which would be considered to be a counting process or, the amount of rain that fell on each day – which would be considered as a measuring process.

So, in the case of the rainfall example, we could have the following tables of data:

Discrete (counting the number of days):

Number of days on which it rained

Month	Jan	Feb	Mar	Apr	May	June
No. of days	4	4	5	9	14	18
Month	July	Aug	Sep	Oct	Nov	Dec
No. of days	19	19	14	13	7	5

Continuous (measuring the amount of rain):

Amount of rain that fell (in mm)

Month	Jan	Feb	Mar	Apr	May	June
Amount of rain	12.5	11.0	14.3	13.7	31.5	53.2
Month	July	Aug	Sep	Oct	Nov	Dec
Amount of rain	73.5	82.9	50.4	30.1	28.7	20.2

From the data, we also observe the different forms that the data takes on. For the discrete data (the number of days) we have whole numbers, i.e. counting numbers. For the continuous data (the amount of rain) we have rational numbers.

Although there are distinct differences between the two types of data, realize that there will be times when a set of data contains whole numbers (i.e. counting numbers) but is recording a measure from a continuous set of data. For example, the amount of rain is recorded to the nearest integer, then the above result for the amount of rain would look like:

Amount of rain that fell (in mm)

Month	Jan	Feb	Mar	Apr	May	June
Amount of rain	13	11	14	14	32	53
Month	July	Aug	Sep	Oct	Nov	Dec
Amount of rain	74	83	50	30	29	20

So, even though the numbers shown are integers, they still reflect a measuring process and so are still considered as continuous.

You have already been exposed to some statistical work in the past, such as representing data in table form or graphical form. Here, we will review this by way of examples so that you may be reminded of the elements involved when dealing with these types of data.

Example 5.1.1

Navneet rolls two dice 26 times, and records the sum showing uppermost on each throw:

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

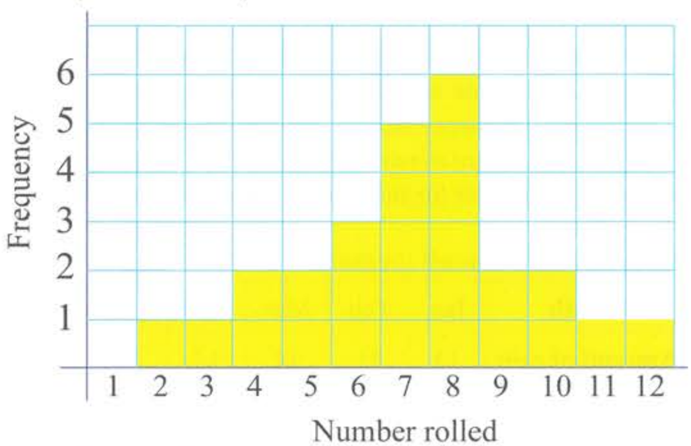
Construct a frequency table for his results and represent the data accordingly.

In this instance, we note that we have a counting process, meaning that we are dealing with a discrete data set. So, we can set up a frequency table.

Sum from rolling two dice

Score (x)	2	3	4	5	6	7	8	9	10	11	12
Frequency (f)	1	1	2	2	3	5	6	2	2	1	1

Based on these results, we can now draw our graph, with the axes labelled appropriately, frequency on the vertical axis and score (number rolled) on the horizontal axis:



Discrete data can also be presented using an interval (or range) of numbers and not individual numbers like in the previous example. The next example illustrates this.

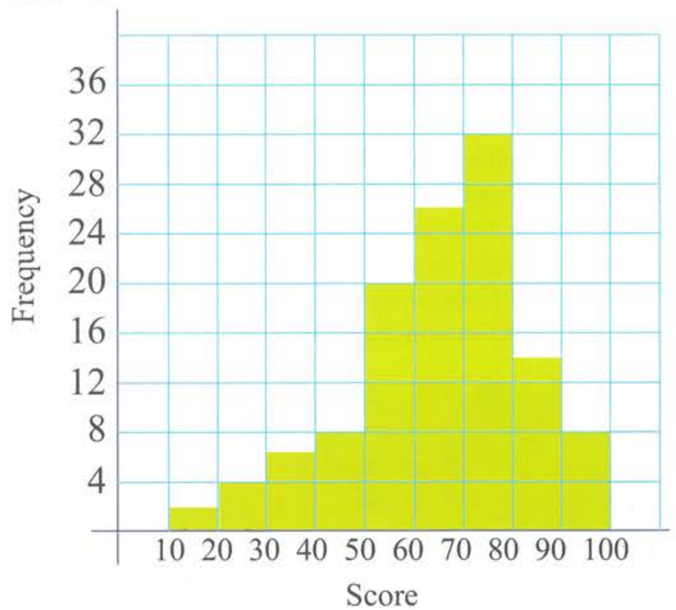
Example 5.1.2

One hundred students sat a test, marked out of 120. Students were only awarded whole marks for their test. The results are shown in the table below. Draw a histogram for the table below.

Student test results

Score range (x)	[0, 10[	[10,20[	[20, 30[	[30, 40[	[40, 50[
Frequency (f)	0	2	4	6	8
Score range (x)	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100]
Frequency (f)	20	26	32	14	8

Again, we have a discrete data set with well defined intervals. As such, we may proceed with our histogram using the given intervals.



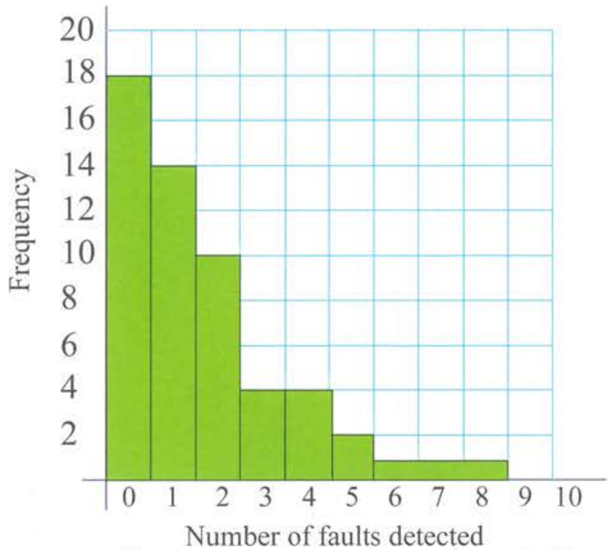
There are a number styles of diagrams available. Selecting the best one for a particular purpose is something of an art.

Our photograph (of a pictogram) at the Immigration Museum at Ellis Island in New York illustrates this.

The first two examples show the bar-chart (in which the height of bar reflects the frequency) and histogram (in which it is the areas of the bars that matter).

Exercise 5.1.2

1. The histogram below displays the number of faults detected during an inspection of car components.



   a How many components were inspected?

   b What percentage of components had no faults?

   c What percentage of components had at least 5 faults?

2. The table below shows the frequency distribution of marks in a science test by 500 students.

Test scores

Score range (x)	[0, 20[	[10, 20[	[20,30[	[30, 40[	[40, 50[
Frequency (f)	0	45	85	145	105
Score range (x)	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Frequency (f)	60	25	15	20	0

   a Draw a histogram for the students marks.

   b If the pass mark is 30, what proportion of students passed?

   c Students are awarded an A-grade if their score is in the top 4%. What is the lowest mark possible in order to attain this grade?

3. The data shown below reflect the rainfall (mm) over a 30-day period:

2.0	3.7	3.2	1.5	2.7	7.5	10.5
8.7	2.2	4.6	3.1	2.5	1.7	7.3
2.2	5.2	4.8	6.2	2.1	7.2	1.2
4.7	2.7	2.1	8.1	1.3	2.5	0.9
5.6	12.2					

   a Is the data continuous or discrete?

   b Draw a suitable histogram for the amount of rainfall that fell over the 30 days.

Statistical Measures

The following figures are the heights (in centimetres) of a group of students:

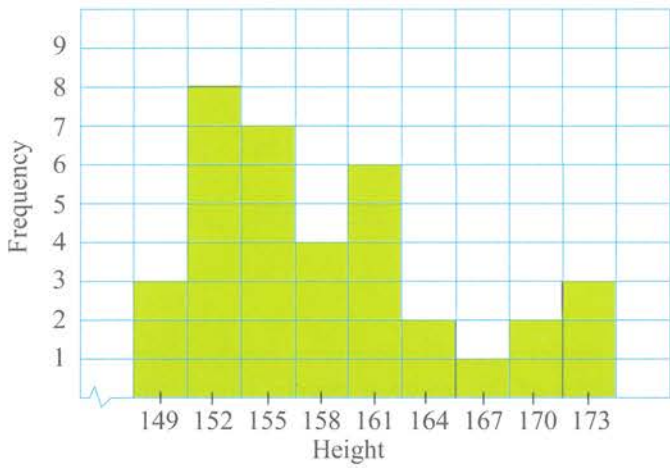
156	172	168	153	170	160	170
156	160	160	172	174	150	160
163	152	157	158	162	154	159
163	157	160	153	154	152	155
150	150	152	152	154	151	151
154						

These figures alone do not give us much information about the heights of this group of people. One of the first things that is usually done in undertaking an analysis is to make a frequency table. In this case, as there are a large number of different heights, it is a good idea to group the height data into the categories (or classes) 148–150, 151–153, 154–156, etc. before making a tally.

Height	Tally	Frequency
148–150	///	3
151–153	////////	8
154–156	///////	7
157–159	////	4
160–162	//////	6
163–165	//	2
166–168	/	1
169–171	//	2
172–174	///	3

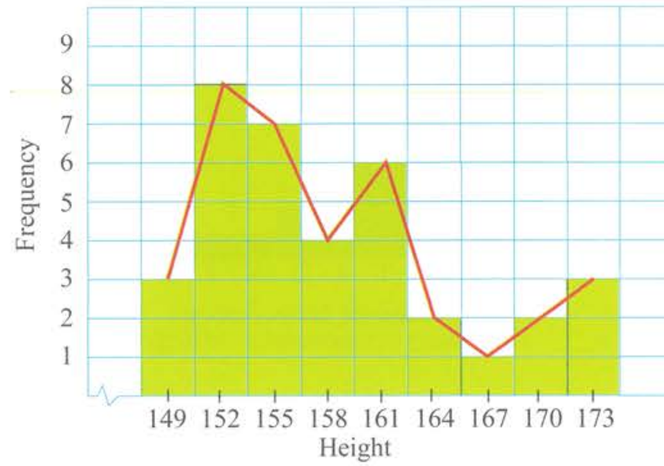
Each height is recorded in the appropriate row of the tally column. Finally, the frequency is the number of tally marks in each row. As a check, the total of the frequency column should equal the count of the number of data items. In this case there are 36 heights.

The choice of class interval in making such a frequency table is generally made so that there are about ten classes. This is not inevitably the case and it is true to say that this choice is an art rather than a science. The objective is to show the distribution of the data as clearly as possible. This can best be seen when the data is shown graphically. There are a number of ways in which this can be done. In the present example, we are dealing with heights. Since heights vary continuously, we would most usually use a histogram to display the distribution.



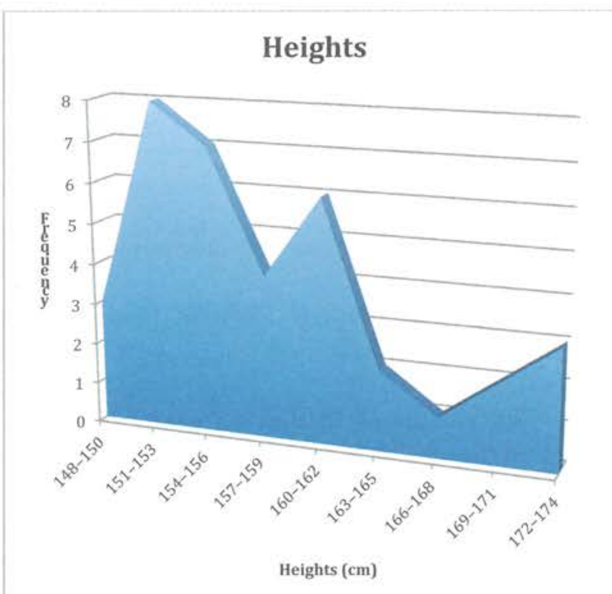
There are two details connected with the construction of histograms that you should not ignore. Firstly, as far as the horizontal scales are concerned, we are representing the continuous variable 'height'. The first class interval represents all the people with heights in the range 148 to 150 cm. Since these have been rounded to the nearest whole centimetre, anyone with a height from 147.5 to 150.5 cm, or [147.5, 150.5], will have been placed in this class. Similarly, anyone with a height in the range [150.5, 153.5) will be categorized in the class 151–153 cm. If you want to label the divisions between the blocks on the histogram, technically these should be 147.5, 150.5 etc. Secondly, in a histogram, it is the area of the bars and not their height that represents the number of data items in each class. To be completely correct, we should give the area as a measure of the vertical scale. This definition allows us to draw histograms with class intervals of varying widths. This is sometimes done to smooth out the variations at the extremes of a distribution where the data is sparse. This aspect will not be considered in this chapter.

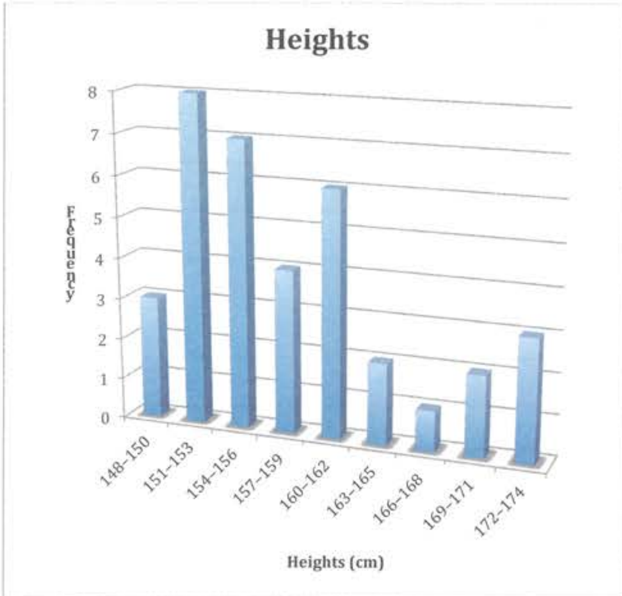
Once we have drawn a histogram, it should be possible to see any patterns that exist in the data. In this case, there is a big group of students with heights from about 150 to 160 cm. There are also quite a few students with heights significantly larger than this and very few with heights below the main group. The distribution has a much larger 'tail' at the positive end than at the negative end and is said to be positively skewed. Patterns can also be seen using other graphical devices such as a line graph:



The same patterns are evident from this diagram as were seen from the histogram.

While on the subject of visual displays, modern spreadsheets offer attractive graphical options. Here these data are displayed in two further ways:





Measure of central tendency

After using a graphical presentation of some sort to look at the general pattern of the data, we would usually calculate some representative 'statistics'. The aim of producing these is to reduce the amount of data to a small number of figures that represent the data as well as possible. In the case of the height data we have been studying, we have already observed that the heights group around the range 150–160 cm. This is sometimes known as a 'central tendency' and we have several ways in which we measure this:

Mode

This is the most frequent class of data. In the present case there were more students in the 151–153 cm class than any other so we would give this class as the mode. It is possible for some data to have more than one mode. We describe this as being bimodal, trimodal etc. The mode tends only to be used when there is no alternative such as when we are collecting data on the television stations that people like best.

Example 5.1.3

The data below records the number of minutes (to the nearest minute) that a student arrives late to school over a period of 45 days. Find the mode of the following data set:

2	5	8	5	7	9	6
5	10	2	8	9	5	8
5	6	5	4	7	9	3
4	3	6	9	8	6	5
10	2	4	9	8	5	5
7	9	4	7	6	5	8
2	9	7				

We start by constructing a frequency table for the data set:

Frequency distribution of minutes late to school

Minutes late	2	3	4	5	6	7	8	9	10
Frequency	4	2	4	10	5	5	6	7	2

From the resulting table we observe that '5 minutes late' occurs the most frequently. This means that the mode is 5.

Mean

This is the measure commonly (and incorrectly) called average. Numeric data is added and the result is divided by the number of items of data that we have.

Notation:

The notation used for the mean depends on whether or not we are claiming to have the mean of all (the **population**) or part (a **sample**) of the possible data set.

In the case of the students, we appear to have a small group of 36 selected from all the possible students in this age group and so we are looking at a sample. It is generally quite clear whether any set of data refers to a population (such as a census) or a sample (such as a poll).

The population mean is denoted by μ and a sample mean by $\bar{x}$.

For a data set x, with n items, both means are calculated in the same way:

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

The symbol $\sum_{i=1}^{n} x_i$ means 'add all the x_i-values'. from x_1 to x_n.

That is, find the sum of all observed values.

If the data is presented in the form of a frequency table in which each item of data x_i is present with a frequency of f_i, then the formula becomes:

$$\mu = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} \text{ or } \mu = \frac{\sum_{i=1}^{k} f_i x_i}{N} = \frac{1}{N}\sum_{i=1}^{k} f_i x_i \text{ where } N = \sum_{i=1}^{k} f_i$$

where k represents the number of groups (or intervals).

Notice that the symbol $\bar{x}$ is used for the sample mean and μ is used for the population mean. The computational calculations, however, provide the same result in this instance. In this course, we use μ as the symbol to represent the mean.

Example 5.1.4

Find, to the nearest minute, the mean number of minutes that the student in Example 5.1.3 arrives late at school.

We could use:

$$\mu = \frac{1}{n}\sum_{i=1}^{45} x_i = \frac{1}{45}(2+5+8+5+\ldots+2+9+7) = \ldots$$

Or, we could make use of the frequency table:

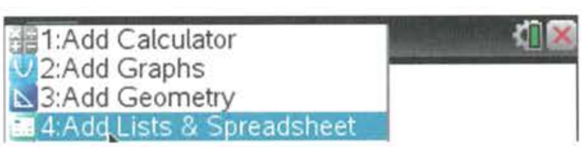
$$\mu = \frac{1}{N}\sum_{i=1}^{9} f_i x_i = \frac{1}{45}(2\times4+3\times2+4\times4+5\times10+6\times5+\ldots$$
$$\ldots+7\times5+8\times6+9\times7+10\times2)$$
$$= \frac{1}{45}\times 276$$
$$= 6.13$$

Therefore, the student is, on average, 6 minutes late.

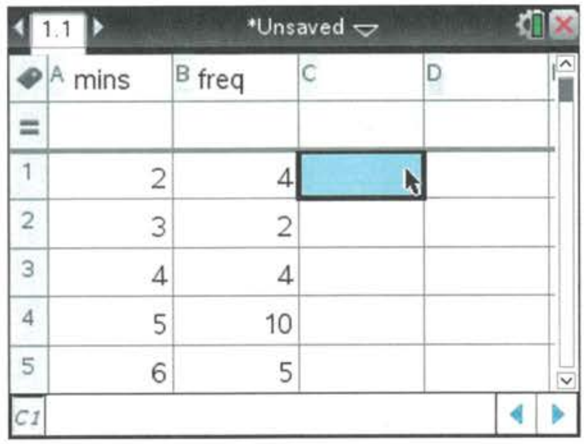
It can be helpful to use a graphics calculator!

Calculators vary and students should be thoroughly familiar with their model. Also, a particular model may be able to handle statistical calculations in more than one way. The TI-NSpire range handles basic statistics in this manner:

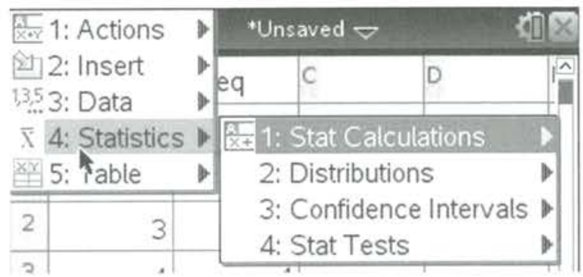
Set up a Lists and Spreadsheet document:



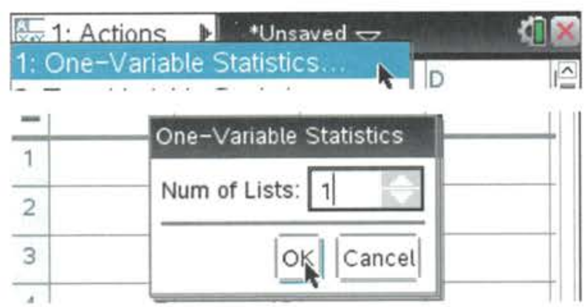
Enter column titles (mins & freq) and the grouped data. The raw data could also be entered as a single column. Select an empty cell (C1).



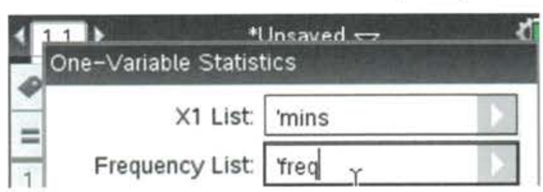
Use MENU / Statistics / Stat Calculations.



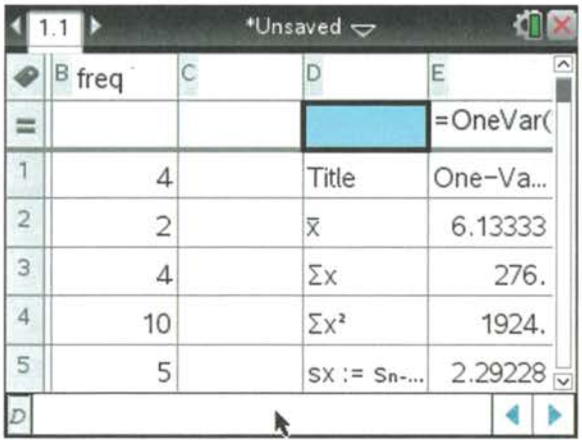
Select the 1-Variable option and number of lists = 1.



Define the lists to be used as data and frequency.



The results will be pasted into the document. Row 2 gives the mean and row 3 the data total.



We now go back to our height data and see how we deal with that situation.

For the height data, we have two ways of approaching this calculation. One way is to return to the original data and add it all up. The total is 5694. There are 36 measurements so:

$$\text{Mean} = \frac{5694}{36} = 158.16667$$

Alternatively we can use the grouped data formula. There is a convenient way of doing this if we add an extra column to the original frequency table:

Height	Mid-height	Frequency	$f \times h$
148–150	149	3	447
151–153	152	8	1216
154–156	155	7	1085
157–159	158	4	632
160–162	161	6	966
163–165	164	2	328
166–168	167	1	167
169–171	170	2	340
172–174	173	3	519
	Totals:	36	5700

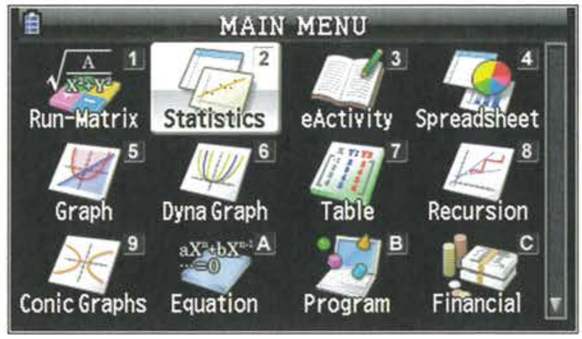
From the table:

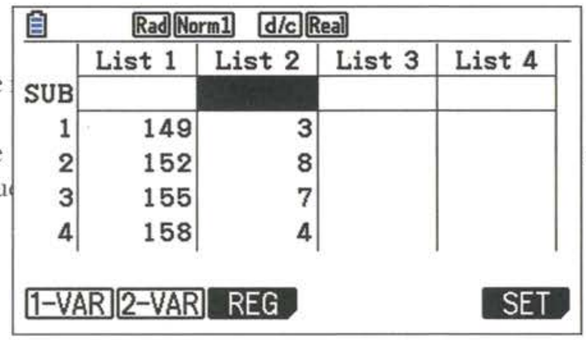
$\sum f_i = 36$ and $\sum f_i \times h_i = 5700$ so

$$\text{Mean} = \frac{\sum f_i \times h_i}{\sum f_i} = \frac{5700}{36} = 158.33333\,.$$

This method of calculating the mean will not necessarily give exactly the same answer as the mean calculated from the original data as we have made the assumption that all the students with heights in the range 148-150 cm had a height of 149 cm. This will not generally be a seriously inaccurate assumption as the students with heights below this figure (148 cm) will be balanced by those with heights above this (150 cm). In this case, the difference is quite small.

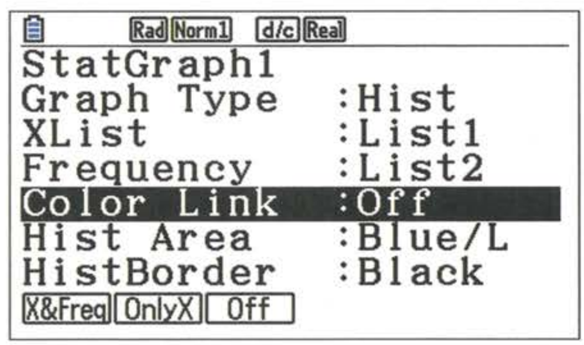
Casio models have a statistics module:



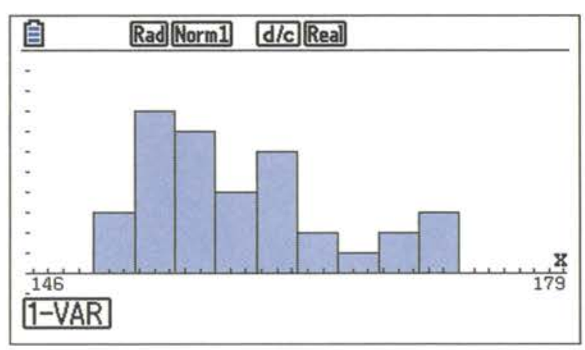


Note that not all the data is shown on the above screen.

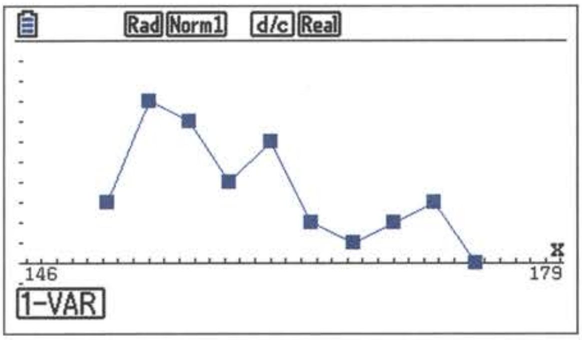
The first thing you may want to do is draw a graph. Press F1-GRAPH and F6-SET. Use the function keys to set the graph as:



Exit to the main data screen and display GRAPH1.

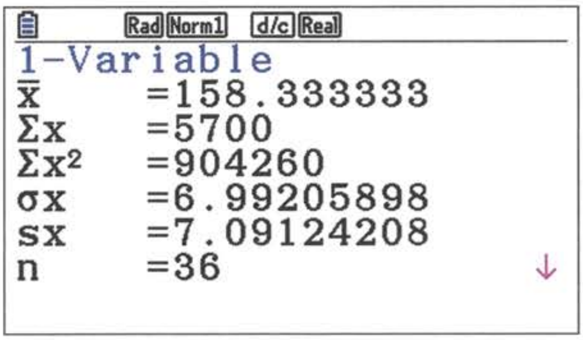


Other types of statistical graphs can be displayed in a similar way. If you amend the settings screen to make the graph type BROKEN. From the main data screen, press F1-GRAPH and F1-GRAPH1:

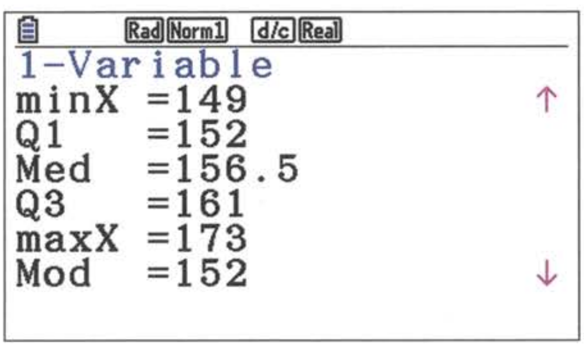


Now that the data has been entered, various statistical measures can be produced.

EXIT to the main data screen and press F2-CALC and F1-1-VAR:



The down arrow indicates that there is more information:



Note that there is now extra information both above and below the current screen. Make sure that you consult the calculator's manual and are familiar with the meaning of 'Q1', 'MED' etc. Don't let the exam be the first moment you find you do not know these things!

Median

The median is found by arranging all the data in order of size and selecting the middle item. For the heights data, there is an even number of figures and so there is not a middle number. In this situation, we take the mean of the middle two data items.

Order:	1	2	3	4	5	6	7	8	9
Height:	150	150	150	151	151	152	152	152	152
	10	11	12	13	14	15	16	17	18
	153	153	154	154	154	154	155	156	156
	19	20	21	22	23	24	25	26	27
	157	157	158	159	160	160	160	160	160
	28	29	30	31	32	33	34	35	36
	162	163	163	168	170	170	172	172	174

The middle heights are the 18th and 19th (156 and 157 cm) so the median is 156.5 cm.

It is usual to take the mean of the two numbers to give an answer to represent the median, however, there are a number of interpolations that can be used. For our purposes, however, we will continue to use the mean of the two observations.

When there are $2n+1$ observation, i.e. there is an odd number of observations, the median corresponds to the $\frac{(2n+1)+1}{2}$ th observation (after they have been placed in order from lowest to highest).

e.g. For the data set {2, 4, 12, 7, 9} we first list the data from lowest to largest: 2, 4, 7, 9, 12.

Here $n = 5$ and so the middle observation is the $\frac{5+1}{2} = 3$ rd observation. i.e. 7.

Note that, for examination purposes, the only measure of central tendency that will be required is the mean.

You will have noticed that the notation $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is often used to calculate the mean for samples – which provides an estimate of the population mean, μ.

However, for examination purposes, all data will be treated as the population.

That is, we will use either $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ (for a list of individual observations) or $\mu = \frac{1}{N}\sum_{i=1}^{k} f_i x_i$, where $N = \sum_{i=1}^{k} f_i$ (for grouped data consisting of k groups).

Exercise 5.1.3

1. The following figures are the weights (in grams) of a group of fish sampled from a reservoir:

226	233	233	244	224	235
238	244	222	239	233	243
221	230	237	240	225	230
236	242	222	235	237	240
220	235	238	243	222	232
232	242	229	231	234	241
228	237	237	245	229	231
237	244	225	236	235	240

Find the mode, mean and median weights.

2. In a study of the weights of a sample of semiprecious gemstones, the following results were obtained (grams):

1.33	1.59	1.82	1.92	1.46
1.57	1.82	2.06	1.59	1.70
1.81	2.02	1.24	1.53	1.69
2.01	1.57	1.62	1.61	1.93
1.11	1.90	1.79	1.91	1.19
1.53	1.90	1.90	1.17	1.97
1.92	2.06	1.41	1.64	1.83
1.90	1.11	1.81	1.83	1.90
1.15	1.68	1.82	1.98	1.39
1.54	1.92	2.04		

Find the mode, mean and median weights.

3. For the data sets below, find the mode, mean and median values.

Set A:

21.1	28.0	26.9	31.9	23.7
28.8	27.9	31.3	21.5	26.8
27.4	31.2	21.4	29.9	29.4
31.5	20.4	25.1	25.8	33.6
23.7	25.6	29.1	30.3	21.5
28.2	28.2	31.3	22.4	25.7
25.1	30.3	21.9	29.1	28.7
30.1	21.8	27.8	29.1	34.3
22.5	25.2	25.5	32.9	22.3
29.0	27.2	33.3		

Set B:

7	6	5	70	9	9
25	72	7	7	4	72
8	9	28	73	9	9
9	72	6	7	27	71
7	7	9	70	6	8
27	73	8	5	26	73
5	6	26	70	9	9
28	73	5	8	26	71

4. The following numbers represent the annual salaries of the employees of a small company.

\$20,910	\$20,110	\$20,390	\$20,170
\$20,060	\$20,350	\$21,410	\$21,130
\$21,340	\$21,360	\$21,360	\$21,410
\$20,350	\$20,990	\$20,690	\$20,760
\$20,880	\$20,960	\$21,240	\$21,060
\$21,190	\$21,400	\$76,000	\$125,000

a Find the mean salary.

b Find the median salary.

c Which of the two figures is the better representative measure of salary?

5. The selling prices for the properties in a suburb over June 2004 were:

\$191,000	\$152,000	\$152,000	\$181,000
\$180,000	\$163,000	\$169,000	\$189,000
\$184,000	\$169,000	\$167,000	\$172,000
\$190,000	\$169,000	\$159,000	\$172,000
\$202,000	\$162,000	\$160,000	\$154,000
\$181,000	\$166,000	\$163,000	\$196,000
\$201,000	\$154,000	\$166,000	\$154,000
\$178,000	\$164,000	\$157,000	\$185,000
\$177,000	\$169,000	\$157,000	\$172,000
\$195,000	\$150,000	\$163,000	\$1,150,000

\$186,000	\$166,000	\$151,000	\$1,155,000
\$185,000	\$151,000	\$168,000	\$1,200,000

a Find the mean selling price.

b Find the median selling price.

c Which of the two figures is the better representative measure of selling price?

Extra questions

Measures of spread

So far we have only looked at ways of measuring the central tendency of a set of data. This is not necessarily the only feature of a data set that may be important. The following sets of data are test results obtained by a group of students in two tests in which the maximum mark was 20.

Test 1:

4	12	11	10	5	10	12	12
6	8	19	13	3	7	11	13
4	9	12	10	6	13	19	11
3	12	14	11	6	13	16	11
5	10	12	13	7	8	13	14
6	10	12	10	7	10	12	10

Test 2:

9	8	10	10	8	9	10	11
8	8	11	10	9	8	11	10
9	8	10	11	8	9	11	10
9	8	11	11	9	9	11	10
8	9	11	10	8	9	11	11
8	8	11	10	8	9	10	10

The means of the two data sets are fairly close to one another (Test 1, 10.1, Test 2, 9.5). However, there is a substantial difference between the two sets which can be seen from the frequency tables.

Test 1:

Mark	3	4	5	6	7	8	9	10	11
Frequency	2	2	2	4	3	2	1	8	5
Mark	12	13	14	15	16	17	18	19	
Frequency	8	6	2	0	1	0	0	2	

Test 2:

Mark	3	4	5	6	7	8	9	10	11
Frequency	0	0	0	0	0	13	11	12	12
Mark	12	13	14	15	16	17	18	19	
Frequency	0	0	0	0	0	0	0	0	

The marks for Test 1 are quite spread out across the available scores whereas those for Test 2 are concentrated around 9, 10 and 11. This may be important as the usual reason for setting tests is to rank students in order of their performance. Test 2 is less effective at this than Test 1 because the marks have a very small spread. In fact, when teachers and examiners set a test, they are more interested in getting a good spread of marks than they are in getting a particular value for the mean. By contrast, manufacturers of precision engineering products want a small spread on the dimensions of the articles they make. Either way, it is necessary to have a way of calculating a numerical measure of the spread of data. The most commonly used measures are variance, standard deviation and interquartile range.

Variance and Standard Deviation

Although statistical computations will usually be carried out using a calculator or computer, we start with a few examples showing the 'background calculations' that are actually carried out. Thereafter, make use of available technology to do the number crunching. We continue with the situation described in Test 1

To calculate the variance of a set of data, the frequency table can be extended as follows:

Mark (M)	Frequency	$M-\mu$	$f(M-\mu)^2$
3	2	−7.10	100.82
4	2	−6.10	74.42
5	2	−5.10	52.02
6	4	−4.10	67.24
7	3	−3.10	28.83
8	2	−2.10	8.82
9	1	−1.10	1.21
10	8	−0.10	0.08
11	5	0.90	4.05
12	8	1.90	28.88
13	6	2.90	50.46
14	2	3.90	30.42
15	0	4.90	0.00
16	1	5.90	34.81
17	0	6.90	0.00
18	0	7.90	0.00
19	2	8.90	158.42
		Total:	640.48

Test 1:

The third column in this table measures the amount that each mark **deviates from the mean** mark of 10.10. Because some of these marks are larger than the mean and some are smaller, some of these deviations are positive and some are negative. If we try to calculate an average deviation using these results, the negative deviations will cancel out the positive deviations. To correct this problem, one method is to square the deviations. Finally, this result is multiplied by the frequency to produce the results in the fourth column.

The last row is calculated:

$$2 \times (3 - 10.10)^2 = 2 \times 50.41 = 100.82\,.$$

The total of the fourth column is divided by the number of data items (48) to obtain the variance of the marks: Variance $= \dfrac{640.48}{48} = 13.34$

The measure most commonly used is the square root of the variance (remember that we squared the deviations). This is a measure known as the standard deviation of the marks. In the previous case: Standard deviation $= \sqrt{13.34} = 3.65$

Repeating this calculation for the second set of marks:

Mark (M)	Frequency	$M-m$	$f(M-m)^2$
8	13	−1.48	28.475
9	11	−0.48	2.534
10	12	0.52	3.245
11	12	1.52	27.725
		Total:	61.979

$$\text{Variance} = \frac{61.979}{48} = 1.291$$

$$\text{Standard deviation} = \sqrt{1.291} = 1.136$$

This figure is about one-third of the figure calculated for Test 1. This reflects the fact that Test 2 has not spread the students very well.

In summary, the variance and population standard deviation are calculated using the formulae:

For individual observations:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

For grouped data:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{1}{N}\sum_{i=1}^{k}f_i(x_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{k}f_ix_i^2-\mu^2$$

where $N = \sum_{i=1}^{k} f_i$.

Then, the standard deviation is calculated as the square root of σ^2.

That is, the standard deviation of a population,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{k}f_i(x_i-\mu)^2}.$$

In the same way that we had a sample mean, $\bar{x}$, and a population mean, μ, when calculating 'the mean', we have a similar situation with the variance. Calculators will have two variance functions; the population variance, σ^2 and the sample variance, s^2. In this course, you will always use the population variance σ^2 (as all data sets will be considered as the population).

The relationship between the sample variance, s^2, and the population variance, σ^2 is $(n-1)s^2 = n\sigma^2$ with $\mu = \bar{x}$.

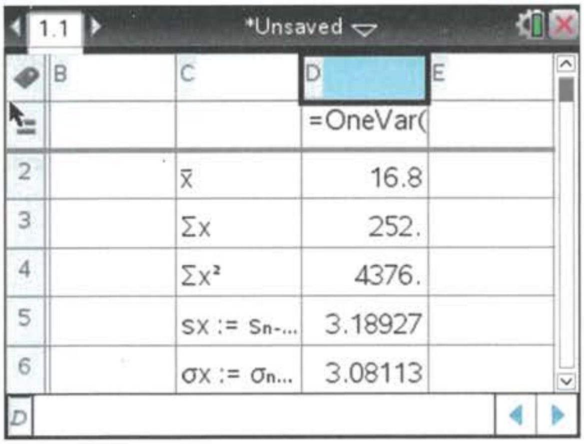
Having done some number-crunching with the illustrated example for test results, we now consider a couple more examples.

Example 5.1.5

Calculate the standard deviation for the following data set.

12	15	11	17	14	16	20
22	15	21	16	17	19	20
17						

With all statistics data, we have the option of making use of a graphics calculator – which we will do in this instance. We start by entering the data as a list and then use the appropriate options to allow the calculator to number crunch,



So, in this case we have that the standard deviation is 3.08.

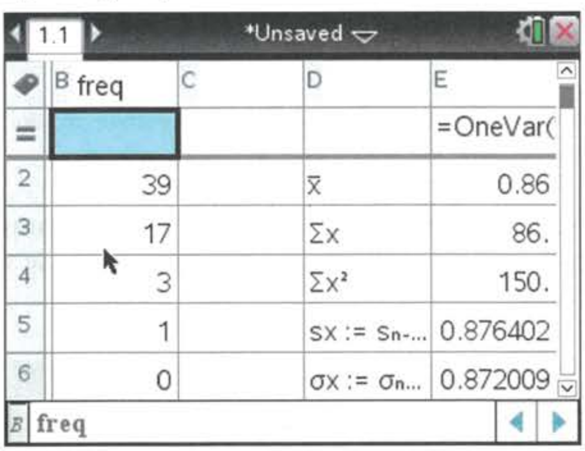
Example 5.1.6

An experiment consists of rolling 5 dice 100 times and recording the number of sixes observed after they land. The result is tabulated below. Calculate the standard deviation of the number of sixes observed.

Results of rolling 5 dice

Number of sixes	0	1	2	3	4	5
Frequency	40	39	17	3	1	0

Again, we use a graphics calculator, which will enable us to quickly work out the standard deviation. We enter the observed values (number of sixes) as the first list, and the corresponding frequencies as the second list:



This gives us the standard deviation as $\sigma = 0.8720$ (the population standard deviation).

Sx is known as the **sample standard deviation**. This is the same as the standard deviation discussed above but with one less than the number of data items in the denominator (47 in this case).

σx is the **population standard deviation** discussed above.

Sample standard deviation? Population standard deviation? What's it all about? Unfortunately there are regional variations (as well as in textbooks) in the notation and the language that is used to define these terms.

When we refer to the *sample variance*, it suggests that we are finding the variance of a sample and, by default, the sample is a subset of a population and so we are in fact finding an estimate of the population variance. This estimate is known as the *unbiased estimate* of the population variance.

The unbiased estimate of the population variance, σ^2, is given by:

$$s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i-\bar{x})^2.$$

The standard deviation of the sample is given by the square root of s_{n-1}^2, i.e. $\sqrt{s_{n-1}^2}$, which

corresponds to the value **Sx** that is produced by the TI-83.

The variance of a population, σ^2, is given by:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} f_i(x_i-\mu)^2$$

The standard deviation then is $\sigma = \sqrt{\sigma^2}$.

To differentiate between division by n and division by $n - 1$ we use s_n^2 for division by n and s_{n-1}^2 for division by $n - 1$.

Giving the relationship $s_{n-1}^2 = \frac{n}{n-1}s_n^2$.

Then, as the population variance, σ^2, is generally unknown, s_{n-1}^2 serves as an estimate of σ^2.

On the TI-83 we have that **Sx** $= s_{n-1}$ and **sx** $= s_n$.

It is therefore important that you are familiar with the notation that your calculator uses for sample standard deviation (unbiased) and population standard deviation.

Exercise 5.1.4

1. The weights (kg) of two samples of bagged sugar taken from a production line.

Sample from machine A:

1.95	1.94	2.02	1.94	2.07	1.95
2.02	2.06	2.09	2.09	1.94	2.01
2.07	2.05	2.04	1.91	1.91	2.02
1.92	1.99	1.98	2.09	2.05	2.05
1.99	1.97	1.97	1.95	1.93	2.03
2.02	1.90	1.93	1.91	2.00	2.03
1.94	2.00	2.02	2.02	2.03	1.96
2.04	1.92	1.95	1.97	1.97	2.07

Sample from machine B:

1.77	2.07	1.97	2.22	1.60	1.96
1.95	2.23	1.79	1.98	2.07	2.32
1.66	1.96	2.05	2.32	1.80	1.96
2.06	1.80	1.93	1.91	1.93	2.25
1.63	1.97	2.08	2.32	1.94	1.93
1.94	2.22	1.76	2.06	1.91	2.39
1.98	2.06	2.02	2.23	1.75	1.95
1.80	1.95	2.09	2.08	2.29	

a Find the mean weights of the bags in each sample.

b Use the formula: $S_x = \sqrt{\dfrac{\sum f_i(x_i - \bar{x})^2}{\sum f_i - 1}}$

to calculate the sample standard deviations of each sample.

c Use the formula: $\sigma_x = \sqrt{\dfrac{\sum f_i(x_i - \mu)^2}{\sum f_i}}$

to calculate the population standard deviations of each sample.

2. The following frequency table gives the numbers of passengers using a bus service over a week-long period.

Passengers	0–4	5–9	10–14	15–19	20–24	25–29
Frequency	3	5	11	15	10	7

a Find the mean number of passengers carried per trip.

b Find the population standard deviation of the number of passengers carried per trip.

3. The number of matches per box in a sample of boxes taken from a packing machine was:

Matches	47	48	49	50	51	52
Frequency	3	6	11	19	12	9

Find the mean and sample standard deviation of the number of matches per box.

4. The weekly expenses paid to a group of employees of a small company were

\$25	\$0	\$10	\$10	\$55	\$0
\$12	\$375	\$75	\$445	\$7	\$2

a Find the mean weekly expense.

b Find the population standard deviation of the expenses.

5. The table shows the numbers of cars per week sold by a dealership over a year.

Cars sold	0	1	2	3	4	5
Number of weeks	2	13	15	12	7	3

a Find the mean weekly sales.

b Find the population standard deviation of the sales.

6. The table shows the weekly turnover of a small shop over a period during Spring and Summer.

Sales (\$)	\$0–\$99	\$100–\$199	\$200–\$299	\$300–\$399
Number of weeks	2	9	15	7

a Find the mean weekly sales.

b Find the population standard deviation of the sales.

7. The frequency distribution of Mathematics and English test results at a local secondary school are shown in the table below:

Test Scores

Mark	Mathematics	English
[0, 10[	7	2
[10, 20[	11	5
[20, 30[	13	11
[30, 40[	17	18
[40, 50[	22	34
[50, 60[	22	34
[60, 70[	16	20
[70, 80[	14	12
[80, 90[	11	2
[90, 100]	7	2

a Draw a histogram showing the test scores for:
i Mathematics ii English.

b

i Draw a table of the cumulative frequencies for each of Mathematics and English.

ii On the same set of axes draw the cumulative frequency graph for Mathematics and English scores.

c The pass mark for Mathematics is the lowest score obtained by 78% of the students. What is the minimum score required for a student to pass Mathematics?

8. Doctors are under pressure to diagnose as many patients as possible, meaning that the session time they allocate to each patient is closely monitored. The table below shows the times that two doctors have spent with their patients.

Session time distribution

Time (minutes)	Number of sessions	
	Doctor A	Doctor B
5–9	5	3
10–14	10	7
15–19	23	x
20–24	16	27
25–29	12	9
30–34	5	6
35–39	3	2

a For doctor A, calculate the:

i mean session time.

ii standard deviation of session times.

b Doctor B has misplaced the tally for the number of times she has seen patients for a period of 15–19 minutes. She decides that she should have the same mean as her colleague for the time spent seeing patients. What value of x should she use (to the nearest integer)?

c A third doctor at the clinic recorded an average of 22 minutes after seeing 60 patients. What is the overall mean time these three doctors spend with patients?

Extra questions

Answers

5.2 Probability

Probability

We are often faced with statements that reflect an element of likelihood, For example, "It is likely to rain later in the day" or "What are the chances that I roll a six?". Such statements relate to a level of uncertainty (or indeed, a level of certainty). It is this element of likelihood in which we are interested. In particular, we need to find a measure of this likelihood — i.e. the associated probability of particular events.

Our title picture is of the *MV Explorer* in Antarctica in December 2001. This voyage was completed safely. Several years later, the ship was holed and sank in Antarctic waters (with no injuries). How did the insurers calculate the premium due for the voyage? The answer is to be found in this and subsequent sections - and in experience.

Roughly, the insurance industry argues:

$$\text{Probability of accident} = \frac{\text{Number of accidents}}{\text{Number of voyages}}$$

The premium is decided by multiplying this probability by the amount insured and adding a profit margin. This is, of course, a considerable simplification of the very complex work of the insurance actuary.

Probability as a long-term relative frequency

An experiment is repeated in such a way that a series of independent and identical trials are produced, so that a particular event A is observed to either occur or not occur. We let N be the total number of trials carried out and $n(A)$ (or $|A|$) be the number of times that the event A was observed.

We then call the ratio $\frac{n(A)}{n}$ $\left(\text{or } \frac{|A|}{N}\right)$ the **relative frequency** of the event A.

This value provides some indication of the likelihood of the event A occurring.

In particular, for large values of N we find that the ratio $\frac{n(A)}{N}$ tends to a number called the **probability** of the event A, which we denote by $p(A)$ or $P(A)$.

As $0 \leq n(A) \leq N$, this number, $P(A)$, must lie between 0 and 1 (inclusive), i.e. $0 \leq P(A) \leq 1$.

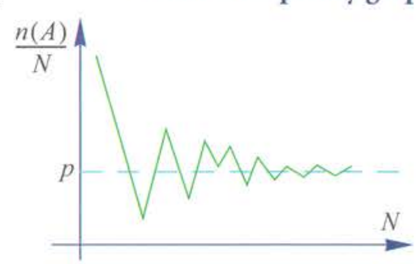
A more formal definition is as follows:

If a random experiment is repeated N times, in such a way that each of the trials is identical and independent, where $n(A)$ is the number of times event A has occurred after N trials, then:

$$\text{As } N \to \infty, \frac{n(A)}{N} \to P(A)$$

It is possible to provide a graph of such a situation, which shows that as N increases, the ratio $\frac{n(A)}{N}$ tends towards some value p, where in fact, $p = P(A)$.

Such a graph is called a **relative frequency graph**.



As far as our actuary is concerned, this demonstrates that the more information they have about a risk, the more reliable the premium calculation will be.

Theoretical probability

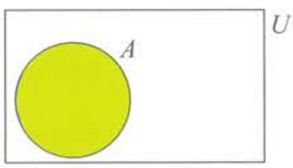
When the circumstances of an experiment are always identical, we can arrive at a value for the probability of a particular event by using mathematical reasoning, often based on an argument reflecting some form of symmetry (i.e. without the need to repeatedly perform the experiment). This type of probability is called **theoretical probability**.

For example, when we roll a die, every possible outcome, known as the **sample space**, can be listed as $U = \{1, 2, 3, 4, 5, 6\}$ (sometimes the letter ε is used instead of U). The probability of obtaining a "four" (based on considerations of **symmetry of equal likelihood**) is given by $^1/_6$. Such a probability seems obvious, as we would argue that:

"Given there are six possible outcomes and each outcome is equally likely to occur (assuming a fair die), then the chances that a 'four' occurs must be one in six, i.e. $^1/_6$."

Laws of probability

We will restrict our arguments to **finite sample spaces**. Recall, that a **sample space** is the set of every possible outcome of an experiment, and that an **event** is any subset of the sample space. This relationship is often represented with a Venn diagram:



The Venn diagram shows the sample A space U, with the event A, as a subset.

Definition of probability

If an experiment has equally likely outcomes and of these the event A is defined, then the **theoretical probability of event A** occurring is given by:

$$P(A) = \frac{n(A)}{n(U)} = \frac{\text{Number of outcomes in which A occurs}}{\text{Total number of outcomes in the sample space}}$$

Where $n(U)$ is the total number of possible outcomes in the sample space, U, (i.e. $n(U) = N$).

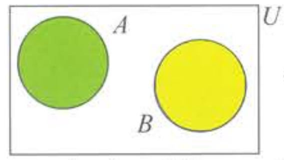
As a consequence of this definition we have what are known as the **axioms of probability**:

1. $0 \le P(A) \le 1$

2. $P(\varnothing) = 0$ and $P(U) = 1$

   That is, if $A = \varnothing$, then the event A can never occur. $A = U$ implies that the event A is a certainty.

3. If A and B are both subsets of U and are mutually exclusive, then $P(A \cup B) = P(A) + P(V)$



Note: Two events A and B are said to be mutually exclusive (or disjoint) if they have no elements in common, i.e. if $A \cap B = \varnothing$.

Example 5.2.1

A fair die is thrown. List the sample space of the experiment and hence find the probability of observing:

a a multiple of 3 b an odd number.

Are these events mutually exclusive?

The sample space is $U = \{1, 2, 3, 4, 5, 6\}$. Let A be the event 'obtaining a multiple of 3'.

We then have that $A = \{3, 6\}$. Therefore, $P(A) = \frac{n(A)}{n(U)} = \frac{2}{6} = \frac{1}{3}$

Let B be the event 'obtaining an odd number'.

Here $B = \{1, 3, 5\}$ and so $P(B) = \frac{n(B)}{n(U)} = \frac{3}{6} = \frac{1}{2}$.

In this case, $A = \{3, 6\}$ and $B = \{1, 3, 5\}$, so that $A \cap B = \{3\}$ Therefore, as $A \cap B \neq \varnothing$ A and B are not mutually exclusive.

Example 5.2.2

Two coins are tossed. Find the probability that:

a two tails are showing b a tail is showing.

Let H denote the event a head is showing and T the event a tail is showing. This means that the sample space (with two coins) is given by $U = \{HH, HT, TH, TT\}$.

The event that two tails are showing is given by the event $\{TT\}$, therefore, we have that:

$$P(\{TT\}) = \frac{n(\{TT\})}{n(U)} = \frac{1}{4}.$$

The event that one tail is showing is given by $\{HT, TH\}$.

Therefore,: $P(\{HT, TH\}) = \frac{n(\{HT, TH\})}{n(U)} = \frac{2}{4} = \frac{1}{2}$.

Example 5.2.3

A card is drawn from a standard deck of 52 playing cards. What is the probability that a diamond card is showing?

Let D denote the event 'a diamond card is selected'.

This means that $n(D) = 13$ as there are 13 diamond cards in a standard deck of cards.

Therefore, $P(D) = \frac{n(D)}{n(U)} = \frac{13}{52} = \frac{1}{4}$.

Problem-solving Strategies in Probability

When dealing with probability problems it is often useful to use some form of diagram to help 'visualize' the situation. **Diagrams** can be in the form of:

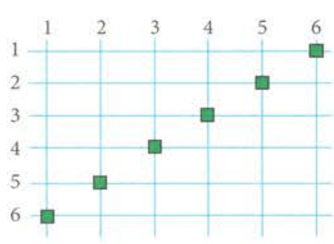
1. Venn diagrams.
2. Tree diagrams.
3. Lattice diagrams.
4. Karnaugh maps (probability tables).
5. As a last resort, any form of diagram that clearly displays the process under discussion (e.g. flow chart).

It is fair to say that some types of diagrams lend themselves well to particular types of problems. These will be considered in due course.

Example 5.2.4

Find the probability of getting a sum of 7 on two throws of a die.

In this instance, we make use of a lattice diagram to display all possible outcomes. From the diagram, we can list the required event (and hence find the required probability):



Let S denote the event 'A sum of seven is observed'. From the lattice diagram, we see that there are 6 possibilities where a sum of seven occurs.

In this case: $S = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$.

Therefore: $P(S) = \frac{n(S)}{n(U)} = \frac{6}{36} = \frac{1}{6}$

Exercise 5.2.1

1. From a bag containing 6 white and 4 red balls, a ball is drawn at random. What is the probability that the ball selected is:

   a red. b white. c not white.

2. From an urn containing 14 marbles of which 4 are blue and 10 are red, a marble is selected at random. What is the probability that:

   a the marble is blue. b the marble is red.

3. A letter is chosen at random from the letters of the alphabet. What is the probability that:

   a the letter is a vowel.

   b the letter is a consonant.

4. A coin is tossed twice. List the sample space and find the probability of observing:

   a two heads.

   b at least one head.

5. A coin is tossed three times. List the sample space and find the probability that:

   a two heads show uppermost.

   b at least two heads show uppermost.

   c three heads or three tails are showing.

6. A letter is chosen at random from the word FERTILITY. Find the probability that the letter chosen is:

   a a T. b an I.

   c a consonant. d a vowel.

7. A bag has 20 coins numbered from 1 to 20. A coin is drawn at random and its number is noted. What is the probability that the coin drawn has:

   a an even number on it?

   b has a number that is divisible by 3?

   c has a number that is divisible by 3 or 5?

8. A die is rolled twice. Use a lattice diagram to illustrate the sample space. What is the probability of observing:

   a at least one five. b a four and a three.

   c a pair. d a sum of eight.

9. A family has three children. List the sample space and hence find the probability that:

   a there are 3 boys.

   b there are 2 boys and 1 girl.

   c there are at least two girls.

10. A card is selected from a pack of 52 cards. Find the probability that the card is:

   a red b a heart

   c red and a heart.

11. A cube is drawn at random from an urn containing 16 cubes of which 6 are red, 4 are white and 6 are black. Find the probability that the cube is:

   a red b white

   c black d red or black.

12. A coin is tossed and a die is rolled simultaneously. Draw a lattice diagram to depict this situation.

   a Using your lattice diagram, list the sample space.

   b What is the probability of observing a tail and an even number?

13. A die is rolled three times. Find the probability of observing:

   a three sixes.

   b three even numbers.

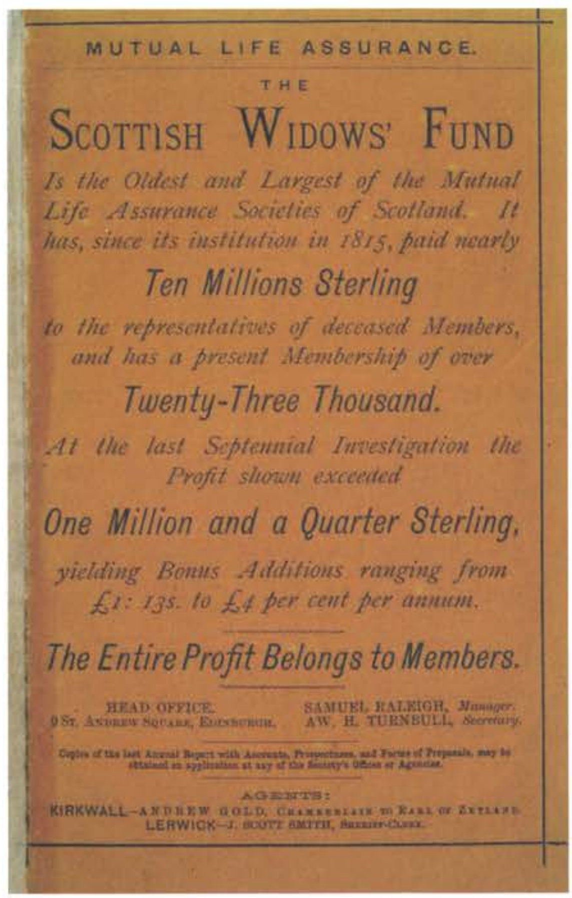
   c two odd numbers.

   (Hint: You may need to draw a three-dimensional lattice diagram.)

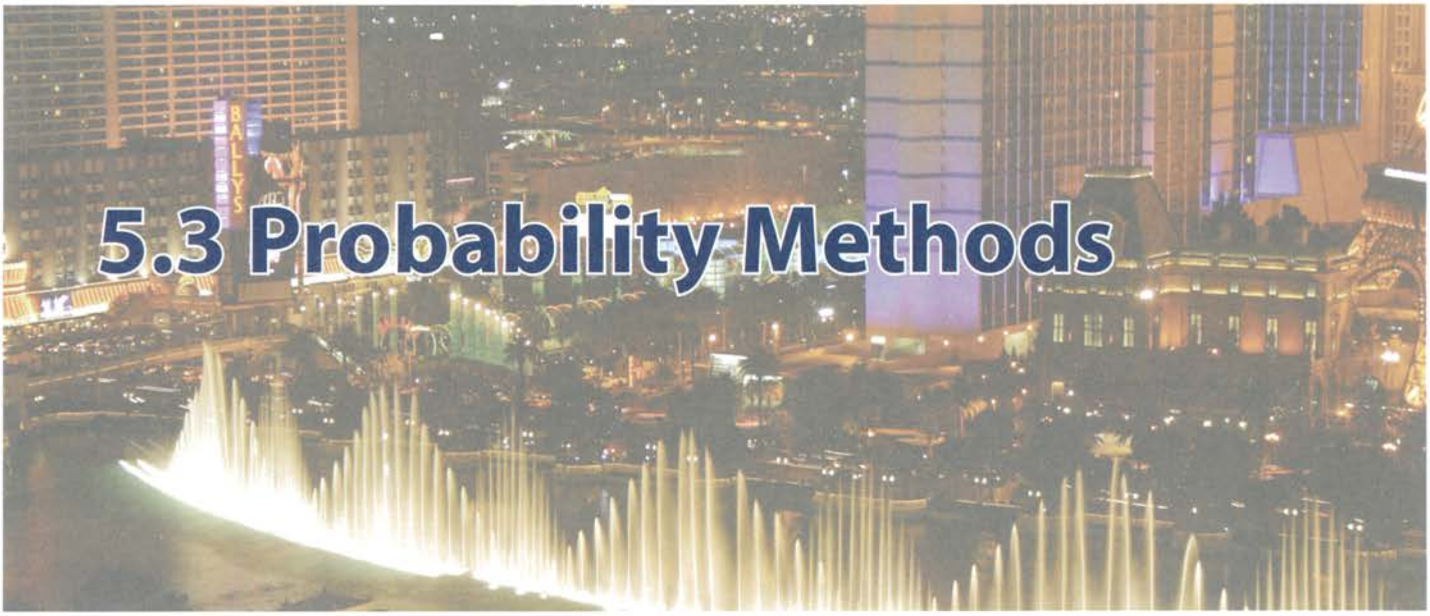
Scottish Widows One of the first life insurance companies in the world was set up in 1812 to care for widows and orphans. Their plan was to collect premiums to create a capital fund. The pensions would be paid from the interest generated by the fund. The capital was intended to be preserved.

They got their calculations right and are trading to this day.

This advertisement dates from 1878.



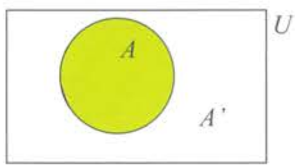
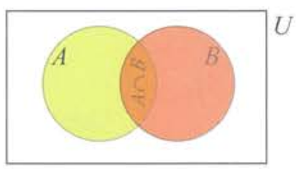
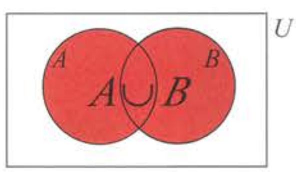
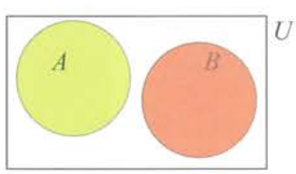
Answers



5.3 Probability Methods

Our title picture is Las Vegas by night - World capital of gambling and heavily reliant on probability theory.

From the axioms of probability we can develop further rules to help solve problems that involve chance. We illustrate these rules with the aid of Venn diagrams.

Event	Set language	Venn diagram	Probability result
The **complement** of A is denoted by A'.	A' is the complement to the set A, i.e. the set of elements that do not belong to the set A.		$\mathrm{P}(A')=1-\mathrm{P}(A)$ $\mathrm{P}(A')$ is the probability that event A does not occur.
The **intersection** of A **and** B : $A \cap B$	$A \cap B$ is the intersection of the sets A and B, i.e. the set of elements that belong to **both** the set A **and** the set B.		$\mathrm{P}(A \cap B)$ is the probability that both A and B occur.
The **union** of events A and B : $A \cup B$	$A \cup B$ is the union of the sets A and B, i.e. the set of elements that belong to A **or** B or both A **and** B.		$\mathrm{P}(A \cup B)$ is the probability that either event A or event B (or both) occur. From this we have what is known as the '**Addition rule**' for probability: $\mathrm{P}(A \cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A \cap B)$
If $A \cap B=\varnothing$ the events A and B are said to be **disjoint**. That is, they have no elements in common.	If $A \cap B=\varnothing$ the sets A and B are **mutually exclusive**.		If A and B are mutually exclusive events then event A and event B cannot occur simultaneously, i.e. $\mathrm{n}(A \cap B)=0$ $\Rightarrow \mathrm{P}(A \cap B)=0$ Therefore $\mathrm{P}(A \cup B)=\mathrm{P}(A)+\mathrm{P}(V)$

Although we now have a number of 'formulae' to help us solve problems that involve probability, using other forms of diagrams to clarify situations and procedures should not be overlooked.

Example 5.3.1

A card is randomly selected from an ordinary pack of 52 playing cards. Find the probability that it is either a 'black card' or a 'king'.

Let B be the event 'A black card is selected.' and K the event 'A king is selected'.

We first note that event B has as its elements the Jack of spades (J♠), the Jack of clubs (J♣), the Queen of spades (Q♠), the Queen of clubs (Q♣) and so on. This means that:

B = {K♠,K♣,Q♠,Q♣,J♠,J♣,10♠,10♣,9♠,9♣,8♠,8♣,7♠,7♣, 6♠,6♣,5♠,5♣,4♠,4♣,3♠,3♣,2♠,2♣,A♠,A♣} and

K = {K♠, K♥, K♣, K♦}, so that $B \cap K$ = {K♠, K♣}.

Using the addition rule, $P(B \cup K) = P(B) + P(K) - P(B \cap K)$

we have $P(B \cup K) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{7}{13}$.

Note the importance of subtracting $\frac{2}{52}$ as this represents the fact that we have included the event {K♠, K♣} twice when finding B and K.

We now consider one of the problems from Exercise 5.2.1, (no. 7) but this time we make use of the addition rule.

Example 5.3.2

A bag has 20 coins numbered from 1 to 20. A coin is drawn at random and its number is noted. What is the probability that the coin has a number that is divisible by 3 or by 5?

Let T denote the event "The number is divisible by 3" and S, the event "The number is divisible by 5".

Using the addition rule we have:

$$P(T \cup S) = P(T) + P(S) - P(T \cap S)$$

Now, $T = \{3, 6, 9, 12, 15, 18\}$ and $S = \{5, 10, 15, 20\}$ so that $T \cap S = \{15\}$.

Therefore, we have $P(T) = \frac{6}{20}$ and $P(S) = \frac{4}{20}$ and $P(T \cap S) = \frac{1}{20}$.

This means that $P(T \cup S) = \frac{6}{20} + \frac{4}{20} - \frac{1}{20} = \frac{9}{20}$.

Example 5.3.3

If $P(A) = 0.6, P(B) = 0.3$ and $P(A \cap B) = 0.2$, find:

a $P(A \cup B)$ b $P(B')$

c $P(A \cap B')$

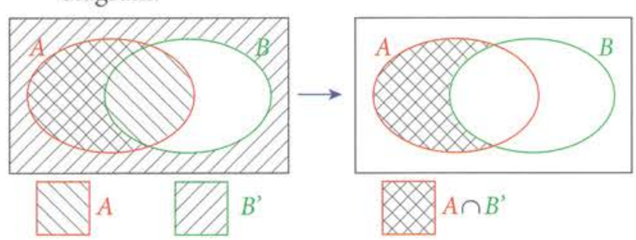
a Using the addition formula, we have:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.6 + 0.3 - 0.2 \\ &= 0.7 \end{aligned}$$

b Using the complementary formula: $P(B') = 1 - P(B)$

$$= 1 - 0.3$$
$$= 0.7$$

c To determine $P(A \cap B')$, we need to use a Venn diagram:



Using the second Venn diagram we are now in a position to form a new formula:

$$\begin{aligned} P(A \cap B') &= P(A) - P(A \cap B) \\ &= 0.6 - 0.2 \\ &= 0.4 \end{aligned}$$

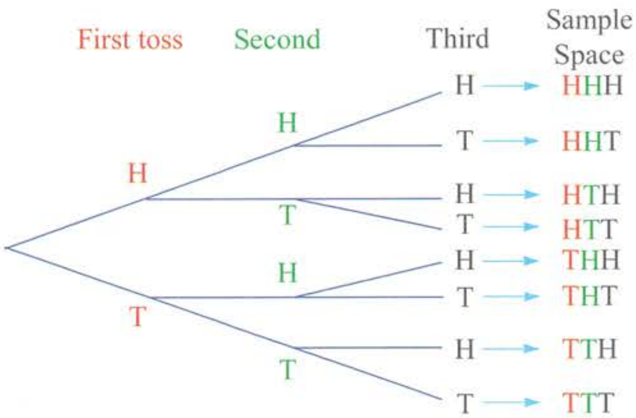
Example 5.3.4

A coin is tossed three times. Find the probability of:

a obtaining three tails

b obtaining at least one head.

We begin by drawing a tree diagram to describe the situation:



From the tree diagram we have a sample space made up of eight possible outcomes:

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

Let X be the event "Obtaining three tails", so $X = \{TTT\}$.

Therefore $P(X) = \frac{1}{8}$.

Although we can answer this question by using the tree diagram, we make use of complementary events to solve this problem.

Notice that 'At least one head' is the complement of no heads.

So, P(At least one head) $= P(X') = 1 - P(X) = 1 - \frac{1}{8} = \frac{7}{8}$.

Exercise 5.3.1

1. A letter is chosen at random from the letters of the word TOGETHER.

 a Find the probability of selecting a T.

 b Find the probability of selecting a consonant.

 c Find the probability of not selecting an E.

2. A card is drawn at random from a standard deck.

 a Find the probability that the card is an ace.

 b Find the probability that the card is black.

 c Find the probability that the card is an ace and black.

 d Find the probability that the card is an ace or black.

3. A letter is selected at random from the alphabet. Find the probability that the letter is a vowel or comes from the word 'helpful'.

4. The events A and B are such that $P(A) = 0.5$, $P(B) = 0.7$ and $P(A \cap B) = 0.2$.

 Find:

 a $P(A \cup B)$.

 b $P(B')$.

 c $P(A' \cap B)$.

5. The events A and B are such that $p(A) = 0.35$, $p(B) = 0.5$ and $p(A \cap B) = 0.15$.

 Using a Venn diagram (where appropriate), find:

 a $p(A')$.

 b $p(A \cup B)$.

 c $p(A \cup B')$.

6. The events A and B are such that $p(A) = 0.45$, $p(B) = 0.7$ and $p(A \cap B) = 0.20$.

 Using a Venn diagram (where appropriate), find:

 a $p(A \cup B)$.

 b $p(A' \cap B')$.

 c $p((A \cap B)')$.

7. A coin is tossed three times.

 a Draw a tree diagram and from it write down the sample space.

 b Use the results from part a to find the probability of obtaining:

 i only one tail.

 ii at least 2 tails.

 iii 2 tails in succession.

 iv 2 tails.

8. In a class of 25 students it is found that 6 of the students play both tennis and chess, 10 play tennis only and 3 play neither. A student is selected at random from this group.

   Using a Venn diagram, find the probability that the student:

   a plays both tennis and chess.

   b plays chess only.

   c does not play chess.

9. A blue and a red die are rolled together (both numbered one to six).

   a Draw a lattice diagram that best represents this experiment.

   b Find the probability of observing an odd number.

   c Find the probability of observing an even number with the red die.

   d Find the probability of observing a sum of 7.

   e Find the probability of observing a sum of 7 or an odd number on the red die.

10. A card is drawn at random from a standard deck of 52 playing cards. Find the probability that the card drawn is:

   a a diamond.

   b a club or spade.

   c a black card or a picture card.

   d a red card or a queen.

11. A and B are two events such that $P(A) = p$, $P(B) = 2p$ and $P(A \cap B) = p^2$.

   a Given that $P(A \cup B) = 0.4$, find p.

   b Use a Venn Diagram to help you find the following:

   i $P(A' \cup B)$.

   ii $P(A' \cap B')$.

12. In a group of 30 students 20 hold an Australian passport, 10 hold a Malaysian passport and 8 hold both passports. The other students hold only one passport (that is neither Australian nor Malaysian). A student is selected at random.

   a Draw a Venn diagram which describes this situation.

   b Find the probability that the student has both passports.

   c Find the probability that the student holds neither passport.

   d Find the probability that the student holds only one passport.

13. This aircraft has one piston engine.

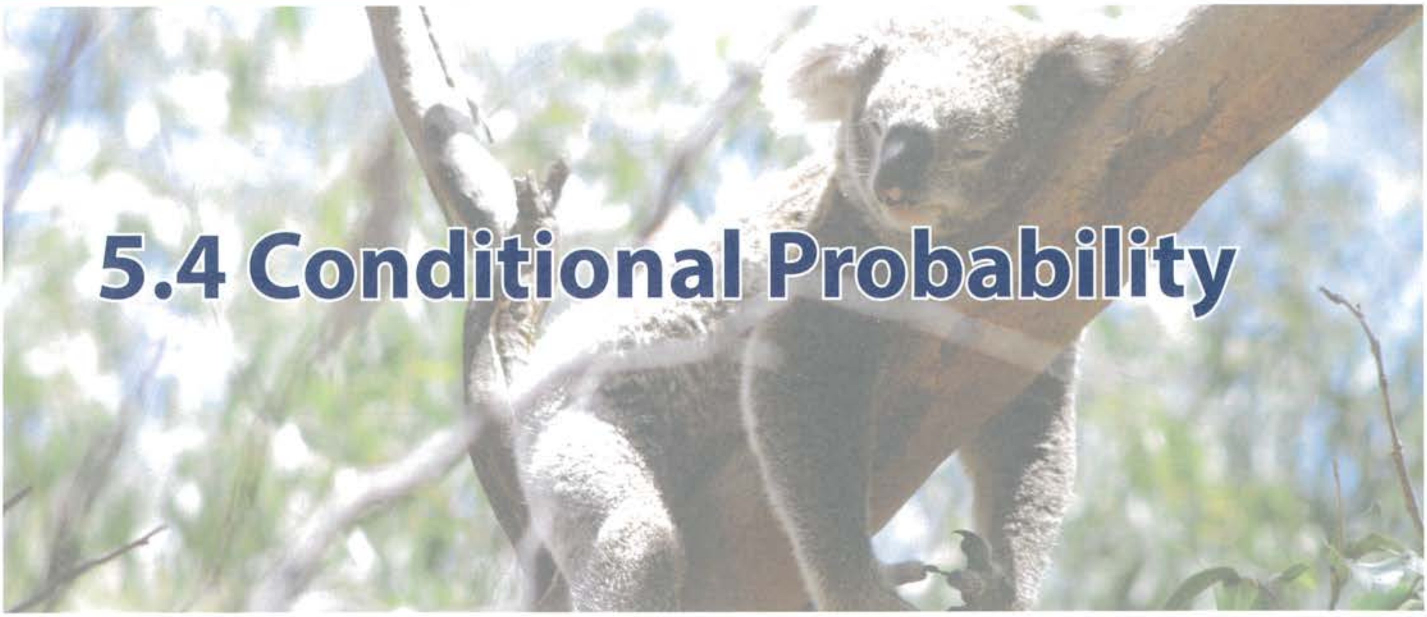
The engine has four horizontally opposed cylinders. Each cylinder has two spark plugs.

The two sets of spark plugs are connected to separate ignition systems (magnetos).

The fuel is stored in wing tanks. As the tanks are below the engine, there is no gravity feed. There are, however, two fuel pumps. One is mechanical (driven by the engine) and the other is electrical and can be turned on and off by the pilot.

Discuss how these two features and the principles of probability, make this aircraft safer than it would be if it was powered by a car engine.

Answers



5.4 Conditional Probability

Every living creature owes its existence to the fact that its parent(s) existed. The koala in our picture can only sleep in the sun because a condition was fulfilled.

Informal Definition of Conditional Probability

Conditional probability works in the same way as simple probability. The only difference is that we are provided with some prior knowledge (or some extra condition about the outcome). So, rather than considering the whole sample space, ε, given some extra information about the outcome of the experiment, we only need to concentrate on part of the whole sample space, ε^*. This means that the sample space is reduced from ε to ε^*. Before formalizing this section, we use an example to highlight the basic idea.

Example 5.4.1

a In the roll of a die, find the probability of obtaining a '2'.

b After rolling a die, it is noted that an even number appeared. What is the probability that it is a '2'?

a This part is solved using the methods of the previous section: $U = \{1, 2, 3, 4, 5, 6\}$, and so P('2') = $^1/_6$.

b This time, because *we know that an even number has occurred*, we have a new sample space, namely $U^* = \{2, 4, 6\}$. The new sample size is $n(U^*)$.

P('2' *given that an even number showed up*) = $^1/_3$.

Formal Definition of Conditional Probability

If A and B are two events, then **the conditional probability of event A given event B** is found using:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

Note:

1. If A and B are mutually exclusive then: $P(A|B) = 0$

2. From the above rule, we also have the general Multiplication rule:

$$P(A \cap B) = P(A|B) \times P(B).$$

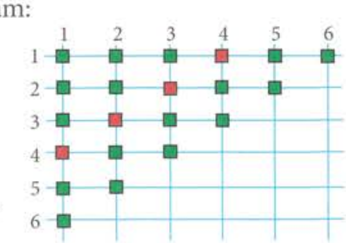
It should also be noted that usually $P(A|B) \neq P(B|A)$.

Example 5.4.2

Two dice numbered one to six are rolled onto a table. Find the probability of obtaining a sum of five given that the sum is seven or less.

We first draw a lattice diagram:

From the diagram we see that the new sample space is made up of 21 outcomes (boxes) and the event we want (red boxes) consists of 4 outcomes.



Then, $P((X=5) \cap (X \le 7)) = \frac{4}{36}$ and $P(X \le 7) = \frac{21}{36}$.

Therefore, $P(X=5|X \le 7) = \dfrac{\frac{4}{36}}{\frac{21}{36}} = \frac{4}{21}$.

Example 5.4.3

A box contains 2 red cubes and 4 black cubes. If two cubes are chosen at random, find the probability that both cubes are red given that:

a the first cube is not replaced before the second cube is selected

b the first cube is replaced before the second cube is selected.

Let A be the event 'the first cube is red' and B be the event 'the second cube is red'. This means that the event $A \cap B$ must be 'both cubes are red'.

Now, $P(A) = {}^2/_6 = {}^1/_3$ (as there are 2 red cubes from a total of 6 cubes in the box). The value of $P(B)$ depends on whether the selection is carried out with or without replacement.

a If the first cube selected is red and it is not replaced, then we only have 1 red cube left in the box out of a total of five cubes.

So, the probability that the second cube is red given that the first is red is ${}^1/_5$.

That is:

$$p(B|A) = \frac{1}{5} \Rightarrow P(A \cap B) = P(B|A) \times P(A) = \frac{1}{5} \times \frac{1}{3} = \frac{1}{15}.$$

b This time, because the cube is replaced, the probability that the second cube is red given that the first one is red is still ${}^1/_3$.

So that:

$$P(B|A) = \frac{1}{3} \Rightarrow P(A \cap B) = P(B|A) \times P(A) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

Example 5.4.4

Two events A and B are such that $P(A) = 0.5$, $P(B) = 0.3$ and $P(A \cup B) = 0.6$. Find:

a $P(A|B)$ b $P(B|A)$ c $P(A'|B)$.

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, therefore we need to find $P(A \cap B)$.

Using the addition rule we have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$0.6 = 0.5 + 0.3 - P(A \cap B)$$

$$\therefore P(A \cap B) = 0.2$$

So, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.3} = \frac{2}{3}$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.2}{0.5} = 0.4$$

$$P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = \frac{0.3 - 0.2}{0.3} = \frac{1}{3}$$

Independence

The events A and B are said to be statistically independent if the probability of event B occurring is not influenced by event A occurring.

Therefore we have the mathematical definition:

Two events **A and B are independent if, and only if,**

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

However, a more convenient definition for independence can be given as follows:

Two events **A and B are independent if, and only if,**

$$P(A \cap B) = P(A) \times P(B)$$

This definition can be used as a test to decide if two events are independent. However, as a rule of thumb, if two events are 'physically independent' then they will also be statistically independent.

There are a few points that should always be considered when dealing with independence:

1. Never assume that two events are independent unless you are absolutely certain that they are independent.

2. How can you tell if two events are independent? A good rule of thumb is: If they are physically independent,

they are mathematically independent.

3. Make sure that you understand the difference between mutually exclusive events and independent events.

 Mutually exclusive means that the events A and B have nothing in common and so there is no intersection, i.e. $A \cap B = \varnothing \Rightarrow P(A \cap B) = 0$.

 Independent means that the outcome of event A will not influence the outcome of event B, i.e. $P(A \cap B) = P(A) \times P(B)$

4. Independence need not be for only two events. It can be extended, i.e. if the events A, B and C are each independent of each other then:

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$$

5. Showing that two events, A and B, are independent, requires three steps:

Step 1 Evaluate the product $P(A) \times P(B)$.

Step 2 Determine the value of $P(A \cap B)$ using any means (other than step 1), i.e. use grids, tables, Venn diagrams, . . . i.e. you must not assume anything about A and B.

Step 3 If the answer using Step 1 is equal to the answer obtained in Step 2, then and only then will the events be independent. Otherwise, they are not independent.

Notice that not being independent does not therefore mean that they are mutually exclusive. They simply aren't independent. That's all.

6. Do not confuse the multiplication principle with the rule for independence:

Multiplication principle is $P(A \cap B) = P(A|B) \times P(B)$.

Independence is given by $P(A \cap B) = P(A) \times P(B)$.

Example 5.4.5

Two fair dice are rolled. Find the probability that two even numbers will show up.

Let the E_1 and E_2 denote the events 'An even number on the first die.' and 'An even number on the second die.' respectively. In this case, the events are physically independent, i.e. the outcome on one die will not influence the outcome on the other die, and so we can confidently say that E_1 and E_2 are independent events.

Therefore, we have:

$$P(E_1 \text{ and } E_2) = P(E_1 \cap E_2) = P(E_1) \times P(E_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Example 5.4.6

Debra has a chance of 0.7 of winning the 100 m race and a 60% chance of winning the 200 m race.

a Find the probability that she only wins one race.

b Find the probability that she wins both races.

a Let W_1 denote the event 'Debra wins the 100 m race' and W_2, the event 'Debra wins the 200 m race'.

If Debra wins only one race she must either:

 win the 100 m **and** lose the 200 m **or**

 win the 200 m **and** lose the 100 m.

That is, we want:

$$P(W_1 \cap W_2') = P(W_1) \times P(W_2') = 0.7 \times 0.4 = 0.28$$

or we can multiply the probabilities because the events are independent (why?):

$$P(W_2 \cap W_1') = P(W_2) \times P(W_1') = 0.6 \times 0.3 = 0.18.$$

Therefore, the required probability is $0.28 + 0.18 = 0.46$

Notice that if W_1 and W_2 are independent, then so too are their complements.

b Winning both races means that Debra will win the 100 m **and** 200 m race.

Therefore, we have:

$$P(W_1 \cap W_2) = P(W_1) \times P(W_2) = 0.7 \times 0.6 = 0.42.$$

Notice how we have made repeated use of the word '**and**'. This emphasizes the fact that we are talking about the intersection of events.

Example 5.4.7

Four seeds are planted, each one having an 80% chance of germinating. Find the probability that:

a all four seeds will germinate

b at least one seed will germinate.

a Let G_i denote the event that the i th seed germinates.

This means that $P(G_1) = P(G_2) = P(G_3) = P(G_4) = 0.8$

It is reasonable to assume that each seed will germinate independently of the other.

Therefore, P(All four seeds germinate) =

$$P(G_1 \cap G_2 \cap G_3 \cap G_4) = P(G_1) \times P(G_2) \times P(G_3) \times P(G_4)$$

$$= (0.8)^4$$

$$= 0.4096$$

b Now, P(At least one seed will germinate) = 1 – p(No seeds germinate).

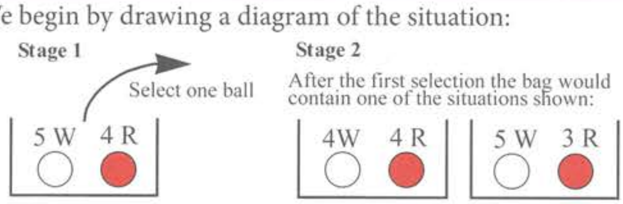
P(Any **one** seed does not germinate) $= P(G_i') = 0.2$

Therefore, P(At least one seed will germinate) $= 1 - (P(G_i'))^4 = 1 - (0.2)^4 = 0.9984$.

Example 5.4.8

A bag contains 5 white balls and 4 red balls. Two balls are selected in such a way that the first ball drawn is not replaced before the next ball is drawn. Find the probability of selecting exactly one white ball.

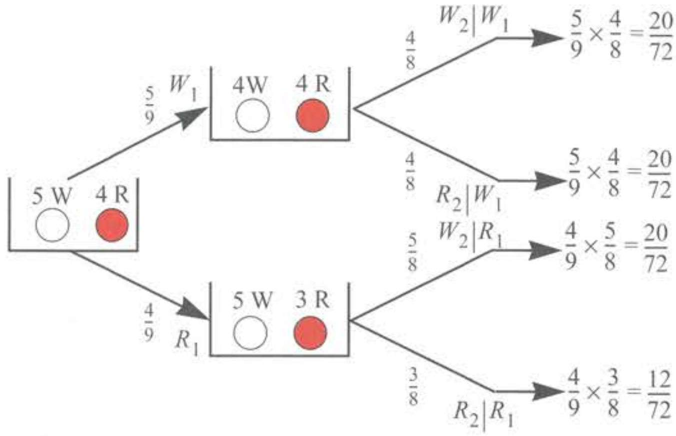
We begin by drawing a diagram of the situation:



From our diagram we notice that there are two possible sample spaces for the second selection.

As an aid, we make use of a tree diagram, where W_i denotes the event 'A white ball is selected on the ith trial' and R_i denotes the event 'A red ball is selected on the ith trial'.

The event 'Only one white' occurs if the first ball is white **and** the second ball is red, **or** the first ball is red **and** the second ball is white.



$$\begin{aligned} P(\text{One White ball}) &= P(W_1 \cap R_2) + P(R_1 \cap W_2) \\ &= P(R_2|W_1) \times P(W_1) + P(W_2|R_1) \times P(R_1) \\ &= \frac{4}{8} \times \frac{5}{9} + \frac{5}{8} \times \frac{4}{9} \\ &= \frac{5}{9} \end{aligned}$$

Exercise 5.4.1

1. Two events A and B are such that $p(A) = 0.6$, $p(B) = 0.4$ and $p(A \cap B) = 0.3$. Find the probability of the following events.

   a $A \cup B$ b $A|B$

   c $B|A$ d $A|B'$

2. A and B are two events such that $p(A) = 0.3$ $p(B) = 0.5$ and $p(A \cup B) = 0.55$. Find the probability of the following events:

   a $A|B$ b $B|A$

   c $A|B'$ d $A'|B'$

3. Urn A contains 9 cubes of which 4 are red. Urn B contains 5 cubes of which 2 are red. A cube is drawn at random and in succession from each urn.

   a Draw a tree diagram representing this process.

   b Find the probability that both cubes are red.

   c Find the probability that only 1 cube is red.

   d If only 1 cube is red, find the probability that it came from urn A.

4. A box contains 5 red, 3 black, and 2 white cubes. A cube is randomly drawn and has its colour noted. The cube is then replaced, together with 2 more of the same colour. A second cube is then drawn.

   a Find the probability that the first cube selected is red.

   b Find the probability that the second cube selected is black.

   c Given that the first cube selected was red, what is the probability that the second cube selected is black?

5. A fair coin, a double-headed coin and a double-tailed coin are placed in a bag. A coin is randomly selected. The coin is then tossed.

   a Draw a tree diagram showing the possible outcomes.

   b Find the probability that the coin lands with a tail showing uppermost.

   c In fact, the coin falls 'heads', find the probability that it is the 'double-headed' coin.

6. Two unbiased coins are tossed together. Find the probability that they both display heads given that at least one is showing a head.

7. A money box contains 10 discs, 5 of which are yellow, 3 of which are black and 2 green. Two discs are selected in succession, with the first disc not replaced before the second is selected.

   a Draw a tree diagram representing this process.

   b Hence find the probability that the discs will be of a different colour.

   c Given that the second disc was black, what is the probability that both were black?

8. Two dice are rolled. Find the probability that the faces are different given that the dice show a sum of 10.

9. Given that $p(A) = 0.6$, $p(B) = 0.7$ and that A and B are independent events.

   Find the probability of the events:

   a $A \cup B$ b $A \cap B$

   c $A|B'$ d $A' \cap B$

10. The probability that an animal will still be alive in 12 years is 0.55 and the probability that its mate will still be alive in 12 years is 0.60. Find the probability that:

    a both will still be alive in 12 years.

    b only the mate will still be alive in 12 years.

    c at least one of them will still be alive in 12 years.

    d the mate is still alive in 12 years given that only one is still alive in 12 years.

11. Tony has a 90% chance of passing his maths test, whilst Tanya has an 85% chance of passing the same test. If they both sit for the test, find the probability that:

    a only one of them passes.

    b at least one passes the test.

    c Tanya passed given that at least one passed.

12. The probability that Roger finishes a race is 0.55 and the probability that Melissa finishes the same race is 0.6. Because of team spirit, there is an 80% chance that Melissa will finish the race if Roger finishes the race. Find the probability that:

    a both will finish the race.

    b Roger finishes the race given that Melissa finishes.

13. If A and B are independent events, show that their complementary events are also independent events.

14. A student runs the 100 m, 200 m and 400 m races at the school athletics day. He has an 80% chance of winning any one given race. Find the probability that he will:

    a win all 3 races.

    b win the first and last race only.

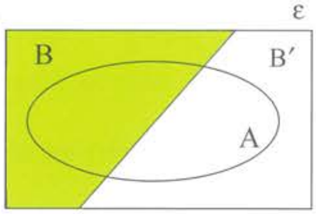
    c win the second race given that he wins at least two races.

Extra questions

Bayes' Theorem

Law of total probability

Using the Venn diagram, for any event A, we have that



$$A = A \cap \varepsilon = A \cap (B \cup B')$$

$$= (A \cap B) \cup (A \cap B')$$

As these two events are mutually exclusive, we have:

$$P(A) = P(A \cap B) + P(A \cap B')$$

However,

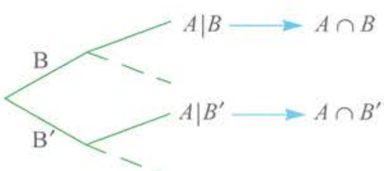
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B) \times P(A|B) \text{ and}$$

$$P(A|B') = \frac{P(A \cap B')}{P(B')} \Rightarrow P(A \cap B') = P(B') \times P(A|B'),$$

which leads to the **Law of Total Probability**.

$$P(A) = P(B) \times P(A|B) + P(B') \times P(A|B')$$

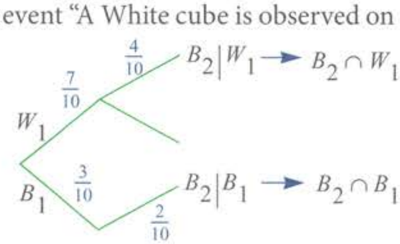
Although this expression may look daunting, in fact, it represents the result that we would obtain if a tree diagram was used.



Example 5.4.9

A box contains 3 black cubes and 7 white cubes. A cube is drawn from the box. Its colour is noted and a cube of the other colour is then added to the box. A second cube is then drawn. What is the probability that the second cube selected is black?

We begin by setting up a tree diagram, where B_i denotes the event "A Black cube is observed on ith selection" and W_i denotes the event "A White cube is observed on ith selection".



A black cube could have been observed on the second selection if:

i the first cube selected was white (i.e. $B_2|W_1$), or

ii the first cube selected was black (i.e. $B_2|B_1$).

$$\text{Therefore, } P(B_2) = P(B_2 \cap W_1) + P(B_2 \cap B_1)$$
$$= \frac{7}{10} \times \frac{4}{10} + \frac{3}{10} \times \frac{2}{10}$$
$$= \frac{17}{50}$$

Bayes' Theorem for two events

As we saw earlier, conditional probability provides a means by which we can adjust the probability of an event in light of new information. Bayes' Theorem, developed by Rev. Thomas Bayes, pictured, (1702–1761), does the same thing, except this time it provides a means of adjusting a set of associated probabilities in the light of new information.

For two events, we have:

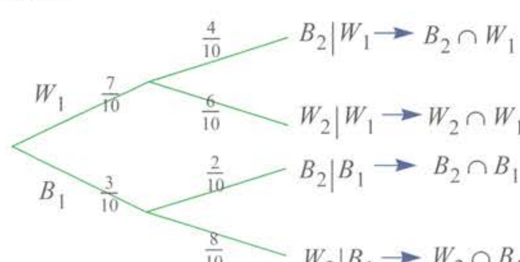
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B|A)}{P(A) \times P(B|A) + P(A') \times P(B|A')}$$

Again, the formula may seem daunting, however, it is only making use of a tree diagram.

Example 5.4.10

A box contains 3 black cubes and 7 white cubes. A cube is drawn from the box. Its colour is noted and a cube of the other colour is then added to the box. A second cube is then drawn. If both cubes are of the same colour, what is the probability that both cubes were in fact white?

Following on from the previous example, we have the same tree diagram:



We require: P(Both white **given** that both are of the same colour)

Now, the probability that they are of the same colour is given by the probability that they are **both white or both black**, i.e. $P((W_2 \cap W_1) \cup (B_2 \cap B_1))$.

Next: P(Both White **given** both are the same colour)

$$= P(W_2 \cap W_1 | (W_2 \cap W_1) \cup (B_2 \cap B_1))$$

$$= \frac{P((W_2 \cap W_1) \cap ((W_2 \cap W_1) \cup (B_2 \cap B_1)))}{P((W_2 \cap W_1) \cup (B_2 \cap B_1))}$$

$$= \frac{P(W_2 \cap W_1)}{P(W_2 \cap W_1) + P(B_2 \cap B_1)}$$

$$= \frac{P(W_2 | W_1) \times P(W_1)}{P(W_2 | W_1)P(W_1) + P(B_2 | B_1)P(B_1)}$$

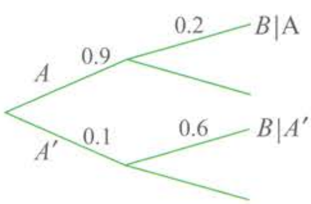
$$= \frac{\frac{6}{10} \times \frac{7}{10}}{\frac{6}{10} \times \frac{7}{10} + \frac{2}{10} \times \frac{3}{10}}$$

$$= \frac{7}{8}$$

Example 5.4.11

In a small country town, it was found that 90% of the drivers would always wear their seatbelts. On 60% of occasions, if a driver was not wearing a seatbelt they would be fined for speeding. If they were wearing a seatbelt, they would be fined for speeding 20% of the time. Find the probability that a driver who was fined for speeding was wearing a seatbelt.

Let the event A denote the event 'driver wears a seatbelt' and B denote the event 'Driver speeds'. Using a tree diagram we have:



We need to find, *Pr*(Driver was wearing a seatbelt| driver was booked for speeding):

$$= P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(A) \times P(B_| A)}{P(A) \times P(B|A) + P(A') \times P(B|A')}$$

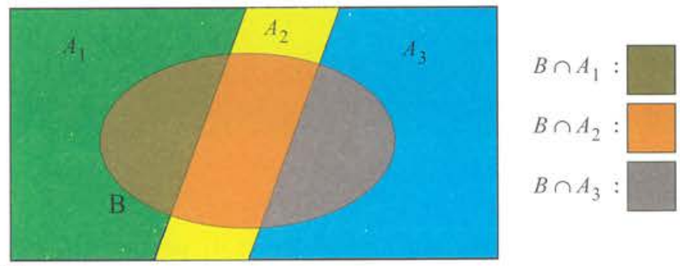
$$= \frac{0.9 \times 0.2}{0.9 \times 0.2 + 0.1 \times 0.6} = \frac{18}{24}$$

So, P(that a driver who was booked for speeding was in fact wearing a seatbelt) = 0.75

Bayes' Theorem for three events

So far we have used Bayes' Theorem for the case when the sample space is partitioned in two events, A and A′, where $A \cup A' = U$. However, this can be easily extended to the situation when the sample may be partitioned into many events. That is, $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = U$ where each of the events A_i are mutually exclusive.

So, let's consider the case when there are 3 events, so that the event A can be partitioned into three exhaustive, mutually exclusive subsets, i.e. $B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3)$.



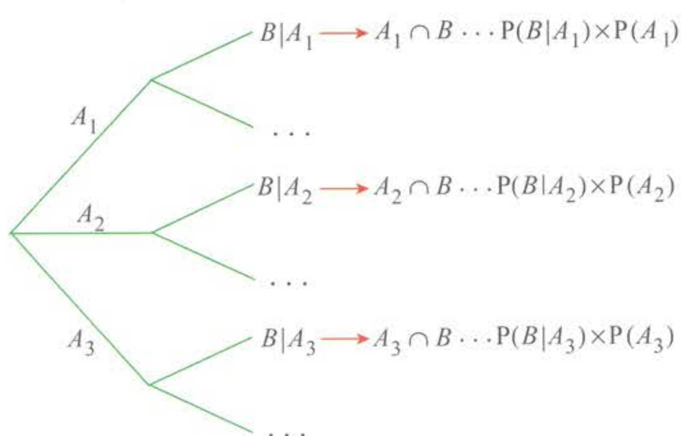
Then,

$$P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3))$$
$$= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$$
$$= \frac{P(B \cap A_1)}{P(A_1)} \times P(A_1) + \frac{P(B \cap A_2)}{P(A_2)} \times P(A_2) + \ldots$$
$$\ldots + \frac{P(B \cap A_3)}{P(A_3)} \times P(A_3)$$
$$= P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3)$$

Therefore, we have that:

$$P(A_1 | B) =$$
$$= \frac{P(A_1 \cap B)}{P(B)}$$
$$= \frac{P(B|A_1) \times P(A_1)}{P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3)}$$

As daunting as this expression may appear, all that we have done is add a new branch to our existing tree diagram. Everything remains the same. Making use of a tree diagram to help us evaluate the required probabilities is always useful. Such a diagram would have the following structure:



Example 5.4.12

It is found that the population within a particular state has a peculiar blood disease. Of this population, 2% have a serious form of the disease; 5% have a mild form; while 93% do not have it at all.

In carrying new trials, a new blood test is used, recording a positive result 92% of the time if the patient has the serious form of the disease; recording a positive result 60% of the time if the patient has the mild form of the disease, and recording a positive result 10% of the time if the patient does not have the disease.

A patient has just tested positive. What is the probability that this patient has the serious form of the disease?

Using the notation just discussed, we let A_1 denote the event 'Patient has disease in serious form', A_2 denote the event 'Patient has disease in mild form' and A_3 denote the event 'Patient does not have the disease'. Let B denote the event 'Records positive blood test'.

This gives, $P(A_1) = 0.02, P(A_2) = 0.05$ and $P(A_3) = 0.93$

$P(B|A_1) = 0.92, P(B|A_2) = 0.60, P(B|A_3) = 0.10$

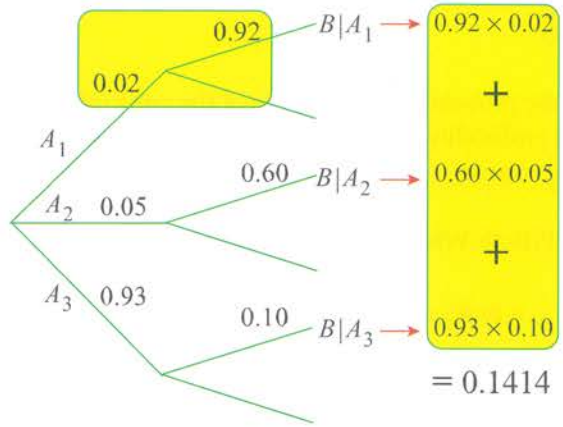
Then, as:

$$\begin{aligned} P(B) &= P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3) \\ &= 0.92 \times 0.02 + 0.6 \times 0.05 + 0.10 \times 0.93 \\ &= 0.1414 \end{aligned}$$

Using Bayes' Theorem, we have:

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1 \cap B)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3)} \\ &= \frac{0.92 \times 0.02}{0.1414} \\ &= 0.1301 \end{aligned}$$

Of course, we could have drawn a tree diagram to help with the example above:



As in the case for two events, drawing a tree diagram works very well. However, one needs to make sure to allocate the correct probabilities to the correct branches.

Next we consider a problem with three consecutive events producing three levels of branches, each identifying two possible outcomes.

Example 5.4.13

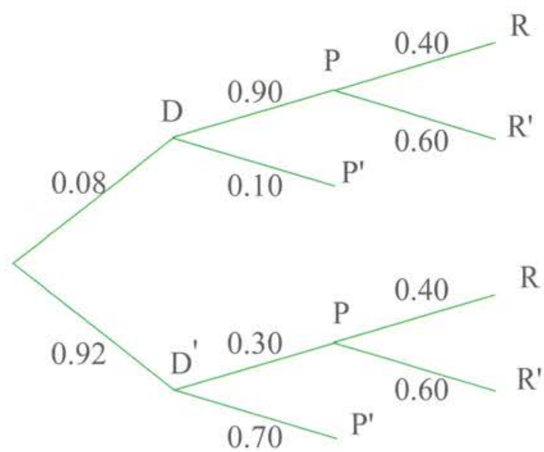
It is found that 8% of the population in a particular city has a disease, 'D'. A test is administered with the following outcomes: test returns a positive result 90% of the time if a person has the disease and returns a positive result 30% of the time if the person does not have the disease.

If a positive result is recorded a second drug is administered, which will cure the disease. However, this drug has the side effect of producing a nasty rash 40% of the time.

A randomly selected person is found to have the rash. What is the probability s/he had the disease to start with?

Let 'D' denote the event that a person has the disease, 'P' denote the event that a person will produce a positive reading and 'R' the event that the person develops a rash.

From the given information, we can produce the following tree diagram (leaving out irrelevant information):



$$P(D|R) = \frac{P(D \cap R)}{P(R)}$$
$$= \frac{0.08 \times 0.90 \times 0.40}{0.08 \times 0.90 \times 0.40 + 0.92 \times 0.30 \times 0.40}$$
$$= 0.2069$$

Again, notice how a tree diagram was most helpful in producing a neat and compact solution.

Exercise 5.4.2

1. Machine A produces 40% of the daily output of a factory but 3% of the items manufactured from this machine are defective. Machine B produces 60% of the daily output of the same factory but 5% of the items manufactured from this machine are defective.

   a An item is selected at random. Find the probability that it is defective.

   b An item is selected and is found to be defective. Find the probability that it came from machine B.

2. At the Heights International School, it is found that 12% of the male students and 7% of the female students are taller than 1.8 m. Sixty per cent of the school is made up of female students.

   a A student selected at random is found to be taller than 1.8m. What is the probability that the student is a female?

   b A second student selected at random is found to be shorter than 1.8m. What is the probability that the student is a male?

3. A box contains 4 black cubes and 6 white cubes. A cube is drawn from the box. Its colour is noted and a cube of the other colour is then added to the box. A second cube is then drawn.

   a If both cubes are of the same colour, what is the probability that both cubes were in fact white?

   b The first cube is replaced before the second cube is added to the box. What is the probability that both cubes were white given that both cubes were of the same colour?

4. An urn, labelled A, contains 8 cards numbered 1 through 8 whilst a second urn, labelled B, contains five cards numbered 1 through five. An urn is selected at random and from that urn a card is selected. Find the probability that the card came from urn A given that it is an even numbered card.

5. An event A can occur only if one of the mutually exclusive events B_1, B_2 or B_3 occurs. Show that $P(A) =$
$P(B_1) \times P(A|B_1) + P(B_2) \times P(A|B_2) + P(B_3) \times P(A|B_3)$

   b Of the daily output, machines A and B produce items of which 2% are defective, whilst machine C produces items of which 4% are defective. Machines B and C produce the same number of items, whilst machine A produces twice as many items as machine B.

   i An item is selected at random. Find the probability that it is defective.

   ii An item is selected and is found to be defective. Find the probability that it came from machine B.

6. A box contains N coins, of which m are fair coins whilst the rest are double-headed coins.

   a A coin is selected at random and tossed.

   i What is the probability of observing a head?

   ii Given that a head was observed, what is the probability that a double-headed coin was selected?

   b This time, a coin is selected at random and tossed n times. What is the probability that it is a fair coin, if it shows up heads on all n tosses?

7. A population of mice is made up of 75% that are classified as 'M+', of which, 30% have a condition classified as 'N–'. Otherwise, all other mice have the 'N–' condition. A mouse selected at random is classified as having the 'N–' condition. What is the probability that the mouse comes from the 'M+' classification group?

8. A survey of the adults in a town shows that 8% have liver problems. Of these, it is also found that 30% are heavy drinkers, 60% are social drinkers and 10% are non-drinkers. Of those that did not suffer from liver problems, 5% are heavy drinkers, 65% are social drinkers and 30% do not drink at all.

a An adult is selected at random. What is the probability that this person is a heavy drinker?

b If a person is found to be a heavy drinker, what is the probability that this person has liver problems?

c If a person is found to have liver problems, what is the probability that this person is a heavy drinker?

d If a person is found to be a non-drinker, what is the probability that this person has liver problems?

9. The probability that a person has a deadly virus is 5 in one thousand. A test will correctly diagnose this disease 95% of the time and incorrectly on 20% of occasions.

a Find the probability of this test giving a correct diagnosis.

b Given that the test diagnoses the patient as having the disease, what is the probability that the patient does not have the disease?

c Given that the test diagnoses the patient as not having the disease, what is the probability that the patient does have the disease?

10. The probability that a patient has a virus is 0.03. A medical diagnostic test will be able to determine whether the person in question actually has the virus. If the patient has the virus, the medical test will produce a positive result 90% of the time whilst if the patient does not have the virus, it will produce a negative result 98% of the time.

a What proportion of all tests provide a positive result?

b If the test shows a positive result, what is the probability that the patient actually has the virus?

c If the test shows a negative result, what is the probability that the patient does not have the virus?

Extra questions

Using Permutations and Combinations in Probability

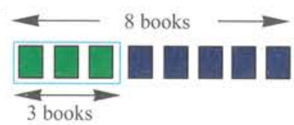
Because enumeration is such an important part of finding probabilities, a sound knowledge of permutations and combinations can help to ease the workload involved.

Example 5.4.14

Three maths books, three chemistry books and two physics books are to be arranged on a shelf. What is the probability that the three maths books are together?

The total number of arrangements of all 8 books is $8! = 40320$

To determine the number of arrangements that contain the three maths books together we make use of the box method:



We now have 6 boxes to arrange, giving a total of $6!$ arrangements.

However, the three maths books (within the blue box) can also be arranged in $3!$ ways.

Therefore, there are $6! \times 3! = 4320$ ways this can be done.

So, P(maths books are together) $= \dfrac{6! \times 3!}{8!} = \dfrac{6! \times 3!}{8 \times 7 \times 6!}$

$$= \frac{6}{8 \times 7}$$

$$= \frac{3}{28}$$

Example 5.4.15

A committee of 5 is randomly chosen from 8 boys and 6 girls. Find the probability that the committee consists of at least 3 boys.

The possibilities are:

Boys	Girls	No. of Selections
3	2	$\binom{8}{3} \times \binom{6}{2} = 840$
4	1	$\binom{8}{4} \times \binom{6}{1} = 420$
5	0	$\binom{8}{5} \times \binom{6}{0} = 56$

The total number of committees with at least 3 boys is 840 + 420 + 56 = 1316

However, the total number of committees of 5 from 14 is $\binom{14}{5} = 2002$.

If X denotes the number of boys on the committee, then $p(X \geq 3) = \frac{1316}{2002} = \frac{94}{143}$.

Exercise 5.4.3

1. Five red cubes and 4 blue cubes are placed at random in a row. Find the probability that:

   a the red cubes are together.

   b both end cubes are red.

   c the cubes alternate in colour.

2. Five books of different heights are arranged in a row. Find the probability that:

   a the tallest book is at the right-hand end.

   b the tallest and shortest books occupy the end positions.

   c the tallest and shortest books are together.

   d the tallest and shortest books are never next to each other.

3. Three cards are dealt from a pack of 52 playing cards. Find the probability that:

   a two of the cards are kings.

   b all three cards are aces.

   c all three cards are aces given that at least one card is an ace.

4. The letters of the word LOTTO are arranged in a row. What is the probability that the Ts are together?

5. A committee of 4 is to be selected from 7 men and 6 women. Find the probability that:

   a there are 2 women on the committee.

   b there is at least one of each sex on the committee.

6. A basketball team of 5 is to be selected from 12 players. Find the probability that:

   a the tallest player is selected.

   b the captain and vice-captain are selected.

   c either one, but not both of the captain or vice-captain are selected.

7. Find the probability of selecting one orange, one apple and one pear at random without replacement from a bag of fruit containing five oranges, four apples and three pears.

8. Three red cubes, four blue cubes and six yellow cubes are arranged in a row. Find the probability that:

   a the cubes at each end are the same colour.

   b the cubes at each end are of a different colour.

9. A sample of three light bulbs is selected from a box containing 15 light bulbs. It is known that five of the light bulbs in the box are defective.

   a Find the probability that the sample contains a defective.

   b Find the probability that the sample contains at least two defectives.

10. Eight people of different heights are to be seated in a row. What is the probability that:

    a the tallest and shortest persons are sitting next to each other?

    b the tallest and shortest occupy the end positions?

    c there are at least three people sitting between the tallest and shortest?

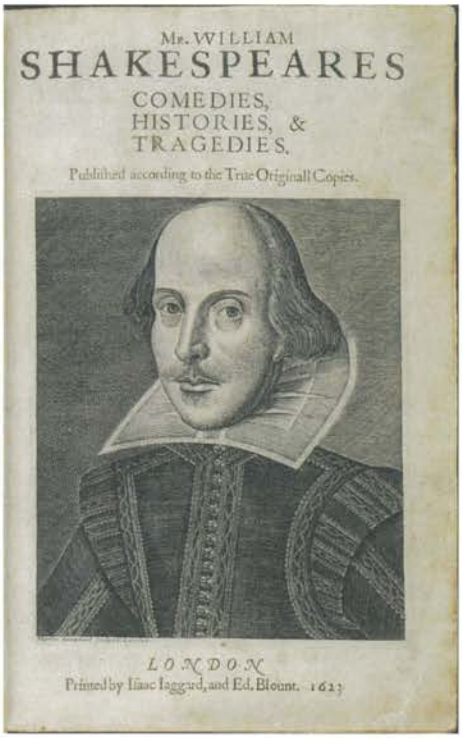
Extra questions

Answers

Theory of Knowledge

Very large numbers and Shakespeare

It has been said that a roomful of monkeys bashing randomly at keyboards will evetually type out the complete works of William Shakespeare.



Is this so?

My keyboard has about 50 keys.

Suppose we set about waiting for our monkey to type "To be or not to be, that is the question".

With spaces, this is 40 characters. The key presses are independent, and so the chance of getting the whole of this bit right on a single 'go' is:

$$\left(\frac{1}{50}\right)^{40} \approx 10^{-68}$$

The number of 'go's needed to type all these permutations is the reciprocal of this number or 10^{68}.

If our monkey got very lucky, it might get the right answer in one thousandth of this number, or 10^{65} 'go's.

How long would this take?

If each 'go' takes ten seconds, this is:

$$\frac{10^{65}}{6} \approx 1.7\times10^{64} \text{ minutes}$$

$$\text{or } \frac{1.7\times10^{64}}{60} \approx 1.7\times10^{62} \text{ hours}$$

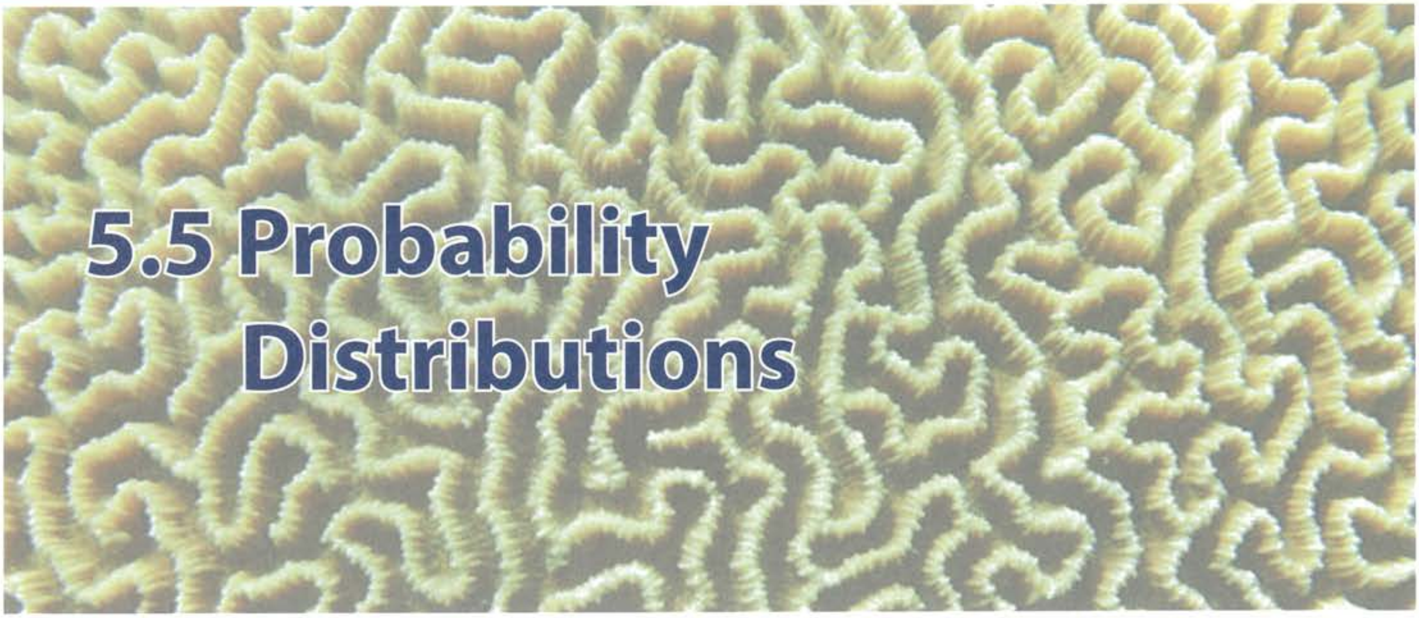
$$\text{or } \frac{1.7\times10^{62}}{24} \approx 1.2\times10^{61} \text{ days}$$

$$\text{or } \frac{1.2\times10^{61}}{365} \approx 3.2\times10^{58} \text{ years}$$

Since the age of the universe is estimated to be 1.4×10^{10} years, it is clear that, even with an army of monkeys typing away frantically, we would stand no chance at all of assembling this one famous quote.

Very large numbers turn up from time to time in these probability calculations. When you find one, is it 'real'?

5.5 Probability Distributions



Discrete Random Variables

Concept of a random variable

Consider the experiment of tossing a coin twice. The sample space, S, (i.e. the list of all possible outcomes) of this experiment can be written as S = {HH, HT, TH, TT}.

We can also assign a numerical value to these outcomes. For example, we can assign the number

0 to the outcome {HH},

1 to the outcomes {HT, TH} and

2 to the outcome {TT}.

These numerical values are used to represent the number of times that a tail was observed **after** the coin was tossed.

The numbers 0, 1 and 2 are **random** in nature, that is, until the coins are tossed we have no idea as to which one of the outcomes will occur. We define a random variable as follows:

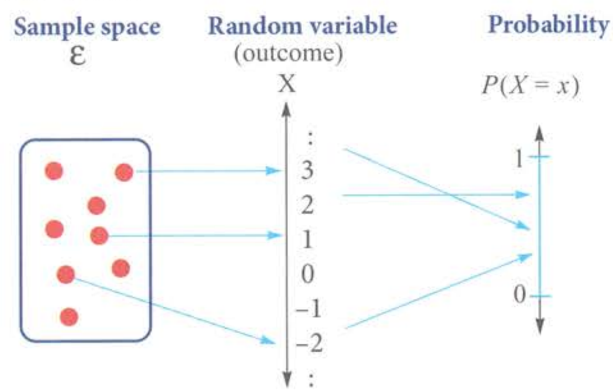
A **random variable, *X*** (random variables are usually denoted by capital letters), which can take on exactly n numerical values, each of which corresponds to one and only one of the events in the sample space is called a **discrete random variable**.

Note that the values that correspond to the random variables X, Y, Z ... are denoted by their corresponding lower case letters, x, y, z ... For the example above, $X = \{x: x = 0, x = 1, x = 2\}$.

Discrete random variable

A **discrete random** variable is one in which we can produce a **countable** number of outcomes. X of these discrete random variables are usually associated with a **counting process**. For example, the number of plants that will flower, the number of defective items in a box or the number of items purchased at a supermarket store.

We can display this concept using a simple diagram such as the one below:



We write
'The **probability** that the random variable **X = x** is **p**' as:

$$P(X = x) = p.$$

Note that the probability of any event must always lie between 0 and 1, inclusive.

To obtain the sample space, we may need to carry out an experiment. However, as we shall see, there are many types of random variables that already possess their own sample spaces, random outcomes and associated probability values. We shall deal with these later.

Example 5.5.1

Consider the experiment of tossing a coin three times in succession.

a List all possible outcomes.

b If the random variable X denotes the number of heads observed, list the values that X can have and find the corresponding probability values.

In the experiment of tossing a coin three times the sample space is given by

S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT},

where the event {HTH} represents the observation, head, tail, head, in that order.

This means that on any one trial of this experiment, we could have obtained 0 heads, 1 head, 2 heads or 3 heads.

Therefore, the random variable X has as its possible values the numbers 0, 1, 2, 3.

That is:

$X = 0$ corresponds to the event {TTT}, that is, no heads.

$X = 1$ corresponds to the events {TTH, THT, HTT}, that is, one head.

$X = 2$ corresponds to the events {THH, HTH, HHT}, that is, two heads.

$X = 3$ corresponds to the event {HHH}, that is, three heads.

Once we have our sample space, we can look at the chances of each of the possible outcomes. In all there are 8 possible outcomes.

The chances of observing the event {HHH} would be $\frac{1}{8}$, i.e. $P(X = 3) = \frac{1}{8}$.

To find $P(X = 2)$, we observe that the outcome 'X = 2' corresponds to {THH, HTH, HHT}.

In this case there is a chance of 3 in 8 of observing the event where '$X = 2$'.

Continuing in this manner we have:

$$P(X = 0) = P(\{\text{HHH}\}) = \frac{1}{8}$$

$$P(X = 1) = P(\{\text{TTH, THT, HTT}\}) = \frac{3}{8}$$

$$P(X = 2) = P(\{\text{HHT, HTH, THH}\}) = \frac{3}{8}$$

$$P(X = 3) = P(\{\text{HHH}\}) = \frac{1}{8}$$

Probability Distributions

We can describe a discrete random variable by making use of its probability distribution. That is, by showing the values of the random variable and the probabilities associated with each of its values.

A probability distribution can be displayed in any one of the following formats:

1. Tabular form

2. Graphical representation

(*With the probability value on the vertical axis, and the values of the random variable on the horizontal axis.*)

3. Function

(*A formula that can be used to determine the probability values.*)

Example 5.5.2

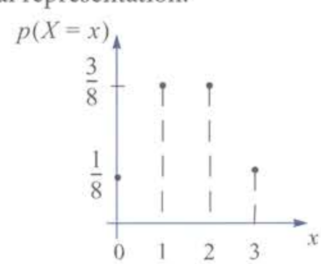
Use each of the probability distribution representations discussed to display the results of the experiment where a coin is tossed three times in succession.

Let the random variable X denote the number of heads observed in three tosses of a coin.

1. Tabular form:

x	0	1	2	3
$p(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

2. Graphical representation:



3. Function:

$p(X = x) = \binom{3}{x}\left(\frac{1}{2}\right)^3, x = 0, 1, 2, 3$, where:

$\binom{3}{x} = \frac{3!}{(3-x)!x!}$.

Properties of the probability function

We can summarise the features of any discrete probability function as follows:

1. The probability for any value of X must **always lie between** 0 and 1 (inclusive).

   That is, $0 \le P(X = x_i) \le 1$ for all values of x_i.

2. For the n mutually exclusive and exhaustive events, $A_1, A_2, \ldots, A_n$ that make up the sample space ε, then, the sum of the corresponding probabilities must be 1.

That is,

$$\sum_{i=1}^{i=n} P(X = x_i) = P(X = x_1) + P(X = x_2) + \ldots + P(X = x_n) = 1$$

Where $P(X = x_i)$ is the probability of event A_i occurring.

Any function that does not abide by these two rules cannot be a probability function.

Example 5.5.3

Consider the random variable X with probability function defined by

$P(X = 0) = \alpha$, $P(X = 1) = 2\alpha$ and $P(X = 2) = 3\alpha$

Determine the value of α.

Because we are given that this is a probability function, then summing all the probabilities must give a result of 1.

Therefore we have that:

$$P(X = 0) + P(X = 1) + P(X = 2) = 1$$

$$\therefore \alpha + 2\alpha + 3\alpha = 1$$

$$\Leftrightarrow 6\alpha = 1$$

$$\Leftrightarrow \alpha = \frac{1}{6}$$

Example 5.5.4

The probability distribution of the random variable X is represented by the function

$$P(X = x) = \frac{k}{x}, x = 1, 2, 3, 4, 5, 6.$$

Find:

a the value of k b $P(3 \le X \le 5)$.

Using the fact that the sum of all the probabilities must be 1, we have:

$$P(X = 1) + P(X = 2) + \ldots + P(X = 6)$$

$$= \frac{k}{1} + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k}{5} + \frac{k}{6} = 1$$

$$\Leftrightarrow \frac{147k}{60} = 1$$

Therefore, $k = \frac{60}{147} = \frac{20}{49}$.

Now, $P(3 \le X \le 5) = P(X = 3) + P(X = 4) + P(X = 5)$

$$= \frac{k}{3} + \frac{k}{4} + \frac{k}{5}$$

$$= \frac{47k}{60}$$

However, we know that $k = \frac{60}{147}$.

Therefore, $P(3 \le X \le 5) = \frac{47}{60} \times \frac{60}{147} = \frac{47}{147}$.

Example 5.5.5

A discrete random variable X has a probability distribution defined by the function:

$P(X = x) = \binom{4}{x}\left(\frac{2}{5}\right)^x\left(\frac{3}{5}\right)^{4-x}$ where $x = 0, 1, 2, 3, 4$

Display this distribution using:

a i a table form ii a graphical form.

b Find: i $P(X = 2)$ ii $P(1 \le X \le 3)$.

a i We begin by evaluating the probability for each value of x: (Using the notation $\binom{4}{x} = {}^4C_x$)

$$p(X = 0) = {}^4C_0\left(\frac{2}{5}\right)^0\left(\frac{3}{5}\right)^{4-0} = \frac{81}{625},$$

$$p(X = 1) = {}^4C_1\left(\frac{2}{5}\right)^1\left(\frac{3}{5}\right)^3 = \frac{216}{625},$$

$$p(X=2) = {}^4C_2\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^2 = \frac{216}{625},$$

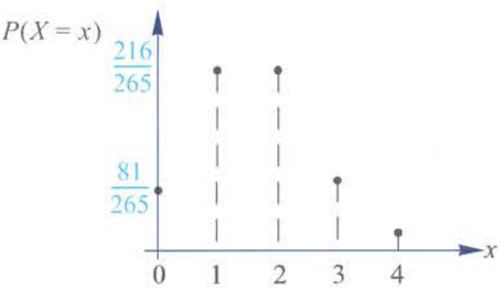
$$p(X=3) = {}^4C_3\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^1 = \frac{96}{625},$$

$$p(X=4) = {}^4C_4\left(\frac{2}{5}\right)^4\left(\frac{3}{5}\right)^0 = \frac{16}{625}$$

We can now set up this information in a table:

x	0	1	2	3	4
$P(X=x)$	$\frac{81}{625}$	$\frac{216}{625}$	$\frac{216}{625}$	$\frac{96}{625}$	$\frac{16}{625}$

ii Using the table found in part i, we can construct the following graph:



i From our probability table: $P(X=2) = \frac{216}{625}$.

ii The statement $P(1 \le X \le 3)$ requires that we find the probability of the random variable X taking on the values 1, 2 or 3. This amounts to evaluating the sum of the corresponding probabilities.

Therefore: $P(1 \le X \le 3) = P(X=1) + P(X=2) + P(X=3)$

$$= \frac{216}{625} + \frac{216}{625} + \frac{96}{625}$$

$$= \frac{528}{625}$$

Constructing Probability Functions

When we are given the probability distribution, we can determine the probabilities of events. However, there is still one issue that we must deal with:

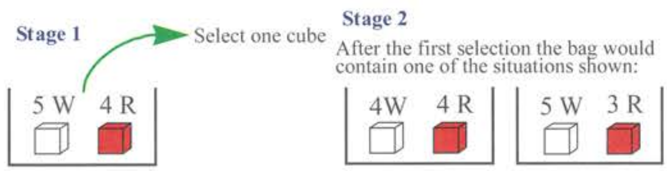
How do we obtain the probabilities in the first place?

Sometimes we recognise a particular problem and know of an existing model that can be used. However, resolving this question is not always an easy task, as this often requires the use of problem-solving skills and modelling skills as well as interpretive skills.

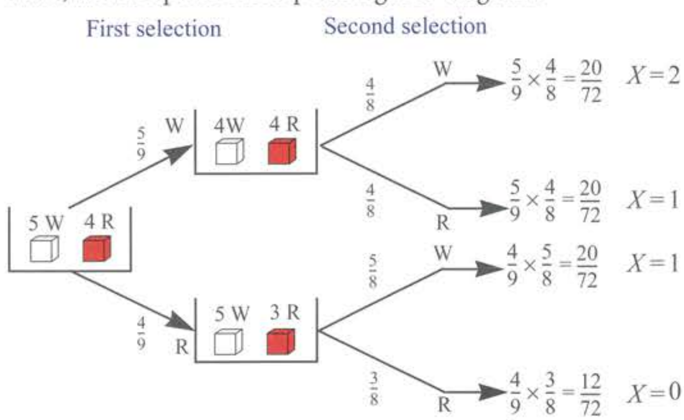
Example 5.5.6

A bag contains 5 white cubes and 4 red cubes. Two cubes are selected in such a way that the first cube drawn is not replaced before the next cube is drawn. Find the probability distribution of X, where X denotes the number of white cubes selected from the bag.

We start by drawing a diagram that will help us visualise the situation:



Next, we set up the corresponding tree diagram:



We are now in a position to complete the probability table.

x	0	1	2
$P(X=x)$	$\frac{12}{72}$	$\frac{40}{72}$	$\frac{20}{72}$

Example 5.5.7

Two friends, Kirsty and Bridget, independently applied for different jobs. The chance that Kirsty is successful is 0.8 and the chance that Bridget is successful is 0.75.

If X denotes the number of successful applications between the two friends, find the probability distribution of X.

Hence find the probability that:

i both are successful

ii that if one is successful, it is Kirsty.

a Let K denote the event that Kirsty is successful, so that $P(K) = 0.8$ and let B denote the event that Bridget is successful, so that $P(B) = 0.75$.

Now, the event '$X = 0$' translates to 'nobody is successful':

That is, $P(X=0) = P(K' \cap B') = P(K') \times P(B')$.

$= 0.2 \times 0.25 = 0.05$

Similarly, the event '$X = 1$' translates to 'only one is successful':

That is, $P(X=1) = P(K \cap B') + P(K' \cap B)$

$= 0.8 \times 0.25 + 0.2 \times 0.75 = 0.35$

Lastly, the event '$X = 2$' translates to 'both are successful':

That is,

$$P(X=2) = P(K \cap B) = P(K) \times P(B) = 0.8 \times 0.75 = 0.6$$

We can now construct a probability distribution for the random variable X:

x	0	1	2
$P(X=x)$	0.05	0.35	0.60

b i $P(\text{Both successful}) = P(X=2) = 0.60$

$$P(\text{K}|\text{Only one is successful}) = \frac{P(K \cap \{X=1\})}{P(\{X=1\})} = \frac{P(K \cap B')}{P(\{X=1\})}$$

$$= \frac{0.20}{0.35} (= 0.5714)$$

Exercise 5.5.1

1. Find the value of k, so that the random variable X describes a probability distribution.

x	1	2	3	4	5
$P(X=x)$	0.25	0.20	0.15	k	0.10

2. The discrete random variable Y has the following probability distribution:

y	1	2	3	4
$P(Y=y)$	β	2β	3β	4β

Find the value of β.

Find: i $P(Y=2)$ ii $P(Y>2)$

3. A delivery of six television sets contains 2 defective sets. A customer makes a random purchase of two sets from this delivery. The random variable X denotes the number of defective sets purchased by the customer.

a Find the probability distribution table for X.

b Represent this distribution as a graph.

c Find $P(X \le 1)$.

4. A pair of dice is rolled. Let Y denote the sum showing uppermost.

a Determine the possible values that the random variable Y can have.

b Display the probability distribution of Y in tabular form.

c Find $P(Y=8)$.

d Sketch the probability distribution of Y.

5. A fair coin is tossed 3 times.

a Draw a tree diagram representing this experiment.

b Display this information using both graphical and tabular representations.

c If the random variable Y denotes the number of heads that appear uppermost, find $P(Y \ge 2 | Y \ge 1)$.

6. The number of customers that enter a small corner newsagency between the hours of 8 p.m. and 9 p.m. can be modelled by a random variable X having a probability distribution given by $P(X=x) = k(3x+1)$, where $x = 0, 1, 2, 3, 4$.

a Find the value of k.

b Represent this distribution in:
i tabular form ii graphical form.

c What are the chances that at least 2 people will enter the newsagency between 8 p.m. and 9 p.m. on any one given day?

7. The number of cars passing an intersection during the hours of 4 p.m. and 6 p.m. follows a probability distribution that is modelled by the function

$$P(X=x) = \frac{(0.1)^x}{x!} e^{-0.1}, x = 0, 1, 2, 3, \ldots,$$

where the random variable X denotes the number of cars that pass this intersection between 4 p.m. and 6 p.m.

a Find: i $P(X = 0)$ ii $P(X = 1)$.

b Find the probability of observing at least three cars passing through this intersection between the hours of 4 p.m. and 6 p.m.

8. The number of particles emitted during a one-hour period is given by the random variable X, having a probability distribution

$$P(X = x) = \frac{(4)^x}{x!}e^{-4}, x = 0, 1, 2, 3, \ldots .$$

Find $P(X > 4)$.

9. A random variable X has the following probability distribution

x	0	1	2	3
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{2}{15}$

a Find the probability distribution of $Y = X^2 - 2X$

b Find: i $P(Y = 0)$ ii $P(Y < 3)$.

10. A bakery has six indistinguishable muffins on display. However, two of them have been filled with strawberry jam and the others with apricot jam. Claire, who hates strawberry jam, purchases two muffins at random. Let N denote the number of strawberry jam muffins Claire buys. Find the probability distribution of the random variable N.

Extra questions

Mean and Variance

Central tendency and expectation

For a discrete random variable X with a probability distribution defined by $P(X = x)$, we define the **expectation of the random variable X** as

$$E(X) = \sum_{i=1}^{i=n} x_i P(X = x_i)$$
$$= x_1 \times P(X = x_1) + x_2 \times P(X = x_2) + \ldots + x_n \times P(X = x_n)$$

Where $E(X)$ is read as "The expected value of X". $E(X)$ is interpreted as the **mean value** of X and is often written as μ_X (or simply μ). Often we write the expected value of X as $\Sigma x.P(X = x)$. This is in contrast to the **mode** which is the most common value(s) and the **median** which is the value with half the probabilities below and half above the median value.

So what exactly does *E*(*X*) measure?

The expected value of the random variable is a **measure of the central tendency** of X. That is, it is an indication of its 'central position' – based not only on the values of X, but also the **probability weighting** associated with each value of X. That is, it is the probability-weighted average of its possible values.

To find the value of $E(X)$ using the formula:

$$E(X) = \sum_{i=1}^{i=n} x_i P(X = x_i)$$

we take each possible value of x_i, multiply it by its associated probability $P(X = x_i)$ (i.e. its 'weight') and then add the results. The number that we obtain can be interpreted in two ways:

As a **probability-weighted average**, it is a summary number that takes into account the relative probabilities of each x_i value.

As a **long-run average**, it is a measure of what one could expect to observe if the experiment were repeated a large number of times.

For example, when tossing a fair coin a large number of times (say 500 times), and the random variable X denotes the number of tails observed, we would expect to observe 250 tails, i.e. $E(X) = 250$.

Note: *Although we would expect 250 tails after tossing a coin 500 times, it may be that we do not observe this outcome! For example, if the average number of children per 'family' in Australia is 2.4, does this mean we expect to see*

2.4 children per 'family'?

In short, we may not be able to observe the value $E(X)$ that we obtain.

Example 5.5.8

For the random variable X with probability distribution defined by:

x	1	2	3	4
$P(X=x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$

Find the mode, median and mean values of X.

The mode is $X = 4$ (the most probable) and the median 3 (half are above and half below 3 – this is probably best done by sketching the cdf of X).

To find the mean of X we use the formula:

$$E(X) = \sum_{i=1}^{i=n} x_i P(X = x_i)$$

$$\begin{aligned} E(X) &= \left(1\times\frac{1}{10}\right)+\left(2\times\frac{2}{10}\right)+\left(3\times\frac{3}{10}\right)+\left(4\times\frac{4}{10}\right) \\ &= \frac{1}{10}+\frac{4}{10}+\frac{9}{10}+\frac{16}{10} \\ &= 3 \end{aligned}$$

Therefore, X has a mean value of 3.

Example 5.5.9

A fair die is rolled once. If the random variable X denotes the number showing, find the expected value of X.

Because the die is fair we have the following probability distribution:

x	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\begin{aligned} E(X) &= \left(1\times\frac{1}{6}\right)+\left(2\times\frac{1}{6}\right)+\left(3\times\frac{1}{6}\right)+\left(4\times\frac{1}{6}\right)+\left(5\times\frac{1}{6}\right)+\left(6\times\frac{1}{6}\right) \\ &= 3.5 \end{aligned}$$

Variance

Although we now have a means by which we can calculate a measure of the central tendency of a random variable, an equally important aspect of a random variable is its spread. For example, the mean of the three numbers, 100, 110 and 120 is 110. Similarly, the mean of the three numbers 10, 100, and 220 is also 110. Yet clearly, the values in the second set of data have a wider spread than those in the first set of data. The **variance** (or more so, the **standard deviation**) provides a better measure of this spread.

The **variance** of a discrete random variable may be considered as the average of the squared deviations about the mean. This provides a measure of the variability of the random variable or the probability dispersion. The variance associated with the random variable X is given by:

$$Var(X) = E((X-\mu)^2) = \sum_{i=1}^{i=n} (x-\mu)^2 P(X=x)$$

However, for computational purposes, it is often better to use the alternative definition:

$$Var(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$$

The variance is also denoted by σ^2 (read as "**sigma squared**"), i.e. $Var(X) = \sigma^2$.

We also have the **standard deviation**, given by

$$Sd(X) = \sigma = \sqrt{Var(X)},$$

which also provides a measure of the spread of the distribution of X.

What is the difference between the $Var(X)$ and the Sd(X)?

Because of the squared factor in the equation for $Var(X)$ (i.e. $Var(X) = E(X^2) - \mu^2$), the units of $Var(X)$ are not the same as those for X. However, because the $Sd(X)$ is the square root of the $Var(X)$, we have "adjusted" the units of $Var(X)$ so that they now have the same units as the random variable X.

The reason for using the $Sd(X)$ rather than the $Var(X)$ is that we can make clearer statistical statements about the random variable X (in particular, statements that relate to an overview of the distribution).

Example 5.5.10

The probability distribution of the random variable X is shown below:

x	-2	-1	0	1	2
$P(X=x)$	$\frac{1}{64}$	$\frac{12}{64}$	$\frac{38}{64}$	$\frac{12}{64}$	$\frac{1}{64}$

Find:

a the variance of X

b the standard deviation of X.

a First we need to find $E(X)$ and $E(X^2)$:

$$E(X) = \sum xP(X=x)$$
$$= -2\times\frac{1}{64}+(-1)\times\frac{12}{64}+0\times\frac{38}{64}+1\times\frac{12}{64}+2\times\frac{1}{64}$$
$$= 0$$
$$E(X^2) = \sum x^2P(X=x)$$
$$= (-2)^2\times\frac{1}{64}+(-1)^2\times\frac{12}{64}+0^2\times\frac{38}{64}+1^2\times\frac{12}{64}+2^2\times\frac{1}{64}$$
$$= \frac{1}{2}$$

b Therefore, $Var(X) = E(X^2)-\mu^2 = \frac{1}{2}-0^2 = \frac{1}{2}$.

Using $Sd(X) = \sigma = \sqrt{Var(X)}$, we have that:

$$Sd(X) = \sigma = \sqrt{\frac{1}{2}} \approx 0.707.$$

Exercise 5.5.2

1. A discrete random variable X has a probability distribution given by

x	1	2	3	4	5
$P(X=x)$	0.25	0.20	0.15	0.3	0.10

a Find the mean value of X.

b Find the variance of X.

2. The discrete random variable Y has the following probability distribution:

y	1	2	3	4
$P(Y=y)$	0.1	0.2	0.3	0.4

a Find the mean value of Y.

b Find: i $Var(Y)$ i $Sd(Y)$.

c Find: i $E(2Y)$ ii $E\left(\frac{1}{Y}\right)$.

3. A random variable X has the following probability distribution:

x	0	1	2	3
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{2}{15}$

a Find: i $E(X)$ ii $E(X^2)$

iii $E(X^2-2X)$.

b Find: i $Sd(X)$

ii $Var(3X+1)$

c If $Y = \frac{1}{X+1}$, find : i $E(Y)$ ii $E(Y^2)$

4. A delivery of six television sets contains 2 defective sets. A customer makes a random purchase of two sets from this delivery. The random variable X denotes the number of defective sets the customer purchased. Find the mean and variance of X.

5. A pair of dice are rolled. Let Y denote the sum showing uppermost.

   a Find $E(Y)$.

   b Find $Var(Y)$.

6. How many tails would you expect to observe when a fair coin is tossed 3 times?

7. The number of customers that enter a small corner newsagency between the hours of 8 p.m. and 9 p.m. can be modelled by a random variable X having a probability distribution given by $P(X = x) = k(2x + 1)$, where $x = 0, 1, 2, 3, 4$.

   a Find the value of k.

   b How many customers can be expected to enter the newsagency between 8 p.m. and 9 p.m.?

   c Find the standard deviation of X.

8. A discrete random variable Y has its probability distribution function defined as

y	-2	-1	0	1
$P(Y = y)$	k	0.2	$3k$	0.4

   a Find k.

   b Given that the function, F, is defined by $F(y) = P(Y \leq y)$, find:

   i $F(-1)$ ii $F(1)$.

   c Find:

   i the expected value of Y.

   ii the variance of Y.

   iii the expected value of $(Y+1)^2$.

9. A dart board consisting of concentric circles of radius 1, 2 and 3 units is placed against a wall. A player throws darts at the board, each dart landing at some random location on the board. The player will receive \$9.00 if the smaller circle is hit, \$7.00 if the middle annular region is hit and \$4.00 if the outer annular region is hit. Should players miss the board altogether, they would lose \$$k$ each time. The probability that the player misses the dart board is 0.5. Find the value of k if the game is to be fair.

10. A box contains 7 black cubes and 3 red cubes. Debra selects three cubes from the box without replacement. Let the random variable N denote the number of red cubes selected by Debra.

   a Find the probability distribution for N.

   b Find: i $E(N)$ ii $Var(N)$.

   Debra will win \$2.00 for every red cube selected and lose \$1.00 for every black cube selected. Let the random variable W denote Debra's winnings.

   c If $W = aN + b$, find a and b. Hence, find $E(W)$.

11.a A new gambling game has been introduced in a casino: A player stakes \$8.00 in return for the throw of two dice, where the player wins as many dollars as the sum of the two numbers showing uppermost.

   How much money can the player expect to walk away with?

   b At a second casino, a different gambling game has been set up: A player stakes \$8.00 in return for the throw of two dice, if two sixes come up, the player wins \$252.

   Which game would be more profitable for the casino in the long run?

12. Given that $Var(X) = 2$, find:

a $Var(5X)$ b $Var(-3X)$

c $Var(1-X)$.

13. Given that $Var(X) = 3$ and $\mu = 2$, find:

Extra questions

a $E(2X^2 - 4X + 5)$ b $Sd\left(4 - \frac{1}{3}X\right)$

c $E(X^2) + 1 - E((X+1)^2)$.

14. A store has eight toasters left in its storeroom. Three of the toasters are defective and should not be sold. A salesperson, unaware of the defective toasters, selects two toasters for a customer. Let the random variable N denote the number of defective toasters the customer purchases.

Answers

Find:

a $E(N)$ b $Sd(N)$.

15. a The random variable Y is defined by:

y	-1	1
$P(Y=y)$	p	$1-p$

Find the mean and variance of Y.

b The random variable X is defined as $X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$ where each Y_i $i = 1, 2, 3, \ldots, n$ is independent and has the distribution defined in part **a**.

Find: i $E(X)$ ii $Var(X)$

5.6 Binomial & Poisson Distribution

The Binomial Distribution

The Binomial Experiment

The binomial distribution is a special type of discrete distribution which finds applications in many settings of everyday life. In this section we summarise the important features of this probability distribution.

Bernoulli Trials

Certain experiments consist of repeated trials where each trial has only two mutually exclusive, possible outcomes. Such trials are referred to as **Bernoulli trials** (after Jacob Bernoulli - pictured). The outcomes of a Bernoulli trial are often referred to as 'a success' or 'a failure'. The terms 'success' and 'failure' in this context do not necessarily refer to the everyday usage of the word success and failure. For example, a 'success' could very well be referring to the outcome of selecting a defective transistor from a large batch of transistors.

We often denote P(Success) by p and P(Failure) by q, where $p + q = 1$ (or $q = 1 - p$).

Properties of the binomial experiment

1. There are a fixed number of trials. We usually say that there are n trials.
2. On each one of the n trials there is only one of two possible outcomes, labelled '*success*' and '*failure*'.
3. Each trial is identical and independent.
4. On each of the trials, the probability of a success, p, is always the same, and the probability of a failure, $q = 1 - p$, is also always the same.

The Binomial Distribution

If a (discrete) random variable X has all of the above mentioned properties, we say that X has a **binomial distribution**. The probability distribution function is given by

$$p(X = x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots n$$

Where X denotes the **number of successes** in n trials such that the **probability of a success on any one trial** is p, $0 \le p \le 1$ and $p + q = 1$ (or $q = 1 - p$)).

We can also express the binomial distribution in a compact form, written as $X \sim B(n,p)$, read as "**X is distributed binomially with parameters n and p**", where n is the number of trials and $p = P(success)$ [it is also common to use $X \sim \mathrm{Bin}(n, p)$].

For example, the probability function for $X \sim B(6,0.4)$ (i.e. 6 trials and $p = 0.4$) would be

$$P(X = x) = \binom{6}{x}(0.4)^x(0.6)^{6-x}, x = 0, 1, 2, \ldots n.$$

Example 5.6.1

If $X \sim B(5, 0.6)$, find $P(X = 4)$.

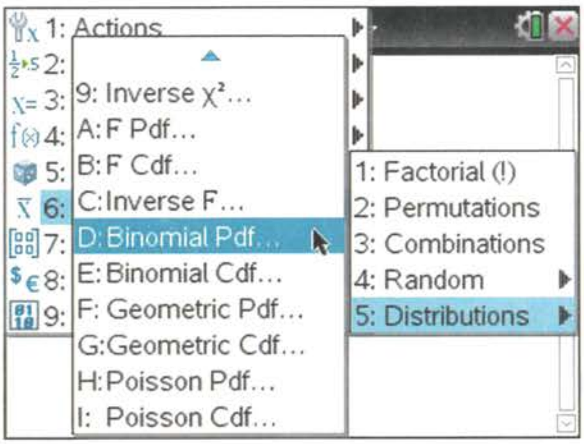
$X \sim B(5, 0.6)$ means that $P(X = 4)$ is the probability of 4 successes in five trials, where each trial has a 0.6 chance of being a success, that is, $n = 5$, $p = 0.6$ and $x = 4$.

$$\therefore P(X=4) = {}^5C_4(0.6)^4(0.4)^{5-4}$$
$$= \frac{5!}{1!4!}(0.6)^4(0.4)^1$$
$$= 0.2592$$

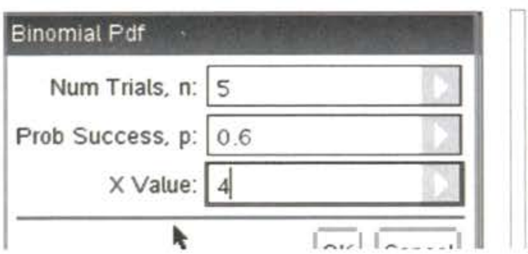
Most modern calculators can perform these calculations. TI NSpire calculators

MENU / 5. Probability / 5. Distributions / D. Binomial PDF

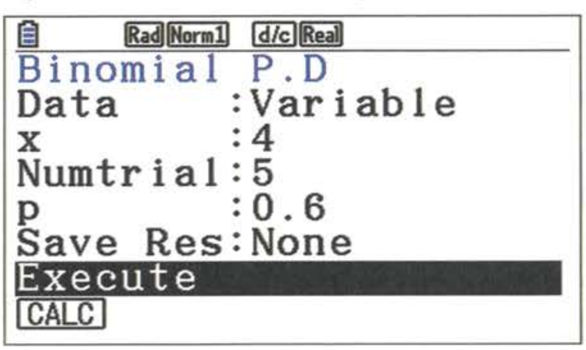
PDF stands for 'Probability Density function'. Make sure you are clear about this and CDF 'Cumulative Density function'.



Example 5.6.1 is solved:



If using the Casio Statistics module, press F5, F5, F1 and set:



F1-CALC completes the calculation.

Example 5.6.2

A manufacturer finds that 30% of the items produced from one of the assembly lines are defective. During a floor inspection, the manufacturer selects 6 items from this assembly line. Find the probability that the manufacturer finds:

a two defectives b at least two defectives.

a Let X denote the number of defectives in the sample of six. Therefore, we have that $n = 6$, $p(success) = p = 0.30$ ($\Rightarrow q = 1 - p = 0.70$), so that $X \sim B(6, 0.3)$.

Note that in this case, a 'success' refers to a defective.

i $P(X=2) = {}^6C_2(0.3)^2(0.7)^4 = 0.3241$.

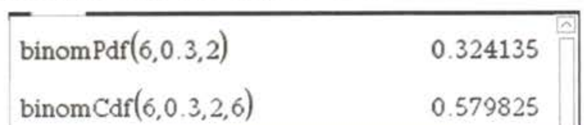
ii $P(X \geq 2) = P(X=2) + P(X=3) + \ldots + P(X=6)$

$$= {}^6C_2(0.3)^2(0.7)^4 + {}^6C_3(0.3)^3(0.7)^3 + \ldots + {}^6C_6(0.3)^6(0.7)^0$$

A second method makes use of the complementary event:

$$P(X \geq 2) = 1 - P(X<2) = 1 - P(X \leq 1)$$
$$= 1 - [P(X=1) + P(X=0)]$$
$$= 1 - [0.1176 + 0.3025]$$
$$= 0.5798$$

Note: Using the cumulative binomial distribution on a calculator, we enter the range of values as 2 to 6.



Example 5.6.3

Sophie has 10 pots labelled one to ten. Each pot, and its contents, is identical in every way. Sophie plants a seed in each pot such that each seed has a germinating probability of 0.8.

a What is the probability that all the seeds will germinate?

b What is the probability that only three seeds will not germinate?

c What is the probability that more than eight seeds do germinate?

d How many pots must Sophie use to be 99.99% sure that at least one seed germinates?

a Let X denote the number of seeds germinating. Therefore we have that $X \sim B(10, 0.8)$,

i.e. X is binomially distributed with parameters $n = 10$ and $p = 0.8$ (and $q = 1 - p = 0.20$).

$P(X = 10) = \binom{10}{10}(0.8)^{10}(0.2)^0 = 0.1074$.

If only three seeds will **not** germinate, then only seven seeds must germinate!

We want, $P(X = 7) = \binom{10}{7}(0.8)^7(0.2)^3 = 0.2013$.

Now, $P(X > 8) = P(X = 9) + P(X = 10)$

$= 0.2684 + 0.1074$

$= 0.3758$

At least one flower means, $X \geq 1$, therefore we need to find a value of n such that $P(X \geq 1) \geq 0.9999$.

$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0}(0.8)^0(0.2)^n = 1 - (0.2)^n$

Solving for n we have: $1 - (0.2)^n \geq 0.9999 \Leftrightarrow (0.2)^n \leq 0.0001$

This inequality can be solved by trial and error, algebraically, or: $(0.2)^n \leq 0.0001$, Sophie would need at least 6 pots, i.e.

Expectation, Mode and Variance for the Binomial Distribution

If the random variable X, is such that $X \sim B(n,p)$, we have:

1. the **expected value of X** is $\mu = E(X) = np$.
2. the **mode of X** is that value of x which has the largest probability
3. the **variance of X** is $\sigma^2 = Var(X) = npq = np(1 - p)$

Notes:

1. Although we can use our earlier definitions of the expected value and the variance of a random variable, the formulae above are in a nice compact form and can only be used when dealing with the binomial distribution.
2. The standard deviation, $Sd(X)$, is still given by $\sigma = \sqrt{Var(X)} = \sqrt{npq}$.

Example 5.6.4

A fair die is rolled six times. If X denotes the number of fours obtained, find:

a $E(X)$ b the mode of X c Sd(X)

a In this case we have that $X \sim B\left(6, \frac{1}{6}\right)$, therefore $q = \frac{5}{6}$

$\mu = E(X) = 6 \times \frac{1}{6} = 1$.

b To find the mode, we need to know the probability of each outcome. We do this by constructing a table of values:

x	0	1	2	3	4	5	6
$P(X=x)$	$\frac{15625}{46656}$	$\frac{18750}{46656}$	$\frac{9375}{46656}$	$\frac{2500}{46656}$	$\frac{375}{46656}$	$\frac{30}{46656}$	$\frac{1}{46656}$

So that the mode of X is 1 (as it has the highest probability value). Notice in this case, the mode of X = expected value of X. Will this always be true?

c $\sigma = \sqrt{Var(X)} = \sqrt{npq} = \sqrt{6 \times \frac{1}{6} \times \frac{5}{6}} \approx 0.9129$.

Example 5.6.5

An urn contains 7 marbles of which 2 are blue. A marble is selected, its colour noted and then replaced in the urn. This process is carried out 50 times. Find;

a The mean number of blue marbles selected

b The standard deviation of the number of blue marbles selected.

Because we replace the marble before the next selection, each trial is identical and independent. Therefore, if we let X denote the number of blue marbles selected, we have that:

$p = \frac{2}{7}, n = 50$ and $q = \frac{5}{7}$.

a $E(X) = np = 50 \times \frac{2}{7} = 14.29$.

b $Var(X) = npq = 50 \times \frac{2}{7} \times \frac{5}{7} = \frac{500}{49} \therefore \sigma = \sqrt{\frac{500}{49}} \approx 3.19$.

Example 5.6.6

The random variable X is such that $E(X) = 8$

and $Var(X) = 4.8$. Find $p(X = 3)$.

This time we are given that $np = 8$ and $npq = np(1-p) = 4.8$.

Therefore, after substituting $\boldsymbol{np} = 8$ into $\boldsymbol{np}(1-p) = 4.8$, we have that

$8(1-p) = 4.8 \quad \therefore (1-p) = 0.6 \Rightarrow p = 0.4$

Substituting $p = 0.4$ back into $np = 8$, we have that $n = 20$.

Therefore, $P(X=3) = \binom{20}{3}(0.4)^3(0.6)^{17} = 0.0123$.

Exercise 5.6.1

1. At an election 40% of the voters favoured the Environment Party. Eight voters were interviewed at random. Find the probability that:

   a exactly 4 voters favoured the Environment Party.

   b a majority of those interviewed favoured the Environment Party.

   c at most 3 of the people interviewed favoured the Environment Party.

2. In the long run, Thomas wins 2 out of every 3 games. If Thomas plays 5 games, find the probability that he will win:

   a exactly 4 games. b at most 4 games.

   c no more than 2 games.

   d all 5 games.

3. A bag consists of 6 white cubes and 10 black cubes. Cubes are withdrawn one at a time with replacement. Find the probability that after 4 draws:

   a all the cubes are black.

   b there are at least 2 white cubes.

   c there are at least 2 white cubes given that there was at least one white cube.

4. An X-ray has a probability of 0.95 of showing a fracture in the leg. If 5 different X-rays are taken of a particular leg, find the probability that:

   a all five X-rays identify the fracture.

   b the fracture does not show up.

   c at least 3 X-rays show the fracture.

   d only one X-ray shows the fracture.

5. A biased die, in which the probability of a '2' turning up is 0.4, is rolled 8 times.

   Find the probability that:

   a a '2' turns up 3 times.

   b a '2' turns up on at least 4 occasions.

6. During an election campaign, 66% of a population of voters are in favour of a food quality control proposal. A sample of 7 voters was chosen at random from this population.

   Find the probability that:

   a there will be 4 voters that were in favour.

   b there will be at least 2 voters who were in favour.

7. During an election 35% of the people in a town favoured the fishing restrictions at Lake Watanaki. Eight people were randomly selected from the town. Find the probability that:

   a 3 people favoured fishing restrictions.

   b at most 3 of the 8 favoured fishing restrictions.

   c there was a majority in favour of fishing restrictions.

8. A bag containing 3 white balls and 5 black balls has 4 balls withdrawn one at a time, in such a way that the first ball is replaced before the next one is drawn. Find the probability of:

   a selecting 3 white balls.

   b selecting at most 2 white balls.

   c selecting a white ball, two black balls and a white ball in that order.

   d selecting two white balls and two black balls.

9. A tennis player finds that he wins 3 out of 7 games he plays. If he plays 7 games straight, find the probability that he will win:

a exactly 3 games.

b at most 3 games.

c all 7 games.

d no more than 5 games.

e After playing 30 games, how many of these would he expect to win?

10. A true–false test consists of 8 questions. A student will sit for the test, but will only be able to guess at each of the answers. Find the probability that the student answers:

a all 8 questions correctly.

b 4 questions correctly.

c at most 4 of the questions correctly.

The following week, the same student will sit another true–false test, this time there will be 12 questions on the test, of which he knows the answer to 4.

e What are the chances of passing this test (assuming that 50% is a pass)?

11. The births of males and females are assumed to be equally likely. Find the probability that in a family of 6 children:

a there are exactly 3 girls.

b there are no girls.

c the girls are in the majority.

d How many girls would you expect to see in a family of 6 children?

12. During any one production cycle it is found that 12% of items produced by a manufacturer are defective. A sample of 10 items is selected at random and inspected. Find the probability that:

a no defectives will be found.

b at least two defectives will be found.

c A batch of 1000 such items are now inspected.

d How many of these items would you expect to be defective?

13. Ten per cent of washers produced by a machine are considered to be either oversized or undersized. A sample of 8 washers is randomly selected for inspection.

a What is the probability that there are 3 defective washers?

b What is the probability that there is at least one defective washer?

14. Over a long period of time, an archer finds that she is successful on 90% of her attempts. In the final round of a competition she has 8 attempts at a target.

a Find the probability that she is successful on all 8 attempts.

b Find the probability that she is successful on at least 6 attempts.

The prize that is awarded is directly proportional to the number of times she is successful, earning 100 fold, in dollars, the number of times she is successful.

c What can she expect her winnings to be after one round?

She draws with another competitor. However, as there can be only one winner, a second challenge is put into place – they must participate in another 3 rounds, with 5 attempts in each round.

d Find the probability that she manages 3 perfect rounds.

15. For each of the random variables:

a $X \sim B(7, 0.2)$

b $X \sim B(8, 0.38)$

Find: i the mean ii the mode

iii the standard deviation

iv $P(X \geq 6 | X > 4)$ v $P(X > 4 | X \leq 6)$

Extra questions

Poisson Distribution Function

The Poisson distribution was first brought to light by the eminent French mathematician Simeon Denis Poisson (1781–1840 - pictured) in his 1837 work *Recherches sur la probabilite de Judgement*, where he included a limit theorem for the binomial distribution. At the time, this was viewed as little more than a welcome approximation for the difficult computations required when using the binomial distribution. However, this was the embryo from which grew what is now one of the most important of all probability models.

However, a more general (and useful) use of the Poisson distribution (as opposed to only seeing it as an approximation to the binomial under certain conditions) is to define the distribution as

the distribution of the number of 'events' in a 'random process'.

The key in identifying a Poisson distribution, then, is to be able to identify the 'random process' and the 'event'. As we shall see, the event can be distributed over time, or distance, or length, or area, or volume, or …

Examples of 'random processes' and their corresponding 'events' are:

Random process	Event
Telephone calls in a fixed time interval.	Number of wrong calls in an hour. (Time dependent)
Accidents in a factory.	Number of accidents in a day. (Time dependent)
Flaws in a glass panel.	Number of flaws per square centimetre (Area dependent)
Flaws in a string.	Number of flaws per 5 metres. (Length dependent)
Bacteria in milk.	Number of bacteria per 2 litres. (Volume dependent)

The above examples serve to highlight the properties associated with the Poisson distribution. These can be best summarized as:

1 An event is as likely to occur in one given interval as it is in another (equally likely).

2. The occurrence of an event at a 'point' – be it a time interval, an area, etc. – is independent of when (or where) other events have occurred.

3. Events occur uniformly, i.e. the expected number of events in a given time interval, or area, or, … is proportional to the size of the time interval, or area, or, …

Note how similar these conditions are to those of the binomial distribution. However, one main difference between the two distributions is that there is, at least theoretically, no upper limit to the number of times an event may occur!

With this in mind, we now provide a statement for the Poisson distribution, incorporating the distribution function.

If $X(t)$ is the number of events in a time interval of length t, corresponding to a random process, with rate λ per unit time, then, we say that $X(t) \sim Pn(\lambda t)$ – read as *the random variable X has a Poisson distribution with parameter λt.*

Setting $\mu = \lambda t$, we define the Poisson probability distribution as:

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$$

Note that the rate λ can be specified as the number of events per unit time, or per unit area, or per unit of volume, or unit of length, etc.

The best way to see how this works is through the following examples.

Example 5.6.7

Cars have been observed to pass a given point on a back road at a rate of 0.5 cars per hour. Find the probability that no cars pass this point in a two-hour period.

The description of the situation fits the conditions under which a Poisson distribution can be assumed. From the information given we have that $\lambda = 0.5$.

Next we define the random variable X as the number of cars that pass the given point in a two-hour period.

This means that our parameter $\mu = \lambda \times 2 = 0.5 \times 2 = 1$ so that the probability function for X is given by:

$$P(X = x) = \frac{e^{-1}1^x}{x!}, x = 0, 1, 2, \ldots$$

$$\therefore P(X = x) = \frac{e^{-1}}{x!}, x = 0, 1, 2, \ldots$$

And so, $P(X=0) = \dfrac{e^{-1}}{0!} = 0.3679$.

Example 5.6.8

Faults occur on a piece of string at an average rate of one every three metres. Bobbins, each containing 5 metres of this string, are to be used. What is the probability that a randomly selected bobbin will contain:

a two faults. b at least two faults.

a The description of the situation fits the conditions under which a Poisson distribution can be assumed. From the information given we have that $\lambda = {}^1/_3$ (i.e. one in three metres).

Next we define the random variable X as the number of faults in a string 5 metres long. That is, number of faults per bobbin.

This means that our parameter $\mu = \lambda \times 5 = {}^1/_3 \times 5 = {}^5/_3$ so that the probability function for X is given by:

$$P(X=x) = \frac{e^{-5/3}\left(\frac{5}{3}\right)^x}{x!}, x = 0, 1, 2, \ldots$$

$$P(X=x) = \frac{e^{-5/3}}{x!}\left(\frac{5}{3}\right)^x, x = 0, 1, 2, \ldots$$

$$P(X=2) = \frac{e^{-5/3}}{2!}\left(\frac{5}{3}\right)^2 = 0.2623$$

The Poisson PDF is found in a similar way to the Binomial PDF on TI calculators.

$\text{poissPdf}\left(\frac{5}{3}, 2\right)$ 0.262327

b $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X=0) + P(X=1)]$

Now, $P(X=0) = \dfrac{e^{-5/3}}{0!}\left(\dfrac{5}{3}\right)^0 = e^{-5/3}$ and

$$P(X=1) = \frac{e^{-5/3}}{1!}\left(\frac{5}{3}\right)^1 = \frac{5}{3}e^{-5/3}$$

$$\begin{aligned}\therefore P(X \geq 2) &= 1 - \left[e^{-5/3} + \frac{5}{3}e^{-5/3}\right] \\ &= 1 - \left[1 + \frac{5}{3}\right]e^{-5/3} \\ &= 1 - \frac{8}{3}e^{-5/3} \\ &= 0.4963\end{aligned}$$

Step 1: Identify that scenario which fits the requirements of a Poisson distribution.

Step 2: Determine the 'base' rate, λ.

Step 3: Define the random variable.

Step 4: Determine the parameter, μ, that corresponds to the random variable in Step 3.

Example 5.6.9

A radioactive source emits particles at an average rate of one every 12 seconds. Find the probability that at most 5 particles are emitted in one minute.

The description of the situation fits the conditions under which a Poisson distribution can be assumed. From the information given we have that $\lambda = {}^1/_{12}$ (i.e. 1 in 12 seconds).

Next we define the random variable X as the number of particles emitted in 1 minute (or 60 seconds).

This means that our parameter $\mu = \lambda \times 60 = {}^1/_{12} \times 60 = 5$ so that the probability function for X is given by:

$$P(X=x) = \frac{e^{-5}5^x}{x!}, x = 0, 1, 2, \ldots$$

Therefore,

$$\begin{aligned}P(X \leq 5) &= P(X=0) + P(X=1) + \ldots + P(X=5) \\ &= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \ldots + \frac{e^{-5}5^5}{5!} \\ &= \left[1 + 5 + \frac{5^2}{2} + \frac{5^3}{6} + \frac{5^4}{24} + \frac{5^5}{120}\right]e^{-5} \\ &= \frac{1097}{12}e^{-5} \\ &= 0.6159 \quad (= 0.6160)\end{aligned}$$

Mean and Variance of the Poisson Distribution

With a little algebra, it can be shown that, for a random variable X having a Poisson distribution with parameter μ, i.e. if $X \sim Pn(\mu)$, then:

1. $E(X) = \mu$

2. $Var(X) = \mu$

Example 5.6.10

A typist finds that, on average, he makes two mistakes every three pages. Assuming that the number of errors per page follows a Poisson distribution, what are the chances that there will be 2 mistakes in the next page they type?

In this case we are given that the average is $^2/_3$. Then, if we let the random variable N denote the number of errors per page we have that:

$E(N) = \frac{2}{3}$. i.e. $\mu = \frac{2}{3}$.

So, $P(N = 2) = \frac{e^{-2/3}}{2!}\left(\frac{2}{3}\right)^2 = 0.1141$.

Example 5.6.11

If the random variable X is defined as $X \sim Pn(1.5)$, find $P(X > \mu + \sigma)$.

As $X \sim Pn(1.5)$ then we have that $\mu = 1.5$ and

$$\sigma^2 = 1.5 \Rightarrow \sigma = \sqrt{1.5} = 1.2247.$$

$$\begin{aligned}\therefore P(X > \mu + \sigma) &= P(X > 1.5 + 1.2247)\\ &= P(X > 2.7247)\\ &= P(X \geq 3)\\ &= 1 - P(X \leq 2)\\ &= 0.1912\end{aligned}$$

Example 5.6.12

The following frequency distribution shows the number of cars that drive over a bridge in a country area over a period of 100 days. Verify that this follows approximately a Poisson distribution.

Number of cars passing over bridge	0	1	2	3	4
Number of days observed	58	29	10	2	1

Let the random variable X denote the number of cars that pass over the bridge per day.

We first determine the average number of cars that pass over the bridge over the 100 days.

$$\begin{aligned}\bar{x} &= \frac{1}{100}(0 \times 58 + 1 \times 29 + 2 \times 10 + 3 \times 2 + 4 \times 1)\\ &= \frac{59}{100} = 0.59\end{aligned}$$

That is, on average there are 0.59 cars that pass over the bridge per day.

If this reflects a Poisson distribution, then we can use this mean as an estimate for the parameter μ in the distribution $X \sim Pn(\mu)$.

With $\mu = 0.59$ we have:

$$P(X = x) = \frac{e^{-0.59}(0.59)^x}{x!}, x = 0, 1, 2, 3, \ldots.$$

We can now produce a corresponding frequency distribution:

x	0	1	2	3	4
$P(X=x)$	0.5543	0.3271	0.0965	0.0190	0.0028

Multiplying these proportions by 100 we deduce the frequency distribution

x	0	1	2	3	4
$100P(X = x)$	55.43	32.71	9.65	1.90	0.28

The frequencies calculated using the Poisson probability function follow a pattern similar to the actual frequencies. We can, therefore, be reasonably sure that the Poisson distribution is an appropriate model for the number of cars that pass over this bridge.

Example 5.6.13

The number of flaws in metal sheets 100 cm by 150 cm is known to follow a Poisson distribution. On inspecting a large number of these metal sheets it is found that 20% of these sheets contain at least one flaw.

a Find the average number of flaws per sheet.

b Find the probability of observing one flaw in a metal sheet selected at random.

a Let the random variable X denote the number of flaws per 100 cm by 150 cm metal sheet.

Then, we have that $X \sim Pn(\mu)$ where μ is to be determined.

Knowing that $P(X \geq 1) = 0.2$ we have, $1 - P(X = 0) = 0.2$

$$\therefore P(X = 0) = 0.8 \Leftrightarrow e^{-\mu} = 0.8$$

$$\Leftrightarrow \mu = 0.2231$$

i.e. average number of flaws per sheet is 0.2232.

b $\quad P(X = 1) = (0.8)(0.2231)^1 = 0.1785$.

Poisson as a limit of binomial

We started this section by mentioning that the Poisson distribution was first encountered as a useful approximation for evaluating probabilities governed by the binomial distribution. We now summarise this relationship:

As $n \to \infty, Bin(n, p) \sim Pn(\mu)$, where $\mu = np$.

Obviously the larger the value of n, the better the approximation. However, a rough guide when considering the use of this approximation is $n > 20$ and $p < 0.05$. (Remember, this is only a guide.)

Example 5.6.14

A manufacturer of electrical components finds that, in the long run, 2.5% of the components are defective in some way. The inspection procedure requires that batches of 50 components are tested. What is the probability that in a batch of 50 there will be at least 3 defectives?

Let the random variable denote the number of defectives in a batch of 50 components.

Then, $X \sim Bin(50, 0.025)$, i.e. $p = 0.025$ and $n = 50$.

The use of the Poisson approximation in this case would be appropriate as $n > 20$ and $p < 0.05$.

Now, $\mu = np = 50 \times 0.025 = 1.25$ and so we have that $X \sim Pn(1.25)$.

Therefore,

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[e^{-1.25} + \frac{1.25}{1}e^{-1.25} + \frac{(1.25)^2}{2}e^{-1.25}\right] \\ &= 1 - [1 + 1.25 + 0.78125]e^{-1.25} \\ &= 1 - 3.03125e^{-1.25} \\ &= 0.1315 \end{aligned}$$

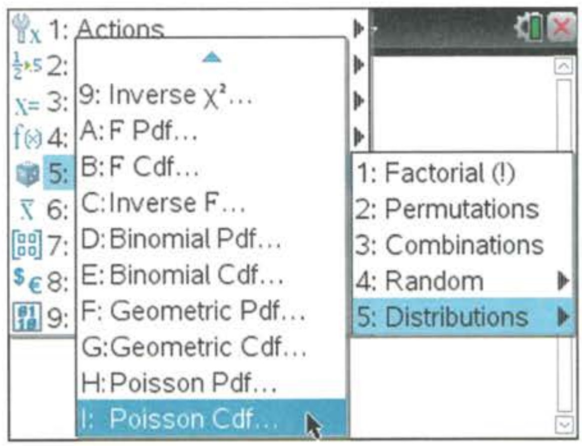
Calculators

When working with probability distributions in this and the next section, you will make heavy use of your calculator. We provide some more examples, but there is no substitute for practice!

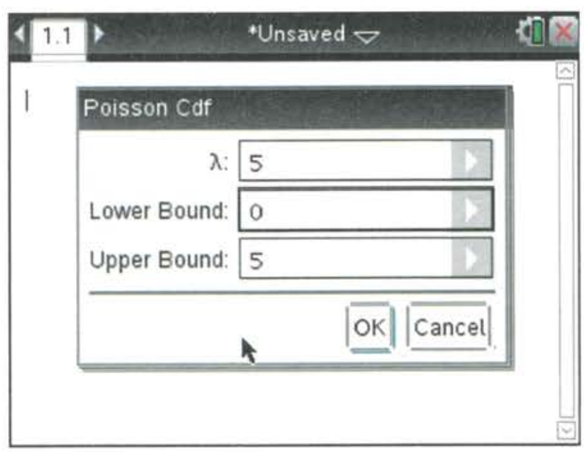
In example 5.6.9, we are looking at the Poisson cumulative density function with a mean of 5.

The question asks for $P(X \leq 5)$

Using TI:



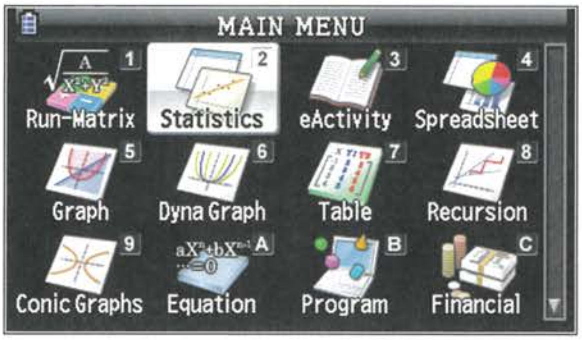
The parameters must now be entered:



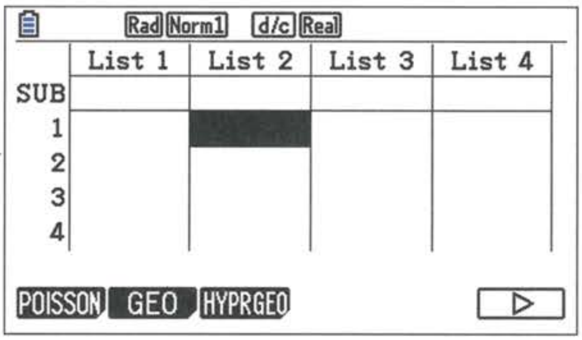
Pressing OK will confirm the answer we obtained earlier in the chapter.

If using Casio:

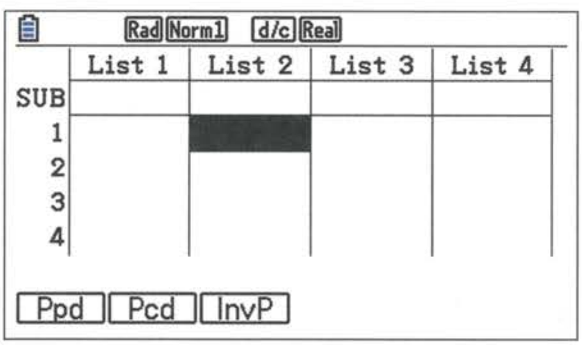
Select the Statistics module:



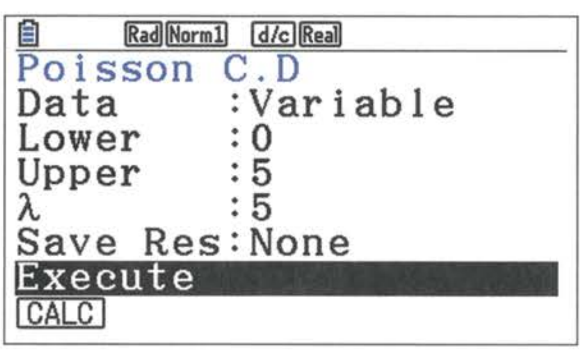
Select F5-DIST and scrolll right to find the Poisson Distribution:



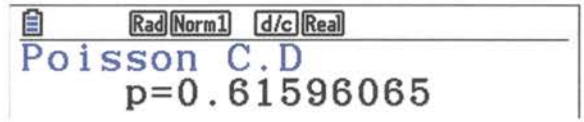
Use F1 to select POISSON.



Then F2 to select the Poisson cumulative and enter the parameters:



Finally, execute the calculation.



Exercise 5.6.2

1. If $X \sim Pn(2)$:

   a write down the probability distribution function for the random variable X.

   b Find: i $P(X=0)$. ii $P(X=2)$.

   iii $P(X>1)$. iv $P(X=2|X>1)$.

2. The flaws in a string occur at a rate of 2 every 5 metres. Find the probability that a string contains 3 flaws in:

   a 2 metres of string.

   b 10 metres of string.

3. Cars that stop at a particular petrol station during weekdays arrive at a rate of 10 cars every hour. Assuming a Poisson distribution, find the probability that:

   a there will be one car at the petrol station during any 15-minute interval.

   b there will be some cars at the petrol station during any 15-minute interval.

4. A switchboard receives an average of 100 calls per hour. Find the probability that:

   a the switchboard receives 2 calls during a one-minute time interval.

   b the switchboard receives at least 2 calls during a two-minute time interval.

5. On average a typist has to correct one word in every 800 words. Each page contains 200 words.

   a Find the probability that the typist makes more than one correction per page.

   b If more than one correction per page is required, the page needs to be retyped. What is the probability that more than two attempts are needed before a page is deemed satisfactory?

6. Cars have been observed to pass a given point on a country road at a rate of 5 cars per hour.

   a Find the probability that no cars pass this point in a 20-minute period.

   b Find the probability that at least 2 cars pass this point in a 30-minute period.

7. Bolts are produced in large quantities and it is expected that there is a 4% rejection rate due to some form of defect. A batch of 40 bolts is randomly selected for inspection.

   Using the Poisson distribution, find the probability that:

   a the batch contains at least one defective.

   b the batch contains no defectives.

   Ten such batches are randomly selected. If it is found that at least 2 batches have at least 4 defective, the total output is considered for the scrap heap to be recycled.

   c Find the probability that the total output is sent to the scrap heap.

8. Road accidents in a certain area occur at an average of 1 every 4 days. Find the probability that during a one-week period there will be:

   a two accidents.

   b at least two accidents.

9. Telephone calls arrive at a switchboard at a rate of 4 every minute. Find the probability that in a two-minute interval there will be fewer than 6 incoming calls.

10. Faults in glass sheets occur at a rate of 2.1 per square metre. If a square metre glass sheet contains at least 3 faults it is returned to the manufacturer.

   a Find the probability that a square metre sheet is returned to the manufacturer.

   b Six such glass sheets are inspected. What is the probability that at least half of them are returned to the manufacturer?

11. The number of faults in a glass sheet is known to have a Poisson distribution. It is found that 5% of sheets are rejected because they contain at least one flaw.

   a Find the probability that a sheet contains at least two flaws.

   b If the random variable X denotes the number of flaws per sheet, find $P(X > \mu + 2\sigma)$.

12. A shopkeeper finds that the number of orders for an electrical good averages 2 per week. At the start of the trading week, i.e. on a Monday, the shopkeeper has 5 such items in stock. Assuming that the orders follow a Poisson distribution, find the probability that during a given 5-day week:

   a there are three orders.

   b there are more orders than he can satisfy from his existing stock.

   If and when his stock level is down to two items during the week, he orders another four items:

   c what are the chances that he will order another four items?

13. Faults occur randomly along the length of a yarn of wool where the number of faults per bobbin holding a fixed length of yarn may be assumed to follow a Poisson distribution. A bobbin is rejected if it contains at least one fault. It is known that in the long run 33% of bobbins are rejected.

   Find the probability that a rejected bobbin contains only one fault.

   The production manager believes that by doubling the length of yarn on each bobbin there will be a smaller rejection rate. Assuming that the manufacturing process has not altered, is the production manager correct? Provide a quantitative argument.

Extra questions

Answers

Theory of Knowledge

At the time of writing, we have just seen several cases in which political polling organisations have got the results of referenda and elections wrong.

The first occurred in the UK General Election of May 2015. The result was predicted as being 'too close to call'. The actual result was a fairly comfortable win for the Conservative party led by David Cameron.

The second was the so called 'Brexit' referendum.

This was a poll to decide whether or not the UK would remain a member of the European Union. The polls predicted a win for the 'Remain' side but the 'Leave' side won reasonably comfortably.

The third was the US Election of November 2016. The polls predicted a narrow win by the Democrats led by Hillary Clinton. The result of the actual election was a win for the Republicans led by Donald Trump.

Has anything gone wrong with polling and the statistical methods that underpin it?

Note that you are being asked to consider the reliability of polling, not whether you like or dislike any of these three results!

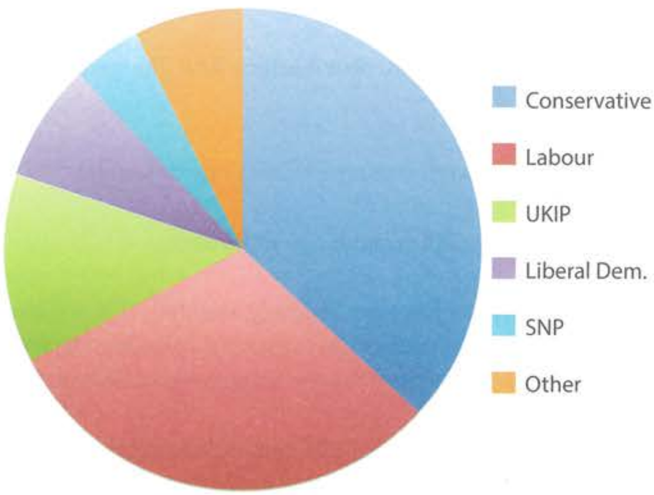
Some explanations are:

1. The 'Shy Tory'. Tory is another word for conservative in the UK. All three of these unexpected swings in voter sentiment were away from the 'official line' pushed by governments and much of the media. Are people reluctant to express views that are seen as 'unpopular' when they are asked by a polster? Are they more likely to express these maverick views in a secret ballot?

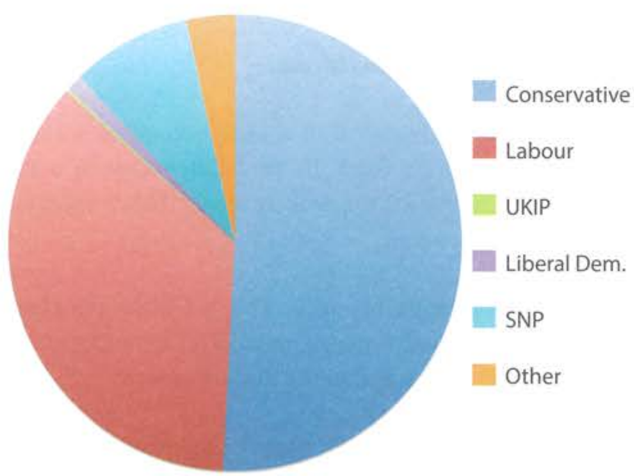
2. Confirmation bias. The detailed wording of questions can have a big effect on the answers. "Are you in favour of a military mission to rescue the impoverished people of Statsland?" and "Are you in favour of an invasion of Statsland?" are similar questions, but they will very likely produce quite different results. Did the writers of the polling questions load the questions (or the sampling method) to produce their favoured result?

3. 'Threshold issues'. The first of our examples is based on a UK General Election. These elections work using a 'first past the post' system. The winner in each constituency effectively gets all the votes as they are elected and the other candidates are not. This works in favour of the large parties and against the smaller ones when it comes to their actual representation in Parliament. These two diagrams show the distribution of votes and the distribution of 'seats' in the Parliament. Small changes in the vote can lead to big changes in the outcome. To what extent did this make life difficult for the pollsters?

Distribution of Votes (UK 2015)



Distribution of Seats in Parliament (UK 2015)



Can you think of other factors? Remember, you are thinking about the polling, not the outcomes.

5.7 Normal Distribution

Why the Normal Distribution?

The examples considered in previous sections mainly dealt with data that was discrete. Discrete data is generally counted and can be found exactly. Discrete data is often made up of whole numbers. For example, we may have counted the number of occupants in each of the cars passing a particular point over a period of two hours. In this case the data is made up of whole numbers. If we collect information on the European standard shoe sizes of a group of people, we will also be collecting discrete data even though some of the data will be fractional: shoe size nine and a half.

Alternatively, sometimes we collect data using measurement. For example, we may collect the birth weights of all the babies delivered at a maternity hospital over a year. Because weight is a continuous quantity (all weights are possible, not just whole numbers or certain fractions), the data collected is continuous. This remains the case even though we usually round continuous data to certain values. In the case of weight, we may round the data to the nearest tenth of a kilogram. In this case, if a baby's weight is given as 3.7 kg it means that the weight has been rounded to this figure and lies in the interval [3.65,3.75). If we are looking at data such as these weights it may seem as if the data is discrete even in cases when it is in fact continuous.

When dealing with continuous data, we use different methods. The most important distinction is that we can never give the number of babies that weigh *exactly* 3.7kg as there may be *none* of these. All that we can give is the number of babies born that have weights in the range [3.65,3.75).

One of the ways in which we can handle continuous data is to use the normal distribution. This distribution is only a model for real data. This means that its predictions are only approximate. The normal distribution generally works best in a situation in which the data clusters about a particular mean and varies from this as a result of random factors. The birth weights of babies cluster about a mean with variations from this mean resulting from a range of chance factors such as genetics, nutrition etc. The variation from the mean is measured by the standard deviation of the data. In examples such as this, the normal distribution is often a fairly good model. The basis of all normal distribution studies is the standard normal curve.

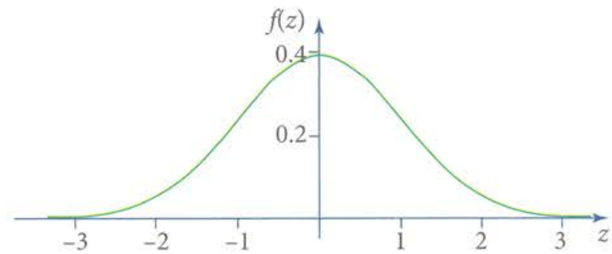
The Standard Normal Curve

The standard normal curve models data that has a mean of zero and a standard deviation of one. The equation of the standard normal curve is:

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{\left(-\frac{z^2}{2}\right)}, -\infty < z < \infty$$

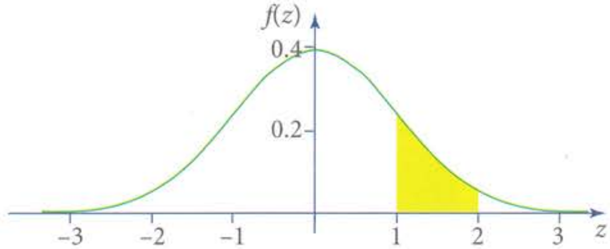
The equation of this distribution is complex and does not directly give us any information about the distribution. The shape of the curve does, however, indicate the general shape of the distribution.

The shape of this curve is often referred to as the 'bell-shaped curve'. On the next page we see how this function behaves.



As a result of the fact that the variable z is continuous, it is not the height of the curve but the areas underneath the curve that represent the proportions of the variable that lie between various values. The total area under the curve is 1 (even though the curve extends to infinity in both directions without actually reaching the axis).

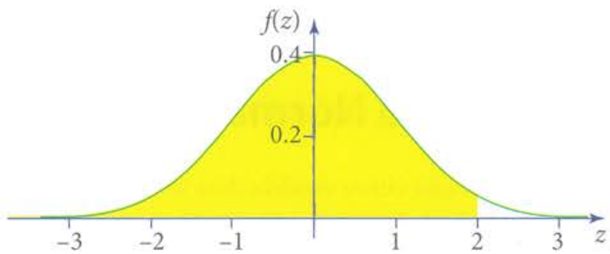
For example, the proportion of the standard normal data that lies between 1 and 2 is represented by the area shown.



Areas under curves are usually found using a method covered in Section 6.5. In the case of the normal curve, the complexity of the equation of the graph makes this impossible at least at this level. Instead, we rely on a graphics calculators.

Using a Calculator

The diagram shows the area that represents the proportion of values for which $z < 2$. This proportion can also be interpreted as the probability that a randomly chosen value of z will have a value of less than 2 or $p(Z < 2)$.



The area to the left is of infinite extent and yet the area is finite.

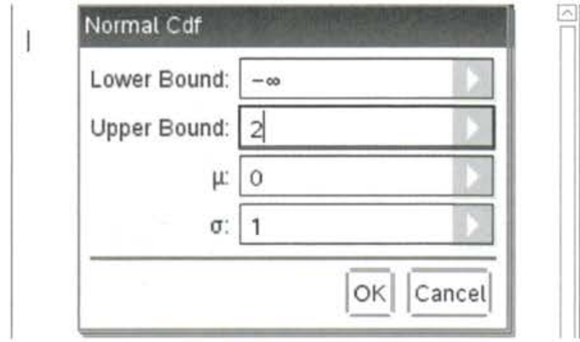
The area is found using a calculator (much as we find the trigonometric ratios etc.

The entries to solve this problem follow the same pattern used with the other probability distributions. If using at TI-NSpire:

MENU / 5. Probability / 5. Distributions / 2. Normal CDF

This allows the calculation of the sort of area (and hence probability) depicted in the diagram above.

The dialog box allows you to fill out the range of the variable $(-\infty, 2]$ and the mean (0) and standard deviation (1).



normCdf(-∞,2,0,1)	0.97725

Example 5.7.1

For the standard normal variable Z, find:

a $p(Z < 1)$ b $p(Z < 0.96)$

c $p(Z < 0.03)$.

All these examples can be solved by direct use of a calculator:

$p(Z < 1) = 0.8413$ $\quad p(Z < 0.96)\ (= 0.8315)$

$p(Z < 0.03)\ (= 0.5120)$

normCdf(-∞,1,0,1)	0.841345
normCdf(-∞,0.96,0,1)	0.831472
normCdf(-∞,0.03,0,1)	0.511967

Example 5.7.2

For the standard normal variable Z, find:

a $p(Z > 1.7)$ b $p(Z > -0.88)$

c $p(Z > -1.53)$.

normCdf(1.7,∞,0,1)	0.044565
normCdf(-0.88,∞,0,1)	0.81057
normCdf(-1.53,∞,0,1)	0.936992

Example 5.7.3

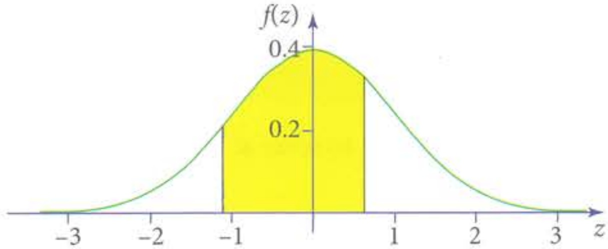
For the standard normal variable Z, find:

a $p(1.7 < Z < 2.5)$ b $(-1.12 < Z < 0.67)$

c $p(-2.45 < Z < -0.08)$.

It is as well when using technology to answer questions of this sort to have an estimate of the correct answer in mind.

For example, part b is represented by this area:



The whole are under the curve is 1, the shaded area looks to be a bit more than half this, so an answer a bit over 0.5 is to be expected. This is confirmed by the calculator.

normCdf(1.7,2.5,0,1)	0.038356
normCdf(-1.12,0.67,0,1)	0.617214
normCdf(-2.45,-0.08,0,1)	0.460976

Exercise 5.7.1

1. For the standard normal variable Z, find:

   a $p(Z < 0.5)$ b $p(Z < 1.84)$

   c $p(Z < 1.62)$ d $p(-2.7 < Z)$

   e $p(-1.97 < Z)$ f $p(Z < -2.55)$

2. For the standard normal variable Z, find:

   a $p(1.75 < Z < 2.65)$

   b $p(0.3 < Z < 2.5)$

   c $p(1.35 < Z < 1.94)$

   d $p(-1.92 < Z < -1.38)$

   e $p(2.23 < Z < 2.92)$

The Normal Distribution

Standardizing any normal distribution

Very few practical applications will have data whose mean is 0 and whose standard deviation is 1. The standard normal curve is, therefore, not directly usable in most cases. We overcome this difficulty by relating every problem to the standard normal curve.

As we have already seen, a general variable, X, is related to the standard normal variable, Z, using the relation:

$$Z = \frac{X - \mu}{\sigma}$$

where μ = the mean of the data and σ is the standard deviation. We use an example to illustrate this.

Example 5.7.4

A production line produces bags of sugar with a mean weight of 1.01 kg and a standard deviation of 0.02 kg:

a Find the proportion of the bags that weigh less than 1.03 kg.

b Find the proportion of the bags that weigh more than 1.02 kg.

c Find the percentage of the bags that weigh between 1.00 kg and 1.05 kg.

When approaching these problems, it is important to estimate answers. In the case of non-standard normal distributions, it is best to think graphically.

This involves relating the distribution you have been given to the standard normal curve. The former is centred on the mean and spreads three standard deviations either side of this.

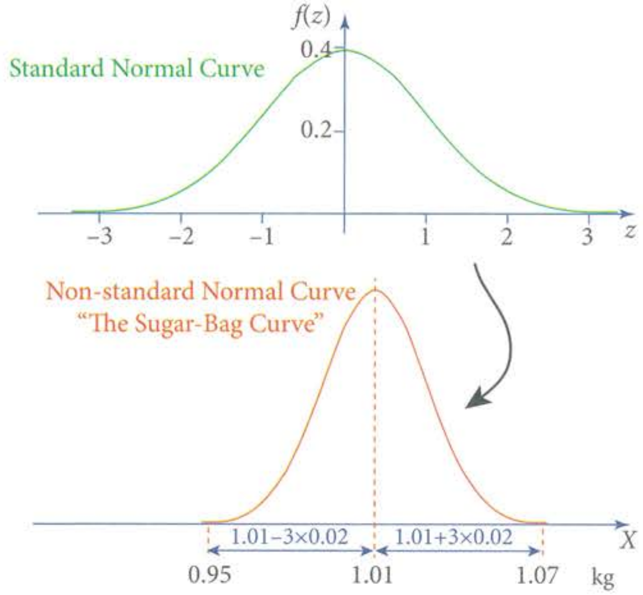
The weight variable, X, is related to the standard normal variable, Z, using the relation:

$$Z = \frac{X - \mu}{\sigma}$$

where μ = the mean of the data and σ is the standard deviation.

In this case: $Z = \dfrac{X - 1.01}{0.02}$.

The curve centres on 1.01 kg and spreads $3\times0.02 = 0.06$ ether side of this mean (i.e 0.95 to 1.07 kg).



The answers to the three parts are:

normCdf(-∞,1.03,1.01,0.02)	0.841345
normCdf(1.02,∞,1.01,0.02)	0.308538
normCdf(1,1.05,1.01,0.02)	0.668712

Rule of Thumb

For normally distributed variables, about two thirds of the values lie within one standard deviation of the mean and more than 99% of the values lie within two standard deviations of the mean.

Inverse Problems

There are occasions when we are told the proportion of the data that we are to consider and asked questions about the data conditions that are appropriate to these proportions.

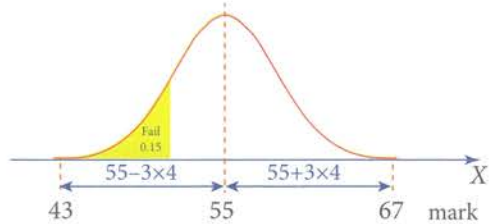
Example 5.7.5

The Board of Examiners have decided that 85% of all candidates sitting Mathematical Methods will obtain a pass grade in the examination. The actual examination marks are found to be normally distributed with a mean of 55 and a variance of 16. What is the lowest score a student can get on the exam to be awarded a pass grade?

This requires used of the inverse normal (option 3 on the TI) distribution.

Again, an estimate is useful. The mean is 55 and the standard deviation (the square root of the variance) is 4.

Thus almost all the students can be expected to score between 55–3×4=43 and 55+3×4=67.



On the basis of the graph, an answer of around 50 marks would seem reasonable:

invNorm(0.15,55,4)	50.8543

Therefore a student needs to score at least 51 marks to pass the exam.

Example 5.7.6

The lifetime of a particular make of television tube is normally distributed with a mean of 8 years, and a standard deviation of σ years. The chances that the tube will not last 5 years is 0.05. What is the value of the standard deviation?

Let X denote the life-time of the television tubes, so that $X \sim N(8, \sigma^2)$.

Given that $p(X < 5) = 0.05 \Rightarrow p\left(Z < \frac{5-8}{\sigma}\right) = 0.05$.

That is, we have that:

$$p\left(Z < -\frac{3}{\sigma}\right) = 0.05 \quad \Leftrightarrow -\frac{3}{\sigma} = -1.6449$$
$$\Leftrightarrow \sigma = 1.8238$$

And so the standard deviation is approximately one year and 10 months.

Example 5.7.7

The weight of a population of men is found to be normally distributed with mean 69.5 kg. Thirteen per cent of the men weigh at least 72.1 kg, find the standard deviation of their weight.

Let the random variable X denote the weight of the men, so that $X \sim N(69.5, \sigma^2)$.

We then have that $p(X \geq 72.1) = 0.13$ or $p(X \leq 72.1) = 0.87$.

$$\therefore p\left(Z \leq \frac{72.1 - 69.5}{\sigma}\right) = 0.87 \Leftrightarrow \frac{72.1 - 69.5}{\sigma} = 1.1264$$

$\therefore \sigma = 2.3083$

Exercise 5.7.2

1. If Z is a standard normal random variable, find:

   a $p(Z > 2)$ b $p(Z < 1.5)$

2. If Z is a standard normal random variable, find:

   a $p(Z > -2)$ b $p(Z < -1.5)$

3. If Z is a standard normal random variable, find:

   a $p(0 \leq Z \leq 1)$ b $p(1 \leq Z \leq 2)$

4. If Z is a standard normal random variable, find:

   a $p(-1 \leq Z \leq 1)$ b $p(-2 \leq Z \leq -1)$

5. If X is a normal random variable with mean $\mu = 8$ and variance $\sigma^2 = 4$, find:

   a $p(X \geq 6)$ b $p(5 < X \leq 8)$

6. If X is a normal random variable with mean $\mu = 100$ and variance $\sigma^2 = 25$, find:

   a $p(X \geq 106)$ b $p(105 < X \leq 108)$

7. If X is a normal random variable with mean $\mu = 60$ and standard deviation $\sigma = 5$, find:

   a $p(X \geq 65)$ b $p(55 < X \geq 65)$

8. Scores on a test are normally distributed with a mean of 68 and a standard deviation of 8. Find the probability that a student scored:

   a at least 75 on the test

   b at least 75 on the test given that the student scored at least 70

   c In a group of 50 students, how many students would you expect to score between 65 and 72 on the test.

9. If X is a normally distributed variable with a mean of 24 and standard deviation of 2, find:

   a $p(X > 28 | X \geq 26)$

   b $p(26 < X < 28 | X \geq 27)$

10. The heights of men are normally distributed with a mean of 174 cm and a standard deviation of 6 cm. Find the probability that a man selected at random:

    a is at least 170 cm tall

    b is no taller than 180 cm

    c is at least 178 cm given that he is at least 174 cm.

11. If X is a normal random variable with a mean of 8 and a standard deviation of 1, find the value of c, such that:

    a $p(X > c) = 0.90$ b $p(X \leq c) = 0.60$

12. If X is a normal random variable with a mean of 50 and a standard deviation of 5, find the value of c, such that:

    a $p(X \leq c) = 0.95$ b $p(X \geq c) = 0.95$

    c $p(-c \leq X \leq c) = 0.95$

13. The Board of Examiners has decided that 80% of all candidates sitting the Mathematical Methods Exam will obtain a pass grade. The actual examination marks are found to be normally distributed with a mean of 45 and a standard deviation of 7. What is the lowest score a student can get on the exam to be awarded a pass grade?

14. The weight of a population of women is found to be normally distributed with mean 62.5 kg. If 15% of the women weigh at least 72 kg, find the standard deviation of their weight.

15. The weights of a sample of a species of small fish are normally distributed with a mean of 37 grams and a standard deviation of 3.8 grams. Find the percentage of fish that weigh between 34.73 and 38.93 grams. Give your answer to the nearest whole number.

16. The weights of the bars of chocolate produced by a machine are normally distributed with a mean of 232 grams and a standard deviation of 3.6 grams. Find the proportion of the bars that could be expected to weigh less than 233.91 grams.

17. For a normal variable, X, $\mu = 196$ and $\sigma = 4.2$. Find:

a $p(X < 193.68)$ b $p(X > 196.44)$

18. The circumferences of a sample of drive belts produced by a machine are normally distributed with a mean of 292 cm and a standard deviation of 3.3 cm. Find the percentage of the belts that have diameters between 291.69 cm and 293.67 cm.

19. A normally distributed variable, X, has a mean of 52. $p(X < 51.15) = 0.0446$. Find the standard deviation of X.

20. The lengths of the drive rods produced by a small engineering company are normally distributed with a mean of 118 cm and a standard deviation of 0.3 cm. Rods that have a length of more than 118.37 cm are rejected. Find the percentage of the rods that are rejected. Give your answer to the nearest whole number.

21. After their manufacture, the engines produced for a make of lawn mower are filled with oil by a machine that delivers an average of 270 mL of oil with a standard deviation of 0.7 mL.

Assuming that the amounts of oil delivered are normally distributed, find the percentage of the engines that receive more than 271.12 mL of oil. Give your answer to the nearest whole number.

22. A sample of detergent boxes have a mean contents of 234 grams with a standard deviation of 4.6 grams. Find the percentage of the boxes that could be expected to contain between 232.22 and 233.87 grams. Give your answer to the nearest whole number.

23. A normally distributed variable, X, has a mean of 259. $p(X < 261.51) = 0.9184$. Find the standard deviation of X.

24. A normally distributed variable, X, has a standard deviation of 3.9. Also, 71.37% of the data are larger than 249.8. Find the mean of X.

25. The times taken by Maisie on her way to work are normally distributed with a mean of 26 minutes and a standard deviation of 2.3 minutes. Find the proportion of the days on which Maisie's trip takes longer than 28 minutes and 22 seconds.

26. In an experiment to determine the value of a physical constant, 100 measurements of the constant were made. The mean of these results was 138 and the standard deviation was 0.1. What is the probability that a final measurement of the constant will lie in the range 138.03 to 139.05?

27. In an experiment to determine the times that production workers take to assemble an electronic testing unit, the times had a mean of 322 minutes and a standard deviation of 2.6 minutes. Find the proportion of units that will take longer than 324 minutes to assemble. Answer to two significant figures.

28. A normally distributed variable, X, has a standard deviation of 2.6. $p(X < 322.68) = 0.6032$. Find the mean of X.

29. The errors in an experiment to determine the temperature at which a chemical catalyst is at its most effective, were normally distributed with a mean of 274°C and a standard deviation of 1.2°C. If the experiment is repeated what is the probability that the result will be between 275°C and 276°C?

30. The weights of ball bearings produced by an engineering process have a mean of 215 g with a standard deviation of 0.1 g. Any bearing with a weight of 215.32 g or more is rejected. The bearings are shipped in crates of 10 000. Find the number of bearings that may be expected to be rejected per crate.

31. If $X \sim N(\mu, 12.96)$ and $p(85.30 < X) = 0.6816$, find μ to the nearest integer.

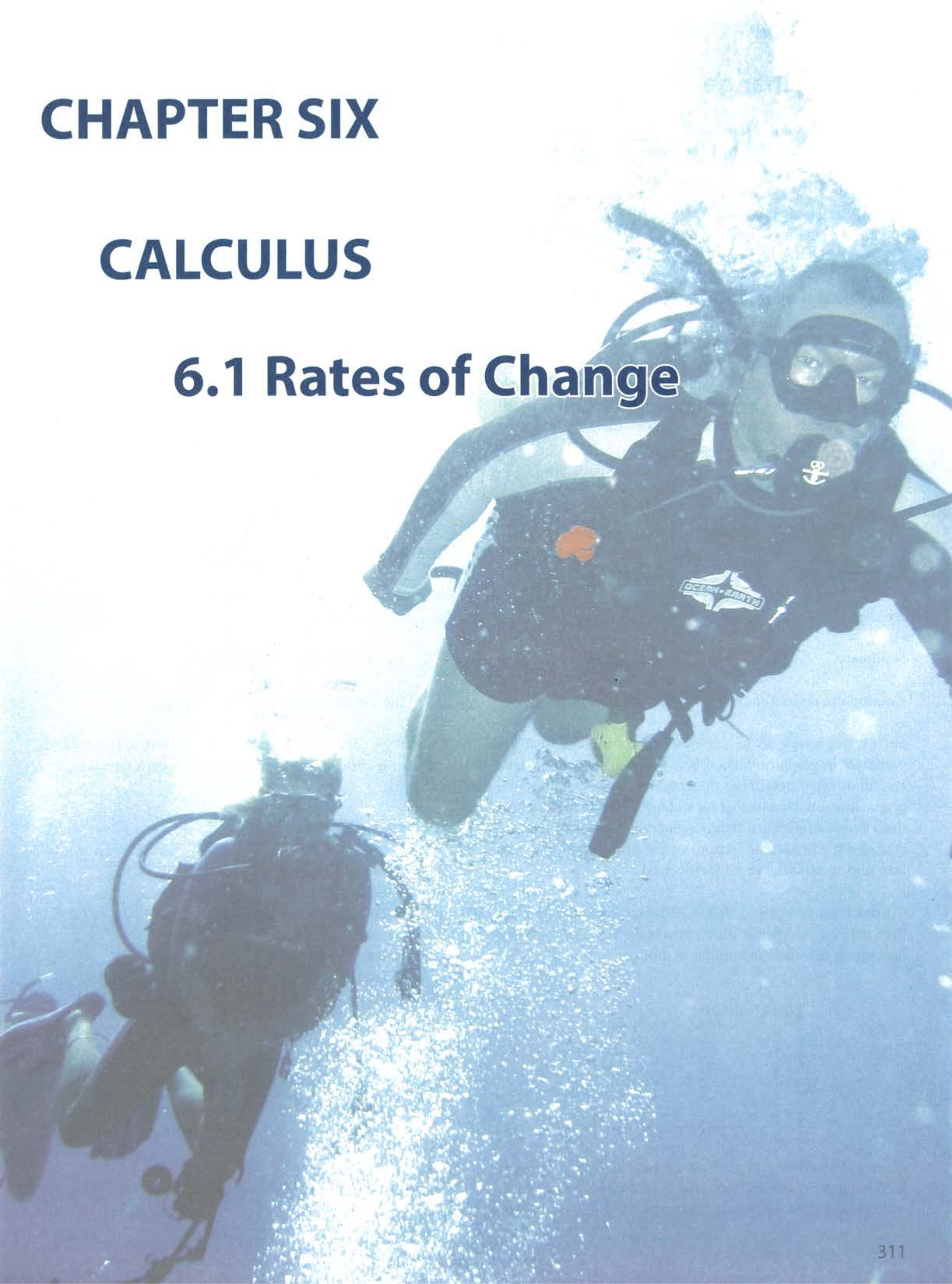
Extra questions

Answers

CHAPTER SIX

CALCULUS

6.1 Rates of Change



Rates of Change

The divers in our cover picture are breathing compressed air. As a result, their bodies are taking up nitrogen.

The rate at which the nitrogen enters and leaves their tissues determines whether or not they are at risk from 'the bends'. In this and many other applications, it is the rate at which things happen that matters. This is the subject of this Chapter.

Functional dependence

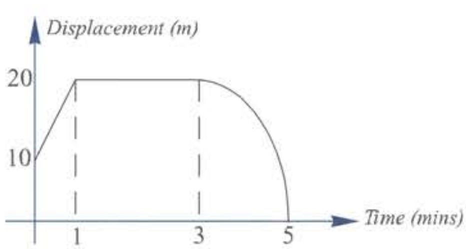
The notion of functional dependence of a function $f(x)$ on the variable x has been dealt with in Chapter 2. However, apart from this algebraic representation, sometimes it is desirable to create a graphical representation using a qualitative rather than quantitative description. In doing so, there are a number of key words that are often used.

Words to be kept in mind are:

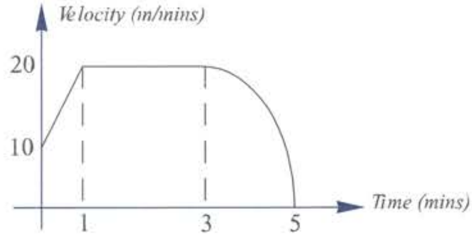
Rate of change (slow, fast, zero)	Increasing, decreasing
Positive, negative	Maximum, minimum
Average	Instantaneous
Stationary	Initial, final
Continuous, discontinuous	Range, domain

Such terms enable us to describe many situations that are presented in graphical form. There is one crucial point to be careful of when describing the graphical representation of a given situation. Graphs that look identical could very well be describing completely different scenarios. Not only must you consider the behaviour (shape) of the graph itself, but also take into account the variables involved.

Consider the two graphs below. Although identical in form, they tell two completely different stories. We describe what happens in the first five minutes of motion:



An object is moving in such a way that its displacement is increasing at a constant rate, that is, the object maintains a constant velocity (or zero acceleration) for the first minute. During the next two minutes the object remains stationary, that is, it maintains its displacement of 20 metres (meaning that it doesn't move any further from its starting position). Finally the particle returns to the origin.



An object is moving at 10 m/min and keeps increasing its velocity at a constant rate until it reaches a velocity of 20 m/min, that is, it maintains a constant acceleration for the first minute. During the next two minutes the object is moving at a constant velocity of 20 m/mins (meaning that it is moving further away from its starting position). Finally, the particle slows to rest, far from the origin.

Although the shape of both graphs is identical, two completely different situations have been described!

Quantitative aspects of change

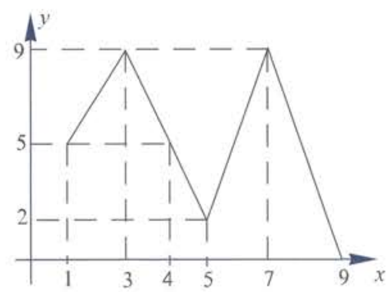
When dealing with the issue of rates of change, we need to consider two types of rates:

1. the average rate of change and
2. the instantaneous rate of change.

We start by considering the first of these terms, the average, and then see how the second, the instantaneous rate, is related to the first.

Average rate of change

The average rate of change can be best described as an 'overview' of the effect that one variable (the independent variable) has on a second variable (the dependent variable). Consider the graph opposite.



We can describe the change in the y-values (relative to the change in the x-values) as follows:

For $x \in [1,3]$:

There is a **constant** increase from $y = 5$ to $y = 9$ as x increases from 1 to 3.

An increase of 2 units in x has produced an increase of 4 units in y.

We say that the **average rate** of change of y with respect to x is $\frac{4}{2} = 2$.

For $x \in [1,4]$:

This time, the overall change in y is 0. That is, although y increases from 5 to 9, it then decreases back to 5. So from its initial value of 5, because it is still at 5 as x increases from 1 to 4, the overall change in y is 0. This time the average rate of change is $\frac{0}{3} = 0$.

For: $x \in [1,5]$:

As x now increases from 1 to 5 we observe that there is an overall decrease in the value of y, i.e. there is an overall decrease of 3 units (y: $5 \to 9 \to 5 \to 2$).

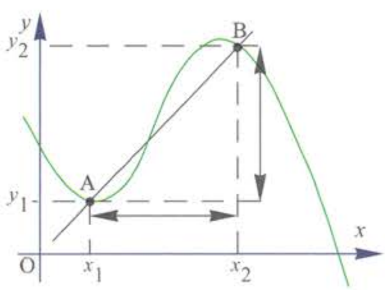
In this instance we say that the average rate of change is $-\frac{3}{4} = -0.75$.

Notice that we have included a negative sign to indicate that there was an overall decrease in the y-values (as x has increased by 4). Similarly for the rest of the graph. Note that we need not start at $x = 1$. We could just as easily have found the change in y for $x \in [3, 5]$. Here, the average rate of change is $-\frac{7}{2} = -3.5$.

The question then remains, is there a simple way to find these average rates of change and will it work for the case where we have non-linear sections? As we shall see in the next sections, the answer is 'yes'.

Determining the average rate of change

To find the average rate of change in y it is necessary to have an initial point and an end point, as x increases from x_1 to x_2.



At A $x = x_1, y = y_1$ and at B $x = x_2, y = y_2$.

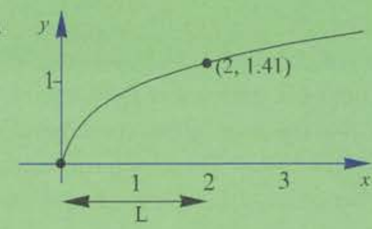
To obtain a numerical value, we find the gradient of the straight line joining these two points.
Average rate of change from A to B = gradient from A to B

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

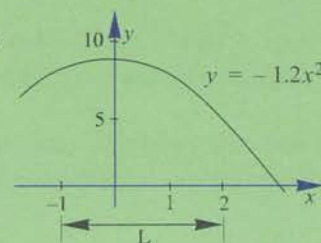
Example 6.1.1

For each of the graphs below, find the average rate of change of y with respect to x over the interval specified (i.e. over the domain L).

a



b



a For this case we have the 'starting point' at the origin (with coordinates (0, 0)) and the 'end' point with coordinates (2, 1.41).

This means that the average rate of change of y with respect to x, over the domain L is given by:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{1.41 - 0}{2 - 0} = 0.705.$$

b This time we will need to first determine the coordinates of the extreme points:

For $x = -1$, $y = -1.2 \times (-1)^2 + 9 = 7.8$ and for $x = 2$, $y = -1.2 \times (2)^2 + 9 = 4.2$.

Therefore, the average rate of change is equal to:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{4.2 - 7.8}{2 - (-1)} = -1.2.$$

It is not always necessary to have a graph in order to find the average rate of change. Often we are given information in the form of a table.

Example 6.1.2

The table below shows the number of bacteria, N, present in an enclosed environment. Find the average growth rate of the population size over the first 4 hours.

Time (hrs)	0	1	2	3	4	5	6	7	9
N	30	36	43	52	62	75	90	107	129

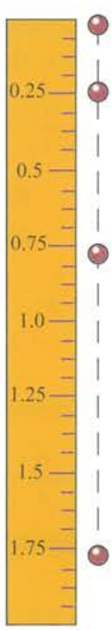
This time we need to consider the time interval $t = 0$ to $t = 4$. From the table we observe that the coordinates corresponding to these values are; (0,30) and (4,62). Therefore, the average rate of growth of the number of bacteria over the first 4 hours is equal to $\frac{62-30}{4-0} = \frac{32}{4} = 8$.

This means that during the first 4 hours, the number of bacteria was increasing (on average) at a rate of 8 every hour.

Notice that in the 1st hour, the average rate was $\frac{36-30}{1-0} = \frac{6}{1} = 6$ (< 8), whereas in the 4th hour the average rate of increase was $\frac{62-52}{4-3} = \frac{10}{1} = 10$ (> 8).

Velocity as a measure of the rate of change of displacement

Consider a marble that is allowed to free fall from a height of 2 metres (see diagram). As the marble is falling, photographs are taken of its fall at regular intervals of 0.25 second.



From its motion, we can tell that the rate at which the marble is falling is increasing (i.e. its velocity is increasing).

What is its average velocity over the first 0.6 second?

Reading from the diagram, we see that the marble has fallen a total distance of 1.75 (approximately), therefore, the average velocity v_{ave} of the marble, given by the rate at which its displacement increases (or decreases), is given by

$$v_{ave} = \frac{1.75-0}{0.6-0} \approx 2.92 \text{ m/sec}$$

Video discussion of average and instantaneous rates of change.

Example 6.1.3

The displacement, x m, of an object, t seconds after it is dropped from the roof of a building is given by $x = 4.9t^2$ metres.

a What is the object's displacement after 4 seconds?

b What is the average velocity of the object over the first 4 seconds of its motion?

a After 4 seconds of free fall, the object's displacement will be $4.9(4)^2 = 78.4$ m.

We obtained this result by substituting the value of $t = 4$ into the equation for the displacement $x = 4.9t^2$.

b The average velocity is given by the average rate of change of displacement, x m, with respect to the time t seconds.

Once we have the starting position and the end position we can determine the average velocity using:

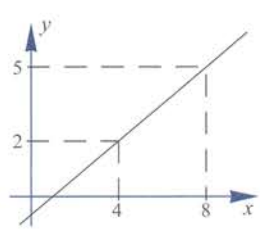
$$v_{ave} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{78.4-0}{4-0} = 19.6.$$

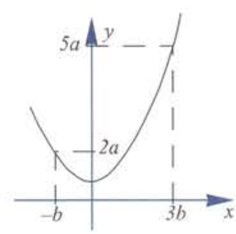
That is, the object's average velocity over the first 4 seconds is 19.6 m/s.

Exercise 6.1.1

1. For each of the following graphs determine the average rate of change over the specified domain.

a $x \in [4, 8]$ b $x \in [-b, 3b]$





2. For each of the following functions, find the average rate of change over the given domain.

a $x \mapsto x^2 + 2x - 1, x \in [0, 2]$

b $x \mapsto \sqrt{x+1}, x \in [3, 8]$

c $x \mapsto 10 - \frac{1}{\sqrt{x}}, x \in [2, 20]$

d $x \mapsto \frac{x}{x+1}, x \in [0.1, 1.1]$

e $x \mapsto \frac{1}{1+x^2} - 1, x \in [0, 100]$

f $x \mapsto x\sqrt{400-x}, x \in [300, 400]$

g $x \mapsto 2^x, x \in [0, 5]$

h $x \mapsto (x-1)(x+3), x \in [-3, 2]$

3. The displacement of an object, t seconds into its motion, is given by the equation, $s(t) = t^3 + 3t^2 + 2t, t \geq 0$.

   Find the average rate of change of displacement during:

   a the first second.

   b the first 4 seconds.

   c the interval when $t = 1$ to $t = 1 + h$.

4. The distance s metres that a particle has moved in t seconds is given by the function $s = 4t + 2t^2, t \geq 0$ Find the particle's average speed over the first 4 seconds.

5. The distance s metres that a particle has moved in t seconds is given by the function $s = 4t + 2t^2, t \geq 0$ Find the particle's average speed during the time interval from $t = 1$ to $t = 1 + h$.

6. The temperature T °C of food placed inside cold storage is modelled by the equation:

   $T = \frac{720}{t^2 + 2t + 25}$, where t is measured in hours.

   Find the average rate of change of the temperature, T°C, with respect to the time, t hours, during the first 2 hours that the food is placed in the cold storage.

7. The volume of water in a hemispherical bowl of radius r is given by:

   $V = \frac{1}{3}\pi h^2(3r - h)$, where h is the height of the water surface inside the bowl.

Extra questions

Instantaneous Rate of Change

Informal idea of limits

As already discussed, the average rate of change between two points on a curve is determined by finding the gradient of the straight line joining these two points. However, we often need to find the rate of change at a particular instant, and so the method used for finding the average rate of change is no longer appropriate. However, it does provide the foundation that leads to obtaining the instantaneous rate of change. We refine our definition of the average rate of change to incorporate the notion of the instantaneous rate of change. The basic argument revolves around the notion of magnifying near the point where we wish to find the instantaneous rate of change, that is, by repeatedly 'closing in' on a section of a curve. This will give the impression that over a very small section, the curve can be approximated by a straight line. Finding the gradient of that straight line will provide us with a very good approximation of the rate of change of the curve (over the small region under investigation). To obtain the exact rate of change at a particular point on the curve we will then need to use a limiting approach.

The process used to determine the rate of change at A is carried out as follows:

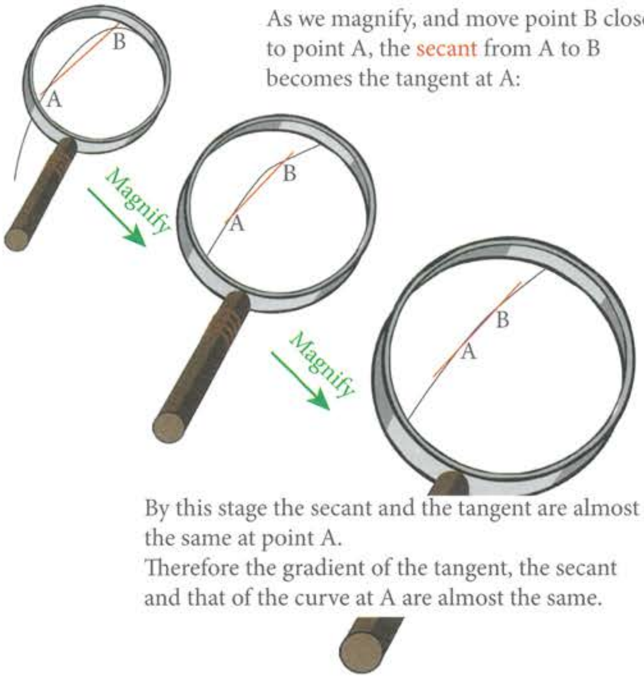
Start by drawing a secant from A to B, where B is chosen to be close to A. This will provide a reasonable first approximation for the rate of change at A. Then, to obtain a better approximation we move point B closer to point A.

Next, zoom–in towards point A, again. We move point B closer to point A, whereby a better measure for the rate of change at point A is now obtained. We then repeat step 2, i.e. move B closer to A and zoom in, move point B closer to A and zoom in, and so on.

Finally, the zooming–in process has reached the stage whereby the secant is now virtually lying on the curve at A. In fact the secant is now the tangent to the curve at the point A.

Using the process of repeatedly zooming in to converge on a particular region lies at the heart of the limiting process. Once we have understood the concepts behind the limiting process, we can move on to the more formal aspect of limits. However, apart from an informal treatment of limits, work on limits is beyond the scope of the core work in HL mathematics.

We now provide a 'visual' representation of steps 1 to 3 described above.



As we magnify, and move point B closer to point A, the secant from A to B becomes the tangent at A:

By this stage the secant and the tangent are almost the same at point A.
Therefore the gradient of the tangent, the secant and that of the curve at A are almost the same.

Example 6.1.4

An object moves along a straight line. Its position, x metres (from a fixed point O), at time t seconds is given by $x(t) = t-\frac{1}{4}t, t \geq 0$. Determine:

a its average velocity over the interval from $t = 1$ to $t = 2$

b its average velocity over the interval $t = 1$ to $t = 1.5$

c its average velocity over the interval $t = 1$ to $t = 1.1$

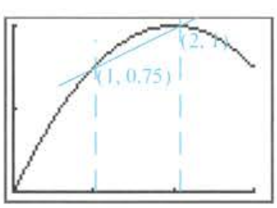
d its average velocity over the interval $t = 1$ to $t = 1 + h$, where h is small.

How can the last result help us determine the object's velocity at $t = 1$?

a The average velocity over the required second (from $t = 1$ to $t = 2$) is found by looking at the slope of the secant joining those two points on the graph of $x(t)$.

At $t = 2$, we have $x(2) = 2-\frac{1}{4}(2)^2 = 1$, and at $t = 1$, $x(1) = 1-\frac{1}{4}(1)^2 = \frac{3}{4}$

Therefore, we have that:



$$v_{ave} = \frac{x(2)-x(1)}{2-1} = \frac{1-0.75}{1} = 0.25$$

Therefore, the average velocity over the second is 0.75m/s.

b For $t = 1$ to $t = 1.5$ we have,

$$v_{ave} = \frac{x(1.5)-x(1)}{1.5-1} = \frac{(1.5-0.25\times 1.5^2)-0.75}{0.5} = 0.375$$

c Similarly, for $t = 1$ to $t = 1.1$, we have,

$$v_{ave} = \frac{x(1.1)-x(1)}{1.1-1} = 0.475$$

d We are now in a position to determine the average rate over the interval $t = 1$ to $t = 1 + h$.

The average velocity is given by $v_{ave} = \frac{x(1+h)-x(1)}{1+h-1}$

$$\begin{aligned}\text{Now, } x(1+h) &= (1+h)-0.25(1+h)^2\\ &= 1+h-0.25(1+2h+h^2)\\ &= 0.75+0.5h-0.25h^2\end{aligned}$$

Therefore,

$$v_{ave} = \frac{0.75+0.5h-0.25h^2-0.75}{1+h-1} = \frac{0.5h-0.25h^2}{h}$$

$$= \frac{h(0.5-0.25h)}{h}$$

$$= 0.5-0.25h, h \neq 0$$

Notice that for part b, (i.e. $t = 1$ to $t = 1.5$) $h = 0.5$, so that substituting $h = 0.5$ into this equation we have, $v_{ave} = 0.5-0.25(0.5) = 0.375$, providing the same result as before.

We can set up a table of values and from it determine what happens as we decrease the time difference.

We notice that, as h becomes very small, the average rate of change from $t = 1$ to $t = 1 + h$ becomes the instantaneous rate of change at t = 1! This is because we are zooming in onto the point where $t = 1$.

h	v_{ave}
0.1	0.475
0.01	0.4975
0.001	0.4999

This means that the rate of change at $t = 1$ (h '= 0') would therefore be 0.5 m/s. This means that the particle would have a velocity of 0.5 m/s after 1 second of motion.

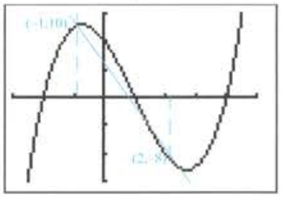
Example 6.1.5

For the graph with equation $f(x) = (x + 2)(x - 1)(x - 4)$,

a Find the average rate of change of f over the interval $[-1,2]$.

b Find the rate of change of f, where $x = 4$.

a We first find the coordinates of the end points for the interval $[-1,2]$:



$x=-1, y=f(-1)=...$

$=(-1+2)(-1-1)(-1-4)=10.$

$x=2,\ y=f(2)=(2+2)(2-1)(2-4)=-8.$

Therefore, the average rate of change in y with respect to x over the interval $[-1,2]$ is given by

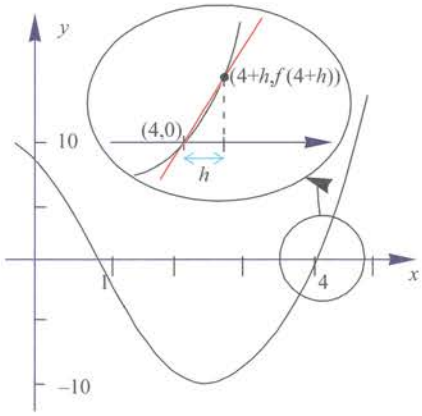
$$\frac{f(2)-f(-1)}{2-(-1)} = \frac{-8-10}{3} = -6$$

b To determine the rate of change at $x = 4$, we choose a second point close to $x = 4$. In this case, we use the point $x = 4 + h$, where h can be considered to be a very small number.

We will look at what happens to the gradient of the secant joining the points $(4, 0)$ and $(4 + h, f(4 + h))$ as h approaches zero.

The gradient of the secant is given by:

$$\frac{f(4+h)-f(4)}{(4+h)-4} = \frac{f(4+h)-f(4)}{h}$$



We now need to determine the value of $f(4+h)$ and $f(4)$. However, we already know that $f(4) = 0$.

We can now find values for $f(4+h)$ as h approaches zero.

For $h=0.1$, $f(4+0.1) = f(4.1) = (4.1+2)(4.1-1)(4.1-4)$
$= 6.1\times 3.1\times 0.1 = 1.891$

Therefore, $\dfrac{f(4+h)-f(4)}{h} = \dfrac{1.891-0}{0.1} = 18.91$, for $h = 0.1$.

We can continue in this same manner by making the value of h smaller still.

We do this by setting up a table of values:

h	$\frac{f(4+h)-f(4)}{h}$
0.01	18.09010000
0.001	18.00900100
0.0001	18.00090001

From the table, it appears that as h approaches zero, the gradient of the secant (which becomes the gradient of the tangent at (4,0)) approaches a value of 18.

Therefore, we have that the rate of change of f at (4,0) is 18.

More formally we write this result as,

$\lim\limits_{h\to 0}\dfrac{f(4+h)-f(4)}{h} = 18$ which is read as

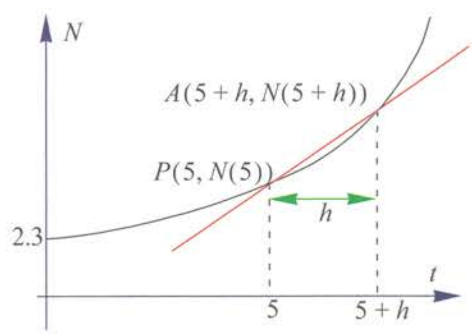
"The limit as h tends to zero of $\dfrac{f(4+h)-f(4)}{h}$ is equal to 18."

Example 6.1.6

The population of a city at the start of 2000 was 2.3 million, and its projected population, N million, is modelled by the equation $N(t)=2.3^{0.0142t}, t\ge 0$ and is measured in years since the beginning of 2000. Find the rate of growth of the population in this city at the start of 2005.

Finding the rate of growth of the population at the start of 2005 as opposed to finding the rate over a period of time means that we are finding the instantaneous rate of change. To do this, we proceed as in the previous example, i.e. we use a limiting approach.

Consider the two points, $P(5, N(5))$ (start of 2005) and $A(5+h, N(5+h))$ on the curve representing the population size:



The gradient of the secant passing through P and A is given by:

$$\frac{N(5+h)-N(5)}{(5+h)-5} = \frac{N(5+h)-N(5)}{h}$$

Now, $N(5) = 2.3e^{0.0142\times 5} = 2.3e^{0.071}$

and $N(5+h) = 2.3e^{0.0142(5+h)}$

Therefore, the gradient of the secant is given by

$$\frac{2.3e^{0.0142(5+h)} - 2.3e^{0.071}}{h} = \frac{2.3e^{0.071+0.0142h} - 2.3e^{0.071}}{h}$$
$$= \frac{2.3e^{0.071}(e^{0.0142h} - 1)}{h}$$

Again we set up a table of values:

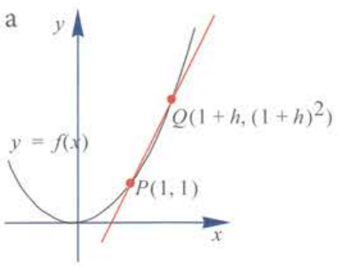
h	$\dfrac{2.3e^{0.071}(e^{0.0142h}-1)}{h}$
0.1	$\dfrac{2.3e^{0.071}(e^{0.0142\times0.1}-1)}{0.1} = 0.035088$
0.01	$\dfrac{2.3e^{0.071}(e^{0.0142\times0.01}-1)}{0.01} = 0.035066$
0.001	$\dfrac{2.3e^{0.071}(e^{0.0142\times0.001}-1)}{0.001} = 0.035063$
0.0001	$\dfrac{2.3e^{0.071}(e^{0.0142\times0.0001}-1)}{0.0001} = 0.035063$

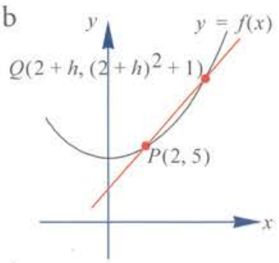
Using limit notation we have: $\lim\limits_{h\to0}\dfrac{N(5+h)-N(5)}{h} = 0.035063$

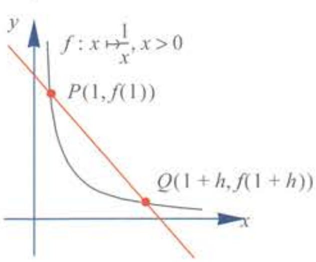
That is, the growth rate at the start of 2005 is 35 063 people per year.

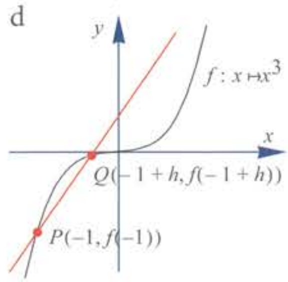
Exercise 6.1.2

1. For each of the graphs shown, find the gradient of the secant joining the points P and Q.

a


b


c


d


2. For each of the graphs in Question 1, use a limiting argument to deduce the instantaneous rate of change of the given function at the point P.

3. For each of the functions, f, given below, find the gradient of the secant joining the points $P(a, f(a))$ and $Q(a+h, f(a+h))$.

a $f(x) = 3 + x^2$ b $f(x) = 1 - x^2$

c $f(x) = (x+1)^2 - 2$ d $f(x) = x^3 + x$

e $f(x) = 2 - x^3$ f $f(x) = x^3 - x^2$

g $f(x) = \dfrac{2}{x}$ h $f(x) = \dfrac{1}{x-1}$

i $f(x) = \sqrt{x}$

4. An object moves along a straight line. Its position, x metres (from a fixed point O), at time t seconds is given by $x(t) = 2t^2 - 3t + 1$, $t \geq 0$.

a Sketch the graph of its displacement function.

b Determine :

i its average velocity over the interval from $t = 1$ to $t = 2$

ii its average velocity over the interval $t = 1$ to $t = 1.5$

iii its average velocity over the interval $t = 1$ to $t = 1.1$

c Show that its average velocity over the interval $t = 1$ to $t = 1 + h$, where h is small, is given by $1 + 2h$.

d How can the last result help us determine the object's velocity at $t = 1$?

e Show that its average velocity over any time interval of length h is given by $4t + 2h - 3$. Hence deduce the object's velocity at any time t during its motion

Extra questions

Answers (includes Ex 6.1.3)

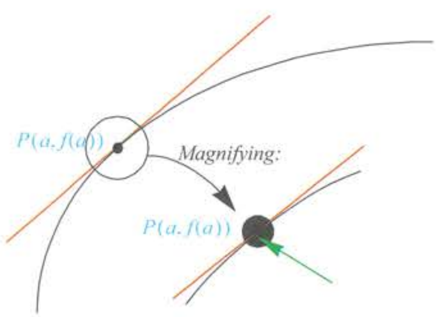
The Derivative and the Gradient Function

In the previous section we concentrated on determining the average rate of change for a function over some fixed interval. We then proceeded to find the instantaneous rate of change at a particular point (on the curve). We now consider the same process, with the exception that we will discuss the instantaneous rate at any point $P(x, f(x))$. The result will be an expression that will enable us to determine the instantaneous rate of change of the function at any point on the curve. Because the instantaneous rate of change at a point on a curve is a measure of the gradient of the curve at that point, our newly found result will be known as the gradient function (otherwise known as the derivative of the function).

For a continuous function, $y = f(x)$, we deduced that the instantaneous rate of change at the point $P(a,f(a))$ is given by:

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

where h is taken to be very small (in fact we say that h approaches or tends to zero).

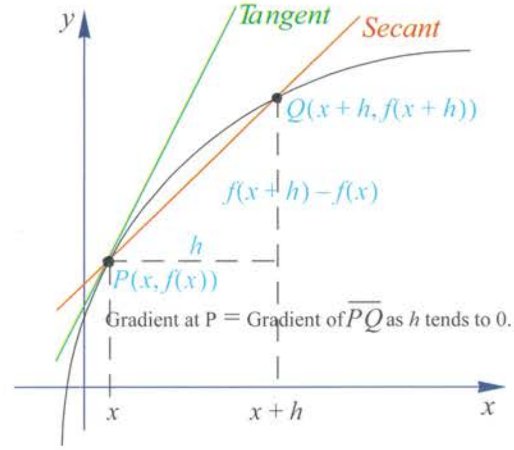


At the point P, the tangent and the line are one and the same.
Therefore, finding the gradient of the tangent at P is the same as finding the rate of change of the function at P.

So, to determine the rate at which a graph changes at a single point, we need to find the slope of the tangent line at that point.

This becomes obvious if we look back at our 'zooming in process—where the tangent line to the function at the point $P(a, f(a))$ is the line that best approximates the graph at that point.

Rather than considering a fixed point $P(a, f(a))$, we now consider any point $P(x, f(x))$ on the curve with equation $y = f(x)$:



The rate of change of the function f at $P(x, f(x))$ is therefore given by the gradient of the tangent to the curve at P.

If point Q comes as close as possible to the point P, so that h approaches zero, then, the gradient of the tangent at P is given by the gradient of the secant joining the points $P(x, f(x))$ and $Q(x+h, f(x+h))$ as $h \to 0$.

In mathematical notation we have:

$$\text{Rate of change at P} = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$$

Notation and language

We now introduce the term derivative of a function:

The **rate of change** of $f(x)$ at $P(x, f(x))$

$= $ **Gradient function** of $f(x)$ at $P(x, f(x))$

$= $ **The derivative** of $f(x)$

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

The derivative of a function $f(x)$ is denoted by $f'(x)$ and is read as "f dash of x".

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

That is, finding the derivative of a function using this approach is referred to as *finding the derivative of f from first principles*.

It is important to realise that in finding $f'(x)$ we have a new function - called the gradient function, because the expression $f'(x)$ will give the gradient anywhere on the curve of $f(x)$. If we want the gradient of the function $f(x)$ at $x = 5$, we first determine $f'(x)$ and then substitute the value of $x = 5$ into the equation of $f'(x)$.

Example 6.1.7

Using the first principles method, find the derivative (*or the gradient function*) of the function $f(x)=3x^2+4$.

Hence, find the gradient of the function at $x = 3$.

Using the first principles method means that we must make use of the expression

$$f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h} \quad -(1)$$

We start by first evaluating the expression $f(x+h)-f(x)$:

$$\begin{aligned} \text{That is: } f(x+h)-f(x) &= 3(x+h)^2+4-[3x^2+4] \\ &= 3(x^2+2xh+h^2)+4-3x^2-4 \\ &= 3x^2+6xh+3h^2-3x^2 \\ &= 6xh+3h^2 \end{aligned}$$

Substituting this result into (1):

$$\begin{aligned} \lim_{h \to 0} \frac{f(x+h)-f(x)}{h} &= \lim_{h \to 0} \frac{6xh+3h^2}{h} \\ &= \lim_{h \to 0} \frac{h(6x+3h)}{h} \\ &= \lim_{h \to 0} (6x+3h), h \neq 0 \\ &= 6x \end{aligned}$$

That is, we now have the gradient function $f'(x) = 6x$.

To determine the gradient of the function at $x = 3$, we need to substitute the value $x = 3$ into the gradient function. That is, $f'(3) = 6 \times 3 = 18$.

Exercise 6.1.3

1. Use a limiting process to find the gradients of these curves at the points indicated:

   a $x \mapsto x^3$ at $x = 1$

   b $v = 2t^2 - 1$ at $t = 2$

   c $f(x) = \frac{1}{x}$ at $x = 3$

   d $x \mapsto 2^x$ at $x = 1$

   e $f = t^2 - 2t + 3$ at $t = 0.5$

   f $t \mapsto \frac{t^2-1}{t}$ at $t = 4$

2. An object is dropped from a high building. The distance, *d* metres, that the object has fallen, *t* seconds after it is released, is given by the formula $d = 4.9t^2, 0 \le t \le 3$.

   a Find the distance fallen during the first second.

   b Find the distance fallen between $t = 1$ and $t = h + 1$ seconds.

   c Hence, find the speed of the object 1 second after it is released.

3. Find, from first principles, the gradient function, $f'(x)$, of the following.

   a $f:x \mapsto 4x^2$ b $f:x \mapsto 5x^2$

   c $f:x \mapsto 4x^3$ d $f:x \mapsto 5x^3$

   e $f:x \mapsto 4x^4$ f $f:x \mapsto 5x^4$

Can you see a pattern in your results?

4. Find, from first principles, the derivatives of the following functions.

   a $f(x) = 2x^2 - 5$ b $g(x) = 2 - x$

   c $g(x) = 2 - x + x^3$ d $f(x) = \frac{1}{x}$

   e $f(x) = \frac{2}{x+1}$ f $f(x) = \sqrt{x}$

5. A particle moving along a straight line has its position at time *t* seconds governed by the equation $x(t) = 2t - 0.5t^2, t \ge 0$, where $x(t)$ is its position in metres from the origin O.

   a Find the particle's velocity after it has been in motion for 1 second.

   b Find the particle's velocity at time $t = a, a > 0$.

6. A particle moving along a straight line has its position at time *t* seconds governed by the equation $x(t) = 4t^2 - t^3, t \ge 0$, where $x(t)$ is its position in metres from the origin O.

   a Sketch the displacement-time graph of the motion over the first five seconds

   b Find the particle's velocity at time:

   i $t = 1$ ii $t = 2$

   c Find the particle's velocity at any time $t, t \ge 0$.

   d When will the particle first come to rest?

6.2 Differentiation

Differential Calculus

Power rule for differentiation

Finding the derivative from first principles can be tedious. The previous two examples clearly show this. However, using the first principles approach produces the results shown in the table below:

Function $y = f(x)$	x^4	x^3	x^2	x^1	x^{-1}	x^{-2}
Derivative $\frac{dy}{dx} = f(x)$	$4x^3$	$3x^2$	$2x^1$	$1x^0$	$-1x^{-2}$	$-2x^{-3}$

Based on these results and following the general pattern, it is reasonable to assume the general result that if:

$$y = x^n, n \in Z, \text{ then } \frac{dy}{dx} = nx^{n-1}$$

In fact this rule is true for any exponent $n \in \mathbb{R}$, i.e. for any real number n.

For example, if we look at the square root function, then we have that $y = \sqrt{x} = x^{1/2}$. So in this case we have that $n = ½$.

Then, using our rule we have:

$$y = \sqrt{x} = x^{\frac{1}{2}} \Rightarrow \frac{dy}{dx} = \frac{1}{2}x^{\frac{1}{2}-1} = \tfrac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}.$$

This result is known as the power rule for differentiation.

Notice that for the case $n = 0$, then $y = x^0$ and so we have that $\frac{dy}{dx} = 0x^{0-1} = 0$.

Example 6.2.1

Use the power rule to differentiate the following functions.

a x^6 b $\frac{1}{\sqrt{x}}$ c $\sqrt[3]{x}$ d $\frac{1}{x^2}$

Before we differentiate these functions, each function must be rewritten in the form x^n before we can use the power rule.

a Let $f(x) = x^6 \Rightarrow f'(x) = 6x^{6-1} = 6x^5$

b Let $y = \frac{1}{\sqrt{x}}$.

Remember, we first need to rewrite it in the form x^n:

Function →	**Step 1** Rewrite: →	**Step 2** Use power rule:
$y = \frac{1}{\sqrt{x}}$	$y = x^{-1/2}$	$\frac{dy}{dx} = -\frac{1}{2}x^{-\frac{1}{2}-1}$

→ **Step 3** Simplify:

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

c Let $y = \sqrt[3]{x}$.

As in the previous example, we rewrite this function in the form x^n so that we can use the power rule:

Function →	**Step 1** Rewrite: →	**Step 2** Use power rule: →	**Step 3** Simplify:
$y = \sqrt[3]{x}$	$y = x^{\frac{1}{3}}$	$\frac{dy}{dx} = \frac{1}{3}x^{\frac{1}{3}-1}$	$\frac{dy}{dx} = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}}$

d Let $f(x) = \frac{1}{x^2}$ so that $f(x) = x^{-2}$

$\therefore f'(x) = -2x^{-2-1} = -2x^{-3}$, that is, $f'(x) = -\frac{2}{x^3}$.

Derivative of a sum or difference

This rule states that the derivative of a sum (or a difference) is equal to the sum (or the difference) of the derivatives.

That is, If $y = f(x) \pm g(x)$ then $\frac{dy}{dx} = f'(x) \pm g'(x)$

Example 6.2.2

Differentiate these functions.

a $y = 2x^3 + 5x - 9$

b $f(x) = \sqrt{x} - \frac{5}{x^3} + x$

c $f(x) = x^{1/3} + x^{5/4} - \sqrt{2x}$

a
$$y = 2x^3 + 5x - 9 \Rightarrow \frac{dy}{dx} = \frac{d}{dx}(2x^3 + 5x - 9) = \frac{d}{dx}(2x^3) + \frac{d}{dx}(5x) - \frac{d}{dx}(9) = 6x^2 + 5$$

Notice we have used a slightly different notation, namely that:

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

We can think of $\frac{d}{dx}$ as the differentiation operator, so that $\frac{d}{dx}(f(x))$ or $\frac{d}{dx}(y)$ is an operation of differentiation done on $f(x)$ or y respectively.

b
$$f(x) = \sqrt{x} - \frac{5}{x^3} + x \Rightarrow f'(x) = \frac{d}{dx}\left(\sqrt{x} - \frac{5}{x^3} + x\right) = \frac{d}{dx}(x^{1/2} - 5x^{-3} + x) = \frac{1}{2}x^{-1/2} - 5 \times -3x^{-3-1} + 1 = \frac{1}{2\sqrt{x}} + \frac{15}{x^4} + 1$$

c
$$\frac{d}{dx}(x^{1/3} + x^{5/4} - \sqrt{2x}) = \frac{d}{dx}(x^{1/3}) + \frac{d}{dx}(x^{5/4}) - \frac{d}{dx}(\sqrt{2x}) = \frac{1}{3}x^{-2/3} + \frac{5}{4}x^{1/4} - \sqrt{2} \times \frac{1}{2}x^{-1/2}$$

Exercise 6.2.1

1. Find the derivative of each of the following.

a	x^5	b	x^9	c	x^{25}
d	$9x^3$	e	$-4x^7$	f	$\frac{1}{4}x^8$
g	$x^2 + 8$	h	$5x^4 + 2x - 1$		
i	$-3x^5 + 6x^3 - x$	j	$20 - \frac{1}{3}x^4 + 10x$		
k	$3x^3 - 6x^2 + 8$	l	$3x - 1 + \frac{x^2}{5} + x^4$		

2. Find the derivative of each of the following.

a	$\frac{1}{x^3}$	b	$\sqrt{x^3}$	c	$\sqrt{x^5}$
d	$\sqrt[3]{x}$	e	$4\sqrt{x}$	f	$6\sqrt{x^3}$
g	$2\sqrt{x} - \frac{3}{x} + 12$	h	$x\sqrt{x} + \frac{1}{\sqrt{x}} + 2$		
i	$5\sqrt[3]{x^2} - 9x$	j	$5x - \frac{x}{\sqrt{x}} + \frac{4}{5x^2}$		
k	$8\sqrt{x} + 3x^{-5} + \frac{x}{2}$	l	$\frac{x}{\sqrt{x^3}} - \frac{2}{x}\sqrt{x^3} + \frac{1}{3}x^3$		

3. Find the derivative of each of the following.

a	$\sqrt{x}(x + 2)$	b	$(x + 1)(x^3 - 1)$
c	$x\left(x^2 + 1 - \frac{1}{x}\right), x \neq 0$	d	$\frac{2x - 1}{x}, x \neq 0$
e	$\frac{\sqrt{x} - 2}{\sqrt{x}}, x > 0$	f	$\frac{x^2 - x + \sqrt{x}}{2x}, x \neq 0$
g	$\frac{3x^2 - 7x^3}{x^2}, x \neq 0$	h	$\left(x - \frac{2}{x}\right)^2, x \neq 0$
i	$\left(x + \frac{1}{x^2}\right)^2, x \neq 0$	j	$\sqrt{3x} - \frac{1}{3\sqrt{x}}, x > 0$
k	$(x - \sqrt[5]{x})^2, x \geq 0$	l	$\left(\frac{1}{\sqrt{x}} - \sqrt{x}\right)^3, x > 0$

4. a Show that if $f(x) = x^2 - x$, then $f'(x) = 1 + \frac{2f(x)}{x}$.

b Show that if $f(x) = \sqrt{2x} - 2\sqrt{x}, x \geq 0$, then $\sqrt{2x}f'(x) = 1 - \sqrt{2}, x > 0$

c Show that if $y = ax^n$ where a is real and $n \in N$, then $\frac{dy}{dx} = \frac{ny}{x}, x \neq 0$

d Show that if $y = \frac{1}{\sqrt{x}}, x > 0$, then $\frac{dy}{dx} + \frac{y}{2x} = 0$.

Differentiating with variables other than *x* and *y*

Although it was convenient to establish the underlying theory of differentiation based on the use of the variables x and y, it must be pointed out that not all expressions are written in terms of x and y. In fact, many of the formulae that we use are written in terms of variables other than y and x, e.g. volume, V, of a sphere is given by $V = \frac{4}{3}\pi r^3$, where r is its radius. The displacement of a particle moving with constant acceleration is given by $s = ut + \frac{1}{2}at^2$. However, it is reassuring to know that the rules are the same regardless of the variables involved. Thus, if we have that y is a function of x, we can differentiate y *with respect to* (w.r.t.) x to find $\frac{dy}{dx}$.

On the other hand, if we have that y is a function of t, we would differentiate y w.r.t. t and write $\frac{dy}{dt}$. Similarly, if W was a function of θ, we would differentiate W w.r.t. θ and write $\frac{dW}{d\theta}$.

Example 6.2.3

Differentiate the following with respect to the appropriate variable.

a $V = \frac{4}{3}\pi r^3$ b $p = 3w^3 - 2w + 20$

c $s = 10t + 4t^2$

a As V is a function of r, we need to differentiate V with respect to r:

$$V = \frac{4}{3}\pi r^3 \Rightarrow \frac{dV}{dr} = \frac{4}{3}\pi(3r^2) = 4\pi r^2.$$

b This time p is a function of w, and so we would differentiate p with respect to w:

$$p = 3w^3 - 2w + 20 \Rightarrow \frac{dp}{dw} = 9w^2 - 2.$$

c In this expression we have that s is a function of t and so we differentiate s w.r.t t:

$$s = 10t + 4t^2 \Rightarrow \frac{ds}{dt} = 10 + 8t.$$

Exercise 6.2.2

1. Differentiate the following functions with respect to the appropriate variable.

a $s = 12t^4 - \sqrt{t}$ b $Q = \left(n + \frac{1}{n^2}\right)^2$

c $P = \sqrt{r}(r + \sqrt[3]{r} - 2)$ d $T = \frac{(\theta - \sqrt{\theta})^3}{\theta}$

e $A = 40L - L^3$ f $F = \frac{50}{v^2} - v$

g $V = 2l^3 + 5l$ h $A = 2\pi h + 4h^2$

Extra questions

Derivatives of Transcendental Functions

A transcendental function is a function that cannot be constructed in a finite number of steps from elementary functions and their inverses. Some examples of these functions are sin$x$, cos$x$, tan$x$, the exponential function, e^x, and the logarithmic function $\log_e x$ (or ln$x$).

Also in this section we look at derivatives of expressions that involve the **product** of two functions, the **quotient** of two functions and the **composite** of two functions. Each of these types of expressions will lead to some standard rules of differentiation.

Derivative of circular trigonometric functions

We begin by considering the trigonometric functions, i.e. the sine, cosine and tangent functions.

There are a number of approaches that can be taken to achieve our goal. In this instance we will use two different approaches to find the derivative of the function $x \mapsto \sin(x)$.

Letting $f(x) = \sin(x)$ and using the definition from first principles we have:

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

From the identities discussed in Section 3.3:

$$\frac{\sin(x+h)-\sin(x)}{h}=\frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h}$$

We are taking the limit as h becomes small.

$h \to 0, \cos(h) \to 1$ so:

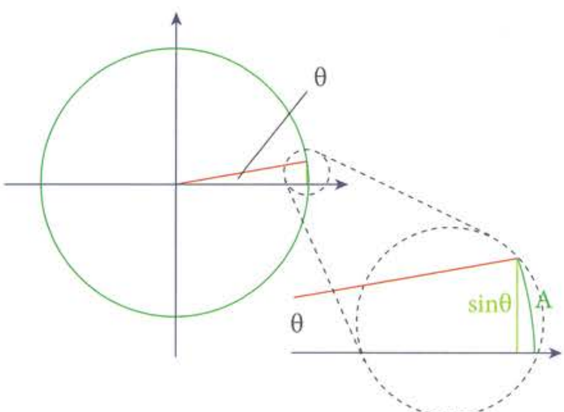
$$f'(x)=\lim_{h\to 0}\left(\frac{\sin(x)\times 1+\cos(x)\sin(h)-\sin(x)}{h}\right)$$
$$=\lim_{h\to 0}\left(\frac{\cos(x)\sin(h)}{h}\right)$$

The result of this limit is not immediately obvious.

It simplifes to: $f'(x)=\lim_{h\to 0}\left(\frac{\sin(h)}{h}\right)\times\cos(x)$

This depends on the result of the limit: $\lim_{h\to 0}\left(\frac{\sin(h)}{h}\right)$

The result of this is suggested (not proved) by the unit circle.



From the definition of angle, the arc length (green) is equal to the angle in radians, θ. The light green line is equal to sin θ. The smaller the angle becomes, the closer these to quantities become to one another.

This strongly suggests that: $\lim_{h\to 0}\left(\frac{\sin(h)}{h}\right)=1$ and hence:

$$f(x)=\sin(x)\Rightarrow f'(x)=\cos(x)$$

This is only true if the angle is measured in radians - a very good reason for using radians rather than degrees!

Similarly, using either approaches, we have:

$f(x) = \cos(x)$ then $f'(x) = -\sin(x)$

$f(x) = \tan(x)$ then $f'(x) = \sec^2(x)$

Derivative of the exponential function

Consider the exponential function $x \mapsto e^x$. Although we could apply the same graphical method which was used to determine the derivative of the sine function, this time we shall make use of the definition of the derivative, i.e.

$$\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\\
&= \lim_{h\to 0}\frac{e^{x+h}-e^x}{h}\\
&= \lim_{h\to 0}\frac{e^x e^h-e^x}{h} \quad \text{using } e^x e^h = e^{x+h}\\
&= \lim_{h\to 0}\frac{e^x(e^h-1)}{h}\\
&= e^x\lim_{h\to 0}\frac{(e^h-1)}{h} \quad \text{Use a numerical method to find that this limit = 1}\\
&= e^x\times 1\\
&= e^x
\end{aligned}$$

Notice that this function has a derivative that is the same as itself.

Derivative of the natural log function

Consider the function $x \mapsto \log_e x$. As in the previous case, we use the definition of the derivative to establish the gradient function of $x \mapsto \log_e x$.

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

$$= \lim_{h\to 0}\frac{\log_e\left(\frac{x+h}{x}\right)}{h} \quad \text{(Using the log laws)}$$

The next step is a little tricky, so we write it down first and then see how we arrive at the result.

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

To get to this step we proceed as follows:

$$\frac{\log_e\left(1+\frac{h}{x}\right)}{h} = \frac{1}{x}\cdot x\frac{\log_e\left(1+\frac{h}{x}\right)}{h} = \frac{1}{x}\cdot\left(\frac{x}{h}\right)\log_e\left(1+\frac{h}{x}\right)$$
$$= \frac{1}{x}\cdot\log_e\left(1+\frac{h}{x}\right)^{\frac{x}{h}}$$

Then, as the argument in the limit is h (i.e. it is independent of x) we have:

$$\lim_{h\to 0}\frac{1}{x}\cdot\log_e\left(1+\frac{h}{x}\right)^{\frac{x}{h}} = \frac{1}{x}\lim_{h\to 0}\log_e\left(1+\frac{h}{x}\right)^{\frac{x}{h}}$$

Then, as the log function is a continuous function, we have that the limit of the log is the same as the log of the limit so that

$$\lim_{h \to 0} \log_e\left(1 + \frac{h}{x}\right)^{\frac{x}{h}} = \log_e\left[\lim_{h \to 0}\left(1 + \frac{h}{x}\right)^{\frac{x}{h}}\right]$$

However, we also have that $\lim_{h \to 0}\left(1 + \frac{h}{x}\right)^{\frac{x}{h}} = e$ and so we end up with the result that

$$\frac{1}{x}\lim_{h \to 0} \log_e\left(1 + \frac{h}{x}\right)^{\frac{x}{h}} = \frac{1}{x}\log_e e = \frac{1}{x} \times 1 = \frac{1}{x}$$

And so, we have that if $f(x) = \log_e x$ then $f'(x) = \frac{1}{x}$

Derivative of a Product of Functions

Many functions can be written as the product of two (or more) functions. For example, the function $y = (x^3 - 2x)(x^2 + x - 3)$ is made up of the product of two simpler functions of x. In fact, expressions such as these take on the general form $y = u \times v$ or $y = f(x) \times g(x)$ where (in this case) we have $u = f(x) = (x^3 - 2x)$ and $v = g(x) = (x^2 + x - 3)$.

To differentiate such expressions we use the product rule, which can be written as:

Function	Derivative
$y = u \times v$	$\frac{dy}{dx} = \frac{du}{dx} \times v + u \times \frac{dv}{dx}$
$y = f(x) \times g(x)$	$\frac{dy}{dx} = f'(x) \times g(x) + f(x) \times g'(x)$

Example 6.2.4

Differentiate the following.

a $x^2\sin(x)$ b $(x^3 - 2x + 1)e^x$

c $\frac{1}{x}\log_e(x)$

a A useful method to find the derivative of a product makes use of the following table:

Function	Derivative
$u = x^2$	$\frac{du}{dx} = 2x$
$v = \sin(x)$	$\frac{dv}{dx} = \cos(x)$

$2x\sin(x)$

$x^2\cos(x)$

Adding

$2x\sin(x) + x^2\cos(x)$

Let $y = x^2\sin(x)$ so that $u = x^2$ and $v = \sin(x)$.

So that $\frac{du}{dx} = 2x$ and $\frac{dv}{dx} = \cos(x)$.

Using the product rule we have $\frac{dy}{dx} = \frac{du}{dx} \times v + u \times \frac{dv}{dx}$

$$= 2x \times \sin(x) + x^2 \times \cos(x)$$
$$= 2x\sin(x) + x^2\cos(x)$$

b Let

$y = (x^3 - 2x + 1)e^x$ so that $u = (x^3 - 2x + 1)$ and $v = e^x$.

Then, $\frac{du}{dx} = 3x^2 - 2$ and $\frac{dv}{dx} = e^x$.

Using the product rule:

$$\frac{dy}{dx} = \frac{du}{dx} \times v + u \times \frac{dv}{dx}$$
$$= (3x^2 - 2) \times e^x + (x^3 - 2x + 1) \times e^x$$
$$= (3x^2 - 2 + x^3 - 2x + 1)e^x$$
$$= (x^3 + 3x^2 - 2x - 1)e^x$$

c This time set up a table:

Let $y = \frac{1}{x}\log_e x$ with $u = \frac{1}{x}$ and $v = \log_e x$.

Function	Derivative
$u = \frac{1}{x}$	$\frac{du}{dx} = -\frac{1}{x^2}$
$v = \log_e x$	$\frac{dv}{dx} = \frac{1}{x}$

$-\frac{1}{x^2} \times \log_e x$

$\frac{1}{x} \times \frac{1}{x}$

Adding: $\frac{dy}{dx} = -\frac{1}{x^2} \times \log_e x + \frac{1}{x} \times \frac{1}{x}$

$$= -\frac{1}{x^2} \times \log_e x + \frac{1}{x^2}$$
$$= \frac{1}{x^2}(1 - \log_e x)$$

Derivative of a Quotient of Functions

In the same way as we have a rule for the product of functions, we also have a rule for the quotient of functions.

For example, the function $y = \frac{x^2}{x^3+x-1}$

is made up of two simpler functions of x. Expressions like this take on the general form

$y = \frac{u}{v}$ or $y = \frac{f(x)}{g(x)}$.

For the example shown above, we have that $u = x^2$ and $v = x^3+x-1$.

As for the product rule, we state the result.

To differentiate such expressions we use the quotient rule, which can be written as:

Function	Derivative
$y = \frac{u}{v}$	$\frac{dy}{dx} = \frac{\frac{du}{dx}\times v - u\times\frac{dv}{dx}}{v^2}$
$y = \frac{f(x)}{g(x)}$	$\frac{dy}{dx} = \frac{f'(x)\times g(x) - f(x)\times g'(x)}{[g(x)]^2}$

Example 6.2.5

Differentiate the following.

a $\frac{x^2+1}{\sin(x)}$ b $\frac{e^x+x}{x+1}$ c $\frac{\sin(x)}{1-\cos(x)}$

a We express $\frac{x^2+1}{\sin(x)}$ in the form $y = \frac{u}{v}$, so that:

$u = x^2+1$ and $v = \sin(x)$.

Giving the following derivatives, $\frac{du}{dx} = 2x$ and $\frac{dv}{dx} = \cos(x)$

Using the quotient rule we have,

$$\frac{dy}{dx} = \frac{\frac{du}{dx}\times v - u\times\frac{dv}{dx}}{v^2}$$

$$= \frac{2x\times\sin(x) - (x^2+1)\times\cos(x)}{[\sin(x)]^2}$$

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

b First express $\frac{e^x+x}{x+1}$ in the form $y = \frac{u}{v}$

so that $u = e^x+x$ and $v = x+1$ and

$\frac{du}{dx} = e^x+1$ and $\frac{dv}{dx} = 1$. Using the quotient rule, we have

$$\frac{dy}{dx} = \frac{\frac{du}{dx}\times v - u\times\frac{dv}{dx}}{v^2} = \frac{(e^x+1)\times(x+1) - (e^x+x)\times 1}{(x+1)^2}$$

$$= \frac{xe^x+e^x+x+1-e^x-x}{(x+1)^2}$$

$$= \frac{xe^x+1}{(x+1)^2}$$

c Express the quotient $\frac{\sin(x)}{1-\cos(x)}$ in the form $y = \frac{u}{v}$,

so that $u = \sin(x)$ and $v = 1-\cos(x)$.

Then $\frac{du}{dx} = \cos(x)$ and $\frac{dv}{dx} = \sin(x)$.

Using the quotient rule, we have

$$\frac{dy}{dx} = \frac{\frac{du}{dx}\times v - u\times\frac{dv}{dx}}{v^2} = \frac{\cos(x)\times(1-\cos(x)) - \sin(x)\times\sin(x)}{(1-\cos(x))^2}$$

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

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

$$= \frac{\cos(x)-1}{(1-\cos(x))^2}$$

$$= -\frac{(1-\cos(x))}{(1-\cos(x))^2}$$

$$= -\frac{1}{(1-\cos(x))}$$

The Chain Rule

To find the derivative of x^3+1 we let $y = x^3+1$ so that $\frac{dy}{dx} = 3x^2$.

Next consider the derivative of the function $y = (x^3+1)^2$ We first expand the brackets, $y = x^6+2x^3+1$, and obtain $\frac{dy}{dx} = 6x^5+6x^2$.

This expression can be simplified (i.e. factorised), giving:

$\frac{dy}{dx} = 6x^2(x^3+1)$.

In fact, it isn't too great a task to differentiate the function $y = (x^3+1)^3$.

As before, we expand $y = x^9 + 3x^6 + 3x^3 + 1$

so that $\frac{dy}{dx} = 9x^8 + 18x^5 + 9x^2$.

Factorizing this expression we now have:

$$\frac{dy}{dx} = 9x^2(x^6 + 2x^3 + 1) = 9x^2(x^3+1)^2.$$

But what happens if we need to differentiate the expression $y = (x^3+1)^8$? We could expand and obtain a polynomial with 9 terms (!), which we then proceed to differentiate and obtain a polynomial with 8 terms … and of course, we can then factorise that polynomial. The question arises, "Is there an easier way to do this?"

We can obtain some idea of how to do this by summarizing the results found so far:

Function	Derivative	(Factored form)
$y = x^3+1$	$\frac{dy}{dx} = 3x^2$	$3x^2$
$y = (x^3+1)^2$	$\frac{dy}{dx} = 6x^5 + 6x^2$	$2 \times 3x^2(x^3+1)$
$y = (x^3+1)^3$	$\frac{dy}{dx} = 9x^8 + 18x^5 + 9x^2$	$3 \times 3x^2(x^3+1)^2$

Continuing $y = (x^3+1)^4$,

$$\frac{dy}{dx} = 12x^{11} + 36x^8 + 36x^5 + 12x^2 \text{ or } 4 \times 3x^2(x^3+1)^3$$

The pattern that is emerging is that if:

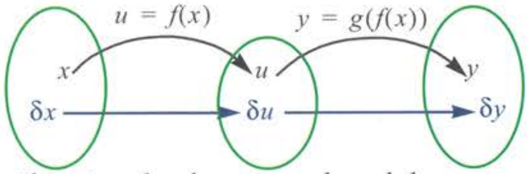
$$y = (x^3+1)^n \text{ then } \frac{dy}{dx} = n \times 3x^2(x^3+1)^{n-1}.$$

In fact, if we consider the term inside the brackets as one function, so that the expression is actually a composition of two functions, namely that of x^3+1 and the power function, we can write $u = x^3+1$ and $y = u^n$.

So that $\frac{dy}{du} = nu^{n-1} = n(x^3+1)^{n-1}$ and $\frac{du}{dx} = 3x^2$, giving the result $\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$.

Is this a 'one–off' result, or can we determine a general result that will always work?

To explore this we use a graphical approach to see why it may be possible to obtain a general result.



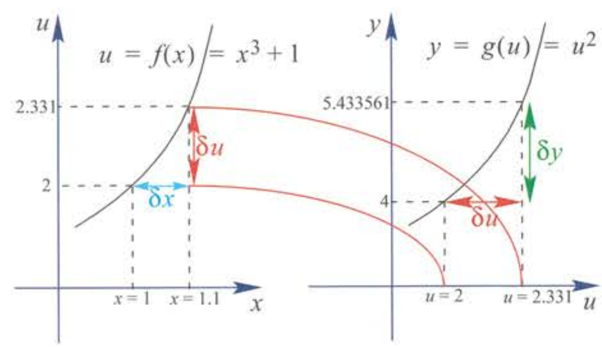
We start by using the above example and then move onto a more general case. For the function $y = (x^3+1)^2$, we let $u = x^3+1$ $(= g(x))$ and so $y = u^2$ $(= f(g(x)))$. We need to find what effect a small change in x will have on the function y (via u), i.e. what effect will δx have on y?

We have a sort of **chain reaction**, that is, a small change in x, δx, will produce a change in u, δu, which in turn will produce a change in y, δy! It is the path from δx to δy that we are interested in.

This can be seen when we produce a graphical representation of the discussion so far.

We start by looking at the effect that a change in x has on u:

Then we observe the effect that the change in u has on y:



We then have $\delta x = 1.1 - 1 = 0.1$ and $\delta u = 2.331 - 2 = 0.331$

Similarly, $\delta u = 2.331 - 2 = 0.331$ and $\delta y = 5.433561 - 4 = 1.433561$

Based on these results, the following relationship can be seen to hold:

$$\boxed{\frac{\delta y}{\delta x} = \frac{\delta y}{\delta u} \times \frac{\delta u}{\delta x}}$$

The basic outline in proving this result is shown in the following argument:

Let δx be a small increment in the variable x and let δu be the corresponding increment in the variable u. This change in u will in turn produce a corresponding change δy in y.

As δx tends to zero, so does δu. We will assume that $\delta u \neq 0$ when $\delta x \neq 0$.

Hence we have that $\frac{\delta y}{\delta x} = \frac{\delta y}{\delta u} \cdot \frac{\delta u}{\delta x} \Rightarrow \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = \lim_{\delta x \to 0} \frac{\delta y}{\delta u} \cdot \frac{\delta u}{\delta x}$

$$= \left(\lim_{\delta x \to 0} \frac{\delta y}{\delta u}\right) \cdot \left(\lim_{\delta x \to 0} \frac{\delta u}{\delta x}\right) \quad \text{Given that:}$$

$$= \left(\lim_{\delta u \to 0} \frac{\delta y}{\delta u}\right) \cdot \left(\lim_{\delta x \to 0} \frac{\delta u}{\delta x}\right) \quad \delta x \to 0 \Rightarrow \delta u \to 0$$

$$\therefore \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

We then have the result: $\boxed{\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}}$

Using the chain rule

We will work our way through an example, showing the critical steps involved when using the chain rule.

This is highlighted by finding the derivative of the function $y = \sin(x^2)$.

Step 1	Recognition
This is the most important step when deciding if using the chain rule is appropriate. In this case we *recognise* that the function $y = \sin(x^2)$ is a composite of the sine and the squared functions.	
Step 2	Define u (or $g(x)$)
	Let the 'inside' function be u. In this case, we have that $u = x^2$
Step 3	Differentiate u (with respect to x)
	$\frac{du}{dx} = 2x$
Step 4	Express y in terms of u
	$y = \sin(u)$
Step 5	Differentiate y (with respect to u)
	$\frac{dy}{du} = \cos(u)$
Step 6	Use the chain rule
	$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \cos(u) \times 2x = 2x\cos(x^2)$

Example 6.2.6

Differentiate the following functions.

a $y = \log_e(x + \cos x)$ b $f(x) = (1 - 3x^2)^4$

a Begin by letting $u = x + \cos(x) \Rightarrow \frac{du}{dx} = 1 - \sin(x)$.

Express y in terms of u, that is,

$$y = \log_e u \Rightarrow \frac{dy}{du} = \frac{1}{u}\left(= \frac{1}{x + \cos(x)}\right).$$

Using the chain rule we have:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \frac{1}{x + \cos(x)} \cdot (1 - \sin(x)) = \frac{1 - \sin(x)}{x + \cos(x)}$$

b This time we let $g(x) = 1 - 3x^2$, so that $g'(x) = -6x$.

Now let $f(x) = (h \circ g)(x)$ so that $h(g(x)) = (g(x))^4$ and $h'(g(x)) = 4(g(x))^3$.

Therefore, using the chain rule we have

$$f'(x) = (h \circ g)'(x) = h'(g(x)) \cdot g'(x)$$
$$= 4(g(x))^3 \times (-6x)$$
$$= -24x(1 - 3x^2)^3$$

Some standard derivatives

Often we wish to differentiate expressions of the form $y = \sin(2x)$ or $y = e^{5x}$ or other such functions, where the x term only differs by a constant factor from that of the basic function. That is, the only difference between $y = \sin(2x)$ and $y = \sin(x)$ is the factor '2'. We can use the chain rule to differentiate such expressions:

Let $u = 2x$, giving $y = \sin(u)$ and so

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \cos(u) \times 2 = 2\cos(2x)$$

Similarly,

Let $u = 5x$, giving $y = e^u$ and so

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = e^u \times 5 = 5e^{5x}$$

Because of the nature of such derivatives, functions such as these form part of a set of functions that can be considered as having derivatives that are often referred to as standard derivatives. Although we could make use of the chain rule to differentiate these functions, they should be viewed as standard derivatives.

These standard derivatives are shown in the table (where k is some real constant):

y	$\frac{dy}{dx}$
$\sin(kx)$	$k\cos(kx)$
$\cos(kx)$	$-k\sin(kx)$
$\tan(kx)$	$k\sec^2(kx)$
e^{kx}	ke^{kx}
$\log_e(kx)$	$\frac{1}{x}$

Notice, the only derivative that does not involve the constant k is that of the logarithmic function. This is because letting $u = kx$, we have $y = \log(u)$ so:

$$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx} = \frac{1}{u}\times k = \frac{1}{\cancel{k}x}\times\cancel{k} = \frac{1}{x}.$$

When should the chain rule be used?

A good **first rule** to follow is: If the expression is made up of a pair of brackets and a power, then the chances are that you will need to use the chain rule.

As a start, the expressions in the list of results that follows would require the use of the chain rule. Notice then that in each case the expression can be (or already is) written in 'power form'. That is, of the form $y = [f(x)]^n$.

Some examples to be on the lookout for:

$y = (2x+6)^5$ - Let $u = 2x+6$ and $y = u^5$

$y = \sqrt{(2x^3+1)}$ - Let $u = 2x^3+1$ and $y = u^{\frac{1}{2}}$

$y = \frac{3}{(x-1)^2}, x \neq 1$ - Let $u = x-1$ and $y = 3u^{-2}$

$y = \frac{1}{\sqrt[3]{e^{-x}+e^x}}$ - Let $u = e^{-x}+e^x$ and $y = u^{-\frac{1}{3}}$

We now look at some of the more demanding derivatives which combine at least two rules of differentiation, for example, the need to use both the quotient rule and the chain rule, or the product rule and the chain rule.

Example 6.2.7

Differentiate the following.

a $y = \sqrt{1+\sin^2 x}$ b $y = e^{x^3}\sin(1-2x)$

c $x \mapsto \frac{x}{\sqrt{x^2+1}}$

a Let $y = \sqrt{(1+\sin^2 x)} = (1+\sin^2 x)^{1/2}$.

Using the chain rule we have

$$\frac{dy}{dx} = \frac{1}{2}\times\frac{d}{dx}(1+\sin^2 x)\times(1+\sin^2 x)^{-1/2}$$

$$= \frac{1}{2}\times(2\sin x\cos x)\times\frac{1}{\sqrt{(1+\sin^2 x)}}$$

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

b Let $y = e^{x^3}\sin(1-2x)$.

Using the product rule first, we have

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

$$= 3x^2e^{x^3}\sin(1-2x) + e^{x^3}\times -2\cos(1-2x)$$

$$= e^{x^3}(3x^2\sin(1-2x) - 2\cos(1-2x))$$

c Let: (Quotient rule)

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

$$= \frac{1\times\sqrt{x^2+1} - x\times\frac{1}{2}\times 2x\times(x^2+1)^{-\frac{1}{2}}}{x^2+1}$$

$$= \frac{\sqrt{x^2+1} - \frac{x^2}{\sqrt{x^2+1}}}{(x^2+1)}$$

$$= \frac{\frac{(\sqrt{x^2+1})^2 - x^2}{\sqrt{x^2+1}}}{(x^2+1)}$$

$$= \frac{1}{(x^2+1)\sqrt{x^2+1}}$$

Example 6.2.8

Differentiate the following.

a $y = \ln\left(\frac{x}{x+1}\right), x > 0$ b $y = \sin(\ln t)$

c $y = x\ln(x^2)$

a $y = \ln\left(\frac{x}{x+1}\right) = \ln(x) - \ln(x+1)$

$$\therefore \frac{dy}{dx} = \frac{1}{x} - \frac{1}{x+1} = \frac{(x+1)-x}{x(x+1)} = \frac{1}{x(x+1)}$$

Notice that using the log laws to first simplify this expression made the differentiation process much easier.

The other approach, i.e. letting $u = \frac{x}{x+1}$, $y = \ln(u)$ and then using the chain rule would have meant more work – as not only would we need to use the chain rule but also the quotient rule to determine $\frac{du}{dx}$.

b Let $u = \ln t$ so that $y = \sin u$.

Using the chain rule we have:

$$\frac{dy}{dt} = \frac{dy}{du}\cdot\frac{du}{dt} = \cos(u)\times\frac{1}{t} = \frac{\cos(\ln t)}{t}$$

c Here we have a product $x\times\ln(x^2)$, so that the product rule needs to be used and then we need the chain rule to differentiate $\ln(x^2)$.

In this case we cannot simply rewrite $\ln(x^2)$ as $2\ln(x)$. Why?

Because the functions $\ln(x^2)$ and $2\ln(x)$ may have different domains. That is, the domain of $\ln(x^2)$ is all real values excluding zero (assuming an implied domain) whereas the domain of $2\ln(x)$ is only the positive real numbers. However, if it had been specified that $x > 0$, then we could have 'converted' $\ln(x^2)$ to $2\ln(x)$.

$$\frac{dy}{dx} = \frac{d}{dx}(x)\times\ln(x^2) + x\times\frac{d}{dx}(\ln(x^2)) = 1\times\ln(x^2) + x\times\frac{2x}{x^2}$$

$$= \ln(x^2) + 2$$

Derivative of reciprocal circular functions

Dealing with the functions $\sec(x)$, $\cot(x)$ and $\operatorname{cosec}(x)$ is achieved by rewriting them as their reciprocals:

$$\sec(x) = \frac{1}{\cos(x)},\ \cot(x) = \frac{1}{\tan(x)}\ \ \operatorname{cosec}(x) = \frac{1}{\sin(x)}.$$

Once this is done, make use of the chain rule.

$$\text{Eg } \frac{d}{dx}(\operatorname{cosec}x) = \frac{d}{dx}\left(\frac{1}{\sin x}\right) = \frac{d}{dx}[(\sin x)^{-1}]$$

$$= -1\times\cos x\times(\sin x)^{-2} = -\frac{\cos x}{(\sin x)^2}$$

We could leave the answer as is or simplify it as follows:

$$-\frac{\cos x}{\sin x\sin x} = -\cot x\operatorname{cosec}x.$$

Rather than providing a table of 'standard results' for the derivative of the reciprocal circular trigonometric functions, we consider them as special cases of the circular trigonometric functions.

Example 6.2.9

Differentiate the following.

a $f(x) = \cot 2x$, $x > 0$ b $y = \sec^2 x$

c $y = \frac{\ln(\operatorname{cosec}x)}{x}$

a
$$f(x) = \cot 2x = \frac{1}{\tan 2x} = (\tan 2x)^{-1}$$
$$\therefore f'(x) = -1\times 2\sec^2 2x\times(\tan 2x)^{-2}$$
$$= -\frac{2\sec^2 2x}{\tan^2 2x}$$
$$= -2\times\frac{1}{\cos^2 2x}\times\frac{1}{\tan^2 2x}$$
$$= -2\times\frac{1}{\cos^2 2x}\times\frac{\cos^2 2x}{\sin^2 2x}$$
$$= -2\operatorname{cosec}^2 2x$$

b
$$y = \sec^2 x = \frac{1}{(\cos x)^2} = (\cos x)^{-2}$$
$$\therefore\frac{dy}{dx} = -2\times -\sin x\times(\cos x)^{-3} = \frac{2\sin x}{(\cos x)^3}$$
$$= 2\times\frac{\sin x}{\cos x}\times\frac{1}{(\cos x)^2}$$
$$= 2\tan x\sec^2 x$$

c
$$y = \frac{\ln(\operatorname{cosec}x)}{x} = \frac{\ln[(\sin x)^{-1}]}{x} = -\frac{\ln(\sin x)}{x}$$
$$\therefore\frac{dy}{dx} = -\frac{\left(\frac{\cos x}{\sin x}\right)\times x - 1\times\ln(\sin x)}{x^2}$$
$$= \frac{\frac{x\cos x - \sin x\ln(\sin x)}{\sin x}}{x^2}$$
$$= \frac{x\cos x - \sin x\ln(\sin x)}{x^2\sin x}$$

An interesting result

A special case of the chain rule involves the case $y = x$.

By viewing this as an application of the chain rule:

$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx}$ we have (after setting $y = x$):

$$\frac{d(x)}{dx} = \frac{dx}{du}\cdot\frac{du}{dx} \Rightarrow 1 = \frac{dx}{du}\cdot\frac{du}{dx}\ \text{ i.e. } \frac{dx}{du} = 1/\frac{du}{dx}$$

This important result is often written in the form: $\boxed{\dfrac{dy}{dx} = \dfrac{1}{\left(\dfrac{dx}{dy}\right)}}$

We find that this result is useful with problems that deal with related rates.

Exercise 6.2.3

1. Use the product rule to differentiate the following and then verify your answer by first expanding the brackets.

   a $(x^2+1)(2x-x^3+1)$

   b $(x^3+x^2)(x^3+x^2-1)$

   c $\left(\dfrac{1}{x^2}-1\right)\left(\dfrac{1}{x^2}+1\right)$

   d $(x^3+x-1)(x^3+x+1)$

2. Use the quotient rule to differentiate the following.

   a $\dfrac{x+1}{x-1}$ b $\dfrac{x}{x+1}$ c $\dfrac{x+1}{x^2+1}$

   d $\dfrac{x^2+1}{x^3-1}$ e $\dfrac{x^2}{2x+1}$ f $\dfrac{x}{1-2x}$

3. Differentiate the following.

   a $e^x\sin x$ b $x\log_e x$

   c $e^x(2x^3+4x)$ d $x^4\cos x$

   e $\sin x\cos x$ f $(1+x^2)\tan x$

4. Differentiate the following.

   a $\dfrac{x}{\sin x}$ b $\dfrac{\cos x}{x+1}$ c $\dfrac{e^x}{e^x+1}$

   d $\dfrac{\sin x}{\sqrt{x}}$ e $\dfrac{x}{\log_e x}$ f $\dfrac{\log_e x}{x+1}$

5. Differentiate the following.

   a $e^{-5x}+x$

   b $\sin 4x - \dfrac{1}{2}\cos 6x$

   c $e^{-\frac{1}{3}x} - \log_e(2x) + 9x^2$

   d $5\sin(5x)+3e^{2x}$

   e $\tan(4x)+e^{2x}$

6. Differentiate the following.

   a $\sin x^2 + \sin^2 x$ b $\tan(2\theta)+\dfrac{1}{\sin\theta}$

   c $\sin\sqrt{x}$ d $\cos\left(\dfrac{1}{x}\right)$

   e $\cos^3\theta$ f $\sin(e^x)$

   g $\tan(\log_e x)$ h $\sqrt{\cos(2x)}$

7. Differentiate the following.

   a e^{2x+1} b $2e^{4-3x}$

   c $2e^{4-3x^2}$ d $\sqrt{e^x}$

   e $e^{\sqrt{x}}$ f $\dfrac{1}{2}e^{2x+4}$

   g $\dfrac{1}{2}e^{2x^2+4}$ h $\dfrac{2}{e^{3x+1}}$

   i e^{3x^2-6x+1} j $e^{\sin(\theta)}$

8. Differentiate the following.

   a $\log_e(x^2+1)$ b $\log_e(\sin\theta+\theta)$

   c $\log_e(e^x-e^{-x})$ d $\log_e\left(\dfrac{1}{x+1}\right)$

   e $(\log_e x)^3$ f $\sqrt{\log_e x}$

   g $\log_e(\sqrt{x-1})$ h $\log_e(1-x^3)$

9. Differentiate the following.

   a $x\log_e(x^3+2)$ b $\sqrt{x}\sin^2 x$

   c $\cos^2\sqrt{\theta}$ d $x^3e^{-2x^2+3}$

   e $\cos(x\log_e x)$ f $\log_e(\log_e x)$

   g $\dfrac{x^2-4x}{\sin(x^2)}$ h $\dfrac{10x+1}{\log_e(10x+1)}$

10. Find the value of x where the function $x \mapsto xe^{-x}$ has a horizontal tangent.

11. Find the gradient of the function $x \mapsto \sin\left(\dfrac{1}{x}\right)$, where $x = \dfrac{2}{\pi}$.

12. Find the gradient of the function $x \mapsto \log_e(x^2+4)$ at the point where the function crosses the y-axis.

13. For what value(s) of x will the function $x \mapsto \ln(x^2+1)$ have a gradient of 1.

14. Find the rate of change of the function $x \mapsto e^{-x^2+2}$ at the point $(1, e)$.

15. Find: a $\frac{d}{dx}(\sin x \cos x)$ b $\frac{d}{dx}(\sin x°)$

16. a

If y is the product of three functions, i.e. $y = f(x)g(x)h(x)$, show that:

$$\frac{dy}{dx} = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x).$$

b Hence, differentiate the following:

i $x^2 \sin x \cos x$ ii $e^{-x^3}\sin(2x)\log_e(\cos x)$

17. a Given that $f(x) = 1 - x^3$ and $g(x) = \log_e x$, find:

i $(f \circ g)'(x)$ ii $(g \circ f)'(x)$

. b Given that $f(x) = \sin(x^2)$ and $g(x) = e^{-x}$, find:

i $(f \circ g)'(x)$ ii $(g \circ f)'(x)$

18. Given that $T(\theta) = \frac{\cos k\theta}{2 + 3\sin k\theta}, k \neq 0$, determine $T\left(\frac{\pi}{2k}\right)$.

19. If $f(x) = (x-a)^m(x-b)^n$, find x such that $f'(x) = 0$

20. If $f(\theta) = \sin\theta^m \cos\theta^n$, find θ such that $f'(\theta) = 0$.

21. Differentiate the following.

a	$f(x) = \cot 4x$	b	$g(x) = \sec 2x$
c	$f(x) = \operatorname{cosec} 3x$	d	$y = \sin\left(3x + \frac{\pi}{2}\right)$

22. Differentiate the following.

a	$\sec x^2$	b	$\sin x \sec x$
c	$\ln(\sec x)$	d	$\cot^3 x$
e	$\frac{x}{\operatorname{cosec} x}$	f	$\frac{\operatorname{cosec} x}{\sin x}$

Extra questions

Derivative of Inverse Trigonometric Functions

In this section, over an appropriate domain, either expression $\text{Sin}^{-1}(x)$ or $\arcsin(x)$ can be used. Similarly we can use for $\text{Cos}^{-1}(x)$ and $\arccos(x)$ as well as $\text{Tan}^{-1}(x)$ and $\arctan(x)$. That is,

$\text{Sin}^{-1}(x) = \arcsin(x), -1 \le x \le 1$,

$\text{Cos}^{-1}(x) = \arccos(x), -1 \le x \le 1$,

$\text{Tan}^{-1}(x) = \arctan(x), -\infty < x < \infty$.

It is important to keep track of how the domain of some functions is not the same as that of their derived function. For example, although the function $y = \arcsin(x)$ is defined for $-1 \le x \le 1$ its derived function, $\frac{dy}{dx}$ is defined for $-1 < x < 1$. i.e. the end points, $x = \pm 1$ are not included.

Derivative of $\text{Sin}^{-1}(x)$

By definition, $\text{Sin}^{-1}(x)$ is defined for $x \in [-1, 1]$. We start by letting $y = \text{Sin}^{-1}(x), -1 \le x \le 1$.

Then we have that $y = \text{Sin}^{-1}(x) \Leftrightarrow x = \sin y, -\frac{\pi}{2} \le y \le \frac{\pi}{2}$.

So, $\frac{dx}{dy} = \cos y, -\frac{\pi}{2} \le y \le \frac{\pi}{2} \Rightarrow \frac{dy}{dx} = \frac{1}{\cos y}, -\frac{\pi}{2} < y < \frac{\pi}{2}$

Note the change in domains!

Now we express $\cos y$ back in terms of x: Using the identity $\cos^2 y + \sin^2 y = 1$ we have:

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

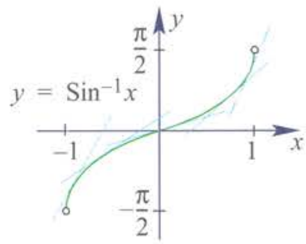
So, at this stage, the derivative of $\text{Sin}^{-1}(x)$ is given by:

$\frac{dy}{dx} = \pm\frac{1}{\sqrt{1-x^2}}, -1 < x < 1$.

However, over the interval $-\frac{\pi}{2} < y < \frac{\pi}{2}$ we have that $\cos y$ is positive and so we only use the positive square root.

We then have the result that $\frac{dy}{dx} = \frac{1}{\sqrt{1-x^2}}, -1 < x < 1$.

Note that we could have arrived at the same conclusion about the sign of the derivative by looking at the graph of $\text{Sin}^{-1}(x)$ for $x \in (-1, 1)$.



Using the graph of $\text{Sin}^{-1}(x)$ for $x \in (-1, 1)$, we can see that over the given interval the gradient anywhere on the curve is always positive and so we have to choose the positive square root.

Derivative of $y = \text{Sin}^{-1}\left(\frac{x}{a}\right), -a \le x \le a$ **where $a > 0$**

Using the chain rule for $y = \text{Sin}^{-1}\left(\frac{x}{a}\right), -a \le x \le a$ we set $u = \frac{x}{a} \Rightarrow y = \text{Sin}^{-1}u$, $-1 \le u \le 1$, which then gives:

$$\frac{dy}{dx} = \frac{dy}{du}\cdot\frac{du}{dx} = \frac{1}{\sqrt{1-u^2}}\times\frac{1}{a} = \frac{a}{\sqrt{a^2-x^2}}\times\frac{1}{a} = \frac{1}{\sqrt{a^2-x^2}}$$

$-1 < \frac{x}{a} < 1$.

Note: $\sqrt{1-u^2} = \sqrt{1-\left(\frac{x}{a}\right)^2} = \sqrt{\frac{a^2-x^2}{a^2}} = \frac{\sqrt{a^2-x^2}}{a}$

$\therefore \frac{1}{\sqrt{1-u^2}} = \frac{a}{\sqrt{a^2-x^2}}$

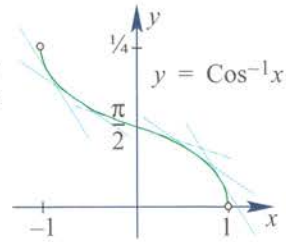
$$\frac{d}{dx}\left(\text{Sin}^{-1}\frac{x}{a}\right) = \frac{1}{\sqrt{a^2-x^2}}, -a < x < a$$

Derivative of $\text{Cos}^{-1}(x)$

Starting with the principal cosine function $f(x) = \cos x, 0 \le x \le \pi$ we define the inverse cosine function, $f^{-1}(x)$ as $f^{-1}(x) = \text{Cos}^{-1}(x), -1 \le x \le 1$. Letting $y = f^{-1}(x)$ we have $y = \text{Cos}^{-1}(x), -1 \le x \le 1$ so that $x = \cos y, 0 \le y \le \pi$ Differentiating both sides with respect to y we have

$\frac{dx}{dy} = -\sin y, 0 \le y \le \pi \Rightarrow \frac{dy}{dx} = -\frac{1}{\sin y}, 0 < y < \pi$

Note the change in domains!



Using the trigonometric identity $\sin^2 y = 1 - \cos^2 y$ we have $\sin y = \pm\sqrt{1-\cos^2 y} = \pm\sqrt{1-x^2}$.

Therefore,
$\frac{dy}{dx} = -\frac{1}{\pm\sqrt{1-x^2}}, -1 < y < 1$.

We now need to determine which sign to choose. Using the graph of $y = \text{Cos}^{-1}(x), -1 < x < 1$ we see that, over this domain, the gradient is always negative and so we choose $\sqrt{1-x^2}$.

$$\frac{d}{dx}(\text{Cos}^{-1}x) = -\frac{1}{\sqrt{1-x^2}}, -1 < x < 1$$
and
$$\frac{d}{dx}\left(\text{Cos}^{-1}\frac{x}{a}\right) = -\frac{1}{\sqrt{a^2-x^2}}, -a < x < a$$

Derivative of $\text{Tan}^{-1}(x)$

Again, we start with a principal tangent function $f(x) = \tan x, -\frac{\pi}{2} < x < \frac{\pi}{2}$.

We define the inverse tangent function, $f^{-1}(x)$ as $f^{-1}(x) = \text{Tan}^{-1}(x), -\infty < x < \infty$. Letting $y = f^{-1}(x)$ we have $y = \text{Tan}^{-1}(x), -\infty < x < \infty$ so that $x = \tan y, -\frac{\pi}{2} < y < \frac{\pi}{2}$. Differentiating both sides with respect to y we have

$\frac{dx}{dy} = \sec^2 y, -\frac{\pi}{2} < y < \frac{\pi}{2} \le \frac{dy}{dx} = \frac{1}{\sec^2 y}, -\frac{\pi}{2} < y < \frac{\pi}{2}$.

However, $\tan^2 y + 1 = \sec^2 y$, therefore:

$\frac{dy}{dx} = \frac{1}{1+\tan^2 y}, -\frac{\pi}{2} < y < \frac{\pi}{2} = \frac{1}{1+x^2}, -\infty < x < \infty$.

$$\frac{d}{dx}(\text{Tan}^{-1}x) = \frac{1}{1+x^2}, -\infty < x < \infty$$
and
$$\frac{d}{dx}\left(\text{Tan}^{-1}\frac{x}{a}\right) = \frac{a}{a^2+x^2}, -\infty < x < \infty$$

Example 6.2.10

Differentiate the following and specify the domain of the derivative.

a $f(x) = \arcsin\left(\frac{x}{2}\right)$ b $y = \arctan(x+2)$

c $y = \arccos(x^2-9)$

a Let $u = \frac{x}{2}$ so that $f(x) = \arcsin(u)$ and as $-1 \le u \le 1 \Rightarrow -2 \le x \le 2$.

Using the chain rule we have,

$$f'(x) = \frac{d}{du}(\arcsin(u)) \cdot \frac{du}{dx} = \frac{1}{\sqrt{1-u^2}} \times \frac{1}{2} = \frac{1}{\sqrt{1-\left(\frac{x}{2}\right)^2}} \times \frac{1}{2}$$

$$= \frac{1}{2\sqrt{\frac{4-x^2}{4}}}$$

That is, $f'(x) = \dfrac{1}{\sqrt{4-x^2}}, -2 < x < 2$.

Note: We could have simply used the standard result,

$\dfrac{d}{dx}\left(\arcsin\left(\dfrac{x}{a}\right)\right) = \dfrac{1}{\sqrt{a^2-x^2}}, -a < x < a$ with $a = 2$!

b Using the chain rule:

$$\frac{d}{dx}(\arctan[x+2]) = \frac{1}{1+[x+2]^2}, -\infty < x+2 < \infty.$$

c Using the chain rule,

$$\frac{dy}{dx} = -\frac{2x}{\sqrt{1-[x^2-9]^2}}, x \in (-\sqrt{10}, -2\sqrt{2}) \cup (2\sqrt{2}, \sqrt{10})$$

Exercise 6.2.4

1. Differentiate with respect to x, each of the following.

a	$\arctan(2x)$	b	$\arcsin\left(\frac{x}{3}\right)$
c	$\mathrm{Cos}^{-1}(2x)$	d	$\mathrm{Sin}^{-1}(4x)$
e	$\arctan\left(\frac{x}{2}\right)$	f	$\arcsin(x-1)$

2. Differentiate with respect to x, each of the following.

a	$\arctan(x^2)$	b	$\mathrm{Sin}^{-1}(\sqrt{x})$
c	$\arccos\left(\frac{1}{\sqrt{x}}\right)$	d	$\arcsin(\cos x)$
e	$\arctan(\sqrt{x-1})$	f	$\ln(\arcsin x)$
g	$\mathrm{Tan}^{-1}(e^x)$	h	$\arccos(e^{-x})$

3. Differentiate with respect to x, each of the following.

a	$x\mathrm{Tan}^{-1}x$	b	$\frac{\arcsin x}{x}$
c	$\frac{x}{\arccos x}$	d	$\frac{\arctan x}{x^2}$
e	$\arcsin(x^2)\ln x$	f	$\frac{1}{\sqrt{x}}\mathrm{Cos}^{-1}\sqrt{x}$

4. Find $\dfrac{dy}{dx}$ if $y = \arctan\left(\dfrac{x}{x+1}\right) + \arctan\left(\dfrac{x+1}{x}\right)$.

   Hence, find the real value of k, if $y = k$.

5. Show that if $y\sqrt{1-x^2} = \mathrm{Sin}^{-1}x$, then $(1-x^2)y' = 1 + xy$.

6.a Show that $\mathrm{Sin}^{-1}x + \mathrm{Cos}^{-1}x = k$, where k is a real number.

   b Find the value of k.

7. Differentiate the following and find the implied domain for each of $f(x)$ and $f'(x)$.

   a $f(x) = \arcsin\left(\dfrac{\pi}{x}\right)$

   b $f(x) = \mathrm{Cos}^{-1}\left(\dfrac{1}{x} - 1\right)$

   c $f(x) = (\mathrm{Sin}^{-1}x)\left(\mathrm{Cos}^{-1}\dfrac{x}{2}\right)$

Extra questions

Differentiating $y = a^x$

We have already considered the derivative of the natural exponential function $y = e^x$. We extend this to a more general form of the exponential function, namely, $y = a^x, a \neq 0, 1$.

The process is straightforward, requiring an algebraic rearrangement of $y = a^x$.

Taking log (base e) of both sides of the equation, we have

$$y = a^x \Leftrightarrow \log_e y = \log_e a^x$$

So that, $\log_e y = x\log_e a$

Next, we differentiate both sides of the equation:

$$\frac{d}{dx}(\log_e y) = \frac{d}{dx}(x\log_e a)$$

Now (this is the tricky bit):

Using the fact that $\dfrac{d}{dx}(\log_e f(x)) = \dfrac{f'(x)}{f(x)}$ or $\dfrac{1}{f(x)} \times f'(x)$ and since y is a function of x, we can write $\dfrac{d}{dx}(\log_e y) = \dfrac{1}{y} \cdot \dfrac{dy}{dx}$.

That is, we have replaced $f(x)$ with y.

This means that we can now replace $\frac{d}{dx}(\log_e y) = \frac{d}{dx}(x\log_e a)$, with $\frac{1}{y}\cdot\frac{dy}{dx} = \frac{d}{dx}(x\log_e a)$.

$\frac{d}{dx}(x\log_e a) = \log_e a \therefore \frac{1}{y}\cdot\frac{dy}{dx} = \log_e a \Leftrightarrow \frac{dy}{dx} = (\log_e a)y$

$= (\log_e a)\times a^x$

i.e. if $y = a^x$ then $\frac{dy}{dx} = (\log_e a)\times a^x$

Example 6.2.11

Differentiate the following:

a $y = 5\times 2^x$ b $y = 3^{4x}$ c $y = 5^{2x+1}$

a Based on our result, we have that:

$$\frac{dy}{dx} = 5\times(\log_e 2)\times 2^x = (\log_e 32)\times 2^x.$$

b Using the result that:

if $y = a^{kx}$ then $\frac{dy}{dx} = k(\log_e a)\times a^{kx}$, we have that

$$\frac{dy}{dx} = 4(\log_e 3)\times 3^{4x} = (\log_e 81)\times 3^{4x}$$

c Letting $u = 2x+1$ gives $y = 5^{2x+1}$ as $y = 5^u$.

Using the chain rule we have:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = (\log_e 5)5^u\times 2$$

$$\therefore \frac{dy}{dx} = (2\log_e 5)\times 5^{2x+1}$$

$$= (\log_e 25)\times 5^{2x+1}$$

Note that from $(2\log_e 5)\times 5^{2x+1}$ a number of different acceptable answers could have been given. For example, $(2\log_e 5)\times 5^{2x+1} = (2\log_e 5)\times 5^{2x}\times 5 = 10(\log_e 5)\times 5^{2x}$.

We must not forget that we could have determined the derivative of $f(x) = a^x$ by using a first principles approach.

That is, $f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{a^{x+h}-a^x}{h}$.

As x is independent of the limit statement, we have:

$$f'(x) = \lim_{h\to 0}\frac{a^x(a^h-1)}{h} = a^x\times\lim_{h\to 0}\frac{a^h-1}{h}.$$

All that remains then is to determine $\lim_{h\to 0}\frac{a^h-1}{h}$.

We leave this as an exercise for you. However, a starting point is to use a numerical approach, i.e. try different values of a (say $a = 2$, $a = 10$) and tabulate your result for a range of (small) values of h (i.e. make h smaller and smaller). Then compare your numerical values to that of $\log_e 2$ for $a = 2$ and $\log_2 10$ for $a = 10$ and so on.

Differentiating $y = \log_a x$

As in the last section, we use a simple algebraic manipulation to convert an expression for which we do not have a standard result (yet!) into one we have met before. In this case we make use of the change of base result.

i.e. given $\log_a x = \frac{\log_e x}{\log_e a}$ the equation $y = \log_a x$ can then be written as $y = \frac{1}{\log_e a}\times\log_e x$.

Now, $\frac{1}{\log_e a}$ is a real constant, and so, we are in fact differentiating an expression of the form $y = k\times\log_e x$,

where $k = \frac{1}{\log_e a}$.

However if $y = k\times\log_e x$ then $\frac{dy}{dx} = k\times\frac{1}{x}$, meaning that we then have:

If $y = \log_a x$ then $\frac{dy}{dx} = \frac{1}{\log_e a}\times\frac{1}{x}$

Example 6.2.12

Find the derivative of:

a $\log_2 x$ b $\log_{10}(2x-1)$ c $y = \log_4 \tan 8x$

a Given that if $y = \log_a x$ then $\frac{dy}{dx} = \frac{1}{\log_e a}\times\frac{1}{x}$ then for $y = \log_2 x$ i.e. $a = 2$ we have that

$$\frac{dy}{dx} = \frac{1}{\log_e 2}\times\frac{1}{x} = \frac{1}{(\log_e 2)x}.$$

b This time we start by letting $u = 2x-1$ so that $y = \log_{10}(2x-1) = \log_{10}u$.

Then, combining the chain rule with the results above (i.e. $a = 10$) we have:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \left(\frac{1}{\log_e 10}\times\frac{1}{u}\right)\times 2$$

$= \dfrac{2}{(\log_e 10)(2x-1)}$

c Again we combine the chain rule with the results of this section, where in this case, $a = 4$.

Let $u = \tan 8x \Rightarrow \dfrac{du}{dx} = 8\sec^2 8x$, then

$y = \log_4 u \Rightarrow \dfrac{dy}{du} = \dfrac{1}{\ln 4} \times \dfrac{1}{u}$.

Therefore, $\dfrac{dy}{dx} = \dfrac{dy}{du}\dfrac{du}{dx} = \left(\dfrac{1}{\ln 4} \times \dfrac{1}{u}\right) \times 8\sec^2 8x = \dfrac{8\sec^2 8x}{(\ln 4)\tan 8x}$

$= \dfrac{8}{(\ln 4)\cos 8x \sin 8x}$

$\left[\text{using } \cos 8x \sin 8x = \frac{1}{2}\sin 16x\right] \quad = \dfrac{8}{(\ln 2)\sin 16x}$

Exercise 6.2.5

1. Differentiate the following.

a $y = 4^x$ b $y = 3^x$

c $y = 8^x$ d $y = 3 \times 5^x$

e $y = 7 \times 6^x$ f $y = 2 \times 10^x$

g $y = 6^{x-2}$ h $y = 2^{3x+1}$

i $y = 5 \times 7^{3-x}$

2. Differentiate the following.

a $y = x \times 3^x$ b $y = 2^{x+1}\sin 2x$

c $y = 5^x e^{-x}$ d $y = \dfrac{x^2}{8^x}$

e $y = \dfrac{x+2}{1+4^x}$ f $y = \dfrac{\cos x}{5^{-x}}$

3. Differentiate the following.

a $y = \log_5 x$ b $y = \log_{10}(5x)$

c $y = \log_4(2x)$ d $y = \log_9(x+1)$

e $y = \log_2(x^2+1)$ f $y = \log_5\sqrt{x-5}$

g $y = x\log_2 x$ h $y = 3^x\log_3 x$

i $y = a^x\log_a x$ j $y = \dfrac{a^x}{\log_a x}$

k $y = \dfrac{(x+1)}{\log_{10}(x+1)}$ l $y = \dfrac{x}{\log_2\sqrt{x}}$

4. Find the value(s) of x where the gradient of $f(x) = \dfrac{4^x}{x^2}$ is zero.

5. For what value(s) of x will a slope of $f(x) = x^2 \times 2^x$ be zero?

6. Given that $g(x) = x\left(\dfrac{1}{3}\right)^x$, find the exact value of $g'(1)$

7. Find $h'\left(\dfrac{\pi}{6}\right)$ where $h(x) = \pi^{\sin x} + \sin^{\pi} x$.

8. Find where $\dfrac{dy}{dx} = 0$ given that $y = 10^x\log_{10}x - 10x$.

9. Find the gradient of the following curves at the point indicated.

a $y = x^2 10^x$ at $x = 1$

b $y = \sin(4^x)$ at $x = 0$

c $y = x^e\log_{e^2}x$ at $y = 0$

d $y = \dfrac{10x}{\log_{10}x}$ at $x = 10$.

10. Differentiate the following.

a 5^{4x+1} b 3^{x-x^3}

c 10^{2x-3} d $9^{\sqrt{x}-x}$

e $4^{\frac{1}{2}\cos 2x}$ f $4^{\sqrt{\cos 2x}}$

Extra questions

Higher Derivatives

Since the derivative of a function f is another function, f', then it may well be that this derived function can itself be differentiated. If this is done, we obtain the second derivative of f which is denoted by f'' and read as "f–double–dash".

The following notation for $y = f(x)$ is used:

First derivative $\dfrac{dy}{dx} = f'(x)$ $[= y']$

Second derivative $\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right) = \dfrac{d^2y}{dx^2} = f''(x)$ $[= y'']$

So, for example, if $f(x) = x^3 - 5x^2 + 10$

then $f'(x) = 3x^2 - 10x$ and $f''(x) = 6x - 10$.

The expression $\frac{d^2y}{dx^2}$ is read as "dee–two–y by dee–x–squared" and the expression y'' is read as "y–double–dash".

Example 6.2.13

Find the second derivative of:

a $x^4 - \sin 2x$ b $\ln(x^2+1)$ c $x\mathrm{Sin}^{-1}x$

a Let $y = x^4 - \sin 2x$ then $y' = 4x^3 - 2\cos 2x$ and $y'' = 12x^2 + 4\sin 2x$.

b Let $f(x) = \ln(x^2+1)$ then $f'(x) = \frac{2x}{x^2+1}$ and

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

c Let $y = x\mathrm{Sin}^{-1}x$ then

$$\frac{dy}{dx} = x \times \frac{1}{\sqrt{1-x^2}} + (1) \times \mathrm{Sin}^{-1}x = \frac{x}{\sqrt{1-x^2}} + \mathrm{Sin}^{-1}x$$

$$\frac{d^2y}{dx^2} = \frac{(1)\times\sqrt{1-x^2} - x \times \frac{1}{2}(-2x)\cdot\frac{1}{\sqrt{1-x^2}}}{(\sqrt{1-x^2})^2} + \frac{1}{\sqrt{1-x^2}}$$

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

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

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

$$= \frac{1}{(1-x^2)\sqrt{1-x^2}} + \frac{1}{\sqrt{1-x^2}}$$

$$= \frac{1}{(1-x^2)\sqrt{1-x^2}} + \frac{1-x^2}{(1-x^2)\sqrt{1-x^2}}$$

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

As we can see from Example 6.2.13c, some second derivatives require the use of algebra to obtain a simplified answer.

Note then that, just as we can find the second derivative, so too can we determine the third derivative and the fourth derivative and so on (of course, assuming that these derivatives exist). We keep differentiating the results. The notation then is extended as follows:

Third derivative is $f'''(x)$ ("f–triple–dash") and so on where the nth derivative is $f^{(n)}(x)$ or $\frac{d^ny}{dx^n}$.

Exercise 6.2.6

1. Find the second derivative of the following functions.

   a $f(x) = x^5$

   b $y = (1+2x)^4$

   c $f:x \mapsto \frac{1}{x}$ where $x \in \mathbb{R}$

   d $f(x) = \frac{1}{1+x}$

   e $y = (x-7)(x+1)$

   f $f:x \mapsto \frac{x+1}{x-2}$ where $x \in \mathbb{R}\backslash\{2\}$

   g $f(x) = \frac{1}{x^6}$

   h $y = (1-2x)^3$

   i $y = \ln x$

   j $f(x) = \ln(1-x^2)$

   k $y = \sin 4\theta$

   l $f(x) = x\sin x$

2. Find the second derivative of the following.

   a $\arctan x$ b $\arcsin x$

   c $\arccos x$ d $x\arctan x$

   e $\arcsin\sqrt{x}$ f $\arccos\left(\frac{1}{\sqrt{x}}\right)$

3. Find the second derivative of the function $f(x) = \frac{\log_e x}{x^2}$

   Find a formula for the second derivative of the function $f(x) = \frac{\log_e x}{x^n}$.

4. Consider the function $f(x) = \frac{1}{x+1}, x \neq -1$.

   Find the first five derivatives by differentiating the function five times. Hypothesise a formula for the nth derivative of this function. Use the method of mathematical induction or other appropriate method to prove that your formula works for all whole numbered values of n.

5. Find a formula for the second derivative of the family of functions $f(x) = \left(\frac{x+1}{x-1}\right)^n$ where n is a real number.

6. Given $y = \frac{1}{1-x}$, prove that $\frac{d^n y}{dx^n} = \frac{n!}{(1-x)^{n+1}}$ for $n \geq 1$.

Extra questions

Implicit Differentiation

Implicit relations

Most of the equations that we have dealt with so far have been expressed in the form $y = f(x)$. For example, $y = \sin(2x)+1$, $y = x^3-2x$, $y = \ln(x-e^x)$, that is, y has been expressed explicitly in terms of x so that for any one given value of x we obtain a unique value of y by substituting the x-value into the given equation.

Expressions such as $x^2y+y-2 = 0$, $\sin(xy) = 1$, $e^{x+y} = x+y$, are called implicit equations because these equations define y implicitly as a function of x. Note then that although $y = x^2$ defines y as an explicit function of x, the equation $y^2+(x+x^2)y+x^3 = 0$ defines y implicitly as functions of x – in fact, we have that two functions are defined implicitly by the equation $y^2+(x+x^2)y+x^3 = 0$ – they are $y = -x$ and $y = -x^2$. We shall see how it is sometimes possible to extract functions from an implicit equation.

It may be possible for an implicit function to be rearranged to form an explicit function. For example, using the equation $x^2y+y-2 = 0$ we have that $(x^2+1)y = 2$ and so, we obtain the equation $y = \frac{2}{x^2+1}$ which defines y explicitly in terms of x.

Using the implicit function $y^2+(x+x^2)y+x^3 = 0$ we have (after expanding and grouping) that

$y^2+(x+x^2)y+x^3 = 0 \Leftrightarrow (y+x^2)(y+x) = 0 \Leftrightarrow y = -x$

or $y = -x^2$. So, we see that in this case two functions are defined implicitly by the equation $y^2+(x+x^2)y+x^3 = 0$.

In fact with more complicated equations it may not be possible to even produce an expression for y, i.e. to solve explicitly for y. Sometimes even simple equations may not define y uniquely as a function of x. For example, if we consider the equation $e^{x+y} = x+y$ we realize that it is not possible to obtain an expression for y explicitly in terms of x. The question then arises, "*How can we differentiate equations such as these?*".

We start by considering the equation $x^2y = 2$. As y is implicitly defined as a function of x, then, one way of finding the derivative of y with respect to x is to first express y explicitly in terms of x:

So, from $x^2y = 2$ we have $y = \frac{2}{x^2} \Rightarrow \frac{dy}{dx} = -\frac{4}{x^3}$.

This method works well, as long as y can be expressed explicitly in terms of x.

Now consider the equation $2x^2+y^3-y = 2$. This time it is not possible to express y explicitly in terms of x and so we use a procedure known as implicit differentiation.

The key to understanding how to find $\frac{dy}{dx}$ implicitly is to realise that we are *differentiating with respect to x* – so that terms in the equation that involve 'x's only can be differentiated as usual but terms that involve 'y's must have the chain rule applied to them (and possibly the product rule or quotient rule) because we are assuming that y is a function of x.

Before we deal with the equation $2x^2+y^3-y = 2$ we discuss some further examples.

To differentiate y^3 with respect to x, with the assumption that y is a function of x we use the chain rule as follows: $\frac{d}{dx}(y^3) = \frac{d}{dy}(y^3)\cdot\frac{dy}{dx} = 3y^2\cdot\frac{dy}{dx}$

To differentiate $\sin y$ with respect to x, with the assumption that y is a function of x we use the chain rule as follows: $\frac{d}{dx}(\sin y) = \frac{d}{dy}(\sin y)\cdot\frac{dy}{dx} = \cos y\cdot\frac{dy}{dx}$

Notice then that to differentiate y^n with respect to x, with the assumption that y is a function of x we have:

$$\frac{d}{dx}(y^n) = \frac{d}{dy}(y^n)\cdot\frac{dy}{dx} = ny^{n-1}\cdot\frac{dy}{dx}$$

To differentiate xy^2 with respect to x, with the assumption that y is a function of x we use the product rule and chain rule as follows:

$$\frac{d}{dx}(xy^2) = \frac{d}{dx}(x)\times y^2 + x\times\frac{d}{dx}(y^2) \text{ (product rule)}$$
$$= 1\times y^2 + x\times\left[\frac{d}{dy}(y^2)\cdot\frac{dy}{dx}\right] \text{ (chain rule for } y^2\text{)}$$
$$= y^2 + x\left[2y\cdot\frac{dy}{dx}\right].$$

And so we have that $\frac{d}{dx}(xy^2) = y^2 + 2xy \cdot \frac{dy}{dx}$.

Now let us return to the equation $2x^2 + y^3 - y = 2$ and find the gradient of the curve at the point (1, 1).

We start by *differentiating both sides of the equation with respect to x*:

i.e. $\frac{d}{dx}(2x^2 + y^3 - y) = \frac{d}{dx}(2)$

Then, we differentiate each term in the expression with respect to x:

$$\frac{d}{dx}(2x^2) + \frac{d}{dx}(y^3) - \frac{d}{dx}(y) = 0$$

Use the chain rule $\quad 4x + \frac{d}{dy}(y^3) \cdot \frac{dy}{dx} - \frac{dy}{dx} = 0$

$$4x + (3y^2) \cdot \frac{dy}{dx} - \frac{dy}{dx} = 0$$

Then we group the $\frac{dy}{dx}$ terms and factorise:

$$4x + (3y^2 - 1)\frac{dy}{dx} = 0$$

Then, we solve for $\frac{dy}{dx}$: $\quad \frac{dy}{dx} = -\frac{4x}{3y^2 - 1}$.

The first thing we notice is that the derivative involves both x and y terms! Now sometimes it is possible to simplify so that there are only x terms in the expression and sometimes it can only be left as is. In this case it will be left in terms of x and y.

Then, to find the gradient of the curve at the point (1, 1) we substitute the values $x = 1$ and $y = 1$ into the equation of the derivative: $\frac{dy}{dx} = -\frac{4}{3-1} = -2$.

Example 6.2.14

Find the derivatives of these relations.

a $\quad 2x^2 + xy = 5 \qquad$ b $\quad \frac{y}{x} + 3y^2 = 2x^3$

c $\quad x\mathrm{Sin}^{-1}y = e^{2y}$

a $\quad$ Differentiating both sides with respect to x *(which can be abbreviated to diff. b.s.w.r.t x)*:

$$\frac{d}{dx}(2x^2 + xy) = \frac{d}{dx}(5) \quad \therefore \frac{d}{dx}(2x^2) + \frac{d}{dx}(xy) = 0$$

$$4x + \left[\frac{d}{dx}(x) \times y + x \times \frac{d}{dx}(y)\right] = 0 \quad \text{(Using product rule)}$$

$$\therefore 4x + \left[1 \times y + x\frac{dy}{dx}\right] = 0$$

$$\Leftrightarrow x\frac{dy}{dx} = -4x - y$$

$$\Leftrightarrow \frac{dy}{dx} = -\frac{(4x+y)}{x}$$

b $\quad$ Here, the first term must be differentiated using the quotient rule. We consider this term on its own first. Its derivative with respect to x is:

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

Then, *diff b.s.w.r.t.x* we have:

$$\frac{d}{dx}\left(\frac{y}{x} + 3y^2\right) = \frac{d}{dx}(2x^3) \Rightarrow \frac{d}{dx}\left(\frac{y}{x}\right) + \frac{d}{dx}(3y^2) = 6x^2$$

$$\therefore \frac{x \times \frac{dy}{dx} - y}{x^2} + 6y \times \frac{dy}{dx} = 6x^2$$

$$\therefore x \times \frac{dy}{dx} - y + 6x^2y \times \frac{dy}{dx} = 6x^4 \text{ (multiplying through by } x^2\text{)}$$

$$\frac{dy}{dx}(x + 6x^2y) = 6x^4 + y \text{ (grouping the } \frac{dy}{dx} \text{ terms)}$$

$$\therefore \frac{dy}{dx} = \frac{6x^4 + y}{x + 6x^2y}$$

c $\quad$ *diff b.s.w.r.t.x*: $\quad \frac{d}{dx}(x\mathrm{Sin}^{-1}y) = \frac{d}{dx}(e^{2y})$

(Using product rule for L.H.S and chain rule for R.H.S)

$$\frac{d}{dx}(x) \times \mathrm{Sin}^{-1}y + x \times \frac{d}{dx}(\mathrm{Sin}^{-1}y) = \frac{d}{dy}(e^{2y})\frac{dy}{dx}$$

$$1 \times \mathrm{Sin}^{-1}y + x\frac{d}{dy}(\mathrm{Sin}^{-1}y)\frac{dy}{dx} = 2e^{2y}\frac{dy}{dx} \text{ (chain rule)}$$

$$\mathrm{Sin}^{-1}y + \frac{x}{\sqrt{1-y^2}}\frac{dy}{dx} = 2e^{2y}\frac{dy}{dx}$$

$$\therefore \mathrm{Sin}^{-1}y = \left(2e^{2y} - \frac{x}{\sqrt{1-y^2}}\right)\frac{dy}{dx} \text{ (grouping the } \frac{dy}{dx} \text{ terms)}$$

$$\frac{dy}{dx} = \frac{\sqrt{1-y^2}(\mathrm{Sin}^{-1}y)}{2e^{2y}\sqrt{1-y^2} - x}$$

Exercise 6.2.7

1. Find the first derivative, $\frac{dy}{dx}$, of the following relations in which y depends on x.

a $\quad 2 + x^2 + y = 6 \qquad$ b $\quad -3 + x^2 + y^2 = 5$

c $\quad \frac{1}{x^2} + y^2 = 14 \qquad$ d $\quad y + xy = 9$

e $4 = y + y \times e^x$ f $\cos x + xy = 12$

g $x + \ln(y) = 8$ h $\frac{1}{x} + x^3 y = -11$

i $2x + y\sin x = 5$ j $(x+y)^2 = 12$

k $x^4 = y + y^3$ l $2\sqrt{x+y} = x$

2. $f(x)$ is a relation on a real variable x such that $e^{f(x)} - f(x) = e^5 - 5$. Find the coordinates of the point for which $x = 1$ and the gradient of the graph of the function at this point.

3. A curve has equation $\frac{e^{x^2y}}{x} + 2x = 3$.

Differentiate the equation implicitly and hence prove that:

$$\frac{e^{x^2y}}{x} = 2x + e^{x^2y}\left(2xy + x^2\frac{dy}{dx}\right).$$

4. Use implicit differentiation to find the coordinates of the points on the circle $x^2 - 3x + y^2 - 4y = 7$ for which the gradient is 2.

5. Consider the conic section with equation:

$$x^2 + xy - y^2 = 20.$$

a Make y the subject of the equation.

b Prove that the domain of the relation is $]-\infty, 4] \cup [4, \infty[$.

c Find an expression for $\frac{dy}{dx}$.

d Use a to eliminate y from your expression for $\frac{dy}{dx}$.

e Hence prove that as $x \to \pm\infty$, $\frac{dy}{dx} \to \frac{5 \pm \sqrt{5}}{2\sqrt{5}}$.

f What type of curve is represented by $x^2 + xy - y^2 = 20$?

6. A curve has equation $x^4 + y^4 = 16$.

a Find the domain and range of the relation.

b Express the gradient, $\frac{dy}{dx}$, in terms of x and y.

c Eliminate y from your expression in part b.

d What is the gradient in the region of the y-axis?

Consider the family of relations $x^{2n} + y^{2n} = k^{2n}$ where k is a constant and n is a positive integer.

e Find the domain and range of the relation.

f Express the gradient, $\frac{dy}{dx}$, in terms of x and y and hence describe the form of the graph of the relation as n becomes large.

7.

a If $pv^{\Upsilon} = c$ where c and Υ are real constants, find $\frac{dv}{dp}$.

b Find $\frac{dy}{dx}$ if $\frac{x^m}{y^n} = \frac{m}{n}xy$.

8. Find the slope of the curve

a $x^3 + y^3 - x^2y = 7$ at $(1, 2)$

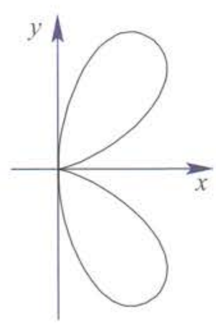
b $x^3 + y^3 - 3kxy = 0$ at $\left(\frac{3}{2}k, \frac{3}{2}k\right)$.

9. Find $\frac{dy}{dx}$ if:

a $\log_e(xy) = y, x > 0$

b $x\mathrm{Tan}^{-1}(y) = x + y$.

10. The graph of the curve $(x^2 + y^2)^2 = 4xy^2$ is shown alongside.



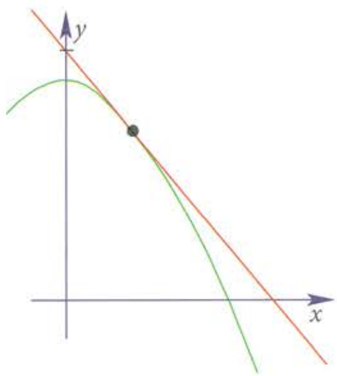
a Find the gradient of the curve at the point where $x = 1$. Explain your result.

b Find the gradients of the curve where $y = ½$, giving your answers to 2 decimal places.

Answers

6.3 Applications of Differentiation

Equation of Tangent



The gradient of a curve $y = f(x)$ at any point (x_1, y_1) is equal to the gradient of the tangent to the curve at that point.

This allows us to find the equation of the straight line with this gradient, through the given point - as in this example:

Example 6.3.1

Find the equation of the tangent to the curve $y = 5 - x^2$ at the point (1, 4).

1. Given that $y = 5 - x^2 \Rightarrow \frac{dy}{dx} = -2x$.
2. Then, for $x = 1$, we have $\frac{dy}{dx} = -2(1) = -2$.

Therefore, using $y - y_1 = m(x - x_1)$, with $m = -2$ and $(x_1, y_1) \equiv (1, 4)$, we have the equation of the tangent given by:

$y - 4 = (-2)(x - 1) \Leftrightarrow y - 4 = -2x + 2$

That is, $y = -2x + 6$

Example 6.3.2

Find the equation of the tangent to the curve $y = x^3 - 8$ where $x = 2$.

Given that $y = x^3 - 8 \Rightarrow y' = 3x^2$. Then, for $x = 2$, $y' = 3 \times 2^2 = 12$, i.e. $m = 12$.

In order to use the equation $y - y_1 = m(x - x_1)$ we need both x- and y-values. As we are only given the x-value, we now determine the corresponding y-value, i.e. $x = 2 \Rightarrow y = 2^3 - 8 = 0$.

With $(x_1, y_1) \equiv (2, 0)$ the equation of the tangent is: $(y - 0) = 12(x - 2) \Leftrightarrow y = 12x - 24$.

Equation of Normal

To find the equation of the normal at the point (x_1, y_1) we first need to determine the gradient of the tangent, m_t, and then use the relationship between the gradients of two perpendicular lines (given that the normal is perpendicular to the tangent).

To find the equation of the normal we need to repeat the gradient calculation but with a gradient determined by the condition that the product of the gradients of perpendicular lines is –1:

$$m \times m' = -1$$

Example 6.3.3

Find the equation of the normal to the curve $y = 2x^3 - x^2 + 1$ at the point $(1, 2)$.

First determine the gradient of the tangent: $\frac{dy}{dx} = 6x^2 - 2x$.

At $x = 1$, we have $\frac{dy}{dx} = 6(1)^2 - 2(1) = 4$. That is, $m_t = 4$.

We can now determine the gradient of the normal:

using $m_N = -\frac{1}{m_t}$ we have $m_N = -\frac{1}{4}$.

Using the equation of a straight line, $y - y_1 = m(x - x_1)$ where $(x_1, y_1) \equiv (1, 2)$ and $m = -\frac{1}{4}$

we have that $y - 2 = -\frac{1}{4}(x-1) \Leftrightarrow 4y - 8 = -x + 1$

Hence the equation of the normal is given by $4y + x = 9$.

Example 6.3.4

Determine the equation of the normal to the curve $y = xe^{-2x} + 2$ at the point where the curve crosses the y-axis.

We first need to determine the y-intercept:
$x = 0 \Rightarrow y = 0 \times e^0 + 2 = 2$.

That is, the curve passes through the point (0, 2).

Next, we need to determine the gradient of the tangent where $x = 0$.

From $y = xe^{-2x} + 2 \Rightarrow \frac{dy}{dx} = 1 \times e^{-2x} + x(-2e^{-2x})$ (using the product rule)

$$= e^{-2x} - 2xe^{-2x}$$

Therefore, at $x = 0$, $\frac{dy}{dx} = e^{-2(0)} - 2(0)e^{-2(0)} = 1 - 0 = 1$.

That is, $m_t = 1 \Rightarrow m_N = -\frac{1}{1} = -1$.

Then, using the general equation of a straight line we have the equation of the normal as $y - 2 = -1(x - 0)$, or $y = -x + 2$.

Exercise 6.3.1

1. Find the equations of the tangents to the following curves at the points indicated:

   a $y = x^3 - x^2 - x + 2$ at (2, 4)

   b $y = x^4 - 4x^2 + 3$ at (1, 0)

   c $y = \sqrt{x+1}$ at (3, 2)

   d $y = \frac{1}{\sqrt{x-1}} + \frac{1}{2}$ at (5, 1)

   e $f(x) = \frac{x}{x+1}, x \neq -1$ at $\left(1, \frac{1}{2}\right)$

   f $f(x) = \frac{2x}{x+2}, x \neq -2$ at (2, 1)

   g $x \mapsto x(x^3 - 4)$ at (2, 8)

   h $x \mapsto \frac{x^2}{x-1}, x \neq 1$ at (2, 4)

2. Find the equation of the normal for each of the curves in Question 1.

3. Find the equations of the tangents to the following curves at the points indicated:

   a $y = xe^x$ at $(1, e)$

   b $x \mapsto \frac{e^x}{x}, x \neq 0$ at $(1, e)$

   c $f(x) = x + \sin(x)$ at (π, π)

   d $y = x\cos(x)$ at $(\pi, -\pi)$

   e $y = \frac{x}{\sin(x)}$ at $\left(\frac{\pi}{2}, \frac{\pi}{2}\right)$

   f $x \mapsto x\log_e(x+1)$ at $(e - 1, e - 1)$

   g $x \mapsto xe^{x^2+1}$ at (0, 0)

   h $f(x) = \sin(2x) + \cos(x)$ at (0, 1)

4. Find the equation of the normal for each of the curves in Question 3.

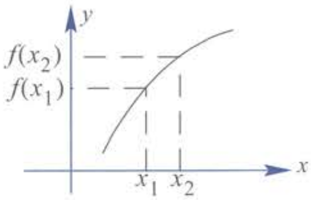
5. Find the equation of the tangent to the curve $y = x^2(x^2 - 1)$ at the point A(2, 12).

   The tangent at a second point, B(–2,12), intersects the tangent at A at the point C. Determine the type of triangle enclosed by the points A, B and C. Show that the tangents drawn at the points X and Y, where $x = a$ and $x = -a$ respectively will always meet at a third point Z which will lie on the y-axis.

Extra questions

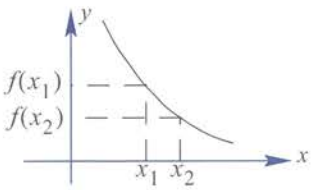
Increasing and Decreasing Functions

A function f is said to be increasing if its graph rises as it is sketched from left to right.



That is, if :

$x_2 > x_1 \Rightarrow f(x_2) > f(x_1)$ (i.e. the y-values increase as the x-values increase).



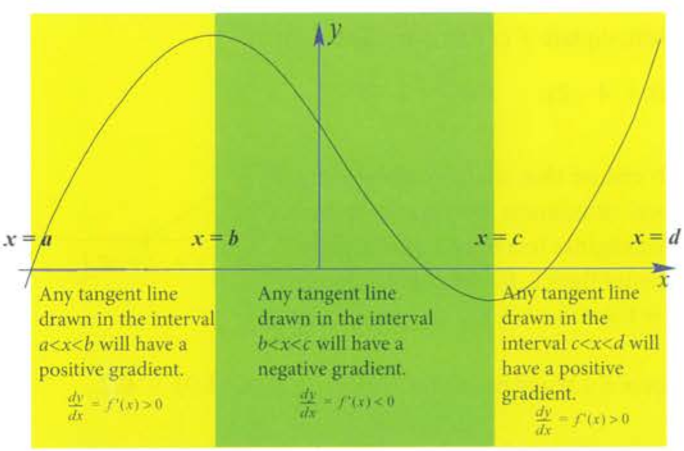
Similarly,

A function f is said to be decreasing if its graph falls as it is sketched from left to right.

That is, if $x_2 > x_1 \Rightarrow f(x_2) < f(x_1)$ (i.e. the y-values decrease as the x-values increase).

... A calculus point of view

The derivative can be used to determine whether a function is increasing or decreasing and so it can be used to help find those values of x for which the function is increasing or decreasing.



This means that, to determine where a function is increasing or decreasing, the values of x for which $f'(x) > 0$ and $f'(x) < 0$ respectively need to be found.

Example 6.3.5

Find the values of x for which the function $f(x) = 1 + 4x - x^2$ is increasing.

By definition, a function is increasing for those values of x for which $f'(x) > 0$.

Therefore find: **1.** $f'(x)$

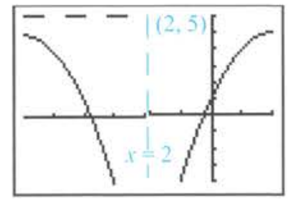
2. the values of x such that $f'(x) > 0$

Now, $f(x) = 1 + 4x - x^2 \Rightarrow f'(x) = 4 - 2x$

Then, $f'(x) > 0 \Leftrightarrow 4 - 2x > 0$

$\Leftrightarrow 4 > 2x$

$\Leftrightarrow x < 2$



We could also have determined this by sketching the graph of $f(x) = 1 + 4x - x^2$. The turning point can be determined by completing the square, i.e. $f(x) = -(x-2)^2 + 5$ giving the axis of symmetry as $x = 2$.

Example 6.3.6

Find the values of x for which the function $f(x) = x\log_e x, x > 0$ is increasing.

Unless you already know what this function looks like, it is difficult to determine the interval for which the function is increasing without using calculus.

First we differentiate (using the product rule):

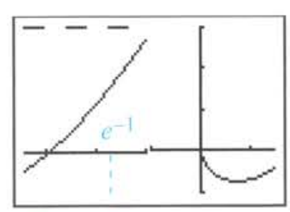
$$f'(x) = 1 \times \log_e x + x \times \frac{1}{x} = \log_e x + 1$$

Now, $f(x)$ is increasing for values of x for which $f'(x) > 0$.

Therefore we need to solve $f(x) = x\log_e x, x > 0$: $\log_e x + 1 > 0$.

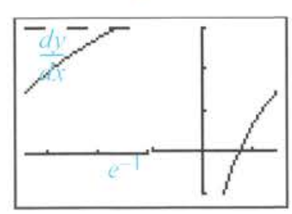
$\log_e x + 1 > 0 \Leftrightarrow \log_e x > -1$

$\Leftrightarrow x > e^{-1}$



(The inequality can be determined by making use of a sketch of $f(x) = x\log_e x, x > 0$.)

$f'(x) = \log_e x + 1$:



That is, $f(x) = x\log_e x, x > 0$ increases for values of x such that $x > e^{-1}$.

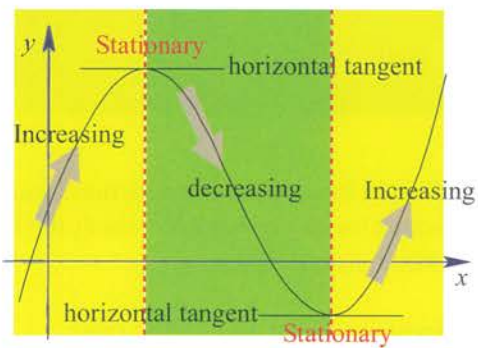
Note that we could have used the graph of the derivative function, $y = f'(x)$, and from it, determined those values of x for which the graph is above the x-axis.

Stationary Points

So far we have discussed the conditions for a function to be increasing $(f'(x) > 0)$ and for a function to be decreasing $(f'(x) < 0)$. What happens at the point where a function changes from an increasing state $(f'(x) > 0)$ to $f'(x) = 0$ and then to a decreasing state $(f'(x) < 0)$ or vice versa?

Points where this happens are known as stationary points. At the point where the function is in a state where it is neither increasing nor decreasing, we have that $f'(x) = 0$. There are times when we can call these stationary points stationary points, but on such occasions, we prefer the terms local maximum and local minimum points.

At the point(s) where $\frac{dy}{dx} = f'(x) = 0$ we have a stationary point.



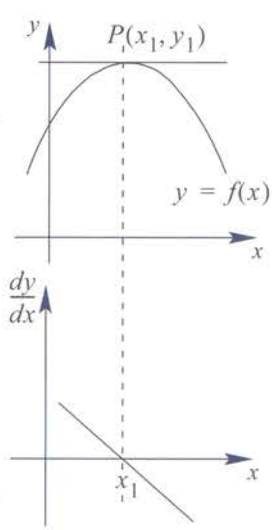
There are three types of stationary points,

namely; local maximum point,

local minimum point and

stationary point of inflection.

1. Local maximum



When sketching a curve, if the following properties hold:

i. At $P(x_1, y_1)$,

$\frac{dy}{dx} = f(x) = 0$ that is $f'(x_1) = 0$.

ii. For $x < x_1$ then $\frac{dy}{dx} > 0$

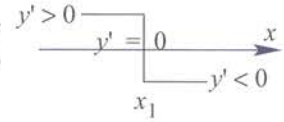
$x > x_1$ then $\frac{dy}{dx} < 0$

where the two chosen values of x are such that one is just slightly less than x_1 and the other is just slightly greater than x_1, then, $y = f(x)$ has a local maximum point (also known as a relative maximum) at the point $P(x_1, y_1)$.

iii. Graph of the gradient function:

Notice that the values of $\frac{dy}{dx}$ are changing from positive to negative. Sometimes this is referred to as the sign of the first derivative. At this stage, it isn't so much the magnitude of the derivative that is important, but that there is a change in the sign of the derivative near $x = x_1$.

In this instance the sign of the derivative changes from positive to negative.



This change in sign is sometimes represented via the diagram above, which is referred to as a sign diagram of the first derivative. Such diagrams are used to confirm the nature of stationary points (in this case, that a local maximum occurs at $x = x_1$).

Example 6.3.7

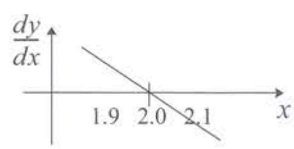
Find the local maximum value of the function whose equation is $f(x) = -3 + 4x - x^2$.

First we differentiate: $f(x) = -3 + 4x - x^2 \Rightarrow f'(x) = 4 - 2x$

Next, equate $f'(x)$ to 0 and solve for x:

$0 = 4 - 2x \quad \Leftrightarrow x = 2$

To ensure that we have obtained a local maximum we choose values of x slightly less than 2 and slightly greater than 2, for example, choose $x = 1.9$ and $x = 2.1$.



For $x = 1.9$, we have that $f'(1.9) = 4 - 2(1.9) = 0.2$.

For $x = 2.1$, we have that $f'(2.1) = 4 - 2(2.1) = -0.2$.

Using the graph of the gradient function, $\frac{dy}{dx}$, confirms that there is a local maximum at $x = 2.0$.

The local maximum value of $f(x)$, is found by substituting $x = 2$ into the given equation: $f(2) = -3 + 4(2) - (2)^2 = 1$. That is, the local maximum occurs at the point (2, 1).

2. Local minimum

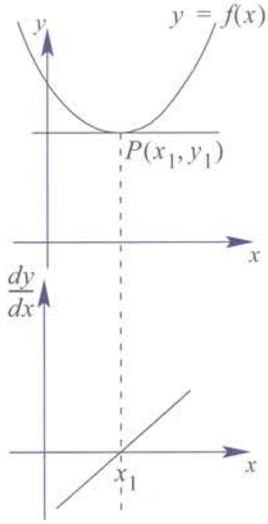
When sketching a curve, if the following properties hold:

i. At $P(x_1, y_1)$,

$\frac{dy}{dx} = f'(x) = 0$ that is $f'(x_1) = 0$.

ii. For $x > x_1$ then $\frac{dy}{dx} > 0$

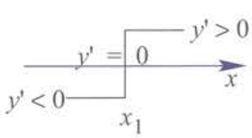
$x < x_1$ then $\frac{dy}{dx} < 0$



where the two chosen values of x are such that one is just slightly greater than x_1 and the other is just slightly less than x_1, then $y = f(x)$ has a local minimum point (also known as a relative minimum) at the point $P(x_1, y_1)$.

iii. Graph of the gradient function:

Notice that the values of $\frac{dy}{dx}$ are changing from negative to positive. Sometimes this is referred to as the sign of the first derivative.



Again we can represent the change in the sign of the first derivative via the diagram alongside, which is referred to as a sign diagram of the first derivative. Such diagrams are used to confirm the nature of stationary points (in this case, that a local minimum occurs at $x = x_1$).

Example 6.3.8

Find the minimum value of $x \mapsto \frac{e^x}{x}, x > 0$

First differentiate (using the quotient rule):

$$\frac{d}{dx}\left(\frac{e^x}{x}\right) = \frac{\frac{d}{dx}(e^x) \times x - e^x \times \frac{d}{dx}(x)}{x^2} = \frac{e^x x - e^x}{x^2} = \frac{e^x(x-1)}{x^2}$$

We solve for $\frac{d}{dx}\left(\frac{e^x}{x}\right) = 0$, i.e. $\frac{e^x(x-1)}{x^2} = 0 \Leftrightarrow e^x(x-1) = 0$

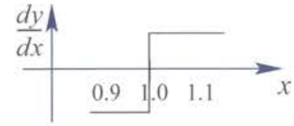
However, $e^x \neq 0$ for all real values of x, therefore, the only possible solution occurs if $x = 1$.

To verify that we have a local minimum we select a value of x slightly less than $x = 1$ and one slightly greater than $x = 1$:

For ($x < 1$): choose $x = 0.9$, we have that $\frac{d}{dx}\left(\frac{e^x}{x}\right) = -0.30$.

For ($x > 1$): choose $x = 1.1$, we have that $\frac{d}{dx}\left(\frac{e^x}{x}\right) = 0.25$.

Therefore, for $x = 1$ we have a local minimum point.



Sign diagram of first derivative:

The minimum value is therefore given by:

$y = \frac{e^1}{1} = e$, and occurs at the point $(1, e)$.

Example 6.3.9

Find the local minimum of the function:

$y = \sin(x) + \frac{1}{2}\sin(2x), 0 \le x \le 2\pi$.

We start by differentiating and finding the stationary points (i.e. solving for $\frac{dy}{dx} = 0$):

Now, $y = \sin(x) + \frac{1}{2}\sin(2x) \Rightarrow \frac{dy}{dx} = \cos(x) + \cos(2x)$.

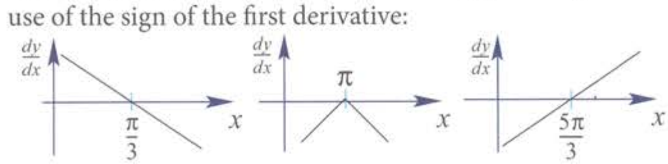
Therefore, solving we have $\cos(x) + \cos(2x) = 0$

$$\cos(x) + (2\cos^2(x) - 1) = 0$$
$$(2\cos(x) - 1)(\cos(x) + 1) = 0$$

Therefore, $\cos(x) = \frac{1}{2}$ or $\cos(x) = -1$, i.e. $x = \frac{\pi}{3}, \frac{5\pi}{3}$ or

$x = \pi$ for $x \in [0, 2\pi]$.

We can check the nature of the stationary points by making use of the sign of the first derivative:



The graph of gradient function (near $x = \frac{5\pi}{3}$) indicates that a local minimum occurs at $x = \frac{5\pi}{3}$. And so, the local minimum value is given by:

$$y = \sin\left(\frac{5\pi}{3}\right) + \frac{1}{2}\sin\left(\frac{10\pi}{3}\right) = -\frac{3\sqrt{3}}{4}.$$

NB: In the process we have come across a new sign diagram (at $x = \pi$). This is dealt with in the next section.

3. Points of inflection

There are two types:

A. Stationary points of inflection

B. Non-stationary points of inflection

A. Stationary point of inflection

The following properties hold at a stationary point of inflection.

i At $P(x_1, y_1)$, $f'(x) = 0$. That is $f'(x_1) = 0$.

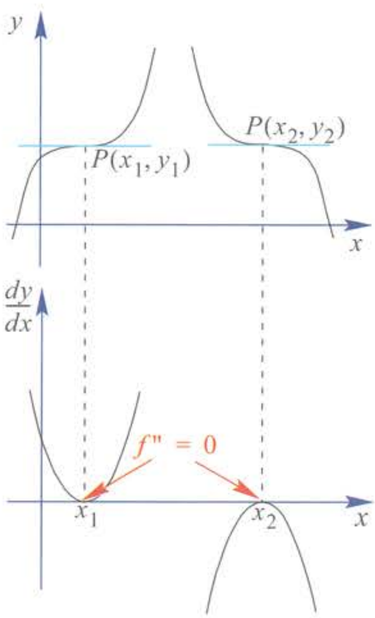
ii For $x < x_1, f'(x) > 0$ and for $x > x_1, f'(x) > 0$.

Similarly,

At $P(x_2, y_2)$, $f'(x) = 0$. That is $f'(x_2) = 0$

and for $x < x_2, f'(x) < 0$ and for $x > x_2, f'(x) < 0$.

That is, the gradient of the curve on either side of x_1 (or x_2) has the same sign.



iii Graph of the gradient function, $y = f'(x)$:

Notice that the values of $f'(x)$ have the same sign on either side of $x = x_1$.

Notice that at $x = x_1$, the gradient of $f'(x)$ is also equal to zero. That is, the derivative of the derivative is equal to zero.

Therefore if there is a stationary point of inflection at $x = x_1$ then $f''(x_1) = 0$.

Example 6.3.10

Find the stationary point of inflection for the graph with equation $y = (x-1)^3(x+2)$.

First we differentiate (using the product rule):

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

$= (x-1)^2[3(x+2) + (x-1)]$

$= (x-1)^2(4x+5)$

Solving for $\frac{dy}{dx} = 0$, we have,

$(x-1)^2(4x+5) = 0 \Leftrightarrow x = 1$ or $x = -\frac{5}{4}(= -1.25)$.

We can now check the sign of the derivative on either side of $x = 1$ and $x = -1.25$

At $x = 1$:

For $x = 0.9$, $\frac{dy}{dx} = (-0.1)^2(8.6) = 0.086 > 0$

For $x = 1.1$, $\frac{dy}{dx} = (0.1)^2(9.4) = 0.094 > 0$.

As the sign of the first derivative is the same on either side of $x = 1$, we have a stationary point of inflection at $x = 1$, i.e. at $(1, 0)$.

A sketch of the graph of $y = (x-1)^3(x+2)$ quickly confirms our result.

For $x = -1.25$, the graph shows a local minimum occurring at this point.

Example 6.3.11

Locate the stationary points of inflection for the curve $f(x) = x^3e^{-x}$.

We begin by determining where stationary points occur:

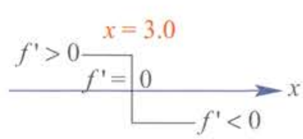
$f(x) = x^3e^{-x} \Rightarrow f'(x) = 3x^2e^{-x} - x^3e^{-x}$

Setting $f'(x) = 0$, we have:

$3x^2e^{-x} - x^3e^{-x} = 0 \Leftrightarrow x^2e^{-x}(3-x) = 0$

$\Leftrightarrow x = 0$ or $x = 3$

We can use the sign of the first derivative to help us determine the nature of the stationary point.



At $x = 3$: Sign diagram:

For $x = 2.9$, $f'(2.9) = (2.9)^2e^{-2.9}(3-2.9) = 0.046 > 0$

For $x = 3.1$, $f'(3.1) = (3.1)^2e^{-3.1}(3-3.1) = -0.043 < 0$

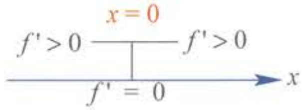
Therefore there exists a local maximum at $x = 3.1$.

At $\boldsymbol{x} = 0$:

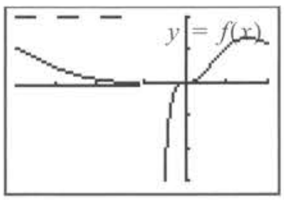
For $x = 0.1$, $f'(0.1) = (0.1)^2e^{-0.1}(3-0.1) = 0.026 > 0$

For $x = -0.1$, $f'(-0.1) = (-0.1)^2e^{0.1}(3+0.1) = 0.034 > 0$

As there is no change in the sign of the first derivative there is a stationary point of inflection at $x = 0$.



Alternatively, we could sketch a graph of the function and use it to help us determine where the stationary point of inflection occurs.



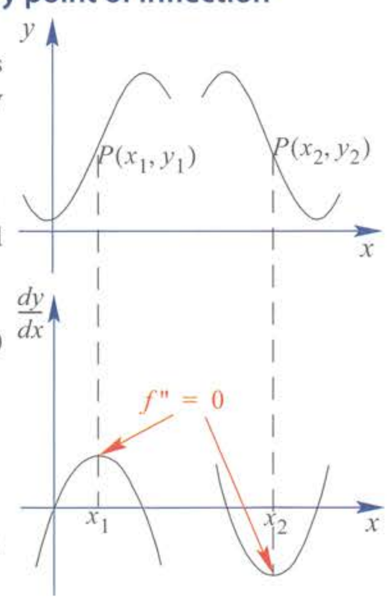
From the graph we can see that there is a local maximum at $x = 3$ and a stationary point of inflection at $x = 0$.

Therefore, the stationary point of inflection occurs at (0, 0).

Notice that the sign diagrams of the first derivative in Examples 6.3.8 and 6.3.11 all look slightly different. We have done this to emphasise that, as long as the diagram provides a clear indication of the sign of the first derivative, then its appearance can vary.

B. Non–stationary point of inflection

The following properties hold at a non-stationary point of inflection:



i. At $P(x_1, y_1)$, $f'(x_1) \neq 0$ and $f''(x_1) = 0$.

ii. For $x < x_1$, $f'(x)>0$ and for $x > x_1$, $f'(x)>0$

Similarly,

At $P(x_2, y_2)$, $f'(x_2) \neq 0$ and $f''(x_2) = 0$.

For $x < x_2, f'(x) < 0$ and for $x > x_2, f'(x) < 0$.

That is, the gradient of the curve on either side of x_1 (or x_2) has the same sign.

iii. Graph of the gradient function, $y = f'(x)$:

Notice that the values of $f'(x)$ have the same sign on either side of $x = x_1$.

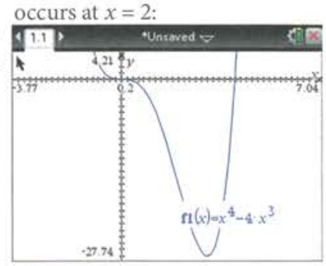
Example 6.3.12

Show that the curve with equation $y = x^4 - 4x^3$ has a non-stationary point of inflection at $x = 2$.

For the curve to have a non-stationary point of inflection at $x = 2$ we need to show that:

1. $\frac{d^2y}{dx^2} = 0$ at $x = 2$ and
2. the sign of the gradient, $\frac{dy}{dx}$, is the same on both sides of $x = 2$ and
3. that $\frac{dy}{dx} \neq 0$ when $x = 2$.

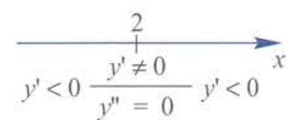
A quick sketch of the function indicates that a point of inflection occurs at $x = 2$:



$$\frac{dy}{dx} = 4x^3 - 12x^2$$

$$\Rightarrow \frac{d^2y}{dx^2} = 12x^2 - 24x$$

Sign diagram:



For $x = 2$, $\frac{d^2y}{dx^2} = 12(2)^2 - 24(2) = 0$ and $\frac{dy}{dx} = -16 \neq 0$.

For $x = 2.1$, $\frac{dy}{dx} = -15.876$ and for $x = 1.9$, $\frac{dy}{dx} = -15.88$.

Therefore there is a non-stationary point of inflection at x=2.

It is important to realize that it is not sufficient to say that "If $f''(x) = 0$ at $x = a$ then there must be a point of inflection at $x = a$."

Does $\boldsymbol{f'(a)} = 0$ imply there is a point of inflection at $\boldsymbol{x = a}$?

The answer is NO!

Although it is necessary for the second derivative to be zero at a point of inflection, the fact that the second derivative is zero at $x = a$ does not mean there must be a point of inflection at $x = a$. That is:

$f''(a) = 0$ is a necessary but not a sufficient reason for there to be an inflection point at $x = a$.

We use the following example to illustrate this.

Consider the case where $f(x) = x^4$.

Now, $f''(x) = 12x^2$, therefore solving for $f''(x) = 0$,

we have $12x^2 = 0 \Leftrightarrow x = 0$.

That is, $f''(0) = 0$. So, do we have a point of inflection at $x = 0$?

A sketch of f shows that although $f''(x) = 0$ at $x = 0$, there is in fact a local minimum and not a point of inflection at $x = 0$.

In other words, finding where $f''(x) = 0$ is not enough to indicate that there is an inflection point. To determine if there is a point of inflection you need to check the sign of the first derivative on either side of the x-value in question.

Example 6.3.13

Find and classify all stationary points (and inflection points) of $f(x) = x^3 - 3x^2 - 9x + 1$.

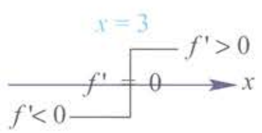
Now, $f(x) = x^3 - 3x^2 - 9x + 1 \Rightarrow f'(x) = 3x^2 - 6x - 9$

Solving for stationary points we have,

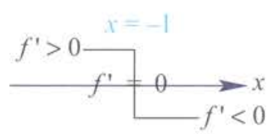
$3x^2 - 6x - 9 = 0 \Leftrightarrow 3(x-3)(x+1) = 0$

$\Leftrightarrow x = 3$ or $x = -1$

So that $f(3) = -26$ and $f(-1) = 6$.



Using the sign of the first derivative, we have:



At $x = 3$:

For $x < 3$ $(x = 2.9)$ $f'(2.9) < 0$ and for $x > 3$ $(x = 3.1)$ $f'(3.1) > 0$

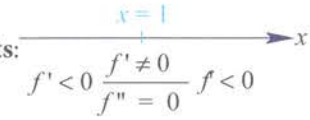
Therefore, there is a local minimum at $(3, -26)$.

At $x = -1$:

For $x < -1$ $(x = -1.1)$ $f'(-1.1) > 0$ and for $x > -1$ $(x = -0.9)$ $f'(-0.9) < 0$.

Therefore, there is a local maximum at $(-1, 6)$.

Checking for inflection points:



We have: $f''(x) = 0 \Leftrightarrow 6x - 6 = 0 \Leftrightarrow x = 1$.

For $x < 1$ $(x = 0.9)$ $f'(0.9) < 0$ and

for $x > 1$ $(x = 1.1)$ $f'(1.1) < 0$.

As the sign of the first derivative remains the same on either side of $x = 1$, there is a point of inflection at $(1, -10)$. Then, as the **first derivative** at $x = 1$ **is not zero**, we have a **non-stationary** point of inflection at $x = 1$.

Curve Sketching

The properties of the basic function are the first things to look at when sketching curves. The main features are summarised below:

- $y = f(x)$ is the equation of the function
- $f(0)$ - intercept(s) for $f(x)$
- $f(x) = 0$ - intercept(s) for $f(x)$
- determine the domain and range for $f(x)$
- identify the asymptotes for $f(x) = \dfrac{p(x)}{q(x)}$
- vertical asymptote(s): the solution(s) for $q(x) = 0$
- horizontal asymptote: if $\deg_{p(x)} = \deg_{q(x)}$ then dividing out the leading (highest power) coefficients between $p(x)$ and $q(x)$, gives the asymptote:
$$y = \frac{\text{leading coeff.}_{p(x)}}{\text{leading coeff.}_{q(x)}}$$
- also, if $\deg_{p(x)} < \deg_{q(x)}$, then the asymptote is:
$$y = \lim_{x \to \pm\infty} f(x)$$
- oblique asymptote: if $\deg_{p(x)} > \deg_{q(x)}$, the quotient $\dfrac{p(x)}{q(x)}$ is the asymptote

First Derivative

The properties of the first derivative also help us sketch curves. The main points are:

- $f'(x)$ suggests the monotonicity (either increasing or decreasing) of $f(x)$ in the interval $S[a,b]$.
  - $f'(x) > 0$: $f(x)$ is increasing in S.
  - $f'(x) < 0$: $f(x)$ is decreasing in S.
- $f'(x)$ detects the existence of inflection point(s) in the interval $S[a,b]$.
  - $f'(k) = 0$ but sign of $f'(x)$ does not change for $x > k$ and $x < k$
- First Derivative Test for Optimum for $k \in S$
- $f(k)$ maximum: $f'(x) > 0$ for $x < k$

  $f'(x) < 0$ for $x > k$

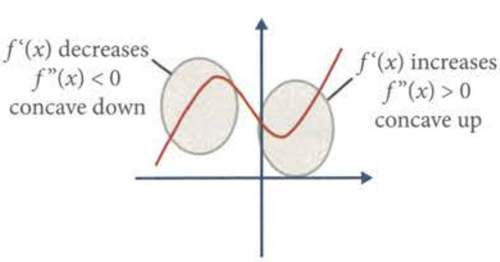
  $f'(k) = 0$
- $f(k)$ minimum: $f'(x) < 0$ for $x < k$

  $f'(x) > 0$ for $x > k$

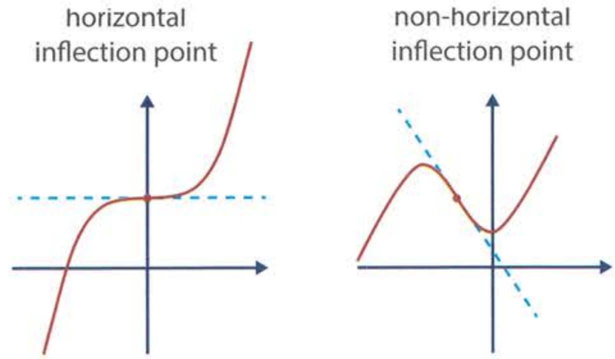
  $f'(k) = 0$

Second Derivative

- $f''(x)$ suggests the concavity of $f(x)$ in the interval $S[a,b]$.
- $f''(x) > 0$: $f(x)$ concaves upward in S.
- $f''(x) < 0$: $f(x)$ concaves downward in S.



- $f''(x)$ determines the nature of an inflection point in the interval in the interval $S[a,b]$.
  - $f''(k)$ changes sign, $f''(k) = 0$ and $f'(k) = 0$, then $f(k)$ is a horizontal inflection point
  - $f''(k)$ changes sign, $f''(k) = 0$ and $f'(k) \neq 0$, then $f(k)$ is a non-horizontal inflection point

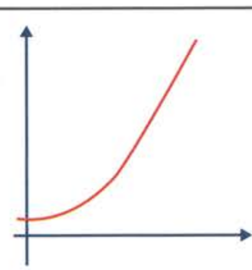
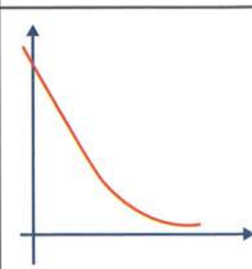
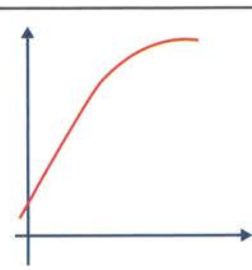
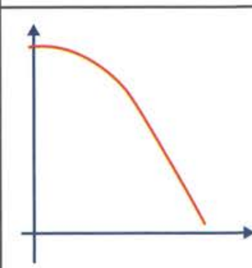


- Second Derivative Test for Optimum for $k \in S$
  - $f(k)_{\max}$ $f''(k) < 0$ and $f'(k) = 0$
  - $f(k)_{\max}$ $f''(k) > 0$ and $f'(k) = 0$

Other Factors

- global minimum is the minimum value of $y = f(x)$ on the entire domain.
- local minimum is the a turning point $y = f(x)$ at $x = k$, for $f'(x)$ changes from negative to positive.
- global maximum is the maximum value of $y = f(x)$ on the entire domain.
- local maximum is the a turning point $y = f(x)$ at $x = k$, for $f'(x)$ changes from positive to negative.

The following table summarises the effects of the first and second derivative on the curvature of a continuous function

	$f'(x) > 0$	$f'(x) < 0$
$f''(x) > 0$		
$f''(x) < 0$		

Exercise 6.3.2

1. Draw a sketch of the graph of the function $f(x), x \in \mathbb{R}$, where:

   a $f(1) = 2, f'(1) = 0, f(3) = -2, f'(3) = 0$, $f'(x) < 0$ for $1 < x < 3$ and $f'(x) > 0$ for $x > 3$ and $x < 1$.

   b $f'(2) = 0, f(2) = 0$, $f'(x) > 0$ for $0 < x < 2$ and $x > 2$, $f'(x) < 0$ for $x < 0$ and $f(0) = -4$.

   c $f(4) = f(0) = 0$, $f'(0) = f'(3) = 0$, $f'(x) > 0$ for $x > 3$ and $f'(x) < 0$ for $x < 0$ and $0 < x < 3$.

   d $f(4) = 4$, $f'(x) > 0$ for $x > 4$, $f'(x) < 0$ for $x < 4$, as $x \to 4^+, f'(x) \to +\infty$ and as $x \to 4^-, f'(x) \to -\infty$.

2. Find the coordinates and nature of the stationary points for the following:

   a $y = 3 + 2x - x^2$

   b $y = x^2 + 9x$

   c $y = x^3 - 27x + 9$

   d $f(x) = x^3 - 6x^2 + 8$

   e $f(x) = 3 + 9x - 3x^2 - x^3$

   f $y = (x-1)(x^2-4)$

   g $f(x) = x - 2\sqrt{x}, x \geq 0$

   h $g(x) = x^4 - 8x^2 + 16$

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

3. Sketch the following functions:

   a $y = 5 - 3x - x^2$

   b $f(x) = x^2 + \frac{1}{2}x + \frac{3}{4}$

   c $f(x) = x^3 + 6x^2 + 9x + 4$

   d $f(x) = x^3 - 4x$

4. Find and describe the nature of all stationary points and points of inflection for the function $f(x) = x^3 + 3x^2 - 9x + 2$.

5. Sketch the graph of $x \mapsto x^4 - 4x^2$.

6. A function f is defined by $f: x \mapsto e^{-x}\sin x$, where $0 \leq x \leq 2\pi$.

   a Find: i $f'(x)$ ii $f''(x)$

   b Find the values of x for which:

   i $f'(x) = 0$ ii $f''(x) = 0$.

   c Using parts a and b, find the points of inflection and stationary points for f.

   d Hence, sketch the graph of f.

7. A function f is defined by $f: x \mapsto e^{x}\sin x$, where $0 \leq x \leq 2\pi$.

   a Find: i $f'(x)$ ii $f''(x)$.

   b Find the values of x for which:

   i $f'(x) = 0$ ii $f''(x) = 0$.

   c Using parts a and b, find the points of inflection and stationary points for f.

   d Hence, sketch the graph of f.

Extra questions

Related Rates

So far we have only dealt with rates of change that involve one independent variable. For example, the volume, V units³, of a sphere of radius r units is given by $V = \frac{4}{3}\pi r^3$. To find the rate of change of the volume with respect to its radius we differentiate with respect to r:

i.e. $\frac{dV}{dr} = \frac{4}{3}\pi \times 3r^2 = 4\pi r^2$.

Now consider this sphere being placed in an acid solution so that it dissolves in such a way that:

1. it maintains its spherical shape, and

2. its radius is decreasing at a rate of 1 cm/hr.

How can we find the rate at which its volume is changing when the sphere's radius is 2 cm?

Note that we are looking for the *rate of change of volume*, that is, we want to find $\frac{dV}{dt}$ (not $\frac{dV}{dr}$ as we found previously – when we specifically requested the rate of change with respect to r). The difference here is that we want the rate of change of one quantity (in this case the volume) which is related to a second variable (in this case the radius r) which is itself changing.

Problems of this type are known as related rates problems and are usually solved by making use of the chain rule.

We now consider the problem at hand. We have:

Want: *rate of change of volume* that is, we want to find $\frac{dV}{dt}$.

When: $r = 2$.

Given: radius is decreasing at a rate of 1 cm/hr, $\frac{dr}{dt} = -1$

Need: This is the tricky bit. Knowing that we will need to use the chain rule, we start by writing down the chain rule with the information we have. Then we try to fill in the missing pieces.

This will often lead to what we need.

Step 1: $\frac{dV}{dt} = \boxed{\frac{\ }{\ }} \times \frac{dr}{dt}$

Step 2: Ask yourself the following question:

"*What do I need in the missing space to complete the chain rule?*"

The missing piece of information in this case is $\frac{dV}{dr}$.

That is, we have $\frac{dV}{dt} = \frac{dV}{dr} \times \frac{dr}{dt}$, which works!

Step 3: Once you have decided on what you need, then find an equation that will enable you to differentiate.

Some warning! Step 3 is the tough bit in the question. Sometimes we are lucky and we know of an equation but sometimes we need to somehow 'create' the equation.

In this case we do have an equation; $V = \frac{4}{3}\pi r^3 \therefore \frac{dV}{dr} = 4\pi r^2$.

And so, using the chain rule we have $\frac{dV}{dt} = 4\pi r^2 \times \frac{dr}{dt}$.

Note: It is **very important** not to substitute any values until the very end.

The last step is to find $\frac{dV}{dt}$ at the specified radius with the given rate, $\frac{dr}{dt} = -1$.

That is, $\frac{dV}{dt} = 4\pi(2)^2 \times -1 = -16\pi$.

So, the volume is decreasing at 16π cm^3/hr.

Example 6.3.14

The radius of a circular oil patch is increasing at a rate of 1.2 cm per minute. Find the rate at which the surface area of the patch is increasing when the radius is 25 cm.

From the data, $\frac{dr}{dt} = 1.2$.

This is the mathematical formulation of the statement 'the radius of a circular oil patch is increasing at a rate of 1.2 cm per minute' where r is the radius and t is the time (in the units given in the question). The radius is increasing and so the rate is positive. The next step is to identify the rate of change that we have been asked to calculate. In this case, the question asks: 'find the rate at which the surface area of the patch is increasing'.

If we define the area as A cm^2, the required rate is $\frac{dA}{dt}$.

So we have: **Want:** $\frac{dA}{dt}$

When: $r = 25$

Given: $\frac{dr}{dt} = 1.2$

Need: (chain rule): $\frac{dA}{dt} = \boxed{\frac{\ }{\ }} \times \frac{dr}{dt}$.

The missing piece must therefore be $\frac{dA}{dr}$!

Therefore, we have, $\frac{dA}{dt} = \frac{dA}{dr} \times \frac{dr}{dt}$.

We need to find an expression for A in terms of r.

This can be done by looking at the geometry of the situation. The oil patch is circular and so the area is given by $A = \pi r^2 \Rightarrow \frac{dA}{dr} = 2\pi r$.

Substituting into the chain rule gives: $\frac{dA}{dt} = 2\pi r \times \frac{dr}{dt}$.

Then, with $r = 25$ and $\frac{dr}{dt} = 1.2$ we have:

$$\frac{dA}{dt} = 2\pi(25) \times 1.2 = 60\pi \approx 188.5 \text{ cm}^2\text{min}^{-1}.$$

That is, the area is increasing at approximately 188.5 cm^{2}min^{-1}.

Note: A useful check that the chain rule has been used appropriately is to use the units of the quantities involved. For Example 6.3.14 we have that:

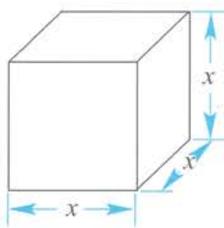
$$\frac{dA}{dt} = \frac{dA}{dr} \times \frac{dr}{dt} = \mathrm{cm^2 cm^{-1}} \times \mathrm{cm^1 min^{-1}} = \mathrm{cm^2 min^{-1}}$$

which is the correct unit for $\frac{dA}{dt}$.

Example 6.3.15

The volume of a cube is increasing at 24 $\mathrm{cm^3 s^{-1}}$. At what rate are the side lengths increasing when the volume is 1,000 $\mathrm{cm^3}$?

We start by determining what variables are involved and see if a diagram might be helpful – usually one is (even if it's only used to visualize the situation). In this case we are talking about a volume and a length, so we let V $\mathrm{cm^3}$ denote the volume of the cube of side length x cm, giving us the expression $V = x^3$.



Next we list all of the information according to our *want*, *when*, *given* and *need*:

Want: $\frac{dx}{dt}$

When: $V = 1{,}000$

Given: $\frac{dV}{dt} = 24$

Need: (chain rule) $\frac{dx}{dt} = \boxed{-} \times \frac{dV}{dt}$ we need $\frac{dx}{dV}$.

So that $\frac{dx}{dt} = \frac{dx}{dV} \times \frac{dV}{dt}$.

However, we have V as a function of x and so it will be easier to first find $\frac{dV}{dx}$ and then use the fact that:

$$\frac{dx}{dV} = \frac{1}{\frac{dV}{dx}}.$$

Then, as $V = x^3 \Rightarrow \frac{dV}{dx} = 3x^2 \Rightarrow \frac{dx}{dt} = \frac{1}{3x^2} \times \frac{dV}{dt}$,

We know $\frac{dV}{dt} = 24$ but, still need a value for x.

From $V = x^3$ we have $1000 = x^3 \therefore x = 10$.

So, $\frac{dx}{dt} = \frac{1}{3(10)^2} \times 24 = \frac{8}{100} = 0.08$.

That is, the side lengths are increasing at 0.08 $\mathrm{cms^{-1}}$.

It is important to realize that when we reach the '**Need:**' stage there are more ways than one to use the chain rule.

For example, with Example 6.3.15, rather than using $\frac{dx}{dt} = \frac{dx}{dV} \times \frac{dV}{dt}$ and then realizing that we need to find $\frac{dV}{dx}$ and then invert it, we could have used the chain rule as follows:

$$\frac{dV}{dt} = \frac{dV}{dx} \times \frac{dx}{dt} \text{ so that } 24 = 3x^2 \times \frac{dx}{dt} \Leftrightarrow \frac{dx}{dt} = \frac{24}{3x^2}.$$

Using the chain rule in this manner has a certain 'logical flow' to it, in that everything seems to 'fit nicely'. But remember, as long as the chain rule expression contains the '*need*', '*want*' and '*given*' it should not make much difference at the end. All that we can say is that as you solve more and more of these problems you will be able to make the 'best' decision available at the time.

Example 6.3.16

A container in the shape of an inverted right circular cone of base radius 10 cm and height 50 cm has water poured into it at a rate of 5 $\mathrm{cm^3 min^{-1}}$. Find the rate at which the level of the water is rising when it reaches a height of 10 cm.

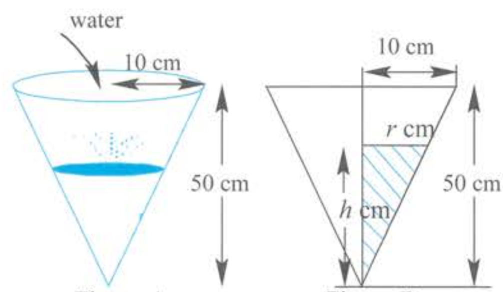


Figure A Figure B

Let the water level at time t min have a height h cm with a corresponding radius r cm and volume V $\mathrm{cm^3}$.

We now list our requirements:

Want: $\frac{dh}{dt}$

When: $h = 10$

Given: $\frac{dV}{dt} = 5$

Need: (chain rule) $\frac{dV}{dt} = \boxed{-} \times \frac{dh}{dt}$ we need $\frac{dV}{dh}$.

Before we can find $\frac{dV}{dh}$ we will need to find an expression for V in terms of h. We do this by making use of Figure B – a cross-section of the inverted cone. The information in Figure B prompts us to make use of similar triangles.

We then have, $\frac{50}{10} = \frac{h}{r} \Leftrightarrow r = \frac{1}{5}h$.

The volume of water in the cone when it reaches a height h cm is given by: $V = \frac{1}{3}\pi r^2 h$.

Then, substituting the expression $r = \frac{1}{5}h$ into the volume equation we have $V = \frac{1}{3}\pi\left(\frac{1}{5}h\right)^2 h = \frac{\pi}{75}h^3 \Rightarrow \frac{dV}{dh} = \frac{\pi}{25}h^2$.

We can now complete the chain rule:

$\frac{dV}{dt} = \frac{dV}{dh} \times \frac{dh}{dt} \therefore \frac{dV}{dt} = \frac{\pi}{25}h^2 \times \frac{dh}{dt}$

$5 = \frac{\pi}{25}h^2 \times \frac{dh}{dt} \quad \Leftrightarrow \frac{dh}{dt} = \frac{125}{\pi h^2}$

Then, when $h = 10$, we have $\frac{dh}{dt} = \frac{125}{100\pi} \approx 0.3978$, i.e. approximately 0.4 cms^{-1}.

Exercise 6.3.3

1. The radius of a circle is increasing at 2 cm/s. Find the rate at which:

   a - its area is increasing and b - its circumference is increasing.

2. The side lengths of a square are increasing at a rate of 3 cm/s. Find the rate at which the area of the square is increasing when its side length is 1 cm.

3. The sides of an equilateral triangle are decreasing at a rate of $\sqrt{6}$ cm/s. Find the rate of change of:

   a - the area of the triangle and b - the altitude of the triangle.

4. A solid 400 g metal cube of side length 10 cm expands uniformly when heated. If the length of its sides expand at 0.5 cm/hr, find the rate at which, after 5 hours:

   a its volume is increasing.

   b its surface area is increasing.

   c its density is changing.

5. A drinking glass is shaped in such a way that the volume of water in the glass when it reaches a height h cm is given by $V = \frac{1}{5}h^3$ cm^3.

   Water is poured into the glass at 2 cm^{3}s^{-1}. At what rate is the water level rising when the depth of water is 3 cm?

6. An ice cube, while retaining its shape, is melting and its side lengths are decreasing at 0.02 cm/min. Find the rate at which the volume is changing when the sides are 2 cm.

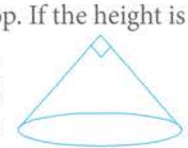
7. A liquid is pumped into an upright cylindrical tank of radius 1.5 m at a rate of 0.25 m^{3}s^{-1}.

   At what rate is the depth of the liquid increasing when it reaches:

   a a depth of 1.25 m?

   b a volume of 10π m^3?

8^: A conical pile of sand with a constant vertical angle of 90° is having sand poured onto the top. If the height is increasing at the rate of 0.5 cm min^{-1}, find the rate at which sand is being poured when the height is 4 cm, giving an exact answer.



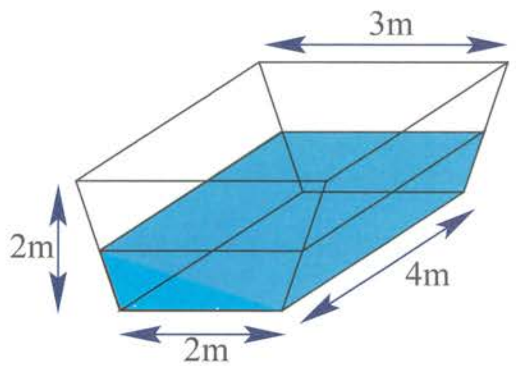
9. An aeroplane flies over an airport at an altitude of 10,000 metres and at a speed of 900 kmh^{-1}. Find the rate at which the actual distance from the airport is increasing 2 minutes after the aeroplane was directly over the airport, correct to the nearest whole number.

10. The temperature inside a chemical reaction vessel, initially 35°C is rising at 7°C per hour.

    The rate at which the reaction happens is modelled by the function: rate $= \frac{t}{12} + 3$, where t is the temperature of the reaction vessel in °C. Find the rate at which the reaction is occurring after 5 hours.

11. A racing car, travelling at 180 km per hour, is passing a television camera on a straight road. The camera is 25 metres from the road. If the camera operator follows the car, find the rate (in radians per second) at which the camera must pan (rotate) at the moment when the car is at its closest to the camera.

12. The diagram shows a water trough. Water is being poured into this trough at 2.4 cubic metres per minute.



Extra questions

Answers

a Find an expression for the volume of water in the trough in terms of its depth.

b Find the rate at which the water level is rising when the depth is 0.5 metres.

c Find the rate at which the exposed surface area of the water is increasing after 1 minute.

13. A square-based pyramid with a fixed height of 20 metres is increasing in volume at 2 m^3min^{-1}. Find the rate at which the side length of the base is increasing when the base has an area of 10 m^2. Give an exact answer with a rational denominator.

14. The length of the edge of a regular tetrahedron is increasing at 2.5 cms^{-1}. Find the rate at which the volume is increasing when the edge is 4 cm.

15. A man 1.8 m tall is walking directly away from a street lamp 3.2 m above the ground at a speed of 0.7 m/s. How fast is the length of his shadow increasing?

16. A ladder 10 m long rests against a vertical wall. The bottom of the ladder, while maintaining contact with the ground, is being pulled away from the wall at 0.8 m/s. How fast is the top of the ladder sliding down the wall, when it is 2 m from the ground?

6.4 Anti-Differentiation

Integration

Antidifferentiation and the indefinite integral

As the name suggests, ***anti*differentiation** is the reverse process of differentiation. We are then searching for the answer to the following:

Given an expression $f'(x)$ (i.e. the derivative of the function $f(x)$), what must the original function $f(x)$ have been?

For example, if $f'(x) = 2x$ then $f(x) = x^2$ is a *possible expression* for the original function. Why do we say ' ... *is a possible expression for the original function*.'?

Consider the following results:

$f(x) = x^2 + 3, \Rightarrow f'(x) = 2x$ [1] & $f(x) = x^2 - 5, \Rightarrow f'(x) = 2x$ [2]

From equations 1 and 2 we see that given an expression $f'(x)$, there are a number of possible different original functions, $f(x)$. This is due to the fact that the derivative of a constant is zero and so when we are given an expression for $f'(x)$, there is no real way of knowing if there was a constant in the original function or what that constant might have been (unless we are given some extra information).

The best that we can do at this stage is to write the following:

Given that $f'(x) = 2x$, then $f(x) = x^2 + c$, where c is some real number that is yet to be determined (it could very well be that $c = 0$).

The antidifferentiation process described above can be summarized as follows:

Given that $\frac{dy}{dx} = f'(x)$, then (after antidifferentiating):

$$y = f(x) + c \text{ where } c \in \mathbb{R}$$

We say that $y = f(x) + c$ where $c \in \mathbb{R}$ is the **antiderivative** of $f'(x)$.

Language and notation

The set of all antiderivatives of a function $h(x)$ is called the **indefinite integral** of $h(x)$, and is denoted by $\int h(x)dx$.

The symbol $\int$ is called the **integral sign**, the function $h(x)$ is the **integrand** of the integral and x is the **variable of integration**.

Once we have found an antiderivative (or indefinite integral) of $h(x)$, $H(x)$ (say) we can then write:

$$\int h(x)dx = H(x) + c, \text{ where } c \in \mathbb{R}$$

The constant c is called the **constant of integration**. The above result is read as:

'The **antiderivative** of $h(x)$ with respect to x is $H(x) + c$, where $c \in \mathbb{R}$'.

or

'The **indefinite integral** of $h(x)$ with respect to x is $H(x) + c$, where $c \in \mathbb{R}$'.

Determining the indefinite integral

So—how do we find the indefinite integral of $h(x)$?

We approach this problem by searching for a pattern (pretty much as we did when dealing with the derivative of a function). Recall the following results (when we were searching for a rule to find the derivative of a function):

$h(x)$	x	x^2	x^3	x^4	x^5	...	x^n
$h'(x)$	1	$2x$	$3x^2$	$4x^3$	$5x^4$	...	nx^{n-1}

This suggests the 'standard form':

Since: $\frac{d}{dx}(x^{n+1}) = (n+1)x^n$ then $\boxed{\int x^n\,dx = \frac{x^{n+1}}{n+1}...(+c)}$

Note that, since we cannot have a zero denominator, n cannot be -1.

The case $\int x^{-1}\,dx = \int \frac{1}{x}\,dx$ will be dealt with later.

A slightly more general result is one where we have ax^n rather than simply x^n. In this case we have that:

$$\int ax^n\,dx = \frac{ax^{n+1}}{n+1}...(+c), n \neq -1$$

Example 6.4.1

Find the indefinite integral of the following.

a $4x^2$ b x^{-3}

c $5\sqrt{x}$ d $\sqrt[5]{x^3}$

In each case we use the standard form. That is, we first increase the power by one and then divide by the new power.

a $\int 4x^2 dx = \frac{4}{2+1}x^{2+1} + c, c \in \mathbb{R} = \frac{4}{3}x^3 + c, c \in \mathbb{R}$

b $\int x^{-3} dx = \frac{1}{-3+1}x^{-3+1} + c, c \in \mathbb{R} = -\frac{1}{2}x^{-2} + c, c \in \mathbb{R}$

$= -\frac{1}{2x^2} + c, c \in \mathbb{R}$

c $\int 5\sqrt{x}dx = \int 5x^{\frac{1}{2}}dx = \frac{5}{\frac{1}{2}+1}x^{\frac{1}{2}+1} + c, c \in \mathbb{R}$

$= \frac{5}{(3/2)}x^{3/2} + c, c \in \mathbb{R} \quad = \frac{10}{3}\sqrt{x^3} + c, c \in \mathbb{R}$

Notice that, before we can start the antidifferentiation process, we must express the integrand in the form ax^n, i.e. in power form.

$\int \sqrt[5]{x^3}\,dx = \int x^{\frac{3}{5}}dx = \frac{1}{\frac{3}{5}+1}x^{\frac{3}{5}+1} + c, c \in \mathbb{R}$

$= \frac{1}{(8/5)}x^{\frac{8}{5}} + c, c \in \mathbb{R} \quad = \frac{5}{8}\sqrt[5]{x^8} + c, c \in \mathbb{R}$

Although we have been working through examples that are made up of only one integrand, we can determine the indefinite integral of expressions that are made up of several terms.

Example 6.4.2

Find: a $\int(2x^2 + x^3 - 4)dx$

b $\int(x-1)(x^4+3x)dx$ c $\int\frac{z^4 - 2z^2 + 3}{z^2}dz$

a $\int(2x^2 + x^3 - 4)dx = \int 2x^2 dx + \int x^3 dx - \int 4dx$

$= \frac{2}{2+1}x^{2+1} + \frac{1}{3+1}x^{3+1} - 4x + c$

$= \frac{2}{3}x^3 + \frac{1}{4}x^4 - 4x + c, c \in \mathbb{R}$

When determining the indefinite integral of 4, we have actually thought of '4' as '$4x^0$'.

So that $\int 4dx = \int 4x^0 dx = \frac{4}{0+1}x^{0+1} = 4x$.

b $\int(x-1)(x^4+3x)dx = \int(x^5 - x^4 + 3x^2 - 3x)dx$

$= \frac{1}{6}x^6 - \frac{1}{5}x^5 + x^3 - \frac{3}{2}x^2 + c$

c $\int\frac{z^4 - 2z^2 + 3}{z^2}dz = \int\left(\frac{z^4}{z^2} - \frac{2z^2}{z^2} + \frac{3}{z^2}\right)dz$

$= \int(z^2 - 2 + 3z^{-2})dz$

$= \frac{1}{3}z^3 - 2z + \frac{3}{-1}z^{-1} + c$

$= \frac{1}{3}z^3 - 2z - \frac{3}{z} + c, c \in \mathbb{R}$

Notice that in part b it was necessary to first multiply out the brackets **before** we could integrate. Similarly, for part c we had to first carry out the division **before** integrating.

Exercise 6.4.1

1. Find the indefinite integral of the following.

a	x^3	b	x^7	c	x^5
d	x^8	e	$4x^2$	f	$7x^5$
g	$9x^8$	h	$\frac{1}{2}x^3$		

2. Find:

a	$\int 5dx$	b	$\int 3dx$	c	$\int 10dx$
d	$\int \frac{2}{3}dx$	e	$\int -4dx$	f	$\int -6dx$
g	$\int -\frac{3}{2}dx$	h	$\int -dx$		

3. Find:

a	$\int(1-x)dx$	b	$\int(2+x^2)dx$
c	$\int(x^3-9)dx$	d	$\int\left(\frac{2}{5}+\frac{1}{3}x^2\right)dx$
e	$\int\left(\frac{2}{4}\sqrt{x}-\frac{1}{x^2}\right)dx$	f	$\int\left(\frac{5}{2}\sqrt{x^3}+8x\right)dx$
g	$\int x(x+2)dx$	h	$\int x^2\left(3-\frac{2}{x}\right)dx$
i	$\int(x+1)(1-x)dx$		

4. Find the antiderivative of the following.

a	$(x+2)(x-3)$	b	$(x^2-3x)(x+1)$
c	$(x-3)^3$	d	$(x+2x^3)(x+1)$
e	$(1-\sqrt{x})(1+x)$	f	$\sqrt{x}(x+1)^2-2$

5. Find:

a $\int\frac{x^2-3x}{x}dx$

b $\int\frac{4u^3+5u^2-1}{u^2}du$

c $\int\frac{(x+2)^2}{x^4}dx$

d $\int\frac{x^2+5x+6}{x+2}dx$

e $\int\frac{x^2-6x+8}{x-2}dx$

f $\int\left(\frac{t^2+1}{t}\right)^2dt$

Extra questions

More Standard Integrals

In the same way that there are rules for differentiating functions other than those of the form ax^n, we also have standard rules for integrating functions other than the ones that we have been dealing with so far. That is, there are standard rules for finding the indefinite integral of circular trigonometric functions, exponential functions and logarithmic functions.

We can deduce many such antiderivatives using the result $\int\frac{d}{dx}(h(x))dx = h(x)+c$ – (1).

For example, if we consider the derivative $\frac{d}{dx}(e^{2x}) = 2e^{2x}$,

We can write $\int\frac{d}{dx}(e^{2x})dx = \int 2e^{2x}dx$.

But from (1) we have $\int\frac{d}{dx}(e^{2x})dx = e^{2x}$.

Therefore, $e^{2x}+c = \int 2e^{2x}dx$.

Or, we could write:

$$e^{2x}+c = 2\int e^{2x}dx \Leftrightarrow \int e^{2x}dx = \frac{1}{2}(e^{2x}+c) = \frac{1}{2}e^{2x}+c$$

Similarly, $\frac{d}{dx}(\ln x) = \frac{1}{x}$ and so, antidifferentiating both sides we have $\int\frac{d}{dx}(\ln x)dx = \int\frac{1}{x}dx$.

Then, as $\int\frac{d}{dx}(\ln x)dx = \ln x$, we can write, $\ln x + c = \int\frac{1}{x}dx$.

We summarize these rules in the table below:

$f(x)$	$\int f(x)dx$	Examples
$ax^n, n\neq -1$	$\frac{a}{n+1}x^{n+1}+c$, $n\neq -1$	$\int 2x^2dx = \frac{2}{3}x^3+c$
$\frac{1}{x}, x\neq 0$	$\log_e\lvert x\rvert + c, x\neq 0$ or $\log_e x + c, x>0$	$\int\frac{6}{x}dx = 6\int\frac{1}{x}dx$ $= 6\log_e x + c, x>0$
$\sin(kx)$	$-\frac{1}{k}\cos(kx)+c$	$\int\sin(5x)dx$ $= -\frac{1}{5}\cos(5x)+c$
$\cos(kx)$	$\frac{1}{k}\sin(kx)+c$	$\int\cos\left(\frac{x}{4}\right)dx = \frac{1}{\left(\frac{1}{4}\right)}\sin\left(\frac{x}{4}\right)+c$ $= 4\sin\left(\frac{x}{4}\right)+c$
$\sec^2(kx)$	$\frac{1}{k}\tan(kx)+c$	$\int\sec^2(2x)dx = \frac{1}{2}\tan(2x)+c$
e^{kx}	$\frac{1}{k}e^{kx}+c$	$\int e^{-3x}dx = -\frac{1}{3}e^{-3x}+c$

NB: In the previous table (and from here on), the constant c is assumed to be a real number.

General power rule

The indefinite integral of $f(ax+b)$:

If $\int f(x)dx = F(x)+c$, then $\int f(ax+b)dx = \frac{1}{a}F(ax+b)+c$.

This means that all of the integrals in the table can be generalized further.

In particular we consider the generalized power rules:

$$\int (ax+b)^n dx = \frac{1}{a(n+1)}(ax+b)^{n+1}+c, n \neq -1$$
$$\int \frac{1}{ax+b}dx = \frac{1}{a}\log_e(ax+b)+c, ax+b>0$$

Example 6.4.3

Find the antiderivative of:

a $(3x+7)^4$ b $3(4-2x)^5$

c $\sqrt{5x+1}$ d $\frac{4}{(2x-3)^2}$

a This question requires the use of the result:

$$\int (ax+b)^n dx = \frac{1}{a(n+1)}(ax+b)^{n+1}+c, n \neq -1$$

$$\int (3x+7)^4 dx = \frac{1}{3\times(4+1)}(3x+7)^{4+1}+c$$
$$= \frac{1}{15}(3x+7)^5+c$$

b $\int 3(4-2x)^5 dx = 3\int (4-2x)^5 dx$

$$= 3\times\frac{1}{(-2)\times(5+1)}(4-2x)^{5+1}+c$$
$$= \frac{3}{-2\times 6}(4-2x)^6+c$$
$$= -\frac{1}{4}(4-2x)^6+c$$

c $\int \sqrt{5x+1}\,dx = \int (5x+1)^{1/2}dx$

(Note: we must first convert into power form!)

$$= \frac{1}{5\times\left(\frac{1}{2}+1\right)}(5x+1)^{\frac{1}{2}+1}+c$$

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

d $\int \frac{4}{(2x-3)^2}dx = 4\int (2x-3)^{-2}dx$

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

Example 6.4.4

Find the antiderivative of:

a $\sin 4x$ b $\cos\left(\frac{1}{2}x\right)$ c $\sec^2(3x)$.

a $\int \sin 4x\,dx = -\frac{1}{4}\cos 4x+c$

$\left[\text{Using } \int \sin(kx)dx = -\frac{1}{k}\cos(kx)+c\right]$

b $\int \cos\left(\frac{1}{2}x\right)dx = \frac{1}{\left(\frac{1}{2}\right)}\sin\left(\frac{1}{2}x\right)+c = 2\sin\left(\frac{1}{2}x\right)+c$

$\left[\text{Using } \int \cos(kx)dx = \frac{1}{k}\sin(kx)+c\right]$

c $\int \sec^2(3x)dx = \frac{1}{3}\tan(3x)+c$

$\left[\text{Using } \int \sec^2(kx)dx = \frac{1}{k}\tan(kx)+c\right]$

We can extend the results for the antiderivatives of circular trigonometric functions as follows:

$f(x)$	$\int f(x)dx$
$\sin(ax+b)$	$-\frac{1}{a}\cos(ax+b)+c$
$\cos(ax+b)$	$\frac{1}{a}\sin(ax+b)+c$
$\sec^2(ax+b)$	$\frac{1}{a}\tan(ax+b)+c$

Example 6.4.5

Find the antiderivative of:

a e^{2x} b $5e^{-3x}$

c $4x-2e^{\frac{1}{3}x}$ d $\frac{5}{x}$

e $\frac{2}{3x+1}$

a $\int e^{2x}dx = \frac{1}{2}e^{2x}+c$. $\left(\text{Using } \int e^{kx}dx = \frac{1}{k}e^{kx}+c\right)$

b
$$\int 5e^{-3x}dx = 5\int e^{-3x}dx = 5\times\frac{1}{-3}e^{-3x}+c = -\frac{5}{3}e^{-3x}+c$$

$\left(\text{Using } \int e^{kx}dx = \frac{1}{k}e^{kx}+c\right)$

c
$$\int\left(4x-2e^{\frac{1}{3}x}\right)dx = \int 4x\,dx - 2\int e^{\frac{1}{3}x}dx = 4\times\frac{1}{2}x^2 - 2\times\frac{1}{\left(\frac{1}{3}\right)}e^{\frac{1}{3}x}+c = 2x^2-6e^{\frac{1}{3}x}+c$$

d $\int\frac{5}{x}dx = 5\int\frac{1}{x}dx = 5\ln(x)+c, x>0$

e $\int\frac{2}{3x+1}dx = 2\int\frac{1}{3x+1}dx = \frac{2}{3}\ln(3x+1)+c, x>-\frac{1}{3}$

$\left[\int\frac{1}{ax+b}dx = \frac{1}{a}\log_e(ax+b)+c\right]$

As we have done for the circular trigonometric functions, we can extend the antiderivative of the exponential function to:

$$e^{ax+b}dx = \frac{1}{a}e^{ax+b}+c$$

Example 6.4.6

Find:

a $\int 2(4x-1)^6dx$ b $\int\frac{7}{4-3x}dx$

c $\int\frac{3}{\sqrt{2x+1}}dx$ d $\int e^{-2x+3}dx$

e $\int(e^{-4x}+\sqrt[5]{3x+2})dx$

f $\int\left(\cos\left(\frac{\pi}{2}x\right)-\frac{2}{6x+5}\right)dx$

a We have $n=6$, $a=4$ and $b=-1$,

$$\therefore \int 2(4x-1)^6dx = \frac{2}{4(6+1)}(4x-1)^{6+1}+c = \frac{1}{14}(4x-1)^7+c$$

b This time we use the case for $n=-1$, with $a=-3$ and $b=4$, giving:

$$\int\frac{7}{4-3x}dx = 7\int\frac{1}{4-3x}dx = -\frac{7}{3}\log_e(4-3x)+c, x<\frac{4}{3}$$

c We first rewrite the indefinite integral in power form:

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

So that $a=2$, $b=1$ and $n=-\frac{1}{2}$.

Therefore we have that:

$$\int\frac{3}{\sqrt{2x+1}}dx = \int 3(2x+1)^{-\frac{1}{2}}dx = \frac{3}{2\left(-\frac{1}{2}+1\right)}(2x+1)^{-\frac{1}{2}+1}+c = \frac{3}{2\times\frac{1}{2}}(2x+1)^{\frac{1}{2}}+c = 3\sqrt{2x+1}+c$$

d $\int e^{-2x+3}dx = -\frac{1}{2}e^{-2x+3}+c$

e
$$\int(e^{-4x}+\sqrt[5]{3x+2})dx = \int\left(e^{-4x}+(3x+2)^{\frac{1}{5}}\right)dx = -\frac{1}{4}e^{-4x}+\frac{1}{3\left(\frac{6}{5}\right)}(3x+2)^{\frac{6}{5}}+c = -\frac{1}{4}e^{-4x}+\frac{5}{18}\sqrt[5]{(3x+2)^6}+c$$

f
$$\int\left(\cos\left(\frac{\pi}{2}x\right)-\frac{2}{6x+5}\right)dx = \frac{1}{\frac{\pi}{2}}\sin\left(\frac{\pi}{2}x\right)-\frac{2}{6}\ln(6x+5)+c = \frac{2}{\pi}\sin\left(\frac{\pi}{2}x\right)-\frac{1}{3}\ln(6x+5)+c$$

Exercise 6.4.2

1. Find the indefinite integral of:

a e^{5x}	b e^{3x}	c e^{2x}
d $e^{0.1x}$	e e^{-4x}	f $4e^{-4x}$
g $0.1e^{-0.5x}$	h $2e^{1-x}$	i $5e^{x+1}$
j $-2e^{2-2x}$	k $e^{\frac{1}{3}x}$	l $\sqrt{e^x}$

2. Find the indefinite integral of:

a	$\frac{4}{x}$	b	$-\frac{3}{x}$
c	$\frac{2}{5x}$	d	$\frac{1}{x+1}$
e	$\frac{1}{2x}$	f	$\left(1-\frac{1}{x}\right)^2$
g	$\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^2$	h	$\frac{3}{x+2}$

3. Find the indefinite integral of:

a	$\sin(3x)$	b	$\cos(2x)$
c	$\sec^2(5x)$	d	$\sin(-x)$

4. Find the indefinite integral of:

a	$\sin(2x)+x$	b	$6x^2-\cos(4x)$
c	e^{5x}	d	$4e^{-3x}+\sin\left(\frac{1}{2}x\right)$
e	$\cos\left(\frac{x}{3}\right)-\sin(3x)$	f	$e^{2x}+\frac{4}{x}-1$
g	$(e^x+1)^2$		
h	$-5\sin(4x)+\frac{x-1}{x}$		
i	$\sec^2(3x)-\frac{2}{x}+e^{\frac{1}{2}x}$		
j	$(e^x-e^{-x})^2$	k	e^{2x+3}
l	$\sin(2x+\pi)$	m	$\cos(\pi-x)$
n	$\sin\left(\frac{1}{4}x+\frac{\pi}{2}\right)$	o	$\frac{e^x-2}{\sqrt{e^x}}$

5. Using the general power rule, find the indefinite integral of:

a	$(4x-1)^3$	b	$(3x+5)^6$
c	$(2-x)^4$	d	$(2x+3)^5$
e	$(7-3x)^8$	f	$\left(\frac{1}{2}x-2\right)^9$
g	$(5x+2)^{-6}$	h	$(9-4x)^{-2}$
i	$(x+3)^{-3}$	j	$\frac{1}{x+1}$
k	$\frac{2}{2x+1}$	l	$\frac{4}{3-2x}$
m	$\frac{-3}{5-x}$	n	$\frac{9}{3-6x}$
o	$\frac{5}{3x+2}$		

6. Find the antiderivative of:

a $\sin(2x-3)-2x$

b $3\cos\left(2+\frac{1}{2}x\right)+5$

c $\frac{1}{2}\cos\left(2-\frac{1}{3}x\right)-\frac{2}{2x+1}$

d $\sec^2(0.1x-5)-2$

e $\frac{4}{2x+3}-e^{-\frac{1}{2}x+2}$

f $\frac{4}{(2x+3)^2}-e^{2x-\frac{1}{2}}$

g $\frac{x+2}{x+1}-\frac{4}{x+2}$

h $\frac{2x+1}{x+2}+\frac{1}{2x+1}$

i $\frac{2}{(2x+1)^2}+\frac{2}{2x+1}$

7. Find $f(x)$ given that:

a $f'(x)=\sqrt{4x+5}$ where $f(-1)=\frac{1}{6}$

b $f'(x)=\frac{8}{4x-3}$ where $f(1)=2$.

c $f'(x)=\cos(2x+3)$ where $f\left(\pi-\frac{3}{2}\right)=1$

d $f'(x)=2-e^{-2x+1}$ where $f(0)=e$.

8. A bacteria population, N thousand, has a growth rate modelled by the equation:

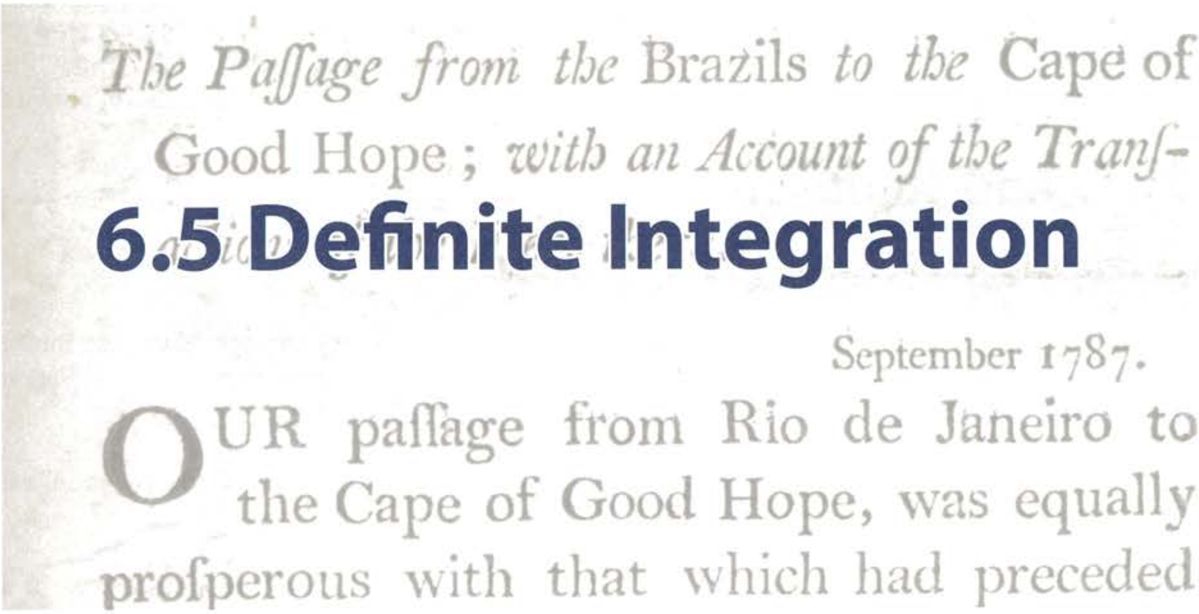
$$\frac{dN}{dt}=\frac{4000}{1+0.5t}, t\geq 0$$

where t is measured in days. Initially there are 250 bacteria in the population.

Find the population size after 10 days.

Extra questions

Answers



The Paſſage from the Brazils to the Cape of Good Hope; with an Account of the Tranſ-

September 1787.

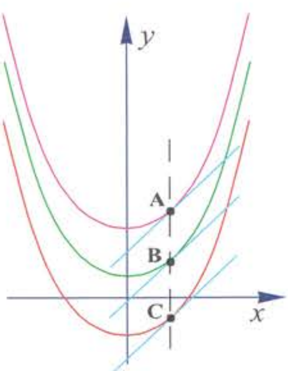
OUR paſſage from Rio de Janeiro to the Cape of Good Hope, was equally proſperous with that which had preceded

6.5 Definite Integration

Boundary Conditions

Our cover picture is from a book printed in 1789. Note the elongated 's's in the word 'passage'. These were common in both handwriting and printing at the time. The integral sign -∫- was chosen as an 'S' for 'summation'. The reason for this is part of the subject of this section.

Although we have already discussed the reason for adding a constant, c, when finding the indefinite integral, it is important that we can also determine the value of c.



We show the family of curves resulting from:

$$\frac{dy}{dx}=2x \Rightarrow y=x^2+c$$

To determine which of these curves is the one that we actually require, we must be provided with some extra information. In this case we would need to be given the coordinates of a point on the curve.

Example 6.5.1

Find $f(x)$ given that $f'(x) = 2x$ and that the curve passes through the point (2, 9).

As $f'(x) = 2x \Rightarrow f(x) = x^2 + c$.

Using the fact that at $x = 2$, $y = 9$, or that $f(2) = 9$, we have $9 = 2^2 + c \Leftrightarrow c = 5$.

Therefore, of all possible solutions of the form $y = x^2 + c$, the function satisfying the given information is $f(x) = x^2 + 5$.

Example 6.5.2

Find $f(x)$ given that $f''(x) = 6x - 2$ and that the gradient at the point (1, 5) is 2.

From the given information we have that $f'(1) = 2$ and $f(1) = 5$.

As $f''(x) = 6x - 2$ we have $f'(x) = 3x^2 - 2x + c_1$ – (1)

But, $f'(1) = 2 \therefore 2 = 3(1)^2 - 2(1) + c_1 \Leftrightarrow c_1 = 1$
(i.e. substituting into (1))

Therefore, we have $f'(x) = 3x^2 - 2x + 1$.

Next, from $f'(x) = 3x^2 - 2x + 1$ we have:

$f(x) = x^3 - x^2 + x + c_2$ – (2)

But, $f(1) = 5 \therefore 5 = (1)^3 - (1)^2 + (1) + c_2 \Leftrightarrow c_2 = 4$
(i.e. substituting into (2))

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

Sometimes, information is not given in the form of a set of coordinates. Information can also be 'hidden' in the context of the problem.

Example 6.5.3

The rate of change in pressure, p units, at a depth x cm from the surface of a liquid is given by $p'(x) = 0.03x^2$. If the pressure at the surface is 10 units, find the pressure at a depth of 5 cm.

Antidifferentiating both sides with respect to x, we have $\int p'(x)dx = \int 0.03x^2 dx$.

$\therefore p(x) = 0.01x^3 + c$ – (1)

At $x = 0$, $p = 10$. Substituting into (1) we have:

$10 = 0.01(0)^3 + c \Leftrightarrow c = 10$

Therefore, the equation for the pressure at a depth of x cm is $p(x) = 0.01x^3 + 10$.

At $x = 5$, we have $p(5) = 0.01(5)^3 + 10 = 11.25$. That is, the pressure is 11.25 units.

Exercise 6.5.1

1. Find the equation of the function in each of the following.

   a $f'(x) = 2x + 1$, given that the curve passes through (1, 5).

   b $f'(x) = 2 - x^2$ and $f(2) = \frac{7}{3}$.

   c $\frac{dy}{dx} = 4\sqrt{x} - x$, given that the curve passes through (4, 0).

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

   e $\frac{dy}{dx} = 3(x+2)^2$, given that the curve passes through (0, 8).

   f $\frac{dy}{dx} = \sqrt[3]{x} + x^3 + 1$, given that the curve passes through (1, 2).

   g $f'(x) = (x+1)(x-1) + 1$, and $f(0) = 1$

   h $f'(x) = 4x^3 - 3x^2 + 2$, and $f(-1) = 3$

2. Find the equation of the function $f(x)$ given that it passes through the point (–1,2) and is such that $f'(x) = ax + \frac{b}{x^2}$, where $f(1) = 4$ and $f'(1) = 0$.

3. The marginal cost for producing x units of a product is modelled by the equation $C'(x) = 30 - 0.06x$. The cost per unit is \$40. How much will it cost to produce 150 units?

4. If $\frac{dA}{dr} = 6 - \frac{1}{r^2}$, and $A = 4$ when $r = 1$, find A when $r = 2$.

5. The rate, in cm^3/sec, at which the volume of a sphere is increasing is given by the relation $\frac{dV}{dt} = 4\pi(2t+1)^2, 0 \le t \le 10$.

   If initially the volume is π cm^3, find the volume of the sphere when $t = 2$.

6. The rate of change of the number of deer, N, in a controlled experiment, is modelled by the equation $\frac{dN}{dt} = 3\sqrt{t^3} + 2t, 0 \le t \le 5$.

   There are initially 200 deer in the experiment. How many deer will there be at the end of the experiment?

7. If $\frac{dy}{dx} \propto \sqrt{x}$, find an expression for y, given that $y = 4$ when $x = 1$ and $y = 9$ when $x = 4$.

8. A function with gradient defined by $\frac{dy}{dx} = 4x - m$ at any point P on its curve passes through the point (2, –6) with a gradient of 4. Find the coordinates of its turning point.

9. The marginal revenue is given by $\frac{dR}{dx} = 25 - 10x + x^2, x \ge 0$, where R is the total revenue and x is the number of units demanded.

   Find the equation for the price per unit, $P(x)$.

10. The rate of growth of a culture of bacteria is modelled by the equation $200t^{1.01}$, $t \ge 0$, t hours after the culture begins to grow. Find the number of bacteria present in the culture at time t hours if initially there were 500 bacteria.

Extra questions

Why the definite integral?

Unlike the previous section where the indefinite integral of an expression resulted in a new expression, when finding the definite integral we produce a numerical value.

Definite integrals are important because they can be used to find different types of measures, for example, areas, volumes, lengths and so on. It is, in essence, an extension of the work we have done in the previous sections.

Language and notation

If the function $f(x)$ is continuous at every point on the interval [a,b] and $F(x)$ is any antiderivative of $f(x)$ on [a,b], then:

$\int_a^b f(x)dx$ is called the definite integral and is equal to $F(b)-F(a)$.

That is, $\int_a^b f(x)dx = F(b)-F(a)$.

Which is read as:

"the integral of $f(x)$ with respect to $\boldsymbol{x}$ from $\boldsymbol{a}$ to $\boldsymbol{b}$ is equal to $F(b) - F(a)$."

Usually we have an intermediate step to aid in the evaluation of the definite integral. This provides a somewhat 'compact recipe' for the evaluation process. This intermediate step is written as $[F(x)]_a^b$.

We therefore write

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

The process is carried out in four steps:

1. Find an indefinite integral of $f(x)$, $F(x)$ (say).
2. Write your result as $[F(x)]_a^b$.
3. Substitute a and b into $F(x)$.
4. Subtract: $F(b)-F(a)$ to obtain the numerical value.

Notice that the constant of integration c is omitted. This is because it would cancel itself out upon carrying out the subtraction: $(F(b)+c)-(F(a)+c) = F(b)-F(a)$.

In the expression $\int_a^b f(x)dx$, x is called the variable of integration, and a and b are called the lower limit and upper limit respectively. It should also be noted that there is no reason why the number b need be greater than the number a when finding the definite integral.

That is, it is just as reasonable to write $\int_{-3}^{2} f(x)dx$ as it is to write $-\int_{2}^{-3} f(x)dx$, both expressions are valid.

Example 6.5.4

Evaluate the following: a $\int_3^5 \frac{1}{x}dx$

b $\int_2^4 \left(x+\frac{1}{x}\right)^2 dx$ c $\int_0^1 \left(e^{2x}+\frac{3}{x+1}\right)dx$

d $\int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \sin 3x dx$ e $\int_5^2 (3x-4)^4 dx$

f $\int_{-2}^{0} (x-e^{-x})dx$

a $\int_3^5 \frac{1}{x}dx = [\log_e x]_3^5 = \log_e 5 - \log_e 3 = \log_e\left(\frac{5}{3}\right) \approx 0.511$

b $\int_2^4 \left(x+\frac{1}{x}\right)^2 dx = \int_2^4 \left(x^2+2+\frac{1}{x^2}\right)dx = \left[\frac{1}{3}x^3+2x-\frac{1}{x}\right]_2^4$

$$= \left(\frac{1}{3}(4)^3+2(4)-\frac{1}{4}\right)-\left(\frac{1}{3}(2)^3+2(2)-\frac{1}{2}\right)$$

$$= \left(\frac{64}{3}+8-\frac{1}{4}\right)-\left(\frac{8}{3}+4-\frac{1}{2}\right)$$

$$= \frac{275}{12}(\approx 22.92)$$

c $\int_0^1 \left(e^{2x}+\frac{3}{x+1}\right)dx = \left[\frac{1}{2}e^{2x}+3\log_e(x+1)\right]_0^1$

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

$$= \frac{1}{2}e^2+3\log_e 2-\frac{1}{2}$$

$$\approx 5.27$$

d $\displaystyle\int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \sin 3x\,dx = \left[-\frac{1}{3}\cos 3x\right]_{\frac{\pi}{6}}^{\frac{\pi}{2}} = -\frac{1}{3}[\cos 3x]_{\frac{\pi}{6}}^{\frac{\pi}{2}}$

$\displaystyle = -\frac{1}{3}\left(\cos\left(\frac{3\pi}{2}\right) - \cos\left(\frac{3\pi}{6}\right)\right)$

$= 0$

e $\displaystyle\int_5^2 (3x-4)^4 dx = \left[\frac{1}{3(5)}(3x-4)^5\right]_5^2 = \frac{1}{15}((2)^5 - (11)^5)$

$\displaystyle = -\frac{161019}{15} = -10734.6$

f $\displaystyle\int_{-2}^{0} (x - e^{-x})dx = \left[\frac{1}{2}x^2 + e^{-x}\right]_{-2}^{0}$

$\displaystyle = (0 + e^0) - \left(\frac{1}{2}(-2)^2 + e^2\right)$

$= -1 - e^2 \approx -8.39$

Example 6.5.5

Differentiate $y = e^{x^2+3}$.

Hence find the exact value of $\displaystyle\int_0^1 5xe^{x^2+3}dx$.

Differentiating we have, $\dfrac{d}{dx}(e^{x^2+3}) = 2xe^{x^2+3}$.

Therefore, $\displaystyle\int_0^1 5xe^{x^2+3}dx = \frac{5}{2}\int_0^1 2xe^{x^2+3}dx = \frac{5}{2}\int_0^1 \frac{d}{dx}(e^{x^2+3})dx$

$\displaystyle = \frac{5}{2}[e^{x^2+3}]_0^1 = \frac{5}{2}(e^4 - e^3)$

Notice that in this case we made use of the fact that if $f'(x) = g(x)$ then $\int g(x)dx = f(x) + c$.

Example 6.5.6

The production rate for radios by the average worker at Bat-Rad Pty Ltd t hours after starting work at 7:00 a.m., is given by $N'(t) = -2t^2 + 8t + 10,\ 0 \le t \le 4$.

How many units can the average worker assemble in the second hour of production?

The second hour starts at $t = 1$ and ends at $t = 2$. Therefore, the number of radios assembled by the average worker in the second hour of production is given by:

$$N = \int_{t=1}^{t=2} N'(t)dt.$$

That is, $\displaystyle N = \int_1^2 (-2t^2 + 8t + 10)dt = \left[-\frac{2}{3}t^3 + 4t^2 + 10t\right]_1^2$

$\displaystyle = \left(-\frac{2}{3}(2)^3 + 4(2)^2 + 10(2)\right) - \left(-\frac{2}{3}(1)^3 + 4(1)^2 + 10(1)\right)$

$\displaystyle = \frac{52}{3}$

Exercise 6.5.2

1. Evaluate the following.

a $\displaystyle\int_1^4 x\,dx$ b $\displaystyle\int_4^9 \sqrt{x}\,dx$

c $\displaystyle\int_2^3 \frac{2}{x^3}dx$ d $\displaystyle\int_{16}^9 \frac{4}{\sqrt{x}}dx$

2. Evaluate the following definite integrals (giving exact answers).

a $\displaystyle\int_1^2 \left(x^2 - \frac{3}{x^4}\right)dx$ b $\displaystyle\int_0^2 (x\sqrt{x} - x)dx$

c $\displaystyle\int_0^2 (1 + 2x - 3x^2)dx$ d $\displaystyle\int_{-2}^0 (x+1)dx$

e $\displaystyle\int_0^{-1} x^3(x+1)dx$ f $\displaystyle\int_{-1}^1 (x+1)(x^2-1)dx$

g $\displaystyle\int_1^4 (\sqrt{x}-1)^2 dx$ h $\displaystyle\int_1^2 \left(x - \frac{1}{x}\right)^2 dx$

i $\displaystyle\int_1^3 \left(\frac{x^3 - x^2 + x}{x}\right)dx$ j $\displaystyle\int_{-1}^1 (x - x^3)dx$

k $\displaystyle\int_1^4 \frac{x+1}{\sqrt{x}}dx$ l $\displaystyle\int_1^4 \left(\sqrt{\frac{2}{x}} - \sqrt{\frac{x}{2}}\right)dx$

3. Use a graphics calculator to check your answers to Question 2.

4. Evaluate the following definite integrals (giving exact values).

a $\int_0^1 (e^x + 1)dx$ b $\int_1^2 \frac{4}{e^{2x}}dx$

c $\int_{-1}^1 (e^x - e^{-x})dx$ d $\int_{-1}^1 (e^x + e^{-x})dx$

e $\int_{-1}^1 (e^x + e^{-x})^2 dx$ f $\int_2^0 e^{2x+1}dx$

g $\int_0^1 (\sqrt{e^x} - 1)dx$ h $\int_0^1 \left(e^{\frac{1}{4}x} - e^{4x}\right)dx$

i $\int_1^{-1} e^{1-2x}dx$

5. Use a graphics calculator to check your answers to Question 4.

6. Evaluate the following definite integrals (giving exact values).

a $\int_1^2 \frac{3}{x}dx$ b $\int_0^4 \frac{2}{x+1}dx$

c $\int_2^6 \frac{x+4}{x}dx$ d $\int_4^5 \left(x^2 + \frac{1}{x}\right)^2 dx$

e $\int_{-1}^0 \frac{3}{1-2x}dx$ f $\int_0^1 \frac{2}{x+1}dx$

7. Use a graphics calculator to check your answers to Question 6.

8. Evaluate the following definite integrals (giving exact values).

a $\int_0^{\frac{\pi}{2}} \sin(2x)dx$ b $\int_{-\pi}^0 \cos\left(\frac{1}{3}x\right)dx$

c $\int_{\frac{\pi}{6}}^{\frac{\pi}{3}} \sec^2 4x dx$ d $\int_0^{\pi} \left(\cos x - \sin\left(\frac{x}{2}\right)\right)dx$

9. Evaluate the following definite integrals (giving exact values).

a $\int_0^1 (x+1)^4 dx$ b $\int_1^3 \sqrt{2x+1}dx$

c $\int_{-1}^2 (1-2x)^3 dx$ d $\int_0^1 \frac{1}{(x+2)^3}dx$

10. Show that $\frac{2x+6}{x^2+6x+5} \equiv \frac{1}{x+1} + \frac{1}{x+5}$.

Hence evaluate $\int_0^2 \frac{2x+6}{x^2+6x+5}dx$.

11. Find $\frac{d}{dx}(x\sin 2x)$.

Hence find the exact value of $\int_0^{\pi} x\cos 2x dx$.

12. Given that $\int_a^b f(x)dx = m$ and $\int_a^b g(x)dx = n$, find:

a $\int_a^b 2f(x)dx - \int_a^b g(x)dx$ b $\int_a^b (f(x) - 1)dx$

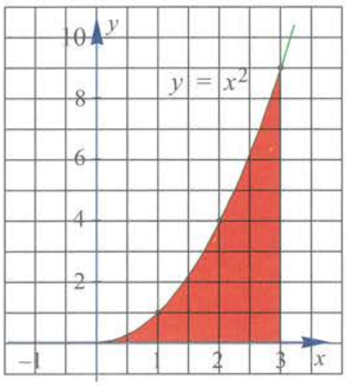
Extra questions

Areas

In Chapter 6.3 we saw that differentiation had a geometric meaning, that is, it provided a measure of the gradient of the curve at a particular point. We have also seen applications of the definite integral throughout the previous sections in this chapter. In this section we will investigate the geometric significance of the integral.

Introduction to the area beneath a curve

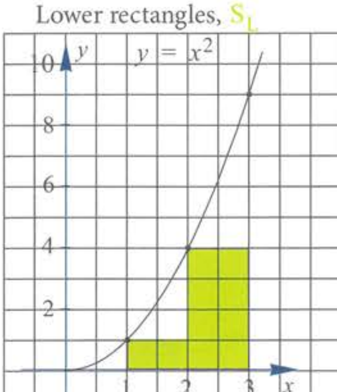
Consider the problem of finding the exact value of the red shaded area, A sq. units, in the diagram shown.



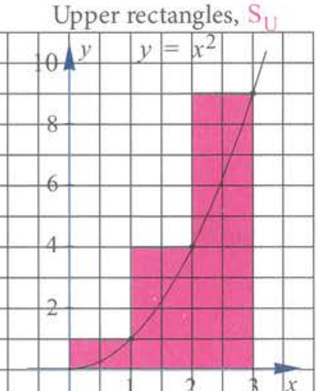
As a first step we make use of rectangular strips as shown below to obtain an approximation of the shaded area. We can set up a table of values, use it to find the area of each strip and then sum these areas.

In the figure, the green rectangles lie below the curve, and so we call these the lower rectangles. In the figure, magenta rectangles lie above the curve, and so we call these the upper rectangles. The red figure above shows that the true area (or exact area) lies somewhere between the sum of the areas of the lower rectangles, S_L, and the sum of the areas of the upper rectangles, S_U.

Lower rectangles, S_L



Upper rectangles, S_U



That is, we have that:

Lower Sum = S_L < Exact Area, A < S_U = Upper Sum

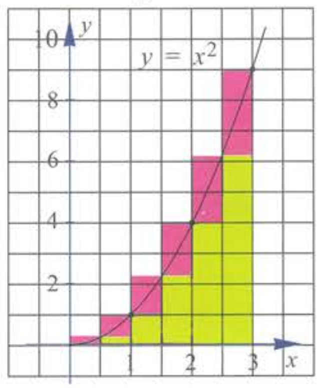
In the case above we have that $S_L = 1\times1 + 1\times4 = 5$ and $S_U = 1\times1 + 1\times4 + 1\times9 = 14$.

Therefore, we can write $5 < A < 14$. However, this does seem to be a poor approximation as there is a difference of 9 sq. units between the lower approximation and the upper approximation. The problem lies in the fact that we have only used two rectangles for the lower sum and three rectangles for the upper sum. We can improve on our approximation by increasing the number of rectangles that are used. For example, we could used 5 lower rectangles and 6 upper rectangles, or 10 lower rectangles and 12 upper rectangles and so on.

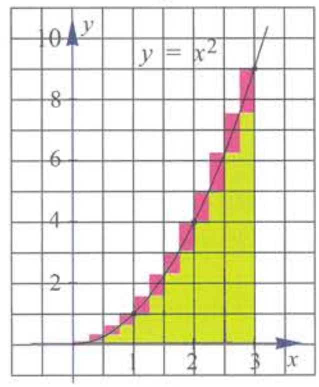
In search of a better approximation

As shown in the diagrams below, as we increase the number of rectangular strips (or decrease the width of each strip) we obtain better approximations to the exact value of the area.

5 lower & 6 upper rectangles



11 lower & 12 upper rectangles



For intervals of width 0.5 we have:

$$S_L = \frac{1}{2}\times[0.25 + 1 + 2.25 + 4 + 6.25] = 6.875$$

$$S_U = \frac{1}{2}\times[0.25 + 1 + 2.25 + 4 + 6.25 + 9] = 11.375$$

So: 6.875 < True area < 11.375

For intervals of width 0.25 we have:

$$S_L = \frac{1}{4}\times[0.25 + 0.5625 + 1 + 1.5625 + 2.25 + \ldots + 7.5625]$$

$$\approx 7.89$$

$$S_U = \frac{1}{4}\times[0.0625 + 0.25 + 0.5625 + \ldots + 7.5625 + 9] \approx 10.16$$

So: 7.56 < True area < 10.16

By continuing in this manner, the value of A will become *sandwiched* between a lower value and an upper value. Of course the more intervals we have the '*tighter*' the sandwich will be! What we can say is that if we partition the interval [0,3] into n equal subintervals, then, as the number of rectangles increases, S_L increases towards the exact value A while S_U decreases towards the exact value A.

That is, $\lim\limits_{n \to \infty} S_L = A = \lim\limits_{n \to \infty} S_U$

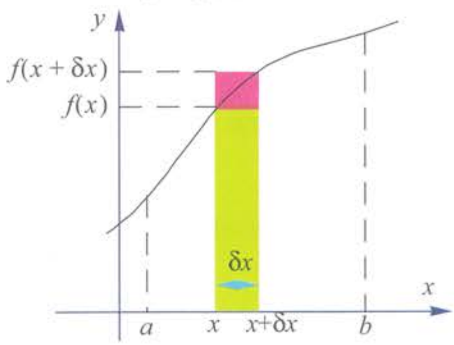
As we have seen, even for a simple case such as $y = x^2$, this process is rather tedious. And as yet, we still have not found the exact area of the shaded region under the curve $y = x^2$ over the interval $[0,3]$.

Towards an exact area

We can produce an algebraic expression to determine the exact area enclosed by a curve. We shall also find that the definite integral plays a large part in determining the area enclosed by a curve.

As a starting point we consider a single rectangular strip.

Consider the function $y = f(x)$ as shown:



Divide the interval from $x = a$ to $x = b$ into n equal parts: $a = x_0, x_1, x_2, \ldots, x_n = b$.

This means that each strip is of width $\dfrac{b-a}{n}$.

We denote this width by δx so that $\delta x = \dfrac{b-a}{n}$.

The area of the lower rectangle is $f(x) \times \delta x$ and that of the upper rectangle is $f(x + \delta x) \times \delta x$.

Then, the sum of the areas of the lower rectangles for $a \le x \le b$ is

$$S_L = \sum_{x=a}^{b-\delta x} f(x)\delta x$$

and the sum of the areas of the upper rectangles for $a \le x \le b$ is

$$S_U = \sum_{x=a}^{b} f(x+\delta x)\delta x$$

Then, if A sq units is the area under the curve $y = f(x)$ over the interval $[a, b]$ we have that:

$$\sum_{x=a}^{b-\delta x} f(x)\delta x < A < \sum_{x=a}^{b} f(x+\delta x)\delta x$$

As the number of strips increases, that is, as $n \to \infty$ and therefore $\delta x \to 0$ the area, A sq units, approaches a common limit, i.e. S_L from below, and S_U from above. We write this result as:

$$A = \lim_{\delta x \to 0} \sum_{x=a}^{b} f(x)\delta x$$

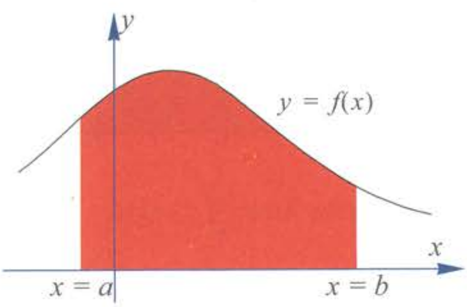
In fact this result leads to the use of the integral sign as a means whereby we can find the required area.

That is, $\lim\limits_{\delta x \to 0} \sum\limits_{x=a}^{b} f(x)\delta x = \int_a^b f(x)dx$

Notice that we've only developed an appropriate notation and a 'recognition' that the definite integral provides a numerical value whose geometrical interpretation is connected to the area enclosed by a curve, the x-axis and the lines $x = a$ and $x = b$. We leave out a formal proof of this result in preference to having developed an intuitive idea behind the concept and relationship between area and the definite integral.

We can now combine our results of the definite integral with its geometrical significance in relation to curves on a Cartesian set of axes.

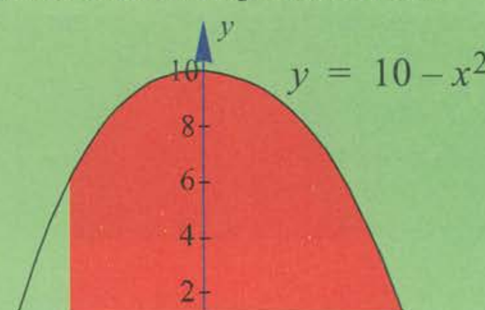
The definite integral and areas



If $y = f(x)$ is **positive** and **continuous** on the interval $[a,b]$, the area, A sq units, bounded by $y = f(x)$, the x-axis and the lines $x = a$ and $x = b$ is given by

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

Example 6.5.7
Find the area of the red region shown below:



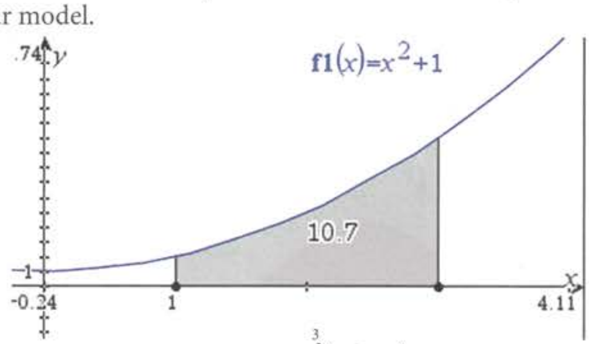
$$\text{Area} = \int_{-2}^{3} (10 - x^2)dx = \left[10x - \frac{1}{3}x^3\right]_{-2}^{3}$$
$$= \left((30 - 9) - \left(-20 + \frac{8}{3}\right)\right)$$
$$= \frac{115}{3}$$

Therefore, the shaded area measures $\frac{115}{3}$ square units.

Example 6.5.8

Find the area enclosed by the curve with equation $f(x) = x^2 + 1$, the x-axis and the lines $x = 1$ and $x = 3$.

Most calculators will produce numeric solutions to these area problems. Make sure you are familiar with the capabilities of your model.



The analytic solution is: $\text{Area} = \int_{1}^{3}(x^2+1)dx$

$$= \left[\frac{x^3}{3} + x\right]_1^3$$
$$= \left(\frac{3^3}{3} + 3\right) - \left(\frac{1^3}{3} + 1\right)$$
$$= 9 + 3 - \frac{1}{3} - 1$$
$$= 10\frac{2}{3}$$

Example 6.5.9

The area enclosed by the curve with equation $y = 4 - e^{-0.5x}$, the x-axis, the y-axis and the line $x = -2$, measures $k\text{–}2e$ sq units. Find the value of k.

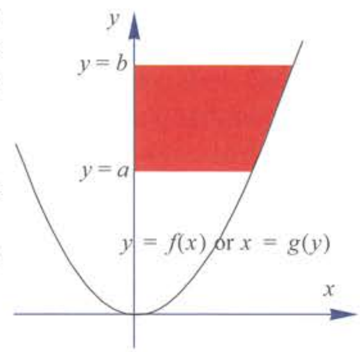
$$\text{Area} = \int_{-2}^{0} (4 - e^{-0.5x})dx = k - 2e$$

$$[4x + 2e^{-0.5x}]_{-2}^{0} = k - 2e$$
$$2 - (-8 + 2e^1) = k - 2e$$
$$10 - 2e = k - 2e$$

Therefore, $k = 10$.

Further observations about areas

To find the area bounded by $y = f(x)$, the y-axis and the lines $y = a$ and $y = b$ we carry out the following process:

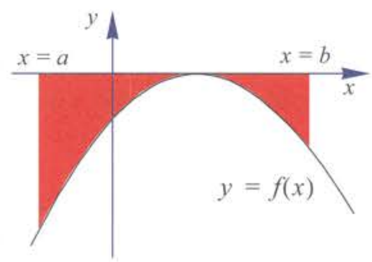


First you need to make x the subject, i.e. from $y = f(x)$ obtain the new equation $x = g(y)$.

Then find the definite integral, $\int_a^b x dy = \int_a^b g(y)dy$ sq units, which will give the red area.

If f is **negative** over the interval [a,b] (i.e $f(x) < 0$ for $a \le x \le b$, then the integral:

$\int_a^b f(x)dx$ is a negative number.



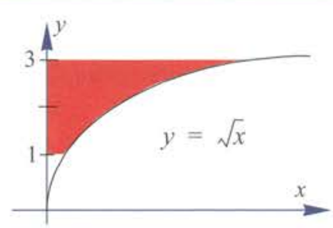
We therefore need to write the area, A, as:

$-\int_a^b f(x)dx$ or, use the **absolute value** of the integral:

$$A = \left|\int_a^b f(x)dx\right|$$

Example 6.5.10

Find the area enclosed by the curve $y = \sqrt{x}$, the y-axis and the lines $y = 1$ and $y = 3$.



We need an expression for x in terms of y:

That is, $y = \sqrt{x} \Rightarrow y^2 = x, x > 0$

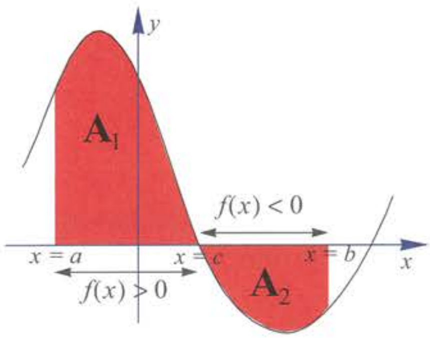
Therefore, the area of the shaded region is:

$$A = \int_1^3 x\,dy = \int_1^3 y^2 dy = \left[\frac{1}{3}y^3\right]_1^3 = \frac{1}{3}(3^3 - 1^3) = \frac{26}{3}.$$

The required area is $\frac{26}{3}$ sq units.

The Signed Area

It is possible for $y = f(x)$ to alternate between negative and positive values over the interval $x = a$ and $x = b$. That is, there is at least one point $x = c$ where the graph crosses the x-axis, and so $y = f(x)$ changes sign when it crosses the point $x = c$.



The integral $\int_a^b f(x)dx$ gives the **algebraic sum** of A_1 and A_2, that is, it gives the **signed area**.

For example, if $A_1 = 12$ and $A_2 = 4$, then the definite integral $\int_a^b f(x)dx = 12 - 4 = 8$.

This is because $\int_a^c f(x)dx = 12$, $\int_c^b f(x)dx = -4$ and so

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx = 12 + (-4) = 8$$

As $\int_c^b f(x)dx$ is a negative value, finding the negative of $\int_c^b f(x)dx$, that is $\left(-\int_c^b f(x)dx\right)$, would provide a positive value and therefore be a measure of the area of the region that is shaded below the x-axis. The red area would then be given by:

$$\int_a^c f(x)dx + \left(-\int_c^b f(x)dx\right)$$

This would provide the sum of two positive numbers.

Steps for finding areas

It follows, that in order to find the area bounded by the curve $y = f(x)$, the x-axis and the lines $x = a$ and $x = b$, we first need to find where (and if) the curve crosses the x-axis at some point $x = c$ in the interval $a \le x \le b$. If it does, we must evaluate the area of the regions above and below the x-axis **separately**.

Otherwise, evaluating $\int_a^b f(x)dx$ will provide the signed area (which only gives the correct area if the function lies above the x-axis over the interval $a \le x \le b$).

Therefore, we need to:

1. Sketch the graph of the curve $y = f(x)$ over the interval $a \le x \le b$. (In doing so you will also determine any x-intercepts).
2. Integrate $y = f(x)$ over each region separately (if necessary). That is, regions above the x-axis and regions below the x-axis.
3. Add the required (positive terms).

Example 6.5.11

Find the area of the region enclosed by the curve $y = x^3 - 1$, the x-axis and $0 \le x \le 2$.

First sketch the graph of the given curve:

x-intercepts (when $y = 0$): $x^3 - 1 = 0 \Leftrightarrow x = 1$

y-intercepts (when $x = 0$): $y = 0 - 1 = -1$.

From the graph we see that y is negative in the region [0,1] and positive in the region [1,2], therefore the area of the region enclosed is given by:

$$A = \left(-\int_0^1 (x^3 - 1)dx\right) + \int_1^2 (x^3 - 1)dx$$

$$= \left(-\left[\frac{x^4}{4} - x\right]_0^1\right) + \left[\frac{x^4}{4} - x\right]_1^2$$

$$= -\left(-\frac{3}{4}\right) + \left((2) - \left(-\frac{3}{4}\right)\right)$$

$$= 3.5$$

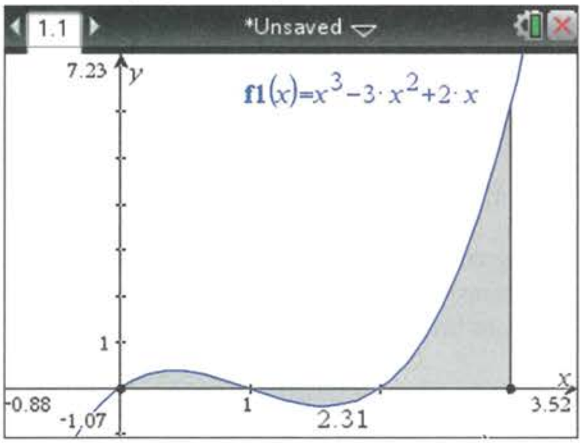
That is, the area measures 3.5 sq units.

Notice that $\int_0^2 (x^3-1)dx = \left[\frac{x^4}{4}-x\right]_0^2 = 2 \ (\neq 3.5)$.

Example 6.5.12

Find the area enclosed by the curve $y = x^3-3x^2+2x$, the x-axis and the lines $x = 0$ and $x = 3$.

First sketch the graph of the given curve:



x-intercepts (when $y = 0$):

$x^3-3x^2+2x = 0 \Leftrightarrow x(x-2)(x-1) = 0 \therefore x = 0, 2, 1$

y-intercepts (when $x = 0$): $y = 0-0+0 = 0$.

From the diagram we have, Area = $A_1 - A_2 + A_3$.

$$A_1 = \int_0^1 (x^3-3x^2+2x)dx = \left[\frac{x^4}{4}-x^3+x^2\right]_0^1 = \frac{1}{4}$$

$$A_2 = \int_1^2 (x^3-3x^2+2x)dx = \left[\frac{x^4}{4}-x^3+x^2\right]_1^2 = -\frac{1}{4}$$

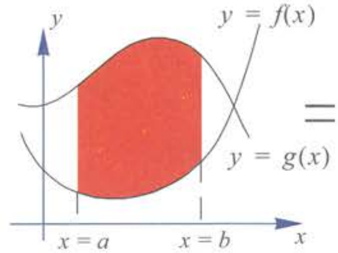
$$A_3 = \int_2^3 (x^3-3x^2+2x)dx = \left[\frac{x^4}{4}-x^3+x^2\right]_2^3 = \frac{9}{4}$$

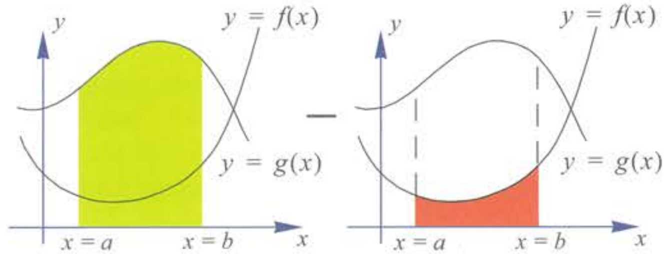
Therefore, the required area is $\frac{1}{4}-\left(-\frac{1}{4}\right)+\frac{9}{4} = \frac{11}{4}$ sq units.

Area between two curves

The use of the definite integral in finding the area of a region enclosed by a single curve can be extended to finding the area enclosed between two curves. Although we do have a compact formula to find such areas, in reality it is a known geometrical observation.

Consider two continuous functions, $f(x)$ and $g(x)$ on some interval $[a,b]$, such that over this interval, $g(x) \geq f(x)$. The area of the region enclosed by these two curves and the lines $x = a$ and $x = b$ is shown next.





Area ***between*** *the curves* =

Area beneath $y = g(x)$ — *Area beneath* $y = f(x)$

That is,

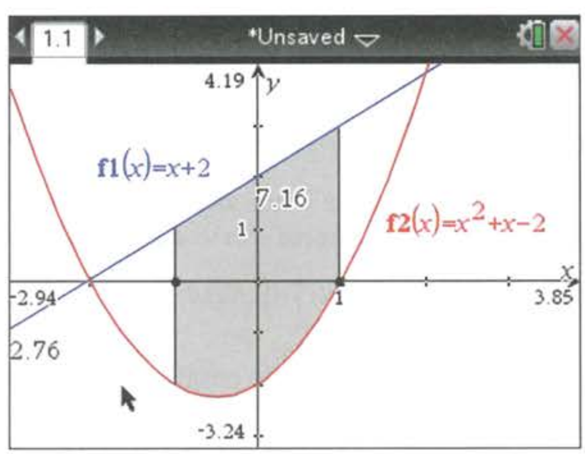
If $g(x) \geq f(x)$ on the interval $[a,b]$, then the area, A square units, enclosed by the two curves and the lines $x=a$ and $x=b$ is given by:

$$A = \int_a^b g(x)dx - \int_a^b f(x)dx = \int_a^b (g(x)-f(x))dx$$

Example 6.5.13

Find the area of the region enclosed by the curves $g(x) = x+2$, $f(x) = x^2+x-2$ and the lines $x = -1$ and $x = 1$.

The first step is to sketch both graphs so that it is clear which one lies above the other.

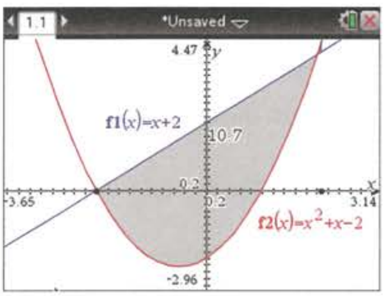


In this case, as $g(x) \geq f(x)$ on $[-1,1]$, we can write the required area, A sq units, as

$$\int_{-1}^{1}((x+2)-(x^2+x-2))dx = \int_{-1}^{1}(4-x^2)dx$$
$$= \left[4x-\frac{x^3}{3}\right]_{-1}^{1}$$
$$= \left(4-\frac{1}{3}\right)-\left(-4+\frac{1}{3}\right)$$
$$= \frac{22}{3} \text{ sq units}$$

Note: If the question had been stated as:

"Find the area enclosed by the curves $g(x) = x+2$ and $f(x) = x^2+x-2$." it would indicate that we want the total area enclosed by the two curves, as shown:



To find such an area we need first find the points of intersection: $x+2 = x^2+x-2 \Leftrightarrow x^2-4 = 0$ $\therefore x = \pm 2$

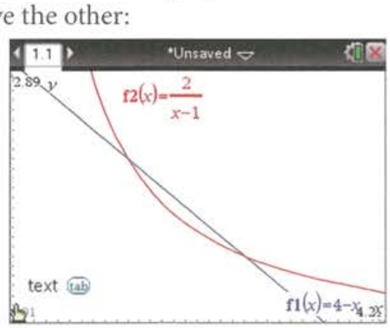
Area $= \int_{-2}^{2}(4-x^2)dx = \left[4x-\frac{x^3}{3}\right]_{-2}^{2} = \frac{32}{3}$ sq. units.

Example 6.5.14

Find the area of the region enclosed by the curves:

$y = 4-x$ and $y = \dfrac{2}{x-1}$.

Again we first sketch both graphs so that we can see which one lies above the other:



Next, we find the points of intersection:

$$4-x = \frac{2}{x-1} \Leftrightarrow (4-x)(x-1) = 2 \quad \Leftrightarrow (x-3)(x-2) = 0$$

Therefore, $x = 3$ or $x = 2$.

Required area $= \displaystyle\int_{2}^{3}\left((4-x)-\frac{2}{x-1}\right)dx$
$$= \left[4x-\frac{x^2}{2}-2\log_e(x-1)\right]_{2}^{3}$$
$$= \frac{3}{2}-2\log_e 2 \text{ sq units}$$

Exercise 6.5.3

1. Find the area of the region bounded by:

    a $y = x^3$, the x-axis, and the line $x = 2$.

    b $y = 4-x^2$, and the x-axis.

    c $y = x^3-4x$, the x-axis, and the lines $x = -2$ and $x = 0$.

    d $y = x^3-4x$, the x-axis, the line $x = 2$ and the line $x = 4$.

    e $y = \sqrt{x}-x$, the x-axis, and the lines $x = 0$ and $x = 1$.

2. Find the area of the region bounded by:

    a $f(x) = e^x+1$, the x-axis, and the lines $x = 0$ and $x = 1$.

    b $f(x) = e^{2x}-1$, the x-axis, the line $x = 1$ and the line $x = 2$.

    c $f(x) = e^x-e^{-x}$, the x-axis, the line $x = -1$ and the line $x = 1$.

    d $y = e^{\frac{1}{2}x+1}-x$, the x-axis, the line $x = 0$ and the line $x = 2$.

3. Find the area of the region bounded by:

    a $y = \dfrac{1}{x}$, the x-axis, the line $x = 4$ and the line

    $x = 5$.

    b $y = \dfrac{2}{x+1}$, the x-axis, the line $x = 0$ and the line

    $x = 4$.

c $f(x) = \frac{3}{2-x}$, the x-axis, the line $x = -1$ and the line $x = 1$.

d $f(x) = \frac{1}{x-1} + 1$, the x-axis, the line $x = -1$ and the line $x = \frac{1}{2}$.

4. Find the area of the region bounded by:

a $f(x) = 2\sin x$, the x-axis, the line $x = 0$ and the line $x = \frac{\pi}{2}$.

b $y = \cos(2x) + 1$, the x-axis, the line $x = 0$ and the line $x = \frac{\pi}{2}$.

c $y = x - \cos\left(\frac{x}{2}\right)$, the x-axis, the line $x = \frac{\pi}{2}$ and the line $x = \pi$.

d $f(x) = \cos(2x) - \sin\left(\frac{x}{2}\right)$, the x-axis, the line $x = \frac{\pi}{2}$ and the line $x = \pi$.

e $y = 3\sec^2\left(\frac{x}{2}\right)$, the x-axis, the line $x = -\frac{\pi}{3}$ and the line $x = \frac{\pi}{3}$.

5. Verify your answers to Questions 1–4, using a graphics calculator.

6. Find the area of the region enclosed by the curve $y = 8 - x^3$, the y-axis and the x-axis.

7. Find the area of the region enclosed by the curve $y = x^2 + 1$, and the lines $y = 2$ and $y = 4$.

8. Find the area of the region enclosed by the curve $f(x) = x + \frac{1}{x}$, the x-axis and the lines $x = -2$ and $x = -1$.

9. Find the area of the region enclosed by the curve $y = x^2 - 1$, the x-axis, the line $x = 0$ and $x = 2$.

10. Find the area of the region enclosed by the curve $y = x(x+1)(x-2)$ and the x-axis.

11. Find the area of the region enclosed by the curve $f(x) = 1 - \frac{1}{x^2}$

a the x-axis, the line $x = 1$ and $x = 2$.

b the x-axis, the line $x = \frac{1}{2}$ and $x = 2$.

c and the lines $y = -\frac{1}{2}$ and $y = \frac{1}{2}$.

12. The area of the region enclosed by the curve $y^2 = 4ax$ and the line $x = a$ is ka^2 sq units. Find the value of k.

13. Differentiate the function $y = \log_e(\cos 2x)$. Hence find the area of the region enclosed by the curve $f(x) = \tan(2x)$, the x-axis and the lines $x = 0$ and $x = \frac{\pi}{8}$.

14.a Find the area of the region enclosed by the curve $y = |2x - 1|$, the x-axis, the line $x = -1$ and the line $x = 2$.

b Find the area of the region enclosed by the curve $y = |2x| - 1$, the x-axis, the line $x = -1$ and the line $x = 2$.

15. Find the area of the region enclosed by the curve $f(x) = \frac{2}{(x-1)^2}$

a the x-axis, the lines $x = 2$ and $x = 3$,

b the y-axis, the lines $y = 2$ and $y = 8$.

16. Differentiate the function $y = x\log_e x$, hence find $\int \log_e x\,dx$.

17. a Find the area of the region enclosed by the curve $y = e^x$, the y-axis and the lines $y = 1$ and $y = e$.

b Find the area of the region bounded by the graphs of $y = x^2 + 2$ and $y = x$, over the interval $0 \le x \le 2$.

18. a Find the area of the region bounded by the graphs of $y = 2 - x^2$ and $y = x$, over the interval $0 \le x \le 1$.

b Find the area of the region bounded by the graphs of $y = 2 - x^2$ and $y = x$.

19. a Find the area of the regions bounded by the following:

i $y = x^3$, $x = 1$, $x = 2$ and $y = 0$.

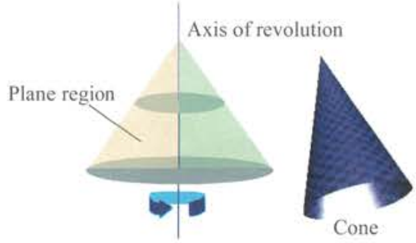
ii $y = x^3$, $y = 1$, $y = 8$ and $x = 0$.

b How could you deduce part ii from part i?

Extra questions

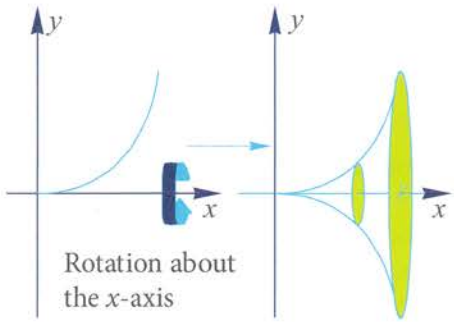
Volumes of Revolution

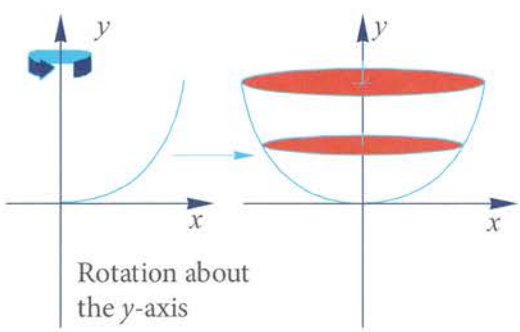
A solid of revolution is formed by revolving a plane region about a line – called the axis of revolution. In this section we will only be using the x-axis or the y-axis.



For example, in the diagram above, if we revolve the triangular plane region about the vertical axis as shown, we obtain a cone.

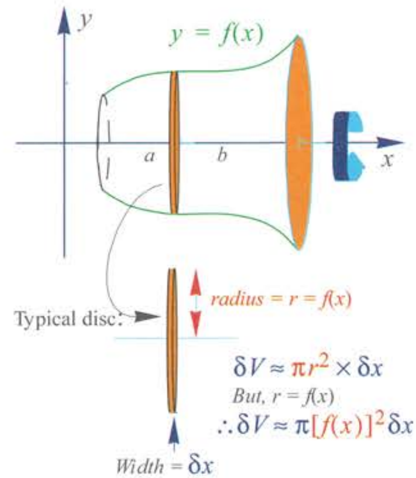
It is important to realize that depending on the axis of revolution, we can obtain very different shapes. For example, if a region bounded by the curve $y = x^2, x \geq 0$ is rotated about the x- and y- axes, two distinct solid shapes are formed:





When the plane region (enclosed by the curve and the x-axis) is rotated about the x-axis, the solid object produced is rather like the bell of a trumpet (with a very narrow mouth piece!) or a Malay hat on its side. However, when the plane region (enclosed by the curve and the y-axis) is rotated about the y-axis, then the solid produced is like a bowl.

Using the same approach as that used when finding the area of a region enclosed by a curve, the x-axis and the lines $x = a$ and $x = b$ we have:



Then, the volume, V units³, of such a solid can be cut up into a large number of slices (i.e. discs) each having a width δx and radius $f(x_i)$. The volume produced is then the sum of the volumes of these discs, i.e.

$$V = \sum_{i=0}^{i=n-1} \pi\left[f(x_i)\right]^2 \delta x \text{ where } \delta x = \frac{b-a}{n}$$

So, as $n \to \infty, \delta x \to 0$ and so,

$$V = \lim_{\delta x \to 0} \sum_{i=0}^{i=n-1} \pi\left[f(x_i)\right]^2 \delta x = \int_a^b \pi\left[f(x)\right]^2 dx$$

Therefore, we have:

The volume, V units³, of a solid of revolution is given by:

1. when a plane region enclosed by the curve $y = f(x)$ and the lines $x = a$ and $x = b$ is revolved about the x-axis.

$$V = \pi \int_{x=a}^{x=b} [f(x)]^2 dx \quad \left[\text{or } V = \pi \int_a^b y^2 dx\right]$$

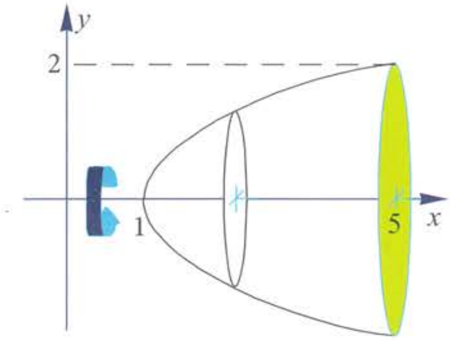
2. when a plane region enclosed by the curve $y = f(x)$ and the lines $y = e$ and $y = f$ is revolved about the y-axis.

$$V = \pi \int_{y=e}^{y=f} [f^{-1}(y)]^2 dy \quad \left[\text{or } V = \pi \int_e^f x^2 dy\right]$$

Example 6.5.15

The curve $y = \sqrt{x-1}, 1 \leq x \leq 5$ is rotated about the x-axis to form a solid of revolution. Sketch this solid and find its volume.

If the same curve is rotated about the y-axis, a different solid is formed. Sketch this second solid and find its volume.



The curve has a restricted domain and is rotated about the x-axis, so, the solid formed has a volume given by:

$$V = \pi\int_1^5 (\sqrt{x-1})^2 dx = \pi\int_1^5 (x-1)dx$$

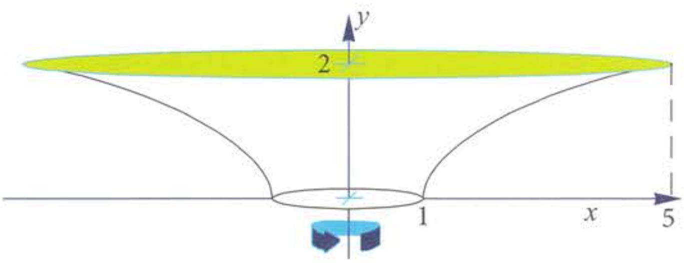
$$= \pi\left[\frac{x^2}{2} - x\right]_1^5$$

$$= \pi\left(\frac{5^2}{2} - 5 - \left(\frac{1^2}{2} - 1\right)\right)$$

$$= 8\pi$$

Therefore, the volume generated is 8π units³.

If the curve is rotated about the y-axis, the solid formed looks like this:



The volume can now be found using the second formula. It is important to realize that the integral limits are in terms of the y variable and so are 0 and 2. Also, x must be made the subject of the rule for the curve:

$$y = \sqrt{x-1} \Rightarrow y^2 = x-1 \Rightarrow x = y^2+1$$

When $x = 1, y = 0$ and when $x = 5, y = 2$, entering these values into the formula gives:

$$V = \pi\int_0^2 (y^2+1)^2 dy = \pi\int_0^2 (y^4+2y^2+1)dy =$$

$$\pi\left[\frac{y^5}{5} + \frac{2y^3}{3} + y\right]_0^2$$

$$= \pi\left(\frac{2^5}{5} + \frac{2(2^3)}{3} + 2\right)$$

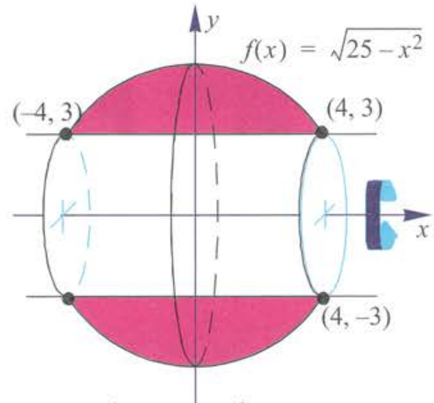
$$= 13\frac{11}{15}\pi$$

i.e. required volume is $13\frac{11}{15}\pi$ units³

Example 6.5.16

Find the volume of the solid formed by revolving the region enclosed by the curve with equation: $f(x) = \sqrt{25-x^2}$ and the line $g(x) = 3$ about the x-axis.

We start by drawing a diagram of this situation. It is a bead.



Next we determine the points of intersection.

Setting $f(x) = g(x)$ we have

$$\sqrt{25-x^2} = 3$$
$$\therefore 25-x^2 = 9$$
$$\Leftrightarrow x^2 = 16$$
$$\therefore x = \pm 4$$

The solid formed is hollow inside, i.e. from $-3 \le y \le 3$.

Next, we find the difference between the two volumes generated (a little bit like finding the area between two curves):

$$V = V_{f(x)} - V_{g(x)} = \pi\int_{-4}^{4} [f(x)]^2 dx - \pi\int_{-4}^{4} [g(x)]^2 dx$$

$$= \pi\int_{-4}^{4} ([f(x)]^2 - [g(x)]^2)dx$$

$$= 2\pi\int_0^4 ([f(x)]^2 - [g(x)]^2)dx \text{ (by symmetry)}$$

$$= 2\pi\int_0^4 ([\sqrt{25-x^2}]^2 - [3]^2)dx$$

$$= 2\pi\int_0^4 (16-x^2)dx$$

$$= 2\pi\left[16x - \frac{1}{3}x^3\right]_0^4$$

$$= \frac{256}{3}\pi$$

i.e. required volume is $\frac{256}{3}\pi$ units³.

Exercise 6.5.4

Finding volumes of revolution is an application of definite integration. Your only restriction will be the limitations on your ability to find integrals.

In the following exercise, you will need to draw on all the techniques you have learned in the preceding sections.

* Unless stated otherwise, all answers should be given as an exact value.

1. The part of the line $y = x+1$ between $x = 0$ and $x = 3$ is rotated about the x-axis. Find the volume of this solid of revolution.

2 A curve is defined by $y = \frac{1}{\sqrt{x}}, x \in [1, 5]$. If this curve is rotated about the x-axis, find the volume of the solid of revolution formed.

3. The curve $y = \frac{1}{x}$ between the x-values $\frac{1}{5}$ and 1 is rotated about the y-axis. Find the volume of the solid of revolution formed in this way.

4. Find the volume of the solid of revolution formed by rotating the part of the curve $y = e^x$ between $x = 1$ and $x = 5$ about the x-axis.

5. A solid is formed by rotating the curve $y = \sin x, x \in [0, 2\pi]$ about the x-axis. Find the volume of this solid.

6. The part of the curve $y = \frac{1}{1-x}$ between the x-values 2 and 3 is rotated about the x-axis. Find the volume of this solid.

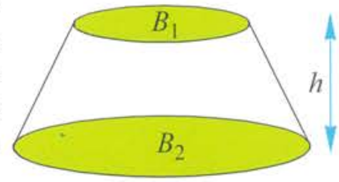
7. The part of the line $y = \frac{x-1}{2}$ between $x = 5$ and $x = 7$ is rotated about the y-axis. Find the volume of the solid of revolution formed in this way.

8. The part of the curve $y = \frac{x}{1+x}$ between the x-values 0 and 2 is rotated about the x-axis. Find the volume of the solid formed in this way.

9. Find the equation of the straight line that passes through the origin and through the point (h,r). Hence use calculus to prove that the volume of a right circular cone with base radius r and height h is given by $V = \frac{1}{3}\pi r^2 h$.

10. Find the equation of a circle of radius r. Use calculus to prove that the volume of a sphere is given by the formula $V = \frac{4}{3}\pi r^3$.

11. The diagram shows a shape known as a frustum. Use calculus to prove that its volume is given by the formula



$V = \frac{h}{3}(B_1 + B_2 + \sqrt{B_1 B_2})$ where B_1 and B_2 are the areas of the circular top and base respectively.

12. The part of the curve $f(x) = \sin\frac{x}{10}$ between $x = 0$ and $x = 5$ is rotated about the x-axis. Find the volume of this solid of revolution.

13. The part of the curve $f(x) = x^2 - x + 2$ between $x = 1$ and $x = 2$ is rotated about the x-axis. Find the volume of this solid of revolution.

14.

a Find the volume generated by the region between the y-axis and that part of the parabola $y = x^2$ from $x = 1$ to $x = 3$ when it is rotated about the y-axis.

b Find the volume generated by the region between the x-axis and that part of the parabola $y = x^2$ from $x = 1$ to $x = 3$ when it is rotated about the x-axis.

15. Find the volume of the solid of revolution that is formed by rotating the region bounded by the curves $y = \sqrt{x}$ and $y = \sqrt{x^3}$ about:

a the y-axis

b the x-axis.

16. Find the volume of the solid of revolution that is formed when the region bounded by the curve with equation $y = 4 - x^2$ and the line $y = 1$ is rotated about:

a the y-axis

b the x-axis.

17. Find the volume of the solid generated by rotating the region bounded by the curves $y^2 = x^3$ and $y^2 = 2 - x$ about the x-axis.

18. The volume of the solid formed when the region bounded by the curve $y = e^x - k$, the x-axis and the line $x = \ln 3$ is rotated about the x-axis is $\pi \ln 3$ units³. Find k.

19. Find the volume of the solid of revolution formed by rotating the region bounded by the axes and the curve $y = \sqrt{3}a\sin x + a\cos x,\ 0 \le x \le 2\pi,\ a > 0$ about the x-axis.

20. If the curve of the function $f(\theta) = \sin k\theta,\ k > 0,\ \theta > 0$ is rotated about the θ-axis, a string of sausages is made. Find k such that the volume of each sausage is π units^3.

21. a On the same set of axes, sketch the curves $y = ax^2$ and:

$$y = 1 - \frac{x^2}{a} \text{ where } a > 0.$$

b Find the volume of the solid of revolution formed when the region enclosed by the curves in part a is:

i rotated about the y-axis

ii rotated about the x-axis.

22. On the same set of axes sketch the two sets of points $\{(x, y) : (x-2)^2 + y^2 \le 4\}$ and $\{(x, y) : (x-a)^2 + y^2 \le 4,\ a \in]-2,6[\}$.

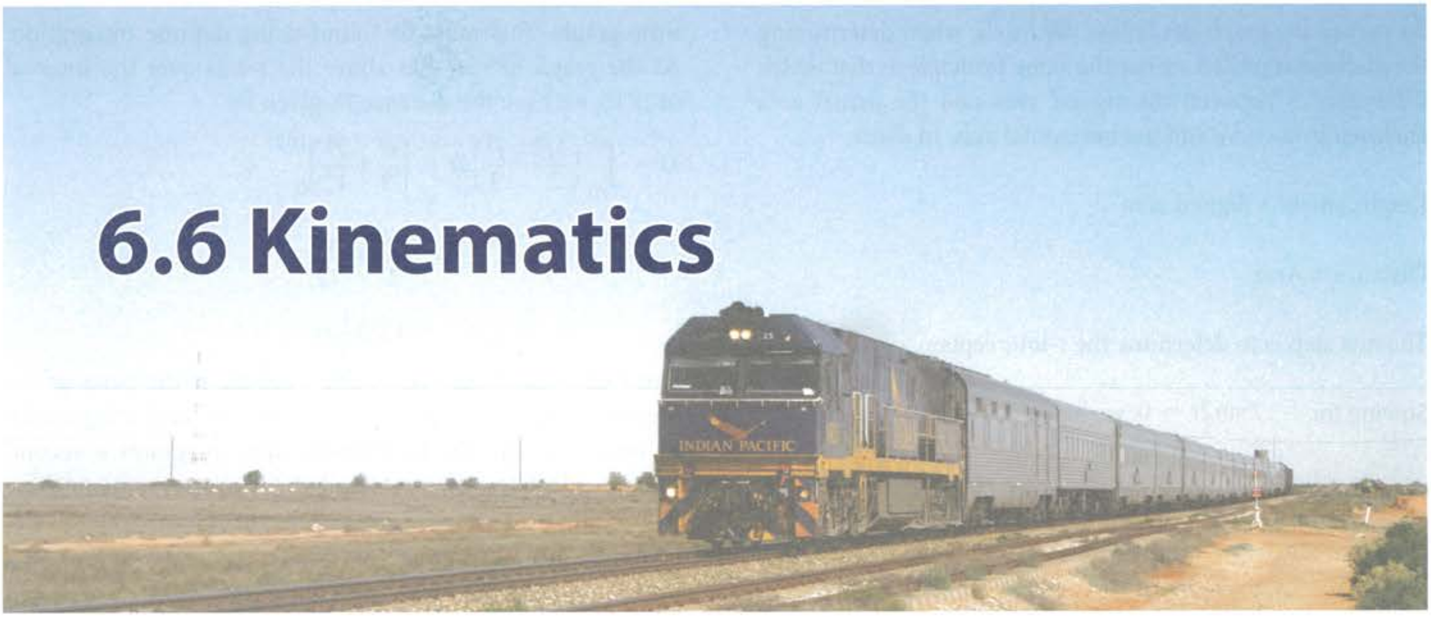
The intersection of these two sets is rotated about the x-axis to generate a solid. Find a if the volume of this solid is π units³. Give your answer to three decimal places.

A donut is formed by rotating the curve $\{(x, y) : (x-a)^2 + y^2 = 1,\ |a| > 1\}$ about the y-axis. Find a if the volume of the donut is 100π units³.

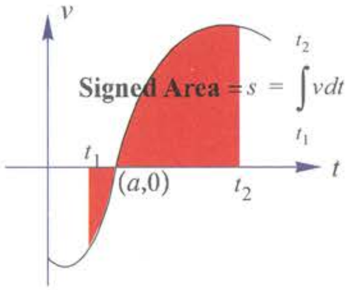
Extra questions

Answers

6.6 Kinematics



An application of integration when relating it to areas is that of kinematics. Just as the gradient of the displacement–time graph produces the velocity–time graph, so too then, we have that the **area beneath the velocity–time graph** produces the **displacement–time** graph. Notice that the area provides the displacement (not necessarily the distance!). Similarly with the acceleration–time graph, i.e. the **area under the acceleration–time graph** represents the **velocity**.



The displacement over the interval $[t_1, t_2]$ is given by

$$\textbf{Displacement} = s = \int_{t_1}^{t_2} v\,dt$$

However, the distance covered over the interval $[t_1, t_2]$ is given by

$$\textbf{Distance} = x = -\int_{t_1}^{a} v\,dt + \int_{a}^{t_2} v\,dt$$

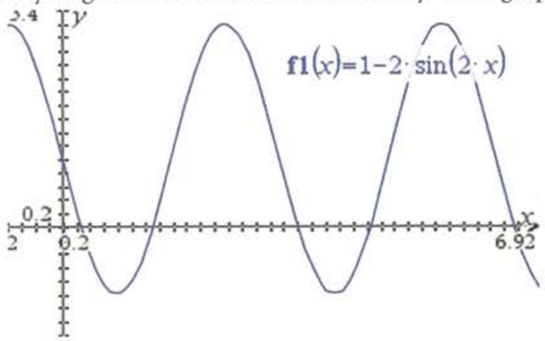
Example 6.6.1

The velocity of an object is v m/s after t seconds, where $v = 1 - 2\sin 2t$.

Find the object's displacement over the first $\frac{3\pi}{4}$ seconds.

Find the distance travelled by the object over the first $\frac{3\pi}{4}$ seconds.

It is always a good idea to sketch the velocity–time graph:



The displacement is then given by:

$$s = \int_0^{\frac{3\pi}{4}} (1 - 2\sin 2t)\,dt = \Big[t + \cos 2t\Big]_0^{\frac{3\pi}{4}}$$

$$= \left(\frac{3\pi}{4} + \cos\left(\frac{6\pi}{4}\right)\right) - (0 + \cos(0))$$

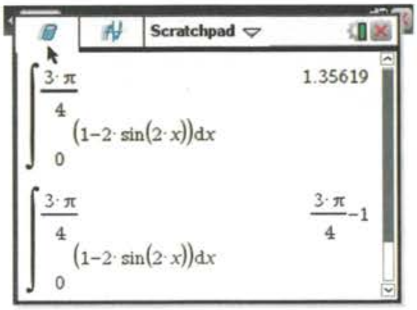
$$= \frac{3\pi}{4} - 1$$

$$\approx 1.36$$

That is, the object's displacement measures (approx.) 1.36 metres.

This time, we will use the scratchpad of the TI calculator followed by MENU 4 Calculus 3 Integral.

The required integral can be entered and either evaluated exactly or approximately.



As part of the graph lies below the t-axis, when determining the distance travelled we use the same principle as that which differentiates between the signed area and the actual area enclosed by a curve and the horizontal axis. In short:

Displacement = Signed area

Distance = Area.

The first step is to determine the t-intercepts:

Solving for $1 - 2\sin 2t = 0$ we have:

$$\sin 2t = \frac{1}{2} \Rightarrow 2t = \frac{\pi}{6}, \frac{5\pi}{6}, \ldots \quad \therefore t = \frac{\pi}{12}, \frac{5\pi}{12}, \ldots$$

Note that we only require the first two intercepts. Therefore, the distance is given by:

$$x = \int_0^{\frac{\pi}{12}} (1 - 2\sin 2t)dt - \int_{\frac{\pi}{12}}^{\frac{5\pi}{12}} (1 - 2\sin 2t)dt + \int_{\frac{5\pi}{12}}^{\frac{3\pi}{4}} (1 - 2\sin 2t)dt$$

Evaluating this expression is rather lengthy, and—unless we require an exact value—we might as well make use of the graphics calculator. There are a number of ways this can be done. We will use absolute value applied to the function.

$$\int_0^{\frac{3\cdot\pi}{4}} |1-2\cdot\sin(2\cdot x)|\,dx \qquad 2.7259$$

Distance travelled: = 0.1278 – (–0.6849) + 1.9132 = 2.7259.

That is, object travelled (approx.) 2.73 m.

Example 6.6.2

A rocket starts from rest and accelerates such that its acceleration is given by the formula $a(t) = 3t + t^2, 0 \le t \le 10$ where distances are measured in metres and time in seconds. Find the distance travelled in the first ten seconds of the rocket's motion.

The information is given as an acceleration. We must find the indefinite integral of this function to get a rule to give us the velocity.

$$a(t) = 3t + t^2, 0 \le t \le 10 \Rightarrow v(t) = \int (3t + t^2)dt = \frac{3t^2}{2} + \frac{t^3}{3} + c$$

Now, when $t = 0, v = 0 \therefore 0 = 0 + 0 + c \Rightarrow c = 0$

The constant is zero because we are told that the rocket starts from rest. The distance travelled is the area under this velocity time graph. This must be found using definite integration. As the graph of $v(t)$ lies above the t-axis over the interval $0 \le t \le 10$, we have the distance D, given by

$$D = \int_0^{10} \left(\frac{3t^2}{2} + \frac{t^3}{3}\right)dt = \left[\frac{t^3}{2} + \frac{t^4}{12}\right]_0^{10}$$

$$= \frac{10^3}{2} + \frac{10^4}{12}$$

$$= 1333\frac{1}{3}\text{m}$$

The technique described in this example is the basis of the inertial navigator. This senses acceleration and integrates it to infer velocity. The instrument then integrates a second time to calculate distance travelled. Of course, none of these quantities are generally expressed as exact mathematical formulae and the calculation has to be performed using numerical approximation.

Exercise 6.6.1

1. Find the displacement equation, $x(t)$, for each of the following:

   a $\frac{d^2x}{dt^2} = 6t$ where $\frac{dx}{dt} = 3$ and $x = 10$ when $t = 0$.

   b $\frac{d^2x}{dt^2} = -(4\sin t + 3\cos t)$ where $\frac{dx}{dt} = 4$ and $x = 2$ when $t = 0$.

   c $\frac{d^2x}{dt^2} = 2 - e^{-\frac{1}{2}t}$ where $\frac{dx}{dt} = 4$ and $x = 0$ when $t = 0$.

2. The acceleration, $a(t)$ ms^{-2}, of a body travelling in a straight line and having a displacement $x(t)$ m from an origin is governed by $a(t) = 6t - 2$ where:

   $\frac{dx}{dt} = 0$ and $x = 0$ when $t = 0$.

   a Find the displacement of the body at any time t.

   b Find the displacement of the body after 5 seconds.

   c Find the distance the body has travelled after 5 seconds.

3. A body moves along a straight line in such a way that its velocity, v ms^{-1}, is given by $v(t) = -\sqrt{t+4}+2$. After 5 seconds of motion the body is at the origin O.

   a Sketch the displacement-time graph for this body.

   b How far will the body have travelled after another 5 seconds.

4. A particle starts from rest and moves with a velocity, v ms^{-1}, where $v = t(t-5)$. Find the distance travelled between the two occasions when the particle is at rest.

5. A stone is thrown vertically upwards from ground level with a velocity of 25 ms^{-1}. If the acceleration of the stone is 9.8 ms^{-2} directed downwards, find the time taken before the stone reaches its highest point and the total distance travelled when the stone falls back to the ground.

6. The velocity of a particle is given by $v(t) = 3 - 3\sin 3t$ which is measured in m/s.

   a Find when the particle first comes to rest.

   b Find the distance travelled by the particle from when it started to when it first comes to rest.

7. An object, starting from rest, moves in a straight line with an acceleration that is given by:

$$a(t) = \frac{12}{(t+1)^2}\ \text{ms}^{-2}.$$

Find the distance travelled during the first 9 seconds.

8. An object has its velocity governed by the equation:

$$v(t) = 10\sin\left(\frac{\pi}{16}t\right)\ \text{m/s}.$$

   a Given that $s(0) = 0$, find its displacement equation.

   b Find its displacement after 20 seconds.

   c Find its displacement during the 20th second.

   d How far has it travelled in twenty seconds?

9. A particle moving in a straight line has its acceleration, a ms^{-2}, defined by the equation:

$$a = \frac{2k}{t^3}\ \text{ms}^{-2}.$$

At the end of the first second of motion, the particle has a velocity measuring 4 ms^{-1}.

   a Find an expression for the velocity of the particle.

   b Given that its velocity approaches a limiting value of 6 ms^{-1}, find k.

   c Find the distance travelled by the particle after a further 9 seconds.

10.

   a Show that:

$$\frac{d}{dx}\left[\frac{e^{ax}}{a^2+b^2}(a\cos bx + b\sin bx)\right] = e^{ax}\cos bx.$$

   b The velocity of a vibrating bridge component is modelled by the function $V = e^{-2t}\cos(3t)$. V ms^{-1} is the velocity of the component and t is the time in seconds after the observations begin.

Find the distance the component travels in the first tenth of a second.

Answers

Theory of Knowledge

Calculus

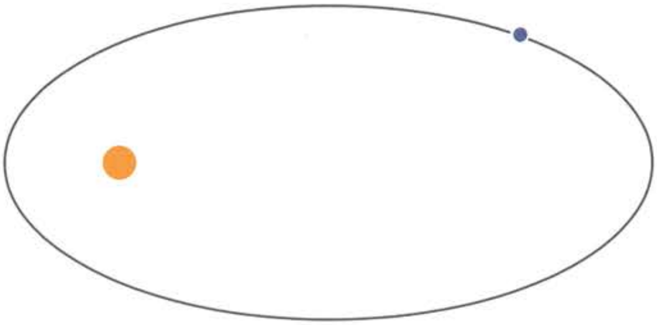
The word 'calculus' is derived from the Latin word meaning 'stone'. The connection between stones and calculation is the stone abacus:

However, the problem that set Isaac Newton thinking was astronomic.

The story that it was a falling apple is, however, Newton's.

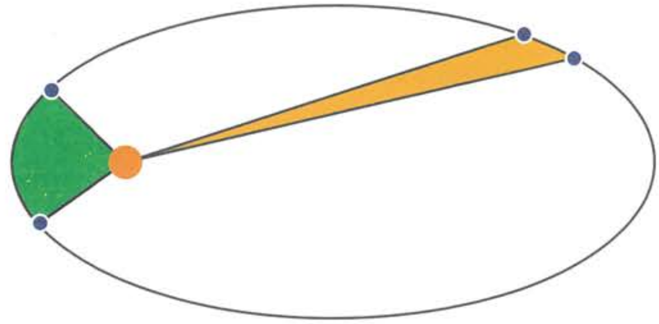
Actually, the problem was the details of planetary motion.

The German astronomer Johannes Kepler (1571-1630) had discovered that the planets move around the sun in elliptical orbits. The sun is at one of the foci of the ellipse.



Kepler had discovered more than that. The planet moves more slowly as its distance from the sun increases,

Specifically, the area it sweeps out in a given period is the same wherever it is in its orbit.



The green and orange areas are the same because the planet moves faster near the sun (green area) than it does at greater distances (orange area).

Animation of planetary motion

Newton was looking at a very complex problem as everthing appeared to be continuously changing.

This was the stimulus to develop the calculus - the study of variation.

His conclusion was that the force between the two bodies was proportional to the masses of the bodies and inversely proportional to the square of the distance between them.

A truly remarkable intellectual achievement.

6.7 Integration Methods

Further Integration

We can obtain the antiderivative, $F(x) + c$, of a function $f(x)$ based on the result that $\frac{d}{dx}(F(x)) = f(x)$.

That is, If $\frac{d}{dx}(F(x)) = f(x)$ then $\int f(x)dx = F(x) + c$

For example, if we know that $\frac{d}{dx}(\sin 5x) = 5\cos 5x$, then:

$\int 5\cos 5x\, dx = \sin 5x + c$.

Similarly, if $\frac{d}{dx}\left(\ln\left(x^2+1\right)\right) = \frac{2x}{x^2+1}$, then:

$\int \frac{2x}{x^2+1}dx = \ln\left(x^2+1\right) + c$.

We are using recognition to obtain antiderivatives. Such a skill is crucial to becoming successful at finding more complex antiderivatives.

One particularly important result is based on the chain rule, from which we obtained the generalised 'power rule' for differentiation;

$$\frac{d}{dx}\left(\left[f(x)\right]^n\right) = n.f'(x)\left[f(x)\right]^{n-1}.$$

From this result we have:

$$\int \frac{d}{dx}([f(x)]^n)dx = \int nf'(x)[f(x)]^{n-1}dx$$

so that $\int nf'(x)[f(x)]^{n-1}dx = [f(x)]^n + c$

This leads to the result:

$$\int g'(x)[g(x)]^n dx = \frac{1}{n+1}[g(x)]^{n+1} + c$$

The use of this result is dependent on an ability to recognise the expressions $g(x)$ and its derivative $g'(x)$ within the integrand. We consider a number of examples.

Example 6.7.1

Find the indefinite integral of the following.

a $2x(x^2+9)^5$ b $(3x^2+1)(x^3+x)^2$

c $-2x\sqrt{1-x^2}$

a We observe that $2x(x^2+9)^5$ can be written as $g'(x)[g(x)]^5$ with $g(x) = x^2+9$.

Therefore, by recognition we have:

$$\int 2x(x^2+9)^5 dx = \frac{1}{5+1}(x^2+9)^{5+1} + c$$
$$= \frac{1}{6}(x^2+9)^6 + c$$

b We observe that $(3x^2+1)(x^3+x)^2$ can be written as $g'(x)[g(x)]^2$ with $g(x) = x^3+x$.

Therefore, by recognition we have:

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

c We first express $-2x\sqrt{1-x^2}$ in the power form, $-2x(1-x^2)^{1/2}$.

We observe that $-2x(1-x^2)^{1/2}$ can be written as $g'(x)[g(x)]^{1/2}$ with $g(x) = 1-x^2$.

Therefore, by recognition we have:

$$\int -2x\sqrt{1-x^2}dx = \int -2x(1-x^2)^{1/2}dx$$

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

$$= \frac{2}{3}(1-x^2)^{3/2}+c$$

$$= \frac{2}{3}\sqrt{(1-x^2)^3}+c$$

Example 6.7.2

Find the indefinite integral of the following.

a $\dfrac{3x^2}{(x^3+4)^4}$ b $\dfrac{2-4x^3}{\sqrt{2x-x^4}}$

c $\dfrac{1}{x+1}\sqrt{\ln(x+1)}$.

a We rewrite $\dfrac{3x^2}{(x^3+4)^4}$ as $3x^2(x^3+4)^{-4}$.

We observe that $3x^2(x^3+4)^{-4}$ can be written as $g'(x)[g(x)]^{-4}$ with $g(x) = x^3+4$.

Therefore, by recognition we have:

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

$$= -\frac{1}{3}(x^3+4)^{-3}+c$$

$$= -\frac{1}{3(x^3+4)^3}+c$$

b First we rewrite $\dfrac{2-4x^3}{\sqrt{2x-x^4}}$ as $(2-4x^3)(2x-x^4)^{-1/2}$.

Then, we observe that $(2-4x^3)(2x-x^4)^{-1/2}$ can be written as $g'(x)[g(x)]^{-1/2}$ with $g(x) = 2x-x^4$.

By recognition we have:

$$\int (2-4x^3)(2x-x^4)^{-1/2}dx = \frac{1}{-\frac{1}{2}+1}(2x-x^4)^{-\frac{1}{2}+1}+c$$

$$= 2\sqrt{2x-x^4}+c$$

c First we rewrite $\dfrac{1}{x+1}\sqrt{\ln(x+1)}$ as $\dfrac{1}{x+1}[\ln(x+1)]^{1/2}$.

We observe that $\dfrac{1}{x+1}[\ln(x+1)]^{1/2}$ can be written as:

$g'(x)[g(x)]^{1/2}$ with $g(x) = \ln(x+1)$.

By recognition we have:

$$\int \frac{1}{x+1}[\ln(x+1)]^{1/2}dx = \frac{1}{\frac{1}{2}+1}[\ln(x+1)]^{\frac{1}{2}+1}+c$$

$$= \frac{2}{3}[\ln(x+1)]^{3/2}+c$$

$$= \frac{2}{3}\sqrt{[\ln(x+1)]^3}+c$$

What happens if the expression is not exactly in the form $\int g'(x)[g(x)]^n dx$, but only differs by some multiple? That is, what happens when we have $\int x(x^2+3)^4dx$ or $\int 5x(x^2+3)^4dx$ rather than $\int 2x(x^2+3)^4dx$?

As the expressions only differ by a multiple, we manipulate them so that they transform into $\int g'(x)[g(x)]^n dx$. For example:

$$\int x(x^2+3)^4dx =$$

$$\frac{1}{2}\int 2x(x^2+3)^4dx = \frac{1}{2}\times\frac{1}{5}(x^2+3)^5+c = \frac{1}{10}(x^2+3)^5+c$$

(i.e. multiply and divide by 2.)

$$\int 5x(x^2+3)^4dx =$$

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

(i.e. 'take' 5 outside the integral sign, then multiply and divide by 2.)

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

These manipulation skills are essential for successfully determining indefinite integrals by recognition.

Exercise 6.7.1

For this set of exercises, use the method of recognition to determine the integrals.

1. Find the following indefinite integrals.

a	$\int 10x\sqrt{5x^2+2}\,dx$	b	$\int \frac{x^2}{(x^3+4)^2}dx$
c	$\int -6x(1-2x^2)^3 dx$	d	$\int 3\sqrt{x}(9+2\sqrt{x^3})^4 dx$
e	$\int 6\cdot x\sqrt[3]{x^2+4}\,dx$	f	$\int \frac{2x+3}{(x^2+3x+1)^3}dx$

2. Find the antiderivatives of the following.

a	$2xe^{x^2+1}$	b	$\frac{3}{\sqrt{x}}e^{\sqrt{x}}$
c	$\sec^2 3x e^{\tan 3x}$	d	$(2ax+b)e^{-(ax^2+bx)}$
e	$3\sin\frac{1}{2}xe^{\cos\frac{1}{2}x}$	f	$\frac{4}{x^2}e^{4+x^{-1}}$

3. Find the antiderivatives of the following.

a	$2x\sin(x^2+1)$	b	$\frac{5}{\sqrt{x}}\sin(\sqrt{x})$
c	$\frac{2}{x^2}\cos\left(2+\frac{1}{x}\right)$	d	$\sin x\sqrt{\cos x}$
e	$\frac{\sin 3x}{\cos 3x}$	f	$\frac{4\sec^2 3x}{1+\tan 3x}$

4. Find the antiderivatives of the following.

a	$\frac{2}{4+x^2}$	b	$\frac{3}{x^2+9}$
c	$\frac{\sqrt{5}}{5+x^2}$	d	$\frac{1}{\sqrt{25-x^2}}$

5. Find the following indefinite integrals.

a	$\int \frac{3}{1+x^2}dx$	b	$\int \frac{5}{\sqrt{1-x^2}}dx$
c	$\int \frac{1}{\sqrt{4-x^2}}dx$	d	$\int \frac{1}{\sqrt{9-x^2}}dx$

6. Evaluate:

a	$\int_1^4 x^{1/2}(1+x^{3/2})^5 dx$	b	$\int_0^1 \frac{e^x}{\sqrt{1+e^x}}dx$
c	$\int_0^{\frac{3\pi}{4}} \frac{3\sin x}{1+\cos x}dx$	d	$\int_0^{\pi} \frac{8}{4+x^2}dx$
e	$\int_{-1}^{1} e^x\cos(e^x)dx$	f	$\int_0^{\frac{\pi}{2}} \sqrt{x}\sin x^{3/2}dx$
g	$\int_0^{\frac{\pi}{4}} \sqrt{\tan x}\sec^2 x\,dx$	h	$\int_{-1}^{1} 3x^2e^{x^3}dx$
i	$\int_e^{e^2} \frac{1}{x\ln x}dx$	j	$\int_3^4 x\sqrt{x^2-9}\,dx$

Extra questions

Substitution Rule

In the previous section, we considered integrals that required the integrand to be of a particular form in order to carry out the antidifferentiation process.

For example, the integral $\int 2x\sqrt{1+x^2}dx$ is of the form $\int h'(x)[h(x)]^n dx$ and so we could proceed by using the result:

$$\int h'(x)[h(x)]^n dx = \frac{1}{n+1}[h(x)]^{n+1}+c.$$

Next consider the integral $\int x\sqrt{x-1}\,dx$. This is not in the form $\int h'(x)[h(x)]^n dx$ and so we cannot rely on the recognition approach we have used so far. To determine such an integral we need to use a formal approach.

Indefinite integrals that require the use of the general power rule can also be determined by making use of a method known as the substitution rule (or change of variable rule). The name of the rule is indicative of the process itself. We introduce a new variable, u (say), and substitute it for an appropriate part (or the whole) of the integrand. An important feature of this method is that it will enable us to find the integral of expressions that cannot be determined by the use of the general power rule.

We illustrate this process using a number of examples (remembering that the success of this method is in making the appropriate substitution). The basic steps in integration by substitution can be summarized as follows:

1. Define u (i.e. let u be a function of the variable which is part of the integrand).

2. Convert the integrand from an expression in the original variable to an expression in u (this means that you also need to convert the 'dx' term to a 'du' term – if the original variable is x).

3. Integrate and then rewrite the answer in terms of x (by substituting back for u).

NB: This is only a guide, you may very well skip steps or use a slightly different approach.

Example 6.7.3

Find:

a $\int(2x+1)^4 dx$

b $\int 2x(x^2+1)^3 dx$ c $\int \frac{x^2}{\sqrt{x^3-4}} dx$.

a Although this integral can be evaluated by making use of the general power rule, we use the substitution method to illustrate the process:

In this case we let $u = 2x+1 \Rightarrow \frac{du}{dx} = 2 \therefore dx = \frac{1}{2}du$.

Having chosen u, we have also obtained an expression for dx and we are now in a position to carry out the substitution for the integrand:

$$\int(2x+1)^4 dx = \int u^4 \times \left(\frac{1}{2}du\right) = \frac{1}{2}\int u^4 du = \frac{1}{2}\times\frac{1}{5}u^5 + c$$

$$= \frac{1}{10}u^5 + c$$

Substituting back, we obtain, in terms of x: $= \frac{1}{10}(2x+1)^5 + c$

b This time, we let $u = x^2+1$. Note the difference between this substitution and the one used in part a. We are making a substitution for a non-linear term!

Now, $u = x^2+1 \Rightarrow \frac{du}{dx} = 2x \therefore \frac{1}{2}du = xdx$.

Although there is an x attached to the dx term, hopefully, when we carry out the substitution, everything will fall into place.

Now, $\int 2x(x^2+1)^3 dx = \int 2(x^2+1)^3 x dx$

(We have moved the x next to the dx.)

$= \int 2u^3 \times \frac{1}{2}du$ (substituting xdx for ½du.)

$$= \frac{1}{4}u^4 + c$$

$$= \frac{1}{4}(x^2+1)^3 + c$$

NB: A second (alternate) method is to obtain an expression for dx in terms of one or both variables. Make the substitution and then simplify. Although there is some dispute as to the 'validity' of this method, in essence it is the same. We illustrate this now:

c Let $u = x^3-4 \Rightarrow \frac{du}{dx} = 3x^2 \therefore dx = \frac{1}{3x^2}du$, making the substitution for u and dx, we have:

$\int \frac{x^2}{\sqrt{x^3-4}}dx = \int \frac{x^2}{\sqrt{u}} \times \frac{1}{3x^2}du = \frac{1}{3}\int u^{-1/2}du$ (Notice the x^2 terms cancel!)

$$= \frac{2}{3}u^{1/2} + c$$

$$= \frac{2}{3}\sqrt{x^3-4} + c$$

Example 6.7.4

Find $\int x\sqrt{x-1}\,dx$.

Letting $u = x-1 \Rightarrow \frac{du}{dx} = 1 \therefore du = dx$.

This then gives $\int x\sqrt{x-1}\,dx = \int x\sqrt{u}\,du$.

We seem to have come at an impasse. After carrying out the substitution we are left with two variables, x and u, and we need to integrate with respect to u! This is a type of integrand where not only do we substitute for the $x - 1$ term, but we must also substitute for the x term that has remained as part of the integrand, from $u = x - 1$ we have $x = u + 1$.

Therefore:

$$\int x\sqrt{x-1}\,dx = \int x\sqrt{u}\,du = \int (u+1)u^{1/2}du$$
$$= \int (u^{3/2} + u^{1/2})du$$
$$= \frac{2}{5}u^{5/2} + \frac{2}{3}u^{3/2} + c$$
$$= \frac{2}{5}\sqrt{(x-1)^5} + \frac{2}{3}\sqrt{(x-1)^3} + c$$

Example 6.7.5

The gradient at any point on the curve $y = f(x)$ is given by the equation:

$$\frac{dy}{dx} = \frac{1}{\sqrt{x+2}}$$

The curve passes through the point (2, 3). Find the equation of this curve.

Integrating both sides of $\frac{dy}{dx} = \frac{1}{\sqrt{x+2}}$ with respect to x, we have:

$$\int \frac{dy}{dx}dx = \int \frac{1}{\sqrt{x+2}}dx.$$

Let $u = x + 2 \Rightarrow \frac{du}{dx} = 1 \therefore du = dx$.

So, $\int \frac{1}{\sqrt{x+2}}dx = \int \frac{1}{\sqrt{u}}du = \int u^{-1/2}du = 2\sqrt{u} + c$

Therefore, we have $y = f(x) = 2\sqrt{x+2} + c$.

Now, $f(2) = 3 \Rightarrow 3 = 2\sqrt{4} + c \Leftrightarrow c = -1$.

Therefore, $f(x) = 2\sqrt{x+2} - 1$.

Example 6.7.6

Find the indefinite integral of the following.

a $x^2e^{x^3+4}$ b $e^x\cos(e^x)$

c $\frac{3x}{x^2+4}$ d $x^2\sqrt{x+1}$.

a Let $u = x^3 + 4 \Rightarrow \frac{du}{dx} = 3x^2 \therefore \frac{1}{3x^2}du = dx$.

Substituting, we have:

$$\int x^2e^{x^3+4}dx = \int \cancel{x^2}e^u \times \frac{1}{3\cancel{x^2}}du = \frac{1}{3}\int e^u du$$
$$= \frac{1}{3}e^u + c$$
$$= \frac{1}{3}e^{x^3+4} + c$$

b Let $u = e^x \Rightarrow \frac{du}{dx} = e^x \therefore dx = \frac{1}{e^x}du$.

Substituting, we have:

$$\int e^x\cos(e^x)dx = \int e^x\cos u \times \frac{1}{e^x}du = \int \cos u\,du$$
$$= \sin u + c$$
$$= \sin(e^x) + c$$

c Let $u = x^2 + 4 \Rightarrow \frac{du}{dx} = 2x \therefore dx = \frac{1}{2x}du$.

Substituting, we have:

$$\int \frac{3x}{x^2+4}dx = \int \frac{3\cancel{x}}{u} \times \frac{1}{2\cancel{x}}du = \frac{3}{2}\int \frac{1}{u}du = \frac{3}{2}\ln u + c$$
$$= \frac{3}{2}\ln(x^2+4) + c$$

d Let $u = x + 1 \Rightarrow \frac{du}{dx} = 1 \therefore dx = du$.

Substituting, we have $\int x^2\sqrt{x+1}\,dx = \int x^2\sqrt{u}\,du$. Then, as there is still an x term in the integrand, we will need to make an extra substitution. From $u = x + 1$ we have $x = u - 1$.

Therefore,

$$\int x^2\sqrt{u}\,du = \int (u-1)^2\sqrt{u}\,du = \int (u^2 - 2u + 1)u^{1/2}du$$
$$= \int (u^{5/2} - 2u^{3/2} + u^{1/2})du$$
$$= \frac{2}{7}u^{7/2} - \frac{4}{5}u^{5/2} + \frac{2}{3}u^{3/2} + c$$
$$= \frac{2}{7}(x+1)^{7/2} - \frac{4}{5}(x+1)^{5/2} + \frac{2}{3}(x+1)^{3/2} + c$$

Example 6.7.7

Find the indefinite integral of the following:

a $\sin 3x\cos^2 3x$ b $\frac{\sin 2x}{5 + \cos 2x}$ c $\frac{\arctan x}{x^2+1}$.

a Let $u = \cos 3x \Rightarrow \frac{du}{dx} = -3\sin 3x \therefore dx = -\frac{1}{3\sin 3x}du$.

Substituting, we have:

$$\int \sin 3x \cos^2 3x dx = \int \cancel{\sin 3x} u^2 \times -\frac{1}{\cancel{\sin 3x}} du$$
$$= -\frac{1}{3}\int u^2 du$$
$$= -\frac{1}{3}\cdot\frac{1}{3}u^3 + c$$
$$= -\frac{1}{9}\cos^3 3x + c$$

b Let $u = 5 + \cos 2x \Rightarrow \frac{du}{dx} = -2\sin 2x \therefore dx = -\frac{1}{2\sin 2x}du$.

Substituting, we have $\int \frac{\sin 2x}{5+\cos 2x}dx = \int \frac{\cancel{\sin 2x}}{u} \times -\frac{1}{2\cancel{\sin 2x}}du$

$$= -\frac{1}{2}\int \frac{1}{u}du$$
$$= -\frac{1}{2}\ln u + c$$
$$= -\frac{1}{2}\ln(5+\cos 2x) + c$$

c Let $u = \arctan x \Rightarrow \frac{du}{dx} = \frac{1}{1+x^2} \therefore dx = (1+x^2)du$.

Substituting, we have:

$$\int \frac{\arctan x}{x^2+1}dx = \int \frac{u}{x^2+1}\times(1+x^2)du = \int u du$$
$$= \frac{1}{2}u^2 + c$$
$$= \frac{1}{2}(\arctan x)^2 + c$$

Example 6.7.8

Evaluate:

a $\int_1^2 xe^{x^2}dx$ b $\int_1^3 \sqrt{2x+3}dx$

c $\int_0^2 \sqrt{4-x^2}dx$.

When using the substitution method to evaluate a definite integral, it is generally more efficient to transform the terminals (limits) of the integral as well as the integrand. This process is illustrated by the following examples.

a This is solved using the substitution $u = x^2, \frac{du}{dx} = 2x$

The integrand is transformed to:

$$\int xe^{x^2}dx = \int \frac{1}{2}e^u du = \frac{1}{2}e^{x^2} + c$$

Having established that the substitution will work, we can now use it to transform the terminals.

The lower terminal is $x = 1 \Rightarrow u = 1^2 = 1$ and the upper terminal is $x = 2 \Rightarrow u = 2^2 = 4$.

Thus: $\int_1^2 xe^{x^2}dx = \int_1^4 \frac{1}{2}e^u du = \frac{1}{2}[e^u]_1^4 = \frac{1}{2}(e^4 - e)$

b Use $u = 2x+3, \frac{du}{dx} = 2$ and:

$x = 1 \Rightarrow u = 5, x = 3 \Rightarrow u = 9$

$$\int_1^3 \sqrt{2x+3}dx = \int_5^9 \frac{1}{2}u^{1/2}du = \frac{1}{2}\left[\frac{2}{3}u^{3/2}\right]_5^9$$
$$= \frac{1}{3}(9^{3/2} - 5^{3/2})$$
$$= \frac{1}{3}(27 - 5\sqrt{5})$$

c Let $x = 2\sin\theta, \frac{dx}{d\theta} = 2\cos\theta$.

The terminals transform to:

$x = 0 \Rightarrow 0 = 2\sin\theta \Rightarrow \theta = 0$

$x = 2 \Rightarrow 2 = 2\sin\theta \Rightarrow \theta = \frac{\pi}{2}$.

$$\therefore \int_0^2 \sqrt{4-x^2}dx = \int_0^{\frac{\pi}{2}} \sqrt{4-4\sin^2\theta}\times 2\cos\theta d\theta$$
$$= \int_0^{\frac{\pi}{2}} 2\sqrt{1-\sin^2\theta}\times 2\cos\theta d\theta$$
$$= 2\int_0^{\frac{\pi}{2}} 2\cos^2\theta d\theta$$
$$= 2\int_0^{\frac{\pi}{2}} (1+\cos 2\theta)d\theta$$
$$= 2\left[\theta + \frac{1}{2}\sin 2\theta\right]_0^{\frac{\pi}{2}}$$
$$= \pi$$

Exercise 6.7.2

1. Find the following, using the given u substitution.

a $\int 2x\sqrt{x^2+1}\,dx$, $u = x^2+1$

b $\int 3x^2\sqrt{x^3+1}\,dx$, $u = x^3+1$

c $\int 2x^3\sqrt{4-x^4}\,dx$, $u = 4-x^4$

d $\int \frac{3x^2}{x^3+1}dx$, $u = x^3+1$

e $\int \frac{x}{(3x^2+9)^4}dx$, $u = 3x^2+9$

f $\int 2xe^{x^2+4}dx$, $u = x^2+4$

g $\int \frac{2z+4}{z^2+4z-5}dz$, $u = z^2+4z-5$

2. Using the substitution method, find:

a $\int x\sqrt{2x-1}\,dx$ b $\int x^2\sqrt{1-x}\,dx$

c $\int (x+1)\sqrt{x-1}\,dx$ d $\int \sec^2 x e^{\tan x}dx$

3. Using an appropriate substitution, evaluate the following, giving exact values.

a $\int_{-1}^{1} \frac{2x}{x^2+1}dx$ b $\int_{0}^{1} \frac{2x^2}{x^3+1}dx$

c $\int_{10}^{12} \frac{2x+1}{x^2+x-2}dx$ d $\int_0^{\frac{\pi}{2}} \frac{\cos x}{1+\sin x}dx$

4. Using an appropriate substitution, evaluate the following, giving exact values.

a $\int_1^2 x\sqrt{x^2+3}\,dx$ b $\int_0^{\frac{\pi}{2}} 3x\sin(4x^2+\pi)dx$

c $\int_{-1}^{1} (3x+2)^4 dx$ d $\int_{-2}^{1} \frac{1}{x+3}dx$

5. Using an appropriate substitution, find the following, giving exact values where required.

a $\int_0^{\frac{\pi}{2}} \sin^3 x\cos x\,dx$ b $\int_{\frac{\pi}{6}}^{\frac{\pi}{3}} \sin x\sec^2 x\,dx$

c $\int_0^{\frac{\pi}{3}} \cos^3 x\sin 2x\,dx$ d $\int_0^{\frac{\pi}{3}} \frac{\sin 2x}{\sqrt{\cos^3 x}}dx$

6. Using an appropriate substitution, find the following, giving exact values where required.

a $\int_{-2}^{-1} x\sqrt{x+2}\,dx$ b $\int_{-1}^{2} x\sqrt{2-x}\,dx$

7. Find the following indefinite integrals.

a $\int \frac{1}{x^2+6x+10}dx$ b $\int \frac{1}{x^2-x+1}dx$

c $\int \frac{1}{\sqrt{1+4x-x^2}}dx$ d $\int \frac{3}{\sqrt{8-2x-x^2}}dx$

8. Given that $\frac{2-x^2}{(x^2+1)(x^2+4)} \equiv \frac{A}{x^2+1} + \frac{B}{x^2+4}$, find A and B.

Hence show that $\int_0^1 \frac{2-x^2}{(x^2+1)(x^2+4)}dx = arctan\left(\frac{1}{3}\right)$.

9. Find $\int_0^k \frac{1}{x^2+1}dx$, $k > 0$.

Evaluate the definite integral in part a for:

i $k = \frac{1}{\sqrt{3}}$ ii $k = 1$.

Find $\lim_{k \to \infty}\int_0^k \frac{1}{x^2+1}dx$. Hence, find $\int_{-\infty}^{\infty} \frac{1}{x^2+1}dx$.

10. Find $\int \frac{1}{\sqrt{x}+1}dx$. Hence evaluate $\int_0^1 \frac{1}{\sqrt{x}+1}dx$.

11. If $z = cis\theta$, use the expansion of $\left(z-\frac{1}{z}\right)^4$ to show that $8\sin^4\theta = \cos 4\theta - 4\cos 2\theta + 3$.

Hence, using the substitution $x = k\sin^2\theta$, evaluate:

$\int_0^k x\sqrt{\frac{x}{k-x}}dx$, $0 < \theta < \pi$.

Extra questions

Integration by Parts

The basics

Consider the indefinite integral $\int x\cos x\,dx$.

Applying any of the techniques we have been using so far will not help us determine the integral. Let us start the process by first finding the derivative of $x\sin x$:

$$\frac{d}{dx}(x\sin x) = \frac{d}{dx}(x)\sin x + x\frac{d}{dx}(\sin x) \text{ (using product rule)}$$

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

We observe that the term $x.\cos x$ has now appeared on the R.H.S. we can then write

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

$$\therefore \int x\cos x\,dx = \int\left[\frac{d}{dx}(x\sin x) - \sin x\right]dx$$

$$= x\sin x + \cos x + c$$

Such a process requires considerable foresight. However, this integrand falls into a category of integrands that can be antidifferentiated via a technique known as integration by parts. The method is identical to that which we have just used in determining $\int x\cos x\,dx$.

We develop a general expression for integrands that involve a product of two functions.

Step 1: Consider the product $u(x)v(x)$.

Step 2: Using the product rule for differentiation we have:

$$\frac{d}{dx}(u(x)v(x)) = u(x)\frac{dv}{dx} + v(x)\frac{du}{dx}.$$

Step 3: Integrating both sides with respect to x gives:

$$u(x)v(x) = \int u(x)\frac{dv}{dx}dx + \int v(x)\frac{du}{dx}dx$$

Step 4: Rearranging, to obtain $\int u(x)\frac{dv}{dx}dx$, we have:

$$\int u(x)\frac{dv}{dx}dx = u(x)v(x) - \int v(x)\frac{du}{dx}dx$$

In the previouscase, we would set, $u(x) = x$ and $\frac{dv}{dx} = \cos x$ and the result would then follow through.

The success of this technique is dependent on your ability to identify the 'correct' $u(x)$ and $v(x)$.

For example, had we used $u(x) = \cos x$ and $\frac{dv}{dx} = x$,

we would have the expression

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

– which is not helpful.

We now consider some examples to highlight the process involved.

Example 6.7.9

Evaluate:

a $\int x\cos x\,dx$ b $\int \frac{x}{3}e^{2x}dx$.

a $\int x\cos x\,dx$.

Applying the parts formula with $u(x) = x$ and $\frac{dv}{dx} = \cos x$, it follows that $v(x) = \sin x$ gives:

$u(x)$ Find $\rightarrow$	$\frac{du}{dx}$
x	1
$v(x)$ Find $\leftarrow$	$\frac{dv}{dx}$
$\sin x$	$\cos x$

$$\int u(x)\frac{dv}{dx}dx = u(x)v(x) - \int v(x)\frac{du}{dx}dx$$

$$\int x\cos x\,dx = x \times \sin x - \int \sin x \times 1\,dx$$

$$= x\sin x - (-\cos x) + c$$

$$= x\sin x + \cos x + c$$

You should check that this is correct by differentiating the answer.

Many people remember the 'parts formula' by thinking of the question as consisting of two parts each of which are functions of the independent variable. One of these functions is to be integrated and the other differentiated. Obviously it pays to select a function that becomes simpler in derivative form to be the 'part' that is differentiated. Often, though not always, this will be the polynomial part.

b $\quad \int \frac{x}{3}e^{2x}dx$.

In this case we choose the function to be differentiated as $u(x) = \frac{x}{3}$ and the function to be integrated as $\frac{dv}{dx} = e^{2x} \Rightarrow v(x) = \frac{1}{2}e^{2x}$

$u(x) \xrightarrow{\text{Find}}$	$\frac{du}{dx}$
$\frac{x}{3}$	$\frac{1}{3}$
$v(x) \xleftarrow{\text{Find}}$	$\frac{dv}{dx}$
$\frac{1}{2}e^{2x}$	e^{2x}

$$\int u(x)\frac{dv}{dx}dx = u(x)v(x) - \int v(x)\frac{du}{dx}dx$$

$$\int \frac{x}{3}e^{2x}dx = \frac{x}{3} \times \frac{1}{2}e^{2x} - \int \frac{1}{2}e^{2x} \times \frac{1}{3}dx$$

$$= \frac{x}{6}e^{2x} - \frac{1}{12}e^{2x} + c$$

Exercise 6.7.3

1. Integrate the following expressions with respect to x.

   a $\quad x\sin x$ $\qquad$ b $\quad x\cos\frac{x}{2}$

   c $\quad 2x\sin\frac{x}{2}$ $\qquad$ d $\quad xe^{-x}$

2. Use integration by parts to antidifferentiate:

   a $\quad x\sqrt{x+1}$ $\qquad$ b $\quad x\sqrt{x-2}$

3. Find:

   a $\quad \int \text{Cos}^{-1}x\,dx$ $\qquad$ b $\quad \int \text{Tan}^{-1}x\,dx$

4. Find:

   a $\quad \int x\text{Cos}^{-1}x\,dx$ $\qquad$ b $\quad \int x\text{Tan}^{-1}x\,dx$

5. Find:

   a $\quad \int_0^{\frac{\pi}{4}} x\sin 2x\,dx$ $\qquad$ b $\quad \int_0^1 xe^{2x}dx$

   c $\quad \int_1^{(e-1)} x\ln(x+1)dx$ $\qquad$ d $\quad \int_1^2 (x-1)\ln x\,dx$

   e $\quad \int_{\frac{\pi}{8}}^{\frac{\pi}{4}} x\cos 2x\,dx$ $\qquad$ f $\quad \int_1^e \frac{\ln x}{x}dx$

6. Find $\int_0^1 x^2\text{Tan}^{-1}x\,dx$.

7. Show that $\frac{d}{dx}[\ln(\sec x + \tan x)] = \sec x$.

   Hence find $\int_0^{\frac{\pi}{4}} \sec^3 x\,dx$.

8. Find:

   a $\quad \int \cos(\ln x)dx$

   b $\quad \int \sin(\ln x)dx$

   c $\quad \int x^3\sqrt{1-x^2}$

Extra questions

Repeated integration by parts

In Exercise 6.7.3, Question 7 required the repeated use of integration by parts. There will be occasions on which you will need to use the 'parts' formula more than once to evaluate an integral as in the following examples.

Example 6.7.10

Evaluate:

a $\int x^2 \cos 2x dx$ b $\int e^{2x} \sin \frac{x}{3} dx$.

a

$$\int x^2 \cos 2x dx = x^2 \times \frac{1}{2}\sin 2x - \int 2x \times \frac{1}{2}\sin 2x dx$$

$$\left[\text{using } u = x^2, \frac{dv}{dx} = \cos 2x\right]$$

$$= \frac{1}{2}x^2 \sin 2x - \int x \sin 2x dx$$

Repeated use of 'parts' formula.

$$= \frac{1}{2}x^2 \sin 2x - \left(x \times -\frac{1}{2}\cos 2x - \int\left(-\frac{1}{2}\cos 2x\right)dx\right)$$

$$= \frac{1}{2}x^2 \sin 2x + \frac{1}{2}x \cos 2x - \frac{1}{4}\sin 2x + c$$

b

$$\int e^{2x} \sin \frac{x}{3} dx = e^{2x} \times -3\cos\frac{x}{3} - \int 2e^{2x} \times -3\cos\frac{x}{3} dx$$

$$= -3e^{2x}\cos\frac{x}{3} - \left(-18e^{2x}\sin\frac{x}{3} - \int -36e^{2x}\sin\frac{x}{3} dx\right)$$

$$= -3e^{2x}\cos\frac{x}{3} + 18e^{2x}\sin\frac{x}{3} - 36\int e^{2x}\sin\frac{x}{3} dx$$

The required integral appears on both sides of this equation, which rearranges to:

$$37\int e^{2x}\sin\frac{x}{3} dx = -3e^{2x}\cos\frac{x}{3} + 18e^{2x}\sin\frac{x}{3}$$

$$\therefore \int e^{2x}\sin\frac{x}{3} dx = -\frac{3}{37}e^{2x}\cos\frac{x}{3} + \frac{18}{37}e^{2x}\sin\frac{x}{3} + c$$

Exercise 6.7.4

1. Find the following integrals (not all are best evaluated using the parts formula).

a $\int x^2 e^x dx$ b $\int 3x^2 \cos(2x) dx$

c $\int x^3 \ln(2x) dx$ d $\int e^x \sin(2x) dx$

e $\int x^2 \cos(3x) dx$ f $\int e^{-2x} \cos(2x) dx$

g $\int 4x^3 \sin\frac{x}{2} dx$ h $\int \frac{1}{x} \ln x dx$

i $\int (\ln(3x))^2 dx$ j $\int \cos x \sin(2x) dx$

k $\int e^{ax} \cos\frac{x}{a} dx$ l $\int x^2\sqrt{x+2} dx$

m $\int x^3 \ln(ax) dx$ n $\int \frac{x^2}{\sqrt{4-x^2}} dx$

o $\int \frac{3x^2 dx}{\sqrt{x^2-9}}$ p $\int \frac{x}{x^2+4} dx$

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

2. Evaluate the following.

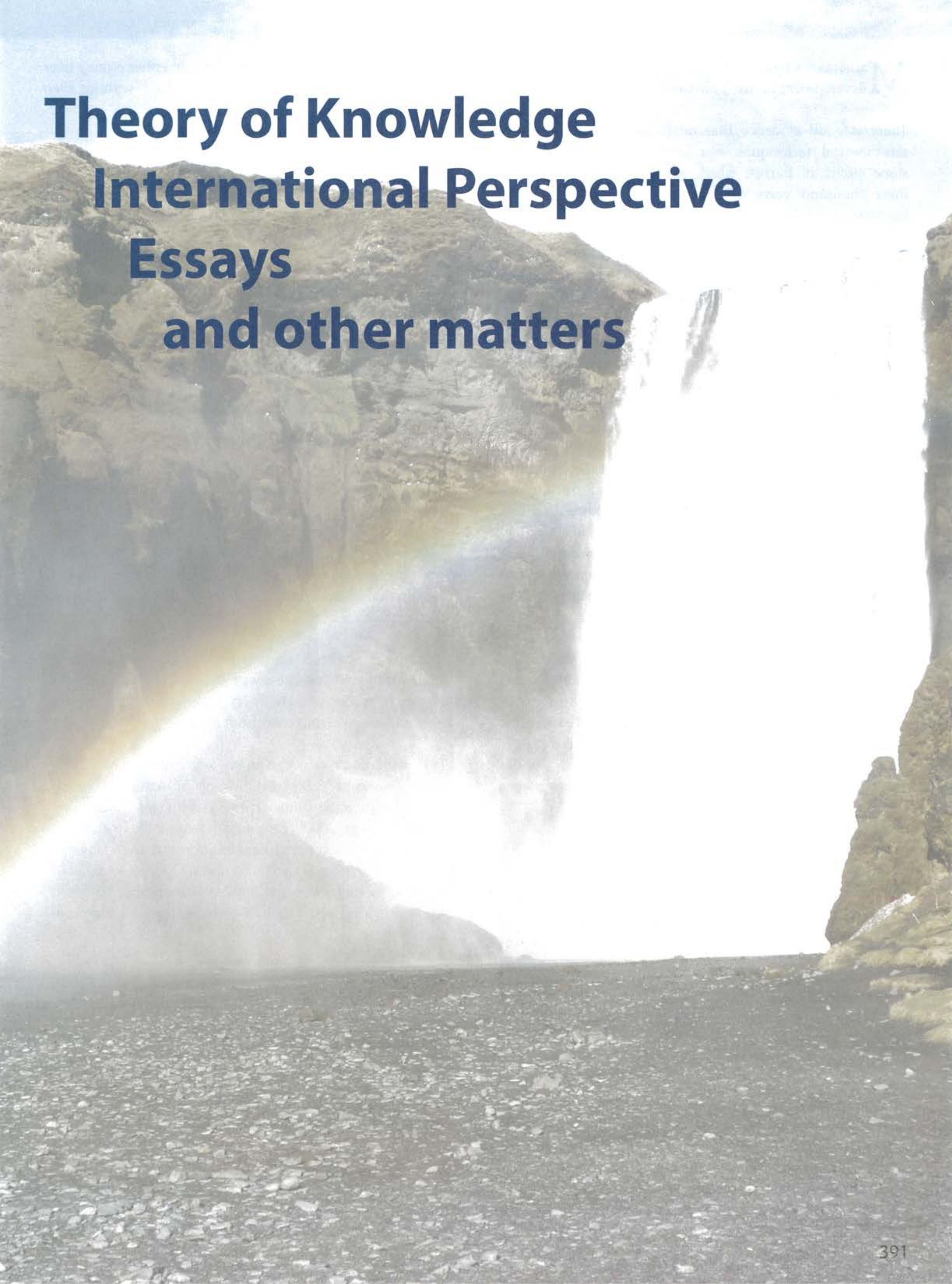
a $\int_0^{\frac{\pi}{2}} x\cos^2 x dx$ b $\int_0^{\frac{\pi}{2}} x \sin x \cos x dx$

c $\int_{\frac{\pi}{2}}^{2\pi} e^x \cos x dx$ d $\int_0^{\ln 2} x^2 e^{-x} dx$

e $\int_{\frac{\pi}{b}}^{\frac{2\pi}{b}} e^{ax} \cos bx dx$ f $\int_1^e (\ln x)^2 dx$

Answers

Theory of Knowledge International Perspective Essays and other matters



Mathematics has clearly played a significant part in the development of many past and present civilisations.

There is good evidence that mathematical, and probably astronomical techniques, were used to build the many stone circles of Europe which are thought to be at least three thousand years old (Thom). It is likely that the Egyptian pyramids and constructions on Aztec and Mayan sites in Central America were also built by mathematically sophisticated architects. Similarly, cultures in China, India and throughout the Middle East developed mathematics a very long time ago. It is also the case that there have been very successful cultures that have found little use for mathematics. Ancient Rome, handicapped, as it was, by a non-place value number system did not develop a mathematical tradition at anything like the same level as did Ancient Greece. Also, the Australian Aborigines, who have one of the most long-lasting and successful cultures in human history, did not find much need for mathematical methods. The same is true of many aboriginal cultures of Africa, Asia and the Americas. This may well be because these aboriginal cultures did not value ownership in the way that western culture does and had no need to count their possessions. Instead, to aboriginal cultures, a responsible and sustainable relationship with the environment is more important than acquisition and exploitation. Maybe we should learn from this before it is too late!

Mathematics has developed two distinct branches: pure mathematics, which is studied for its own sake, and applied mathematics which is studied for its usefulness. This is not to say that the two branches have not cross-fertilised each other, for there have been many examples in which they have.

The pure mathematician Pierre de Fermat (1601–1665) guessed that the equation $x^n + y^n = z^n$ has whole numbered solutions for $n = 2$ only. To the pure mathematician, this type of problem is interesting for its own sake. To study it is to look for an essential truth, the 'majestic clockwork' of the universe. Pure mathematicians see 'beauty' and 'elegance' in a neat proof. To pure mathematicians, their subject is an art.

Applied mathematics seeks to develop mathematical objects such as equations and computer algorithms that can be used to predict what will happen if we follow a particular course of action. This is a very valuable capability. We no longer build bridges without making careful calculations as to whether or not they will stand. Airline pilots are able to experience serious failures in commercial jets without either risking lives or the airline's valuable aeroplanes or, indeed, without even leaving the ground.

Silk, rabbits and Pisa

The term 'Silk Road' is applied to a network of trade routes linking China and the Spice Isles (now Indonesia) through India and Arabia to Africa and Europe. Traders and their products have been passing along these routes on both land and sea for millennia. It was not just silk and spices and other goods that travelled the Silk Road. It is virtually certain that ideas, games, folk tales etc. also travelled with the traders.

This makes it difficult to attribute inventions with certainty. It appears likely that chess was invented in India and was carried by traders as a good way of passing the evenings in a stimulating way, but we cannot be sure.

It is virtually certain that many key mathematical ideas passed along the Silk Road. Schools of mathematics that were using a place-value decimal system were flourishing in China over 2000 years ago. It seems likely that the decimal number system (including zero) we use today was developed in India in the 2nd century. This spread and had reached Persia by the year 800. Al-Khwarizmi's book On the Calculation with Hindu Numerals appeared at around this time.

Leonardo of Pisa (c1170 – c1250), known as Fibonacci, was a trader. He saw that Arab traders using a place-value system for their calculations found them easier than Europeans who used Roman numerals. Leonardo travelled the Mediterranean studying the work of Arab mathematicians. The result was the book Liber Abaci in which Fibonacci introduced the modus Indorum (method of the Indians) to a European audience.

[continued]

Many, for reasons best known to themselves continued to use Roman numeration. The diarist Samuel Pepys (1633–1703), who held the very important job of running England's Royal Navy, made extensive use of Roman numerals. However, the superiority of what has become known as the Hindu–Arabic system became evident and now is a 'universal language'.

And what of the rabbits? Fibonacci was interested in modelling the population of rabbits assuming that they spend one year as non-reproducing infants. The result is the sequence that bears his name: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55... Why this sequence relates to rabbits is, as they say, left to the reader.

Axioms

Mathematics is based on axioms. These are 'facts' that are assumed to be true. An axiom is a statement that is accepted without proof. Early sets of axioms contained statements that appeared to be obviously true. Euclid postulated a number of these 'obvious' axioms.

'Things equal to the same thing are equal to each other';
That is, if $y = a$ and $x = a$ then $y = x$.

Euclid was mainly interested in geometry and we still call plane geometry 'Euclidean'. In Euclidean space, the shortest distance between two points is a straight line. We will see later that it is possible to develop a useful, consistent mathematics that does not accept this axiom.

Most axiom systems have been based on the notion of a 'set', meaning a collection of objects. An example of a set axiom is the 'axiom of specification'. In crude terms, this says that if we have a set of objects and are looking at placing some condition or specification on this set, then the set thus specified must exist. We consider some examples of this axiom.

Assume that the set of citizens of China is defined. If we impose the condition that the members of this set must be female, then this new set (of Chinese females) is defined.

As a more mathematical example, if we assume that the set of whole numbers exists, then the set of even numbers (multiples of 2) must also exist.

A second example of a set axiom is the 'axiom of powers'.

For each set, there exists a collection of sets that contains amongst its elements all the subsets of the original set. If we look at the set of cats in Bogotá, then there must be a set that contains all the female cats in Bogotá, another that contains all the cats with green eyes in Bogotá, another that contains all the Bogotá cats with black tails etc. A good, but theoretical, account of axiomatic set theory can be found in Halmos, 1960.

Mathematics has, in some sense, been a search for the smallest possible set of consistent axioms. In the section on paradox, we will look further at the notion of axioms and the search for a set of assumptions that does not lead to contradictions. There is a very strong sense in which mathematics is an unusual pursuit in this respect. Pure mathematics is concerned with absolute truth only in the sense of creating a self-consistent structure of thinking.

As an example of some axioms that may not seem to be sensible, consider a geometry in which the shortest path between two points is the arc of a circle and all parallel lines meet. These "axioms" do not seem to make sense in "normal" geometry. The first mathematicians to investigate non-Euclidean geometry were the Russian, Nicolai Lobachevsky (1792–1856) and the Hungarian, Janos Bolyai (1802–1860).

Independently, they developed self-consistent geometries that did not include the so called parallel postulate which states that for every line AB and point C outside AB there is only one line through C that does not meet AB.

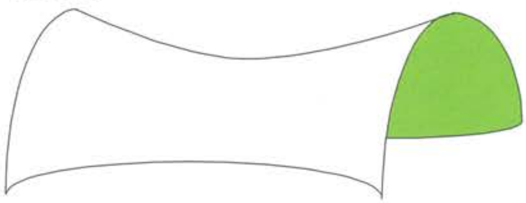
Since both lines extend to infinity in both directions, this seems to be obvious' Non-Euclidean geometries do not include this postulate and assume either that there are no lines through C that do not meet AB or that there is more than one such line. It was the great achievement of Lobachevsky and Bolyai that they proved that these assumptions lead to geometries that are self consistent and thus acceptable as 'true' to pure mathematicians. In case you are thinking that this sort of activity is completely useless, one of the two non-Euclidean geometries discussed above has actually proved to be useful; the geometry of shapes drawn on a sphere. This is useful because it is the geometry used by the navigators of aeroplanes and ships.

The first point about this geometry is that it is impossible to travel in straight lines. On the surface of a sphere, the shortest distance between two points is an arc of a circle centred at the centre of the sphere (a great circle). The shortest path from London to Delhi is circular. If you want to see this path on a geographer's globe, take a length of sewing cotton and stretch it tightly between the two cities. The cotton will follow the approximate great circle route between the two cities.

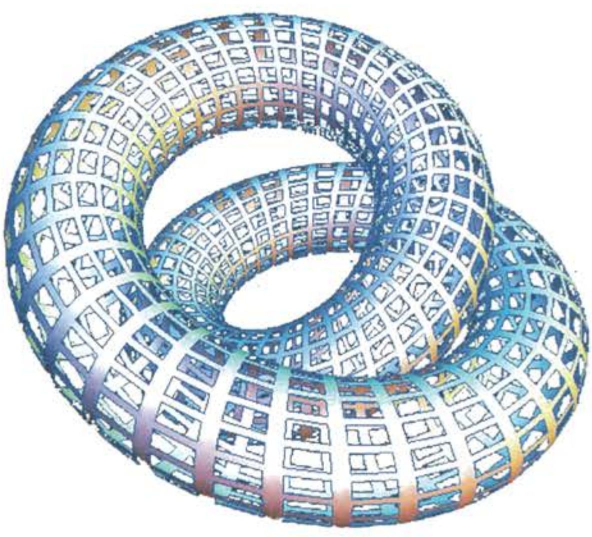
If we now think of the arcs of great circles as our 'straight lines', what kind of geometry will we get? You can see some of these results without going into any complex calculations. For example, what would a triangle look like?

The first point is that the angles of this triangle add up to more than 180°. There are many other 'odd' features of this geometry. However, fortunately for the international airline trade, the geometry is self consistent and allows us to navigate safely around the surface of the globe. Thus non-Euclidean geometry is an acceptable pure mathematical structure.

While you are thinking about unusual geometries, what are the main features of the geometry of shapes drawn on the 'saddle surface'?



Or the even more exotic 'toroid':



One final point on the subject of non-Euclidean geometries; it seems to be the case that our three-dimensional universe is also curved. This was one of the great insights of Albert Einstein (1879–1955). We do not yet know if our universe is bent back on itself rather like a sphere or whether another model is appropriate. A short account of non-Euclidean Geometries can be found in Cameron (pp. 31–40).

By contrast, applied mathematics is judged more by its ability to predict the future, than by its self-consistency. Applied mathematics is also based on axioms, but these are judged more on their ability to lead to calculations that can predict eclipses, cyclones, whether or not a suspension bridge will be able to support traffic loads, etc. In some cases such mathematical models can be very complex and may not give very accurate predictions. Applied mathematics is about getting a prediction, evaluating it (seeing how well it predicts the future) and then improving the model.

In summary, both branches of mathematics are based on axioms. These may or may not be designed to be 'realistic'. What matters to the pure mathematician is that an axiom set should not lead to contradictions. The applied mathematician is looking for an axiom set and a mathematical structure built on these axioms that can be used to model the phenomena that we observe in nature. As we have seen, useful axiom sets need not start out being 'sensible'.

The system of deduction that we use to build the other truths of mathematics is known as **proof**.

Numbers and the transcendental

Many mathematical words have been coined to describe mathematical ideas and objects. 'Logarithm' is an example of such a word. It is derived from the Greek word logos which means 'reckoning'. It does not have a commonly used meaning outside mathematics. However, mathematicians sometimes use everyday words to describe mathematical ideas in ways that may sometimes be confusing.

For example, the word 'prime' in everyday usage means 'of first importance, main, ...'. In mathematics, a prime number is one with exactly two factors. It is true that this makes prime numbers primarily interesting. The point is that the two meanings are not the same.

When studying this chapter on logarithms, you will have encountered Euler's number, e. This number has been shown to be irrational (like $\sqrt{2}$). However, it also has the property that it is not the solution of any polynomial equations with rational coefficients. This is a big claim given that the polynomials can have any number of terms going up to any power and that we also have an infinite choice for each coefficient. This was first proved by Charles Hermite in the 1870s.

All transcendental numbers are irrational, but not all irrational numbers are transcendental. $\sqrt{2}$, a solution to the equation $x^2 - 2 = 0$, is irrational but not transcendental. The list of known transcendental numbers is quite short (it includes π) but not, oddly, $(\pi + e)$. However, the transcendental numbers are the most numerous of all the types of real numbers.

[continued]

By now, you should be thinking that the transcendental numbers should be called the 'weird numbers' or somesuch. However, the word has been well chosen even if its mathematical meaning is different from its dictionary definition: '... of or relating to a spiritual or nonphysical realm'.

Proof

Proof has a very special meaning in mathematics. We use the word generally to mean "proof beyond reasonable doubt" in situations such as law courts when we accept some doubt in a verdict. For mathematicians, proof is an argument that has *no* doubt at all. When a new proof is published, it is scrutinised and criticized by other mathematicians and is accepted when it is established that every step in the argument is legitimate. Only when this has happened does a proof become accepted.

Technically, every step in a proof rests on the axioms of the mathematics that is being used. As we have seen, there is more than one set of axioms that could be chosen. The statements that we prove from the axioms are known as **theorems**. Once we have a theorem, it becomes a statement that we accept as true and which can be used in the proof of other theorems. In this way we build up a structure that constitutes a "mathematics". The axioms are the foundations and the theorems are the superstructure. In the previous section we made use of the idea of consistency. This means that it must not be possible to use our axiom set to prove two theorems that are contradictory.

There are a variety of methods of proof available. This section will look at three of these in detail. We will mention others.

Rules of inference

All proofs depend on rules of inference. Fundamental to these rules is the idea of 'implication'.

As an example, we can say that $2x = 4$ (which is known as a **proposition**) implies that $x = 2$ (provided that x is a normal real number and that we are talking about normal arithmetic). In mathematical shorthand we would write this statement as $2x = 4 \Rightarrow x = 2$.

This implication works both ways because $x = 2$ implies that $2x = 4$ also. This is written as $x = 2 \Rightarrow 2x = 4$ or the fact that the implication is both ways can be written as $x = 2 \Leftrightarrow 2x = 4$. The $\Leftrightarrow$ symbol is read as '**If and only if**' or simply as '**Iff**', i.e. If with two fs.

Not every implication works both ways in this manner. If $x = 2$ then we can conclude that $x^2 = 4$. However, we cannot conclude the reverse, i.e. $x^2 = 4$ implies that $x = 2$ is false because x may be -2.

So that $x = 2 \Rightarrow x^2 = 4$ is all that can be said in this case.

There are four main rules of inference:

The rule of detachment: from a is true and $a \Rightarrow b$ is true we can infer that b is true. a and b are propositions.

If the following propositions are true:

It is raining.

If it is raining, I will take an umbrella.

We can infer that I will take an umbrella.

The rule of syllogism: from $a \Rightarrow b$ is true and $b \Rightarrow c$ is true, we can conclude that $a \Rightarrow c$ is true. a, b & c are propositions.

If we accept as true that:

if x is an odd number then x is not divisible by 4 ($a \Rightarrow b$)and,

if x is not divisible by 4 then x is not divisible by 16 ($b \Rightarrow c$)

We can infer that the proposition;

if x is an odd number then x is not divisible by 16 ($a \Rightarrow c$) is true.

The rule of equivalence: at any stage in an argument we can replace any statement by an equivalent statement.

If x is a whole number, the statement x is even could be replaced by the statement x is divisible by 2.

The rule of substitution: If we have a true statement about all the elements of a set, then that statement is true about any individual member of the set.

If we accept that all lions have sharp teeth then Benji, who is a lion, must have sharp teeth.

Now that we have our rules of inference, we can look at some of the most commonly used methods of proof.

Proof by exhaustion

This method can be, as its name implies, exhausting! It depends on testing every possible case of a theorem.

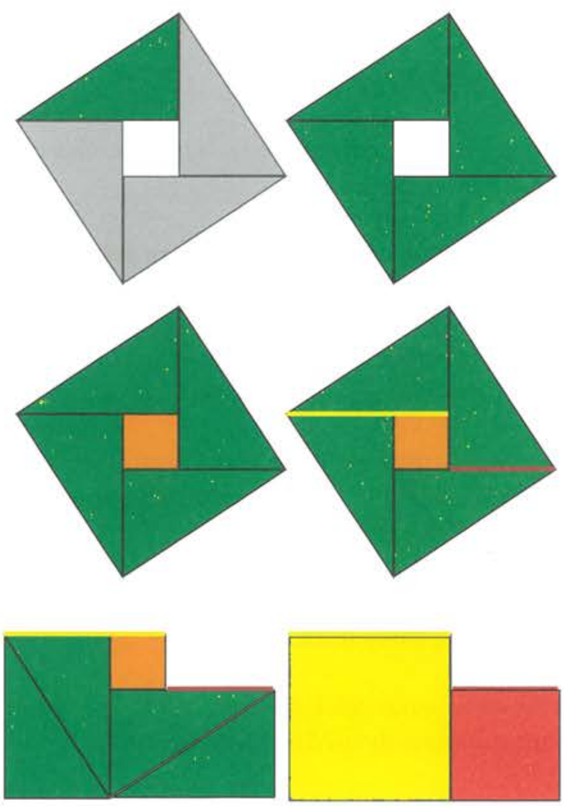
Consider the theorem: Every year must contain at least one 'Friday the thirteenth'.

There are a limited number of possibilities as the first day of every year must be a Monday or a Tuesday or a Wednesday.... or a Sunday (7 possibilities). Taking the fact that the year may or may not be a leap year (with 366 days) means that there are going to be fourteen possibilities.

Once we have established all the possibilities, we would look at the calendar associated with each and establish whether or not it has a 'Friday the thirteenth'. If, for example, we are looking at a non-leap year in which January 1st is a Saturday, there will be a 'Friday the thirteenth' in May. Take a look at all the possibilities (an electronic organiser helps!). Is the theorem true?

Direct proof

The diagrams on the following page represent a proof of the theorem of Pythagoras described in *The Ascent of Man* (Bronowski, pp. 158–161). The theorem states that the area of a square drawn on the hypotenuse of a right-angled triangle is equal to the sum of the areas of the squares drawn on the two shorter sides. The method is direct in the sense that it makes no assumptions at the start. Can you follow the steps of this proof and draw the appropriate conclusion?



Proof by contradiction

This method works by assuming that the proposition is false and then proving that this assumption leads to a contradiction.

The number $\sqrt{2}$ greatly interested classical Greek mathematicians who were unable to find a number that, when it was squared, gave exactly 2.

Modern students are often fooled into thinking that their calculators give an exact square root for 2 as when 2 is entered and the square root button is pressed, a result (depending on the model of calculator) of 1.414213562 is produced. When this is squared, exactly 2 results – but not because we have an exact square root. It results from the way in which the calculator is designed to calculate with more figures than it actually displays.

The first answer is stored to more figures than are shown, the result is rounded and then displayed. The same is true of the second result which only rounds to 2. Try squaring 1.414213562, the answer is not 2.

The theorem we shall prove is that there is *no* fraction that when squared gives 2. This also implies that there is no terminating or recurring decimal that, when squared, gives exactly 2, but this further theorem requires more argument.

The method begins by assuming that there *is* a fraction p/q (p and q are integers) which has been cancelled to its lowest terms, such that $p/q = \sqrt{2}$. From the assumption, the argument proceeds:

$$\frac{p}{q} = \sqrt{2} \Rightarrow \frac{p^2}{q^2} = 2 \Rightarrow p^2 = 2q^2 \Rightarrow p^2 \text{ is even} \Rightarrow p \text{ is even}$$

As with most mathematical proofs, we have used simple axioms and theorems of arithmetic. The most complex theorem used is that if p^2 is even, then p is even. Can you prove this?

The main proof continues with the deduction that if p is even there must be another *integer*, r, that is half p.

$$p = 2r \Rightarrow p^2 = 4r^2 \Rightarrow 2q^2 = 4r^2$$
$$\Rightarrow q^2 = 2r^2 \Rightarrow q^2 \text{ is even} \Rightarrow q \text{ is even}$$

We now have our contradiction as we assumed that p/q was in its lowest terms so p and q cannot both be even. This proves the result, because we have a contradiction.

This theorem is a very strong statement of *impossibility*.

There are very few other areas of knowledge in which we can make similar statements. We may be virtually certain that we will never travel faster than the speed of light but it would be a brave physicist who would state with certainty that it is *impossible*. Other methods of proof include proof by induction which is mainly used to prove theorems involving sequences of statements.

Whilst on the subject of proof, it is worth noting that it is much easier to disprove a statement than to prove it. When we succeed in disproving a statement, we have succeeded in proving its negation or reverse. To disprove a statement, all we need is a single example of a case in which the theorem does not hold. Such a case is known as a **counter-example**.

The theorem 'all prime numbers are odd' is false. This can be established by noting that 2 is an even prime and, therefore, is the only counter-example we need to give. By this method we have proved the theorem that 'not every prime number is odd'.

This is another example of the way in which pure mathematicians think in a slightly different way from other disciplines. Zoo-keepers (and indeed the rest of us) may be happy with the statement that "all giraffes have long necks" and would not be very impressed with a pure mathematician who said that the statement was false because there was one giraffe (with a birth defect) who has a very short neck. This goes back to the slightly different standards of proof that are required in mathematics.

Counter-examples and proofs in mathematics may be difficult to find.

Consider the theorem that every odd positive integer is the sum of a prime number and twice the square of an integer. Examples of this theorem that do work are:

$$5 = 3 + 2 \times 1^2,\ 15 = 13 + 2 \times 1^2,\ 35 = 17 + 2 \times 3^2.$$

The theorem remains true for a very large number of cases and we do not arrive at a counter-example until 5777.

Another similar "theorem" is known as the Goldbach Conjecture. Christian Goldbach (1690–1764) stated that every even number larger than 2 can be written as the sum of two primes. For example, $4 = 2 + 2$, $10 = 3 + 7$, $48 = 19 + 29$ etc. No-one has ever found a counter-example to this simple conjecture and yet no accepted proof has ever been produced, despite the fact that the conjecture is not exactly recent!

Finally, whilst considering proof, it would be a mistake to think that mathematics is a complete set of truths that has nothing which needs to be added. We have already seen that there are unproved theorems that we suspect to be true. It is also the case that new branches of mathematics are emerging with a fair degree of regularity. During this course you will study linear programming which was developed in the 1940s to help solve the problems associated with the distribution of limited resources. Recently, both pure and applied mathematics have been enriched by the development of "Chaos Theory". This has produced items of beauty such as the Mandelbrot set and insights into the workings of nature. It seems, for example, that the results of Chaos Theory indicate that accurate long-term weather forecasts will never be possible (Mandelbrot).

Counting rabbits

Mathematicians are searchers after pattern. This reflects an innate human proclivity for looking for connections even when none exist. There is nothing the tabloid press loves more than a peasant who finds a the face of the US president when they slice open a watermelon. However, most of these "connections" have no actual meaning.

Can the same be said of mathematical connections?

Here are the first few rows of what is variously called the "Chinese triangle" or "Pascal's triangle":

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

Now displace each row to the right to produce the echelon form shown below and sum the columns(only the first seven columns are complete).

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 .....
1 1 2 3 5 8 13 ...

It looks like we have the Fibonacci sequence (and the rabbits) again. How can you be certain that this is not just chance and that the pattern continues forever.

What distinguishes the true mathematician from the presidential watermelloners is that a mathematician will demand a proof. Can you supply it?

And once you have a proof, does this imply that the polynomial coefficients are really connected to the mating habits of rabbits?

Paradox

What is a paradox?

Pure mathematics is a quest for a structure that does not contain internal contradictions. A satisfactory mathematics will contain no 'nonsense'.

Consider the following proof:

Let $x = 1$ Then $x^2 - 1 = x - 1$

Try substituting $x = 1$ to check this line.

$$(x+1)(x-1) = x-1$$

Factorizing using the difference of two squares.

Dividing both sides by $x - 1$: $x + 1 = 1$

Substituting $x = 1$. .... $2 = 1$

There is obviously something wrong here as this is the sort of inconsistency that we have discussed earlier in this chapter, but what is wrong? To discover this, we must check each line of the argument for errors or faulty reasoning.

must be acceptable as we are entitled to assign a numerical value to a pronumeral.

is true because the left-hand and right-hand sides are the same if we substitute the given value of the pronumeral.

is a simple factorisation of the left-hand side.

is obtained from line 3 by dividing both sides of the equation by $x - 1$ and should be acceptable as we have 'done the same thing' to both sides of the equation.

is obtained from line 4 by substituting $x = 1$ and so should give the correct answer.

Obviously we have an unacceptable conclusion from a seemingly watertight argument. There must be something there that needs to be removed as an acceptable operation in mathematics.

The unacceptable operation is dividing both sides by $x - 1$ and then using a value of 1 for x. What we have effectively done is divide by a quantity that is zero. It is this operation that has allowed us to prove that $2 = 1$, an unacceptable result. When a **paradox** of this sort arises, we need to look at the steps of the proof to see if there is a faulty step. If there is, then the faulty step must be removed. In this case, we must add this rule to the allowed operations of mathematics:

Never divide by a quantity that is, or will become, zero. This rule, often ignored by students, has important implications for algebra and calculus.

Some paradoxes are arguments that seem to be sound but contain a hidden error and thus do not contain serious implications for the structure of mathematical logic. An amusing compilation of simple paradoxes can be found in Gardner (1982). An example is the "elevator paradox".

Why does it always seem that when we are waiting for an elevator near the bottom of a tall building and wanting to go up, the first elevator to arrive is always going down? Also, when we want to go back down, why is the first elevator to arrive always going up? Is this a real phenomenon or is it just a subjective result of our impatience for the elevator to arrive? Or is it another example of Murphy's Law – "whatever can go wrong will go wrong"?

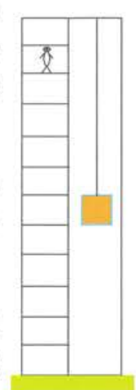
This is quite a complex question, but a simple explanation may run as follows:

If we are waiting near the bottom of a tall building, there are a small number of floors below us from which elevators that are going up may come and then pass our floor.

By contrast, there are more floors above us from which elevators may come and then pass our floor going down.

On the basis of this and assuming that the elevators are randomly distributed amongst the floors, it is more likely that the next elevator to pass will come from above and will, therefore, be going down.

By contrast, if we are waiting near the top of a tall building, there are a small number of floors above us from which elevators that are going down may come and then pass our floor.



Also, there are more floors below us from which elevators may come and then pass our floor going up.

It is more likely that the next elevator to pass will come from below and will, therefore, be going up.

A fuller analysis of this paradox can be found in Gardner (pp. 96–97).

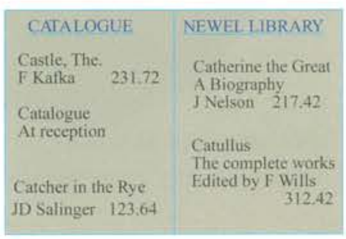
The elevator paradox does not contain serious implication for the structure of mathematics like our first example. We will conclude this section with a look at a modern paradox that did cause a re-evaluation of one of the basic ideas of mathematics, the set.

Russell's Paradox

Bertrand Russell (1872–1970) looked in detail at the basic set axioms of mathematics. We do regard the existence of sets as axiomatic in all mathematical structures. Does this mean that we can make a set that contains 'everything'? There would seem to be no difficulty with this as we just move around the universe and sweep everything that we meet into our set, numbers, words, whales, motorcycles etc. and the result is the set that contains everything.

Russell posed the following question which we will relate in the context of library catalogues.

Every library has a catalogue. There are various forms that this catalogue may take; a book, a set of cards, a computer disc etc. Whatever form the catalogue in your local library takes, there is a sense in which this catalogue is a book (or publication) owned by the library and, as such, should appear as an entry in the catalogue:



Of course, many librarians will decide that it is silly to include the catalogue as an entry in the catalogue because people who are already looking at the catalogue know where to find it in the library! It follows that library catalogues can be divided into two distinct groups:

Catalogues that do contain an entry describing themselves.

Catalogues that do not contain an entry describing themselves.

Next, let us make a catalogue of *all* the catalogues of type two, those that do not contain themselves.

This gives us a problem. Should we include an entry describing our new catalogue? If we do, then our catalogue ceases to be a catalogue of all those catalogues that do not contain themselves. If we do not, then our catalogue is no longer a complete catalogue of all those catalogues that do not contain themselves.

The conclusion is that making such a catalogue is impossible. This does not mean that the library catalogues themselves cannot exist. We have, however, defined an impossible catalogue.

In set terms, Russell's paradox says that sets are of two types:

Sets that do contain themselves.

Sets that do not contain themselves.

The set of all sets of type 2 cannot be properly defined without reaching a contradiction.

The most commonly accepted result of Russell's paradox is the conclusion that we have to be very careful when we talk about sets of everything. The most usual way out is to work within a carefully defined universal set, chosen to be appropriate to the mathematics that we are undertaking. If we are doing normal arithmetic, the universal set is the set of real numbers.

Induction

It has already been observed that the mathematical use of some words may differ from their vernacular use. In this chapter, you have seen the particular meaning that induction has in mathematics. In everyday use, it (and its related words) can have other meanings: "Babe Ruth was inducted into the Baseball Hall of Fame". "If Kylie has not had her baby by the end of the month, the doctors will induce it." "The induction stroke of a petrol engine draws a fuel/air mix into the cylinder." In Physics, when a magnet moves near an electrical conductor, an electric potential is said to be induced in the conductor. This is the principle behind the dynamo.

On a more philosophical note, you will have seen that the method of mathematical induction requires you to assume the truth of what you are trying to prove. This is a surprisingly common error in general discourse – watch out for it in your own writing as well as in the writing of others!

Mathematics and other Disciplines

When writing Theory of Knowledge essays, students are required to develop their arguments in a cross-disciplinary way. For more details on this, you are strongly advised to read the task specifications and the assessment criteria that accompany the essay title. You are reminded that it is these statements that define what is expected of a good essay, not the contents of this Chapter which have been provided as a background resource. A good essay will only result if you develop your own ideas and examples in a clear and connected manner. Part of this process may include comparing the 'mathematical method' described earlier with the methods that are appropriate to other systems of knowledge.

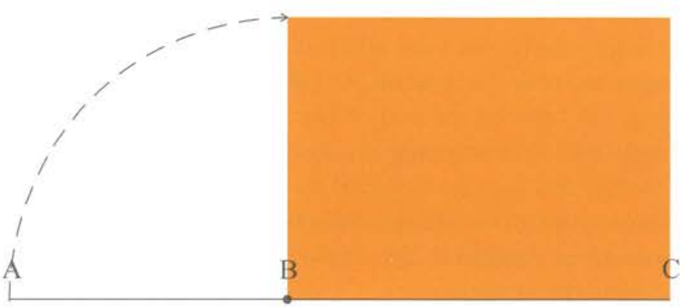
As we have seen, mathematics rests on sets of axioms. This is true of many other disciplines. There is a sense in which many ethical systems also have their axioms such as 'Thou shalt not kill'.

The Ancient Greeks believed that beauty and harmony are based, almost axiomatically, on mathematical proportions. The golden mean is found by dividing a line in the following ratio:

The ratio of the length AB to the length BC is the same as the ratio of the length BC to the whole length AC.

The actual ratio is $1{:}\frac{1}{2}(1+\sqrt{5})$ or about 1:1.618.

The Greek idea was that if this line is converted into a rectangle, then the shape produced would be in perfect proportion:

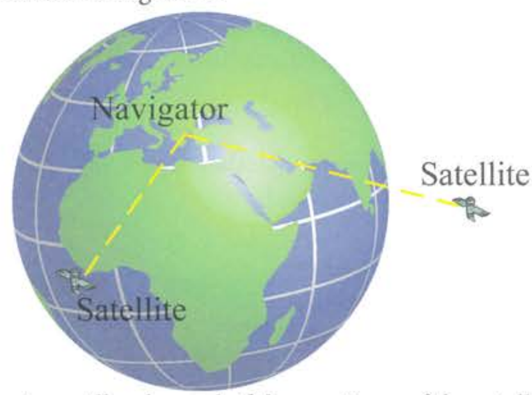


Likewise, the correct place to put the centre of interest in a picture is placed at the golden mean position between the sides and also at the golden mean between top and bottom. Take a look at the way in which television pictures are composed to see if we still use this idea.

In a similar way, the Ancient Greeks believed that ratio determined harmony in music. If two similar strings whose lengths bear a simple ratio such as 1:2 or 2:3 are plucked together the resulting sound will be pleasant (harmonious). If the ratio of string lengths is 'awkward', such as 17:19, then the notes will be discordant. The same principle of simple ratios is used in tuning musical instruments (in most cultures) today.

The most common connection between mathematics and other disciplines is the use of mathematics as a tool. Examples are: the use of statistics by insurance actuaries, probability by quality control officers and the use of almost all branches of mathematics by engineers. Every time mathematics is used in this way, there is an assumption that the calculations will be done using techniques that produce consistent and correct answers. It is here that pure mathematical techniques, applied mathematical modelling and other disciplines interface.

In some of these examples, we apply very precise criteria to our calculations and are prepared to accept only very low levels of error. Navigation satellite systems work by measuring the position of a point on or above Earth relative to the positions of satellites orbiting Earth.



This system will only work if the positions of the satellites are known with very great precision.

By contrast, when calculations are made to forecast the weather, whilst they are done with as much precision as necessary, because the data is incomplete and the atmospheric models used are approximate, the results of the calculations are, at best, only an indication of what may happen. Fortunately, most of us expect this and are much more tolerant of errors in weather forecasting than we would be if airlines regularly failed to find their destinations!

There are, therefore a large number of ways in which mathematics complements other disciplines. In fact, because computers are essentially mathematical devices and we are increasingly dependent on them, it could be argued that mathematics and its methods underpin the modern world.

That is not to say that mathematics is 'everywhere'. Many very successful people have managed to avoid the subject altogether. Great art, music and poetry has been produced by people for whom mathematical ideas held little interest.

In using mathematical ideas in essays, remember that you should produce original examples, look at them in a mathematical context and then compare the ways in which the example may appear to a mathematician with the way in which the same example may appear to a thinker from another discipline.

As a very simple example, what should we think of gambling?

To the mathematician (Pascal was one of the first to look at this activity from the mathematical perspective), a gambling game is a probability event. The outcome of a single spin of a roulette wheel is unknown. If we place a single bet, we can only know the chances of winning, not whether or not we *will* win. Also, in the long run, we can expect to lose one thirty-seventh of any money that we bet every time we play. To the mathematician, (or at least to this mathematician) this rather removes the interest from the game!

Other people look at gambling from a different standpoint. To the politician, a casino is a source of revenue and possibly a focus of some social problems. To a social scientist, the major concern may be problem gamblers and the effect that gambling has on the fabric of society. A theologian may look at the ethical issues as being paramount. Is it ethical to take money for a service such as is provided by a casino? Many of these people may use mathematics in their investigations, but they are all bringing a slightly different view to the discussion.

As we can see, there are many sides to this question as there are many sides to most questions. Mathematics can often illuminate these, but will seldom provide all the answers. When you choose an essay title, you do not have to use mathematical ideas or a mathematical method to develop your analysis. However, we hope that if you do choose to do this, you will find the brief sketch of the mathematical method described in this chapter helpful.

We will finish with one observation.

Mathematics and mathematicians are sometimes viewed as dry and unimaginative. This may be true in some cases, but definitely not all.

We conclude with some remarks by the mathematician Charles Dodgson (1832–1898), otherwise known as Lewis Carroll:

'The time has come', the Walrus said,

'To talk of many things:

Of shoes and ships and sealing wax,

Of cabbages and kings,

Of why the sea is boiling hot

And whether pigs have wings'.

Through the Looking Glass

Essays

We would like to encourage students to consider Mathematics as a choice of subject for their extended essay.

Whilst there is a requirement that these have solid academic content, it is not necessary to record an original discovery to produce an excellent essay! That said, many of the great original discoveries of mathematics are the work of comparatively young individuals with a modest level of 'experience'.

An excellent example is Evariste Galois who struggled to enter university and whose life was cut short by a duel in 1832 at age 21. Galois left a set of 'memoirs', many of which were written on the night before the duel, that are regarded as some of the most original ideas ever contributed to mathematics.

It is fashionable today to regard this sort of original thought as the preserve of 'experts'. We assert that it is not and encourage all our students to believe that they are capable of original ideas and hope that, if they do have a new idea, they have the courage to explore it.

Students may choose to look at some of the many simply-stated but as yet unproved conjectures of mathematics:

1 There are an infinite number of prime numbers. Pairs of primes such as 5 & 7, 11 & 13 that are separated by one even number are called 'twin primes'. How many twin primes are there?

2 The Goldbach conjecture: 'Every even natural number greater than 2 is equal to the sum of two prime numbers' remains unproved.

3 If an infinite number of canon-balls are stacked in an infinite pyramid, what is the biggest proportion of the space they can fill?

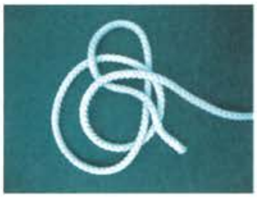
4 What are Mersenne primes and can you find a new one? The 25th & 26th Mersenne primes were found by high school students.

... and a lot more!

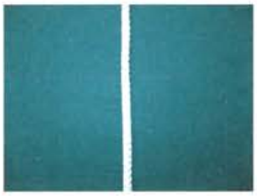
As a short case study, we will outline the work of a student who undertook a mathematical essay. The topic she chose was "The Mathematics of Knots".

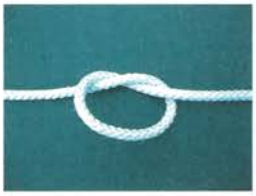
To begin with, **the student displayed an understanding** of two of the major "ways of thinking" that are characteristic of mathematicians: EXISTENCE and CLASSIFICATION.

EXISTENCE means developing tests for when a knot does or does not exist. It is neither possible nor appropriate to explore all these ideas in an introduction such as this, but the essentials are illustrated below.



These look similar, but if the ends are pulled apart, the results are quite different:

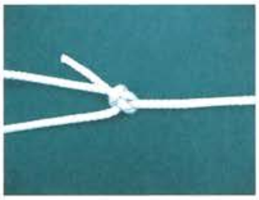




The left-hand arrangement was a tangle whereas the right was a knot.

The student investigated and skillfully explained the tests that can be applied to such rope arrangements to determine if a knot EXISTS.

Sheet bend



Bowline

These knots have two different uses. The sheet bend is used to join two lengths of rope and is designed not to slip. The bowline (pronounced "bo-lin") is the knot you tie around your waist if you are drowning and a rescuer throws you a rescue rope. It will not slip off or tighten around you and, irrespective of mathematics, is well worth knowing!

These knots have different uses and belong to different CLASSES of knot. There are knots similar to the bowline that are intended to slip (such as the noose) that belong to yet another class of knots. What are the classes of knots and what are their mathematical characteristics?

Complex numbers

The term 'imaginary' is often applied to the 'i' part of a complex number. To what extent are all numbers imaginary?

It is a common misunderstanding to think that four cars parked in a line (which are indisputably 'real') are the same as the more abstract construct 'four', which does not only apply to the process of counting four cars. Four has a life of its own without needing to be used to count things.

Moving on to the rather more challenging and very significant invention of zero (which cannot be used for counting), we could look at its invention as necessary to solve some equations in counting numbers. An example is: Solve $x + 7 = 7$. Similarly, it could be argued that it was necessary to invent negative numbers to solve equations such as: $x + 10 = 7$. If the coefficients in the equation are 'real' surely the solution must be every bit as real?

In a similar way, the rational numbers are the solutions to equations using counting numbers: $2x = 7$ and so must be as 'real' as 2 and 7.

So what about the irrational numbers? These could be viewed as the solutions to equations in numbers that we have already discovered and think of as 'real': $x^2 = 7$. Can these be any less 'real' than the counting numbers?

And now we reach the question, 'Just how real are the solutions of equations such as: $x^2 + 7 = 0$'. These numbers are no more or less imaginary than any of the other number we use in everyday life.

The question arises: just how far do we need to go in inventing new numbers to solve equations? This question was answered by the Irish mathematician William Hamilton. Hamilton was walking along a canal towpath with his wife (other accounts say his dog) when he had a 'eureka moment 'and saw the solution. He then scribbled this on a bridge lest he forget it. Hamilton's answer will have to wait until you get to university – but it is worth waiting for. The rest of the story may be just 'blarney'.

\References:

Megalithic Sites in Britain, Thom, A. (1967). U.K. Oxford University Press.

Heritage Mathematics, Cameron, M. (1984). U.K. E.J. Arnold.

The Ascent of Man, Bronowski, J. (1973). U.K. BBC.

The Fractal Geometry of Nature, Mandelbrot, B. (1977). U.S.A. W.H. Freeman & Co.

Gotcha!, Gardner, M. (1977). U.S.A. W.H. Freeman & Co.

Index

T

U

V

Z