

IB Mathematics Analysis and Approaches

Course Companion

Liam McQuay

Contents

0	About the course companion	3
1	Preliminaries	8
1.1	Symbols	8
1.2	Number systems	10
1.3	Approximations	13
1.4	Counting principles [HL]	19
2	Sequences and Series	28
2.1	Arithmetic sequences and series	28
2.2	Geometric sequences and series	34
2.3	Infinite geometric series	40
2.4	Applications: simple and compound interest	43
3	Polynomials	47
3.1	Lines	47
3.2	Quadratics	54
3.3	The binomial theorem	63
3.4	Polynomial theorems [HL]	68
4	Functions	73
4.1	Exponents and logarithms	73
4.2	Properties of functions	79
5	Appendix A	88
6	Bibliography	94

0 About the course companion

Overview

This course companion covers all of IB Mathematics Analysis and Approaches at both SL and HL. This is **not** your typical textbook. What most textbooks do is teach the content, provide exam tips, worked examples, exercises and maybe a sample exam-style paper. This book does all that, but also includes extra material outside the syllabus to enhance your learning (and provide further context for syllabus content when needed). Extra content will be clearly indicated so you know what is and isn't in the syllabus. However, I highly recommend going through all the extra content anyway. That being said, this textbook is recommended to those who have a great interest in mathematics and would like to learn more than just what the course teaches them.

This book also approaches the content more formally than other textbooks. It will contain mathematical proofs for some of the major theorems to provide you both reasoning for what you learn as well as an idea of what formal proofs look like in University mathematics and beyond. From my personal experience learning mathematics, I found most textbooks were only somewhat helpful and didn't fully explain concepts the best way. I am aware of how most students understand mathematics and will do my best to present the content as such.

The Analysis and Approaches (AA) content is covered for both SL and HL students. Throughout the book, I have indicated whether content is HL content (and labelled such sections **HL**). If a section doesn't have any indication of this kind, it will be SL content too and you'll need to know it. Note that the textbook will be in an order in which you can follow linearly from start to finish. This ensures all (or most) prerequisites are covered before you move on to the next topic.

Moreover, there will be sections labelled **FC** for **Further Content**. This will typically be out-of-syllabus topics which may be useful. I've done this because unfortunately I'm not very happy with the lack of rigor provided in the AA course (and I certainly feel the same way about the AI course too). I recommend you look into this content to further your learning about mathematics. If you do want to skip it however, you can.

Key Points

Any key points (formulas, equations, identities or equivalent) which need to be displayed for clarity shall be displayed like so:

Euler's identity

$$e^{i\pi} + 1 = 0$$

Worked Examples

The book will have a number of worked examples per chapter which span the types of short-answer questions you may get in an examination and how a candidate can respond to such questions. Worked examples look like this:

Worked example 0.0.0

An arithmetic sequence $\{u_n\}$ has first three terms: $u_1 = 1, u_2 = 5, u_3 = 9$.

(a) Write down the common difference

(b) Hence find u_8

(a) The common difference, d , is obtained by subtracting one term from the previous:

$$d = u_3 - u_2 = 9 - 5 = 4 \text{ or equivalently } d = u_2 - u_1 = 5 - 1 = 4$$

(b) Using the fact that $u_n = u_1 + d(n - 1)$, we can simply substitute the values of d and n :

$$u_8 = 1 + 4(8 - 1) = 1 + 4 \cdot 7 = 1 + 28 = 29$$

You are encouraged to find other methods of answering these questions (if any) in order to determine and use the methods which are best for you.

Proofs

As stated, this book will contain mathematical proofs for most of the major theorems encountered in the course. Most of these you will not need to know. However, you are encouraged to read these proofs as they will help you better understand the content you are learning. It's one thing having formulas given to you in a booklet where you know how and when to use them, but another if you don't **understand** them. Proofs look like this:

Theorem. $\sqrt{2}$ is irrational

Proof. Suppose $\sqrt{2}$ is rational (i.e. $\sqrt{2} \in \mathbb{Q}$).

Then there exist $m, n \in \mathbb{Z}$ such that $\sqrt{2} = \frac{m}{n}$ is irreducible (i.e. $\gcd(m, n) = 1$), as per the definition of a rational number.

Squaring both sides yields: $2 = \frac{m^2}{n^2}$, which implies $m^2 = 2n^2$. (1)

This means m^2 is even and thus that m must be even (since an odd number squared is odd).

m can hence be expressed as $2k_1$ for some $k_1 \in \mathbb{Z}$. Putting m back into equation (1) yields:

$(2k_1)^2 = 2n^2 \implies 4k_1^2 = 2n^2 \implies n^2 = 2k_1^2$, which means n^2 must be even and hence n must be even.

Thus n can be expressed as $2k_2$ for some $k_2 \in \mathbb{Z}$.

Now $m = 2k_1$ and $n = 2k_2$ and hence, $\sqrt{2} = \frac{2k_1}{2k_2}$, which is clearly reducible by cancelling the 2s from the numerator and denominator, implying $\frac{m}{n}$ is reducible.

This is a contradiction to the definition of rational numbers, as an irreducible fraction can't be found. Hence, $\sqrt{2}$ is irrational. ■

Notice the ■ symbol at the bottom-right corner. This symbol means *Quod Erat Demonstrandum* (Latin for “what was to be shown”). Mathematicians use this as a way of telling a reader that their proof is finished. If you need to give a proof for an examination, you can put this at the end of your response as a formality, but this isn't required.

Exercises

Most textbooks include drill exercises at the end of each chapter. This book will have a limited number of drill exercises at the end of each chapter section and exam-style questions and/or word problems as exercises at the end of each chapter. Each exercise will have a difficulty level indicated:

 = Easy  = Medium  = Hard  = Out of syllabus

The out of syllabus questions will naturally be more challenging. However, you are advised to attempt them to the best of your ability.

Further Reading

You will also find boxes with further information outside the syllabus about the content of the book. This may discuss history, some interesting mathematical fact(s) or something else. You are **not** expected to know them for the exams. These will look like this:

Further reading: Peano's axioms

An **axiom** in mathematics is a statement which we assume to be true (without proof) so we can use it as a basis for further arguments made. An example includes *Peano's Axioms* (by mathematician *Giuseppe Peano*), which define the set of natural numbers \mathbb{N} . Define 0 to be unity. Define a *successor* function $S(n)$ such that $1 = S(0), 2 = S(1) = S(S(0)), \dots$

Then Peano's Axioms are:

1. 0 is a natural number*
2. For any natural number n , $S(n)$ is also a natural number
3. If m and n are natural numbers such that $m = n$, then $S(m) = S(n)$
4. There is no such natural number n for which $S(n) = 0$ *
5. If K is a subset of \mathbb{N} which contains 0 and the successor of every number in K , then $K = \mathbb{N}$ (induction axiom)*

* - Note that some start the natural numbers at 1. If so, one can simply define 1 to be unity and make the necessary amendments to the axioms.

The axioms above define the nature of natural numbers and are a basis for arithmetic in \mathbb{N} . For any math we do, we have justification for our steps and justification for our justification. However, if we keep trying to find justification for anything, there has to be a point where we must stop; we encounter statements which have no justification. If these unjustifiable statements break down (somehow), all of math would break down.

Calculators

Due to overlap with the two courses, this textbook will **not** indicate questions when a calculator (GDC) is disallowed. Generally, calculator use is allowed for any question in the textbook, but I advise you to try questions without a calculator first and use it if it seems appropriate.

1 Preliminaries

This chapter recaps some information on symbols, numbers and approximations which you may or may not be familiar with. You should also be aware of the constants π and e and be able to find and use them on your calculator. This chapter also looks at counting principles for HL students.

1.1 Symbols

The list of symbols used in this book and the IB is given below. This does **not** cover everything; only symbols which you must be familiar with before learning the course. You can find the full notation list in the course guide.

Set symbols

Symbol	Meaning	Example (if applicable)
\mathbb{N}	The set of natural numbers	-
\mathbb{Z}	The set of integers	-
\mathbb{Z}^+	The set of positive integers	-
\mathbb{Z}^-	The set of negative integers	-
\mathbb{Q}	The set of rational numbers	-
\mathbb{Q}^+	The set of positive rational numbers	-
\mathbb{Q}^-	The set of negative rational numbers	-
\mathbb{R}	The set of real numbers	-
\mathbb{R}^+	The set of positive real numbers	-
\mathbb{R}^-	The set of negative real numbers	-
\mathbb{C}	The set of complex numbers	-
\emptyset	The empty set	-
U	The universal set	-
$\{x \mid \dots\}$	Set of all elements x for which ...	$\{x \mid x \geq 5\}$
\in	Element of	$-5 \in \mathbb{Z}$
\notin	Not an element of	$-5 \notin \mathbb{N}$
$n(A)$	Number of elements in A	-
\cap	Intersection of sets	$A \cap B$
\cup	Union of sets	$A \cup B$
A'	Complement of A (set of all elements not in A)	-

\subset	Proper subset of	$A \subset B$
\subseteq	Subset of or equal to	$A \subseteq B$
$\not\subset$	Not a subset of	$A \not\subset B$
$\not\subseteq$	Not a subset of or equal to	$A \not\subseteq B$
$[a, b]$	The closed interval of values between a and b inclusive	$x \in [a, b]$
(a, b)	The open interval of values between a and b non-inclusive	$y \in (a, b)$

Relation symbols

Symbol	Meaning	Example (if applicable)
$=$	Equal	$a = b$
\neq	Not equal	$a \neq b$
\equiv	Identical	$a \equiv b$
\approx	Approximately equal to	$a \approx b$
$>$	Greater than	$a > b$
\geq	Greater than or equal to	$a \geq b$
$<$	Less than	$a < b$
\leq	Less than or equal to	$a \leq b$

Function symbols

Symbol	Meaning	Example (if applicable)
$f : A \rightarrow B$	The function f maps the set A to the set B	$f : \mathbb{R} \rightarrow \mathbb{R}^+$
$f : x \mapsto y$	The function f maps the element x to the element y	$f : x \mapsto e^x$
$f(x)$	The function f of element x	$f(x) = e^x$
$f^{-1}(x)$	The inverse function of f	$f^{-1}(x) = \ln(x)$
$f \circ g(x)$	The composite function of f and g	$f \circ g(x) = e^{g(x)}$

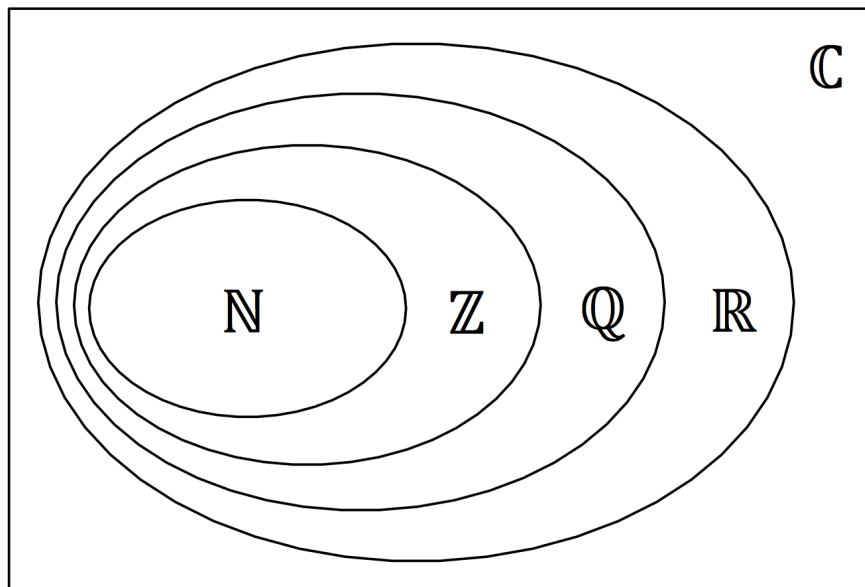
Geometry symbols

Symbol	Meaning	Example (if applicable)
--------	---------	-------------------------

$P(x, y)$	The point with coordinates (x, y)	$P(1, 2)$
$[AB]$	The line segment with endpoints A and B	-
AB	The length of the segment $[AB]$	$AB = 5$
(AB)	The line containing the points A and B	-
\hat{A}	The angle of point A	$\hat{A} = 45^\circ$
\hat{ABC}	The angle between segments $[AB]$ and $[BC]$	$\hat{ABC} = 30^\circ$
$\triangle ABC$	The triangle with vertices A, B and C	-

1.2 Number systems

A number system refers to a set of numbers which have a certain property. The Venn diagram below shows common number systems looked at in the IB course.



Some common number systems and their set relations

\mathbb{N} is the set of *natural numbers*: $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ (sometimes 0 is excluded). These are numbers which we can “count naturally” (such as the number of books on a bookshelf). \mathbb{N} holds some importance in mathematics (see **Peano’s axioms** on page 6 and **Countability** on page 11).

\mathbb{Z} is the set of *integers*: $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$. These are numbers which are formed by taking \mathbb{N} and adding 0 (if not already in \mathbb{N}) and $-\mathbb{N}$, which are *negative* natural numbers. Integers hold importance in topics such as discrete mathematics. Note that the natural numbers are contained in the integers, so we say $\mathbb{N} \subset \mathbb{Z}$, which is read as “ \mathbb{N} is a subset of \mathbb{Z} ”.

\mathbb{Q} is the set of *rational numbers*: $\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\}$. This is the set of all numbers which can be expressed as fractions, such as $\frac{1}{2}$ or $\frac{4}{5}$. The rationals are what we describe as *dense*. This means between any two rationals p and q where $p < q$, we can find another rational r for which $p < r < q$. Despite this, we can still count the rationals like we can count the natural numbers or the integers (see **Countability** on page 11). We also have that $\mathbb{Z} \subset \mathbb{Q}$, because we can express any integer a as a fraction: $\frac{a}{1}$.

\mathbb{R} is the set of *real numbers*. This is the set of all numbers on the number line, including numbers which are **not** rational (which we say are *irrational*). Think of the real numbers as the rational numbers with the “gaps filled” or the set of all possible decimal numbers. Some examples of irrational numbers are $\sqrt{2} = 1.4142135624\dots$ and $\pi = 3.1415926536\dots$, neither of which can be expressed as fractions (the proof of $\sqrt{2}$ being irrational is on page 5). In fact, a real number is rational if and only if it has a terminating digit or repeating decimal digits. We clearly have that $\mathbb{Q} \subset \mathbb{R}$.

\mathbb{C} is the set of *complex numbers*. This is the set of all numbers of the form $a + bi$, where $a, b \in \mathbb{R}$ and i is an *imaginary* number defined as $\sqrt{-1}$. The complex numbers serve a purpose in finding roots of polynomials. It can be proven that every polynomial of degree n with complex coefficients has exactly n complex roots (some of which may be repeated). They are also looked at further in Complex Analysis at university. Complex numbers are covered at HL, so SL students don’t need to know this.

Further reading: Countability

As stated, the natural numbers \mathbb{N} are numbers which we can count with. Therefore, we can use these to determine whether other sets are countable. What we mean by countable is that we can count each element in an *infinite* set one at a time; even though it would take us forever! Some definitions may say a finite set is countable; there’s no problem with this. However this text will take countable to exclude the finite case so as to use “countable” to mean countably-infinite sets distinct from finite sets, avoiding ambiguity.

The formal definition for a set Y to be countable (countably-infinite) is that there exists a **bijection** from \mathbb{N} to Y . That is, we can find a function $f : \mathbb{N} \rightarrow Y$ such that f is both **injective** (one-to-one) and **surjective** (onto). The definitions of these are given below:

- A function $f : X \rightarrow Y$ is *injective* when for $x_1, x_2 \in X$, we have that $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

- A function $f : X \rightarrow Y$ is *surjective* when for all $y \in Y$, we can find some $x \in X$ for which $f(x) = y$.

We can now use these definitions to show that some familiar sets are countable (please be aware though that there is no need to be able to do this at IB, no matter what level you're taking):

Theorem. \mathbb{Z} is countable

Proof. Note: in this proof, \mathbb{N} starts at 0

We must find a function $f : \mathbb{N} \rightarrow \mathbb{Z}$ such that f is bijective (injective and surjective).

$$\text{Consider } f(n) = \begin{cases} \frac{n}{2}, & n \text{ is even} \\ -\frac{n+1}{2}, & n \text{ is odd} \end{cases}$$

Injection: suppose $f(m) = f(n)$. Then either:

$$\frac{m}{2} = \frac{n}{2} \tag{1}$$

$$\text{or } -\frac{m+1}{2} = \frac{n}{2} \tag{2}$$

(without loss of generality).

For (1), it is clear that $m = n$ as required.

For (2), $-m-1 = n \implies m+n = -1$, which is not possible in \mathbb{N} .

Hence, f must be injective.

Surjection: let $z \in \mathbb{Z}$; we must show there is an $n \in \mathbb{N}$ for which $z = f(n)$

Let $z \geq 0$ (z is positive or 0) and consider $n = 2z$. Then $f(n) = f(2z) = \frac{2z}{2} = z$

Now let $z < 0$ (z is negative) and consider $n = -2z - 1$ (n is odd). Then $f(n) = f(-2z - 1) = -\frac{-2z}{2} = z$

We have found values of n for which any $z \in \mathbb{Z}$ can be produced. Hence, f is surjective.

Since f is injective and surjective, it is bijective. Hence a bijection between \mathbb{N} and \mathbb{Z} exists, which means \mathbb{Z} is countable. ■

Theorem. \mathbb{Q} is countable

Proof. Since we have established that \mathbb{Z} is countable, we can use this property to prove \mathbb{Q} is.

Consider the set $S_n = \left\{ \frac{m}{n} \mid m \in \mathbb{Z} \right\}$ for $n \in \mathbb{Z}^+$.

It is clear that $\mathbb{Q} = \bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cup S_3 \cup \dots$

$S_n = \frac{1}{n} \{m \mid m \in \mathbb{Z}\} = \frac{1}{n} \mathbb{Z}$. Since \mathbb{Z} is countable and we are simply multiplying each term of \mathbb{Z} by a constant, so S_n must be countable.

We will use the fact (without proof) that the countable union of countable sets is countable.

This proves that $\bigcup_{n=1}^{\infty} S_n$ is countable and hence that \mathbb{Q} is countable. ■

Note that if you aren't happy with the proof using the fact that the countable union of countable sets is countable, I've included a proof of this fact in **Appendix A** on page 88. Note also that there are other ways to prove \mathbb{Q} is countable; one is written in **Appendix A** on page 90 which uses (and proves!) the *Fundamental Theorem of Arithmetic* instead (**Appendix A**, page 89). We can also prove that \mathbb{R} is **uncountable** (**Appendix A**, page 92).

1.3 Approximations

While this topic is officially covered in the AI course, it may be worth looking into anyway (as an AA student).

In mathematics, we may find values which have a large number of trailing decimal digits, some of which may not even terminate (such as irrational numbers). It may come handy to express such numbers either in *exact form* (if possible) or in *approximated form*.

Exact form

A number is in exact form if it is written such that there is complete certainty (that is, the expression provides all information about that number). Some examples are:

Number in exact form	Number not in exact form
$\sqrt{2}$	1.414213562
π	3.141592654
e	2.718281828

A number in approximated form is usually written as a decimal number with a terminating digit, but can also be written as a fraction. For example, 3.14 is an approximation of π and so is $\frac{22}{7} = 3.\overline{142857}$ (note that a horizontal bar over the decimal digits indicates that the group of digits is recurring). We say that $\pi \approx 3.14$ and $\pi \approx \frac{22}{7}$.

When approximating, we usually provide the *degree* of an approximation; how close it is to the exact value. The course looks at two of these: *decimal places* and *significant figures*.

Decimal places are the number of decimal digits which are written in the approximated form (the number of digits which come after the whole part). For example, 3.1416 is correct to 4 decimal places (4 d.p.) while 103.4 is correct to 1 d.p.

Note that when rounding to a number of decimal places, the last digit to be written (d_n) must be rounded up if the digit to the right of d_n (d_{n+1}) is greater than or equal to 5. Also, d_{n+1} must **not** be rounded.

Decimal places

A number in approximated form given correct to n decimal places has n decimal digits provided. Some approximations of $\sqrt{2}$ are given below:

Approximation	Decimal places
1	0
1.4	1

1.41	2
1.414214	6
1.4142136	7
1.41421356	8

Significant figures are the digits which “have meaning” to the approximation. Note that these don’t necessarily include only the decimal digits; this applies to all digits of a number. Significant figures have a few rules associated with them.

Significant figures

- Any non-zero digit is significant
- Any zeros between two significant zeros are significant
- Trailing zeros are significant **only** when they are in the decimal part of a number

Additionally, the rounding rule described with decimal places applies here as well. Some approximations of 1006.50089 are given below:

Approximation	Significant figures
1000	1
1000	2
1010	3
1007	4
1006.5	5
1006.50	6
1006.501	7
1006.5009	8
1006.50089	9
1006.500890	10
1006.5008900	11

It is also worth knowing how to express numbers in *standard form*:

Standard form

A number is in standard form if it is expressed as $a \times 10^k$, where $1 \leq a < 10$ and $k \in \mathbb{Z}$.

For example, 0.0000456 is expressed as 4.56×10^{-5} and 12300000 is expressed as 1.23×10^7 .

Another property to consider is the *upper and lower bounds* of an approximation. When a number is rounded, these bounds describe the **range** of possible values for the exact value. For instance, suppose $a \approx 4.23$ approximated to 2 d.p. Then the range of values of a is: $4.225 \leq a < 4.235$. Note that the last part is given a strict inequality. This is because a can't be 4.235 exactly (otherwise it would've been rounded to 4.24), but it can still be anything less than that (such as 4.2349999).

Upper and lower bounds

Given an approximation of a value, one can find the upper and lower bounds of the exact number by considering the range of possible values which produce the approximation when rounded. Some examples are given below:

Approximation	d.p. / s.f.	Range of values
1.41	2 d.p.	$1.405 \leq x < 1.415$
56.789	3 d.p.	$56.7885 \leq x < 56.7895$
1200	2 s.f.	$1150 \leq x < 1250$
1200	3 s.f.	$1195 \leq x < 1205$

Finally, we can calculate how off our approximation is to an exact value, given that we have an exact value. We denote this as a *percentage error*, calculated like so:

Percentage errors

Given an approximation a_0 of an exact value a , the percentage error of the approximation is calculated by:

$$\left| \frac{a_0 - a}{a} \right| \cdot 100$$

The lower the percentage error, the better the approximation.

Of course, you need to assume you have an exact value or something extremely close to it. Even calculators aren't exact when it comes to irrational numbers.

Notice the vertical bars surrounding the expression. This denotes the *absolute value*.

Absolute value notation

For some value $x \in \mathbb{R}$, we define the *absolute value* of x as:

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

In informal terms, $|x|$ is x with the “minus sign removed” if there is a minus sign.

Worked example 1.3.1

Calculate the percentage errors of the following approximations of π and hence determine which approximation is the best. You may make use of the predefined constant π on your calculator as the exact value. Give your answers to 3 s.f. in standard form.

(a) $\frac{22}{7}$

(b) $\sqrt[4]{\frac{2143}{22}}$

(c) $\sqrt{2} + \sqrt{3}$

(a) $\left| \frac{\pi - \frac{22}{7}}{\pi} \right| \cdot 100 = 4.02 \times 10^{-2} \%$

(b) $\left| \frac{\pi - \sqrt[4]{\frac{2143}{22}}}{\pi} \right| \cdot 100 = 3.21 \times 10^{-8} \%$

(c) $\left| \frac{\pi - (\sqrt{2} + \sqrt{3})}{\pi} \right| \cdot 100 = 1.49 \times 10^{-1} \%$

(b) is the best approximation since the percentage error is the lowest out of the three.

Final word of advice: make sure not to round any values during your working. Only round the final answer (if needed) to ensure it is rounded correctly.

Further reading: Finding π

Although there are a number of ways to express the exact value of π in theory, there is no way to express it numerically. This is due to the fact that π is irrational and hence can't be expressed as a fraction. Here are some ways π can be expressed exactly (along with the mathematician(s) who discovered it):

$$\pi = 4 \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \quad (\text{Madhava/Leibnitz})$$

$$\pi = \sqrt{\sum_{n=1}^{\infty} \frac{6}{n^2}} \quad (\text{Euler})$$

$$\pi = \frac{9801}{2\sqrt{2} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4(396^{4k})}} \quad (\text{Ramanujan})$$

$$\pi = \int_{-1}^1 \frac{dx}{\sqrt{1-x^2}} \quad (\text{Weierstra\ss})$$

Drill Exercises

1 Round the following numbers:

(a) 4.55666 to 2 d.p.

(d) π to 1 s.f.

(b) e to 5 d.p.

(e) $\sqrt{135}$ to 4 s.f.

(c) $\frac{13}{17}$ to 3 d.p.

(f) $\frac{13}{17}$ to 3 s.f.

2 For the following approximations, find the upper and lower bounds and hence write down the range of possible exact values.

(a) 1.41 to 2 d.p.

(d) 13400 to 3 s.f.

(b) 4.056 to 3 d.p.

(e) 13400 to 5 s.f.

(c) 4.53453 to 5 d.p.

(f) 0.00406 to 3 s.f.

3 Find the percentage error of each of the following approximations with respect to their given exact value.

(a) 1.41421 as an approximation of $\sqrt{2}$

(d) $\sqrt{11}$ as an approximation of $\sqrt{12}$

(b) $\sqrt{2}$ as an “approximation” of 1.41421

(e) e as an approximation of π

(c) 2.8 as an approximation of $\sqrt{8}$

(f) π as an approximation of e

Word Problems

4 Show that $0.\overline{9} = 1$ (recall that the bar over 9 indicates a recurring decimal digit, so $0.\overline{9}$ means 0.99999999...).

1.4 Counting principles [HL]

Students may find this particular topic difficult to get the hang of. It looks at methods of counting in certain situations and how we can easily find the number of ways an event occurs.

The first concept to introduce is the *product* and *sum* principles. If two choices are to be made, these principles can be used to combine these choices in different ways. Specifically, the product principle is used for situations where one choice **and** another choice must be made, while the sum principle is used for situations where one choice **or** another choice must be made.

Consider the situation where you need to choose one option for each of these categories:

Breakfast	Lunch	Dinner
Boiled egg	Sandwich	Steak
Scrambled egg	Salmon	Pork
Cereal	Salad	
Yogurt		

How many ways could you do this? To explain the process, we must look at what must be done. We have to choose 1 breakfast item **and** 1 lunch item **and** 1 dinner item. Let the number of ways we can choose breakfast, lunch and dinner be $n(B)$, $n(L)$ and $n(D)$, respectively.

Then we have that $n(B) = 4$, $n(L) = 3$ and $n(D) = 2$.

The *product* principle requires us to **multiply** our numbers of choices to obtain the total value. Hence, the total number of ways these can be chosen is $n(B) \cdot n(L) \cdot n(D) = 4 \cdot 3 \cdot 2 = 24$. You can verify that this is indeed the correct number by considering all possible meal combinations and counting them.

Now suppose you are in the (strange) situation where you must choose either a breakfast and a dinner item or a lunch and a dinner item. Then the sum principle can be used. We must choose either (1 breakfast item **and** 1 dinner item) **or** (1 lunch item **and** 1 dinner item), which is equivalent to saying we must choose (1 breakfast item **or** 1 lunch item) **and** 1 dinner item.

The *sum* principle requires us to **add** our numbers of choices to obtain the total value. Hence, the total number of ways we can do this is $n(B) \cdot n(D) + n(L) \cdot n(D) = (n(B) + n(L)) \cdot n(D) = (4 + 3) \cdot 2 = 7 \cdot 2 = 14$.

To summarize:

Product and sum principles

Let $n(A)$ be the number of ways some selection A can occur and $n(B)$ be the number of ways some selection B can occur. Then:

- The number of ways A **and** B can occur is $n(A) \cdot n(B)$
- The number of ways A **or** B can occur is $n(A) + n(B)$

Worked example 1.4.1

A camp offers a number of activities during the day for campers to choose from:

Creative	Sports	Group
Arts and crafts	Archery	Board games
Cooking	Canoeing	Campfire
Music	Climbing	Scavenger hunt
	Kayaking	Theater
	Swimming	

How many ways can a camper select their activities, given that:

- (a) They must choose one activity from each category.
- (b) They must select one activity from any of the categories.
- (c) They must select a group activity and one of either a creative or sports activity.

We have that $n(C) = 3$, $n(S) = 5$ and $n(G) = 4$. Hence:

- (a) $n(C) \cdot n(S) \cdot n(G) = 3 \cdot 5 \cdot 4 = 60$
- (b) $n(C) + n(S) + n(G) = 3 + 5 + 4 = 12$
- (c) $n(G) \cdot (n(C) + n(S)) = 4(3 + 5) = 4 \cdot 8 = 32$

While the product and sum principles are used to combine different ways of counting, the more interesting counting principles are *permutations* and *combinations*.

Suppose you have 2 distinct objects $\{A, B\}$. How many ways can you arrange them in a line? We have:

- $\{A, B\}$
- $\{B, A\}$

2 ways in total. Now suppose we add a third object C :

- $\{A, B, C\}$

- $\{A, C, B\}$
- $\{B, A, C\}$
- $\{B, C, A\}$
- $\{C, A, B\}$
- $\{C, B, A\}$

6 ways in total. This comes from the fact that there are 3 locations we can place object C among A and B and then count the ways we can arrange A and B . In other words, $2 \cdot 3 = 6$ (using the product principle).

If we added a fourth object D , we would have 4 possible locations to place it multiplied by the number of ways we can arrange A , B and C . Thus, $6 \cdot 4 = 24$. This pattern follows as we introduce more objects. Also, the number of ways we can arrange one object A is obviously 1: $\{A\}$.

We also call an arrangement a *permutation*. To get the number of permutations of n objects, we use the following:

The factorial function

For $n \in \mathbb{Z}$ where $n \geq 0$, we define $n!$ to be the *factorial* of n given by the function:

$$n! = f(n) = n f(n-1), \quad f(0) = 1$$

or in other words:

$$n! = 1 \cdot 2 \cdot 3 \cdots n \text{ for } n \geq 1 \text{ and } 0! = 1$$

To understand why $0! = 1$, consider the case where you have 0 objects $\{\}$. How many ways can you arrange them? Simply **one** way, which is $\{\}$.

Worked example 1.4.2

How many different words (don't need to be real ones) can be formed with the letters of the word "RELATION"?

The word has 8 distinct letters. Hence, we can make $8! = 40\,320$ different words.

Another principle to look at is the principle of combinations. This definition is not the same as the usual one. In mathematics, combinations are ways we can choose **without replacement** a number of objects out of a larger (or equal) number of objects without being concerned with how the choices are arranged. Suppose you are given five books $\{A, B, C, D, E\}$. How many ways could you choose three books out of the five?

Think about it this way. Let's suppose you considered all arrangements of the five books in a line and only concerned yourself with the first three books (whatever they were) which you are going to choose. Changing the arrangement of the books **might** yield another arrangement of the first three books, but this is not guaranteed. For instance $\{A, B, C, D, E\}$ and $\{A, B, C, E, D\}$ both have the first three books in the same order. We consider the number of ways all 5 can be arranged, which is $5! = 120$.

With this, we must *cancel out* the arrangements we don't want to consider. This simply means dividing by $(5 - 3)! = 2! = 2$ (we are *reversing* the product principle). We now have $\frac{5!}{2!} = \frac{120}{2} = 60$, but we still aren't done. This counts the number of ways the 3 books can be arranged, which is $3!$. Hence, we must cancel this out as well. This will give us $\frac{5!}{2! \cdot 3!} = \frac{60}{6} = 10$.

Therefore, there are 10 ways to choose 3 books out of the 5. We can form a general formula from this process:

Combinations formula

The number of ways r objects can be chosen out of n objects is given by $\binom{n}{r}$ or nC_r with formula:

$$\binom{n}{r} = {}^nC_r = \frac{n!}{r!(n-r)!}$$

Similarly if we consider the case where we **want** to permute the chosen objects (i.e. the order in which we choose the objects **does** matter), we simply don't divide the $r!$ in the nC_r formula:

Partial permutations formula

The number of ways r objects can be chosen **and permuted** out of n objects is given by nP_r with formula:

$${}^nP_r = \frac{n!}{(n-r)!}$$

It is more likely you will use the nC_r formula a lot more than the nP_r formula. But nP_r can be derived from nC_r by using the product principle anyway.

Worked example 1.4.3

A basketball team of 5 people is to be formed by choosing from 6 boys and 4 girls. How many ways can a team be formed given that:

- (a) The team must have only boys.
- (b) The team must have 3 boys and 2 girls.
- (c) The team must have at least 1 girl.

- (a) Simply choose 5 out of 6 boys:

$${}^6C_5 = \frac{6!}{5!(6-5)!} = \frac{6}{1} = 6$$

- (b) Choose 3 boys out of 6 **and** choose 2 girls out of 4:

$$\begin{aligned} {}^6C_3 \cdot {}^4C_2 &= \frac{6!}{3!(6-3)!} \cdot \frac{4!}{2!(4-2)!} \\ &= \frac{4 \cdot 5 \cdot 6}{2 \cdot 3} \cdot \frac{3 \cdot 4}{2} = 20 \cdot 6 = 120 \end{aligned}$$

- (c) **At least** 1 girl means (1 girl **and** 4 boys) **or** (2 girls **and** 3 boys) **or** (3 girls **and** 2 boys) **or** (4 girls **and** 1 boy)

$$\begin{aligned} &({}^6C_4 \cdot {}^4C_1) + ({}^6C_3 \cdot {}^4C_2) + ({}^6C_2 \cdot {}^4C_3) + ({}^6C_1 \cdot {}^4C_4) \\ &= 15 \cdot 4 + 20 \cdot 6 + 15 \cdot 4 + 6 \cdot 1 \\ &= 60 + 120 + 60 + 6 = 246 \end{aligned}$$

We looked at how we can reverse the product principle by dividing. We can also do the same for the sum principle by subtracting. This comes in handy if you want to find the number of ways some selection **can't** occur.

For example, consider a safe with a 4-digit password of digits 0 to 9. How many combinations are there which **don't** start with the digit 1?

We can go through all digits 0XXX, 2XXX, 3XXX, ..., 9XXX, which would be time-consuming or we could simply calculate the number of codes which **do** start with 1 and subtract it from the total number of possible codes, thus excluding the cases where the code starts with 1, implying the other codes don't start with 1.

The total number of codes can be calculated simply by using the product principle. There are 10 digits to choose from and we want to choose 4 of them **with replacement**. This means we choose from 10 and 10 and 10 and 10, which results in $10 \cdot 10 \cdot 10 \cdot 10 = 10^4 = 10\,000$ total possibilities.

Similarly, we choose the digit 1 to be the first digit and then choose 3 trailing digits (with replacement again) out of the 10. This means $1 \cdot 10^3 = 1\,000$. Hence, there are 1000 codes which start with 1.

Finally, we subtract the total with this value: $10\,000 - 1\,000 = 9\,000$ codes which **don't** start with 1.

Subtraction principle

If $n(T)$ is the total number of ways **any** selection in a given context can be made and $n(A)$ is the total number of ways a selection A can be made within that context, then the number of ways A **can't** be made will be:

$$n(T) - n(A)$$

As you can see, we need some sort of total value if we're going to use the subtraction principle. This total T requires some sort of context which can limit the selection A . In other words, $n(T) \geq n(A)$ for this to work.

It is recommended that you get used to the types of questions which may appear on the exams for this topic. Some can be made very difficult to think about. With no obvious approach given to you, you may find yourself struggling. Luckily, these types of questions rarely come up and when they do, they are not very difficult 80% of the time.

If you are interested in this topic and want to learn further, look into *combinatorics*.

Further reading: The Gamma function

We know that factorials for integers are intuitive to calculate. For instance, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$ and similarly $5! = 5 \cdot 4! = 120$. But what about real-valued factorials such as $4.5!$. Can they even exist? There is a technique in advanced mathematics called *analytic continuation*, which is used to extend the domains of functions such that the extension has a nice property (called *analyticity*). Moreover, this extension is **unique**. In this case, the Gamma function extends the factorial function from the domain \mathbb{N} to \mathbb{C} (which of course includes \mathbb{R}). The definition of the Gamma function for some $z \in \mathbb{C}$ is:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

Don't worry if you can't understand this. You will learn about what \int means later on in the course. The Gamma function is actually shifted by 1 with respect to the factorial function, which means $\Gamma(n) = (n-1)!$. This means that $4.5! = \Gamma(5.5) \approx 52.3428$. If you want to calculate more with the Gamma function, you can do so on Wolfram|Alpha. The function has some strange (but interesting) behavior when $z \in \mathbb{R}^-$. It's worth taking a look at.

Word problems

- 1** How many **distinct** words can be formed with the letters of the word "FUNCTION"?
- 2** How many **distinct** words can be formed with the letters of the word "QUADRATIC" if "QU" must be kept together in that order?
- 3** A bookshelf has three volumes of Naruto, four volumes of Dragon Ball, six volumes of One Piece and a cookbook.
 - (a) How many ways can these books be arranged such that the books of each series are grouped together?
 - (b) How many of the ways found in part (a) will have the cookbook next to the Naruto books?
- 4** 4 people attend a social gathering. Each person shakes hands with every other person. Find the total number of handshakes. Can you come up with a general formula for n people in terms of n ?

(This plays an important role in *graph theory* when finding the number of edges in a *complete graph*).

5 You are given 8 letters: $\{A, B, C, D, E, F, G, H\}$. How many ways can you arrange the letters such that each letter is **not** in its original position (i.e. A isn't the first letter, B isn't the second letter and so on)?

6 This question follows up from question 5.

(a) Find a general formula for when you have n letters. This formula is called the *inclusion-exclusion principle*.

(b) Continue your formula by letting k in your denominator tend to infinity. Show that this limit is $\frac{n!}{e}$.

(c) For what values of n would $\frac{n!}{e}$ be a good approximation for your formula in **(a)**?

2 Sequences and Series

A sequence is a list of numbers in a particular order (we call these numbers *terms*). There are many types of sequences which can be defined. For example, $\{1, 1, 2, 3, 5, 8, 13, 21, \dots\}$ is a well-known sequence called the *Fibonacci Sequence*, where the next term is found by adding the previous two terms. Let's call the n 'th term u_n . Then we can write u_n in terms of previous terms like so: $u_n = u_{n-1} + u_{n-2}$. Note that we must also explicitly define $u_1 = 1$ and $u_2 = 1$, since otherwise $u_2 = u_1 + u_0$ and $u_1 = u_0 + u_{-1}$, where u_0 and u_{-1} don't exist. We call these definitions the *initial conditions*. This sequence is special because it is what we call *recursive*, which means the next term is expressed in terms of the previous term(s). However, we can find a way to write u_n in terms of just n . In particular:

$$u_n = \left(\frac{1}{2} + \frac{\sqrt{5}}{10}\right) \left(\frac{1 + \sqrt{5}}{2}\right)^n + \left(\frac{1}{2} - \frac{\sqrt{5}}{10}\right) \left(\frac{1 - \sqrt{5}}{2}\right)^n$$

How this was obtained is left to you to find out for yourself. However, we don't need to be concerned with recursive sequences. The IB will test you on two types of sequences: arithmetic and geometric. These will be looked at in this chapter.

2.1 Arithmetic sequences and series

An *arithmetic sequence* is a sequence in which the next term is obtained by adding a constant value to the previous term. A very basic example is the sequence of whole numbers: $\{1, 2, 3, 4, 5, \dots\}$, where we obtain the next term by adding 1 to the previous one. Formally, we need to determine two values: the *first term*, which we call u_1 and the *common difference*, which we call d . We define the n 'th term to be u_n . It is common to denote a sequence itself as $\{u_n\}$ (i.e. for this case, $\{u_n\} = \{1, 2, 3, 4, 5, \dots\}$). The first term is the initial condition required to begin the arithmetic sequence. In the case of the sequence of whole numbers, $u_1 = 1$ is the first term. The value of d is what we add to the previous term to obtain the next one. To obtain 2, we simply need to add 1 to 1. Hence, $d = 1$ is the common difference. Note that $d = u_{n+1} - u_n$ for any $n \geq 1$, as this must remain constant and thus consistent with the sequence. Using our values of u_1 and d , we can express u_n in terms of them: $u_n = u_1 + (n - 1)d$. For our particular sequence, $u_n = 1 + (n - 1) = n$. This formula should be easy enough to understand and thus require no proof.

Arithmetic sequences

For an arithmetic sequence $\{u_n\}$ with first term u_1 and common difference d , the n 'th term u_n is obtained by:

$$u_n = u_1 + (n - 1)d$$

Worked example 2.1.1

Consider a less trivial sequence: $\{3, 7, 11, 15, \dots\}$. Find the common difference and write down an expression for the n 'th term of the sequence.

The common difference is $d = 7 - 3 = 4$.

Then $u_n = 3 + 4(n - 1) = 4n - 1$.

We can also have arithmetic sequences such as $\{12, 7, 2, -3, \dots\}$, where the sequence is decreasing. For this example, the common difference is $d = -5$ and thus the n 'th term is $u_n = 12 - 5(n - 1) = 17 - 5n$.

If there is a sequence with a common difference of 0 and first term u_1 , clearly all terms will be u_1 .

Increasing and decreasing arithmetic sequences

For an arithmetic sequence with common difference d :

- If $d > 0$, the sequence is **increasing**
- If $d < 0$, the sequence is **decreasing**
- If $d = 0$, the sequence is **constant**

Worked example 2.1.2

An arithmetic sequence $\{u_n\}$ is such that $u_1 = a$ and the common difference is d . You are given that $u_5 = 15$. You are also given that the **sum** of the first three terms is 0. Find a and d and hence determine whether the sequence is increasing or decreasing.

$$u_5 = a + 4d = 15, \text{ so } a = 15 - 4d \quad (1)$$

$$u_1 + u_2 + u_3 = 3a + 3d = 3(a + d) = 0, \text{ so } a + d = 0 \implies a = -d$$

Putting $a = -d$ into equation (1):

$$-d = 15 - 4d \implies 3d = 15 \implies d = 5 \text{ and hence } a = -5$$

Since $d = 5 > 0$, $\{u_n\}$ must be **increasing**.

Worked example 2.1.3

Let $\{u_n\}$ be an arithmetic sequence and let $a \in \mathbb{R}$ be a constant. Show that $\{au_n\}$ is also an arithmetic sequence.

Let the common difference of $\{u_n\}$ be d .

Consider $au_{n+1} - au_n$

$$= a(u_{n+1} - u_n)$$

$$= ad$$

Since $au_{n+1} - au_n$ is in terms of a and d , which are constants, $\{au_n\}$ has a common difference and is hence an arithmetic sequence.

An *arithmetic series* is the **sum** of an arithmetic sequence $\{u_n\}$ obtained by adding consecutive terms (i.e. $u_1 + u_2 + u_3 + \dots$). We will look at how the sum of the first n terms of an arithmetic sequence can be expressed in terms of u_1, d and n as well as *Sigma notation*.

Sigma is a letter from the Greek alphabet. σ is lowercase and Σ is upper case. For sums, Σ is used like so:

Sigma notation

For an arithmetic sequence $\{u_n\}$, the sum of the first N terms is:

$$\sum_{k=1}^N u_k = u_1 + u_2 + u_3 + \dots + u_N$$

The $k = 1$ below the Σ describes the variable k which is iterated starting from 1. The N above the Σ describes the value for which k terminates (so the sum ends when $k = N$).

We shall now find a way to express the sum of the first N terms of an arithmetic sequence in closed form in terms of u_1 , d and N . Before this, it is important to consider how to obtain an expression for the sum of the first N **natural numbers** in closed form in terms of N .

Theorem.
$$\sum_{k=1}^N k = \frac{N(N+1)}{2}$$

Proof. Let $S = 1 + 2 + 3 + \dots + N$. Then:

$$\begin{aligned} S &= N + (N-1) + (N-2) + \dots + (N - (N-2)) + (N - (N-1)) \\ &= N^2 - (1 + 2 + 3 + \dots + (N-1)) \\ &= N^2 - (1 + 2 + 3 + \dots + (N-1) + N) + N \\ &= N^2 + N - S \end{aligned}$$

$$2S = N^2 + N = N(N+1) \implies S = \frac{N(N+1)}{2}$$

■

Now consider:

$$\begin{aligned} \sum_{k=1}^N u_k &= u_1 + u_2 + \dots + u_N \\ &= u_1 + (u_1 + d) + (u_1 + 2d) + \dots + (u_1 + (N-1)d) \\ &= Nu_1 + (1 + 2 + \dots + (N-1))d \\ &= Nu_1 + \frac{(N-1)N}{2}d \\ &= \frac{N}{2}(2u_1 + (N-1)d) \end{aligned}$$

This is indeed the sum in closed form. Notice also that:

$$\frac{N}{2}(2u_1 + (N-1)d) = \frac{N}{2}(u_1 + (u_1 + (N-1)d)) = \frac{N}{2}(u_1 + u_N)$$

Thus we have another way of expressing the sum in closed form.

Arithmetic series

For an arithmetic sequence $\{u_n\}$ with common difference d , the sum of the first n terms of the sequence is given by:

$$S_n = \frac{n}{2}(2u_1 + (n-1)d)$$

or equivalently:

$$S_n = \frac{n}{2}(u_1 + u_n)$$

Worked example 2.1.4

Let $\{u_n\}$ be an arithmetic sequence with common difference d and let $\{S_n\}$ be a sequence such that $S_n = \sum_{k=1}^n u_k$. You are given that $u_7 = 14$ and $S_{12} = 126$.

- (a) Find u_1 and d
(b) Find the least value of n for which $S_n > 500$

$$(a) \quad u_7 = u_1 + 6d = 14 \tag{1}$$

$$\text{and } S_{12} = 6(2u_1 + 11d) = 12u_1 + 66d = 126 \tag{2}$$

Thus, by solving (1) and (2) simultaneously (via GDC or otherwise):

$$a = -28 \text{ and } d = 7$$

$$(b) \quad S_n = \frac{n}{2}(2 \cdot -28 + 7(n-1)) = \frac{7n^2 - 63n}{2}.$$

Let $f(x) = \frac{7}{2}x^2 - \frac{63}{2}x$ and consider $f(x) = 500$.

From GDC or otherwise, $x = -8.27$ or $x = 17.3$. But $n > 0$. Thus, $x = 17.3$.

Finally, $S_{17} = 476$ and $S_{18} = 567$. Hence, the least value of n is 18 for which $S_n > 500$.

Questions like the above example appear frequently in paper 2. It is imperative that you understand how to use your GDC to solve these types of questions.

Drill Exercises

1 Determine whether each of the following can be consecutive terms of an arithmetic sequence. If they can, write down their common difference.

(a) $\{1, 3, 5, 7, 9\}$

(d) $\{-5, -10, -15, -20, -30\}$

(b) $\{1, 1, 2, 3, 5\}$

(e) $\{-2, -8, -14, -20, -26\}$

(c) $\{1, 4, 9, 16, 25\}$

(f) $\{1, 1, 1, 1, 1\}$

2 Let $u_n = 3n - 1$ and $v_n = 8 - 5n$ for $n \in \mathbb{Z}^+$. Find the following:

(a) $\sum_{k=1}^5 u_k$

(d) $\sum_{k=3}^5 v_k$

(b) $\sum_{k=1}^8 v_k$

(e) $\sum_{k=1}^{10} u_k + v_k$

(c) $\sum_{k=5}^8 u_k$

(f) $\sum_{k=1}^{12} u_k - v_{k+1}$

Word Problems

3 Consider an arithmetic sequence $\{u_n\}$ with first term $u_1 = 2$ and common difference $d_u \neq 0$ and an arithmetic sequence $\{v_n\}$ with first term $v_1 = 5$ and common difference $d_v \neq 0$. Let $S_n = \sum_{k=1}^n u_k$ and $T_n = \sum_{k=1}^n v_k$.

Find d_u and d_v given that:

(a) $u_4 = v_4$ and $S_7 = T_9$

(b) $u_2 = v_2$ and $u_5 = T_4$

4 Let $\{u_n\}$ and $\{v_n\}$ be arithmetic sequences. Show that $\{u_n \pm v_n\}$ is also an arithmetic sequence.

5 A **doubly-infinite** arithmetic sequence is an arithmetic sequence which has no first term (e.g. $\{..., -16, -12, -8, -4, 0, 4, 8, 12, 16, ...\}$). Let α and β be doubly-infinite arithmetic sequences and let $\gamma = \alpha \cap \beta$

Prove that either $\gamma = \emptyset$ or γ is an arithmetic sequence. Does this also hold for arithmetic sequences which *have* a first term? Prove or disprove.

6 The **product** of the first n terms of an arithmetic sequence $\{u_n\}$ is given in *Pi notation* (not to be confused with π) like so:

$$\prod_{k=1}^n u_k = u_1 u_2 u_3 \dots u_n$$

Find a formula for the product of the first n terms of an arithmetic sequence in closed form (in terms of the first term, the common difference and n). What restrictions/conditions do you need for the formula to work?

2.2 Geometric sequences and series

A *geometric sequence* is a sequence in which the next term is calculated by **multiplying** the previous term by a constant value $r \neq 0$ (which is referred to as the *common ratio*). For example: $\{1, 2, 4, 8, 16, 32, \dots\}$ (where $r = 2$). It should be clear that for such sequences, it's harder to determine whether one is increasing or decreasing (in fact, some sequences may do neither). With these types of sequences, we will see that there is a different classification of behavior we may use.

With a geometric sequence $\{u_n\}$ with common ratio $r \neq 0$ and first term u_1 , it's clear from the definition above that $u_2 = u_1 r$, $u_3 = u_2 r = u_1 r^2$ and so on. In fact, $u_n = u_1 r^{n-1}$.

Geometric sequences

For a geometric sequence $\{u_n\}$ with first term u_1 and common ratio $r \neq 0$, the n 'th term u_n is obtained by:

$$u_n = u_1 r^{n-1}$$

Notice also that $\frac{u_{n+1}}{u_n} = \frac{u_1 r^n}{u_1 r^{n-1}} = r$.

Worked example 2.2.1

A geometric sequence $\{u_n\}$ has first term 3 and common ratio -2 . Find u_4 and u_7 .

$u_n = 3 \cdot (-2)^{n-1}$. Hence:

$$u_4 = 3 \cdot (-2)^3 = -24 \text{ and } u_7 = 3 \cdot (-2)^6 = 192$$

Alternating geometric sequences

A geometric sequence is an *alternating sequence* if $r < 0$.

The common ratio of a geometric sequence can be used to describe the sequence's nature. Let $\{u_n\}$ be a geometric sequence with common ratio r . If $r > 1$, it's clear that $u_1 < u_1 r < u_1 r^2 < \dots$ and thus $u_1 < u_2 < u_3 < \dots$. In simpler terms, $u_n < u_{n+1}$ for all n . If $r > 1$, it is clear (or should be clear) that r^n approaches infinity as n approaches infinity. This means that u_n approaches infinity as n approaches infinity, assuming $u_1 > 0$. Note that if $u_1 < 0$, the same effect occurs, but u_n approaches **negative** infinity. Either way, we can say that $\lim_{n \rightarrow \infty} |u_n| = \infty$. If you aren't familiar with this notation:

Limit notation

For a sequence $\{u_n\}$, if u_n approaches some value L as n approaches infinity, we write:

$$\lim_{n \rightarrow \infty} u_n = L$$

It is common (and in fact proper) to say that “no such limit exists” if L is infinity or negative infinity. If L is finite, we can say $|L| < \infty$.

Similarly if $r < -1$, we have that $|u_n| < |u_{n+1}|$ and $|r|^n$ approaches infinity as n approaches infinity. Thus, $\lim_{n \rightarrow \infty} |u_n| = \infty$. Note that in this case the sequence is alternating, so we say that u_n is approaching infinity **absolutely**.

Now if $|r| < 1$ (i.e. $-1 < r < 1$), we have that $|u_{n+1}| < |u_n|$ for all values of n . This means that each absolute term is decreasing. We also have that if $|r| < 1$, then $|r| = \frac{1}{|s|}$ and $s > 1$ as a result. So $|u_n| = \frac{u_1}{|s|^{n-1}}$. As we let n approach infinity, we know from above that $|s|^{n-1}$ approaches infinity. Thus, $|u_n|$ must approach 0 as n approaches infinity. In other words, $\lim_{n \rightarrow \infty} u_n = 0$.

If $r = -1$, we simply get that $|u_n| = |u_{n+1}|$. This does not mean that a limit exists, though. While the sequence doesn't approach infinity, notice that it also doesn't approach a finite value either.

The sequence has alternating constant terms. This is a case where if $\lim_{n \rightarrow \infty} u_n = L$, $|L| \neq \infty$ and $L \notin \mathbb{R}$. L doesn't exist whatsoever.

Converging and diverging sequences

For a geometric sequence $\{u_n\}$ with common ratio $r \neq 0, 1$:

- If $|r| < 1$, $\lim_{n \rightarrow \infty} u_n = 0$ and we say the series is **converging**
- If $|r| \geq 1$, $\lim_{n \rightarrow \infty} u_n$ doesn't exist and we say the series is **diverging**

A *geometric series* is the **sum** of a geometric sequence $\{u_n\}$ obtained by multiplying consecutive terms (i.e. $u_1 u_2 u_3 \dots$). Like with arithmetic series, we will look at how the sum of the first n terms of a geometric sequence can be expressed in terms of u_1, r and n in closed form.

Let $\{u_n\}$ be a geometric sequence with common ratio $r \neq 0, 1$ and consider:

$$\begin{aligned}
 \sum_{k=1}^N u_k &= u_1 + u_2 + u_3 + \dots + u_N \\
 &= u_1 + u_1 r + u_1 r^2 + \dots + u_1 r^{N-1} \\
 &= u_1 (1 + r + r^2 + \dots + r^{N-1}) \\
 &= u_1 (1 + r + r^2 + \dots + r^{N-1}) \frac{(r-1)}{(r-1)} \\
 &= u_1 \frac{(r + r^2 + r^3 + \dots + r^N) - (1 + r + r^2 + \dots + r^{N-1})}{r-1} \\
 &= \frac{u_1(r^N - 1)}{r-1} = \frac{u_1(1 - r^N)}{1-r}
 \end{aligned}$$

The formula was obtained by using a property of polynomial factorization.

Geometric series

For a geometric sequence $\{u_n\}$ with common ratio $r \neq 0, 1$, the sum of the first n terms of the sequence is given by:

$$S_n = \frac{u_1(r^n - 1)}{r - 1}$$

or equivalently:

$$S_n = \frac{u_1(1 - r^n)}{1 - r}$$

Worked example 2.2.2

Let $n, k \in \mathbb{Z}$ where $n \geq 0$ and $k \geq 2$.

- (a) Find an expression for the first n powers of k
- (b) Hence find the sum of the first 10 powers of 5

- (a) The powers of k are $k^0, k^1, k^2, k^3, \dots$

The n 'th power of k is k^n ; we're interested in the first n powers (including 0), so:
 $k^0 + k^1 + \dots + k^{n-1}$.

Each term is in a geometric sequence $\{u_n\}$ where $u_n = k^{n-1}$ and the common ratio is k . Hence:

$$k^0 + k^1 + \dots + k^{n-1} = \frac{k^{n-1} - 1}{k - 1}$$

- (b) Using the above, this sum will be $\frac{5^{10-1} - 1}{5 - 1} = \frac{5^9 - 1}{4} = 488\,281$

Further reading: Formal definition of convergence

We have determined that a geometric sequence converges if $|r| < 1$. Intuitively, this is quite clear due to the fact that as n approaches infinity, $|r|^n$ approaches infinity whenever $|r| > 1$ and $|r|^n$ approaches 0 whenever $|r| < 1$. However, this may not be the case for all types of sequences.

Consider the sequence $\{e_n\}$ where $e_n = \left(1 + \frac{1}{n}\right)^n$. Indeed, it looks difficult to determine $\lim_{n \rightarrow \infty} e_n$. Notice also that $e_n < e_{n+1}$ too, so e_n is an increasing sequence. This sequence does

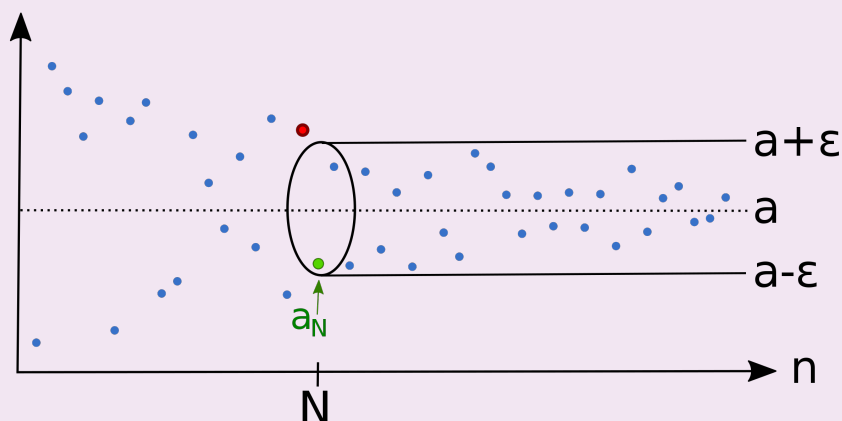
indeed approach a limit; we shall find out what this limit is later in the book.

The formal definition of sequence convergence is given as follows:

Let $\{a_n\}$ be some sequence where $a_n \in \mathbb{R}$. Then $\lim_{n \rightarrow \infty} a_n = a$ exists if and only if:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ such that } |a_n - a| < \varepsilon \quad \forall n \geq N$$

This is read as: “for all epsilon greater than zero, there exists some natural number N such that $|a_n - a|$ is less than epsilon for every $n \geq N$ ”. In other words, if we choose any real number greater than zero, we can find some term a_N for which the absolute difference between a_n and a is less than the real number chosen for any term a_n which comes after or is equal to a_N . Visualizing this may help you understand how this definition works:



All terms after (and including) a_N are contained within the bounds of $a - \varepsilon$ and $a + \varepsilon$.

Drill Exercises

1 Determine whether each of the following geometric sequences are converging or diverging.

(a) $\{1, 2, 4, 8, 16, \dots\}$

(d) $\{4, -4, 4, -4, 4, \dots\}$

(b) $\{400, 200, 100, 50, 25, \dots\}$

(e) $\{-1, -4, -16, -64, -256, \dots\}$

(c) $\{-3^{-1}, 3^{-2}, -3^{-3}, 3^{-4}, -3^{-5}, \dots\}$

(f) $\{4, 2, 1, \frac{1}{2}, \frac{1}{4}, \dots\}$

2 Let $u_n = 2 \cdot 5^n$ and $v_n = \pi^k$ for $n \in \mathbb{Z}^+$. Find the following (write your answers in terms of π where appropriate):

(a) $\sum_{k=1}^3 u_k$

(d) $\sum_{k=2}^5 v_k$

(b) $\sum_{k=1}^3 v_k$

(e) $\sum_{k=1}^4 u_k v_k$

(c) $\sum_{k=3}^6 u_k$

(f) $\sum_{k=1}^4 \frac{u_k}{v_k}$

Word Problems

3 Let $\{u_n\}$ be a geometric sequence with common ratio $r_u \neq 0$ and $\{v_n\}$ be a geometric sequence with common ratio $r_v \neq 0$. Also, let $a \in \mathbb{R}$. Show that:

(a) $\{u_n v_n\}$ is a geometric sequence

(b) $\{u_n^a\}$ is a geometric sequence

4 Let $\{u_n\}$ be a geometric sequence with first term u_1 and common ratio $r \neq 0, 1$.

Find u_1 and r , given that:

(a) $u_2 = -2u_4$

(b) $u_4 = S_3$

5 Let $\{u_n\}$ be a geometric sequence with common ratio r .

Show that $\prod_{k=1}^n u_k = (\sqrt{u_1 u_n})^n$ and explain what conditions for u_1 and r must be set for this formula to work.

2.3 Infinite geometric series

We have looked at convergence of geometric *sequences* and have established that if a common ratio r of a geometric sequence is such that $|r| < 1$, then the series converges (otherwise, it diverges). Specifically for a geometric sequence $\{u_n\}$, if $|r| < 1$ then $\lim_{n \rightarrow \infty} u_n = 0$. This also implies that the **series** converges as well, which means $\sum_{k=1}^{\infty} u_k$ has a finite value.

Theorem. *Let $\{u_n\}$ be a geometric sequence with common ratio $r \neq 0$. Then if $|r| < 1$, the series $\{u_n\}$ converges.*

Proof. Let $S_n = \sum_{k=1}^n u_k$ and consider the sequence $\{S_n\}$. We want to show that this sequence converges.

We know that $S_n = \frac{u_1(1-r^n)}{1-r}$, so $\sum_{k=1}^{\infty} u_k = \lim_{n \rightarrow \infty} S_n$.

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \frac{u_1(1-r^n)}{1-r} = \frac{u_1}{1-r} \cdot \lim_{n \rightarrow \infty} (1-r^n)$$

Since $|r| < 1$, we have established (in section 2.2) that r^n tends to 0 as n approaches infinity. Hence: $\lim_{n \rightarrow \infty} r^n = 0$, which implies:

$$\frac{u_1}{1-r} \cdot \lim_{n \rightarrow \infty} (1-r^n) = \frac{u_1}{1-r}$$

Since this expression is defined (as $r \neq 1$), S_{∞} exists and thus the series $\{u_n\}$ converges. ■

From our convergence proof above, we have found a closed-form expression for $\sum_{k=1}^{\infty} u_k$:

Infinite geometric series

For a geometric series $\{u_n\}$ with first term u_1 and common ratio r , if $|r| < 1$ then

$$S_{\infty} = \sum_{k=1}^{\infty} u_k = \frac{u_1}{1-r}$$

We can consider an example. Let $u_n = \frac{3}{2^{n-1}}$. Then $u_1 = 3$ and $r = \frac{1}{2}$.
 $= \frac{1}{2} < 1$. Hence, $S_{\infty} = \frac{3}{1 - \frac{1}{2}} = 6$.

Worked example 2.3.1

Let $u_n = \frac{5}{(-3)^{n-1}}$ denote the n 'th term of a geometric sequence. Find S_∞ , the sum to infinity of $\{u_n\}$.

$u_1 = 5$ and $r = -\frac{1}{3}$. It's clear that $|r| = \frac{1}{3} < 1$. Hence:

$$S_\infty = \frac{5}{1 + \frac{1}{3}} = \frac{15}{4}$$

For situations where $|r| \geq 1$, the sum to infinity formula may still give you an answer. Be aware though that this will be false. For example, consider $u_n = 2^n$. This obviously diverges to infinity as n approaches infinity. However, $S_\infty = \frac{2}{1-2} = -2$, which is absurd! In an exam, be sure to check and state that $|r| < 1$ before finding a sum to infinity (if not already stated in the question). This ensures you have justification for finding a sum to infinity.

Further reading: Cesàro sums

A *Cesàro sum* is the limit as n approaches infinity of the mean of the first n *partial sums* of a sequence (if it exists). A partial sum is simply $S_n = \sum_{k=1}^n u_k$ for a sequence $\{u_n\}$. Cesàro sums were named after the Italian mathematician *Ernesto Cesàro*.

Consider the sequence $\{1, -1, 1, -1, 1, -1, \dots\}$, with n 'th term $u_n = (-1)^{n-1}$. You can probably deduce (easily) that this sequence diverges and hence the series also diverges. However, a Cesàro sum C_s exists:

$$S_n = \sum_{k=1}^n u_k = \begin{cases} 1, & n \text{ is odd} \\ 0, & n \text{ is even} \end{cases}$$

$$C_n = \frac{1}{n} \sum_{k=1}^n S_k$$

$$\{C_n\} = \left\{1, \frac{1}{2}, \frac{2}{3}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, \frac{4}{7}, \dots\right\}$$

$$C_n = \begin{cases} \frac{n+1}{2n}, & n \text{ is odd} \\ \frac{1}{2}, & n \text{ is even} \end{cases}$$

$$\lim_{n \rightarrow \infty} \frac{n+1}{2n} = \frac{1}{2}$$

Hence, $C_s = \lim_{n \rightarrow \infty} C_n = \frac{1}{2}$, which is the middle value between 0 and 1.

Note that this does **not** imply that the series $\{1, -1, 1, -1, \dots\}$ converges! Only that it is *Cesàro summable*.

There are also sequences which aren't Cesàro summable (i.e. a Cesàro sum doesn't exist). An example of a sequence with this property is $\{1, 2, 4, 8, 16, \dots\}$. Is there some criteria for which a sequence is or isn't Cesàro summable? It's worth exploring.

Drill Exercises

1 Find the sums to infinity of the following geometric sequences, expressing your answers in exact form:

(a) $\{400, 200, 100, 50, 25, \dots\}$

(d) $\{\pi, \sqrt{\pi}, 1, \frac{1}{\sqrt{\pi}}, \frac{1}{\pi}, \dots\}$

(b) $\{-3^{-1}, 3^{-2}, -3^{-3}, 3^{-4}, -3^{-5}, \dots\}$

(e) $\{9, 3, 1, \frac{1}{3}, \frac{1}{9}, \dots\}$

(c) $\{4, -2, 1, -\frac{1}{2}, \frac{1}{4}, \dots\}$

(f) $\{40, 20, 10, 1, \frac{1}{10}, \dots\}$

Word Problems

3 Komi drops a rubber ball on the ground from a height of 1.5 m. Each time it bounces back up, it reaches 80% of the height reached on the last bounce. Calculate the total distance the ball travels.

4 Using the property of geometric series convergence, show that any number x (where $0 < x < 1$) with a recurring group of decimal digits (i.e. $x = 0.\overline{d_1 d_2 d_3 \dots d_n}$) is rational.

5 Use question 4 to convert the following decimal numbers to fractional form:

(a) 0.33333333...

(c) 1.307692307692307692...

(b) 0.6969696969696969...

(d) 0.86538461538461538461...

2.4 Applications: simple and compound interest

Arithmetic and geometric sequences have applications in the real world. An example is calculating *interest*. This section will introduce the concept of interest and how to make such calculations.

If someone wants to take a loan from a bank in order to do something like buy a house, they may do so. However, they must pay the bank back at some point in time and usually must pay back **more** than they loaned. Interest is the charge for loaned money; the additional amount you must pay back on top of what you have loaned.

Another situation where interest may apply is when someone makes an investment. They invest a certain amount of money and in time, the money they put in grows. After a certain amount of time, they may withdraw what they have invested and make a profit.

The amount charged on a loan or investment depends on the *interest rate*. This describes the percentage of the initial loan or investment (such as 5%) which must be paid in specific time intervals (such as yearly or monthly). Note that it may help to express the rate as a decimal value rather than a percentage to make calculations easier (e.g. $5\% = 0.05$).

The initial value loaned or invested is called the *principal amount*.

There are two types of interest calculations covered in the course. These types are explained below:

Simple interest

Suppose a principal amount of $\$P$ is loaned/invested at a rate of r (so $100r\%$) for t periods (e.g. t months or t years). Then the *simple interest* I is calculated by:

$$I = Prt$$

and the total amount owed/obtained A is calculated by:

$$A = P + I = P + Prt = P(1 + rt)$$

The simple interest is calculated as a fraction r of the principal amount $\$P$ and is constant ($\$Pr$). Multiplying this constant by a period t yields the total simple interest after that period of time.

Simple interest is usually used for short-term loans. Observe that if $u_t = P(1 + r(t - 1))$, then $\{u_t\}$ is an arithmetic sequence with first term P and common difference Pr .

Compound interest

Suppose a principal amount of $\$P$ is loaned/invested at a rate of r for a period of t . Then the *compound interest* is calculated as a percentage of the previously-calculated balance and the total amount owed/obtained A is calculated by:

$$A = P(1 + r)^t$$

Additionally, if the interest is compounded n times per period t , then A is calculated by:

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

Compound interest is usually used for long-term loans and grows at a faster rate than simple interest. Observe that if $u_t = P \left(1 + \frac{r}{n} \right)^{n(t-1)}$, then $\{u_t\}$ is a geometric sequence with first term P and common ratio $\left(1 + \frac{r}{n} \right)^n$.

Also, interests can be compounded a fixed number of times during a period. Here are some common frequency terms for a period of one year:

Term	Meaning	n
Annually	Once per year	1
Semi-annually	Twice per year	2
Quarterly	Four times per year	4
Monthly	Twelve times per year	12

Worked example 2.4.1

Simon takes a loan of $\$15\,000$ from his bank to buy a car, with a monthly simple interest rate of 5%. Calculate how much Simon owes the bank after a year.

Note that the simple interest is calculated monthly for a year. Hence, $t = 12$. $r = 0.05$, so using the simple interest formula:

$$A = P(1 + rt) = 15000(1 + 0.05 \cdot 12) = 24000$$

Hence, Simon owes the bank \$24 000 after a year.

Worked example 2.4.2

Chika invests ¥11 000 (Japanese yen) in a bank with an annual compound interest rate of 7%, compounded quarterly. Calculate the profit she will make after 10 years. Give your answer correct to the nearest yen.

Total amount after 10 years:

$$A = P\left(1 + \frac{r}{n}\right)^{nt} = 11000\left(1 + \frac{0.07}{4}\right)^{4 \cdot 10} = 22017.57$$

$$22017.57 - 11000 = 11017.57$$

Hence, the profit Chika makes is ¥11 018 rounded to the nearest yen.

Further reading: e

Consider the compound interest formula:

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

Suppose \$1 is invested into a bank account for a year, compounded n times at a rate of 100%:

$$A = \left(1 + \frac{1}{n}\right)^n$$

Now try inputting this expression into a calculator for different values of $n \geq 1$:

n	A
1	2
2	2.25
5	2.49
10	2.59
20	2.65
100	2.70

1000	2.717
10000	2.718

As n is increased, A is indeed getting closer to the constant e . This is actually a **definition** of e . More formally:

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

We shall see other definitions of e later in the course.

Word Problems

1 Walker wants to invest \$6000 in a bank which provides an annual compound interest rate of $r\%$, compounded **monthly**. Given that Walker's investment doubles in value after 5 years, find the value of r correct to 3 s.f.

2 Kashif takes a loan of \$14000 from his bank which has an annual compound interest rate of 4%, compounded at the end of each year. To help pay off his loan, he puts in $\$D$ at the end of every year into a savings account with an annual compound interest rate of 3%, also compounded yearly. He starts doing this the same year he takes the loan.

(a) Find an expression for the total amount in Kashif's savings account after n years.

(b) Hence find the value of D for which Kashif can pay off his loan after 10 years.

3 Polynomials

Polynomials are probably the most fundamental expressions in algebra. In fact, the *Fundamental Theorem of Algebra* is a statement about polynomials. A general polynomial has *variables* and *coefficients*. Some examples are:

- $x^3 + 3x^2 - 6x + 4$
- $x^2 + 2xy + y^2$
- 1

To remove ambiguity, it helps to identify the variables and coefficients (especially if a polynomial has multiple variables). This course will only study polynomials of one variable (usually x). These will have the form:

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

Where $x \in \mathbb{R}$ is the variable and $\{a_0, a_1, \dots, a_n\} \subseteq \mathbb{R}$ are the coefficients (constants in \mathbb{R}). Of course, we can use another set instead of \mathbb{R} such as \mathbb{Z} , \mathbb{Q} or \mathbb{C} .

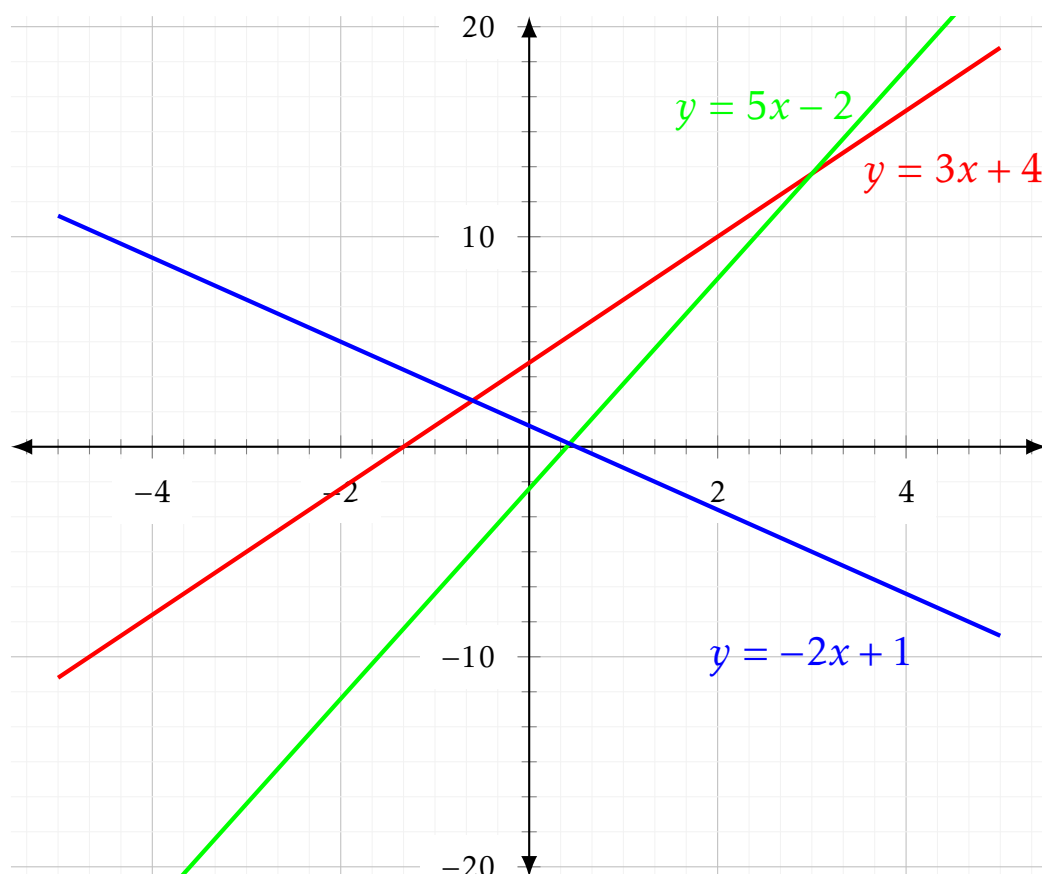
Some common polynomials you may be familiar with are:

- $ax + b$ (line)
- $ax^2 + bx + c$ (quadratic)

By setting a polynomial to a variable y or a function $f(x)$, we can plot polynomials on a graph and analyze their behavior. We will have a look at what some of these look like in this chapter.

3.1 Lines

You may (or should) be familiar with what a straight line equation looks like. A line, by definition, is the set of all points (x, y) for which $y = mx + c$ (you may be familiar with this form), where m is the *gradient* and c is the y -intercept of the line. Of course, x is the variable and m and c are coefficients. This is in fact a type of polynomial, as it has the form described above in the introduction. Here are some examples of straight line equations plotted on a graph:



You may ask yourself “where do these coefficients come from and how do they form a straight line?” Firstly, it may be obvious that c is the y -intercept because if you set $x = 0$, you get $y = c$, which is the point $(0, c)$. Thus, it is the point of the line which lies on the y -axis. Similarly if you set $y = 0$, you get $mx + c = 0$ which implies $x = -\frac{c}{m}$ (assuming $m \neq 0$). This point $(-\frac{c}{m}, 0)$ is the point of the line which lies on the x -axis.

As for why m is the gradient, we must recall what is meant by a gradient. You may know it as the “rise over run” of a line. In other words, if you take two **distinct** points (x_1, y_1) and (x_2, y_2) on a line, the gradient is $m = \frac{y_2 - y_1}{x_2 - x_1}$. This describes how much y changes when x changes by 1 unit. In fact, take the point $(0, c)$ and the arbitrary point (x, y) (for $x \neq 0$) and observe that:

$$m = \frac{y - c}{x - 0} = \frac{y - c}{x} \implies mx = y - c \implies y = mx + c$$

Indeed, we get the straight line equation. A few other forms of the equation are:

- $ax + by + d = 0$ where $a, b, d \in \mathbb{R}$ are constants
- $y - y_1 = m(x - x_1)$ where m is the gradient and (x_1, y_1) is a point on the line

The derivations of these forms are left as an exercise if you are interested; they are not at all hard to find.

Worked example 3.1.1

Let $m = 3$ be a gradient of a line with a point $(\frac{1}{3}, \frac{4}{3})$. Write down the equation of the line in the form $ax + by + d = 0$ where $a, b, d \in \mathbb{Z}$ are constants to be found.

From the form $y - y_1 = m(x - x_1)$, we have that $y - \frac{4}{3} = 3(x - \frac{1}{3})$, so $y = 3x + \frac{1}{3}$.

Multiplying both sides by 3, we get $3y = 3x + 1$.

Thus $3x - 3y + 1 = 0$ or $-3x + 3y - 1 = 0$ (both equations are equivalent).

We can also find the point for which two lines L_1 and L_2 might *intersect*. The procedure is to take two equations: $L_1 : y = m_1x + c_1$ and $L_2 : y = m_2x + c_2$ and simply equate the right-hand sides like so: $m_1x + c_1 = m_2x + c_2$. This means we have equated the y values and must find an x value which satisfies the equation. Then we use our found x value and plug it back into either of L_1 or L_2 to find the corresponding y value. To illustrate this, consider the example:

Worked example 3.1.2

Find the point of intersection, P , of the two lines L_1 and L_2 defined by:

$$L_1 : y = 3x + 5$$

$$L_2 : y = x - 5$$

Equate the two right-hand sides: $3x + 5 = x - 5$

Then solve for x : $2x = -10 \implies x = -5$

Finally, plug this value of x back into L_1 : $y = 3 \cdot (-5) + 5 = -10$

Hence, the point of intersection is $P(-5, -10)$

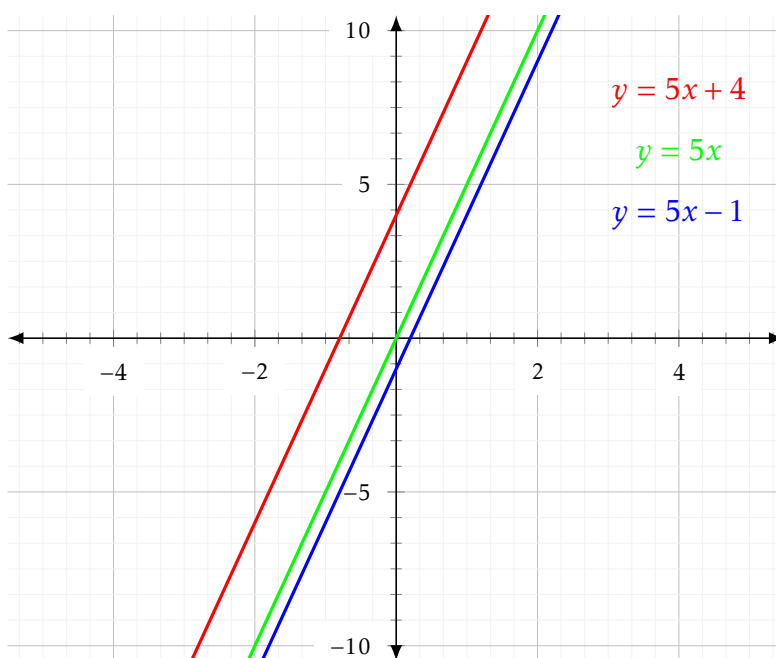
You may have noticed (or not) that if two equations had the same gradient, problems may arise when trying to find points of intersection. Suppose we have $y = mx + c_1$ and $y = mx + c_2$, both of

which have the same gradient m . Then equating these would yield: $mx + c_1 = mx + c_2 \implies c_1 = c_2$, which wouldn't be a problem as long as $c_1 = c_2$, but clearly that's not always the case. If $y = 3x + 4$ and $y = 3x - 5$ and we equate them: $3x + 4 = 3x - 5 \implies 4 = -5$, which is a clear contradiction. This simply means that the two lines never meet. We denote this as *parallelism* and say that the two lines are *parallel* to each other.

Parallel lines

Two lines $L_1 : y = m_1x + c_1$ and $L_2 : m_2x + c_2$ are *parallel* if and only if $m_1 = m_2$ and $c_1 \neq c_2$.

Some parallel lines are graphed below:

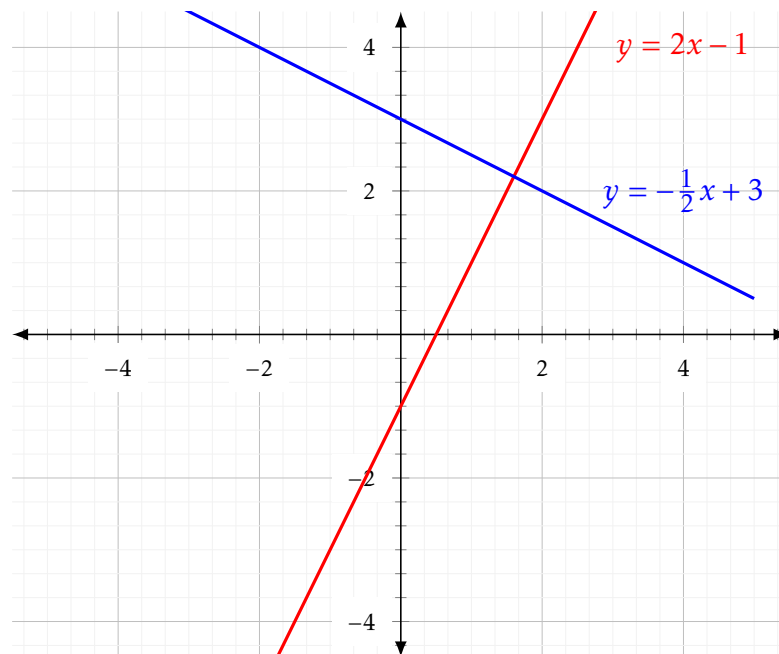


If two lines are such that they meet at a right angle, we say they are *perpendicular*. There is indeed a gradient condition for this property to hold:

Perpendicular lines

Two lines $L_1 : y = m_1x + c_1$ and $L_2 : m_2x + c_2$ are *perpendicular* if and only if $m_1 m_2 = -1$.

Two perpendicular lines are graphed below:

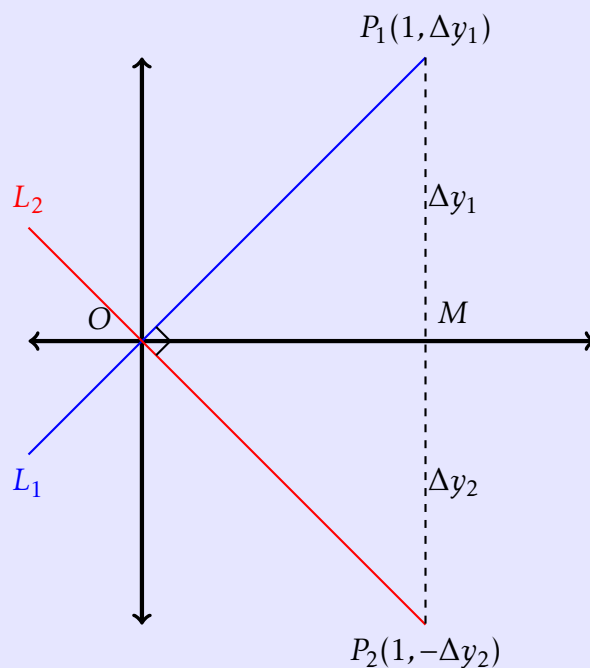


The proof of the perpendicular line fact is given below (you will **not** need to know it for the exam):

Theorem. For two lines $L_1 : y = m_1x + c_1$ and $L_2 : y = m_2x + c_2$ ($m_1, m_2 \neq 0$), L_1 and L_2 are perpendicular if and only if $m_1m_2 = -1$

Proof. Note: \iff means 'if and only if'

Suppose L_1 and L_2 meet at a right angle:



Where $m_1 = \Delta y_1$ and $m_2 = -\Delta y_2$.

By Pythagoras' theorem, $OP_1 = \sqrt{\Delta y_1^2 + 1}$ and $OP_2 = \sqrt{\Delta y_2^2 + 1}$.

$$\text{Also, } (OP_1)^2 + (OP_2)^2 = (\Delta y_1 + \Delta y_2)^2$$

$$\iff (\Delta y_1)^2 + 1 + (\Delta y_2)^2 + 1 = (\Delta y_1)^2 + (\Delta y_2)^2 + 2(\Delta y_1)(\Delta y_2)$$

$$\iff 2(\Delta y_1)(\Delta y_2) = 2$$

$$\iff m_1(-m_2) = 1$$

$$\iff m_1 m_2 = -1 \text{ as required.}$$

The converse clearly holds due to equality of the statements, so the theorem must be true. ■

Worked example 3.1.3

Let $L_1 : y = kx + 4$ and $L_2 : y = 2x + 3$. Find the value of k for which:

- (a) L_1 and L_2 are parallel
- (b) L_1 and L_2 are perpendicular

(a) We have that $4 \neq 3$, so we need $k = 2$ for the lines to be parallel.

(b) We must have that $2k = -1 \implies k = -\frac{1}{2}$ for the lines to be perpendicular.

Further reading: Equivalence relations

There is some debate on whether or not we need the condition that $c_1 \neq c_2$. That is, whether a line is **parallel to itself**. Euclid (the 300 BC Greek mathematician and founder of geometry) has defined parallelism such that a line can't be parallel to itself. However, others have allowed lines to be parallel to themselves so that parallelism could be an *equivalence relation*.

What we mean by a relation is an operator which describes how two (or more) objects are related. For example, equality ($=$) is a relation between two values x and y (if $x = y$, we say x is related to y by equality). Other relations include $>$ and \equiv .

A relation R is an *equivalence relation* if R is *reflexive* (xRx), *symmetric* ($xRy \implies yRx$) and *transitive* (xRy and $yRz \implies xRz$) for all values x, y, z in a specified set. It is obvious that $=$ is an equivalence relation, but not $>$ (since it isn't reflexive or symmetric).

As for parallelism, denote the relation to be \parallel (i.e. $L_1 \parallel L_2$ means the lines L_1 and L_2 are parallel) and suppose we say that for $L_1 : y = m_1x + c_1$ and $L_2 : y = m_2x + c_2$, $L_1 \parallel L_2$ if and only if $m_1 = m_2$. Let $L_3 = m_3x + c_3$. Then we have:

Reflexivity: $L_1 \parallel L_1 \implies m_1 = m_1$, which is clearly true.

Symmetry: $L_1 \parallel L_2 \implies m_1 = m_2 \implies m_2 = m_1 \implies L_2 \parallel L_1$.

Transitivity: $L_1 \parallel L_2 \implies m_1 = m_2$ and $L_2 \parallel L_3 \implies m_2 = m_3 \implies m_1 = m_2 = m_3 \implies m_1 = m_3 \implies L_1 \parallel L_3$

By that definition of parallelism, it is indeed an equivalence relation. However, you can see that the reflexive property doesn't hold when we introduce the condition that $c_1 \neq c_2$.

Drill Exercises

1 Convert the following line equations to the form $ax + by + d = 0$ where $a, b, d \in \mathbb{Z}$.

(a) $y = 3x + 4$

(c) $y = \frac{3}{8}x - \frac{2}{3}$

(b) $y = \frac{5}{4}x - 1$

(d) $y - 3 = 4(x - 1)$

(e) $y + 3 = 2(x + 5)$

(f) $\sqrt{2}y + \frac{\sqrt{2}}{2}x = \frac{1}{\sqrt{2}}$

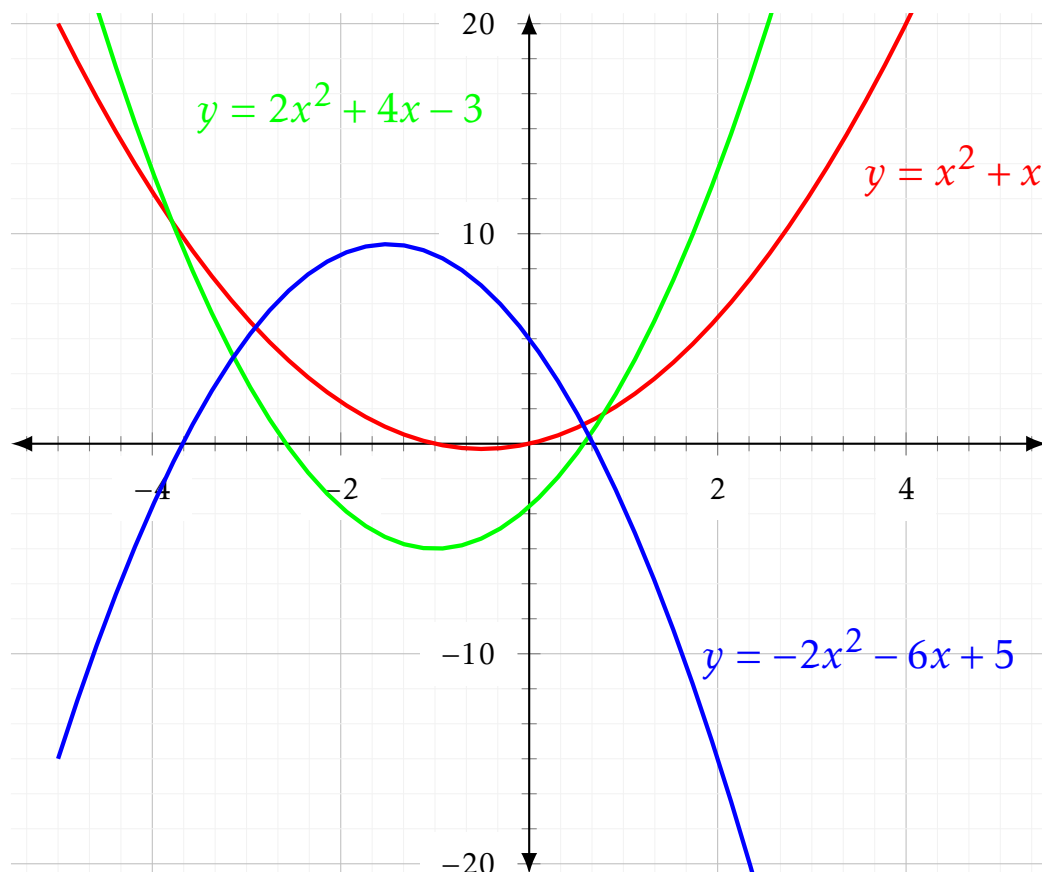
Word Problems

3 While the proof of perpendicularity shown above is correct, it assumes that the gradients of the lines are non-zero. If **one** of the lines has a gradient of zero (i.e. $L_1 : y = c_1$), explain what equation the other line L_2 must have for L_1 and L_2 to be perpendicular.

2 Given any arbitrary point (x_1, y_1) and a gradient $m \neq 0$, explain how a straight line equation of the form $y = mx + c$ can be found and prove that this line is unique (that is, no other line of the same gradient can pass through that point).

3.2 Quadratics

A quadratic is a polynomial of the form $ax^2 + bx + c = 0$ where $a \neq 0$. If $a = 0$, clearly we would have a straight line. Some quadratics are plotted below:

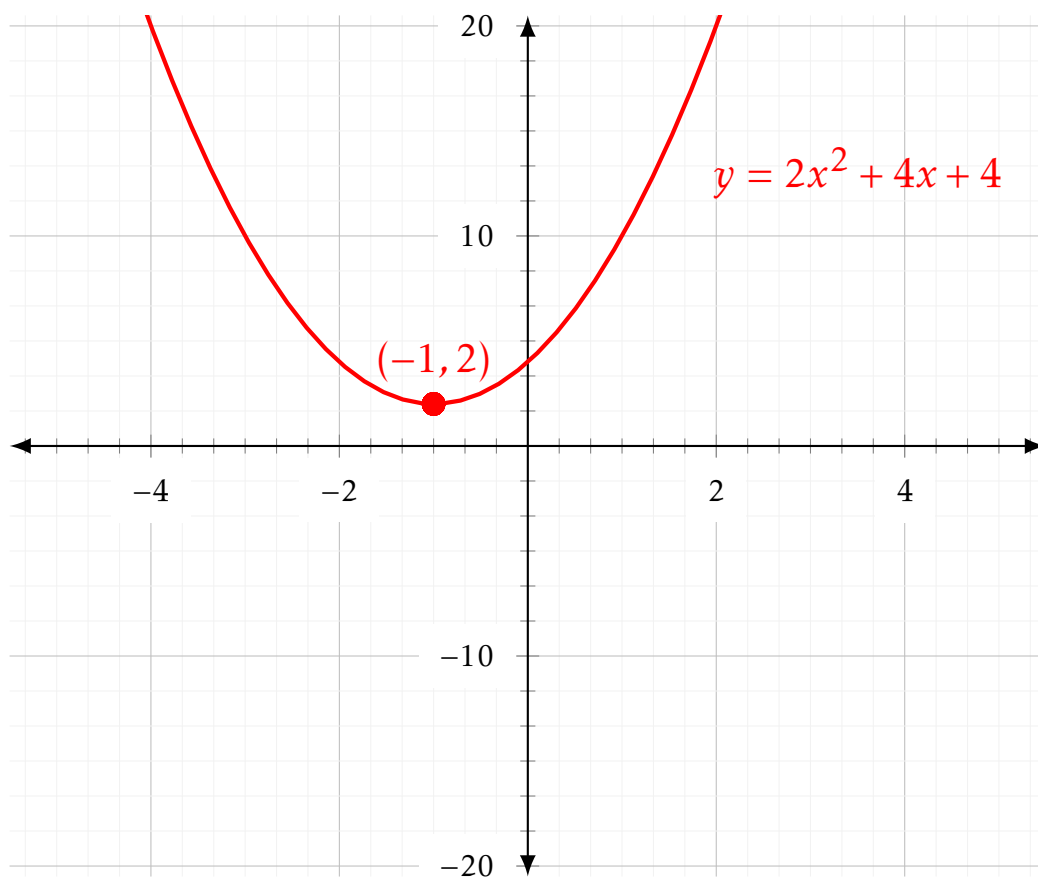


From an analytic point of view, there is a lot more to quadratics than straight lines. You will notice that quadratics have a variable gradient for starters (we will see how we can obtain this using calculus later on). Another is that quadratics in \mathbb{R} will not span the whole of \mathbb{R} . That is, every quadratic in \mathbb{R} has either a least value or a greatest value (the point with this value is called the *vertex*).

Vertex of a quadratic

For a quadratic ax^2+bx+c , the vertex is the point for which the quadratic attains a maximum (when $a < 0$) or minimum (when $a > 0$) value. It is also the only point on the quadratic for which the gradient is 0.

The vertex for the quadratic equation $y = 2x^2 + 4x + 4$ is shown below:



There is a way to obtain the exact coordinates of a vertex algebraically using a quadratic expression. It involves a process you may be familiar with called *completing the square*.

Completing the square

A quadratic of the form $ax^2 + bx + c$ (for $a, b, c \in \mathbb{R}$, $a \neq 0$) can be written in the form $a(x - h)^2 + k$ (for $h, k \in \mathbb{R}$) where the vertex has coordinates (h, k) . The process is as follows:

1. Factor out the coefficient a from the x^2 and x terms: $a(x^2 + \frac{b}{a}x) + c$
2. Notice that $(x + \frac{b}{2a})^2 = x^2 + \frac{b}{a}x + (\frac{b}{2a})^2$
3. Hence $x^2 + \frac{b}{a}x$ can be written as $(x + \frac{b}{2a})^2 - (\frac{b}{2a})^2$
4. Hence the quadratic can be written as $a((x + \frac{b}{2a})^2 - (\frac{b}{2a})^2) + c$
5. Multiply a and simplify (if needed): $a(x + \frac{b}{2a})^2 - \frac{b^2}{4a} + c$
6. The vertex will have coordinates $(-\frac{b}{2a}, c - \frac{b^2}{4a})$

Notice the $-h$ in the expression and take care when writing down the coordinates of the vertex (for example, a quadratic $2(x + 3)^2 + 2$ will have its vertex at $(-3, 2)$).

Worked example 3.2.1

Let $y = 3x^2 - 6x + 2$ be a quadratic equation.

(a) Convert the quadratic to completed square form.

(b) Hence write down the coordinates of its vertex.

(a) $3x^2 - 6x + 2$

$$= 3(x^2 - 2x) + 2$$

$$= 3((x - 1)^2 - 1) + 2$$

$$= 3(x - 1)^2 - 1$$

(b) The vertex has coordinates: $(1, 0)$

We can also *factorize* some quadratics by writing them in the form $a(x - p)(x - q)$ where $a, p, q \in \mathbb{R} (a \neq 0)$. Not all quadratics in \mathbb{R} have this property (but they all do in \mathbb{C}). Factorizing is a skill which takes some intuition and a bit of “trial and error”. By setting the factorized quadratic to zero, we can easily find values of x for which the quadratic is zero (these will clearly be $x = p$ and $x = q$ and p and q could be the same value as well). We call these the *roots* or the *zeros* of the quadratic. This key point may help:

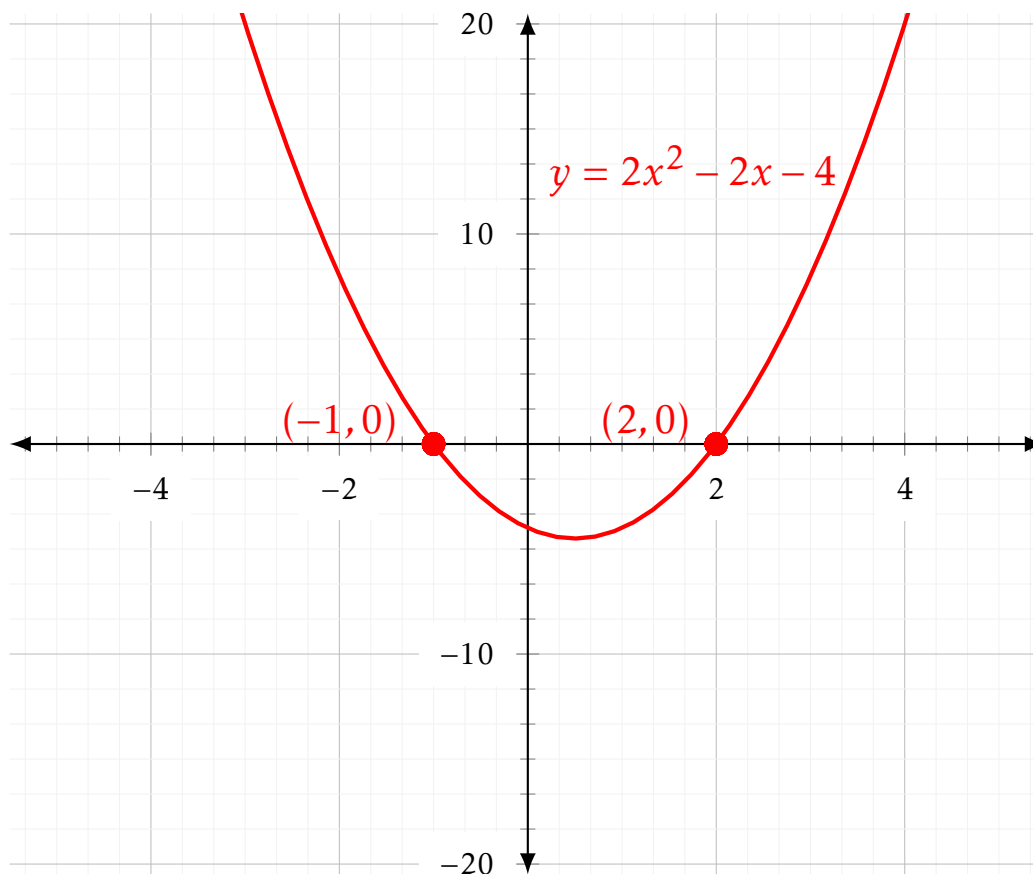
Product and sum of quadratic roots

For a quadratic $ax^2 + bx + c$ with roots p and q :

- $pq = \frac{c}{a}$
- $p + q = -\frac{b}{a}$

Note that these formulas will still hold even when $p = q$, and the roots will be counted twice.

This fact is easy to see when multiplying out the quadratic’s factorized form and equating the coefficients. Some roots of a quadratic are shown below on a graph:



Worked example 3.2.2

By factorizing $x^2 + 5x + 6$, find the values of x for which $x^2 + 5x + 6 = 0$.

Let $x^2 + 5x + 6 = a(x-p)(x-q)$. Since $a = 1$, we know that $pq = 6$ and $p+q = 3$. By intuition, $p = 3$ and $q = 2$. Hence, the roots are $x = 3$ and $x = 2$.

If a particular quadratic looks difficult to factorize, you can complete the square to solve for the roots:

Worked example 3.2.3

Find the roots of $x^2 + x - 1$ in exact form.

Complete the square:

$$\begin{aligned}x^2 + x - 1 &= \left(x + \frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - 1 \\&= \left(x + \frac{1}{2}\right)^2 - \frac{5}{4}\end{aligned}$$

Set the expression to 0 and solve for x :

$$\begin{aligned}\left(x + \frac{1}{2}\right)^2 - \frac{5}{4} &= 0 \\ \left(x + \frac{1}{2}\right)^2 &= \frac{5}{4} \\ x + \frac{1}{2} &= \pm \frac{\sqrt{5}}{2} \text{ (since squaring a negative number gives a positive number)} \\ x &= \frac{-1 \pm \sqrt{5}}{2}\end{aligned}$$

Hence the roots are $x = \frac{\sqrt{5}-1}{2}$ and $x = -\frac{\sqrt{5}+1}{2}$

You may find it helpful to plug the roots you've found back into the original equation. The expression will evaluate to 0 if you've found the correct ones. You can also find a factorized expression for a quadratic using the roots found. For the above example, the quadratic $x^2 + x - 1$ can be written as $\left(x - \frac{\sqrt{5}-1}{2}\right)\left(x + \frac{\sqrt{5}+1}{2}\right)$.

While completing the square is one way of finding quadratic roots, there is a method to obtain them directly which comes from this process. This is a formula in terms of quadratic coefficients

called the *quadratic formula* which you may be familiar with already. Here's the formula and the proof:

Theorem. For a quadratic of the form $ax^2 + bx + c$ where $a \neq 0$, the roots are:

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

Proof. Start with $ax^2 + bx + c$ and complete the square to obtain:

$$a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c$$

Set this expression to 0 and solve for x :

$$a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c = 0$$

$$a\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a} - c$$

$$a\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a}$$

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

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

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

■

This equations works for any quadratic in \mathbb{R} or even \mathbb{C} as long as $a \neq 0$.

We can now analyze the quadratic equation; in particular the expression $b^2 - 4ac$. We call this the *discriminant* and use the variable Δ (i.e. $\Delta = b^2 - 4ac$).

Discriminant of a quadratic

For a quadratic $ax^2 + bx + c$, the *discriminant* $\Delta = b^2 - 4ac$ can be used to determine the nature of the quadratic roots as follows:

- If $\Delta > 0$, there are exactly **two distinct real roots**.
- If $\Delta = 0$, there is exactly **one real root** (a repeated root).
- If $\Delta < 0$, there are **no real roots**.

This is quite clear to see due to the symmetry of the square root in the quadratic equation. Also, the square root of a negative number is undefined in \mathbb{R} , so $\Delta < 0$ results in no real roots.

Worked example 3.2.4

Consider the quadratic $kx^2 + (k-1)x + (k-1)$. Find the range of values of k for which the quadratic has exactly two distinct real roots.

We can use the discriminant condition that $\Delta > 0$:

$$\begin{aligned}\Delta &= (k-1)^2 - 4(k)(k-1) \\ &= (k-1)(k-1-4k) \\ &= (1-k)(3k+1)\end{aligned}$$

Then: $(1-k)(3k+1) > 0$, so either:

1. $1-k > 0$ and $3k+1 > 0$
2. $1-k < 0$ and $3k+1 < 0$

For the first case, $k < 1$ and $k > -\frac{1}{3}$

For the second case, $k > 1$ and $k < -\frac{1}{3}$, which is not possible

Hence, the range of values of k for which the quadratic has exactly two roots is $-\frac{1}{3} < k < 1$

Speaking of symmetry, the vertex's x -value gives the equation of the *axis of symmetry* of a quadratic. This denotes the line of symmetry for a quadratic graph.

Axis of symmetry

For a quadratic $ax^2 + bx + c$, the *axis of symmetry* is the line $x = -\frac{b}{2a}$

Worked example 3.2.5

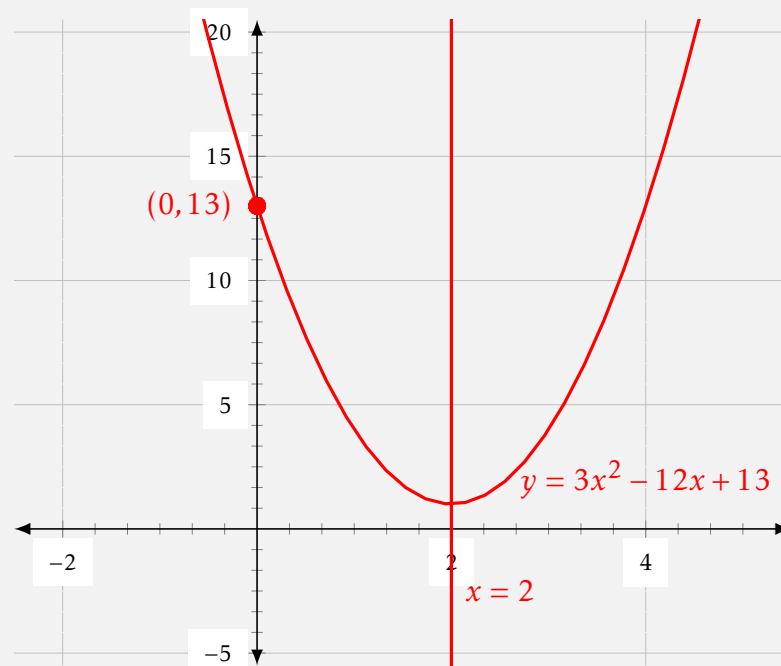
Consider the quadratic $3x^2 - 12x + 13$.

- (a) Express the quadratic in completed square form
- (b) Hence write down the equation of the quadratic's axis of symmetry
- (c) Hence sketch the graph of $y = 3x^2 - 12x + 13$ and draw the axis of symmetry.

(a) $3x^2 - 12x + 13$
 $= 3(x^2 - 4x) + 13$
 $= 3((x - 2)^2 - 4) + 13$
 $= 3(x - 2)^2 + 1$

(b) The equation of the axis of symmetry is $x = 2$

(c) The graph is shown below:



Drill Exercises

1 Find the real roots of the following polynomials using any method:

(a) $4x - 5$

(d) $15x^2 - 45x + 26$

(b) $x^2 + 4x + 4$

(e) $-x^2 + 16x + 19$

(c) $2x^2 - 4x - 6$

(f) $3x^3 + 5x^2 - 4x$

2 Find the point(s) of intersection (if any) for the following pairs of quadratic equations:

(a) $y = x^2 - 3x + 2$ and $y = x^2 + 5x - 7$

(d) $y = 5x^2 + 15x + 3$ and $y = -8x^2 + 4x + 6$

(b) $y = x^2 + 4x + 5$ and $y = -x^2 - 4x - 3$

(e) $y = x^2$ and $y = 2x^2 + 8$

(c) $y = 3x^2 + 8x + 1$ and $y = -x^2 + 5x - 5$

(f) $y = x^2 - 6x + 9$ and $x = y^2 - 6y + 9$

3 For each of the following quadratics, use the discriminant to determine the nature of its roots (i.e. two distinct real roots, one real root or no real roots):

(a) $3x^2 - 6x - 9$

(d) $y = 5x^2 + 15x + 3$ and $y = -8x^2 + 4x + 6$

(b) $2x^2 + 2x + 2$

(e) $y = x^2$ and $y = 2x^2 + 8$

(c) $2x^2 + 2x + 2$

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

Word Problems

4 Consider the quadratic $4x^2 + (3k - 2)x - (k + 3)$ where $k \in \mathbb{R}$.

(a) Determine the range of values of k for which the quadratic has two distinct real roots.

(b) Hence determine the value of k for which the roots are closest together.

5 For the quadratic $4x^2 - 10x + k$ ($k \in \mathbb{R}$), one root is four times the other root. Find the value of k .

6 Mahir kicks a football. The height of the football, h , can be modeled with respect to its horizontal distance, d , by the equation $h = d - 0.04d^2 + 0.2$. Assume both h and d are measured in meters. Find the maximum height reached by the ball.

3.3 The binomial theorem

Consider the expansion of $(a + b)^2$. It should be fairly obvious that this will be $a^2 + 2ab + b^2$. What about $(a + b)^3$? With some work, you'll find it is $a^3 + 3a^2b + 3ab^2 + b^3$. For higher integral powers, you will find it becomes harder to expand these increasing expressions. But there is an easier way of finding coefficients of each term for higher integral powers of n .

First, it is worth looking at what a factorial is:

The factorial function

For $n \in \mathbb{Z}$ where $n \geq 0$, we define $n!$ to be the *factorial* of n given by the function:

$$n! = f(n) = nf(n-1), \quad f(0) = 1$$

or in other words:

$$n! = 1 \cdot 2 \cdot 3 \cdots n \text{ for } n \geq 1 \text{ and } 0! = 1$$

For example, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. Then the binomial theorem is as follows:

Binomial theorem

For $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$

$$(a + b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r}$$

Where $\binom{n}{r}$ is the *binomial coefficient* defined by $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ for $0 \leq r \leq n$.

You do not need to know the proof, but it is shown below if you are interested:

Theorem. For $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$, $(a + b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r}$

Proof. An inductive proof seems best. For more information on how this works, look up *proof by induction*. HL students will learn this proof technique later in the course.

Let $P(n)$ denote the above statement for $n \in \mathbb{N}$ and consider: $P(0)$:

$$\text{LHS: } (a+b)^0 = 1$$

$$\text{RHS: } \sum_{r=0}^0 \binom{0}{r} a^r b^{-r} = 1 \cdot a^0 b^0 = 1$$

LHS = RHS, so $P(0)$ is true.

$$\text{Assume } P(k) \text{ is true for some } k \in \mathbb{N}: (a+b)^k = \sum_{r=0}^k \binom{k}{r} a^r b^{k-r}$$

$$\begin{aligned} \text{Then consider } (a+b)^{k+1} &= (a+b)(a+b)^k = (a+b) \sum_{r=0}^k \binom{k}{r} a^r b^{k-r} \\ &= \sum_{r=0}^k \binom{k}{r} a^{r+1} b^{k-r} + \sum_{r=0}^k \binom{k}{r} a^r b^{k+1-r} \\ &= \sum_{r=1}^{k+1} \binom{k}{r-1} a^r b^{k+1-r} + \sum_{r=0}^k \binom{k}{r} a^r b^{k+1-r} \\ &= \sum_{r=0}^k \binom{k}{r-1} a^r b^{k+1-r} + \sum_{r=0}^{k+1} \binom{k}{r} a^r b^{k+1-r} - \binom{k}{-1} a^0 b^{k+1} - \binom{k}{k+1} a^{k+1} b^0 \\ &= \sum_{r=0}^k \left(\binom{k}{r-1} + \binom{k}{r} \right) a^r b^{k+1-r} \\ &= \sum_{r=0}^k \binom{k+1}{r} a^r b^{k+1-r} \end{aligned}$$

Therefore, $P(k+1)$ is true assuming $P(k)$ is true.

Hence, $P(n)$ is true for all values of $n \in \mathbb{N}$ by induction. ■

Notice that we had $\binom{k}{k+1}$ and $\binom{k}{-1}$ in the proof. These, by convention, both evaluate to zero.

To see why, consider:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\cdots(n-(r-1))}{r(r-1)(r-2)\cdots 1}$$

$$\text{Then } \binom{k}{k+1} = \frac{k(k-1)(k-2)\cdots 0}{k(k-1)\cdots 1} = 0$$

$$\text{Similarly } \binom{k}{-1} = \binom{k}{k+1} = 0 \text{ (by symmetry)}$$

The binomial coefficient should be explored further. It seems difficult to generate $\binom{n}{r}$ for some $n, r \in \mathbb{N}$ by hand, but actually it is easy. In fact, you can **very** easily generate all the coefficients of the expansion of $(a + b)^n$. It is a pattern called *Pascal's triangle*:

$$\begin{array}{ccccccc}
 & & & & 1 & & \\
 & & & 1 & & 1 & \\
 & & 1 & & 2 & & 1 \\
 & 1 & & 3 & & 3 & & 1 \\
 & & 1 & & 4 & & 6 & & 4 & & 1 \\
 & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
 & & & & & 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
 \end{array}$$

The above shows Pascal's triangle for $n = 0$ to $n = 7$, but can be continued. Start the top of the triangle at 1 and then each element below is generated at the midpoint of two upper elements by adding them together. 1 is appended on the left and right of each row. Notice the third row ($n = 2$) gives the coefficients of the expansion of $(a + b)^2 = a^2 + 2ab + b^2$ and similarly the fourth row ($n = 3$) gives the coefficients of the expansion of $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$. You can verify that this pattern continues further. Notice the symmetry as well. The k 'th term coefficient is the same as that of the $n - k$ 'th term, which algebraically makes sense too.

This pattern comes from the fact that $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, which can be verified with some algebra. Pascal's triangle comes in handy during a paper 1 exam when you need to quickly obtain binomial coefficients for high powers.

Worked example 3.3.1

Find the coefficient of x^5 in the expansion $(x - 3)^8$.

By the binomial theorem, the x^5 term of the expansion will be:

$$\binom{8}{5} x^5 \cdot (-3)^{8-5} = 56 \cdot (-3)^3 \cdot x^5 = -1512x^5$$

Hence, the coefficient of x^5 is -1512 .

Most of questions you will get in the exam regarding the binomial theorem will be like the one above. They can make some of them trickier, though:

Worked example 3.3.2

Find the constant term in the expansion $\left(1 - \frac{2}{x}\right)^5 (x^2 + 1)^3$.

The above is equivalent to $(1 - 2x^{-1})^5 (x^2 + 1)^3$.

The possible powers of x for $(1 - 2x^{-1})^5$ are: $0, -1, -2, -3, -4, -5$

The possible powers of x for $(x^2 + 1)^3$ are: $0, 2, 4, 6$

We need to find combinations of power terms which multiply to give x^0 . These are:

- $Ax^0 \cdot Bx^0$
- $Ax^{-2} \cdot Bx^2$
- $Ax^{-4} \cdot Bx^4$

The constant term for the left expression is 1

The constant term for the right expression is 1

The x^{-2} term for the left expression is $\binom{5}{2}(-2)^2 x^{-2}$

The x^2 term for the right expression is $\binom{3}{1} x^2$

The x^{-4} term for the left expression is $\binom{5}{4}(-2)^4 x^{-4}$

The x^4 term for the right expression is $\binom{3}{2} x^4$

The constant term will be the sum of these multiplied combinations:

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

$$= 1 + 10 \cdot 4 \cdot 3 + 5 \cdot 16 \cdot 3$$

$$= 1 + 120 + 240 = 361$$

By narrowing down term powers which result from multiplication using the binomial theorem, we can avoid performing nasty expansions to obtain specific coefficients. We will look at the *extended binomial theorem* later on, which looks at binomial expansions with rational and negative integer powers.

Drill Exercises

1 Use Pascal's triangle (or otherwise) to expand the following:

(a) $(x + 1)^4$

(e) $(x^2 + 1)^5$

(b) $(x + 1)^5$

(f) $(x^3 - 4)^6$

(c) $(x + 1)^6$

(g) $\left(1 + \frac{1}{x}\right)^6$

(d) $(x + 1)^{10}$

(h) $\left(4 - \frac{2}{5x^2}\right)^9$

Word Problems

2 Show that $\binom{n}{r} = \prod_{k=1}^r \frac{n+1-k}{k}$

3 Show that $\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$

4 Show that $\sum_{r=0}^n \binom{n}{r}^2 = \binom{2n}{n}$

Hint: consider the expansion of $(x + 1)^{2n}$ and compare it to the expansion of $((x + 1)^n)^2$.

3.4 Polynomial theorems [HL]

This topic will cover a few major theorems for generic polynomials of one variable. Before looking at these, here's an important note:

Degree of a polynomial

The *degree* of a single-variable polynomial $P(x)$ is the highest power of x for which the coefficient is non-zero (this is denoted by $\deg(P)$ normally).

In general, for $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ where $a_n \neq 0$, $\deg(P) = n$.

The first theorem is the *remainder theorem*, which states the following:

Theorem. For a polynomial $P(x)$, the remainder when $P(x)$ is divided by $x - r$ is $P(r)$

Proof. This proof uses the *Euclidean Division Theorem*, which will not be proven here (as it isn't useful in this course). You can find the proof in **Appendix A** on page 93.

From this algorithm, there are polynomials $Q(x)$ and $R(x)$ for which $P(x) = (x-r)Q(x) + R(x)$, where $\deg(R) < \deg(Q)$

Since $P(r)$ is constant, $R(x)$ is constant (as $P(r) = R(x)$). This is the remainder when $P(x)$ is divided by $x - r$. ■

The other theorem is the *factor theorem*:

Theorem. For a polynomial $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ where x is a variable and a_0, \dots, a_n are coefficients, $x - r$ is a factor of $P(x)$ if and only if r is a root of $P(x)$.

Proof. \implies Suppose $x - r$ is a factor of $P(x)$. Then

$P(x) = (x - r)(b_{n-1} x^{n-1} + \dots + b_1 x + b_0)$ for some coefficients b_0, \dots, b_{n-1} .

Clearly $P(r) = 0$, so r is a root.

\Leftarrow Now suppose r is a root of $P(x)$, so $P(r) = 0$.

By the remainder theorem, there is some polynomial $Q(x)$ and a constant R for which:
 $P(x) = (x - r)Q(x) + R$.

Since $P(r) = 0$, $R = 0$ and therefore $P(x) = (x - r)Q(x)$, which means $x - r$ is a factor of $P(x)$. ■

To summarize:

Remainder and factor theorems

Let $P(x)$ be a polynomial. Then:

- When $P(x)$ is divided by $x - r$, the remainder is $P(r)$
- $x - r$ is a factor of $P(x)$ if and only if $P(r) = 0$

Worked example 3.4.1

Let $f(x) = 3x^4 + 3x^3 - 15x^2 - 9x + 18$

- (a) Show that $x - 1$ is a factor of $f(x)$.
- (b) Find the remainder when $f(x)$ is divided by $x + 2$

(a) $f(1) = 3 + 3 - 15 - 9 + 18 = 0$, which must imply $x - 1$ is a factor (by the factor theorem)

(b) By the remainder theorem, this will be

$$f(-2) = 3(-2)^4 + 3(-2)^3 - 15(-2)^2 - 9(-2) + 18 = 12$$

We can also obtain the product and sum of the roots of a polynomial directly in terms of its coefficients. Note that these count root multiplicities. For example if $P(x) = (x - r)^2$, P has one root r with multiplicity 2, so r is counted or multiplied twice to give a sum of $2r$ and product of r^2 .

Theorem. For a polynomial $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, the sum of the roots, S_r , is given by:

$$S_r = -\frac{a_{n-1}}{a_n}$$

Proof. Let $\{r_1, \dots, r_n\}$ be the roots of the polynomial (there will be exactly n roots in \mathbb{C} including repeated ones). Then the polynomial can be written as $P(x) = a_n(x - r_1)(x - r_2)\dots(x - r_n)$ (by the factor theorem).

Expand the factors and notice that the x^{n-1} -coefficient is: $-a_n(r_1 + r_2 + \dots + r_n)$. Let the sum of the roots be denoted by S_r . Then the x^{n-1} coefficient is $-a_n S_r$.

Equate this coefficient with the original polynomial's coefficient to get: $-a_n S_r = a_{n-1}$

This implies, by rearranging, that $S_r = -\frac{a_{n-1}}{a_n}$. ■

Theorem. For a polynomial $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, the product of the roots, P_r , is given by:

$$P_r = \frac{(-1)^n a_0}{a_n}$$

Proof. $P(x) = a_n(x - r_1)(x - r_2)\dots(x - r_n)$ (by the factor theorem).

Expand the factors and notice that the constant term is: $(-1)^n r_1 r_2 \dots r_n$. Let the product of the roots be denoted by P_r . Then the constant term is $a_n(-1)^n P_r$.

Equate this term with the original polynomial's constant term to get: $a_n(-1)^n P_r = a_0$

This implies, by rearranging, that $P_r = \frac{(-1)^n a_0}{a_n}$. ■

Worked example 3.4.1

Let $f(x) = x^3 + bx^2 + cx + d$, where $b, c, d \in \mathbb{R}$. 4 is a root of f . Given that the sum of the roots is 4 and the product of the roots is -8 , find b , c and d .

$$4 = -b \implies b = -4$$

$$-8 = -d \implies d = 8$$

$$\text{Finally, } f(4) = 0 = 4^3 - 4^3 + 4c + 8 \implies c + 2 = 0 \implies c = -2$$

All these theorems covered apply to \mathbb{C} as well. We will look at this in a later chapter.

Drill Exercises

1 Find the remainders of the following polynomial divisions:

(a) 2 when divided by $x - 4$

(d) $3x^3 + 3x^2 + 3x + 3$ when divided by $x - 5$

(b) $4x + 9$ when divided by $x + 2$

(e) $-x^4 + 2x^3 - 3x^2 + 4x$ when divided by $x + 1$

(c) $2x^2 - x - 6$ when divided by $x - 2$

(f) $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ when divided by x .

2 Factorize the following polynomials into linear factors (factors of the form $x - a$) given the additional information provided:

(a) $x^3 - 6x^2 + 11x - 6$ given that $f(1) = 0$

(c) $x^3 - 5x^2 - 8x + 48$ given that two of the roots are equal

(b) $x^3 + x^2 - 2x - 2$ given that the sum of two of the roots is 0

(d) $x^3 + 3x^2 - x - 3$ given that the roots form an arithmetic sequence with first term -3

Word Problems

3 Consider the cubic $6x^3 - 24x^2 + 30x - k$ and let the product and sum of its roots be P_r and S_r , respectively. Given that $P_r + S_r = 0$, find the value of k .

4 The polynomial $x^3 + ax^2 + bx + 12$ is divisible by $x + 3$ and $x + 2$. Find the values of a and b .

5 Let $q(x)$ be a polynomial of degree n and let S_q denote the sum of its roots. Let $r(x) = q(x+k)$ for $k \in \mathbb{R}$.

Find an expression for the sum of the roots of $r(x)$, S_r , in terms of S_q , k and n .

4 Functions

4.1 Exponents and logarithms

This section introduces the concept of exponential and logarithmic functions. Both of these types of functions prove to have significance in topics such as analysis and modeling.

An exponential function has the form $f(x) = a^x$. Instead of the variable being raised to the power of a constant like with polynomials, an exponential function has a constant raised to the variable. This leads to a much quicker increase or decrease of the function (generally) than polynomials have.

A logarithmic function has the form $g(x) = \log_a(x)$ and is the **inverse** of an exponential function. Hence, you can think about logarithmic functions as a reflection of a corresponding exponential function on the line $y = x$. We also have that $a^{\log_a(x)} = \log_a(a^x) = x$.

We will come across the constant e often. This section will introduce the usefulness of e (to some extent).

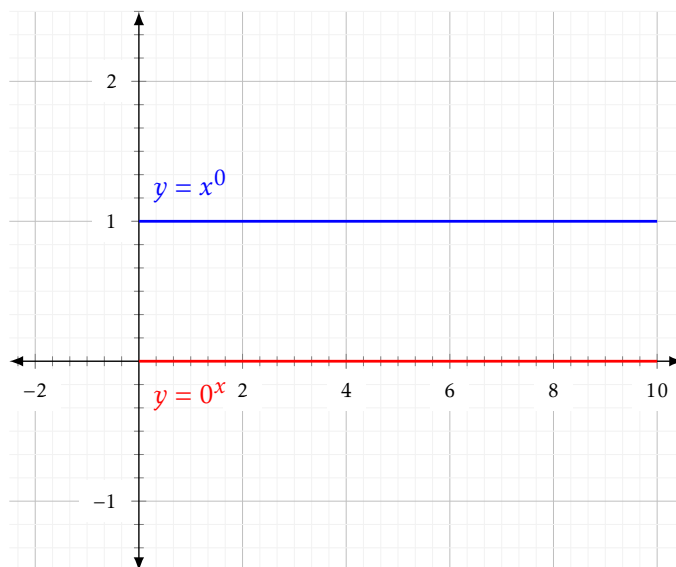
An *exponent* is the part of a number which denotes the power some number is raised by. For the expression a^n , we say that a is the *base* and n is the *exponent* and we read this number as “ a to the power of n ” or more simply “ a to the n ”. Before jumping into the analysis and use of exponential functions, it would be useful to review the laws of exponents. They are as follows:

Laws of exponents

Let $a, b, m, n \in \mathbb{R}$. Then:

- $a^m \cdot a^n = a^{m+n}$ (multiplication law)
- $\frac{a^m}{a^n} = a^{m-n}$ (division law)
- $(a^m)^n = a^{mn}$ (exponentiation law)
- $(ab)^m = a^m b^m$ (distribution law)
- $a^0 = 1$ if $a \neq 0$
- $0^m = 0$ if $m \neq 0$
- 0^0 is **undefined**

It may be useful to know why 0^0 is undefined. It's because we can obtain different values from it depending on how we approach it. Consider the two graphs below:



It can be seen that two different functions approach two distinct values of 0^0 . Hence, 0^0 is undefined (or indeterminate).

We can also use the laws of exponents to recognize that $\sqrt[n]{a} = a^{\frac{1}{n}}$. For example, we can write $\sqrt{2} = 2^{\frac{1}{2}}$. Converting roots to this form can help make exponential calculations easier to perform.

Worked example 4.1.1

Express the following values in the form a^n where $a, n \in \mathbb{Q}$

(a) $4 \cdot 4^5$

(b) $2 \cdot 8^2$

(c) $3^8 \cdot 9^3$

(d) $\sqrt{2} \cdot 4$

(a) $4 \cdot 4^5 = 4^{1+5} = 4^6$

(b) $2 \cdot 8^2 = 2 \cdot (2^3)^2 = 2 \cdot 2^{3 \cdot 2} = 2^{1+6} = 2^7$

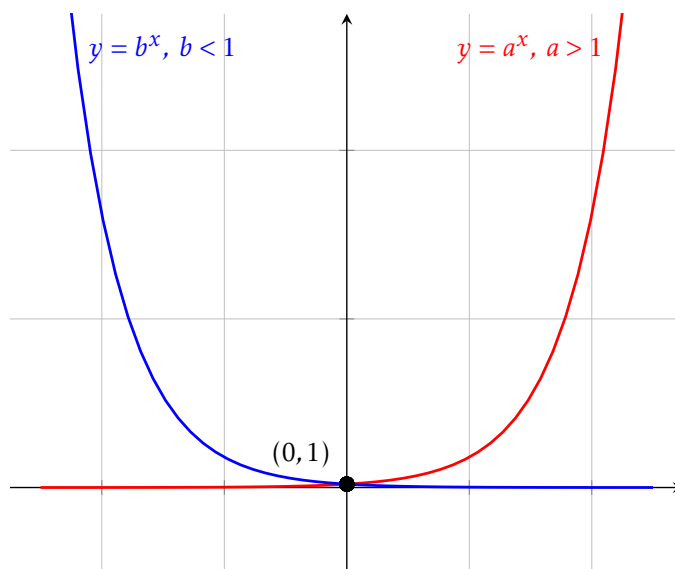
(c) $3^8 \cdot 9^3 = 3^8 \cdot (3^2)^3 = 3^{8+6} = 3^{14}$

(d) $\sqrt{2} \cdot 4 = 2^{\frac{1}{2}} \cdot 2^2 = 2^{\frac{1}{2}+2} = 2^{\frac{5}{2}}$

It may be useful to know the general shape of the graph of $f(x) = a^x$ for $a > 0$ ($a \neq 1$). Before we plot it, it may be worth determining some points:

- $f(0) = a^0 = 1$
- $f(1) = a^1 = a$
- $\lim_{x \rightarrow \infty} f(x) = \infty$ if $a > 1$
- $\lim_{x \rightarrow -\infty} f(x) = 0$ if $a > 1$
- $\lim_{x \rightarrow \infty} f(x) = 0$ if $a < 1$
- $\lim_{x \rightarrow -\infty} f(x) = \infty$ if $a < 1$

The last four points refer to the function behavior when x approaches infinity (positive or negative) depending on the value of a . If you have read the section on geometric sequences already, you should be familiar with this type of behavior. Of course if $a = 1$, $f(x) = 1$ for all values of x .



Notice that no such value x in \mathbb{R} gives $a^x \leq 0$ for any value of a in \mathbb{R} . a^x will always be positive.

It is common to use the base e for a generic exponential function (i.e. $f(x) = e^x$). e is a very special number in mathematics; probably the most important constant. As for why that is the case, later chapters will shed more light on this. However, you will be working with e^x very frequently in the course. Note that $e > 1$, so the shape of e^x can be easily deduced using the facts from above.

Further reading: Expressing e^x as a power series

While values such as 2^3 and 5^6 are intuitive to evaluate, the same can't be said for e^x . This is especially difficult since e is irrational.

We need to find a way to obtain approximations of e^x in terms of x . Luckily, there are ways of doing so. For instance, we know that $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$, so larger values of n will mean closer rational approximations to e , which we can take to the power of some x . However, this is very tedious to do by hand.

An easier way to find an approximation for e^x in terms of x is to perform a *binomial expansion* on the limit definition of e^x (which we will prove later on in the course):

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$$

Notice that for $x = 1$, we obtain $e^1 = e$, as per the definition of e . Additionally, $x = 0$ yields $e^0 = 1$, which also makes sense.

If we expand this expression using the binomial theorem:

$$\begin{aligned} e^x &= \lim_{n \rightarrow \infty} 1 + \binom{n}{1} \frac{x}{n} + \binom{n}{2} \frac{x^2}{n^2} + \binom{n}{3} \frac{x^3}{n^3} + \dots \\ &= \lim_{n \rightarrow \infty} 1 + \frac{n}{n}x + \frac{n(n-1)}{n^2} \frac{x^2}{2!} + \frac{n(n-1)(n-2)}{n^3} \frac{x^3}{3!} + \dots \\ &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ &= \sum_{k=0}^{\infty} \frac{x^k}{k!} \end{aligned}$$

We have expressed e^x as an infinite sum of powers of x . Note that this is **not** a polynomial; a polynomial must have a finite degree. Instead, this is called a **power series**. By choosing a good-enough k to stop our terms at, we can approximate e^x quite well. For larger or smaller (negative) values of x , you can see that the approximation becomes less accurate.

As stated, a logarithm function is an inverse for a corresponding exponential function. The general form is $\log_a(x)$ (read as “log to base a of x ”), where a is the base of its corresponding exponential function. It is such that if $y = a^x$, then $x = \log_a(y)$. In other words, log can be used to obtain the exponential part of an expression. That being said, $a^{\log_a(x)} = x$ and similarly $\log_a(a^x) = x$.

Just as exponential functions have algebraic laws, logarithms do as well:

Laws of logarithms

Let $a, bx, y \in \mathbb{R}^+$ where $a, b \neq 1$ and let $p \in \mathbb{R}$. Then:

- $\log_a(xy) = \log_a(x) + \log_a(y)$ (multiplication law)
- $\log_a\left(\frac{x}{y}\right) = \log_a(x) - \log_a(y)$ (division law)
- $\log_a(x^p) = p \log_a(x)$ (exponentiation law)
- $\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$ (change of base law)
- $\log_a(a) = 1$
- $\log_a(1) = 0$

These laws follow from the exponential laws and can be proven using them.

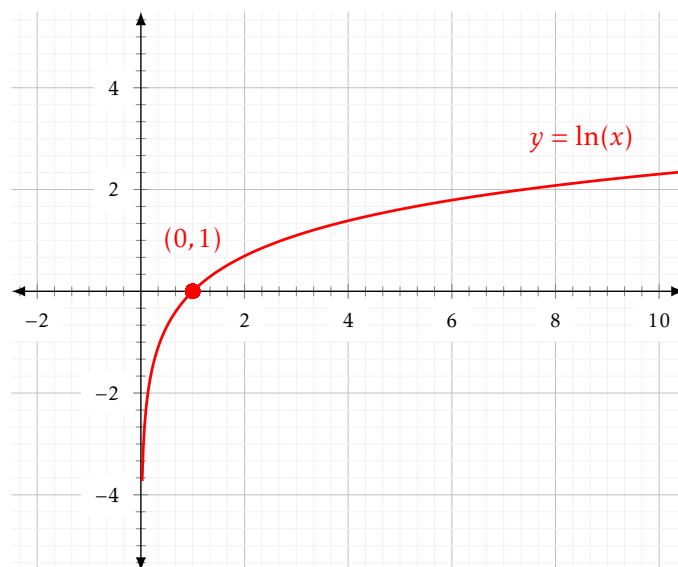
Worked example 4.1.2

Solve the equation $5^x = 125$

If $y = a^x$, then $x = \log_a(y)$. Hence, $x = \log_5(125)$. We are looking for a power of 5 which gives 125.

Intuitively (or using a calculator), this will be $x = 3$.

Since e^x is used a lot, \ln denotes logarithms to base e for brevity (i.e. $\log_e(x) = \ln(x)$). It is also common to write $\log_a(x)$ as $\frac{\ln(x)}{\ln(a)}$ (using the change of base rule) for consistency. Here is a plot of $y = \ln(x)$ on a graph:



Graphs of generic logarithms of the form $\log_a(x)$ will look similar to the above, except for when $0 < a < 1$. Remember that $y = \log_a(x)$ is just a reflection of $y = a^x$ on the line $y = x$. While it doesn't look like it, note that $\ln(x)$ approaches infinity as x does; just very slowly. This is evident from the exponential function.

Worked example 4.1.3

Solve the equation $e^{2x} - 4e^x - 1 = 0$ for $x \in \mathbb{R}$, giving your answer(s) in exact form.

Notice that $e^{2x} = (e^x)^2$ and substitute $y = e^x$ to obtain $y^2 - 4y - 1 = 0$. By substituting a common expression for a variable, we have converted the equation to a recognizable form; in this case, a quadratic. We just have to make sure our final answer is x and not y .

$$y^2 - 4y - 1 = (y - 2)^2 - 1 - 4 = 0$$

$$(y - 2)^2 = 5 \implies y - 2 = \pm\sqrt{5} \implies y = 2 \pm \sqrt{5}$$

Then back-substitute $y = e^x$:

$$e^x = 2 \pm \sqrt{5} \text{ so } e^x = 2 + \sqrt{5} \text{ or } 2 - \sqrt{5}.$$

However, notice that $2 - \sqrt{5} < 2 - \sqrt{4} = 2 - 2 = 0$ and e^x can't be negative for any real value x .

Hence, $e^x = 2 + \sqrt{5}$ is the only solution, so $x = \ln(2 + \sqrt{5})$

These types of questions can appear frequently. With a GDC, it will be easier. But by hand, it will require some intuition. You will see more of these variable substitution questions later on.

Drill Exercises

1 Solve the following equations, giving your answers in exact, simplest form:

Note: $\ln^p(x)$ means $(\ln(x))^p$

(a) $5^{x^2} \cdot 4^x = 2$

(d) $\log_{x^2}(3) = 3$ (be careful here)

(b) $3^x + 9^x = 4$

(e) $\ln(\ln^2(x)) = 0$

(c) $\ln^2(x) + \ln(x) - 1 = 0$

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

2 Find all solutions to the following equations using a GDC or otherwise. Give your answers to at least 3 s.f. where appropriate:

(a) $3^x = x^3$

(d) $\frac{e^x - e^{-x}}{2} = x^5 - x^4 + x^3 - x^2 + x + \frac{1}{10}$

(b) $e^x = x^2 + 3x + 1$

(e) $\sqrt[3]{x} = x^{ex}$

(c) $\ln(x) + e^x = 3^{\ln(x)}$

(f) $\sqrt{x^x} = x!$ (you may find Desmos particularly useful for this one)

Word Problems

3 We have seen that $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$.

Prove that $\lim_{h \rightarrow 0} \frac{x^h - 1}{h} = \ln(x)$

4.2 Properties of functions

As the title says, this section looks at properties of functions. In particular, we look at what it means to be a function, the domain, codomain and range of a function, function composition, inverse functions and function parity.

Let $f : X \rightarrow Y$. This notation says “let f be a function which maps the set X to the set Y ”. What this means is f takes any value x in the set X and *maps* it to **exactly one** value y in Y . We write $f(x) = y$ to mean f maps x to y . This tells us the first condition which a function must

satisfy: it must map each value in A to **only one** value in B . This may seem intuitive enough, but it's important to be aware of this. The reason for why this is so is because it's natural to define an equation $y = f(x)$ for graphing. This means each value of x on the x -axis is mapped to a corresponding value $y = f(x)$ on the y -axis. But such equations may not be functions.

For instance, consider the equation $y^2 = 1 - x^2$ for $(x, y) \in \mathbb{R}^2$. This equation when graphed will give a unit circle centered at the origin. If we try to solve for y , we can see that $y = \pm\sqrt{1 - x^2}$; two possible values. If we make a choice of $+$ or $-$ and write $y = f(x) = \sqrt{1 - x^2}$ or $y = f(x) = -\sqrt{1 - x^2}$, we will lose half of the graph (the function will only plot half the circle). We call an equation such as $y^2 = 1 - x^2$ a **relation**, but not a function. Given a function $f(x)$, setting $y = f(x)$ gives a relations, but not all relations give functions.

Relations

A relation on two sets X and Y is a collection of ordered pairs $(x, y) \in X \times Y$ satisfying a certain property $P(x, y)$. We say x is *related to* y if and only if $P(x, y)$ is true.

For example, we can set $X = Y = \mathbb{R}$ and define $P(x, y) : x^2 + y^2 = 1$. Then we have that x and y are related if and only if $y^2 = 1 - x^2$. It's easy to verify that the set of solutions $(x, y) \in \mathbb{R}^2$ to the equation gives the set of values satisfying the relation.

An easy way to check if a graph represents a function plot $y = f(x)$ is to verify that every vertical line passing through the plane intersects no more than 1 point on the graph.

We actually have names for the sets X and Y in relation to a function f .

Domain, codomain and range

For $f : X \rightarrow Y$ a function, we say:

- X is the **domain** of f
- Y is the **codomain** of f
- $f(X) = \{y \in Y \mid \text{there is some } x \in X \text{ such that } f(x) = y\}$ is the **range** of f

The one which stands out is the **range**. The range describes the *subset* of the codomain of values which f is able to reach with the given domain X . For example, we can define $f : \mathbb{R} \rightarrow \mathbb{R}$

by $f(x) = x^2$. It's clear that the domain and codomain of f is \mathbb{R} , but the range of f is **not** \mathbb{R} because no negative values of \mathbb{R} can be reached by f given that its domain is \mathbb{R} . We say that the range is $f(\mathbb{R}) = \mathbb{R}^+ \cup \{0\} \subseteq \mathbb{R}$ or equivalently $f(\mathbb{R}) = \{y \in \mathbb{R} \mid y \geq 0\}$. Note that the codomain of a function can be chosen to be any superset of the range of the function (or the range itself, of course). For the rest of the chapter, we will denote $\mathbb{R}_{\geq 0} = \mathbb{R}^+ \cup \{0\}$ and $\mathbb{R}_{\leq 0} = \mathbb{R}^- \cup \{0\}$.

Another point to make is a function must map **all** values in its domain to some value in the codomain. For example, we can't define $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(x) = \sqrt{x}$, because $g(x) \notin \mathbb{R}$ for any $x < 0$. We would need to write $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, then it would make sense.

Worked example 4.2.1

Write down the domain and range of the following functions:

- $f(x) = e^x$ with codomain \mathbb{R}
- $g(x) = \ln(x)$ with codomain \mathbb{R}
- $h(x) = \sqrt[3]{x}$ with codomain \mathbb{R}

- $f(x) = e^x$ can map any $x \in \mathbb{R}$ to \mathbb{R} , but $e^x > 0$ for all $x \in \mathbb{R}$. So the domain is \mathbb{R} and the range is \mathbb{R}^+ .
- $g(x) = \ln(x)$ maps to \mathbb{R} only when $x > 0$, so the domain is \mathbb{R} . It's clear that $\ln(\mathbb{R}^+) = \mathbb{R}$, since we can choose any $y \in \mathbb{R}$ and use $x = e^y \in \mathbb{R}^+$ to give $\ln(x) = y$ so the range is \mathbb{R} .
- $h(x) = \sqrt[3]{x}$ is defined on all of \mathbb{R} and maps any value in \mathbb{R} to a value in \mathbb{R} . Moreover, given any $y \in \mathbb{R}$, we can choose $x = y^3$ to give $h(x) = y$. So the domain and range are both \mathbb{R} .

The preceding exercise specified that the codomain for each provided function. This is because if the codomain was larger, the corresponding domain and range would be different. For example if we chose \mathbb{C} to be the codomain for $f(x) = e^x$ and $g(x) = \ln(x)$, the range of f and the domain of g would both be $\mathbb{C} \setminus \{0\}$ (which includes $\mathbb{R} \setminus \{0\}$). It's not necessary to make this specification given the context of functions mapping reals to reals, but one should nonetheless be aware of this fact.

On page 11, we looked at the concept of injective, surjective and bijective functions as part of a further reading section. We formalize it here:

Injective, surjective and bijective functions

Let $f : X \rightarrow Y$ be a function. We say:

- f is **injective (one-to-one)** when for every $x_1, x_2 \in X$, $f(x_1) = f(x_2) \implies x_1 = x_2$
- f is **surjective (onto)** when for every $y \in Y$, we can find some $x \in X$ for which $f(x) = y$
- f is **bijective** when it is both injective and surjective.

Here are some examples for $f : \mathbb{R} \rightarrow \mathbb{R}$:

- $f(x) = e^x$ is injective but not surjective
- $f(x) = x^3 + 2x^2$ is surjective but not injective
- $f(x) = x^3 + 2x$ is both injective and surjective
- $f(x) = x^2$ is neither injective nor surjective

It's important to know that the codomain of a function determines whether or not a function is surjective. Certainly if a function's codomain is the range itself, the function will be surjective (in fact, this is the definition of surjectivity). Also the domain of a function determines whether or not a function is injective or surjective. One can, for example, restrict $f(x) = x^2$ to the domain $[0, \infty)$ and it's clear that the function will be injective.

Another important topic is the composition of functions. This is where, given two function f and g , we may define a new function $h(x) = f(g(x))$ or $h(x) = g(f(x))$ **if permitted**.

Function composition

Define $f : W \rightarrow X$ and $g : Y \rightarrow Z$. If $f(W) \subseteq Y$, then we can define a function $h : W \rightarrow Z$ by $h(w) = g(f(w))$. We write $h = g \circ f$.

It's easy to see that if $f : X \rightarrow Y$ and $g : Y \rightarrow Z$, we can always define $h = g \circ f$ because $f(X) \subseteq Y$ (the range of f is contained in the domain of g).

Define $f : \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) = e^x$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(x) = x^2$. We have $f(\mathbb{R}) = \mathbb{R}^+$ and $g(\mathbb{R}) = \mathbb{R}_{\geq 0}$, both contained in the domain of f and g . Thus we can define $h = g \circ f : \mathbb{R} \rightarrow \mathbb{R}$ and $k = f \circ g : \mathbb{R} \rightarrow \mathbb{R}$. We have that $h(x) = e^{2x}$ and $k(x) = e^{x^2}$. They clearly aren't the same function, which shows \circ isn't necessarily a commutative operation. That is, $f \circ g = g \circ f$ doesn't always hold.

Worked example 4.2.2

Let $f(x) = \sqrt{x}$ and $g(x) = -x^2 + 4kx - 3k^2$ be real functions with $k \neq 0$. Find the domain and range of $f \circ g$ in terms of k .

We need to find all x such that $g(x) \geq 0$. Observe that $g(x) = -(x^2 - 4kx + 3k^2) = -(x - 3k)(x - k)$, so the roots of g are k and $3k$. Also, the x^2 coefficient of g is negative. Thus we have that $g(x) \geq 0$ on the interval $[k, 3k]$, which means the domain of $f \circ g$ must be $[k, 3k]$ to be defined.

Next, we can see that $g(x) = -(x^2 - 4kx + 3k^2) = -[(x - 2k)^2 + 3k^2 - 4k^2] = -(x - 2k)^2 + k^2$. Thus the vertex of g is the point $(2k, k^2)$, with $k^2 > 0$. This means $g([k, 3k]) = [0, k^2]$. Then $(f \circ g)([k, 3k]) = f([0, k^2]) = [0, |k|]$, so the range of $f \circ g$ is $[0, |k|]$.

We know that function composition isn't commutative. Define $f(x) = x^2$ and $g(x) = \sqrt{x}$. We have that $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(\mathbb{R}) = \mathbb{R}_{\geq 0}$ and $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ with $g(\mathbb{R}_{\geq 0}) = \mathbb{R}_{\geq 0}$. We can see that $f(\mathbb{R})$ is exactly the domain of g , which means the function $h : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ given by $h = g \circ f$ is well-defined. Indeed, we have $h(x) = \sqrt{x^2} = |x|$. Conversely we observe that $g(\mathbb{R}_{\geq 0})$ is contained in the domain of f , but isn't exactly the domain itself. This is fine when it comes to defining $k : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ by $k(x) = (\sqrt{x})^2 = x$. Although h and k have the same range, it's clear that they aren't the same function (they have different domains). But you've probably have noticed that there's something special with the particular functions chosen.

If we're given $x^2 \in \mathbb{R}$, we can get back x by taking the square root of x^2 , but this may yield two possible results. For example if $x^2 = 4$, we have $x = 2$ or $x = -2$ as possible values. We can, however, find x given $\sqrt{x} \in \mathbb{R}_{\geq 0}$ by squaring the root. That is if $\sqrt{x} = 4$, we have $x = 16$; and this is the only value of x for which $\sqrt{x} = 4$. This brings us to the concept of **identity** and **inverse** of functions.

Identity function

We call the function $I_X : X \rightarrow X$ defined by $I_X(x) = x$ the **identity function** on X . Moreover for any $f : X \rightarrow Y$, we have $f \circ I_X = I_Y \circ f$.

It can be seen that an identity depends on the set on which it is defined. For instance if we take $f(x) = x^2$ and $g(x) = \sqrt{x}$ with $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ and $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, it's clear that $f \circ g = I_{\mathbb{R}_{\geq 0}}$ (note that we've changed the codomain of f). We say that f is a **left inverse** of g which this happens, or equivalently that g is a **right inverse** of f .

Left and right inverses

Let $f : X \rightarrow Y$. A **left inverse** of f is a function $g : Y \rightarrow X$ satisfying $g \circ f = I_X$. A **right inverse** of f is a function $h : Y \rightarrow X$ satisfying $f \circ h = I_Y$.

We say a function is **left-invertible** when it has a left inverse and is **right-invertible** when it has a right inverse.

For the same f and g , we can see that g isn't a left inverse of f . That is, $g \circ f \neq I_{\mathbb{R}}$; it's easy to see because we've established that $(g \circ f)(x) = |x| \neq x$ on \mathbb{R} . Thus, $f(x) = x^2$ is not left-invertible. If we restrict the domain of f to $\mathbb{R}_{\geq 0}$ however, g *would be* a left-inverse, with $g \circ f = I_{\mathbb{R}_{\geq 0}}$.

Now consider $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ defined by $f(x) = e^x$ and $g(x) = \ln(x)$. It's apparent that $f \circ g : \mathbb{R}^+ \rightarrow \mathbb{R}$ is $e^{\ln(x)} = x$, which means $f \circ g = I_{\mathbb{R}^+}$. You probably have noticed that the codomain of f isn't \mathbb{R}^+ though. However, we know that $f(\mathbb{R}) = \mathbb{R}^+ \subseteq \mathbb{R}$, so we can always restrict the codomain of f to \mathbb{R}^+ if we want. So by making $f : \mathbb{R} \rightarrow \mathbb{R}^+$, we can satisfy the inverse requirement and say that g is a right inverse of f . We typically aren't picky about this, though. As long as the range of f is the domain of g , we can say that $f : \mathbb{R} \rightarrow \mathbb{R}$ has a right inverse $g : \mathbb{R}^+ \rightarrow \mathbb{R}$. We can also see that $(g \circ f)(x) = \ln(e^x) = x$, so $g \circ f = I_{\mathbb{R}}$. Therefore, g is also a left inverse of f . When a function has a left inverse **and** a right inverse, we say the function is **invertible**.

Invertible functions

A function $f : X \rightarrow Y$ is **invertible** when f has both a left inverse $g : Y \rightarrow X$ and a right inverse $h : Y \rightarrow X$. If a function is invertible, we have that $g = h$ and say that g (or h) is the **inverse** of f , denoted f^{-1} .

So for $f : \mathbb{R} \rightarrow \mathbb{R}^+$ given by $f(x) = e^x$, we have $f^{-1}(x) = \ln(x)$ is the inverse (which can be applied to either the left side or the right side of f via composition). We also have another property:

Theorem. *Let $f : X \rightarrow Y$ be a function. Then:*

- *f has a left inverse if and only if f is injective (one-to-one)*
- *f has a right inverse if and only if f is surjective (onto)*
- *f invertible if and only if f is bijective (one-to-one and onto)*

Proof. Exercise. ■

Worked example 4.2.3

Let $f : [k, \infty) \rightarrow \mathbb{R}$ be defined by $f(x) = x^2 + 3x + 4$. Find the least value of k for which f is invertible and find f^{-1} .

We need f to be such that its left inverse and right inverse exist and are (consequently) the same. If f is to have a left inverse g , we need to make sure f is one-to-one on $[k, \infty)$ (by the preceding theorem). Hence, k must at least be the x -coordinate of the vertex of f . We observe that:

$$f(x) = (x + \frac{3}{2})^2 + 4 - \frac{9}{4} = (x + \frac{3}{2})^2 + \frac{7}{4}$$

Thus $k = -\frac{3}{2}$. The range of f is $[\frac{7}{4}, \infty)$, so by treating $f : [-\frac{3}{2}, \infty) \rightarrow [\frac{7}{4}, \infty)$ (restricting the codomain), it's clear that f is surjective onto $[\frac{7}{4}, \infty)$. We can look for a function $g : [\frac{7}{4}, \infty) \rightarrow [-\frac{3}{2}, \infty)$ such that $(f \circ g)(x) = x$. Write $x = f(g(x))$ and we have:

$$x = (g(x) + \frac{3}{2})^2 + \frac{7}{4}$$

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

$$g(x) + \frac{3}{2} = \sqrt{x - \frac{7}{4}}$$

$$g(x) = -\frac{3}{2} + \sqrt{x - \frac{7}{4}}$$

where we choose the positive square root because we want $g(x) \geq -\frac{3}{2}$. We conclude that $f^{-1}(x) = -\frac{3}{2} + \sqrt{x - \frac{7}{4}}$.

We finish this section with function parity. While this is technically an HL concept, it's not a difficult concept and is worth knowing even as an SL student.

Function parity

Let f be a real-valued function. We say:

- f is **odd** when $f(-x) = -f(x)$ for all x in the domain of f
- f is **even** when $f(-x) = f(x)$ for all x in the domain of f

This gives us a notion of symmetry in a function. An example of an odd function is $f(x) = x$. This isn't a very interesting one though. Another one is $f(x) = x^3$. An example of an even function is $f(x) = x^2$ or $f(x) = |x|$. Note that not all functions have these properties. For example, $f(x) = e^x$ isn't odd nor even, neither is $f(x) = (x - 2)^2$.

Worked example 4.2.4

Show that:

- the composition of two even functions is even
- the composition of two odd functions is odd
- the composition of an odd function with an even function is even (and vice versa)

Let f and g be even functions and let h and k be odd functions. Then:

- $(f \circ g)(-x) = f(g(-x)) = f(g(x)) = (f \circ g)(x)$
- $(h \circ k)(-x) = h(k(-x)) = h(-k(x)) = -h(k(x)) = -(h \circ k)(x)$
- $(h \circ f)(-x) = h(f(-x)) = h(f(x)) = (h \circ f)(x)$
- $(f \circ h)(-x) = f(h(-x)) = f(-h(x)) = f(h(x)) = (f \circ h)(x)$

Drill Exercises

1 Coming soon

(a) ?

Word Problems

? Coming soon

5 Appendix A

Below are the proofs of some additional theorems which were left out of the main body either for the sake of brevity or because they went off-topic.

Theorem. *Let $U_1, U_2, \dots, U_n, U_{n+1}, \dots$ be a countable collection of countable sets. Then the set $U = \bigcup_{i \in \mathbb{N}} U_i$ is countable.*

Proof. Since each U_k ($k \in \mathbb{Z}^+$) is countable, there is a bijection $\phi_k : U_k \rightarrow \mathbb{N}$. Let $\psi : U \rightarrow \mathbb{N} \times \mathbb{N}$ be defined by $\psi(x) = (m_x, \phi_{m_x})$ where $m_x \in \mathbb{Z}^+$ is the least value such that $x \in U_{m_x}$. We observe that for $x, y \in U$, $\psi(x) = \psi(y) \iff (m_x, \phi_{m_x}) = (m_y, \phi_{m_y}) \iff m_x = m_y$ and $\phi_{m_x} = \phi_{m_y} \iff m_x = m_y$. Thus, ψ is injective.

Now let $\omega : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ be defined by $\omega(a, b) = 2^a(2b - 1)$. We observe that $\omega(a, b) = \omega(c, d) \iff 2^a(2b - 1) = 2^c(2d - 1)$. If $a = b$, then $2b - 1 = 2d - 1 \iff b = d$ and we're done. Otherwise suppose $a < c$. Then $2b - 1 = 2^{c-a}(2d - 1)$. The LHS is odd while the RHS is even, which is a contradiction. We thus conclude that $a = c$ is required, which proves ω is injective.

Finally, we have that $\omega \circ \psi : U \rightarrow \mathbb{N}$ is injective, which proves U is countable (either finite or countably-infinite). ■

Theorem. *Fundamental Theorem of Arithmetic:*

*For all $n \in \mathbb{N}$ with $n > 1$, n can be written **uniquely** as a product of prime numbers.*

Proof. We proceed by strong induction. The base case is $n = 2$, which is clearly true because 2 is prime. Suppose the theorem holds for all $n < k$ for some $k > 1$. If k is prime, we're done. Otherwise, suppose k is not prime. Then k can be written as a product of two integers a and b with $2 \leq a, b < k$. That is, $k = ab$. Now $a, b < k$ means the inductive hypothesis holds for a and b . So we can write a and b as a product of primes, which means $k = ab$ is also a product of primes. This means the theorem holds for k , thus proving that every integer $n > 1$ can be written as a product of primes.

For uniqueness, suppose $a = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n$ where $p_1, \dots, p_m, q_1, \dots, q_n$ are primes. Then p_1 divides a , which means it divides one of q_1, \dots, q_n as well. Since they are prime, we must have that $p_1 = q_i$ for some i . WLOG, suppose $i = 1$ so $p_1 = q_1$. Then we have that $p_2 p_3 \cdots p_m = q_2 q_3 \cdots q_n$. Repeating this process, suppose (WLOG) that $m > n$. Then we get that $p_{n+1} p_{n+2} \cdots p_m = 1$. But this isn't possible, since 1 can't be written as a product of primes. Therefore, we need $m = n$ and consequently $p_i = q_i$ for each $i \in \{1, \dots, m = n\}$. This proves uniqueness. ■

Theorem. \mathbb{Q} is countable (alternative proof using the Fundamental Theorem of Arithmetic)

Proof. We take a bijection $f : \mathbb{N} \rightarrow \mathbb{Z}$ satisfying $f(0) = 0$ (it's been shown that \mathbb{Z} is countable on page 12). We wish to find a bijection $g : \mathbb{Z} \rightarrow \mathbb{Q}$. By the Fundamental Theorem of Arithmetic, we know that every positive integer in \mathbb{Z}^+ can be written either as 1 or a product of primes $p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$ (with p_1, \dots, p_k distinct primes and non-negative powers r_1, \dots, r_k).

We can extend this to the positive rational numbers. Every $\frac{m}{n} \in \mathbb{Q}^+$ (with $m, n \in \mathbb{Z}^+$) can be written as either 1 or a quotient of a product of primes $\frac{p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}}{q_1^{s_1} q_2^{s_2} \dots q_l^{s_l}} = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k} q_1^{-s_1} q_2^{-s_2} \dots q_l^{-s_l}$. WLOG, we can assume $\gcd(m, n) = 1$ and we have that $p_i \neq q_j$ for any $i = 1, \dots, k$ and $j = 1, \dots, l$ (otherwise, we may cancel them out). Observe that this gives a unique prime factorization for each $q \in \mathbb{Q}^+$, where we include negative powers as well; and by setting all powers to zero, we would get 1.

Now define a function $h : \mathbb{Z}^+ \rightarrow \mathbb{Q}^+$ by:

$$h(m) = h(p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}) = p_1^{f(r_1)} p_2^{f(r_2)} \dots p_k^{f(r_k)}$$

This function is **injective** because if $p_1^{f(r_1)} p_2^{f(r_2)} \dots p_k^{f(r_k)} = q_1^{f(s_1)} q_2^{f(s_2)} \dots q_l^{f(s_l)}$, uniqueness of the prime factorization and injectivity of f tell us that we require $k = l$ and (WLOG) $p_i^{f(r_i)} = q_i^{f(s_i)} \implies p_i = q_i$ and $r_i = s_i$ for all $i = 1, \dots, k$. Thus, we get that $p_1^{r_1} p_2^{r_2} \dots p_k^{r_k} = q_1^{s_1} q_2^{s_2} \dots q_l^{s_l}$.

This function is also **surjective**. For each $q \in \mathbb{Q}^+$, we can write down $q = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$ for p_1, \dots, p_k primes and $r_1, \dots, r_k \in \mathbb{Z}$. Since f is surjective, we can find $n_i \in \mathbb{N}$ for which $f(n_i) = r_i$ for each $i = 1, \dots, k$. If $r_1 = r_2 = \dots = r_k = 0$, we have $q = 1$ and $n_1 = n_2 = \dots = 0$ (by how we chose f). We then see that $h(1) = 1$. Otherwise, suppose $r_1, \dots, r_k \neq 0$ (zero-powers can be ignored because their values will end up being 1 anyway). Then we can find $n_i \in \mathbb{Z}^+$ for all $i = 1, \dots, k$ such that $f(n_i) = r_i$ by surjectivity of f , and we have that:

$$q = p_1^{f(n_1)} p_2^{f(n_2)} \dots p_k^{f(n_k)}$$

Thus we can set $m = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$ to give $h(m) = q$, which proves h is indeed surjective.

We've therefore constructed a bijection $h : \mathbb{Z}^+ \rightarrow \mathbb{Q}^+$. Now write:

$$g(m) = \begin{cases} h(m), & m \in \mathbb{Z}^+ \\ -h(-m), & m \in \mathbb{Z}^- \\ 0, & m = 0 \end{cases}$$

It's clear that $g : \mathbb{Z} \rightarrow \mathbb{Q}$ is a bijection, which proves \mathbb{Q} is countable. ■

Theorem. \mathbb{R} is uncountable.

Proof. We need only prove a subset of \mathbb{R} is uncountable. Let $B \subseteq \mathbb{R}$ be the subset of \mathbb{R} defined by:

$$B = \{0.b_1b_2b_3\ldots \mid b_i \in \{0,1\} \text{ for } i \in \mathbb{Z}^+\}$$

Suppose B is **countable**. Then we can write **all** elements $\beta_1, \beta_2, \beta_3, \ldots \in B$ as:

$$\beta_1 = 0.b_{11}b_{12}b_{13}\ldots$$

$$\beta_2 = 0.b_{21}b_{22}b_{23}\ldots$$

$$\beta_3 = 0.b_{31}b_{32}b_{33}\ldots$$

$$\vdots$$

for $b_{i,j} \in \{0,1\}$. Now construct a number $\beta = 0.\overline{b_{11}}\overline{b_{22}}\overline{b_{33}}\ldots$ where $\overline{b_{ii}} = 1$ if $b_{ii} = 0$ and $\overline{b_{ii}} = 0$ if $b_{ii} = 1$. It's clear that $\beta \in B$ but apparent that $\beta \neq \beta_k$ for any $k \in \mathbb{Z}^+$ by construction (since $\overline{b_{kk}} \neq b_{kk}$). Hence, we have that $\beta_1, \beta_2, \beta_3, \ldots$ isn't the entire set B , which proves B must be uncountable.

Since $B \subseteq \mathbb{R}$, it follows that \mathbb{R} is uncountable as well. ■

Theorem. *Euclidean Division Theorem for Polynomials:*

Given two non-zero polynomials $f(x)$ and $g(x)$ with $\deg(f) \geq \deg(g)$, there are polynomials $q(x)$ and $r(x)$ with $\deg(r) < \deg(g)$ or $r(x) = 0$ such that $f(x) = g(x)q(x) + r(x)$.

Proof. Write $f(x) = \sum_{k=0}^m a_k x^k$ and $g(x) = \sum_{k=0}^n b_k x^k$ for $a_m, b_n \neq 0$ and $m \geq n$. Let $d_1(x) = \frac{a_m}{b_n} x^{m-n}$ and let $h_1(x) = f(x) - d_1(x)g(x)$. Observe that either $h_1(x) = 0$ or $\deg(h_1) < \deg(f)$, which is clear by construction. If $h_1(x) = 0$, we can set $r(x) = 0$ and $q(x) = d_1(x)$ to give $f(x) = q(x)g(x) + r(x)$, as desired.

Otherwise, we have $\deg(h_1) < \deg(f)$. If $\deg(h_1) < \deg(g)$, we can write $q(x) = d_1(x)$ and $r(x) = h_1(x)$ to give $f(x) = q(x)g(x) + r(x)$ as desired. If $\deg(h_1) \geq \deg(g)$, repeat the process with $d_2(x)$ chosen such that $h_2(x) = h_1(x) - d_2(x)g(x)$ and either $h_2(x) = 0$ or $\deg(h_2) < \deg(h_1)$.

By repeating this process further, we eventually reach a point where $\deg(h_k) < \deg(g)$ for some k and we have $h_k(x) = h_{k-1}(x) - d_k(x)g(x) = f(x) - g(x)[d_1(x) + \dots + d_k(x)]$. We can set $q(x) = d_1(x) + \dots + d_k(x)$ and $r(x) = h_k(x)$ to give $f(x) = q(x)g(x) + r(x)$ as desired. ■

6 Bibliography

- Herstein, I.N. (1964), Topics in Algebra, Ginn and Company, ISBN 0-471-02371-X
- Remmert, Reinhold (1991), "What is π ?", Numbers, Springer, p. 129
- Arndt, Jörg; Haenel, Christoph (2006). Pi Unleashed. Springer-Verlag. ISBN 978-3-540-66572-4. Retrieved 5 June 2013. English translation by Catriona and David Lischka.
- Posamentier, Alfred S.; Lehmann, Ingmar (2004). Pi: A Biography of the World's Most Mysterious Number. Prometheus Books. ISBN 978-1-59102-200-8.
- <https://upload.wikimedia.org/wikipedia/commons/c/cb/Epsilonschlauch.svg>