A Toolchest of Mathematics Notes

for Mathematical Olympiads

1 May 2024

Contents

CHAPTER 1

Sets and Logic

In this chapter, we establish the *notation* and fundamental *notions* of *sets*, and discuss the correspondence of *sets* with *logic*.

1.1 Sets and membership

A *set* is a "assemblage"* of objects. These objects are called the *elements* or *members* of the *set,* where the objects can be anything: numbers, people, other sets, etc. E.g., 12 is an element of the *set of even integers*.

If x is *an element* of A, then it is also said that x *belongs to* A, or that x *is in* A, or that A *contains* x. In this case, we write $x \in A$. The symbol $\in$ was introduced by Peano in 1889, and is derived from the Greek letter epsilon ϵ (the letter in the Greek alphabet corresponding to the Roman letter 'e' which is of course the initial letter of *element*).

We may also say A *contains* x, and write $A \ni x$. If x *is not in* A, we write: $x \notin A$.

1.2 Specifying sets

The simplest way to describe a *set* is by *enumeration,* i.e. by listing its elements explicitly between *curly braces*. Thus $\{1, 2\}$ denotes the set whose only elements are 1 and 2. Note the following two properties of sets:

- Order of elements is immaterial, e.g. $\{1, 2\} = \{2, 1\}$.
- Repetition (multiplicity) of elements is irrelevant, e.g. $\{1, 1, 2, 2\} = \{1, 2, 2, 2\} = \{1, 2\}$.

A *set* can have *no elements*. In this case we say the set is *empty* and denote the *empty set* by $\{\}$. We also use the term *null set* and the notation $\emptyset$ for the *empty set*.

The alternative way to represent a set is with *set-builder notation,* which has the form

$$\{\textit{pattern} \mid \textit{condition(s)}\}.$$

Typically, we write $\{x \mid P(x)\}$, or $\{x : P(x)\}$, to denote *the set containing all objects* x *such that the condition or property* P *holds for* x, e.g. we may write

$$\{x \mid x \text{ is a prime}\},$$

literally read as:

the set of all x *such that* x *is a prime*,

which, in this case, we could say more succinctly as: *the set of prime numbers*. Usually, we read the symbol '|' as *'such that'*. The *pattern* may also be an expression, e.g.

$$\{p^2 \mid p \text{ is a prime}\}$$

is the *set of all numbers that are the squares of prime numbers*.

*We might have called a *set* a "collection" of objects, but, a *collection* already means something else; a **collection** is a *set* whose members are themselves sets, e.g. $\{\{1\}, \{1, 2\}\}$ is a collection.

1.3 Special sets

Some sets turn up so often that we have special symbols for them in a special typeface known as *Blackboard Bold:*

$$\begin{aligned}
\mathbb{N} &= \{1, 2, 3, \dots\} && = (\text{the set of}) \textit{ Natural Numbers} \\
\mathbb{Z} &= \{\dots, -2, -1, 0, 1, 2, 3, \dots\} && = (\text{the set of}) \textit{ Integers} \\
\mathbb{Q} &= \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ with } b \neq 0 \right\} && = (\text{the set of}) \textit{ Rational Numbers} \\
\mathbb{R} & && = (\text{the set of}) \textit{ Real Numbers}
\end{aligned}$$

The last set contains all numbers that can be represented on the *number line* including the *rational numbers* and numbers such as $\sqrt{2}, \sqrt{3}, \pi, \dots$ that are termed *irrational*.

1.4 Variants of set-builder notation

Set builder notation may also be used as follows:

- $\{x \in A \mid P(x)\}$ denotes the set of all x that are already in A such that x has the property P, e.g. $\{x \in \mathbb{Z} \mid x \text{ is even}\}$ is *the set of all even integers*.
- $\{f(x) \mid x \in A\}$ denotes the set of all objects with pattern $f(x)$ such that x is in A. We saw this form above, in the definition of the *rational numbers* $\mathbb{Q}$. For a simpler example, consider: $\{2x \mid x \in \mathbb{Z}\}$ is another way of specifying *the set of all even integers*.
- $\{f(x) \mid P(x)\}$ is the most general form of set builder notation, e.g. above we saw: $\{p^2 \mid p \text{ is a prime}\}$, *the set of squared prime numbers*.

1.5 Subsets and supersets

Given two sets A and B, we say that A is a *subset* of B, and write $A \subseteq B$, if every element of A is also an element of B. Notice that in particular, B is a subset of itself. If a subset A of B is *not equal* to B, we say A is a *strict subset* of B or that A is a *proper subset* of B.

If A is a subset of B, then one can also say that B is a *superset* of A, and write: $B \supseteq A$. We also say that A *is contained in* B, or that B *contains* A.

Note that two sets A, B are equal, written $A = B$, if and only if both $A \subseteq B$ and $B \subseteq A$. Usually when trying to prove that two sets are equal, one shows each set is contained in the other.

Let $\mathbb{P} = \{p \in \mathbb{N} \mid p \text{ is prime}\}$, then we have the following sequence of inclusions, i.e. a sequence of sets for which each set is a subset of the next:

$$\mathbb{P} \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

1.6 Universal sets and complements

Often we consider all sets as being subsets of some given *universal set*. E.g., if we are investigating properties of the real numbers $\mathbb{R}$ (and subsets of $\mathbb{R}$), then we may take $\mathbb{R}$ as our *universal set*.

Given a universal set U and a subset A of U, we may define the *complement* of A (in U) as

$$A^{\mathrm{c}} = \{x \in U \mid x \notin A\}.$$

In other words, A^{c} ('A-complement') is the set of all elements of U which are not elements of A. The notations A' and $\overline{A}$ are also commonly used to represent the *complement of* A. Thus, the complement E^{c} of the set $E = \{2x \mid x \in \mathbb{Z}\}$ (the set of all even integers) in $\mathbb{Z}$, is the *set of all odd integers*, while the complement of E in $\mathbb{R}$ is the set of all real numbers that are either odd integers or not integers at all.

1.7 Unions, intersections, and relative complements

Given two sets A and B, their *union*, written $A \cup B$, is the set consisting of all objects which are elements of A *or* of B (or of both).

The *intersection* of A and B, written $A \cap B$, is the set of all objects which are both in A *and* in B.

Finally, the *relative complement* of B *relative to* A, also known as the *set(-theoretic) difference* of A and B, is the set of all objects that belong to A but not to B. It is written as $A \setminus B$.

Formally, these sets are:

$$\begin{aligned} A \cup B &= \{x \mid (x \in A) \text{ or } (x \in B)\}, \\ A \cap B &= \{x \mid (x \in A) \text{ and } (x \in B)\}, \\ A \setminus B &= \{x \in A \mid x \notin B\}. \end{aligned}$$

The *(absolute) complement* A^{c} of a set A (in a universal set U) using the *set difference* notation is $U \setminus A$.

1.8 De Morgan's Laws

The following statements, known as *de Morgan's Laws,* are true for any sets A and B. Each is easy to prove using Venn diagrams.

$$\begin{aligned} (A \cup B)^{\mathrm{c}} &= A^{\mathrm{c}} \cap B^{\mathrm{c}} \\ (A \cap B)^{\mathrm{c}} &= A^{\mathrm{c}} \cup B^{\mathrm{c}} \end{aligned}$$

Note that when we draw a Venn diagram that is supposed to represent a general situation each set should be drawn to intersect each other set (this represents a general situation, since it can still happen that any region may actually be empty). Thus, a Venn diagram with two sets A and B should be drawn as two intersecting circles in a rectangle representing the *universal set* U.

1.9 Symbols Summary

$x \in A$	means that x is an **element** of A.
$A \cup B$	is the **union** of A and B, i.e. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ (in mathematics "or" means "and/or").
$A \cap B$	is the **intersection** of A and B, i.e. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.
$A \subseteq B$	means that A is a **subset** of B, i.e. if $x \in A$ then $x \in B$.
$A \subset B$	means that A is a **proper subset** of B, i.e. $A \subseteq B$ and $A \neq B$.
$A \setminus B$	(or $A - B$) is A **take** B, i.e. the set $\{x \in A \mid x \notin B\}$.
$\{\}, \emptyset$	is the **empty set**, i.e. the set with no elements. Note that $\emptyset$ is a subset of every set.
$A^c, A', \overline{A}$	is the **complement** of A, i.e. $A' = \{x \mid x \notin A\}$. Here x belongs to some "universal set" which should be clear from the context.

1.10 The connection between Set Theory and Logic

Logic deals with *statements* that are either `true` or `false`, whereas Set Theory deals with *elements* of sets – a given *element* can either be *in* a given set or *not in* the set. Let A, B represent *sets*. Also, let p, q represent *statements* and p' represents the **negation** of p (if p is `true` then p' is `false`, and vice-versa).

A right arrow ($\rightarrow$) denotes **implies**. If $p \rightarrow q$, then, when p is `true` so is q.

A double-arrow ($\leftrightarrow$) is the corresponding symbol for equals; it denotes its operands are **logically equivalent**. If $p \leftrightarrow q$ then, when p is `true`, so is q, and when p is `false`, so is q).

A way of viewing the connection between Set Theory and Logic: is to say that the *statements* in Logic tell us which regions of a given *Universal set* are *non-empty*. With this view, one can often convert a Set Theory statement to a Logic one by sticking '$x \in$' in front of it, e.g.

$$\begin{aligned} A \cup B &\longrightarrow x \in (A \cup B) \\ &\leftrightarrow x \in A \text{ or } x \in B \end{aligned}$$

Now, $x \in A$ and $x \in B$ are examples of statements p and q, respectively.

We give a few examples of this correspondence between Set Theory and Logic:

Set Theory	Logic
$A \cup B$	p or q
$A \cap B$	p and q
$A = B$	$p \leftrightarrow q$
$A \subseteq B$	$p \rightarrow q$
$(A \cup B)' = A' \cap B'$	$(p \text{ or } q)' \leftrightarrow p'$ and q'
$(A \cap B)' = A' \cup B'$	$(p \text{ and } q)' \leftrightarrow p'$ or q'

The statement '$p \rightarrow q$' has particular importance. From the Set Theory perspective, we are saying that $A \cap B' = \emptyset$ or equivalently that the *Universal set* is $A' \cup B$, so that we have the following logical equivalence:

$$p \rightarrow q \quad \leftrightarrow \quad p' \text{ or } q.$$

In Logic it is customary to use $\vee$ (vee) for 'or', $\wedge$ (wedge) for 'and' and $\neg$ for **negation** (i.e. we write $\neg p$ for p'). So the above statement may also be written as it appears in the first exercise.

1.11 Logic with Quantifiers

The following are known as the **universal quantifier** and **existential quantifier**, respectively.

$\forall$ (which is an inverted **A**) means "for **A**ll".

$\exists$ (a back-to-front **E**) means "there **E**xists".

If $P(x)$ represents a statement that depends on x, then the following are *logical equivalences:*

$$\exists x P(x) \;\leftrightarrow\; \neg\forall x\big(\neg P(x)\big)$$
$$\forall x P(x) \;\leftrightarrow\; \neg\exists x\big(\neg P(x)\big).$$

Sometimes we write the quantifiers at the back of an expression rather than at the front. The meaning is the same, but a trailing '$\exists x$' reads better as: '*for some* x'.

Exercise Set 1.

1. Use a truth table to prove that $p \to q$ is logically equivalent to $\neg p \vee q$, i.e. show
$$(p \to q) \leftrightarrow (\neg p \vee q)$$
is a **tautology** (i.e. has value `true` for all possible inputs of p and q).

2. Use 1. to show $p \to q$ is logically equivalent to its contrapositive $\neg q \to \neg p$.

3. Write the statement,

   In every class there is at least one student who gets everything right.

   symbolically using quantifiers.
   Hint. For example, let $R(c, s, p)$ represent "Student s in class c gets problem p correct."

   Then write the *negation* of this statement as an *existential* statement (i.e. beginning with $\exists$), symbolically, and then translate that statement back into English.

4. Noting that we often add "s.t." (abbreviating "*such that*") to improve the readability of a clause following $\exists$, rewrite
$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R} \text{ s.t. } y = \cos(x)$$
so that the first quantifier in the sentence is $\exists$.

CHAPTER 2

Proof Techniques

2.1 Propositions

If P and Q are statements then

$$P \implies Q$$

means that P **implies** Q, i.e. that whenever P is true then Q must also be true. There are several other ways of saying this. Simplest is just "If P then Q", and there is "P only if Q". You can also say that P is a **sufficient condition** for Q; or that Q is a **necessary condition** for P. For example, if A is the statement "n is a prime" and B is the statement "n is a natural number" then $A \implies B$ is a **proposition**.

The sort of propositions that are important in mathematics are **theorem**s, **corollaries** (singular: **corollary**) and **lemma**s. Theorems are the most important: everybody knows about Pythagoras' Theorem. In this course, (and in other advanced mathematics) a theorem appears in two parts. The first is a statement of the theorem, with some sort of label like "Theorem 5" or "Lagrange's Theorem". The second part is the proof of the theorem which begins with the word "Proof" and ends with the symbol □. Lemmas are (usually short, easy) propositions that are used to prove theorems; corollaries are propositions that follow easily from theorems.

$\neg P$ is the negation of P, e.g. if A is the statement "n is a prime" then $\neg A$ would mean "n is not a prime."

If $P \implies Q$ and $Q \implies P$ we write '$P \iff Q$' or 'P **if and only if** Q'. We can also say P is a **necessary and sufficient condition** for Q. A shorthand way of writing "if and only if" is the odd-looking word **iff**.

Definition 2.1.1. The **converse** of '$P \implies Q$' is '$Q \implies P$'.

Note that we might have $P \implies Q$ being true but $Q \implies P$ being false. Can you think of statements P and Q for which this is so?

Definition 2.1.2. The **contrapositive** of '$P \implies Q$' is '$\neg Q \implies \neg P$'.

If $P \implies Q$ is true then its contrapositive is true, and vice versa. Thus we can say for any statements P and Q,

$$P \implies Q \qquad \text{iff} \qquad \neg Q \implies \neg P.$$

2.2 Direct Proofs

The most straightforward way of proving a theorem is by **direct proof**. To prove $P \implies Q$ we show

$$\begin{aligned} P &\implies A_1 \\ A_1 &\implies A_2 \\ A_2 &\implies A_3 \\ &\vdots \\ A_{n-1} &\implies A_n \\ A_n &\implies Q \end{aligned}$$

In writing such a proof, we generally omit the repetition of A_i on the next line, and we usually preface the initial P with 'Assume'. And, often the statement we are asked to prove is given in a convoluted way, so that it it's advisable to reorganise it by collecting *all the given conditions* as P and *what we have to deduce* as Q, and write this succinctly before we start the proof as:

RTP: $P \implies Q$

where 'RTP' stands for *Required To Prove*. And finally we like to say *we have finished the proof* with a little box: □. This is the modern fashionable way to end proofs; the old-fashioned way was to write QED which is short for "Quod erat demonstrandum," Latin for "which was to be demonstrated".

Putting all that together, our proof looks like this:

RTP: $P \implies Q$

Proof. Assume P
$\implies A_1$
$\implies A_2$
$\implies A_3$
$\vdots$
$\implies A_n$
$\implies Q$ □

Example 2.2.1.

Proposition. *The sum of two even integers is even.*

RTP: *a, b are even* $\implies$ *$a + b$ is even.*

Proof. Assume a, b are even.

$\implies a, b$ are multiples of 2
$\implies \exists m, n \in \mathbb{Z}$ such that $a = 2m$ and $b = 2n$
$\implies a + b = 2m + 2n$
$= 2(m + n)$
$\implies a + b$ is even. □

2.3 Proof by Contraposition

We said above that '$P \implies Q$' is equivalent to its contrapositive '$\neg Q \implies \neg P$'. Sometimes it's easier to prove the contrapositive proposition than the original one. Thus, to prove we can show that $\neg Q \implies \neg P$ by direct proof.

Example 2.3.1.

Proposition. *If $a \in \mathbb{Z}$ and a^2 is odd then a is odd.*

Proof. Suppose $\neg(a \text{ is odd})$.

$$\begin{aligned}
&\implies a \text{ is even} \\
&\implies \exists m \text{ such that } a = 2m \\
&\implies a^2 = 4m^2 = 2(2m^2) \\
&\implies a^2 \text{ is even} \\
&\implies \neg(a^2 \text{ is odd})
\end{aligned}$$

Thus the proposition is proved by contraposition. □

2.4 Proof by Contradiction

To prove a statement by contradiction, we show that the statement being false leads to an absurdity, leaving us to conclude that if the statement is not false, then in fact it is true.

Now a statement of form $P \implies Q$ is *vacuously true* whenever P is *false*. It is also *true* if Q is *true*.

This leaves the case when both P is *true* and Q is *false*. If we can show this combination of possibilities cannot occur, we will have shown that a statement of form: $P \implies Q$ is *true*.

Thus a proof of $P \implies Q$ generally starts by assuming P and $\neg Q$ and proceeds until a condition known to be untrue occurs, in which case we have what's termed a contradiction (which we can signal with a lightning bolt: $\lightning$), and are then able to deduce that the original statement was in fact true.

It is enormously helpful to the reader to add the phrase "*for a contradiction*" before the assumption $\neg Q$. Thus a contradiction proof is shaped like this:

Proof. Assume P and, for a contradiction, suppose $\neg Q$.

$$\begin{aligned}
&\implies \ldots \\
&\vdots \\
&\implies R \lightning \qquad [\text{where } R \text{ is something known to be untrue}]
\end{aligned}$$

Hence in fact, $P \implies Q$ (is true). □

If the reason for the contradiction is clear, then the symbol $\lightning$ is enough; often however a reason should be given in brackets after the $\lightning$ symbol.

It may happen that the statement is of form: Q. In this case, one can think of P as being true, and there is no need to write "Assume true".

The following famous result demonstrates a proof by contradiction.

Theorem 2.4.1. *The set of primes is infinite.*

Proof. For a contradiction, suppose there are only finitely many primes and label them

$$p_1 = 2 < p_2 = 3 < p_3 = 5 < \cdots < p_n.$$

Let $N = p_1 p_2 \cdots p_n + 1$. Now N is a product of primes,

$$\begin{aligned}
&\implies N \text{ is divisible by some prime in our list, say } p_i \\
&\implies \exists m \in \mathbb{Z} \text{ such that } N = p_i m \\
&\implies p_i m = p_1 p_2 \cdots p_n + 1 \\
&\implies p_i m - p_1 p_2 \cdots p_n = 1 \\
&\implies p_i(m - p_1 p_2 \cdots p_{i-1} p_{i+1} \cdots p_n) = 1 \;\lightning\; (p_i \,|\, \text{LHS, but } p_i \nmid \text{RHS})
\end{aligned}$$

Thus, in fact, the set of primes is infinite. □

2.5 Identities

An equation that it is true for any choice of the variables is called an **identity**. Some well-known identities that should be familiar to you are

$$\begin{aligned}
(a+b)^2 &= a^2 + 2ab + b^2 \\
(a-b)^2 &= a^2 - 2ab + b^2 \\
(a+b)(a-b) &= a^2 - b^2
\end{aligned}$$

To *prove* an identity, one starts with one side, e.g. the *lefthand side* (LHS), and by a sequence of steps reduces it to the other side, e.g. the *righthand side* (RHS). The proof should have this shape:

$$\begin{aligned}
\text{LHS} &= \cdots \\
&\vdots \\
&= \text{RHS}
\end{aligned}$$

As an example, we prove the first of the identities above:

$$\begin{aligned}
\text{LHS} &= (a+b)^2 \\
&= (a+b)(a+b) \\
&= a(a+b) + b(a+b) \\
&= a^2 + ab + ba + b^2 \\
&= a^2 + 2ab + b^2, && \text{since } ab = ba \\
&= \text{RHS}
\end{aligned}$$

2.6 Proof by Mathematical Induction

In this case we want to prove an infinite set of statements $P(1), P(2), P(3), \ldots$, or equivalently we want to prove,

$$\forall n \in \mathbb{N},\ P(n) \text{ is true.}$$

We prove this in two steps:

(i) We show that $P(1)$ is true.

(ii) We show that $\forall k \in \mathbb{N},\ P(k) \implies P(k+1)$.

Example 2.6.1.

Proposition. *If $n \in \mathbb{N}$ then*

$$P(n):\ 1+3+5+\cdots+(2n-1) = n^2 \tag{2.6.1}$$

Proof. (i) If $n = 1$ then the lefthand side of (2.6.1) equals 1, and the righthand side is $1^2 = 1$. So the proposition is OK for $n = 1$, i.e. $P(1)$ is true.
(ii) Suppose (2.6.1) is OK for k, i.e. suppose that $P(k)$ is true. Then

$$\begin{aligned} \text{LHS of } P(k+1) &= 1+3+\cdots+(2k-1)+(2(k+1)-1) \\ &= \text{LHS of } P(k) + 2k + 1 \\ &= k^2 + 2k + 1 \\ &= (k+1)^2 = \text{RHS of } P(k+1) \end{aligned}$$

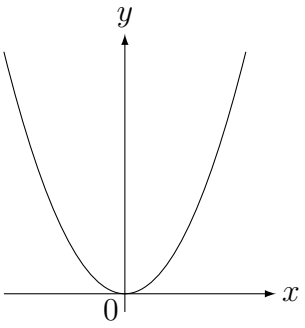
Thus $P(k+1)$ follows from $P(k)$.
Now from (i) and (ii) together, we have

$$1+3+5+\cdots+(2n-1) = n^2\ \forall n \in \mathbb{N}. \qquad \square$$

Note 2.6.2. Above we just introduced the idea of *Mathematical Induction*. In general, when we are asked to prove something by induction, the $P(n)$ won't be given. A good practice is to define the statement $P(n)$ at the beginning of your proof. Then step (i) is usually called the *base case* and step (ii) is called the *inductive step*. Finally, we usually have a *conclusion*, essentially saying that from (i) and (ii), by the *Principle of Mathematical Induction* we have the desired conclusion. For readability, it's nice to have the headings, *base case*, *inductive step*, and *conclusion* in your proof. You will see these embellishments in the next chapter.

2.7 Another example of a Direct Proof

Below we will need the following useful idea. Recall that the graph of $y = x^2$ looks like:



The important observation to make is that for any real number x, x^2 is never negative, and is only zero when x itself is zero, i.e.

$$x^2 \geqslant 0 \text{ for all real } x, \text{ and}$$
$$\text{if } x^2 = 0 \text{ then } x = 0.$$

This innocuous statement is *incredibly useful.* Keep it in your box of tricks!! We use it in the third line of the proof below.

Exercise 2.7.1. Prove that: If $a, b \geqslant 0$ then $\sqrt{ab} \leqslant \dfrac{a+b}{2}$.

Proof. Assume $a, b \geqslant 0$.

$$\begin{aligned}
&\implies \sqrt{a}, \sqrt{b} \text{ exist, i.e. are real numbers} \\
&\implies (\sqrt{a} - \sqrt{b})^2 \geqslant 0 \\
&\implies a - 2\sqrt{a}.\sqrt{b} + b \geqslant 0 \\
&\implies a + b \geqslant 2\sqrt{ab} \\
&\implies \frac{a+b}{2} \geqslant \sqrt{ab} \\
&\implies \sqrt{ab} \leqslant \frac{a+b}{2}.
\end{aligned}$$

□

2.8 More examples of Proofs by Contradiction

We gave a scheme for setting out a *proof by contradiction* of a statement of form '$P \implies Q$' earlier. The exercise below doesn't have a P part. One can think of the P as being *true* and just omit it; otherwise the structure is as it was given.

Exercise 2.8.1. Prove that: $\sqrt{2}$ is irrational.

Proof. For a contradiction assume that $\sqrt{2}$ is rational.

$$\begin{aligned}
&\implies \sqrt{2} = \frac{p}{q} \text{ for some } p, q \in \mathbb{Z} \text{ with } \gcd(p,q) = 1, \text{ i.e. } p/q \text{ is reduced to "lowest terms"} \\
&\implies 2 = \frac{p^2}{q^2} \\
&\implies 2q^2 = p^2 \\
&\implies 2 \,|\, p^2, \text{ since } 2 \,|\, 2q^2 \\
&\implies 2 \,|\, p \\
&\implies p = 2r \text{ for some } r \in \mathbb{Z} \\
&\implies 2q^2 = (2r)^2 = 4r^2 \\
&\implies q^2 = 2r^2 \\
&\implies 2 \,|\, q^2, \text{ since } 2 \,|\, 2r^2 \\
&\implies 2 \,|\, q \\
&\implies \gcd(p,q) \geqslant 2 \; \lightning (\gcd(p,q) = 1), \text{ since now } 2 \,|\, p \text{ and } 2 \,|\, q
\end{aligned}$$

$\therefore$ in fact, $\sqrt{2}$ is not rational.
$\therefore$ $\sqrt{2}$ is irrational. □

In many of the examples up to now we have not rewritten the problem with an RTP statement. When the "If part" and "then part" are clear, it's unnecessary. However, a problem where this device is particularly important is Exercise 7 (in the exercise set at the end of the chapter) which could be set out in the following fashion.

RTP: $x \in \mathbb{Q}, y \notin \mathbb{Q} \implies x + y \notin \mathbb{Q}$

Proof. Assume $x \in \mathbb{Q}, y \notin \mathbb{Q}$ and, for a contradiction, assume $x + y \in \mathbb{Q}$.

$$\begin{aligned} &\implies \dots \\ &\implies \dots \\ &\vdots \\ &\implies y \in \mathbb{Q} \lightning \\ &\therefore \text{ in fact, } x + y \notin \mathbb{Q} \\ &\therefore x \in \mathbb{Q}, y \notin \mathbb{Q} \implies x + y \notin \mathbb{Q} \end{aligned}$$

□

2.9 Relations

We now define various types of **binary relations**.

Definition 2.9.1. An **equivalence relation** is a binary relation $\sim$ on a set S that satisfies $\forall x, y, z \in S$:

E1: $x \sim x$, (**reflexivity**)

E2: $x \sim y \implies y \sim x$, (**symmetry**)

E3: $x \sim y$ and $y \sim z \implies x \sim z$. (**transitivity**)

Remark 2.9.2. An *equivalence relation* is essentially something that behaves like "=". Other examples are $\cong$ (*congruence*) on the set of triangles, $\sim$ (*similarity*) on the set of triangles, $\iff$ (*if and only if*) on the set of true-false statements.

The main reason we have includes the concept of an *equivalence relation* here, is that it helps explain a desirable way for setting out the proof of an **identity**. An *identity* is a statement of the form

$$\text{LHS} = \text{RHS},$$

that's generally true. The idea is that if we have a sequence of statements,

$$\text{LHS} = a,\ a = b,\ b = c, \dots,\ z = \text{RHS}, \tag{2.9.1}$$

then we can deduce LHS = RHS, by repeated application of the *transitivity* property, and we set out the proof, in the following way:

$$\begin{aligned} \text{LHS} &= a \\ &= b \\ &= c \\ &\vdots \\ &= z \\ &= \text{RHS}. \end{aligned}$$

Moreover, such a proof is just as valid, by the *symmetry* property, if we instead start with RHS and finish with the LHS.

Definition 2.9.3. A **partial order** is a binary relation $\preceq$ on a set S that satisfies $\forall x, y, z \in S$:

PO1: $x \preceq x$, (**reflexivity**)

PO2: $x \preceq y$ and $y \preceq x \implies x = y$, (**antisymmetry**)

PO3: $x \preceq y$ and $y \preceq z \implies x \preceq z$. (**transitivity**)

Remark 2.9.4. A *partial order* is essentially something that behaves like "$\leqslant$". Other examples are $\geqslant$, $\subseteq$, $\supseteq$.

Definition 2.9.5. A **strict partial order** is a binary relation $\prec$ on a set S such that $\forall x, y, z \in S$:

SPO1: $x \not\prec x$, (**irreflexivity**)

SPO2: $x \prec y \implies y \not\prec x$, (**asymmetry**)

SPO3: $x \prec y$ and $y \prec z \implies x \prec z$. (**transitivity**)

Remark 2.9.6. A *strict partial order* is essentially something that behaves like "$<$". Other examples are $>, \subset, \supset$.

This is not the end of the story; $<$ and $>$ are also **total order** (also called **full order**) relations, where SPO2 can be replaced by:

FO2: $x \prec y$ or $y \prec x$ or $x = y$, (**trichotomy**)

The relations: $\subset, \supset$ don't satisfy this stronger axiom, since two sets can be *incomparable* (none of the three given possibilities).

The reason for mentioning all three of the relations: *equivalence relation*, *partial order* and *strict partial order* here, is that they all possess the property of *transitivity*, so that if the relations of some of the equations in (2.9.1) are replaced by "$\leqslant$" and "$<$" then we can deduce LHS $\sim$ RHS where $\sim$ is the strictest of "$=$", "$\leqslant$" and "$<$" that appears in the sequence (where "$\leqslant$" is *stricter* than "$=$", and "$<$" is *stricter* than "$\leqslant$"). Thus,for example, from

$$\begin{aligned} \text{LHS} &= a \\ &\leqslant b \\ &< c \\ &= z \\ &\leqslant \text{RHS}, \end{aligned}$$

we can deduce LHS $<$ RHS.

Similarly, if we have a sequence of such statements where the relations are "$=$", "$\geqslant$" and "$>$" (which are again in order of increasing *strictness*), then we can deduce LHS $\sim$ RHS where $\sim$ is the strictest of "$=$", "$\geqslant$" and "$>$" that appears. It's important that the relations that occur in such a sequence have the same *directionality*.

Formalising above, we say that $(=, \leqslant, <)$ is a **supertransitive** triple, meaning that if we have an alternating sequence of expressions a_i and relations rel_i (from the triple):

$$a_1 \ \text{rel}_1 \ a_2 \ \text{rel}_2 \ a_3 \ \text{rel}_3 \cdots \text{rel}_{n-1} \ a_n \ \text{rel}_n \ a_{n+1}$$

then the relationship between a_1 and a_{n+1} is the strictest of the relations $\text{rel}_1, \ldots, \text{rel}_n$.

Similarly, $(=, \geqslant, >)$, $(=, \subseteq, \subset)$ and $(=, \supseteq, \supset)$ are **supertransitive** triples.

2.10 Application of functions to inequalities

We discuss inequalities more generally in Chapter 11 and functions in Section 14.6. In particular, we define **domain** in Section 14.6. For now, when we mention the word *domain* we mean an interval of the real line for which the statement (an *inequality*) is *valid*, i.e. a *set* of values x for which the *inequality* holds.

Here we are interested in the answer to the following question:

When can one apply a function to both sides of an inequality?

The key idea is embedded in the following definitions:

Definition 2.10.1. A function f is **increasing** if, for x, y in the *domain* of f,

$$x < y \implies f(x) < f(y).$$

From this definition, we see *increasing functions* **preserve** the *directionality* of an inequality.

Definition 2.10.2. A function f is **decreasing** if, for x, y in the *domain* of f,

$$x < y \implies f(x) > f(y).$$

From this definition, we see that *decreasing functions* **reverse** the *directionality* of an inequality.

Thus the answer to the question is:

> The properties a function f must have, in order that it can be applied to both sides of an inequality (so that the statement remains true), are summed up as follows:
>
> - If a function f is *increasing* on the *domain of the inequality*, then its application to both sides of the inequality *preserves* the directionality of the inequality.
> - If a function f is *decreasing* on the *domain of the inequality*, then its application to both sides of the inequality *reverses* the directionality of the inequality.
> - Otherwise, application of the function will result in nonsense. Don't do it!

Since we have introduced above some ideas that can refine proofs, it makes sense to give some further examples that demonstrate their usage.

Exercise 2.10.3. Prove that if $a, b > 0$ such that $ab \geqslant a + b$ then $a + b \geqslant 4$.

RTP: $a, b > 0, ab \geqslant a + b \implies a + b \geqslant 4$

Proof. Assume $a, b > 0$ (1)

$ab \geqslant a + b$ (2)

$\implies a + b \geqslant 2\sqrt{ab}$, by Exercise 2.7.1

$\implies (a + b)^2 \geqslant 4ab$, since $u \mapsto u^2$ is increasing for $u \geqslant 0$

$\geqslant 4(a + b)$, by (2)

$\implies a + b \geqslant 4$, since $a, b > 0 \implies a + b > 0$ □

Finally, where symmetry exists and introducing an extra condition makes no essential difference to the proof, we can introduce that condition with the phase **without loss of generality** (abbreviated as: **w.l.o.g.**).

Exercise 2.10.4. Prove that if $0 < a, b \leqslant 1$ then $a/(b+1) + b/(a+1) \leqslant 1$.

RTP: $0 < a, b \leqslant 1 \implies \dfrac{a}{b+1} + \dfrac{b}{a+1} \leqslant 1.$

Proof. Assume $0 < a, b \leqslant 1$ and w.l.o.g. assume $a \geqslant b$. Then

$$\begin{aligned} \frac{a}{b+1} + \frac{b}{a+1} &\leqslant \frac{a}{b+1} + \frac{b}{b+1} \\ &\leqslant \frac{1}{b+1} + \frac{b}{b+1} \\ &= \frac{b+1}{b+1} \\ &= 1. \end{aligned}$$

□

Exercise Set 2.

1. Let n be a positive integer and d its smallest divisor greater than 1. Prove that d is prime.
2. Prove that if x, y are positive numbers then $x + y \geqslant 2\sqrt{xy}$.
3. A right-angled triangle has sides of length a, b, and c with c being the hypotenuse. If the triangle has area $c^2/4$, show that it is isosceles.
4. Prove that if p is a prime and q is a positive integer less than p, then p and q are relatively prime.
5. Prove that if a, b and c are real numbers such that $a^2 + b^2 + c^2 = 0$, then $a = b = c = 0$.
6. Prove that the numbers $n - 1$ and $n + 15$ are relatively prime for every even integer n.
7. Prove that the sum of two real numbers one of which is rational and the other irrational is an irrational number.
8. Prove that if m is an integer and $2^m - 1$ is a prime, then m is also a prime.
9. Prove that $\sqrt{2}$ is irrational, by the following approach.
   (i) Assume for a contradiction $\sqrt{2} \in \mathbb{Q}$.
   (ii) Deduce that there exist $a, b \in \mathbb{Z}$ with $b \neq 0$ such that $(a/b)^2 = 2$.
   (iii) Show that w.l.o.g. we may assume $a, b > 0$, and b as small as possible.
   (iv) Show that $\left(\dfrac{2b-a}{a-b}\right)^2 = 2$.
   (v) Show that $2b - a, a - b > 0$ with $a - b < b$, and thereby deduce a contradiction.

CHAPTER 3

Mathematical Induction

3.1 Revisiting induction

We already visited *induction* as a proof technique in the last chapter. We describe it again; this time, in terms of a ladder. How might we prove we can climb a ladder?

One way to show we can climb a ladder is to show we can do two things:

(i) We can get on the *first* rung, and

(ii) we can get from any rung to the next rung.

Knowing we can do these two things, we can deduce that we can get to any rung of the ladder. This is essentially the *inductive* way to prove we can climb a ladder.

How might we express the above mathematically? Firstly, let $P(n)$ be the statement:

"we can get onto the n^{th} rung."

Now being able to climb the ladder is the same as saying

"we can get onto the n^{th} rung, for any n",

or, in terms of $P(n)$, that:

$P(n)$ is true, for any n.

Now, (i) above is: $P(1)$ is true.

And, another way of saying (ii) is that:

"***if*** we can get to the k^{th} rung, ***then*** we can get to the $(k+1)^{\text{st}}$ rung",

which in terms of $P(n)$ is:

$$P(k) \implies P(k+1).$$

Lastly, mathematicians' ladders tend to have an *infinite* number of rungs. So *"for any n"* becomes: *"for all $n \in \mathbb{N}$"*.

Expressing the above formally, we have:

Definition 3.1.1. The **Principle of Mathematical Induction**, states:

To prove the statement $P(n)$ is true for all $n \in \mathbb{N}$, *it is sufficient* to prove:

(i) $P(1)$ is true, and

(ii) $P(k) \implies P(k+1)$, for general k.

Note 3.1.2. In fact, what we have described above is **simple induction**. Revisiting the ladder concept, we can imagine beasts with various afflictions that need to be catered for, and customise other forms of induction to show those beasts can still climb the ladder.

One way induction arises naturally is in proving a guessed relationship is true, e.g. from

$$\begin{aligned} 1 &= 1 \\ 1+2 &= 3 \\ 1+2+3 &= 6 \\ &\vdots \end{aligned}$$

we might guess that: $1+2+\cdots+n = \frac{1}{2}n(n+1)$, and then proceed to prove it by induction; this is done in the example below. We'll do the proof twice; the 1^{st} proof is accompanied by much explanation. The 2^{nd} proof omits superfluous explanation and follows protocols in Note 3.1.3.

Note 3.1.3. When we introduced *Mathematical Induction*, last chapter, we explained that, for readability, it's nice to begin an induction proof by defining a $P(n)$. Usually, you will be just asked to prove some statement in terms of an integer parameter n; introducing $P(n)$, effectively gives that statement a short name $P(n)$.

Then, it's tidy to have headings, *base case* (for step (i)), *inductive step* (for step (ii)), and a *conclusion* that states: "by **PMI**, from (i) and (ii) we deduce $P(n)$ is true for all $n \in \mathbb{N}$", where *PMI* abbreviates *Principle of Mathematical Induction*).

Also, it's a good practice to label the *assume* statement in the *inductive step*, **(I.A.)** (for **Inductive Assumption**), and to have *"by (I.A.)"* to the right of the step *(I.A.)* is applied. Finally, *"holds"* is a little nicer than *"is true"*.

Example 3.1.4. *Prove* $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$ *for all* $n \in \mathbb{N}$.

Proof. First we let $P(n) : 1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$.


Notice, a ':' was used here to indicate that $P(n)$ is short-hand for *everything* that follows the ':'. Use of a symbol like '=' instead of ':' would have been too confusing!

(i) Show $P(1)$ is true:

Proof. $P(1)$ is of the form LHS = RHS. To show it is true we start with one side and *reduce* it to the other side. Now the *LHS of P(1)* is just 1 and the *RHS of P(1)* is $\frac{1}{2} \cdot 1(1+1)$, i.e.

$$\begin{aligned} \text{LHS of } P(1) &= 1 \\ &= \tfrac{1}{2} \cdot 1(1+1) \\ &= \text{RHS of } P(1) \end{aligned}$$

So $P(1)$ is true. □

(ii) Show, for a general natural number k: if $P(k)$ is true then $P(k+1)$ is also true;

Proof. To prove a statement of form:

If *hypothesis* ***then*** *conclusion*

by Direct Proof (the most usual way, for an induction), we ***assume*** the *hypothesis* and deduce from it, the *conclusion*. Hence, we assume $P(k)$ is true, i.e. we assume

$$\text{LHS of } P(k) = \text{RHS of } P(k).$$

Now we wish to deduce that $P(k+1)$ is true, where $P(k+1)$ is of the form LHS = RHS. So to show $P(k+1)$ is true we start with one side and *reduce* it to the other side. (Somewhere along the way we expect to use our *assumption* that $P(k)$ is true – incidentally, this assumption is called the ***inductive assumption***). Thus, starting with one side ...

$$\begin{aligned} \text{LHS of } P(k+1) &= 1 + 2 + \cdots + k + k + 1 \\ &= \big(\text{LHS of } P(k)\big) + k + 1 \\ &= (\text{RHS of } P(k)) + k + 1, \quad \text{(using the inductive assumption)} \\ &= \tfrac{1}{2}k(k+1) + k + 1 \\ &= \tfrac{1}{2}(k+1)\big(k+2\big) \\ &= \tfrac{1}{2}(k+1)\big((k+1)+1\big) \\ &= \text{RHS of } P(k+1) \end{aligned}$$

So, *if* $P(k)$ *is true* then $P(k+1)$ is true. □

Thus we may now deduce from (i) and (ii), that, by *PMI*, $P(n)$ is true for all $n \in \mathbb{N}$. □

Now let's do the whole proof again, following the protocols of Note 3.1.3.

Proof. Let $P(n) : 1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$.

(base case) We prove $P(1)$.

$$\begin{aligned} \text{LHS of } P(1) &= 1 \\ &= \tfrac{1}{2} \cdot 1(1+1) \\ &= \text{RHS of } P(1) \\ \therefore\ P(1) &\text{ holds.} \end{aligned}$$

(inductive step) We prove: $P(k) \implies P(k+1)$ for general k.
Assume $P(k)$ holds, i.e. $1 + 2 + \cdots + k = \frac{1}{2}k(k+1)$ (I.A.)
Consider $P(k+1)$:

$$\begin{aligned} \text{LHS of } P(k+1) &= 1 + 2 + \cdots + k + k + 1 \\ &= \tfrac{1}{2}k(k+1) + k + 1, \quad \text{by (I.A.)} \\ &= \tfrac{1}{2}(k+1)\big(k+2\big) \\ &= \tfrac{1}{2}(k+1)\big((k+1)+1\big) \\ &= \text{RHS of } P(k+1) \\ \therefore\ P(k+1) &\text{ holds, if } P(k) \text{ holds.} \end{aligned}$$

(conclusion) So, we have proved:

$$\begin{aligned} &P(1) \text{ holds, and} \\ &P(k) \implies P(k+1), \text{for general } k, \end{aligned}$$

and hence, by PMI, $P(n)$ holds for all $n \in \mathbb{N}$, i.e.
$1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$ for all $n \in \mathbb{N}$. □

Example 3.1.5. *Prove that:*

If $x + \dfrac{1}{x}$ *is an integer then* $x^n + \dfrac{1}{x^n}$ *is an integer for all positive integers* n.

You will notice differences between the structures of our proof below and that of our elementary example above, but you will notice also great similarities. One of the differences is that the ***inductive step*** *of the proof requires* two *previous "rungs", which means that the* ***base case*** *of the proof must be replaced with a proof that one* "can get onto the first two rungs" – *think of the* ladder. *This is one form of* ***secondary induction***.

Proof. Assume $x + \dfrac{1}{x} = N \in \mathbb{Z} \qquad (*)$

We show that $x^n + \dfrac{1}{x^n} \in \mathbb{Z}$ for all $n \in \mathbb{N}$, by induction.

Let $f(n) = x^n + \dfrac{1}{x^n}$.

Let $P(n) : f(n) \in \mathbb{Z}$.

(base case) We prove $P(1)$ and $P(2)$.

$$\begin{aligned} f(1) &= x^1 + \frac{1}{x^1} \\ &= x + \frac{1}{x} \\ &\in \mathbb{Z}, && \text{by } (*) \end{aligned}$$

$\therefore$ $P(1)$ holds.

$$\begin{aligned} f(2) &= x^2 + \frac{1}{x^2} \\ &= \left(x + \frac{1}{x}\right)^2 - 2x \cdot \frac{1}{x} \\ &= N^2 - 2 \\ &\in \mathbb{Z}, && \text{since by } (*),\ N \in \mathbb{Z} \end{aligned}$$

$\therefore$ $P(2)$ holds.

(inductive step) We prove: $P(k-1)$ and $P(k) \implies P(k+1)$ for $k \geqslant 2$.
Assume $P(k-1)$ and $P(k)$ hold, i.e. $f(k-1), f(k) \in \mathbb{Z}$ (I.A.)
Consider $P(k+1)$:

$$\begin{aligned} f(k+1) &= x^{k+1} + \frac{1}{x^{k+1}} \\ &= \left(x^k + \frac{1}{x^k}\right)\left(x + \frac{1}{x}\right) - x^k \cdot \frac{1}{x} - \frac{1}{x^k} \cdot x \\ &= \left(x^k + \frac{1}{x^k}\right)\left(x + \frac{1}{x}\right) - \left(x^{k-1} + \frac{1}{x^{k-1}}\right) \\ &= f(k) \cdot N + f(k-1), \quad \text{by } (*) \\ &\in \mathbb{Z}, && \text{since } f(k-1), f(k) \in \mathbb{Z} \text{ by (I.A.)} \end{aligned}$$

$\therefore$ $P(k+1)$ holds, if $P(k-1), P(k)$ hold.

(conclusion) Hence, if $x + 1/x \in \mathbb{Z}$, we have proved:

$$\begin{aligned} &P(1) \text{ and } P(2) \text{ hold, and} \\ &P(k-1) \text{ and } P(k) \implies P(k+1), \text{for } k \geqslant 2, \end{aligned}$$

so that, by PMI, $P(n)$ holds for all $n \in \mathbb{N}$, i.e.
if $x + 1/x \in \mathbb{Z}$ then $x^n + 1/x^n \in \mathbb{Z}$ for all $n \in \mathbb{N}$. □

Remark 3.1.6. In the example above, we defined $f(n)$. We did this since, with the relational operator of $P(n)$ being "$\in$", "LHS of P(n)" is somewhat unnatural, and $f(n)$ is, in any case, much shorter.

3.2 Variants of induction

We have already seen two variants of induction, in the two examples of this chapter. The most common variant, **simple induction** was what was actually given in Definition 3.1.1.

As suggested in Note 3.1.2, we are now going to have a little fun, revisiting the ladder concept, and imagining beasts with various afflictions that need to be catered for, to customise other forms of induction, and thereby show those beasts can still climb the ladder. For each variant, we give a bit of a fun "back story" to explain any modifications to the *base case*, *inductive step*, and the *conclusion*:

Simple induction. This is just the usual induction, with:
(base case) Show $P(1)$ holds.
(inductive step) Show $P(k) \implies P(k+1)$, for general k.
(conclusion) $P(n)$ holds for all $n \in \mathbb{N}$.

Simple induction – Aladdin's version.
Here, the first rungs of the ladder are broken, but Aladdin can use a wish to get to rung n_0:
(base case) Show $P(n_0)$ holds.
(inductive step) Show $P(k) \implies P(k+1)$, for general $k \geqslant n_0$.
(conclusion) $P(n)$ holds for all $n \in \mathbb{N}$, such that $n \geqslant n_0$.

Secondary induction – giant version.
The giant can't bend his knees and can only do two rungs at a time.
If he can only start by getting to the first rung, then he is only able to get to odd rungs.
So he must be able to start by getting to both rungs 1 and 2:
(base case) Show $P(1)$ and $P(2)$ hold.
(inductive step) Show $P(k) \implies P(k+2)$, for general k.
(conclusion) $P(n)$ holds for all $n \in \mathbb{N}$.

Secondary induction – Jake the Peg version.
Jake the Peg is a 3-legged man. (A variant of this one was used in Example 3.1.5.)
To climb the ladder Jake needs to have 2 feet on the 2 previous rungs,
to push up with his 3^{rd} leg to get on the next rung.
To start he needs to be able to somehow get on rungs 1 and 2:
(base case) Show $P(1)$ and $P(2)$ hold.
(inductive step) Show $P(k), P(k+1) \implies P(k+2)$, for general k.
(conclusion) $P(n)$ holds for all $n \in \mathbb{N}$.

Complete* induction – polypus version.
The polypus is a strange animal. As it climbs new legs spring into existence.
To climb the ladder it needs to have k feet on the k previous rungs
to push up with its newly sprung $(k+1)^{\text{st}}$ leg to get on the next rung.
At the start it only has 2 legs. So to start, it only needs to be able to get on the first rung:
(base case) Show $P(1)$ holds.
(inductive step) Show $P(1), P(2), \ldots, P(k) \implies P(k+1)$, for general k.
(conclusion) $P(n)$ holds for all $n \in \mathbb{N}$.

* *Complete induction* is also called **strong induction**, though *strong induction* often refers to any induction "stronger" than *simple induction*.

Exercise Set 3.

1. Prove for any natural number n that
   (a) $1+3+5+\cdots+2n-1=n^2$;
   (b) $1^2+2^2+3^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1)$;
   (c) $1^3+2^3+3^3+\cdots+n^3=\frac{1}{4}n^2(n+1)^2$;
   (d) $1^2+4^2+7^2+\cdots+(3n-2)^2=\frac{1}{2}n(6n^2-3n-1)$;
   (e) $2^2+5^2+8^2+\cdots+(3n-1)^2=\frac{1}{2}n(6n^2+3n-1)$.

2. Prove that for any natural number n,
$$2(\sqrt{n+1}-1)<1+\frac{1}{\sqrt{2}}+\cdots+\frac{1}{\sqrt{n}}<2\sqrt{n}.$$

3. Prove $3^n>2^n$ for all natural numbers n.

4. Prove *Bernoulli's Inequality* which states:
   If $x\geqslant -1$ then $(1+x)^n\geqslant 1+nx$ for all natural numbers n.

5. Prove that for any natural number $n\geqslant 2$,
$$\big(1-\frac{1}{\sqrt{2}}\big)\big(1-\frac{1}{\sqrt{3}}\big)\cdots\big(1-\frac{1}{\sqrt{n}}\big)<\frac{2}{n^2}.$$

6. Prove that for any natural number n,
$$\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdots\frac{2n-1}{2n}\leqslant\frac{1}{\sqrt{3n+1}}.$$

7. Prove that $7^{2n}-48n-1$ is divisible by 2304 for every natural number n.

8. For every natural number n, show that
$$u_n=\frac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n.\sqrt{5}}$$
   is a natural number.
   In fact, u_n is the nth Fibonacci number.

9. Prove that for all $n\in\mathbb{N}$ and $x\in\mathbb{Z}\setminus\{1\}$, x^n-1 is divisible by $x-1$.

10. Given that $a_1=-2$, $a_2=-16$ and $a_{n+2}=8a_{n+1}-15a_n$, prove $a_n=3^n-5^n$ for all $n\in\mathbb{N}$.

11. Consider all possible subsets of the set $\{1,2,\ldots,n\}$ which do not contain any consecutive numbers.
    Prove that the sum of squares of the products of the numbers in these sets is $(n+1)!-1$.

12. Use Mathematical Induction to prove the following propositions, for $n \in \mathbb{N}$ (unless further restricted).

    (a) $2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 2$.

    (b) $\dfrac{n^3}{3} + \dfrac{n^5}{5} + \dfrac{7n}{15}$ is an integer.

    (c) For all $n > 2$, the sum of the interior angles of a convex polygon of n sides is $180(n-2)^\circ$.

    (d) If a set A contains n elements then the power set of A contains 2^n elements.

    (e) The Fibonacci numbers are defined by:

    $$F_1 = F_2 = 1, \text{ and } F_{n+2} = F_n + F_{n+1}, n \geqslant 1.$$

    Show that

    $$F_1 + F_2 + \cdots + F_n = F_{n+2} - 1.$$

    (f) $1^2 + 3^2 + \cdots + (2n-1)^2 = \frac{n}{3}(2n-1)(2n+1)$.

    (g) $2^n \geqslant n^2$ for all $n \geqslant 4$.

    (h) $2^n \geqslant n^5$ for all $n \geqslant 23$.

    (i) $e^n \geqslant 10n$ for all $n \geqslant 4$.

    (j) $\dfrac{1}{1 \times 3} + \dfrac{1}{3 \times 5} + \cdots + \dfrac{1}{(2n-1)(2n+1)} = \dfrac{n}{2n+1}$.

13. Prove that

    $$\frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1$$

    for all integers $n > 1$.

    
    The lesson with this example is that sometimes one needs to prove something stronger. A naive (perhaps, that should be: *first*) attempt at the problem, might be to define

    $$P(n) : \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1$$

    and try to prove $P(n)$ by induction for all $n \geqslant 2$, but at the inductive step we seem to be stuck because we want to increase the righthand side.
    So define the RHS of $P(n)$ to be slightly smaller than 1 such that it depends on n.

14. (AMO 2005, Problem 4) Show that, for $n \in \mathbb{N}$, there exists an $x \in \mathbb{N}$ such that

    $$\sqrt{x + 2004^n} + \sqrt{x} = (\sqrt{2005} + 1)^n.$$

15. Given $x_1 = 1$

    $$x_{n+1} = x_n + \frac{1}{2x_n}, n \geqslant 1,$$

    prove that $\lfloor 25x_{625} \rfloor = 625$.

16. (You may like to come back to this problem after having done some Number Theory.) What are the last two digits of 2^{222}?

Some ideas: You might be wondering what Euler's Theorem or Fermat's Little Theorem have to say about this. You should be able to see that the problem is equivalent to asking what 2^{222} *modulo* 100. Since 100 is not prime, we cannot use Fermat's Little Theorem, and we cannot use Euler's Theorem directly since 2 and 100 are not coprime, but one could find what 2^{222} *modulo* 25 by Euler's Theorem and lift that to a result *modulo* 100.

Another approach is repeated squaring. This certainly works, and on a computer is quite fast, but for a human it's a bit tedious.

Yet another approach is to recover something that looks a little like a Fermat's Little Theorem result, namely finding a cycle such that:

$$2^{k+\ell} \equiv 2^k \pmod{100}.$$

In fact, one finds that

$$2^2 \equiv 4 \pmod{100} \quad \text{and} \quad 2^{12} = 4096 \equiv -4 \pmod{100}.$$

Note that this does *not* imply that 2^{10} is congruent to -1 *modulo* 100, since 2^2 does *not* have a multiplicative inverse *modulo* 100, but nevertheless one can show (by induction):

$$2^{2+10n} \equiv (-1)^n 2^2 \pmod{100}$$

and use it to solve the problem.

17. Prove any natural number $n \geqslant 2$ is the product of primes.
Hint. Define a $P(n)$ and use an Aladdin's version of *complete induction* where the *base case* proves $P(2)$ holds.

18. Prove that with 3c and 5c stamps, any postage denomination greater than 7c is possible.
Hint. Define a $P(n)$ and use a variant of the *giant's version,* where the *base case* proves $P(8)$, $P(9)$ and $P(10)$ hold, and the *inductive step* proves $P(k) \implies P(k+3)$.

19. Prove $10 \,|\, (n^5 - n)$ for all nonnegative integers n.
Hint. Define a $P(n)$ and as per the traditional *giant's version,* prove $P(k) \implies P(k+2)$. What's the *base case?*

20. Prove an equilateral triangle can be partitioned into n subtriangles that are all equilateral, for all $n \geqslant 6$.
Hint. Define a $P(n)$. Prove $P(4)$, $P(6)$, $P(7)$, $P(8)$. Prove $P(k) \implies P(k+3)$, using $P(4)$, as per a variant of the *giant's version.* What's the *base case?*

CHAPTER 4

Algebra: Quadratic polynomials and Viète's Theorem

4.1 Polynomials

A **polynomial** in x **over** $\mathbb{Z}$ (or **over** $\mathbb{Q}$, or **over** $\mathbb{R}$, or ...) is an *expression* of the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n$ are fixed numbers called **coefficient**s, that are elements of $\mathbb{Z}$ (or $\mathbb{Q}$, or $\mathbb{R}$, or ...).
For convenience, we usually like to give such expressions some sort of label; so that we will usually write something like:

Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$.

(Here $p(x)$ is said: *p of x*).
If $p(x)$ is identically 0, $p(x)$ is called the **zero polynomial**.
If $p(x)$ is not the *zero polynomial*, we insist that $a_n \neq 0$; then

n is the **degree** of $p(x)$,
a_n is the **leading coefficient** of $p(x)$, and
$a_n x^n$ is the **leading term** of $p(x)$.

Note that we haven't yet defined the *degree* of the *zero polynomial*. More generally,

a_k is the k^{th} *coefficient* of $p(x)$,
$a_k x^k$ is the k^{th} *term* of $p(x)$; and
a_0 is both the **constant coefficient** and **constant term** of $p(x)$.

Writing $p(3)$ is a shorthand way of writing:

$$a_n \cdot 3^n + a_{n-1} \cdot 3^{n-1} + \cdots + a_2 \cdot 3^2 + a_1 \cdot 3 + a_0$$
(the expression $p(x)$ with 3 substituted for each occurrence of x).

In words, we say for $p(3)$, *the value of the polynomial* $p(x)$ *at* $x = 3$ or $p(x)$ *evaluated at* $x = 3$ or just simply, *p of 3*.
If the leading coefficient $a_n = 1$ then the polynomial $p(x)$ is said to be **monic**.
An **equation** is something of the form

$$\textit{expression} = \textit{expression}.$$

So a **polynomial equation** is something of the form

$$\textit{polynomial} = \textit{expression},$$

where the *expression* on the right-hand side is usually just a *value* (i.e. a number) and that *value* is usually 0. A *polynomial equation* (in x) usually only holds true for a small number of x values. For example,

$$x^2 - 1 = 0$$

only holds true for the x-values, 1 and -1. We say, the *polynomial equation* $x^2 - 1$ *is satisfied by 1 and* -1; or we say, it has **roots** 1 and -1; or we say, it has **solutions** 1 and -1.

On the other hand, a *polynomial* $p(x)$ is simply an expression in x. It makes sense, (i.e. has a *value*) for *any* choice of x. Sometimes this value is 0. For this reason, a choice of x such that $p(x)$ evaluates to 0, is called a **zero** of $p(x)$.

Example 4.1.1. *(a) The* ***polynomial*** $x^2 - 3x + 2$ *has two* ***zeros****, namely 1 and 2.*

(b) The ***polynomial equation*** $x^2 - 3x + 2 = 0$ *has two* ***roots****, namely 1 and 2.*
(In this context, the term ***root*** *can be used interchangeably with* ***solution****.)*

So we see that the terms *zero* and *root* (or *solution*) generally amount to the same thing but from different viewpoints!

Two polynomials $p(x), q(x)$ are said to be **equal** if they have the same value for every value of x. This occurs *if and only if* $p(x)$ and $q(x)$ have the same degree and identical coefficients. *Note that two polynomials can have the same value at several* x*-values without being* equal *polynomials.*

Example 4.1.2. *(a) The polynomials* $p(x) = x^2 + 1$ *and* $q(x) = x^4 + 1$ *are equal at* $x = 0$*, i.e.*

$$p(0) = q(0),$$

but $p(x), q(x)$ *are* not *equal polynomials.*

(b) The polynomials $p(x) = (x-1)^2$ *and* $q(x) = x^2 - 2x + 1$ are *equal polynomials, and often this is written:* $p(x) = q(x)$*. Usually, the context makes it clear what is meant, but to avoid any ambiguity it is a good idea to get into the practice of writing:*

$$p(x) = q(x) \quad \textit{for all } x.$$

4.2 Degree of a polynomial

We will use the notation $\partial(p)$ for the **degree** of polynomial $p(x)$, which we effectively defined to be n, if the highest power of x occurring in $p(x)$ is n. As yet we haven't defined the *degree* of the *zero polynomial*. Considering, sums and products of polynomials, it would seem ∂ has the following properties:

Properties of ∂. For polynomials $p(x)$ and $q(x)$,

1. $\partial\big(p(x) + q(x)\big) \leqslant \max\big(\partial(p), \partial(q)\big)$.

2. $\partial\big(p(x) \cdot q(x)\big) = \partial(p) + \partial(q)$.

Note that in 1., the "$\leqslant$" is necessary for the case where the leading terms of $p(x)$ and $q(x)$ cancel.


It would seem useful that these properties be generally true. So this brings into question what the *degree* of the *zero polynomial* should be. To make the properties above work with the *zero polynomial*, we *invent* a special integer $-\infty$ which is smaller than any genuine integer and has the property that:

$$-\infty + k = -\infty$$

for any integer k. The zero polynomial is then said to have degree $-\infty$ (rather than 0). All other **constant polynomial**s (i.e. *polynomial*s consisting of only a nonzero *constant* term) have degree 0.

Division Algorithm for polynomials.
*For polynomials $p(x), u(x)$ with $u(x) \neq 0$ there exist polynomials $q(x)$ (the **quotient**) and $r(x)$ (the **remainder**) such that*

$$p(x) = u(x)\,q(x) + r(x) \text{ where } \partial(r) < \partial(u).$$

Essentially $q(x), r(x)$ are the polynomials that make the following division work:

$$u(x)\,)\,\overline{p(x)}\quad \begin{matrix} q(x) & \text{rem.} & r(x) \end{matrix}$$

(Compare the above with the corresponding Division Algorithm *for* integers *in Chapter 6.)*

Often $u(x) = x - a$ for some fixed number a, in which case $\partial(u) = 1$ so that $r(x)$, being necessarily of lower degree, is a **constant polynomial**. If $r(x)$ is the zero polynomial then $u(x)$ (and also $q(x)$) is a **factor** of $p(x)$. Now that we have a concrete idea of what polynomial division is, we are in a position to make sense of the following theorems.

Theorem (Remainder Theorem). *If the polynomial $p(x)$ is divided by the polynomial $x - a$ then the remainder is $p(a)$.*

Proof. Write $p(x) = (x - a)q(x) + r(x)$. Then $r(x)$ is a *constant* polynomial (and so we may as well drop the x and simply write r). Substituting a for x gives

$$\begin{aligned} p(a) &= (a - a)q(a) + r \\ &= r, \end{aligned}$$

i.e. the remainder is $p(a)$. □

An immediate consequence of the Remainder Theorem is the following.

Theorem (Factor Theorem). *Let $p(x)$ be a polynomial.*
Then a is a zero *of $p(x)$ $\iff$ $x - a$ is a factor of $p(x)$.*

Proof. Let $p(x)$ be a polynomial.

($\Longrightarrow$) First assume a is a zero of $p(x)$.
Then $p(a) = 0$, and so by the Remainder Theorem, for some polynomial $q(x)$,

$$\begin{aligned} p(x) &= (x - a)q(x) + p(a) \\ &= (x - a)q(x), \end{aligned}$$

and so $x - a$ is a factor of $p(x)$.

($\Longleftarrow$) Now assume $x - a$ is a factor of $p(x)$.

$$\begin{aligned} &\Longrightarrow p(x) = (x - a)q(x) \text{ for some polynomial } q(x) \\ &\Longrightarrow p(a) = (a - a)q(a) \\ &\qquad\quad = 0 \end{aligned}$$

and so a is a zero of $p(x)$. □

Definition. A polynomial $p(x)$ is said to **factor over** $\mathbb{Z}$ (resp. *over* $\mathbb{Q}$, or *over* $\mathbb{R}$, or ...) if $p(x) = u(x)\,q(x)$ for some non-constant polynomials $u(x), q(x)$ that are both *polynomials over* $\mathbb{Z}$ (resp. *over* $\mathbb{Q}$, or *over* $\mathbb{R}$, or ...).

The next result shows that finding rational zeros of monic polynomials over $\mathbb{Z}$ is relatively easy.

Theorem (Rational Zero Theorem). *If $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0$ is a* monic *polynomial over $\mathbb{Z}$, and α is a rational zero of $p(x)$, then $\alpha \in \mathbb{Z}$.*

Proof. Assume $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0$ where $a_0, a_1, \ldots, a_{n-1} \in \mathbb{Z}$, and suppose $\alpha \in \mathbb{Q}$ is a zero of $p(x)$. Then $\alpha = s/t$ for some $s, t \in \mathbb{Z}$ with $t \neq 0$.
We may assume s/t is reduced to lowest terms so that $(s,t) = 1$.
For a contradiction, suppose $t > 1$, and hence has a prime divisor q.

$$\begin{aligned}
\implies \quad 0 &= p(\alpha) \\
&= \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 \\
&= \frac{s^n}{t^n} + a_{n-1}\frac{s^{n-1}}{t^{n-1}} + \cdots + a_1\frac{s}{t} + a_0 \\
\implies \quad 0 &= s^n + a_{n-1}s^{n-1}t + \cdots + a_1st^{n-1} + a_0t^n, \text{ multiplying through by } t^n \\
\implies \quad -s^n &= (a_{n-1}s^{n-1} + \cdots + a_1st^{n-2} + a_0t^{n-1})t \\
\implies \quad t &\,\big|\, s^n \\
\implies \quad q &\,\big|\, s^n \\
\implies \quad q &\,\big|\, s \\
\implies \quad (s,t) &\geqslant q > 1 \text{ (contradiction)}
\end{aligned}$$

So, in fact, $t = 1$ and hence $\alpha \in \mathbb{Z}$. □

Remark. By the Rational Zero Theorem, to find the rational zeros of a *monic* polynomial with integer coefficients we only need to check all integer divisors of the *constant* coefficient a_0.

4.3 Quadratic polynomials

Definition. A **quadratic polynomial** is simply a *polynomial* of *degree* 2, i.e. a polynomial of form

$$ax^2 + bx + c$$

where a, b, c are fixed numbers and $a \neq 0$. Let $p(x) = ax^2 + bx + c$. Below, we will show many key properties of $p(x)$ are associated with the value of its **discriminant** $\Delta := b^2 - 4ac$.

Theorem 4.3.1. *Let $p(x) = ax^2 + bx + c$ be a quadratic polynomial over $\mathbb{R}$ with discriminant $\Delta = b^2 - 4ac$. Then*

$$p(x) \text{ has real zeros } \frac{-b \pm \sqrt{\Delta}}{2a} \iff \Delta \geqslant 0.$$

Proof. First using the identity

$$(A + B)^2 = A^2 + 2AB + B^2$$

we rewrite $p(x)$ to put "x in one place" (this process is called **completing the square**)

$$\begin{aligned} p(x) &= ax^2 + bx + c \\ &= a\Big(x^2 + \frac{b}{a}x + \frac{c}{a}\Big) \\ &= a\Big(x^2 + \frac{b}{a}x + \Big(\frac{b}{2a}\Big)^2 + \frac{c}{a} - \Big(\frac{b}{2a}\Big)^2\Big) \\ &= a\Big(\Big(x + \frac{b}{2a}\Big)^2 + \frac{4ac - b^2}{4a^2}\Big) \\ &= a\Big(\Big(x + \frac{b}{2a}\Big)^2 - \frac{b^2 - 4ac}{4a^2}\Big) \\ &= a\Big(\Big(x + \frac{b}{2a}\Big)^2 - \frac{\Delta}{4a^2}\Big) \end{aligned}$$

Since the square $\big(x + b/(2a)\big)^2$ is non-negative, the expression

$$\Big(x + \frac{b}{2a}\Big)^2 - \frac{\Delta}{4a^2}$$

is strictly positive, if Δ is negative, and hence for $p(x)$ to have zeros, it is necessary that $\Delta \geqslant 0$. Thus, assuming $\Delta \geqslant 0$, by using (in reverse) the identity

$$(A + B)(A - B) = A^2 - B^2,$$

we proceed further

$$\begin{aligned} p(x) &= a\Big(\Big(x + \frac{b}{2a}\Big)^2 - \frac{\Delta}{4a^2}\Big) \\ &= a\Big(\Big(x + \frac{b}{2a}\Big)^2 - \Big(\frac{\sqrt{\Delta}}{2a}\Big)^2\Big) \\ &= a\Big(x + \frac{b}{2a} + \frac{\sqrt{\Delta}}{2a}\Big)\Big(x + \frac{b}{2a} - \frac{\sqrt{\Delta}}{2a}\Big) \\ &= a\Big(x - \frac{-b - \sqrt{\Delta}}{2a}\Big)\Big(x - \frac{-b + \sqrt{\Delta}}{2a}\Big). \end{aligned}$$

Thus, $p(x)$ has real zeros $(-b - \sqrt{\Delta})/(2a)$ and $(-b + \sqrt{\Delta})/(2a)$ if and only if $\Delta \geqslant 0$. □

Remark 4.3.2. Observe that the quadratic polynomial $p(x) = ax^2 + bx + c$ (whose discriminant is $\Delta = b^2 - 4ac$), has

- two *distinct* real zeros, if $\Delta > 0$;
- two *equal* real zeros, if $\Delta = 0$;
- no real zeros, if $\Delta < 0$.

Further, a *real* square is always either positive or zero. So

$$\Big(x + \frac{b}{2a}\Big)^2 - \frac{\Delta}{4a^2} \geqslant -\frac{\Delta}{4a^2} \quad \text{for all } x.$$

Hence, $p(x)$ is **bounded below** (resp. **bounded above**) by $-\Delta/(4a)$, if $a > 0$ (resp. if $a < 0$), and since this value is attained at $x = -b/(2a)$, $-\Delta/(4a)$ is the **minimum value** (resp. **maximum value**) of $p(x)$.

Over $\mathbb{Z}$, we can say a little more, but in order to prove, we need some results that follow easily.

Definition 4.3.3. A **perfect square** is the square of an integer.

Definition 4.3.4. The **parity** of an integer is its *oddness/evenness*. Let $m, n \in \mathbb{Z}$. If m, n are both even or both odd, we say m and n *have the same parity*. On the other hand, if one of m or n is odd and the other is even, we say m and n *have opposite parity*.

Lemma 4.3.5. *Let $m, n \in \mathbb{Z}$. Then m, n have the same parity (resp. opposite parity) if and only if $m - n$ and $m + n$ are even (resp. odd).*

Proof. If $m, n \in \mathbb{Z}$ have the same parity then $m = 2k + \varepsilon$ and $n = 2\ell + \varepsilon$ for some $k, \ell \in \mathbb{Z}$ and $\varepsilon \in \{0, 1\}$, in which case, $m - n = 2(k - \ell)$ is even.
If $m, n \in \mathbb{Z}$ have opposite parity then without loss of generality $m = 2k + 1$ (odd) and $n = 2\ell$ (even) for some $k, \ell \in \mathbb{Z}$ and $m - n = 2(k - \ell) + 1$ is odd.
Observe that $-n$ has the same parity as n; so the corresponding result for $m + n$ follows. □

Lemma 4.3.6. *An integer and its square have the same parity.*

Proof. Let $n \in \mathbb{Z}$. Then $n^2 - n = n(n - 1)$ which is even, since it is the product of two consecutive integers, one of which is even and the other odd. □

Theorem 4.3.7. *Let $p(x) = x^2 + bx + c$ be a monic quadratic polynomial over $\mathbb{Z}$. Then*

$$p(x) \text{ has integer zeros} \iff \Delta = b^2 - 4c \text{ is a perfect square.}$$

Proof. Let $p(x) = x^2 + bx + c$ be a monic polynomial over $\mathbb{Z}$.

($\Longrightarrow$) First assume $p(x)$ has integer zeros.

$$\begin{aligned}
&\Longrightarrow \frac{-b \pm \sqrt{\Delta}}{2} \in \mathbb{Z} \\
&\Longrightarrow -b \pm \sqrt{\Delta} \in 2\mathbb{Z} \subset \mathbb{Z}, \text{ where } 2\mathbb{Z} \text{ is the set of even integers} \\
&\Longrightarrow \sqrt{\Delta} \in \mathbb{Z} \\
&\Longrightarrow \Delta \text{ is a perfect square.}
\end{aligned}$$

($\Longleftarrow$) Now assume $\Delta = b^2 - 4c$ is a perfect square, and hence $\sqrt{\Delta} \in \mathbb{Z}$.

$$\begin{aligned}
&\Longrightarrow -b, b^2, \Delta = b^2 - 4c, \sqrt{\Delta} \text{ have the same parity by Lemmas 4.3.5, 4.3.6} \\
&\Longrightarrow -b \pm \sqrt{\Delta} \text{ is even, by Lemma 4.3.5} \\
&\Longrightarrow \frac{-b \pm \sqrt{\Delta}}{2} \in \mathbb{Z}.
\end{aligned}$$

□

Theorem (Viète's Theorem). *Let $p(x) = x^2 + bx + c$ be a* monic quadratic polynomial *with* zeros α, β. *Then*

$$\begin{aligned}
-b &= \alpha + \beta \\
c &= \alpha.\beta
\end{aligned}$$

Proof. First observe that $(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta$.
Now α, β are the zeros of $p(x)$, and $p(x)$ is monic; so $p(x)$ must factor as per the left-hand side of the above equation. The result now follows by comparison of coefficients of $p(x)$ with the right-hand side of the above equation. □

4.4 Horner's Method and Synthetic Division

A quick method for evaluating a polynomial at a point x_0 is given by **Horner's Method**, which is based on writing the polynomial in "nested form":

$$\begin{aligned} p(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 \\ &= ((\cdots(a_n x + a_{n-1})x + \cdots + a_2)x + a_1)x + a_0 \end{aligned}$$

To evaluate $p(x_0)$ then, we evaluate a sequence of numbers as follows

$$\begin{aligned} b_n &= a_n \\ b_{n-1} &= b_n x_0 + a_{n-1} \\ &\vdots \\ b_0 &= b_1 x_0 + a_0. \end{aligned}$$

Then the final number b_0 is $p(x_0)$, since

$$\begin{aligned} p(x_0) &= ((\cdots(a_n x_0 + a_{n-1})x_0 + \cdots + a_2)x_0 + a_1)x_0 + a_0 \\ &= ((\cdots(b_n x_0 + a_{n-1})x_0 + \cdots + a_2)x_0 + a_1)x_0 + a_0 \\ &= ((\cdots b_{n-1}x_0 + \cdots + a_2)x_0 + a_1)x_0 + a_0 \\ &\vdots \\ &= b_1 x_0 + a_0 \\ &= b_0. \end{aligned}$$

Of course, from above we saw that $p(x_0)$ is the *remainder* when $p(x)$ is divided by $x - x_0$. The way the above process is usually performed is via a tableau as below:

	a_n	a_{n-1}	a_{n-2}	$\cdots$	a_0
x_0		$b_n x_0$	$b_{n-1}x_0$	$\cdots$	b_1
	b_n	b_{n-1}	b_{n-2}	$\cdots$	b_0

As a by-product, $b_n, b_{n-1}, \ldots, b_1$ are the coefficients of the *quotient* when $p(x)$ is divided by $x - x_0$, i.e.

$$p(x) = (x - x_0)(b_n x^{n-1} + b_{n-1}x^{n-2} + \cdots + b_1) + b_0.$$

Example 4.4.1. Find $p(3)$ by Horner's Method/Synthetic Division for $p(x) = x^3 + 2x^2 - 13x + 10$, and hence find the *quotient* and *remainder* when $p(x)$ is divided by $x - 3$.

Exercise Set 4.

1. The *quadratic equation* $x^2 - 3x - 5 = 0$ has *roots* α, β. Determine $\alpha^2 + \beta^2$ and $\alpha^{-2} + \beta^{-2}$.

2. The *quadratic polynomial* $x^2 + 4x - 1$ has *zeros* α, β. Determine $\alpha^3 + \beta^3$ and $\alpha^{-3} + \beta^{-3}$. *Hint.* $(\alpha + \beta)^3 = \alpha^3 + \beta^3 + 3\alpha^2\beta + 3\alpha\beta^2 = \alpha^3 + \beta^3 + 3\alpha\beta(\alpha + \beta)$.

3. Solve $2\big(x + \dfrac{1}{x}\big)^2 - \big(x + \dfrac{1}{x}\big) = 10$.

4. Use the Remainder and Factor Theorems to factorise

   (i) $x^3 - 2x^2 - 5x + 6$ (ii) $x^3 - 5x^2 + 3x + 1$

5. The *quadratic polynomial* $ax^2 + bx - 4$ leaves remainder 12 on division by $x - 1$ and has $x + 2$ as a factor. Find a, b and the *zeros* of the polynomial.

6. Find a *quadratic equation* with *roots* $2 + \sqrt{3}$ and $2 - \sqrt{3}$.

7. June solved a *quadratic equation* of the form:

$$ax^2 + bx + c = 0$$

   and got 2 as a root. Kay switched the b and the c and got 3 as a root. What was June's equation?

8. The equation $x^2 + ax + (b + 2) = 0$ has real roots. What is the least value that $a^2 + b^2$ could be?

∗9. If a, b are odd integers, prove that the equation

$$x^2 + 2ax + 2b = 0$$

   has no *rational* roots.

10. Find $p(1)$ and $p(-2)$ via Horner's method, given $p(x) = 2x^4 - 3x^3 - 2x^2 + x - 4$.

11. Find the quotient and remainder when

    (i) $3x^5 + 2x^4 - 3x^2 + 2x - 7$ is divided by $x + 2$;
    (ii) $x^3 + 3x^2 - 5x + 6$ is divided by $(x - 1)(x + 2)$;
    (iii) $x^4 - 2x^3 + x^2 - 5x + 11$ is divided by $x^2 + 3x + 2$.

CHAPTER 5

Combinatorics, Factorisation Background

Note that in this chapter, $n, k, r \in \mathbb{N}$.

5.1 Factorisation of $a^n \pm b^n$

Let us start with something familiar, summing

$$1 + r + r^2 + \cdots + r^{n-1}.$$

This is a *geometric series* with *common ratio* r (with 1 as the first term). It's useful to give the series a name. Lets call it S_n (the sum to n terms). Then

$$\begin{aligned} S_n &= 1 + r + r^2 + \cdots + r^{n-1} && (5.1.1)\\ rS_n &= \quad r + r^2 + \cdots + r^{n-1} + r^n && (5.1.2)\\ (1-r)S_n &= 1 - r^n, \qquad \text{subtracting (5.1.2) from (5.1.1).} && (5.1.3) \end{aligned}$$

In the context of *geometric series,* we would then proceed to isolate S_n, but in the current context, substituting (5.1.1) back in (5.1.3) and swapping the lefthand and righthand sides, gives us the factorisation

$$1 - r^n = (1-r)(1 + r + r^2 + \cdots + r^{n-1}).$$

Now replace r by b/a, and then multiply through both sides by a^n:

$$\begin{aligned} 1 - \left(\frac{b}{a}\right)^n &= \left(1 - \left(\frac{b}{a}\right)\right)\left(1 + \left(\frac{b}{a}\right) + \left(\frac{b}{a}\right)^2 + \cdots + \left(\frac{b}{a}\right)^{n-1}\right)\\ a^n - b^n &= (a-b)(a^{n-1} + a^{n-2}b + \cdots + a^{n-k}b^{k-1} + \cdots + b^{n-1}). \qquad (5.1.4) \end{aligned}$$

In multiplying the righthand side by a^n, we multiply the first bracket by a and the second bracket by a^{n-1}, and note that that means the sum of the indices appearing in each term in the second bracket is $n-1$. We now have a factorisation of $a^n - b^n$.

To get a similar factorisation of $a^n + b^n$, we replace b by $-b$ in (5.1.4). Note that, in order for this to actually give us a factorisation of $a^n + b^n$, we need $-(-b)^n = b^n$, i.e. n must be *odd.* Thus,

$$\begin{aligned} a^n - (-b)^n &= \big(a - (-b)\big)\big(a^{n-1} + a^{n-2}(-b) + \cdots + a^{n-k}(-b)^{k-1} + \cdots + (-b)^{n-1}\big)\\ &= \big(a - (-b)\big)\big(a^{n-1} + a^{n-2}(-b) + \cdots + a^{n-k}(-b)^{k-1} + \cdots + (-b)^{n-1}\big)\\ a^n + b^n &= (a+b)(a^{n-1} - a^{n-2}b + \cdots + (-1)^{k-1}a^{n-k}b^{k-1} + \cdots + b^{n-1}), \quad \text{for } n \text{ odd.} \end{aligned} \qquad (5.1.5)$$

Observe that $(-b)^{n-1} = b^{n-1}$ in (5.1.5), since $n-1$ is even.

Summarising, we have

$$\boxed{\begin{aligned} a^n - b^n &= (a-b)(a^{n-1} + a^{n-2}b + \cdots + a^{n-k}b^{k-1} + \cdots + b^{n-1})\\ a^n + b^n &= (a+b)(a^{n-1} - a^{n-2}b + \cdots + (-1)^{k-1}a^{n-k}b^{k-1} + \cdots + b^{n-1}), \quad \text{for } n \text{ odd.} \end{aligned}}$$

5.2 Expansion of the binomial $(a+b)^n$

Now consider the expansion of $(a+b)^n$. For completeness (it gives the first line of Pascal's Triangle), also consider $n=0$:

$$\begin{aligned}(a+b)^0 &= 1\\ (a+b)^1 &= a+b\\ (a+b)^2 &= a^2+2ab+b^2\\ (a+b)^3 &= a^3+3a^2b+3ab^2+b^3\end{aligned}$$

and in general,

Theorem 5.2.1 (Binomial Theorem). *For $n \in \mathbb{N}\cup\{0\}$,*

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n}b^n,$$

where

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

Proof. In expanding $(a+b)^n$ we note that there are 2^n terms, since there are 2 terms in each of n brackets, where each term is a product of some number of as and some number of bs. Since the number of brackets is n, and in forming a particular term we take either an a or a b from each bracket, the sum of the numbers of as and bs making up the term is n, i.e. each term of $(a+b)^n$ is of form

$$a^{n-r}b^r,$$

for some $r \in \{0,1,2,\ldots,n\}$. Evidently, a term of form $a^{n-r}b^r$ occurs many times, and we will see that this number is dependent on both n and r; we call this number a **binomial coefficient**, and write it as follows:

$$\binom{n}{r}.$$

Thus,

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n}b^n.$$

Now we do must establish a value for $\binom{n}{r}$; in forming $a^{n-r}b^r$ we may choose a b from r of the n brackets, then the brackets we take the as from are determined.

The first b can come from any of n brackets, leaving $n-1$ brackets from which to choose the second b, and so on, until the r^{th} b can be taken from $n-(r-1) = n-r+1$ brackets.

It would appear then that there are

$$n(n-1)(n-2)\cdots(n-r+1)$$

ways of choosing which of the r brackets we take a b from. However, we can choose the same r brackets in

$$r(r-1)(r-2)\cdots 1 = r!$$

different orders. Thus counting in the above way, overcounts every possible way by a factor of $r!$. Thus we have

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

□

Note that there are r factors in both the numerator and denominator of the expression

$$\frac{n(n-1)(n-2)\cdots(n-r+1)}{r(r-1)(r-2)\cdots 1},$$

for $\binom{n}{r}$. The formula above doesn't appear to given an expression for $r = 0$. Just know that empty products are 1 (we will revisit this idea in a later chapter); so that $\binom{n}{0} = 1$.

Properties of binomial coefficients. Binomial coefficients have the following properties. You are asked to prove the fourth of these in Exercise 11.

$$\begin{aligned}
\binom{n}{0} &= 1 \\
\binom{n}{1} &= n \\
\binom{n}{r} &= \binom{n}{n-r}, \quad \text{(the symmetry property)} \\
\binom{n+1}{r+1} &= \binom{n}{r} + \binom{n}{r+1}
\end{aligned}$$

Pascal's Triangle. The *binomial coefficients* form a triangle known as **Pascal's triangle**, where n is the row number, starting at the 0^{th} row, and r is fixed along diagonals.

$$\begin{array}{ccccccccccccccc}
 & & & & & & 1 & & & & & & & & \\
 & & & & & 1 & & 1 & & & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & & & \\
\binom{n}{0} & & \binom{n}{1} & \cdots & \binom{n}{r} & & \binom{n}{r+1} & \cdots & \binom{n}{n-1} & & \binom{n}{n}
\end{array}$$

5.3 Counting

Fundamental Counting Principle. If one event can occur in m ways and another event in n ways, then:

FC1: *one* of the events can occur in $m + n$ ways. **(addition principle)**
In this case, the events are said to be **mutually exclusive**.

FC2: *both* events can occur in mn ways. **(multiplication principle)**
In this case, the events are said to be **independent**.

Definition 5.3.1. An **r-sequence** is a sequence of r elements, represented by enclosing the r elements separated by commas within parentheses, e.g. $(a_1, a_2, \ldots, a_r)$.

An **r-set** is a set of r (distinct) elements; it differs from an *r-sequence* in that it an *r-sequence* is ordered, whereas an r-set is not, in the sense, that if one rearranges the elements of an r-set, it is the same r-set, e.g. the 3-set, $\{1, 2, 3\} = \{1, 3, 2\}$.

A **permutation** of an n-set, is an n-sequence of that n-set. From a given n-set, one can form $n!$ permutations of that n-set.

Remark 5.3.2. From a given r-set, one can form $r!$ distinct r-sequences, e.g. from the 3-set $\{a, b, c\}$ one can form $3! = 6$ 3-sequences,

$$(a, b, c),\ (a, c, b),\ (b, a, c),\ (b, c, a),\ (c, a, b),\ (c, b, a).$$

Notation 5.3.3. We will use # followed by a quoted string, to denote **number of** ways of forming the thing represented by the quoted string. In particular,

$$\begin{aligned} \#\text{“}r\text{-sequences formed from an } n\text{-set”} &= n(n-1)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!} \\ \#\text{“}r\text{-sets formed from an } n\text{-set”} &= \frac{n(n-1)\ldots(n-r+1)}{r!} \\ &= \frac{n!}{r!(n-r)!} \\ &= \binom{n}{r}. \end{aligned}$$

Remark 5.3.4. In the literature, #“r-sequences formed from an n-set” is denoted by ${}^nP_r$, and is vaguely referred to as the “number of permutations”, but this is an abbreviation of “number of permutations of r of n objects”. Without these last five words it’s gibberish.
Also, in the literature, #“r-sets formed from an n-set” is commonly denoted by ${}^nC_r$, and is vaguely referred to as the “number of combinations”. However, ${}^nC_r = \binom{n}{r}$, and *mathematicians* read both notations as *“n choose r”*, which is beautifully succinct and precise, and mnemonically uses the C.

Exercise Set 5.

1. Expand:
   (a) $(1 + a)(1 + b)(1 + c)$.
   (b) $(1 + x)^3$.
   (c) $(a + b + c)^3$.
   (d) $(a + b)^4$.

2. Factor:
   (a) $a - b$ as the difference of two squares.
   (b) $x + 2\sqrt{xy} + y$, where $x, y \geqslant 0$.

3. (i) Factor $a^6 - b^6$ as the difference of two squares.
   (ii) Factor $a^6 - b^6$ as the difference of two cubes.
   (iii) Factor $a^6 - b^6$ as the difference of two sixth powers.
   (iv) Fully factor $a^6 - b^6$.

4. Factor fully (over $\mathbb{Q}$):

   (a) $(a+b)^2 - c^2$.
   (b) $a^4 + 2a^2b^2 + b^4$.
   (c) $a^4 + a^2b^2 + b^4$.
   (d) $x^4 + 3x^2 + 4$.
   (e) $x^4 - 15x^2y^2 + 9y^4$.
   (f) $a^2 - 2a - b^2 + 1$.

5. Find all possible integers solutions to the equation,

$$x^2 + y^2 + z^2 = 10(x + y + z).$$

   Prove that there are no other integer solutions.

6. Express

$$2(a-b)(a-c) + 2(b-c)(b-a) + 2(c-a)(c-b)$$

   as the sum of three squares.

7. If $x + 3$ divides $3x^2 + x + k$ without remainder, find the value of k.

8. Factor $n^4 + 4$ as the product of two quadratics.
   For what positive integer values of n is $n^4 + 4$ a prime number?

9. Factor:

   (a) $1 + y(1 + x)^2(1 + xy)$.
   (b) $1 - b - a^2 + a^3b + a^2b^3 - a^3b^3$.

10. Given that

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

    (a) find $\binom{21}{3}$ and $\binom{12}{5}$.
    (b) Show that $\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$.

11. (i) Show that the sum of the coefficients of the r^{th} and $(r+1)^{\text{st}}$ term in the expansion of $(1+x)^n$ is equal to the $(r+1)^{\text{st}}$ term in the expansion of $(1+x)^{n+1}$. (To avoid awkward numbering of the terms, let us call the x^0 term, the *zeroth* term, so that the r^{th} term is the term involving x^r.)

    (ii) Prove the same result using the factorial expression of $\binom{n}{r}$.

    (iii) How is the result connected with Pascal's Triangle?

    (iv) Use the result of (i) (or equivalently, (ii)), to prove

$$\sum_{m=r}^{n} \binom{m}{r} = \binom{n+1}{r+1},$$

    for $r, n \in \mathbb{N}$ such that $n \geqslant r$.
    Hints. Use induction, with fixed r. Call the displayed equation $P(n)$. The *'base case'* is $P(r)$. It's easy!

12. Given the Binomial Theorem result:
$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n,$$
prove each of the following.

(a) $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$

(b) $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n\binom{n}{n} = 0.$

(c) $\binom{n}{0} + 2\binom{n}{1} + 4\binom{n}{2} + 8\binom{n}{3} + \cdots + 2^n\binom{n}{n} = 3^n.$

(d) $\binom{n}{0} - 2\binom{n}{1} + 4\binom{n}{2} - 8\binom{n}{3} + \cdots + (-1)^n 2^n\binom{n}{n} = \begin{cases} 1, & \text{if } n \text{ is even;} \\ -1, & \text{if } n \text{ is odd.} \end{cases}$

13. Prove that
$$\binom{n}{r} = \frac{n-r+1}{r} \cdot \binom{n}{r-1},$$
and hence find the value of r that maximises $\binom{n}{r}$.

14. Observe that in any row of Pascal's Triangle that the sum of the odd-indexed elements is equal to the sum of the even-indexed elements, i.e.
$$\binom{n}{1} + \binom{n}{3} + \cdots = \binom{n}{0} + \binom{n}{2} + \cdots .$$
Prove this result.
Hint. You've already done it!

15. The relation *divides* (written: $\big|$) is defined and discussed in detail in the next chapter. For now, we say,

> For $a, b \in \mathbb{Z}$, a **divides** b (written: $a \,\big|\, b$) if $b = aq$ for some $q \in \mathbb{Z}$.

In (a), we assume $a, b \in \mathbb{Z}$, so that $a+b, a-b \in \mathbb{Z}$.
However, the statements in (a) still make sense in a *polynomial* context, in which case, '$\big|$' should be interpreted to mean *is a factor of*, i.e. there is *zero remainder*.

(a) Show
  (i) $(a-b) \,\big|\, (a^n - b^n)$, for all $n \in \mathbb{N}$.
  (ii) $(a+b) \,\big|\, (a^n + b^n)$, for all odd $n \in \mathbb{N}$.

(b) Show
  (i) $3 \,\big|\, 5^{39} - 2^{39}$.
  (ii) $5 \,\big|\, 2^{99} + 3^{99}$.
  (iii) $5 \nmid 2^{98} + 3^{98}$.
  (iv) $7 \,\big|\, 2^{99} + 3^{99} + 4^{99} + 5^{99}$.
  (v) $10 \,\big|\, 2^{99} - 4^{99} - 7^{99} + 9^{99}$.

(c) Prove that $1991 \,\big|\, 3500^n - 728^n - 785^n + 4^n$ for all $n \in \mathbb{N}$.

16. Find a short expression for the following.

    (a) $1 + x + x^2 + \cdots + x^n$ for all positive integers n.

    (b) $1 - x + x^2 - \cdots + x^n$ for all even positive integers n.

17. Factor $a^2(b-c) + b^2(c-a) + c^2(a-b)$.

18. Find all positive integer pairs (x, y) that satisfy

$$x^2 - 871 = y^6.$$

19. If $\left(a - \frac{1}{a}\right)^2 = 3$ and $a - \frac{1}{a} > 0$, evaluate

    (i) $a^3 - \frac{1}{a^3}$. (ii) $a^4 + \frac{1}{a^4}$.

20. If a is the difference between any quantity and its reciprocal, and b is the difference between the square of the same quantity and the square of its reciprocal, show that

$$a^2(a^2 + 4) = b^2.$$

CHAPTER 6

Number Theory – introduction

Number Theory is mainly concerned with properties of $\mathbb{N}$ and more generally $\mathbb{Z}$. Throughout this chapter and the next two (also about Number Theory) we will use $a, b, c, d, m, n, q, r, N$ for integers, sometimes without explicitly stating that they are.

Division Algorithm. *For integers a, b with $a > 0$ there exist integers q (the **quotient**) and r (the **remainder**) such that*

$$b = aq + r \text{ and } 0 \leqslant r < a.$$

Essentially q, r are the numbers that make the following division work:

$$\begin{array}{r|l} & q \quad \textit{rem.} \quad r \\ \hline a & b \end{array}$$

(Often $b > 0$, but this is not necessary.)

Let's apply the Division Algorithm to a few examples:

$$\begin{array}{llll} \text{if } a = 7 & \text{and } b = 22 & \text{we write } 22 = 7 \cdot 3 + 1 & (\text{so } q = 3 \text{ and } r = 1); \\ \text{if } a = 113 & \text{and } b = 355 & \text{we write } 355 = 113 \cdot 3 + 16 & (\text{so } q = 3 \text{ and } r = 16); \\ \text{if } a = 8 & \text{and } b = 72 & \text{we write } 72 = 8 \cdot 9 + 0 & (\text{so } q = 9 \text{ and } r = 0). \end{array}$$

6.1 Divisibility

In the special case where the **division algorithm** applied to two integers a, b as described above yields a *remainder* of 0, we have the following.

Definition 6.1.1. The following are equivalent.

(i) $a, b \in \mathbb{Z}$ and $b = aq$ for some $q \in \mathbb{Z}$;

(ii) b is a **multiple** of a;

(iii) b is **divisible by** a (written: $b \vdots a$);

(iv) a is a **divisor** of b;

(v) a **divides** b (written: $a \,|\, b$).

If $a \,|\, b$ does not hold, then a does **not divide** b (written: $a \nmid b$). For example,

$$7 \,|\, 35, \quad -3 \,|\, 21, \quad 4 \,|\, 0 \quad \text{and } (a+1) \,|\, a^2 - 1 \text{ for any integer } a$$

but

$$7 \nmid 33, \quad -3 \nmid 22, \quad 0 \nmid 4.$$

Finally, the **divisors** of b are all d such that $d \,|\, b$, e.g. the *divisors* of 12 are those

$$d \in \{\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12\}.$$

Don't confuse the *divides* symbol: $\mid$ (which is a *vertical* stroke with a little space around it) with the *slash* symbol: / (which separates the numerator and denominator of a fraction). Also, note that despite the symmetry of the symbol: '$\mid$' cannot be used in reverse, i.e. $a \mid b$ and $b \mid a$ mean *different* things (see: Property 6.1.3(iv)). What may also be confusing is that $a \mid b$ means the same as $b/a \in \mathbb{Z}$; so *don't* slant '$\mid$'!

Note 6.1.2. While the dictionary meaning of **factor** is a synonym for *divisor*, we prefer to reserve *factor* (of an integer) for the *multiplicands* that make up a product, e.g. $6 = 2 \cdot 3$ writes 6 as the product of *factors* 2 and 3, whereas the *positive divisors* of 6 are 1, 2, 3, and 6.

Properties 6.1.3 (Properties of Divides).

(i) *If* $d \mid a$ *then* $d \mid ax$ *for all* $x \in \mathbb{Z}$.

(ii) *If* $d \mid a$ *and* $d \mid b$ *then* $d \mid a + b$.

(iii) *If* $d \mid a$ *and* $d \mid b$ *then* $d \mid ax + by$ *for all* $x, y \in \mathbb{Z}$.

(iv) *If* $a \mid b$ *and* $b \mid a$ *then* $a = \pm b$.

(v) *If* $a \mid b$ *and* $b \mid c$ *then* $a \mid c$. ***(transitivity)***

Proof of (i)–(iii). Suppose $d \mid a$, and choose arbitrary $x \in \mathbb{Z}$. Then

$$\begin{aligned} & a = dq, \text{ for some } q \in \mathbb{Z} \\ \implies\ & ax = dqx, \text{ where } qx \in \mathbb{Z} \\ \implies\ & d \mid ax. \end{aligned}$$

So (i) holds.
Now suppose also $d \mid b$. Then

$$\begin{aligned} & b = ds, \qquad \text{for some } s \in \mathbb{Z} \\ \implies\ & a + b = dq + ds \\ & \quad = d(q + s), \text{ where } q + s \in \mathbb{Z} \\ \implies\ & d \mid a + b. \end{aligned}$$

So (ii) holds.
Now suppose $d \mid a$ and $d \mid b$. Then,

$$\begin{aligned} & d \mid ax \text{ and } d \mid by \ \ \forall x, y \in \mathbb{Z}, && \text{by (i)} \\ \implies\ & d \mid ax + by, && \text{by (ii)} \end{aligned}$$

So (iii) holds. □

Definition 6.1.4. If p is a prime and α is a positive integer then we write $p^\alpha \parallel m$ if $p^\alpha \mid m$ but $p^{\alpha+1} \nmid m$. In this case we say p^α **exactly divides** m (or p^α **divides** m **exactly**).

6.2 Prime numbers

Definition 6.2.1.

1. $p \in \mathbb{N}$ is **prime** $\iff$ $p > 1$ and the only positive divisors of p are: 1 and p.
2. d is a **unit** $\iff$ $d \mid n$ for all $n \in \mathbb{N}$; 1 is a *unit*.
3. d is a **proper divisor** of N $\iff$ $d \mid N$ and $1 < d < N$.
4. $N \in \mathbb{N}$ is **composite** $\iff$ $N = ab$ for some $a, b \in \mathbb{N}$ such that $1 < a \leqslant b < N$
   $\iff$ N has a proper divisor.

Primality. The property of being prime is a number's **primality**. In the first 100 natural numbers, 25 are prime, namely,

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, and 97.

Lemma 6.2.2. *If $1 < n \in \mathbb{N}$ and n has no prime divisor $p \leqslant \sqrt{n}$ then n is prime.*

Proof. Let $1 < n \in \mathbb{N}$. Consider n is not prime, then n is composite, and so $n = ab$ for some $a, b \in \mathbb{N}$ such that $1 < a \leqslant b < n$ which implies $a \leqslant b = n/a$, i.e. $a^2 \leqslant n$ or $a \leqslant \sqrt{n}$. Thus having no prime divisors $p \leqslant \sqrt{n}$ implies n has no divisors $a \leqslant \sqrt{n}$ and hence n is not composite, and since $n > 1$, n is prime. □

What makes *primes* so interesting is that every *natural number* (other than 1) can be expressed in just one way (except that we may be able to arrange the factors in many ways) as the product of prime divisors, e.g.

$$74844 = 2^2 \cdot 3^5 \cdot 7 \cdot 11.$$

Such a factorisation is called a **prime decomposition** or **prime factorisation**.

If we were to include 1 as a prime then $74844 = 1^5 \cdot 2^2 \cdot 3^5 \cdot 7 \cdot 11$, say, would be "another prime decomposition". Excluding 1 as a prime ensures the *uniqueness* of *prime decompositions*.

The above fact is so important it is given a special name. Let's give it its name and recap what it says, and follow it up by an equally important result from the Greek, Euclid:

Theorem 6.2.3 (Fundamental theorem of arithmetic). *Any natural number n, can be written* uniquely *as follows:*

$$n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k},$$

where $k \in \mathbb{N}$, each p_i is a prime number and $1 < p_1 < p_2 < \cdots < p_k$, and each $e_i \in \mathbb{N}$.

Lemma 6.2.4 (Euclid's Lemma). *If p is prime and $p \mid ab$ then $p \mid a$ or $p \mid b$.*

Algorithm 6.2.5 (Sieve of Eratosthenes). * *For some $N \in \mathbb{N}$, to find all the* primes *p such that $p \leqslant N$, perform the following steps.*

1. *Start by writing down all the* natural numbers *from 1* upto *N.*
2. *Cross out 1 . . . 1 is* not *prime (by definition).*
3. *The first number* not *crossed out is 2 . . . it must be* prime; *put* a box *around it and cross out* all *multiples of 2 in the list . . . i.e. cross out 4, 6, 8, ...*
4. *Go back to the start of the list and* box *the first number that is* not *crossed out or boxed . . . it must be* prime; *and cross out* all *multiples of that number in the list. (*Note. *Some multiples may already have been crossed out.)*
5. *Repeat* Step 4. *until every number in the list is either* boxed *or crossed out.*

At the termination of the algorithm, the list of all *primes p such that $p \leqslant N$ are the numbers that are* boxed.

* *Eratosthenes* is pronounced: error-toss-the-knees.


Eratosthenes (c. 276 BC–194 BC) was a Greek mathematician, historian, astronomer, poet and geographer. Born at Cyrene in northern Africa he lived much of his life in Alexandria where he was the chief librarian. (At the time, Alexandria was famous for its library.) Eratosthenes was also famous for estimating the circumference of the earth using elementary *trigonometry* (i.e. *geometry*) and the lengths of shadows in two different places (measured at the same time of day.)

Example 6.2.6. *Let's use the* Sieve of Eratosthenes *to find all the primes less than or equal to 30. Below is the list of numbers from 1 to 30, after the method has been applied.*

~~1~~	[2]	[3]	~~4~~	[5]	~~6~~	[7]	~~8~~	~~9~~	~~10~~
[11]	~~12~~	[13]	~~14~~	~~15~~	~~16~~	[17]	~~18~~	[19]	~~20~~
~~21~~	~~22~~	[23]	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	[29]	~~30~~

The following are the steps required to obtain this.

1. *List* natural numbers *from 1* upto 30.
2. *Cross out 1.*
3. *The first number* not *crossed out is 2;* box *it and cross out 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 (*all *multiples of 2 – other than 2 itself – in the list).*
4. *The first number* not *crossed out or boxed is now 3;* box *it and cross out 9, 15, 21, 27 (*all *multiples of 3 – other than 3 itself – in the list; 6, 12, 18, 24, 30 are also multiples of 3 but have already been crossed out).*
5. *The first number* not *crossed out or boxed is now 5;* box *it and cross out 25 (the only multiple of 5 left that hasn't already been crossed out or boxed; 10, 15, 20, 30 are also multiples of 5 but have already been crossed out).*
   On further repeats of Step 4. *we box 7, 11, 13, 17, 19, 23, 29 . . . it turns out that on each of these occasions there are no multiples left to cross out.*

So the list of primes less than or equal to 30 is: $2, 3, 5, 7, 11, 13, 17, 19, 23, 29$.

Observations.
1. *On each execution of Step 4. (or Step 3.), the first prime multiple crossed out is the square of the prime just boxed, since all smaller multiple of that prime were already crossed out earlier.*
2. *When performing the Sieve of Eratosthenes in Example 6.2.6 we observed in Step 5. that when we came to box 7, no further multiples of anything were left to cross out, and the reason for this is that* $7 > \sqrt{30}$.
3. *The Sieve of Eratosthenes is impractical for determining the primality of an integer* N*, especially as* N *gets larger. On the other hand it is relatively easy to determine divisibility by any given number.*

Example 6.2.7. *97 is prime, since*

- $97 < 100 = 10^2$;
- $2, 3, 5, 7$ *are the primes less than* $\sqrt{97} < 10$ *(we don't need to find square-roots* exactly!*);*
- *and none of* $2, 3, 5, 7$ *divides* 97*. Here, we use divisibility rules, or determine a remainder:*
  - $2 \nmid 7$ *(7 is last digit of* 97*)* $\implies 2 \nmid 97$.
  - $3 \nmid 16 = S(97) \implies 3 \nmid 97$ *(*$S(N)$ = *"the sum of digits of* N*", see Lemma 7.2.6).*
  - $5 \nmid 7$ *(7 is last digit of* 97*)* $\implies 5 \nmid 97$.
  - $97 = 7 \cdot 13 + 6 \implies 7 \nmid 97$.

Exercise Set 6.

1. Determine *simple* rules for divisibility by each of the following natural numbers:

   (i) 2 (iv) 5 (vii) 9 (x) 12
   (ii) 3 (v) 6 (viii) 10 (xi) 15
   (iii) 4 (vi) 8 (ix) 11

   Note: there is a rule for 7, but it's complicated and it is not much better than straight division.

2. The number $739ABC$ is divisible by 7, 8 and 9. What values can A, B and C take?

3. Show that $x^2 - y^2 = 2$ has no integer solutions.

4. Prove that for every integer n:

   (i) $3 \mid n^3 - n$; (iii) $30 \mid n^5 - n$; (v) $4 \nmid n^2 + 2$;
   (ii) $6 \mid n(n-1)(2n-1)$; (iv) $120 \mid n^5 - 5n^3 + 4n$; (vi) $121 \nmid n^2 + 3n + 5$.

5. Prove that for all integers a and b: 3 divides $(a+b)^3 - a^3 - b^3$.

6. Is 167 prime?

7. Show that if $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ is the prime decomposition of the positive integer n, then the number of positive divisors of n (including 1 and n) is $(e_1+1)(e_2+1)\cdots(e_k+1)$.

8. Which positive integers have exactly three positive divisors?

9. Which positive integers have exactly four positive divisors?

10. Show that a natural number n is an exact square if and only if it has an odd number of positive divisors.

∗11. There are 50 prisoners in a row of locked cells. With the return of the King from the Crusades, a partial amnesty is declared and it works like this. When the prisoners are still asleep, the jailer walks past the cells 50 times, each time walking from left to right. On the first pass, he turns the lock in every cell (so that every cell is now open). On the second pass he turns the lock on every second cell (meaning that these cells are now locked again). On the third pass, he turns the lock on every third cell, and so on. In general, on the kth pass, he turns the lock on every kth cell. The question is: which cells are unlocked at the end of the process so that the prisoner is free to go?

12. Is the following statement true or false? *The number $n^2 + n + 41$ is prime for all positive integers n.*

13. Is the list of prime numbers *finite?* i.e. is there a *largest* prime number?

14. Suppose p is prime.

    (i) Show that if $p \mid a^3$ then $p \mid a$.
    (ii) Show that if $p \mid b$ and $p \mid a^2 + b^2$ then $p \mid a$.

15. Obtain a complete list of primes less than 1000.
[*Hint.* There are 168 of them!]

Answer. Using the *Sieve of Eratosthenes,* the primes less than 1000 are:

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, and 997.

If you avoided this problem because you thought it would take too long, note that $32^2 > 1000$; so ... once you have boxed 31 (the 11^{th} prime) all remaining numbers not crossed out must be prime. (So you only need to run through the algorithm 11 times.)

CHAPTER 7

Number Theory – divisors and multiples

7.1 Greatest common divisor

Definition 7.1.1. The **greatest common divisor** (or **highest common factor**) of two integers a, b, written $\gcd(a, b)$ or $\mathrm{hcf}(a, b)$ or just (a, b) is the largest natural number that divides both a and b. For example, $(10, 16) = 2$.

Observe that, if d is any *common divisor* of a and b then $d \mid a - bm$, for any integer m. Conversely, any *common divisor* of $a-bm$ and b, where $m \in \mathbb{Z}$, is a divisor of $a = (a-bm)+mb$. So the set of common divisors of a and b is also the set of common divisors of $a - bm$ and b, for any $m \in \mathbb{Z}$. In particular, this means that $(a, b) = (a - bm, b)$ for any integer m. This fact is the central idea behind the Euclidean Algorithm (Algorithm 7.1.3).

Definition 7.1.2. The **lowest common multiple** (or **least common multiple**) of a and b, written $\mathrm{lcm}(a, b)$, is the least natural number m such that $a \mid m$ and $b \mid m$.

It can be shown that

$$\gcd(a, b)\,\mathrm{lcm}(a, b) = |ab|.$$

Below we introduce an efficient way of calculating the gcd of pairs of integers, that remains useful even when the integers are very large. The method we'll use is an old one.

The Euclidean Algorithm

Algorithm 7.1.3 (Euclidean Algorithm). *To find* $\gcd(a, b)$ *where* a *and* b *are positive integers. We'll assume that* $a \geqslant b$*. The algorithm proceeds by finding pairs of integers* (q_i, r_i) *as follows:*

$$\begin{aligned}
a &= q_1 b + r_1, && 0 \leqslant r_1 < b \\
b &= q_2 r_1 + r_2, && 0 \leqslant r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3, && 0 \leqslant r_3 < r_2 \\
&\vdots \\
r_k &= q_{k+2} r_{k+1} + r_{k+2}, && 0 \leqslant r_{k+2} < r_{k+1} \\
&\vdots
\end{aligned}$$

The algorithm continues until $r_{k+2} = 0$*, so that we finish with*

$$r_{n-1} = q_{n+1} r_n,$$

with $r_{n+1} = 0$*, for some positive integer* n*.*
We will see later that being able to trace the algorithm backwards is just as valuable as obtaining $\gcd(a, b)$. In doing so, it will be useful to assign the labels:

$$a = r_{-1}, \quad b = r_0.$$

Theorem 7.1.4. *When the Euclidean Algorithm terminates we have*

$$r_n = (a, b).$$

*The assumption $a \geqslant b$ is actually unnecessary. If $a < b$ one obtains $q_1 = 0, r_1 = a$, so that the first iteration effectively swaps a and b. Thus a computer program does not need to (and for efficiency, should not) check whether or not $a \geqslant b$.

Proof. This follows from the observation we made after Definition 7.1.1, namely that $(a, b) = (a - bm)$ for any $m \in \mathbb{Z}$. At step $k + 2$, we have

$$r_{k+2} = r_k - q_{k+2}r_{k+1},$$

so that we have

$$\begin{aligned}(a, b) = (r_{-1}, r_0) &= (r_1, r_0) = (r_0, r_1)\\ &= (r_1, r_2)\\ &\vdots\\ &= (r_k, r_{k+1})\\ &\vdots\\ &= (r_{n-1}, r_n)\\ &= (r_n, r_{n+1}) = (r_n, 0) = r_n.\end{aligned}$$

Hence the last non-zero remainder r_n, of the Euclidean Algorithm, is (a, b), as claimed. □

Example 7.1.5. *Find the greatest common divisor of 1547 and 560.*
We set $a = 1547, b = 560$. Then

$$\begin{aligned}1547 &= 2 \times 560 + 427\\ 560 &= 1 \times 427 + 133\\ 427 &= 3 \times 133 + 28\\ 133 &= 4 \times 28 + 21\\ 28 &= 1 \times 21 + 7\\ 21 &= 3 \times 7 + 0\end{aligned}$$

So $(1547, 560) = 7$.

A convenient way to execute the algorithm that helps one keep track of things is via a **division table**. We demonstrate this technique below, redoing the above example:

	1547	560	
2	1120	427	1
	427	133	
3	399	112	4
	28	21	
1	21	21	3
	7	0	

Remark 7.1.6. Observe that for each i we have $r_{i+2} < r_i$. So the algorithm finishes fairly quickly.

Theorem 7.1.7 (Bézout's Lemma). *If $(a, b) = d$ then there exist integers x and y such that*

$$d = xa + yb.$$

Proof. Tracing the Euclidean Algorithm backwards, we have that r_n is a linear combination of r_{n-1} and r_{n-2}:

$$r_n = r_{n-2} - q_n r_{n-1}.$$

Substituting for r_{n-1}, using the previous step of the Euclidean Algorithm, one has r_n as a linear combination of r_{n-2} and r_{n-3}. Continuing in this way, one eventually has r_n as a linear combination of $r_{-1} = a$ and $r_0 = b$. (See the final remark of the description of Algorithm 7.1.3.)

We demonstrate the procedure, using the pair $a = 1547, b = 560$ from the previous example:

$$\begin{aligned}
7 &= 28 - 1 \times 21 \\
&= 28 - 1(133 - 4 \times 28) \\
&= 5 \times 28 - 133 \\
&= 5 \times (427 - 3 \times 133) - 133 \\
&= 5 \times 427 - 16 \times 133 \\
&= 5 \times 427 - 16 \times (560 - 427) \\
&= 21 \times 427 - 16 \times 560 \\
&= 21 \times (1547 - 2 \times 560) - 16 \times 560 \\
&= 21 \times 1547 - 58 \times 560.
\end{aligned}$$

So $(1547, 560) = 1547x + 560y$ with $x = 21$ and $y = -58$. □

Remark 7.1.8. In Theorem 7.1.7, the pair of integers (x, y) is not unique, for suppose (x_0, y_0) satisfies

$$d = ax + by, \tag{7.1.1}$$

then

$$\begin{aligned}
d &= ax_0 + by_0 \\
&= ax_0 + \lambda ab/d + by_0 - \lambda ab/d \\
&= a(x_0 + \lambda b/d) + b(y_0 - \lambda a/d).
\end{aligned}$$

Hence

$$x = x_0 + \lambda b/d, \quad y = y_0 - \lambda a/d, \tag{7.1.2}$$

where $\lambda \in \mathbb{Z}$, is a more general solution of (7.1.1). In fact, when $d = (a, b)$ one can show (7.1.2) is the most general solution of (7.1.1).

Definition 7.1.9. An equation of form

$$ax + by = c, \tag{7.1.3}$$

where x, y are unknown integers is a **linear Diophantine equation**.
Exercise. Show that (7.1.3) has solutions if and only if $\gcd(a, b) \mid c$.

Definition 7.1.10. If $(a, b) = 1$ then a and b are **relatively prime** (or **coprime**).

Corollary 7.1.11. *Integers a and b are relatively prime if and only if $\exists x, y \in \mathbb{Z}$ such that*

$$ax + by = 1.$$

Proof. The "only if" part is immediate from Theorem 7.1.7. In the other direction, suppose that $(a, b) = d$ and for a contradiction suppose that $d > 1$. Then since $d \mid ax + by$ we cannot have $ax + by = 1 \not\downarrow$, and hence, in fact $d = 1$. □

Extended Euclidean Algorithm

By reorganising our **division table**, and doing a few extra calculations on the forward trace of the Euclidean Algorithm, we can obtain coefficients x, y such that $d = (a, b)$ and

$$ax + by = d.$$

The table is organised with the following column headings:

i	x_i	y_i	r_i	q_i

Defining $a = r_{-1}$, $b = r_0$ and q_i and r_i as before, so that we have

$$r_i = r_{i-2} - q_i r_{i-1}, \text{ for } i \geqslant 1,$$

we define the x_i and y_i to satisfy the same relation as r_i, i.e.

$$x_i = x_{i-2} - q_i x_{i-1}$$
$$y_i = y_{i-2} - q_i y_{i-1}$$

for $i \geqslant 1$. As with the r_i, in order to be able to start, we must define x_{-1}, x_0, y_{-1}, y_0. We define these in such a way (and we show this below) that

$$ax_i + by_i = r_i, \text{ for } -1 \leqslant i \leqslant n,$$

so that when $i = n$ we have

$$ax_n + by_n = r_n = d,$$

i.e. the final line of the table has coefficients $x = x_n$, $y = y_n$ satisfying

$$ax + by = d.$$

Magic! The starting values for x_i and y_i are

$$x_{-1} = 1, \quad x_0 = 0, \quad y_{-1} = 0, \quad y_0 = 1.$$

One can sometimes save a step by allowing *negative* remainders. Before proving the algorithm works, let us demonstrate using our earlier example:

i	x_i	y_i	r_i	q_i	Comments
-1	1	0	1547		
0	0	1	560		
1	1	-2	427	2	$1547 - 2 \cdot 560 = 427$, $1 - 2 \cdot 0 = 1$, $0 - 2 \cdot 1 = -2$
2	-1	3	133	1	$560 - 1 \cdot 427 = 133$, $0 - 1 \cdot 1 = -1$, $1 - 1 \cdot -2 = 3$
3	4	-11	28	3	$427 - 3 \cdot 133 = 28$, $1 - 3 \cdot -1 = 4$, $-2 - 3 \cdot 3 = -11$
4	-21	58	-7	5	$133 - 5 \cdot 28 = -7$, $-1 - 5 \cdot 4 = -21$, $3 - 5 \cdot -11 = 58$

Observe that we may stop since $-7 \mid 28$. Thus from the last line of the table we have

$$-21 \cdot 1547 + 58 \cdot 560 = -7$$
$$\therefore 21 \cdot 1547 - 58 \cdot 560 = 7.$$

Proof (of Extended Euclidean Algorithm). We need to prove the claim that

$$ax_i + by_i = r_i, \text{ for } -1 \leqslant i \leqslant n,$$

given that each of x_i, y_i, r_i satisfy the recurrence

$$u_i = u_{i-2} - q_i u_{i-1}, \text{ for } i \geqslant 1,$$

and for $i = -1, 0$,

$$x_{-1} = 1, \quad x_0 = 0, \quad y_{-1} = 0, \quad y_0 = 1, \quad r_{-1} = a, \quad r_0 = b.$$

Thus, define

$$P(i) : r_i = ax_i + by_i.$$

We will prove $P(i)$ for $-1 \leqslant i \leqslant n$ by a (finite) induction by proving each of $P(-1)$, $P(0)$, and $P(k)$ and $P(k+1) \implies P(k+2)$ for general k.

For $P(-1)$, we have

$$\begin{aligned} \text{LHS of } P(-1) = r_{-1} &= a \\ &= a \cdot 1 + b \cdot 0 \\ &= a \cdot x_{-1} + b \cdot y_{-1} = \text{RHS of } P(-1) \end{aligned}$$

So $P(-1)$ holds.

For $P(0)$, we have

$$\begin{aligned} \text{LHS of } P(0) = r_0 &= b \\ &= a \cdot 0 + b \cdot 1 \\ &= a \cdot x_0 + b \cdot y_0 = \text{RHS of } P(0) \end{aligned}$$

So $P(0)$ holds.

Now we show $P(k)$ and $P(k+1) \implies P(k+2)$, for general k.
So assume $P(k)$ and $P(k+1)$, i.e.

$$\begin{aligned} r_k &= ax_k + by_k \\ r_{k+1} &= ax_{k+1} + by_{k+1} \end{aligned}$$

and consider $P(k+2)$:

$$\begin{aligned} \text{LHS of } P(k+2) &= r_{k+2} \\ &= r_k - q_{k+2} r_{k+1}, && \text{by the given recurrence} \\ &= ax_k + by_k - q_{k+2}(ax_{k+1} + by_{k+1}), && \text{by the inductive assumption} \\ &= a(x_k - q_{k+2}x_{k+1}) + b(y_k - q_{k+2}y_{k+1}) \\ &= ax_{k+2} + by_{k+2}, && \text{by the given recurrence} \\ &= \text{RHS of } P(k+2) \end{aligned}$$

So we have shown that $P(k)$ and $P(k+1) \implies P(k+2)$.

So the induction is complete and the claim

$$ax_i + by_i = r_i, \text{ for } -1 \leqslant i \leqslant n,$$

is proved. □

7.2 Congruence modulo m

Definition 7.2.1. Let $m \in \mathbb{N}$. Then we say,

$$\begin{aligned} & a \text{ is } \textbf{congruent} \text{ to } b \textbf{ modulo } m \\ \text{written: } & a \equiv b \pmod{m} \\ \iff & m \,|\, a - b \\ \iff & a - b = qm \text{ for some integer } q \\ \iff & a = b + qm \text{ for some integer } q \\ \iff & a = b + \text{“some multiple of } m\text{”}. \end{aligned}$$

Properties 7.2.2 (Properties of Congruence modulo m).

(i) $a \equiv a \pmod{m}$ *for all* $a \in \mathbb{Z}$ ***(reflexivity)***

(ii) $a \equiv b \pmod{m} \implies b \equiv a \pmod{m}$ ***(symmetry)***

(iii) $a \equiv b \pmod{m}$ *and* $b \equiv c \pmod{m} \implies a \equiv c \pmod{m}$ ***(transitivity)***

(iv) If

$$\begin{aligned} a &\equiv b \pmod{m} \text{ and} \\ c &\equiv d \pmod{m} \end{aligned}$$

then

$$\begin{aligned} (1) \qquad a + c &\equiv b + d \pmod{m}, \\ (2) \qquad ac &\equiv bd \pmod{m}, \\ (3) \qquad a^n &\equiv b^n \pmod{m} \text{ for all } n \in \mathbb{N}. \end{aligned}$$

Proof. Properties (i)–(iii) prove *congruence modulo* m is an equivalence relation on $\mathbb{Z}$.
(i) holds, since $m \,|\, 0 = a - a$.
(ii) holds, since $m \,|\, (a - b) \implies m \,|\, (b - a)$.
(iii) holds, since $m \,|\, (a - b)$ and $m \,|\, (b - c) \implies m \,|\, (a - c) = (a - b) + (b - c)$.
Now, for the parts of (iv), assume $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$. Then for some $k, \ell \in \mathbb{Z}$,

$$a = b + km \qquad (\S)$$
$$c = d + \ell m \qquad (\dagger)$$

So,

$$\begin{aligned} (\S) + (\dagger) : a + c &= b + d + (k + \ell)m \\ &\equiv b + d \pmod{m} \end{aligned}$$

Hence, (1) holds. Also,

$$\begin{aligned} (\S) \cdot (\dagger) : ac &= (b + km)(d + \ell m) \\ &= bd + (kd + b\ell + k\ell m)m \\ &\equiv bd \pmod{m} \end{aligned}$$

So, (2) holds.
Finally, (3) follows from (2) by induction. □

Definition 7.2.3. If $a \in \mathbb{Z}$ and $m \in \mathbb{N}$, the binary operations **div** and **mod** are defined to return the *quotient* and *remainder* when a is divided by m according to the *Division Algorithm*, i.e. if $q, r \in \mathbb{Z}$ with $0 \leqslant r < m$ and

$$a = mq + r,$$

then

$$\begin{aligned} a \operatorname{div} m &= q, \text{ and} \\ a \bmod m &= r. \end{aligned}$$

Essentially, div does "integer division" of a by m. (Sometimes **quo** is used in place of div.) Most computing languages define some sort of mod operation, extended to allow negative m. The key property of interest to us, is that,

If $a \bmod m = r$ then $a \equiv r \pmod{m}$.

Note 7.2.4. Divisibility statements can be written in terms of congruences. In particular,

$$\begin{aligned} b \vdots m &\iff m \,|\, b \\ &\iff b \equiv 0 \pmod{m}. \end{aligned}$$

Notation 7.2.5. Suppose $N \in \mathbb{N}$ has decimal digit expansion

$$N = 10^n a_n + 10^{n-1} a_{n-1} + \cdots + 10a_1 + a_0 = \sum_{k=0}^{n} 10^k a_k,$$

where $a_k \in \{0, 1, \ldots, 9\}$ for $0 \leqslant k \leqslant n$. Then:

(i) $N = \overline{a_n a_{n-1} \ldots a_1 a_0}$ is called the **decimal representation** of N;

(ii) $S(N) = \sum_{k=0}^{n} a_k$ is the **sum of digits of** N; and

(iii) $A(N) = \sum_{k=0}^{n} (-1)^k a_k$ is the **alternating sum of digits of** N.

Lemma 7.2.6. *$N \equiv S(N) \pmod{m}$ where $m \in \{3, 9\}$.*

Proof. Let $\overline{a_n a_{n-1} \ldots a_1 a_0}$ be the decimal representation of N, and observe that 10 is congruent to 1 modulo each of 3 or 9. Then

$$\begin{aligned} N &= \overline{a_n a_{n-1} \ldots a_1 a_0} \\ &= \sum_{k=0}^{n} 10^k a_k \\ &\equiv \sum_{k=0}^{n} 1^k a_k \pmod{3} \\ &\equiv \sum_{k=0}^{n} a_k \pmod{3}. \end{aligned}$$

But $\sum_{k=0}^{n} a_k = S(N)$. The argument is identical if '(mod 3)' is replaced by '(mod 9)'. □

Corollary 7.2.7. *(i)* $3 \,|\, N \iff 3 \,|\, S(N)$. *(ii)* $9 \,|\, N \iff 9 \,|\, S(N)$.

Proof. This is the special case of Lemma 7.2.6 when $N \equiv 0 \pmod{m}$, for each value of m, written in terms of divisibility. □

Lemma 7.2.8. $N \equiv A(N) \pmod{11}$.

Corollary 7.2.9. $11 \,|\, N \iff 11 \,|\, A(N)$.

Remark 7.2.10. One usually describes $A(N)$ as the '*difference of the sums of even-place and odd-place digits*'. Lemma 7.2.8 is proved by observing that $10 \equiv -1 \pmod{11}$, deducing $N \equiv A(N) \pmod{11}$ analogously to Lemma 7.2.6; then the corollary is the special case of Lemma 7.2.8 when $N \equiv 0 \pmod{11}$ written in terms of divisibility.

Lemma 7.2.11. *If* $ac \equiv bc \pmod{m}$ *and* $(c, m) = 1$ *then* $a \equiv b \pmod{m}$.

Proof. First, assume $(c, m) = 1$. By Corollary 7.1.11,

$$\begin{aligned} & cx + my = 1, \quad \text{for some } xy \in \mathbb{Z} \\ \implies \quad & cx \equiv 1 \pmod{m}. \end{aligned}$$

So now,

$$\begin{aligned} & ac \equiv bc \pmod{m} \\ \implies \quad & acx \equiv bcx \pmod{m} \\ \implies \quad & a \equiv b \pmod{m}. \end{aligned}$$

□

Corollary 7.2.12. *If* $ac \equiv bc \pmod{m}$ *and* $(c, m) = d$ *then* $a \equiv b \pmod{m/d}$.

Lemma 7.2.13. *If* $m_1, m_2 \in \mathbb{N}$ *such that* $(m_1, m_2) = 1$ *and*

$$\begin{aligned} a &\equiv b \pmod{m_1}, \\ a &\equiv b \pmod{m_2} \end{aligned}$$

then

$$a \equiv b \pmod{m_1 m_2}.$$

Theorem 7.2.14 (Chinese Remainder Theorem). *If* $m_1, m_2, \ldots, m_k \in \mathbb{N}$ *are pairwise coprime, i.e.* $(m_i, m_j) = 1$ *for* $i \neq j$, *then the* k *simultaneous congruences*

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \\ &\vdots \\ x &\equiv a_k \pmod{m_k} \end{aligned}$$

have a unique solution modulo $M = \prod_{i=1}^{k} m_i$.
Moreover, the solution may be constructed as follows.

- *Define* $M_i := \prod_{j \neq i} m_j = M/m_i$, *for* $i = 1, 2, \ldots, k$.
- *Define* b_i *to be the multiplicative inverse of* M_i *modulo* m_i, *for each* i, *i.e.*

  $$b_i M_i \equiv 1 \pmod{m_i}, \text{ for } i = 1, 2, \ldots, k,$$

  noting that: $b_i M_i \equiv 1 \pmod{m_i} \iff b_i M_i + q m_i = 1$ *for some integer* $q \in \mathbb{Z}$, *so that* b_i *(and* q*) can be constructed via the Euclidean Algorithm.*
- *The solution is now given by:* $x \equiv \sum_{j=1}^{k} a_j b_j M_j \pmod{M}$.

Proof. The solution above means that, for some $q \in \mathbb{Z}$,

$$x = \sum_{j=1}^{k} a_j b_j M_j + qM.$$

Consider this expression modulo m_i. Since $m_i \mid M_j$ when $j \neq i$, $b_i M_i \equiv 1 \pmod{m_i}$, and $m_i \mid M$ for all i, we have

$$\begin{aligned} x &= \sum_{j=1}^{k} a_j b_j M_j + qM \\ &= \sum_{j \neq i} a_j b_j M_j + a_i b_i M_i + qM \\ &\equiv \sum_{j \neq i} a_j b_j \cdot 0 + a_i \cdot 1 + q \cdot 0 \pmod{m_i} \\ &\equiv a_i \pmod{m_i} \end{aligned}$$

i.e. the expression for x reduces to each of the given congruences. Hence the constructed expression for x satisfies all the given congruences. □

Euler's Totient Function

Definition 7.2.15. For a given natural number n, let the set $\mathcal{S}_n$ of natural numbers $k < n$ that are coprime to n, i.e.

$$\mathcal{S}_n = \{k \in \mathbb{N} \mid k < n \text{ and } (k, n) = 1\}.$$

Then **Euler's totient function** $\varphi(n)$ is defined to be the *cardinality* of $\mathcal{S}_n$.

Example 7.2.16. *For $n = 12$, the set of natural numbers less than n is*

$$\{1, 2, \ldots, 11\}.$$

Eliminating all the elements that are divisible by 2 or 3 (the prime divisors of 12), we have

$$\mathcal{S}_{12} = \{1, 5, 7, 11\},$$

and so

$$\varphi(12) = |\,\mathcal{S}_{12}\,| = 4.$$

In Definition 14.1.1 we define an *abelian group*. In Number Theory, we have some key examples of *abelian groups*.

Example 7.2.17. *$(\mathbb{Z}, +)$ is an abelian group, i.e. the set $\mathbb{Z}$ of integers, with ordinary addition $+$ as the binary operation, is an abelian group.*
Below, we do a "run-through check" that the axioms hold.

G1: $\forall m, n \in \mathbb{Z}$, we have $m + n \in \mathbb{Z}$.

G2: $\forall \ell, m, n \in \mathbb{Z}$, we have $\ell + (m + n) = (\ell + m) + n$.

G3: The identity is 0, since $\forall m \in \mathbb{Z}, 0 + m = m + 0 = m$.

G4: The inverse of each $m \in \mathbb{Z}$ is $-m$ since $m + (-m) = (-m) + m = 0$.

G5: $\forall m, n \in \mathbb{Z}$, we have $m + n = n + m$.

Example 7.2.18. $(\mathcal{S}_n, \cdot)$ *where* $\cdot$ *is multiplication* modulo n *is an abelian group.*

When $n = p$ prime, we get the following special case of Example 7.2.18 is the following, which we need later for Theorem 7.2.33.

Example 7.2.19. *The set* $\mathbb{Z}_p \setminus \{0\} = \{1, 2, \ldots, p-1\}$ *is an abelian group under multiplication modulo* p*, where* p *is a prime.*

Definition 7.2.20. The **order of a group** G is the number of elements it contains, and is denoted by $|G|$ (essentially, $|G|$ is the cardinality of G when regarded as a set).
The **order of an element** x of a group G is the least number of times that it can be composed with itself to obtain the identity. The order of x is similarly denoted by $|x|$. If the group operation is $\cdot$, in which case the identity is usually represented by 1, then $|x|$ is the least $m \in \mathbb{N}$ for which $x^m = 1$ in G.

Theorem 7.2.21. *If* $(G, \cdot)$ *is a finite group and* $x \in G$ *then* $|x| \,|\, |G|$.

Corollary 7.2.22. *If* $(G, \cdot)$ *is a finite group and* $x \in G$ *then* $x^{|G|} = 1$.

One of the main applications of Euler's totient function is the following generalisation of a theorem we have yet to meet: Fermat's Little Theorem.

Theorem 7.2.23 (Euler's Theorem). *If* $a \in \mathbb{Z}$, $n \in \mathbb{N}$ *and* $(a, n) = 1$, *then*

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Proof. Using the information in the dangerous bend, $\mathcal{S}_n$ is an abelian group under multiplication *modulo* n, and so Corollary 7.2.22 gives

$$a^{|\,\mathcal{S}_n\,|} \equiv 1 \pmod{n},$$

and the result follows. □

Example 7.2.24. *We have* $(3, 8) = 1$ *and since* $\mathcal{S}_8 = \{1, 3, 5, 7\}$ *is the set of natural numbers less than* 8 *that are coprime to* 8*, we have* $\varphi(8) = |\,\mathcal{S}_8\,| = 4$.
Thus Euler's Theorem tells us $3^4 \equiv 1 \pmod 8$, and indeed we have: $3^4 = 81 \equiv 1 \pmod 8$.

We now obtain a formula for $\varphi(m)$, which depends on two essential properties:

(i) $\varphi(p^k) = p^k - p^{k-1}$, if $k \in \mathbb{N}$ and p is prime;

(ii) φ is a *multiplicative* function (we define what this means below, but we won't bother to prove it).

Definition 7.2.25. A function $f : \mathbb{N} \to \mathbb{N}$ is said to be **multiplicative** if for $a, b \in \mathbb{N}$,

$$(a, b) = 1 \implies f(ab) = f(a)f(b).$$

Theorem 7.2.26. φ *is multiplicative, i.e. if* $m_1, m_2 \in \mathbb{N}$ *then*

$$\varphi(m_1 m_2) = \varphi(m_1)\varphi(m_2).$$

Example 7.2.27. *Let's find $\varphi(12)$ using the two properties:*

$$\begin{aligned}\varphi(12) &= \varphi(3)\varphi(4), && \text{by property (ii), since } (3,4)=1\\ &= (3-3^0)(2^2-2^1), && \text{by property (i), twice}\\ &= 2\cdot 2 = 4\end{aligned}$$

as we observed before.

The elements of $\mathcal{S}_p = \{1, 2, \ldots, p-1\}$ are all coprime to p, if p is prime. So it follows immediately that for a prime p, $\varphi(p) = p-1$. We can generalise this idea to get a proof of property (i).

Theorem 7.2.28. *If $n = p^k$ where p is a prime and $k \in \mathbb{N}$ then*

$$\varphi(n) = p^k - p^{k-1} = p^k\Big(\frac{p-1}{p}\Big).$$

Proof. Observe that for a number in $\{1, 2, \ldots, p^k\}$ not to be coprime to $n = p^k$, it must have p as a divisor. So the set $\mathcal{S}_{p^k}$ of natural numbers less than p that are coprime to $n = p^k$ can be written as

$$\mathcal{S}_{p^k} = \{1, 2, \ldots, p^k\} \setminus \big\{pm \mid m \in \{1, 2, \ldots, p^{k-1}\}\big\}.$$
$$\therefore \varphi(n) = |\,\mathcal{S}_{p^k}\,| = p^k - p^{k-1}.$$ □

Theorem 7.2.29. *If N has prime factorisation*

$$N = \prod_{i=1}^{n} p_i^{\varepsilon_i}$$

then

$$\varphi(N) = N\prod_{i=1}^{n}\frac{p_i-1}{p_i}.$$

Proof. This follows by induction on n. □

Example 7.2.30. $\varphi(180) = \varphi(2^2 3^2 5) = 180\Big(\frac{2-1}{2}\Big)\Big(\frac{3-1}{3}\Big)\Big(\frac{5-1}{5}\Big) = 48.$

The following theorem follows from Euler's Theorem as a special case since $\varphi(p) = p-1$, if p is prime, but we give an independent proof.

Theorem 7.2.31 (Fermat's Little Theorem). *If $n \in \mathbb{N}$, p is a prime and $p \nmid n$ then*

$$n^{p-1} \equiv 1 \pmod{p}.$$

Proof. Suppose $r \in \mathbb{Z}$ such that $0 < r < p$. Then $rn \equiv s \pmod{p}$ for some $s \in \mathbb{Z}$, also satisfying $0 < s < p$ (since $rn \equiv 0 \pmod{p}$ would imply $p \mid n$ contrary to assumption). Furthermore, for distinct values of r we obtain different values of s, since if

$$r_1 n \equiv s \pmod{p} \quad \text{and} \quad r_2 n \equiv s \pmod{p},$$

then $r_1 n \equiv r_2 n \pmod{p}$ whence $r_1 = r_2$ by Lemma 7.2.11. Hence

$$1n \cdot 2n \cdot 3n \cdots (p-1)n \equiv 1\cdot 2\cdot 3\cdots(p-1) \pmod{p}$$

which on cancellation (using Lemma 7.2.11) gives

$$n^{p-1} \equiv 1 \pmod{p}.$$ □

Corollary 7.2.32. *If $n \in \mathbb{N}$, p is a prime then $n^p \equiv n \pmod{p}$.*

Proof. If $p \nmid n$ then the result follows from Fermat's Little Theorem by multiplying both sides of the congruence by n. Otherwise $n \equiv 0 \pmod{p}$, in which case the result is trivially true. □

Theorem 7.2.33 (Wilson's Theorem). *Let $1 < n \in \mathbb{N}$. Then*

$$(n-1)! \equiv -1 \pmod{n} \iff n \text{ is prime.}$$

Proof. ($\Longrightarrow$) Suppose that $(n-1)! \equiv -1 \pmod{n}$ and let $d \in \mathbb{Z}$ such that $1 \leqslant d \,|\, n$. Then $d \in \{1, 2, \ldots, n-1\}$ and hence

$$d \,|\, (n-1)!.$$

By the congruence, we have $(n-1)! + 1 \equiv 0 \pmod{n}$. Hence, we also have

$$d \,|\, n \,|\, (n-1)! + 1.$$

Hence $d \,|\, 1 = \big((n-1)! + 1\big) - (n-1)!$, and so $d = 1$, which implies n has no **proper divisors**[†], so that n is prime.
($\Longleftarrow$) Suppose $n = p$ prime. Consider $G = \mathbb{Z}_p \setminus \{0\} = \{1, 2, \ldots, p-1\}$ as a group under multiplication modulo p.

Case 1: $p = 2$. Then

$$(p-1)! \equiv 1 \equiv -1 \pmod{2}.$$

Case 2: $p \geqslant 3$. Then p is odd. Each element $a \in G$ has an inverse b, so that we have $ab \equiv 1 \pmod{p}$. First consider the case, $a = b$.

$$\begin{aligned} \Longrightarrow \quad & a^2 \equiv 1 \pmod{p} \\ \Longrightarrow \quad & a^2 - 1 \equiv 0 \pmod{p} \\ \Longrightarrow \quad & (a-1)(a+1) \equiv 0 \pmod{p} \end{aligned}$$

So that we have $a \equiv \pm 1 \pmod{p} \implies a = 1$ or $a = p-1$. Otherwise we have $a \neq b$. Thus except for 1 and $p-1$, the elements of G can be collected in pairs whose product is 1, so that

$$\begin{aligned} (p-1)! &\equiv 1^{(p-3)/2} \cdot 1 \cdot (p-1) \pmod{p} \\ &\equiv (p-1) \pmod{p} \\ &\equiv -1 \pmod{p} \end{aligned}$$

So in all cases, $(n-1)! \equiv -1 \pmod{n}$. □

[†]**Definition.** If $d \in \mathbb{N}$ then d is a **proper divisor** of n if $d \,|\, n$ and $1 < d < n$.

Exercise Set 7.

1. For each of the following pairs of integers a, b use the *Euclidean Algorithm* to find $d = (a, b)$ and find a pair of integers x, y such that $ax + by = d$.

   (i) $a = 85$, $b = 41$; (ii) $a = 2613$, $b = 637$.

2. Show that if there exist integers x, y such that $ax + by = 1$ then $(a, b) = 1$.

3. Show that $(3k + 2, 5k + 3) = 1$ for any integer k.

4. Show that $(a, a + 2) = 2$ if a is even and $(a, a + 2) = 1$ otherwise.

5. Show that if $(a, b) = 1$ then $(a + b, a - b) = 1$ or 2.

6. Find all solutions to the following *Diophantine Equations*.

   (i) $2x + 5y = 11$. (ii) $12x + 18y = 50$. (iii) $202x + 74y = 7638$.

   Does equation (iii) have a solution in *positive* integers x, y?

7. A grocer orders apples and oranges at a total cost of \$8.39. If apples cost 25c each and oranges cost 18c each, how many of each type of fruit did the grocer order?

8. An apartment block has units at two rates: most rent at \$87/week, but a few rent at \$123/week. When all are rented the gross income is \$8733/week. How many units of each type are there?

∗9. When Jane is one year younger than Betty will be when Jane is half as old as Betty will be when Jane is twice as old as Betty is now, Betty will be three times as old as Jane was when Betty was as old as Jane is now.
One is in her teens and ages are in completed years. How old are they?

10. Solve the adjacent *alphametic* (an addition in which: each letter stands for a different digit; and left-most digits of a number are not allowed to be 0).

$$\begin{array}{cccc} & A & H & A \\ & A & H & A \\ & & & A \\ & W & A & G \\ \hline H & A & H & A \end{array}$$

   Answer. HAHA = 1717 ($W = 2$, $G = 6$). The solution is unique.

11. About all we know of Diophantus' life is his epitaph from which his age at death is to be deduced:

    > Diophantus spent one-sixth of his life in childhood, one-twelfth in youth, and another one-seventh in bachelorhood. A son was born five years after his marriage and died four years before his father at half his father's age.

12. Augustus de Morgan, a nineteenth-century mathematician, stated:

    > I was x years old in the year x^2.

    When was he born?

13. Prove that for every integer n:

    (i) $3 \mid n^3 - n$; (iii) $7 \mid n^7 - n$;
    (ii) $5 \mid n^5 - n$; (iv) $11 \mid n^{11} - n$.

    Show that $n^9 - n$ is not necessarily divisible by 9. *Hint:* Try $n = 2$.
    What general result is suggested by the above?

14. Prove that $3^{6n} - 2^{6n}$ is divisible by 35, for every positive integer n.

∗15. What is the final digit of $7^{7^{7^{7^{7^7}}}}$.

Exponentiation convention. Note that a^{b^c} means $a^{(b^c)}$. Mathematicians chose the "from the right" definition, since

$$(a^b)^c = a^{bc},$$

i.e. the "from the left" expression can be written without power towers and without brackets.

16. Prove that for any $n \in \mathbb{N}$, 17 divides $2^n.3^{2n} - 1$.

17. Prove that for any $n \in \mathbb{N}$, $17^n - 12^n - 24^n + 19^n$ is divisible by 35.

18. Using Fermat's Little Theorem, prove that for all positive integers a and b:

    (i) 3 divides $(a+b)^3 - a^3 - b^3$; (ii) 5 divides $(a+b)^5 - a^5 - b^5$.

∗19. Prove that $5^{99} + 11^{99} + 17^{99}$ is divisible by 33.

∗20. What is the final digit of $((((((((((7^7)^7)^7)^7)^7)^7)^7)^7)^7)^7)$? (7 occurs as a power 10 times.)

∗21. *(17th International Olympiad, 1975, Problem 4)* When 4444^{4444} is written in decimal notation, the sum of its digits is A. Let B be the sum of the digits of A. Find the sum of the digits of B.
Hints: First show that the sum of the digits of B is fairly small (in fact: less than 16). Then use the fact that, for any natural number N,

$$N \equiv (\text{sum of the digits of } N) \pmod 9.$$

22. Show $7 \mid 2222^{5555} + 5555^{2222}$.

23. For which a does the congruence $ax \equiv 1 \pmod m$ have a solution, when ...

    (i) $m = 4$? (ii) $m = 5$? (iii) $m = 6$? (iv) $m = 7$?

24. Solve $58x \equiv 1 \pmod{127}$.
    [*Hint.* Use the Euclidean Algorithm as one of your steps.]

25. Use Wilson's Theorem to show that $(p-1)! \equiv p - 1 \pmod{p(p-1)}$ if p is prime.

26. Find the remainder when 97! is divided by 101.

CHAPTER 8

Number Theory – cryptosystems

8.1 Review

Let's now look at expressing Property 1, Euclid's Lemma and Fermat's Little Theorem in terms of congruences.

Property 1. *If* $b \equiv 0 \pmod{a}$ *and* $c \equiv 0 \pmod{a}$ *then* $bm + cn \equiv 0 \pmod{a}$*, for all* $m, n \in \mathbb{Z}$*.*

Lemma 8.1.1 (Euclid's Lemma). *If* p *is prime and* $ab \equiv 0 \pmod{p}$ *then*

$$a \equiv 0 \pmod{p} \quad \text{or} \quad b \equiv 0 \pmod{p}.$$

Theorem 8.1.2 (Fermat's Little Theorem). *If* p *is prime and* $a \not\equiv 0 \pmod{p}$ *then* $a^{p-1} \equiv 1 \pmod{p}$.

Also, recall that two integers a, b are *coprime* if their greatest common divisor (a, b) is 1. We also saw the following lemma previously as two separate results: part (i) is Lemma 7.2.11, and (ii) is Corollary 7.2.12; in fact, (ii) generalises (i).

Lemma 8.1.3. *(i) If* $ac \equiv bc \pmod{m}$ *and* $(c, m) = 1$ *then* $a \equiv b \pmod{m}$*.*

(ii) If $ac \equiv bc \pmod{m}$ *and* $(c, m) = d$ *then* $a \equiv b \pmod{m/d}$*.*

8.2 Cryptosystems

A *code* is an algorithm used to reformulate a message in a convenient form for transmission, e.g. the *Morse code encodes* a message as dots and dashes, convenient for transmission along a wire. It was designed before it was discovered how to transmit voice messages. When one *encodes* a message, one is not trying to keep the message *secret*.

A *cryptosystem* is an algorithm used to *encrypt* a message to keep it *secret*. To describe some of these it will be useful to start with a numerical encoding of the alphabet and the blank space between words:

letter	space	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
encoding	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

One of the simplest cryptosystems is the *Caesar cipher* which replaces each letter (with encoding u) of a message by the letter with encoding v, where:

$$v \equiv au + b \pmod{m}$$

where a, b, m are fixed integers such that a and m are coprime and m is at least as large as the number of letters in your alphabet. For our alphabet above we would need m to be at least 27.

The cipher is named for Julius Caesar who used such a cryptosystem with $a = 1$ and $b = 3$. If we choose $m = 27$ with Julius Caesar's choice for a and b encrypting a message amounts to a *cyclic shift* of each letter right by 3 letters. This simple example is easy to break ... by determining the encryption of just one letter of the message, by trial-and-error (there are only 27 possibilities to check ... and by picking on the most frequently occurring letter of the encrypted message we might try *'e'* first, etc.).

Another simple cryptosystem is the *one-time pad*. The way this works is that both the receiver and sender have a long sequence of random numbers, $(b_1, b_2, \dots)$. If a message with the simple numerical encoding of the letters is:

$$u_1, u_2, \dots, u_i, \dots$$

then the i^{th} letter of the encoded message is encrypted as

$$u_i + b_i \pmod{m},$$

where m may be 27 if spaces are encoded or 26 if they are not encoded. Each sequence of random numbers is used just *once* which makes the cryptosystem *unbreakable*. The method is however extremely cumbersome, because both sender and receiver must keep a very long sequence of numbers.


A one-time pad is used for the hot-line between Washington and Moscow.

For frequent computer-based communication among several parties it is desirable to have a cryptosystem with neither of the faults of the above systems, i.e.

(i) the *encryption* and *decryption* algorithms are easy to compute and reusable; and

(ii) each person's *decryption* algorithm cannot be obtained from his/her *encryption* algorithm in any reasonable amount of time.

The second property means that the *encryption* algorithm can in fact be made public, and so such a system, together with the following property, is called a *public-key system*.

(iii) For a *message* a, *encryption* algorithm E and *decryption* algorithm D both

$$D\big(E(a)\big) = a \quad \text{and} \quad E\big(D(a)\big) = a.$$

8.2.1 The RSA Cryptosystem

The RSA system is an example of a *public-key system* that was developed in 1977 by Rivest, Shamir and Adleman. It is based on the following result that follows from Fermat's Little Theorem.

Theorem 8.2.1.1 (RSA Theorem). *Let p, q be distinct primes;*

$$\begin{aligned} &\text{let } n = pq, \\ &\text{let } k = (p-1)(q-1), \\ &\text{choose } d \text{ coprime to } k, \text{ and} \\ &\text{choose } e \text{ such that } de \equiv 1 \pmod{k}. \end{aligned}$$

Then $a^{ed} \equiv a \pmod{n}$ for any integer a.

Proof. Firstly, the hypotheses of the theorem can be satisfied, since, if one chooses $d = k - 1$ then $(d, k) = 1$ and so, the Euclidean Algorithm guarantees a solution of

$$dx + ky = 1,$$

whence, for $e = x$ we have

$$de \equiv 1 \pmod{k}.$$

By Fermat's Little Theorem, if $a \not\equiv 0 \pmod{p}$ then

$$a^{p-1} \equiv 1 \pmod{p}.$$

So, for any nonnegative integer ℓ,

$$a^{\ell(p-1)+1} \equiv a \pmod{p},$$

and this is also true if $a \equiv 0 \pmod{p}$. Since $de \equiv 1 \pmod{k}$, where $k = (p-1)(q-1)$, we have $ed = de = \ell(p-1) + 1$ for some integer ℓ. So

$$a^{ed} \equiv a \pmod{p},$$

and hence

$$a^{ed} \equiv a + m_1 q \pmod{pq}, \tag{8.2.1.1}$$

for some integer m_1. Similarly,

$$a^{ed} \equiv a \pmod{q},$$

and hence

$$a^{ed} \equiv a + m_2 p \pmod{pq}, \tag{8.2.1.2}$$

for some integer m_2. Since p and q are distinct primes, it follows from (8.2.1.1) and (8.2.1.2) that

$$a^{ed} \equiv a \pmod{pq}. \qquad \square$$

Here is an example to show how we use this result to come up with a *cryptosystem.*

Example 8.2.1.2. *Suppose our message is 'GO WEST'. Then we perform the following steps.*

1. *Encode the letters numerically, e.g. by the encoding given in the table on page 67. This gives: 07150023051920. Call this number a.*
2. *We need p, q such that $n = pq > a$.*
3. *To encrypt the message, compute: $E(a) = a^e \bmod n$.*
4. *To decrypt the message, the receiver computes: $D\big(E(a)\big) = (a^e \bmod n)^d \bmod n$ which by the RSA Theorem is congruent to $a \pmod{n}$, and since $n > a$ we know the message was a.*

In computing, the *binary operators* div and mod are defined as follows: if for integers a, m, with m positive, the *quotient* and *remainder* when a is divided by m are q and r respectively, i.e. $a = mq + r$, then

$$a \operatorname{div} m = q,$$
$$a \bmod m = r.$$

In particular, if $r = a \bmod m$ then $r \equiv a \pmod{m}$ and $0 \leqslant r < m$. Take careful notice of how the *computing* syntax for mod differs from the usual *mathematics* usage. We defined these operations in Definition 7.2.3.

Observe that in our example the message a was a rather large number and that n needed to be even larger. In practice, p and q are large primes (in 2023, RSA primes are 600–1200 digits each). Observe that choosing appropriate p, q, d determines n, k, e. The numbers e and n for the cryptosystem are publicly announced. The system is secure since determining the decryption algorithm is at least as difficult as factoring n.


With the technology of 1990, the expected time to factor a 200-digit number that is the product of two equal length primes, was approximately 4 million years. Ironically, 33 years later, it is estimated with 2023 technology, the same task would take 75 years.

Observe, also that Property (iii) of a *public-key system* is satisfied since

$$(a^d)^e = (a^e)^d \equiv a \pmod{pq}.$$


The significance of this is that a sender may use their *decryption* algorithm to encrypt a signature s as $s^d \bmod n$. Then, the receiver (and in fact, anybody) can use the publicly available *encryption* algorithm to decrypt it again, and so verify the source of the message.

Very long messages may still give a number larger than the simple encoding a, in which case one needs to break up the message into modules and encrypt each module separately. Here is one of Rivest, Shamir and Adleman's own examples.

Example 8.2.1.3. *Take the message: "IT'S ALL GREEK TO ME" and suppose* $n = 2773$, $d = 157$ *and* $e = 17$. *Since we can only encode numbers less than 2773, we choose blocks of length 2.*

1. *Numerical encoding of the blocks gives (where # denotes a blank space and other punctuation is ignored):*

I T	*S #*	*A L*	*L #*	*G R*
09 20	*19 00*	*01 12*	*12 00*	*07 18*
E E	*K #*	*T O*	*# M*	*E #*
05 05	*11 00*	*20 15*	*00 13*	*05 00*

2. *Observe each block is encoded with a number less than 2773.*

3. *Encrypting the first block we have:* $920^e = 920^{17} \equiv 948 \pmod{2773}$. *Encrypting all the blocks we get the following coded form of the message:*

09 48	*23 42*	*10 84*	*14 44*	*26 63*
23 93	*07 78*	*07 74*	*02 19*	*16 55*

4. *The receiver would now decrypt the message by applying the decryption algorithm to each block, e.g. for the first block*

$$948^d = 948^{157} \equiv 920 \pmod{2773}$$

which is the numerical encoding for the first original block: 'IT'.

Exercise Set 8.

1. Using the *Caesar cipher,* with a, b, m as defined in the dangerous bend on page 67 of the notes, encrypt: CRYPTOLOGY.

   Answer. With $a = 1$, $b = 3$ and $m = 27$, the *Caesar cipher* amounts to being a cyclic shift of each letter by three letters. Hence CRYPTOLOGY is encrypted as:
   FUASWRORJA

∗2. Decrypt the following message. Spaces are also encoded. There is one space in the encrypted output.

   RUOELTWK EINHXFEQHZEYTDJPEHVONERUOEBGCAEMHS

3. Find the values of p, q, k for Example 8.2.1.3 ($n = 2773$, $d = 157$ and $e = 17$).

4. Use $e = 3$ and $n = 2773$ to encrypt the following message using the RSA cryptosystem:

   CODING IS EASY

   Use 2-letter blocks and don't omit spaces.

∗5. Decrypt the following message. Spaces are also encoded. (It just so happens that no spaces appear after the encryption.)

   BKDAKUNFKDWTDBJKNWKFNANTTNLKWNTKBKIDS
   CKMCCUKYCMFCTJDYKDUJKBKHNSCKFNJDY

   Note that a *Caesar cipher* has been used (i.e. ⟨SPACE⟩, A, . . . , Z are encoded as $00, 01, \ldots, 26$ (as per the table on page 67), the *Caesar cipher* algorithm

   $$v \equiv au + b \pmod{27}$$

   has been applied for each letter u of the message for some a, b (which you essentially have to find), and the encoded letter v has been changed back to a letter using the table on page 67 again.)
   Note: Letters and spaces occurring in English text, arranged approximately in order of highest frequency to lowest frequency are

   ⟨SPACE⟩, E, T, A, I, O, N, S, H, R, D, L, U, . . .

   Also, use the fact that inter-word spaces occur on average every 4–5 letters and use what you know about the possibilities of letters in short words of 1, 2 or 3 letters.
   If this problem seems too hard, try doing it without using the fact that a *Caesar cipher* has been used.
   Hint. Since you want to *decrypt* you really want to express u in terms of v, i.e. you really want to find a c, d such that

   $$u \equiv cv + d \pmod{27}.$$

6. Use $e = 3$ and $n = 2773$ to encrypt the following message using the RSA cryptosystem:

   THE HUNS ARE COMING

   Use 2-letter blocks and don't omit spaces.

∗7. Find the decryption algorithm for the previous problem.

CHAPTER 9

The Pigeon-Hole Principle

The *Pigeon-Hole Principle (PHP)* is easily illustrated by a simple example:

> If 5 pigeons fly into 4 pigeon-holes then at least one pigeon-hole contains two or more pigeons.

i.e. if there is at least one more *pigeon* than there are *pigeon-holes* then at least one of the *pigeon-holes* has more than one *pigeon*.

Example. *Suppose that in my dresser drawer I have socks of three colours . . . loose. A bit silly, because I have to get up this morning while it's still dark. How do I ensure that I get a matching pair of socks in the most economical way . . . without disturbing my partner?*

Solution. I take 4 socks from the drawer . . . since then, by PHP, I must have at least one pair. The idea is that the *colours* (3 of them) are the *pigeon-holes* and the *socks* are the *pigeons*. □

Of course, this idea can be generalised a bit:

Theorem (Extended Pigeon-Hole Principle). *If there are* N pigeon-holes *and more than* kN pigeons *then at least one* pigeon-hole *has at least* $k+1$ pigeons.

Proof. Assume there are N pigeon-holes and $\geqslant kN+1$ *pigeons*, and for a contradiction, suppose each pigeon-hole has $\leqslant k$ pigeons. Then

$$\begin{aligned} \#\text{pigeons} &\leqslant kN \\ &< kN+1 \ \lightning \end{aligned}$$

So, in fact, there is a pigeon-hole with (at least) $k+1$ pigeons. □

Elections! Elections!

Voting for a member of the House of Representatives

Members for the House of Representatives are elected using the *preferential voting system*. Suppose that in a given *electorate* there are six *candidates*. Then, a *formal vote* (i.e. a vote that obeys the rules and so one that will be counted) numbers the candidates from 1 to 6 in some order.

So . . . how is the winning candidate determined? Well . . . first the number 1 votes are counted. In our example, this results in 6 piles of ballot papers: one for each candidate. The candidate with the least number 1 votes is then excluded; and that candidate's pile of ballot papers are re-distributed to the other 5 piles according to the number 2 votes on those ballot papers.

Question 9.1. How many of the number 1 (*primary*) votes must a candidate get to ensure they are not excluded at the first round?

At the second round the candidate with the smallest pile of ballot papers after the first round is excluded; and that candidate's pile of ballot papers are re-distributed to the remaining (4, in our example) piles according to the *next* number preference on those ballot papers, i.e. the *smallest* number vote that doesn't correspond to an excluded candidate – which will be either number 2 votes or number 3 votes.

Subsequent rounds are analogous to the second round. The natural conclusion of this process is a single pile corresponding to the winning candidate.

Question 9.2. How many of the number 1 (*primary*) votes must a candidate get to ensure they are not excluded at the kth round, (where k is less than the number of candidates)?

Question 9.3. At the k^{th} round, how big must a candidate's pile be at the beginning of the round to ensure they are not excluded at that round, (where k is less than the number of candidates)?

In answering these questions you will see a number of short-cuts to the process described above, e.g. if after any round a candidate's pile contains more than half the number of ballot papers then that candidate is certainly the winner.

Question 9.4. Can you think of another short-cut ... using the PHP?

Voting for a member of the Senate

Usually at a *general election* there is a *half-senate* election, i.e. as well as voting for all the *House of Representatives* we vote for *half* the Senate, the other half keep their jobs until there is another general election.* On 2 March 1996, in Western Australia there were 29 candidates for the 6 (of the 12) senator positions that had become vacant. So a *formal* Senate vote numbered the candidates in some order from 1 to 29.†

So ... how are the winning 6 candidates determined? Well ... first as for the *House of Representatives* the number 1 votes are counted. In the recent election, this would have resulted in 29 piles of formal ballot papers. Let V be the total number of *formal votes*. Then a candidate is elected once their pile of ballot papers achieves a *quota*,

$$Q = \left\lfloor \frac{V}{6+1} \right\rfloor + 1$$

of the formal votes cast.

Question 9.5. What is the significance of the *quota* Q? (Ideas explored with regard to House of Representatives voting should help you.)

Now ... what happens? Like the *House of Representatives* there are a number of *rounds*, but unlike the *House of Representatives* each round has two parts. Firstly, any candidate who has achieved the quota Q is declared elected. Then the ballot papers of these elected candidates are re-distributed to the next not-so-far elected preference but at a *reduced* value: they are scaled by the factor‡

$$\frac{c-Q}{c},$$

where c is the number of ballot papers in the candidate's pile. Once no more candidates can be lifted to a quota this way, the second part of the round begins. This proceeds *exactly* in

*Occasionally, exceptional circumstances bring about a *double dissolution*, where both houses of parliament are dissolved and there is a *full-senate* election.

†Voters had to either put a 1 in one box *above the line* or to number all boxes below the line. Each party corresponding to a box above the line logged with the Electoral Commission how *they* would number the boxes *below the line*. So really all *formal* votes cast, number the 29 candidates.

‡The idea is that the *surplus* votes for elected candidates are passed on to the remaining candidates; but it would be *unfair* to simply take any $c-Q$ votes as the surplus. So *all* c votes are re-distributed but at a reduced weighting.

the way a *House of Representatives* round does: the candidate with the smallest pile of ballot papers is excluded and those ballot papers are re-distributed to the remaining piles according to the *next* preference on those ballot papers, at *full* value (where "*next*" preference this time means the least number vote that corresponds neither to an *already-elected* candidate nor to an *excluded* candidate).

Now, to be convinced that this is a viable means of voting we really need only consider the following two questions:

Question 9.6. Why can no more than 6 be elected in this way?

Question 9.7. Why are (at least) 6 candidates elected by this method . . . unless there is a tie?

Exercise Set 9.

1. In a group of 8 people show that at least two have their birthday on the same day of the week.

2. Three natural numbers are chosen at random. Their sum is 19. Show at least one number is 7 or more.

3. A box contains 10 French books, 20 Spanish books, 8 German books, 15 Russian books, and 25 Italian books. How many must we choose to ensure that we have 12 books in the same language?

4. There are 30 students in a class. While doing a keyboarding test one student made 14 mistakes, while the rest made fewer mistakes. Show that at least 3 students made the same number of mistakes.

5. A teacher starts each year with 3 jokes. Over 12 years the teacher never repeated the same triple of jokes. What is the smallest number of jokes the teacher must have in her repertoire for this to be possible?

6. A canteen has 95 tables with a total of 465 chairs. Can we be sure that there is a table with at least 6 chairs?

7. Prove that of any 5 points chosen in an equilateral triangle of side-length 1, there are two points whose distance apart is at most $\frac{1}{2}$.

8. Suppose we have 27 *distinct* positive odd numbers . . . all less than 100. Show there is a pair of numbers whose sum is 102.

9. The integers 1 to 10 are arranged in random order around a circle. Show that there are three consecutive numbers whose sum is *at least* 17.

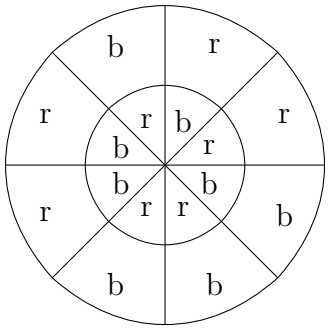
10. Six swimmers training together either swam in a race or watched the others swim. At least how many races must have been scheduled if every swimmer had opportunity to watch all of the others?

11. There are 11 people at a party. Some of them exchange handshakes with some of the others. Prove that at least two people have shaken the same number of hands.

12. A computer is used for 99 billable hours over a period of 12 days. Prove that on some pair of consecutive days the computer was used at least 17 billable hours.
Note. If the actual usage is x hours, the *billable hours* are $\lceil x \rceil$.

13. Show that given any 17 natural numbers it is possible to choose 5 whose sum is divisible by 5.

14. A circle is divided into 8 equal sectors. Half are coloured red and half are coloured blue. A smaller circle is also divided into equal sectors, half coloured red and half coloured blue. The smaller circle is placed concentrically on the larger. Prove that no matter how the red and blue sectors are chosen it is always possible to rotate the smaller circle so that at least 4 colour matches are obtained. (The diagram below shows an example.)



15. Five microcomputers are to be connected to three printers. How many connections are necessary between computers and printers in order to ensure that whenever any three computers require a printer the printers are available?

16. Prove that, of any 5 points chosen within a square of side-length 2, there are two whose distance apart is at most $\sqrt{2}$.

17. A disk of radius 1 is completely covered by 7 identical smaller disks. (They may overlap.) Show that the radius of each of the smaller disks must not be less than $\frac{1}{2}$.

18. A *graph* consists of *vertices* (singular: *vertex*) and *edges*. *Vertices* are usually represented by filled-in dots and each *edge* starts and finishes at a *vertex*. The *degree* of a *vertex* is the number of *edges* that start (or finish) at that *vertex*.
Suppose a *graph* has 9 *vertices* such that each vertex has *degree* 5 or 6. Prove that at least 5 vertices have degree 6 or at least 6 vertices have degree 5.

19. How many trees can farmer Fred plant on his 100 m square field if they are to be no closer than 10 m apart? (Neglect the thickness of the trees.)

CHAPTER 10

Summation and Product Notation

This is intended as a somewhat informal discussion of $\sum$ (summation) and $\prod$ (product) notation.

Firstly, Σ is the capital Greek letter *sigma* that corresponds to the Roman letter 'S', which, of course, is the first letter of *S*um. Thus, it makes sense to use Σ to represent summations.

Definition 10.1.

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n$$

i.e. $\sum_{k=1}^{n} a_k$ is the *'sum of all the a_ks from $k = 1$ up to $k = n$'.*

The sum need not start at $k = 1$; it can start at *any* integer. Typical other starting *indices* (in the context above, k is called an **index** variable; **indices** is the plural of **index**) are $k = 0$, $k = 2$, etc., but $k = 1$ and $k = 0$ are the most common starting indices. If the above sum needed to also include the term a_0 (each of the a_k is called a **term** of the summation), it would instead be written: $\sum_{k=0}^{n} a_k$.

Note that the indices k *always* step *upwards* and always in steps of 1, e.g.

$$\sum_{k=1}^{3} a_k = a_1 + a_2 + a_3.$$

To get the effect of steps larger than 1, we can use a function of the index variable to generate the terms, e.g.

$$\sum_{k=1}^{3} a_{3k-2} = a_1 + a_4 + a_7.$$

If the *beginning* and *end* indices are the *same,* the summation consists of just *one* term, e.g.

$$\sum_{k=0}^{0} a_k = a_0.$$

If the *end* index is *less than* the *beginning* index, then the summation is *empty,* i.e.

$$\sum_{k=1}^{0} a_k$$

is a *sum to no terms. Empty sums* are zero.

An *empty sum* is essentially a **degenerate** sum. A clue as to why an *empty sum* should be *zero,* comes from writing some computer code to implement $\sum_{k=1}^{n} a_k$:

Input: a_k, n *[The terms: $a_1, a_2, \ldots$ and end index n]*
Output: *sum*

Set *sum* := 0; *[Initialise sum]*
for $k := 1$ **to** n **do**
    sum := *sum* + a_k;
od;
return *sum*;

Observe that the variable *sum* is initialised to 0, and each time round the **for** loop the next term a_k is added. Once all the terms a_k have been added the final value of *sum* is returned as the **output** value. In particular, observe that if $n < 1$ then the **for** loop is *never* executed, in which case the returned value of *sum* is just the initial value of 0. Thus the computer algorithm demonstrates that an *empty sum* being *zero* makes sense.

The principal reasons for inventing such notation are: the notation is compact, and, more importantly, it is *unambiguous*. Use of an ellipsis ('$\cdots$' or '...') presumes we can identify a pattern, and is somewhat awkward for small sums, e.g. what does

$$1^2 + 2^2 + \cdots + n^2$$

mean, if $n = 1$? What it does mean is *'a sum to one term',* namely: 1^2 (which is of course, 1, in this case), which is to say, there is no 2^2 term. In such cases, the sigma-summation notation is more transparent:

$$\sum_{k=1}^{1} k^2 = 1^2 = 1.$$

Properties of $\sum$ for *finite* summations

The following properties are fairly straightforward to verify, simply by writing out in long-hand. A *finite* summation is one for which the upper index n is finite. (For *infinite* summations, the same rules may also apply, but there's an issue about *convergence,* a concept that's beyond the intended scope of these notes.)

$$\sum_{k=1}^{n}(a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k \qquad (10.1)$$

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{\ell} a_k + \sum_{k=\ell+1}^{n} a_k \qquad (10.2)$$

$$\sum_{k=1}^{n} ca_k = c\sum_{k=1}^{n} a_k \qquad (10.3)$$

$$\sum_{k=1}^{n} 1 = n \qquad (10.4)$$

$$\sum_{k=1}^{n} c = cn \qquad (10.5)$$

$$\sum_{k=1}^{n} a_k = \sum_{m=1}^{n} a_m \qquad (10.6)$$

$$\sum_{k=\ell+1}^{n} a_k = \sum_{k=1}^{n-\ell} a_{\ell+k} \qquad (10.7)$$

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} a_{n+1-k} \qquad (10.8)$$

$$\sum_{k=1}^{n} a_k = a_1 + \sum_{k=2}^{n} a_k \qquad (10.9)$$

$$\sum_{k=1}^{n} a_k = \Big(\sum_{k=1}^{n-1} a_k\Big) + a_n \qquad (10.10)$$

$$\sum_{k=1}^{n}\sum_{\ell=1}^{m} a_{k\ell} = \sum_{\ell=1}^{m}\sum_{k=1}^{n} a_{k\ell} \qquad (10.11)$$

$$\sum_{k=1}^{n}\sum_{\ell=1}^{m}(a_k b_\ell) = \Big(\sum_{k=1}^{n} a_k\Big)\Big(\sum_{\ell=1}^{m} b_\ell\Big) \qquad (10.12)$$

The first and second rules allow us to split certain sums into more than one sum.
The third rule says we can take out a common factor.
The fourth rule says the sum of n 1s is n.
The fifth rule is an obvious corollary of (10.3) and (10.4), generalising (10.4) to: the sum of n cs is nc.
The sixth rule says we can use *any* variable for the *index* variable. For this reason, it's often called a **dummy variable**.
The seventh rule says we can re-index the terms so that the *starting index* is 1.
The eighth rule says that we can sum the terms in reverse order.
The ninth and tenth rules say that we can split off an individual term from a sum. They are special cases of (10.2). The ninth splits off the bottom term. The tenth splits off the top term.
The eleventh rule says that we can swap the order of a double sum.
The twelfth rule says that we can split a double sum of products, so long as each of factors of the products being summed depends on only one of the index variables.

Common series with $\sum$ notation

One of the main uses of sigma-summation notation is for writing down *series* in a compact way. We will also write S_n for $\sum_{n=1}^{k}$, mnemonic for the *'sum to n terms'*. Some (finite) series you have met are:

arithmetic series

These have a **common difference** d, i.e. the terms satisfy the **recurrence (relation)**

$$a_{k+1} = a_k + d.$$

One can show, by induction, for example, that the terms of an *arithmetic series* therefore satisfies an **explicit relation**

$$a_k = a_1 + (k-1)d.$$

The usual way to sum *arithmetic series* is by the observation that if we write the series down *twice,* first *forwards* and then *backwards,* observing that the sum of each *pair* of corresponding terms of the series is the same, so that the total of the *forwards* and *backwards* series is n times the sum of the *first* and *last* terms of the series, i.e. observe that

$$\begin{aligned} a_k + a_{n-k+1} &= a_1 + (k-1)d + a_1 + (n-k+1-1)d \\ &= a_1 + a_1 + (n-1)d \\ &= a_1 + a_n \end{aligned}$$

independently of k, so that

$$S_n = \sum_{k=1}^{n} a_k \qquad = a_1 + a_2 \quad + \cdots + a_k \qquad + \cdots + a_n, \qquad (\textit{forwards})$$

$$S_n = \sum_{k=1}^{n} a_{n-k+1} \qquad = a_n + a_{n-1} + \cdots + a_{n-k+1} + \cdots + a_1, \qquad (\textit{backwards})$$

$$\begin{aligned} 2S_n = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} a_{n-k+1} &= (a_1 + a_n) + \cdots + (a_k + a_{n-k+1}) + \cdots + (a_n + a_1) \\ &= \sum_{k=1}^{n} (a_1 + a_n) \\ &= n(a_1 + a_n), \qquad \text{using property (10.5)} \\ \therefore S_n = \sum_{k=1}^{n} a_k &= \frac{n}{2}(a_1 + a_n) \\ &= \frac{n}{2}\big(2a_1 + (n-1)d\big) \end{aligned}$$

geometric series

These have a **common ratio** r, i.e. the terms satisfy the **recurrence (relation)**

$$a_{k+1} = a_k r.$$

One can show, by induction, for example, that the terms of an *arithmetic series* therefore satisfies an **explicit relation**

$$a_k = a_1 r^{k-1}.$$

The usual way to sum a *geometric series* is to observe that if one finds the difference of S_n and rS_n most terms cancel.

$$\begin{aligned} S_n &= \sum_{k=1}^{n} a_1 r^{k-1} = a_1 + a_1 r + \cdots + a_1 r^{k-1} + \cdots + a_1 r^{n-1} \\ rS_n &= \sum_{k=1}^{n} a_1 r^{k} \quad = \qquad a_1 r + \cdots + a_1 r^{k-1} + \cdots + a_1 r^{n-1} + a_1 r^n \\ (1-r)S_n &= a_1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - a_1 r^n \\ \therefore S_n &= \sum_{k=1}^{n} a_1 r^{k-1} = \begin{cases} a_1 \cdot \dfrac{1-r^n}{1-r} & \text{if } r \neq 1 \\ n a_1 & \text{if } r = 1. \end{cases} \end{aligned}$$

The $r = 1$ case, follows from property (10.5) directly, by observing that in this case all the terms are a_1.

binomial expansion

The **Binomial Theorem** states that

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \cdots + \binom{n}{r} a^{n-r} b^r + \cdots + \binom{n}{n} b^n,$$

where

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

Note that there are r factors in both the numerator and denominator of $\binom{n}{r}$, and as we'll see later, since empty products are 1, we have $\binom{n}{0} = 1$.

Binomial coefficients have the following five properties, where $r, n \in \mathbb{Z}$ and $0 \leqslant r \leqslant n$:

$$\binom{n}{0} = 1 \qquad\qquad \binom{n}{1} = n$$

$$\binom{n}{r} = \binom{n}{n-r}, \text{ (the symmetry property)} \qquad\qquad \binom{n+1}{r+1} = \binom{n}{r} + \binom{n}{r+1}$$

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

power series

The power series of interest to us are those of form $\sum_{k=1}^{n} k^m$ for various m, which we will find convenient to abbreviate to $\mathcal{S}_m$.

$$\mathcal{S}_1 = \sum_{k=1}^{n} k = 1 \ + 2 \ + 3 \ + 4 \ + \cdots + n \ = \frac{1}{2} n(n+1)$$

$$\mathcal{S}_2 = \sum_{k=1}^{n} k^2 = 1^2 + 2^2 + 3^2 + 4^2 + \cdots + n^2 = \frac{1}{6} n(n+1)(2n+1)$$

$$\mathcal{S}_3 = \sum_{k=1}^{n} k^3 = 1^3 + 2^3 + 3^3 + 4^3 + \cdots + n^3 = \frac{1}{4} n^2 (n+1)^2$$

 It's useful also to consider the degenerate case $m = 0$, (which is (10.4)):

$$\mathcal{S}_0 = \sum_{k=1}^{n} k^0 = \sum_{k=1}^{n} 1 = n.$$

We are also already familiar with the case $m = 1$, since it is an **arithmetic series**. The remaining cases can be proved recursively, by looking at $\sum_{k=1}^{n}\big((k+1)^{m+1} - k^{m+1}\big)$ in two ways. Firstly, we find the **binomial expansion** of $(1 + k)^{m+1}$ and observe that k^{m+1} is the last $(r = m + 1)$ term of that expansion.

$$\begin{aligned}
(k+1)^{m+1} = (1+k)^{m+1} &= \sum_{r=0}^{m+1} \binom{m+1}{r} k^r \\
&= \sum_{r=0}^{m} \binom{m+1}{r} k^r + k^{m+1}, && \text{by property (10.10)} \\
(k+1)^{m+1} - k^{m+1} &= \sum_{r=0}^{m} \binom{m+1}{r} k^r \\
\sum_{k=1}^{n}\big((k+1)^{m+1} - k^{m+1}\big) &= \sum_{k=1}^{n} \left(\sum_{r=0}^{m} \binom{m+1}{r} k^r \right) \\
&= \sum_{r=0}^{m} \left(\binom{m+1}{r} \sum_{k=1}^{n} k^r \right), && \text{by properties (10.11), (10.7) and (10.3)} \\
&= \sum_{r=0}^{m} \binom{m+1}{r} \mathcal{S}_r \\
&= \left(\sum_{r=0}^{m-1} \binom{m+1}{r} \mathcal{S}_r \right) + (m+1)\mathcal{S}_m, && \text{by property (10.10)}
\end{aligned}$$

Looking at the sum another way,

$$\begin{aligned}
\sum_{k=1}^{n}\big((k+1)^{m+1} - k^{m+1}\big) &= \sum_{k=1}^{n}(k+1)^{m+1} - \sum_{k=1}^{n} k^{m+1}, && \text{by property (10.1)} \\
&= \sum_{k=2}^{n+1} k^{m+1} - \sum_{k=1}^{n} k^{m+1}, && \text{by property (10.7)} \\
&= (n+1)^{m+1} - 1^{m+1}, && \text{by properties (10.10) and (10.9)}
\end{aligned}$$

Equating the two expressions we derived for $\sum_{k=1}^{n}\big((k+1)^{m+1} - k^{m+1}\big)$ we have:

$$\left(\sum_{r=0}^{m-1} \binom{m+1}{r} \mathcal{S}_r \right) + (m+1)\mathcal{S}_m = (n+1)^{m+1} - 1$$

so that on rearrangement we have the *recursive rule*

$$\boxed{\mathcal{S}_m = \frac{1}{m+1}\left((n+1)^{m+1} - 1 - \sum_{r=0}^{m-1} \binom{m+1}{r} \mathcal{S}_r \right), \quad m \geqslant 1.}$$

Actually, the above rule still works for the case $m = 0$, if we remember that empty sums are zero.

The notation for *products* is analogous to *sums:* Π is the capital Greek letter *pi* that corresponds to the Roman letter 'P', which, of course, is the first letter of *P*roduct. Thus, it makes sense to use Π to represent products.

Definition 10.2.

$$\prod_{k=1}^{n} a_k = a_1 \cdot a_2 \cdot \cdots \cdot a_n$$

i.e. $\prod_{k=1}^{n} a_k$ is the *'product of all the a_ks from $k = 1$ up to $k = n$'.*

The rules for products are similar to those for sums. As with sums, a product need not start at $k = 1$; it can start at *any* integer, and we have the following rules and properties.

Note that the indices k *always* step *upwards* and always in steps of 1, e.g.

$$\prod_{k=1}^{3} a_k = a_1 \cdot a_2 \cdot a_3.$$

To get the effect of steps larger than 1, we can use a function of the index variable to generate the terms, e.g.

$$\prod_{k=1}^{3} a_{3k-2} = a_1 \cdot a_4 \cdot a_7.$$

If the *beginning* and *end* indices are the *same,* the product consists of just *one* term, e.g.

$$\prod_{k=0}^{0} a_k = a_0.$$

If the *end* index is *less than* the *beginning* index, then the product is *empty,* i.e.

$$\prod_{k=1}^{0} a_k$$

is a *product to no terms. Empty products* are one.

An *empty product* is essentially a **degenerate** product. A clue as to why an *empty product* should be *one,* comes from writing some computer code to implement $\prod_{k=1}^{n} a_k$:

Input: a_k, n *[The terms $a_1, a_2, \ldots$ and end index n]*
Output: *product*

Set *product* := 1; *[Initialise product]*
for $k := 1$ **to** n **do**
    product := *product* $*$ a_k;
od;
return *product*;

Observe that the variable *product* is initialised to 1, and each time round the **for** loop the next term a_k is multiplied (the multiplication operator is $*$. Once all the terms a_k have been multiplied the final value of *product* is returned as the **output** value. In particular, observe that if $n < 1$ then the **for** loop is *never* executed, in which case the returned value of *product* is just the initial value of 1. Thus the computer algorithm demonstrates that an *empty product* being *one* makes sense.

The principal reasons for inventing such notation are exactly the reasons for the summation notation: the notation is compact, *unambiguous*, and, in particular, avoids the awkward cases encountered with using an ellipsis.

Properties of $\prod$ for *finite* products

A *finite* product is one for which the upper index n is finite. The ten rules below are what correspond to the first ten rules listed for summations. Note where there are differences. As with the summation properties, they are straightforward to verify, simply by writing out in long-hand.

$$\prod_{k=1}^{n} a_k b_k = \prod_{k=1}^{n} a_k \cdot \prod_{k=1}^{n} b_k \qquad (10.13)$$

$$\prod_{k=1}^{n} a_k = \prod_{k=1}^{\ell} a_k \cdot \prod_{k=\ell+1}^{n} a_k \qquad (10.14)$$

$$\prod_{k=1}^{n} c a_k = c^n \prod_{k=1}^{n} a_k \qquad (10.15)$$

$$\prod_{k=1}^{n} 1 = 1 \qquad (10.16)$$

$$\prod_{k=1}^{n} c = c^n \qquad (10.17)$$

$$\prod_{k=1}^{n} a_k = \prod_{m=1}^{n} a_m \qquad (10.18)$$

$$\prod_{k=\ell+1}^{n} a_k = \prod_{k=1}^{n-\ell} a_{\ell+k} \qquad (10.19)$$

$$\prod_{k=1}^{n} a_k = \prod_{k=1}^{n} a_{n+1-k} \qquad (10.20)$$

$$\prod_{k=1}^{n} a_k = a_1 \cdot \prod_{k=2}^{n} a_k \qquad (10.21)$$

$$\prod_{k=1}^{n} a_k = \Big(\prod_{k=1}^{n-1} a_k\Big) \cdot a_n \qquad (10.22)$$

Rules (10.13) and (10.14) allow us to split certain products into a product of products.
Rule (10.15) says we can take out a common factor, but unlike summations, since each term has a c, we have to take out n cs.
Rule (10.16) says the product of n 1s is 1.
Rule (10.17) is an obvious corollary of (10.15) and (10.16), generalising (10.16) to: the product of n cs is c^n.
Rule (10.18) says we can use *any* variable for the *index* variable, i.e. it is a **dummy variable**.
Rule (10.19) says we can re-index the terms so that the *starting index* is 1.
Rule (10.20) says that we multiply the terms in reverse order.
Rules (10.21) and (10.22) say that we can split off an individual term from a product. They are special cases of (10.14). Rule (10.21) splits off the bottom term. Rule (10.22) splits off the top term.

Common expressions with $\prod$ notation

Some (finite) products you have met are:

factorials

Factorial n, written $n!$, is the product of all the integers from 1 to n, i.e.

$$n! = 1 \cdot 2 \cdots n = \prod_{k=1}^{n} k.$$


Note that 0! is defined as 1, as you might expect (since the product notation shows that it is an *empty product*.)

prime factor decompositions

The *Fundamental Theorem of Arithmetic* says that any natural number can be decomposed as the product of primes, and, up to ordering, this decomposition is unique, i.e.

For any natural number N we may write

$$N = \prod_{k=1}^{n} p_k^{e_k}$$

for some primes p_k and integer exponents $e_k \geqslant 0$, $k = 1, \ldots, n$.

binomial coefficients

In product notation, for non-negative integers n, r such that $0 \leqslant r \leqslant n$,

$$\begin{aligned}
\binom{n}{r} &= \frac{n(n-1)(n-2)\cdots(n-r+1)}{r(r-1)(r-2)\cdots 1} \\
&= \frac{\prod_{k=1}^{r}(n+1-k)}{\prod_{k=1}^{r}(r+1-k)} = \frac{\prod_{k=1}^{r}(n-r+k)}{\prod_{k=1}^{r} k}, && \text{by property (10.20)} \\
&= \frac{\prod_{k=n-r+1}^{n} k}{\prod_{k=1}^{r} k}, && \text{by property (10.19)} \\
&= \frac{\left(\prod_{k=n-r+1}^{n} k\right) \cdot \prod_{k=1}^{n-r} k}{\prod_{k=1}^{r} k \cdot \prod_{k=1}^{n-r} k} = \frac{\prod_{k=1}^{n} k}{\prod_{k=1}^{r} k \cdot \prod_{k=1}^{n-r} k}, && \text{by property (10.14)} \\
&= \frac{n!}{(n-r)!r!} = \frac{n!}{r!(n-r)!}
\end{aligned}$$

Note that $\binom{n}{0}$ is 1, which is again expected, since the product notation shows that it is an *empty product*.

Note that it makes sense to define $\binom{n}{r} = 0$ for $r < 0$ or $r > n$, but in this case naive computer algorithms would give the wrong answer. (You can only take the concept of degeneracy so far. What could -1 factors mean? Usually one can make no sense of such a concept.)

power series

Related series to the power series we saw earlier are

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

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

$$\vdots$$

$$\sum_{k=1}^{n}\prod_{j=0}^{m-1}(k+j) = \sum_{k=1}^{n} k(k+1)\cdots(k+m-1) = \frac{1}{m+1}n(n+1)\cdots(n+m)$$

$$= \frac{1}{m+1}\prod_{j=0}^{m}(n+j)$$

In fact, the cases $m = 0$ and $m = 1$ which we have omitted above, are just the series we denoted by $\mathcal{S}_0$ and $\mathcal{S}_1$, respectively. A proof of the above general result is obtained by evaluating the sum

$$\sum_{k=1}^{n} k(k+1)\cdots(k+m) - \sum_{k=0}^{n-1} k(k+1)\cdots(k+m),$$

in two ways.

Exercise Set 10.

1. Find the corresponding *explicit relation* for the recurrence equations:

   (i) $a_{n+1} = 5a_n,\ a_0 = 1.$
   (ii) $a_{n+1} = 2a_n + 5,\ a_0 = 4.$
   (iii) $a_{n+1} = 2a_n + n + 1,\ a_0 = 4.$
   (iv) $x_{n+1} = x_n - n + 4,\ x_1 = 4.$
   (v) $u_n = u_{n-1} + n^3,\ u_1 = 1.$
   (vi) $u_{n+1} = u_n + \dfrac{1}{(4n+1)(4n+5)},\ u_1 = \frac{1}{5}.$

2. Find the corresponding *explicit relation* for the recurrences:

   (i) $a_{n+2} - 4a_n = 0,\ a_0 = 1,\ a_1 = 2.$
   (ii) $a_{n+2} - a_{n+1} - 6a_n = 0,\ a_0 = a_1 = 1.$
   (iii) $x_{n+2} - 5x_{n+1} + 6x_n = 0,\ x_0 = x_1 = 1.$
   (iv) $3u_{n+2} + 5u_{n+1} - 2u_n = 0,\ u_0 = 1,\ u_1 = 2.$

3. Evaluate

   (i) $\displaystyle\sum_{k=0}^{n} \frac{2}{(k+1)(k+2)}$
   (ii) $\displaystyle\sum_{k=0}^{n}(2k+1)$
   (iii) $\displaystyle\sum_{k=0}^{n}(4k^2+4k+8)$
   (iv) $\displaystyle\sum_{k=0}^{n} k^3$

CHAPTER 11

Algebra: Inequalities

11.1 Introduction

At school when you met the topic of *Inequalities* you were interested in finding the set of solutions for which a given inequality is satisfied, e.g. you might be asked:

For which $x \in \mathbb{R}$, is $x^2 \geqslant 3x - 2$.

One technique for *solving* such an inequality is rearrange it to have right hand side 0, and factorise the resulting left hand quadratic. It's then straightforward to determine the sign of that left hand side expression, and hence find the solution of the inequality as interval(s) of $\mathbb{R}$:

$$\begin{aligned} x^2 &\geqslant 3x - 2 \\ x^2 - 3x + 2 &\geqslant 0 \\ (x-2)(x-1) &\geqslant 0 \end{aligned}$$

Now, we observe that $(x-2)(x-1)$ is *positive* if both factors are *negative* or both factors are *positive*, and is *zero* when either factor is *zero*, i.e. the solution is

$$x \leqslant 1 \text{ or } x \geqslant 2.$$

Of course, one needs to start this way to gain some familiarity with how *inequations* differ from *equations*.

However, our interest in these lectures is to prove certain *Inequalities* hold for all $x \in \mathbb{R}$. One technique for this might be to *solve* an inequality as above, and show that the solution interval is all of $\mathbb{R}$, but for the sorts of inequalities with which we will consider, often involving several variables, this is generally not a useful approach. Instead, we will build up an armoury of *Standard Inequalities* and use these to prove the results we are after. Before we do that, let's start near the beginning.

11.2 Symbols and Elementary Rules

No doubt, you are very familiar with the symbols

$$> \quad \geqslant \quad < \quad \leqslant$$

but you probably have not thought much about the rules they obey. Let us start with some properties of *real* numbers.

- A real number can only be one of *positive*, *negative* or 0. Put another way, for a real number r, one of r or $-r$ is *positive* or else $r = 0$.
- The sum or product of two *positive* numbers is *positive*.
- Of course, for any real number r, $r + 0 = r$ and $r \cdot 0 = 0$.

Now, recognise that $a > b$ *means* that $a - b$ is *positive*. Also $a \geqslant b$ *means* that *either* $a > b$ *or* $a = b$. (Sometimes, it is useful to interpret $a = b$ as: $a - b$ is 0.) Of course, $a < b$ *means* $b > a$; and $a \leqslant b$ *means* $b \geqslant a$.

So now let's look at some rules that involve $>$ and $\geqslant$ (and $<$ and $\leqslant$). In each rule a, b, c, d are real numbers. The proofs will seem obvious – notice in each case we have used just *real* number properties (the main ones we use are mentioned above.)

- If $a > b$ then $a + c > b + c$. *(Note that c is allowed to be negative.)*

  Proof. Let $a > b$, i.e. $a - b$ is *positive*. Now $a - b = (a + c) - (b + c)$. So $(a + c) - (b + c)$ is *positive*, i.e. $a + c > b + c$. □

- If $a > b$ and c is *positive* then $ac > bc$.

  Proof. Let $a > b$, i.e. $a - b$ is *positive*. Also, let c be *positive*. Thus, $(a - b)c = ac - bc$ is *positive*, i.e. $ac > bc$. □

- If $a > b$ and c is *negative* then $ac < bc$.

  Proof. Let $a > b$ and c be *negative*, i.e. $a - b$ and $-c$ are *positive*. Thus, $(a - b)(-c) = bc - ac$ is *positive*, i.e. $bc > ac$ (or equivalently $ac < bc$). □

- Always $a^2 \geqslant 0$. (The *minimum value* property of a square.)

  Proof. If a is *positive* then $a \cdot a = a^2$ is *positive*. If $-a$ is *positive* then $(-a) \cdot (-a) = a^2$ is *positive*. If a is 0 then $a \cdot a = a^2$ is 0. Hence a^2 is *positive* or 0, i.e. $a^2 \geqslant 0$. □

- If $a > b$ and $b > c$ then $a > c$. *(Transitivity property)*

  Proof. Let $a > b$ and $b > c$, i.e. $a - b$ and $b - c$ are *positive*. Hence $(a - b) + (b - c)$ is *positive*. But $(a - b) + (b - c) = a - c$. Hence $a - c$ is *positive*, i.e. $a > c$. □

- If $a > b$ and $c > d$ then $a + c > b + d$.

  Proof. Let $a > b$ and $c > d$, i.e. $a - b$ and $c - d$ are *positive*. Hence $(a - b) + (c - d)$ is *positive*. But $(a - b) + (c - d) = (a + c) - (b + d)$. Hence $(a + c) - (b + d)$ is *positive*, i.e. $a + c > b + d$. □

- If $0 < a < b$ then $\frac{1}{a} > \frac{1}{b} > 0$.

  Proof. *Exercise.* □

- If $0 < a < 1$ and n is a natural number then $0 < a^n < 1$.

  Proof. *Exercise. (Hint: use Mathematical Induction.)* □

Observe that if we let $a = x/y$, $b = 1$ and $c = y$ then the second rule becomes:

If $\frac{x}{y} > 1$ and y is *positive* then $x > y$.

Thus, we may prove that $x > y$ by showing *either*

- $x - y$ is *positive; or*
- $\frac{x}{y} > 1$ provided that y is *positive.*

Example 11.2.1. *(i) If x, y are* distinct positive *numbers then*

$$x^3 + y^3 > x^2y + xy^2.$$

Proof. We will show that $(x^3 + y^3) - (x^2y + xy^2)$ is *positive*. Now

$$\begin{aligned}(x^3 + y^3) - (x^2y + xy^2) = x^3 - x^2y + y^3 - xy^2 &= x^2(x - y) + y^2(y - x)\\ &= (x^2 - y^2)(x - y)\\ &= (x + y)(x - y)^2.\end{aligned}$$

Now, by our properties of *real* numbers and our rules, both $x + y$ and $(x - y)^2$ are *positive*, and hence their product is *positive*, i.e. $x^3 + y^3 > x^2y + xy^2$. □

(ii) If $x > y > 0$ then

$$4x^3(x - y) > x^4 - y^4.$$

Proof. Since $x > y > 0$ we have $x > 0$ (using the *transitivity property*). Now $x^4 - y^4 = (x - y)(x + y)(x^2 + y^2)$ and each of $x - y$, $x + y$ and $x^2 + y^2$ is *positive*. (Check the details!) Hence $x^4 - y^4$ is *positive*. We are now in a position to prove the result by showing that

$$\frac{4x^3(x - y)}{x^4 - y^4} > 1.$$

But,

$$\begin{aligned}\frac{4x^3(x - y)}{x^4 - y^4} &= \frac{4x^3(x - y)}{(x - y)(x^3 + x^2y + xy^2 + y^3)}\\ &= \frac{4x^3}{x^3 + x^2y + xy^2 + y^3} \quad \text{since } x - y \neq 0\\ &= \frac{4}{1 + \dfrac{y}{x} + \dfrac{y^2}{x^2} + \dfrac{y^3}{x^3}} \quad \text{since } x \neq 0\\ &> 1\end{aligned}$$

The last step is valid since $0 < \frac{y}{x} < 1$. (Check all the skipped details!)
Thus, we may deduce that $4x^3(x - y) > x^4 - y^4$. □

11.3 Absolute values

Absolute values are often most easily treated from a geometric point of view. In particular, *the absolute value of a number measures its distance from* 0. We can extend this idea to interpret

$$|x - a|$$

as *the distance of x from a*. Thus to solve

$$|x + 1| < 3$$

we may first rewrite it as

$$|x - (-1)| < 3$$

and interpret it as: *the distance of x from -1 is less than 3* giving us $-4 < x < 2$. (To see this, draw a number line.)

Algebraically, we have the following definition for $|x|$,

$$|x| = \begin{cases} x, & \text{if } x \geqslant 0,\\ -x, & \text{if } x < 0.\end{cases}$$

and note that for a positive real number a we have that

$$|x| < a \quad \text{if and only if} \quad -a < x < a.$$

To gain some familiarity with manipulating absolute values, try the following exercises.

Exercises – absolute values.

1. Find the solution interval(s) for the following inequalities.

   (i) $|x+7| > 3$ (iii) $|x-2| \geqslant |2x+3|$

   (ii) $|2x-7| < 2$ (iv) $1-x \geqslant |x-1|$

11.4 Triangle Inequality

The name of this inequality comes from the geometric observation that the length of a side of triangle must lie between the difference and sum of the other two sides:

Theorem 11.4.1 (Triangle Inequality).

$$\big||x|-|y|\big| \leqslant |x+y| \leqslant |x|+|y|$$

for any real numbers x, y.

11.5 Squares are never negative

We identified this property earlier, but it's so important it bears repeating and putting it in its own section.

The square of a real number is never negative, i.e.

$$x^2 \geqslant 0, \quad \text{with equality} \iff x = 0,$$

or more generally

$$x_1^2 + x_2^2 + \cdots + x_n^2 \geqslant 0, \quad \text{with equality} \iff x_1 = x_2 = \cdots = x_n = 0.$$

Exercises – squares are non-negative.

2. Prove that for any non-negative a, b,

   $$\frac{a+b}{2} \geqslant \sqrt{ab}.$$

   This result is AM-GM (which we will discuss further later) for the case $n = 2$.

3. For arbitrary $a, b, c \in \mathbb{R}$, prove $a^2 + b^2 + c^2 \geqslant ab + bc + ca$.

4. (1990 USSR MO Q1) Prove that for arbitrary $t \in \mathbb{R}$, the inequality $t^4 - t + \frac{1}{2} > 0$ holds.

5. Let $a, b, c, d \in \mathbb{R}$. Prove that the numbers $a - b^2$, $b - c^2$, $c - d^2$, $d - a^2$ cannot all be larger than $\frac{1}{4}$.

6. Prove that $(a+5b)(3a+2b) \geqslant (a+9b)(2a+b)$ for all $a, b \in \mathbb{R}$.

7. Prove $(p+2)(q+2)(p+q) \geqslant 16pq$ for all $p, q \geqslant 0$.

8. Prove that $a^2(1+b^4) + b^2(1+a^4) \leqslant (1+a^4)(1+b^4)$ for all $a, b \in \mathbb{R}$.

9. If $x, y, z \in \mathbb{R}$ such that $x + y + z = 1$, prove that $x^2 + y^2 + z^2 \geqslant \frac{1}{3}$.

11.6 Arithmetic, Geometric and Harmonic Means

For a positive real number sequence $x_1, x_2, \ldots, x_n$, these means are defined by

$$\text{The } \textbf{Arithmetic Mean} \text{ (AM)} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\text{The } \textbf{Geometric Mean} \text{ (GM)} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

$$\text{The } \textbf{Harmonic Mean} \text{ (HM)} = \left(\frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}{n} \right)^{-1} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}.$$

Theorem (AM-GM-HM). *Let $x_1, x_2, \ldots, x_n$ be positive real numbers. Then*

$$\begin{aligned} & \mathrm{AM}(x_1, x_2, \ldots, x_n) \geqslant \mathrm{GM}(x_1, x_2, \ldots, x_n) \geqslant \mathrm{HM}(x_1, x_2, \ldots, x_n) \\ \textit{i.e.} \quad & \frac{x_1 + x_2 + \cdots + x_n}{n} \geqslant \sqrt[n]{x_1 x_2 \cdots x_n} \qquad \geqslant \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \end{aligned}$$

with equality $\iff x_1 = x_2 = \cdots = x_n$.

Proof. The statement can be proved by induction. The case $n = 1$ is trivially true. The case $n = 2$ follows after starting with

$$(\sqrt{x_1} - \sqrt{x_2})^2 \geqslant 0.$$

This gives $\mathrm{AM}(x_1, x_2) \geqslant \mathrm{GM}(x_1, x_2)$, from which can be deduced $\mathrm{GM}(x_1, x_2) \geqslant \mathrm{HM}(x_1, x_2)$.

Following Cauchy's approach, we deduce the AM-GM inequality for $n = 2k$ from the cases $n = 2$ and $n = k$. Similarly, we deduce the GM-HM inequality for $n = 2k$ from the cases $n = 2$ and $n = k$.

At this stage, one has AM-GM-HM for all powers of 2. To deduce for general n, first let $\alpha = \mathrm{AM}(x_1, x_2, \ldots, x_n)$. Then add in $(m - n)$ extra αs, where m is a power of 2. Then

$$\begin{aligned} \alpha = \mathrm{AM}(x_1, x_2, \ldots, x_n) &= \mathrm{AM}(x_1, x_2, \ldots, x_n, \alpha, \ldots, \alpha) \\ &\geqslant \mathrm{GM}(x_1, x_2, \ldots, x_n, \alpha, \ldots, \alpha) \\ &= \sqrt[m]{x_1 x_2 \cdots x_n \alpha^{m-n}} \\ \alpha^m &\geqslant x_1 x_2 \cdots x_n \alpha^{m-n} \\ \alpha^n &\geqslant x_1 x_2 \cdots x_n \\ \mathrm{AM}(x_1, x_2, \ldots, x_n) = \alpha &\geqslant \sqrt[n]{x_1 x_2 \cdots x_n} = \mathrm{GM}(x_1, x_2, \ldots, x_n) \end{aligned}$$

The proof of the GM-HM inequality for general n is similar. Start with $\alpha = \mathrm{HM}(x_1, x_2, \ldots, x_n)$, again add in $(m - n)$ extra αs, and deduce that $\mathrm{HM}(x_1, x_2, \ldots, x_n) \leqslant \mathrm{GM}(x_1, x_2, \ldots, x_n)$. □

Exercises – AM-GM Examples.

10. (1995 AIC* Q3) If $1 \leqslant n \in \mathbb{Z}$, prove that $(n + 1)^n \geqslant 2^n n!$. When does equality hold?

*AIC (Australian Intermediate Contest) was a 5-question fore-runner of the AIMO.

11. Prove that $(a+b)(b+c)(c+a) \geqslant 8abc$ for all nonnegative $a, b, c \in \mathbb{R}$.

12. Prove that $a^2+b^2+c^2 \geqslant ab+bc+ca$ for all $a, b, c \in \mathbb{R}$.

13. Prove that $x(a-x) \leqslant a^2/4$ if $a, x \in \mathbb{R}$, $x > 0$.

14. Prove that $a + 1/a \geqslant 2$, for all positive $a \in \mathbb{R}$.

15. (1961 Swedish MO Q2) For all positive $x_1, x_2, \ldots, x_n \in \mathbb{R}$, prove that

$$\frac{x_1}{x_2} + \frac{x_2}{x_3} + \cdots + \frac{x_n}{x_1} \geqslant n.$$

16. If $0 < a, b, c, d \in \mathbb{R}$ such that $a+b+c+d = 1$, prove that

$$\sqrt{4a+1} + \sqrt{4b+1} + \sqrt{4c+1} + \sqrt{4d+1} < 6.$$

Exercises – AM-HM Examples.

17. (1998 Irish MO Q7) Prove that if $0 < a, b, c \in \mathbb{R}$ then

$$\frac{9}{a+b+c} \leqslant \frac{2}{a+b} + \frac{2}{b+c} + \frac{2}{c+a} \leqslant \frac{1}{a} + \frac{1}{b} + \frac{1}{c}.$$

18. (1976 British MO Q2) Prove that if $0 < a, b, c \in \mathbb{R}$ then

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \geqslant \frac{3}{2}.$$

19. For positive $x_1, x_2, x_3, x_4 \in \mathbb{R}$, prove that

$$\frac{x_1+x_3}{x_1+x_2} + \frac{x_2+x_4}{x_2+x_3} + \frac{x_3+x_1}{x_3+x_4} + \frac{x_4+x_2}{x_4+x_1} \geqslant 4.$$

11.7 The Cauchy-Schwarz Inequality

Theorem 11.7.1 (Cauchy-Schwarz Inequality). *For $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{R}$,*

$$(a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2) \geqslant (a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2$$

with equality if and only if

$$a_1 : b_1 = a_2 : b_2 = \cdots = a_n : b_n.$$

The Cauchy-Schwarz Inequality is most easily remembered in terms of vectors:

$$\|\underset{\sim}{\boldsymbol{a}}\|^2 \|\underset{\sim}{\boldsymbol{b}}\|^2 \geqslant |\underset{\sim}{\boldsymbol{a}} \cdot \underset{\sim}{\boldsymbol{b}}|^2$$

$$\textit{i.e.} \quad \sum_i a_i^2 \cdot \sum_i b_i^2 \geqslant \Big(\sum_i a_i b_i\Big)^2$$

where the $a_i, b_i \in \mathbb{R}$ for all i, with equality if and only if $\underset{\sim}{\boldsymbol{a}} \parallel \underset{\sim}{\boldsymbol{b}}$.

Proof. Firstly, we give a proof without using vectors.
Since $(a_ix + b_i)^2 \geqslant 0$,

$$\sum_{i=1}^{n}(a_ix + b_i)^2 \geqslant 0$$
$$\Big(\sum_{i=1}^{n} a_i^2\Big)x^2 + 2\Big(\sum_{i=1}^{n} a_ib_i\Big)x + \Big(\sum_{i=1}^{n} b_i^2\Big) \geqslant 0$$

The lefthand side of the last inequality is a quadratic polynomial in x. Since it is nonnegative its graph either touches the x-axis at one point (i.e. the polynomial has *exactly* one zero) or is entirely above the x-axis (i.e. the polynomial has no real zeros). Consequently, polynomial's discriminant is *non-positive:*

$$4\Big(\sum_{i=1}^{n} a_ib_i\Big)^2 - 4\Big(\sum_{i=1}^{n} a_i^2\Big)\Big(\sum_{i=1}^{n} b_i^2\Big) \leqslant 0$$
$$\Big(\sum_{i=1}^{n} a_i^2\Big)\Big(\sum_{i=1}^{n} b_i^2\Big) \geqslant \Big(\sum_{i=1}^{n} a_ib_i\Big)^2.$$

Equality occurs when the polynomial has *exactly* one zero, which is to say that there is an x such that $a_ix + b_i = 0$ for all i, which is equivalent to saying the ratios $a_i : b_i$ are equal for all i.


In terms of vectors, we have the identity

$$\underset{\sim}{a} \cdot \underset{\sim}{b} = \|\underset{\sim}{a}\|\,\|\underset{\sim}{b}\| \cos\theta,$$

where θ is the angle between the 'tails' of the vectors. Squaring and using $|\cos\theta| \leqslant 1$, we have

$$\|\underset{\sim}{a}\|^2\|\underset{\sim}{b}\|^2 \geqslant \|\underset{\sim}{a}\|^2\|\underset{\sim}{b}\|^2\cos^2\theta$$
$$= |\underset{\sim}{a} \cdot \underset{\sim}{b}|^2.$$

Equality occurs if and only if

$$\cos\theta = 1 \iff \underset{\sim}{a} \parallel \underset{\sim}{b}.$$ □

Exercises – Cauchy-Schwarz.

20. Prove that for $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and positive $h_1, h_2, \ldots, h_n \in \mathbb{R}$, where $n \in \mathbb{N}$,

$$\sum_{i=1}^{n} \frac{a_i^2}{h_i} \geqslant \frac{(\sum_{i=1}^{n} a_i)^2}{\sum_{i=1}^{n} h_i}.$$

21. If $0 < a, b, c, d \in \mathbb{R}$, prove that

$$\frac{1}{a} + \frac{1}{b} + \frac{4}{c} + \frac{16}{d} \geqslant \frac{64}{a+b+c+d}.$$

22. For all $a, b, c \in \mathbb{R}$, prove that $a^2 + b^2 + c^2 \geqslant ab + bc + ca$.

23. If $0 < a, b, c, d \in \mathbb{R}$ such that $(a^2 + b^2)^3 = c^2 + d^2$, prove that
$$\frac{a^3}{c} + \frac{b^3}{d} \geqslant 1.$$

24. (1990 USSR MO Q10) If $0 < a_1, a_2, \ldots, a_n \in \mathbb{R}$ such that $a_1 + a_2 + \cdots + a_n = 1$, prove
$$\frac{a_1^2}{a_1 + a_2} + \frac{a_2^2}{a_2 + a_3} + \cdots + \frac{a_n^2}{a_n + a_1} \geqslant \frac{1}{2}.$$

11.8 Rearrangements

Theorem 11.8.1 (Rearrangement Inequality). *Let $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in \mathbb{R}$ such that $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$ and $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$, where $n \in \mathbb{N}$ and let $z_1, z_2, \ldots, z_n$ be any permutation (rearrangement) of $y_1, y_2, \ldots, y_n$. Then*
$$x_1y_n + x_2y_{n-1} + \cdots + x_ny_1 \leqslant x_1z_1 + x_2z_2 + \cdots + x_nz_n \leqslant x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Remark. Suppose we have two number sequences (terms x_i and y_j, respectively) of length n. Then the Rearrangement Inequality says that of the expressions one can form that are sums of n product pairs x_iy_j, the *minimum* value is achieved when the largest x_i is paired with the smallest y_j, the second-largest x_i is paired with the second-smallest y_j, and so on; and the *maximum* value is achieved when the largest x_i is paired with the largest y_j, the second-largest x_i is paired with the second-largest y_j, etc. Any other choice of pairings gives a value that lies between these minimum and maximum values.

Partial proof of Rearrangement Inequality. The following is a 'start' giving the general idea. Suppose $x_1 \leqslant x_2 \leqslant x_3$ and $y_1 \leqslant y_2 \leqslant y_3$. Then
$$\begin{aligned}(x_3 - x_2)(y_3 - y_2) &\geqslant 0\\ x_2y_2 + x_3y_3 &\geqslant x_2y_3 + x_3y_2\\ x_1y_1 + x_2y_2 + x_3y_3 &\geqslant x_1y_1 + x_2y_3 + x_3y_2\end{aligned}$$
Proceeding in this way leads to a general proof. □

Exercises – rearrangements.

25. For all $a, b, c \in \mathbb{R}$ prove $a^2 + b^2 + c^2 \geqslant ab + bc + ac$.

26. (1935 Eötvös Competition Q1) Let $y_1, y_2, \ldots, y_n$ be any permutation of the positive real numbers $x_1, x_2, \ldots, x_n$. Prove that
$$\frac{x_1}{y_1} + \frac{x_2}{y_2} + \cdots + \frac{x_n}{y_n} \geqslant n.$$

27. (1976 British MO Q2) For positive $a, b, c \in \mathbb{R}$, prove that
$$\frac{a}{b + c} + \frac{b}{c + a} + \frac{c}{a + b} \geqslant \frac{3}{2}.$$

28. (2002 Canadian MO Q3) For positive $x, y, z \in \mathbb{R}$, prove that
$$\frac{x^3}{yz} + \frac{y^3}{zx} + \frac{z^3}{xy} \geqslant x + y + z.$$

11.9 Optimisation applications

Later you will probably use calculus almost exclusively, when you need to find the maximum or minimum of a function, but you shouldn't forget that you can often use inequalities techniques for this purpose, and such solutions are often exquisitely short and *elegant!*

Example 11.9.1. *(i) Find the minimum value of $x^2+8x+23$ and the value(s) of x for which this minimum is attained.*

Solution. *We complete the square:*

$$\begin{aligned} x^2+8x+23 &= (x+4)^2+7 \\ &\geqslant 7, \qquad \text{since the square } (x+4)^2 \geqslant 0. \end{aligned}$$

Thus the expression is bounded below by $7 = 0+7$, and since at $x=-4$ we have $(x+4)^2=0$, in fact the lower bound is attained, i.e. the expression has a minimum *value 7 that is attained at $x=-4$.*

(ii) (Adapted from AIMO 2008 Q10) Find the maximum value of E satisfying

$$A + B + C + D + E = 0 \tag{11.9.1}$$
$$A^2+B^2+C^2+D^2+E^2 = 80. \tag{11.9.2}$$

Solution. *By AM-GM, for $n=2$, we have*

$$\frac{A^2+B^2}{2} \geqslant AB, \ \frac{A^2+C^2}{2} \geqslant AC, \ \ldots, \ \frac{C^2+D^2}{2} \geqslant CD,$$

with these all becoming equalities *if $A=B=C=D$. We will use this in step* (11.9.4) *below. Isolating E in* (11.9.1) *we have*

$$E = -(A+B+C+D) \tag{11.9.3}$$
$$\begin{aligned} E^2 &= (A+B+C+D)^2 \\ &= A^2+B^2+C^2+D^2+2AB+2AC+\cdots+2CD \\ &\leqslant A^2+B^2+C^2+D^2+(A^2+B^2)+(A^2+C^2)+\cdots+(C^2+D^2) \qquad (11.9.4) \\ &= 4(A^2+B^2+C^2+D^2), \qquad \text{since from } 2AB,\ 2AC \text{ and } 2AD \text{ we obtain } 3A^2 \text{ and by symmetry there are as many } A^2\text{s as } B^2\text{s, } C^2\text{s and } D^2\text{s} \\ &= 4(80-E^2), \qquad \text{using (11.9.2)} \\ \therefore 5E^2 &\leqslant 4\cdot 80 \\ E^2 &\leqslant 4\cdot 16 \\ E &\leqslant 8, \qquad \text{with equality if } A=B=C=D \ (=-E/4 \text{ by (11.9.3)}) \end{aligned}$$

Therefore, the maximum value of E is 8, attained when $A=B=C=D=-2$.

Later, in your mathematics career you will learn how to do the following problem by *Lagrange multipliers,* but AM-GM is quicker!

Exercises – Optimisation Application of AM-GM.

29. Given $a,b,c>0$, find the minimum value of $a+2b+7c$ such that $a^2b^5c=1$.

Hint. Let $x_1=x_2=\frac{a}{2}$, $x_3=\cdots=x_7=\frac{2b}{5}$, and $x_8=7c$, and note that minimum value is a nasty surd, but it's not difficult to obtain. We want the exact expression, but once you've found it you can find an approximate value with a calculator if you like `;-)`.

11.10 Generalising AM-GM-HM

The following theorem generalises the AM-GM-HM Theorem. Firstly, the **Quadratic Mean** (QM) is the 2-Power Mean:

$$\mathrm{QM}(x_1, x_2, \ldots, x_n) = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}.$$

In general, the k-**Power Mean** (PM_k), $k \in \mathbb{Z}$, is given by

$$\mathrm{PM}_k(x_1, x_2, \ldots, x_n) = \begin{cases} \sqrt[k]{\dfrac{x_1^k + x_2^k + \cdots + x_n^k}{n}}, & \text{if } k \neq 0 \\ \sqrt[n]{x_1 \cdot x_2 \cdots x_n}, & \text{if } k = 0. \end{cases}$$

With this definition, the AM is the 1-Power Mean, the GM is the 0-Power Mean, and the HM is the -1-Power Mean.

Theorem 11.10.1 (Power Mean (Hölder Mean)). *Let $0 < x_1, x_2, \ldots, x_n \in \mathbb{R}$. Then*

$$k \geqslant \ell \implies \mathrm{PM}_k(x_1, x_2, \ldots, x_n) \geqslant \mathrm{PM}_\ell(x_1, x_2, \ldots, x_n),$$

with equality $\iff x_1 = x_2 = \cdots = x_n$.

In particular,

$$\mathrm{QM}(x_1, \ldots, x_n) \geqslant \mathrm{AM}(x_1, \ldots, x_n) \geqslant \mathrm{GM}(x_1, \ldots, x_n) \geqslant \mathrm{HM}(x_1, \ldots, x_n).$$

Exercises – QM-AM-HM and Power Mean.

30. Prove the QM-AM part of the above theorem for the case $n = 2$.

31. Show that, if $a, b > 0$ and $a + b = 1$ then $\left(a + \dfrac{1}{a}\right)^2 + \left(b + \dfrac{1}{b}\right)^2 \geqslant \dfrac{25}{2}$.

32. (1976 Vietnam MO B3 Q6) For positive $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that $x_1 + x_2 + \cdots + x_n = 1$ and nonnegative $k \in \mathbb{Z}$, prove that

$$\frac{1}{x_1^k} + \frac{1}{x_2^k} + \cdots + \frac{1}{x_n^k} \geqslant n^{k+1}.$$

11.11 The Chebyshev Inequality

The following theorem essentially extends the Rearrangement Inequality, and we show that it follows from the Rearrangement Inequality.

Theorem 11.11.1 (Chebyshev Inequality). *If $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ then*

$$\frac{a_1b_1 + a_2b_2 + \cdots + a_nb_n}{n} \geqslant \frac{(a_1 + a_2 + \cdots + a_n)}{n} \cdot \frac{(b_1 + b_2 + \cdots + b_n)}{n} \geqslant \frac{a_1b_n + a_2b_{n-1} + \cdots + a_nb_1}{n}.$$

Proof. Assume $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$. Then by the Rearrangement Inequality, we have the following n inequalities:

$$\begin{aligned}
a_1b_1 + a_2b_2 + \cdots + a_nb_n = a_1b_1 + a_2b_2 + \cdots + a_nb_n &\geqslant a_1b_n + a_2b_{n-1} + \cdots + a_nb_1 \\
a_1b_1 + a_2b_2 + \cdots + a_nb_n \geqslant a_1b_2 + a_2b_3 + \cdots + a_nb_1 &\geqslant a_1b_n + a_2b_{n-1} + \cdots + a_nb_1 \\
a_1b_1 + a_2b_2 + \cdots + a_nb_n \geqslant a_1b_3 + a_2b_4 + \cdots + a_nb_2 &\geqslant a_1b_n + a_2b_{n-1} + \cdots + a_nb_1 \\
\vdots & \\
a_1b_1 + a_2b_2 + \cdots + a_nb_n \geqslant a_1b_n + a_2b_1 + \cdots + a_nb_{n-1} &\geqslant a_1b_n + a_2b_{n-1} + \cdots + a_nb_1.
\end{aligned}$$

Now, adding these n inequalities, followed by dividing through by n^2 gives the result:

$$\begin{aligned}
n(a_1b_1 + a_2b_2 + \cdots + a_nb_n) &\geqslant (a_1 + a_2 + \cdots + a_n)(b_1 + b_2 + \cdots + b_n) \\
&\geqslant n(a_1b_n + a_2b_{n-1} + \cdots + a_nb_1) \\
\therefore \frac{a_1b_1 + a_2b_2 + \cdots + a_nb_n}{n} &\geqslant \frac{(a_1 + a_2 + \cdots + a_n)}{n} \cdot \frac{(b_1 + b_2 + \cdots + b_n)}{n} \\
&\geqslant \frac{a_1b_n + a_2b_{n-1} + \cdots + a_nb_1}{n}.
\end{aligned}$$

□

Exercises – Chebyshev Inequality.

33. (2002 TT† Northern Autumn SO Q4) If $x, y, z \in \mathbb{R}$ such that $0 < x, y, z < \pi/2$, prove

$$\frac{x\cos x + y\cos y + z\cos z}{x + y + z} \leqslant \frac{\cos x + \cos y + \cos z}{3}.$$

11.12 There's more than one way!

We prove

$$a^2 + b^2 + c^2 \geqslant ab + bc + ca, \quad \text{for all } a, b, c \in \mathbb{R},$$

four different ways, in order to demonstrate the usage of some inequalities.

Proof 1 (using $x^2 \geqslant 0, x \in \mathbb{R}$). Assume $a, b, c \in \mathbb{R}$. Then

$$\begin{aligned}
(a - b)^2 &\geqslant 0 && (11.12.1)\\
(b - c)^2 &\geqslant 0 && (11.12.2)\\
(c - a)^2 &\geqslant 0 && (11.12.3)\\
\therefore\ a^2 - 2ab + b^2 & \\
+ b^2 - 2bc + c^2 & \\
+ c^2 - 2ca + a^2 &\geqslant 0, && \text{adding (11.12.1)–(11.12.3)}\\
\therefore\ 2a^2 + 2b^2 + 2c^2 &\geqslant 2ab + 2bc + 2ca \\
\therefore\ a^2 + b^2 + c^2 &\geqslant ab + bc + ca
\end{aligned}$$

□

†TT (Tournament of the Towns).

Proof 2 (using AM-GM). Assume $a, b, c \in \mathbb{R}$. Then $a^2, b^2, c^2 \geqslant 0$, so that by AM-GM we have the following three inequalities:

$$\frac{a^2+b^2}{2} \geqslant \sqrt{a^2b^2} \tag{11.12.4}$$

$$\frac{b^2+c^2}{2} \geqslant \sqrt{b^2c^2} \tag{11.12.5}$$

$$\frac{c^2+a^2}{2} \geqslant \sqrt{c^2a^2} \tag{11.12.6}$$

$$\begin{aligned} \therefore\ a^2+b^2+c^2 &\geqslant |ab|+|bc|+|ca|, \text{ adding (11.12.4)–(11.12.6),} \\ &\qquad \text{noting that } \sqrt{x^2} = |x| \text{ for any } x \in \mathbb{R}. \\ &\geqslant ab+bc+ca \end{aligned}$$

□

Proof 3 (using Cauchy-Schwarz). Assume $a, b, c \in \mathbb{R}$ and let

$$\underset{\sim}{\boldsymbol{x}} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \underset{\sim}{\boldsymbol{y}} = \begin{pmatrix} b \\ c \\ a \end{pmatrix}$$

then

$$\begin{aligned} \left\| \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right\|^2 \cdot \left\| \begin{pmatrix} b \\ c \\ a \end{pmatrix} \right\|^2 &\geqslant \left(\begin{pmatrix} a \\ b \\ c \end{pmatrix} \cdot \begin{pmatrix} b \\ c \\ a \end{pmatrix} \right)^2 \\ (a^2+b^2+c^2)^2 &\geqslant (ab+bc+ca)^2 \\ a^2+b^2+c^2 &\geqslant |ab+bc+ca| \\ &\geqslant ab+bc+ca \end{aligned}$$

□

Proof 4 (using Rearrangement). W.l.o.g. assume $a \leqslant b \leqslant c$ then (vacuously)

$$a \leqslant b \leqslant c$$

so that with b, c, a as a permutation of a, b, c, using the latter part of the Rearrangement Inequality, we have

$$\begin{aligned} ab+bc+ca &\leqslant a^2+b^2+c^2 \\ \text{i.e. } a^2+b^2+c^2 &\geqslant ab+bc+ca. \end{aligned}$$

□

Exercises – Miscellaneous examples.

34. Prove that for any natural number $n \geqslant 2$,

$$\frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} < 1.$$

Hint: First use the observation that

$$\frac{1}{k} - \frac{1}{k+1} = \frac{k+1-k}{k(k+1)} = \frac{1}{k(k+1)}$$

to prove

$$\frac{1}{1.2} + \frac{1}{2.3} + \cdots + \frac{1}{(n-1)n} = \frac{n-1}{n}.$$

35. Prove that for any positive a and b

$$\frac{2}{\frac{1}{a}+\frac{1}{b}} \leqslant \sqrt{ab} \leqslant \frac{a+b}{2} \leqslant \sqrt{\frac{a^2+b^2}{2}}.$$

Problem Set 11.

1. For all $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that $x_i \geqslant i^2, i = 1, 2, \ldots, n$, prove that
$$\frac{x_1 + x_2 + \cdots + x_n}{2} \geqslant \sqrt{x_1 - 1^2} + 2\sqrt{x_2 - 2^2} + \cdots + n\sqrt{x_n - n^2}.$$

2. Prove that for all $x, y, z \in \mathbb{R}$,
$$x^2 + y^2 + z^2 - xy - yz - zx \geqslant \tfrac{3}{4}(x - y)^2.$$

3. For any positive $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in \mathbb{R}$, prove that
$$\sum_{i=1}^{n} \frac{1}{x_i y_i} \geqslant \frac{4n^2}{\sum_{i=1}^{n} (x_i + y_i)^2}.$$

4. If $0 < a, b, c \in \mathbb{R}$, show that
$$(ab)^2 + (bc)^2 + (ca)^2 \geqslant abc(a + b + c).$$

5. For all $x \in \mathbb{R}$, prove that
$$\frac{x^2 + 2}{\sqrt{x^2 + 1}} \geqslant 2.$$

6. For $a, b, c, d \geqslant 0$, prove that $\sqrt{(a + c)(b + d)} \geqslant \sqrt{ab} + \sqrt{cd}$.

7. Let $a, b, c > 0$. Show that
$$\frac{a + b + c}{abc} \leqslant \frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}.$$

8. Show that, if $a, b > 0$ and $a + b = 1$ then
$$\left(a + \frac{1}{a}\right)^2 + \left(b + \frac{1}{b}\right)^2 \geqslant \frac{25}{2}.$$

9. For $x, y, z > 0$, prove that

   (a) $\dfrac{x^2}{y^2} + \dfrac{y^2}{z^2} + \dfrac{z^2}{x^2} \geqslant \dfrac{y}{x} + \dfrac{z}{y} + \dfrac{x}{z}$, (b) $\dfrac{x^2}{y^2} + \dfrac{y^2}{z^2} + \dfrac{z^2}{x^2} \geqslant \dfrac{x}{y} + \dfrac{y}{z} + \dfrac{z}{x}$.

10. Prove that, if $a, b, c \in \mathbb{R}$ then
$$a^4(1 + b^4) + b^4(1 + c^4) + c^4(1 + a^4) \geqslant 6a^2b^2c^2,$$
and determine when equality occurs.

11. Let $1 < n \in \mathbb{N}$. Prove that
$$\frac{1}{n} + \frac{1}{n + 1} + \cdots + \frac{1}{n^2} > 1.$$

12. Prove that, if $0 < a_1, a_2, \ldots, a_n \in \mathbb{R}$ such that $a_1 a_2 \cdots a_n = 1$ then

$$(1 + a_1)(1 + a_2) \cdots (1 + a_n) \geqslant 2^n.$$

13. Prove that if $0 < a, b, c, d, e \in \mathbb{R}$ then

$$\left(\frac{a}{b}\right)^4 + \left(\frac{b}{c}\right)^4 + \left(\frac{c}{d}\right)^4 + \left(\frac{d}{e}\right)^4 + \left(\frac{e}{a}\right)^4 \geqslant \frac{b}{a} + \frac{c}{b} + \frac{d}{c} + \frac{e}{d} + \frac{a}{e}.$$

14. For all $1 \leqslant x_1, x_2, \ldots, x_n \in \mathbb{R}$, prove that

$$\frac{(1 + x_1)(1 + x_2) \cdots (1 + x_n)}{1 + x_1 x_2 \cdots x_n} \leqslant 2^{n-1}.$$

15. (1997 Melb. Uni. Maths Comp. Senior Q4) If $0 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 1$, show that

$$\frac{x}{1 + y} + \frac{y}{1 + x} \leqslant 1.$$

16. Let a, b, c be the side-lengths of a triangle. Prove that

$$\frac{a}{b + c - a} + \frac{b}{c + a - b} + \frac{c}{a + b - c} \geqslant 3.$$

17. (2002 Mentor Set) Let a, b, c be the side-lengths of a triangle. Prove that

$$(a+b-c)(b+c-a)+(b+c-a)(c+a-b)+(c+a-b)(a+b-c) \leqslant \sqrt{abc}(\sqrt{a}+\sqrt{b}+\sqrt{c}).$$

18. (2002 Mentor Set) For positive $a_1, a_2, \ldots, a_n \in \mathbb{R}$, $n \geqslant 2$, show that

$$(a_1^3 + 1)(a_2^3 + 1) \cdots (a_n^3 + 1) \geqslant (a_1^2 a_2 + 1)(a_2^2 a_3 + 1) \cdots (a_n^2 a_1 + 1).$$

19. (1974 USA MO Q2/1995 Canadian MO Q2) For positive $x, y, z \in \mathbb{R}$, prove that

$$x^x y^y z^z \geqslant (xyz)^{(x+y+z)/3}.$$

20. (1998 Canadian MO Q3) Let $n \in \mathbb{N}$ such that $n \geqslant 2$. Show that

$$\frac{1}{n+1}\Big(1 + \frac{1}{3} + \cdots + \frac{1}{2n-1}\Big) \geqslant \frac{1}{n}\Big(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\Big).$$

21. (1995 Irish MO Q6) Show that for all $n \in \mathbb{N}$,

$$n^n \leqslant (n!)^2 \leqslant \Big(\frac{(n+1)(n+2)}{6}\Big)^n.$$

22. (1997 Irish MO Q4) Let a, b, c be nonnegative real numbers such that $a + b + c \geqslant abc$. Prove that

$$a^2 + b^2 + c^2 \geqslant abc.$$

23. (2000 Irish MO Q6) Let $0 \leqslant x, y \in \mathbb{R}$ such that $x + y = 2$. Prove that
$$x^2y^2(x^2 + y^2) \leqslant 2.$$

24. (1990 British MO) For any positive $x, y, z \in \mathbb{R}$, prove that
$$\sqrt{x^2 - xy + y^2} + \sqrt{y^2 - yz + z^2} \geqslant \sqrt{z^2 + zx + x^2}.$$

25. (1997/8 Iranian MO Round 2 Q5) If $1 < x, y, z \in \mathbb{R}$ such that $\frac{1}{x} + \frac{1}{y} + \frac{1}{z} = 2$, prove
$$\sqrt{x + y + z} \geqslant \sqrt{x - 1} + \sqrt{y - 1} + \sqrt{z - 1}.$$

26. For $0 < x \in \mathbb{R}$ and $n \in \mathbb{N}$, prove that
$$\frac{x^n}{1 + x + x^2 + \cdots + x^{2n}} \leqslant \frac{1}{2n + 1}.$$

27. (1976 Vietnam MO B3 Q6) For positive $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that $x_1 + x_2 + \cdots + x_n = 1$ and nonnegative $k \in \mathbb{Z}$, prove that
$$\frac{1}{x_1^k} + \frac{1}{x_2^k} + \cdots + \frac{1}{x_n^k} \geqslant n^{k+1}.$$

28. (1975 Swedish MO Q3) For positive $a, b, c \in \mathbb{R}$ and $n \in \mathbb{N}$, prove that
$$a^n + b^n + c^n \geqslant ab^{n-1} + bc^{n-1} + ca^{n-1}.$$

29. Prove that for all $a \geqslant b \geqslant 0$,
$$\frac{(a - b)^2}{8a} \leqslant \frac{a + b}{2} - \sqrt{ab} \leqslant \frac{(a - b)^2}{8b}.$$

30. Let $0 < x, y, z \in \mathbb{R}$. Prove that
$$\frac{x}{x + \sqrt{(x + y)(x + z)}} + \frac{y}{y + \sqrt{(y + z)(y + x)}} + \frac{z}{z + \sqrt{(z + x)(z + y)}} \leqslant 1.$$

31. (1989 USSR MO Q21) Find the least value of $(x + y)(y + z)$, given that $0 < x, y, z \in \mathbb{R}$ such that $xyz(x + y + z) = 1$.

32. (1991 USSR MO Q9) For all nonnegative a, b, c, prove that
$$\frac{(a + b + c)^2}{3} \geqslant a\sqrt{bc} + b\sqrt{ca} + c\sqrt{ab}.$$

33. (1992 CIS[‡] MO Q1) Prove that for all positive $a, b, c \in \mathbb{R}$,
$$a^4 + b^4 + c^2 \geqslant 2\sqrt{2}abc.$$

[‡]CIS (Commonwealth of Independent States) MO, was previously the USSR MO.

34. (1992 CIS[§] MO Q9) Prove that for any $a > 1$, $b > 1$,

$$\frac{a^2}{b-1} + \frac{b^2}{a-1} \geqslant 8.$$

35. (2002 TT Northern Autumn SO Q4) If $x, y, z \in \mathbb{R}$ such that $0 < x, y, z < \pi/2$, prove

$$\frac{x\cos x + y\cos y + z\cos z}{x+y+z} \leqslant \frac{\cos x + \cos y + \cos z}{3}.$$

36. (1986 AMO) Given $1 < n \in \mathbb{N}$ and $0 < a \in \mathbb{R}$, determine the maximum value of

$$\sum_{i=1}^{n-1} x_i x_{i+1}$$

taken over all sets of n nonnegative numbers x_i with sum a.

37. (1987 AMO) Prove that for each $n \in \mathbb{N}$ such that $n > 1$,

$$\sqrt{n+1} + \sqrt{n} - \sqrt{2} > 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}}.$$

38. (1992 AMO) Let $n \in \mathbb{N}$. Show that

$$\frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n-1} > n(\sqrt[n]{2} - 1).$$

39. (1992 AMO) Let $n \in \mathbb{N}$, $0 < a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $s = a_1 + a_2 + \cdots + a_n$. Prove that

$$\sum_{i=1}^{n} \frac{a_i}{s - a_i} \geqslant \frac{n}{n-1} \text{ and } \sum_{i=1}^{n} \frac{s - a_i}{a_i} \geqslant n(n-1).$$

40. (1997 AMO Q2) Let $a_1, a_2, \ldots, a_k \in \mathbb{R}$ satisfying

(i) $0 \leqslant a_1 \leqslant a_2 \leqslant \cdots \leqslant a_k$, and

(ii) $a_1 + a_2 + \cdots + a_k = 1$.

Prove that $\dfrac{a_1 + a_2 + \cdots + a_n}{n} \leqslant \dfrac{1}{k}$ for $n = 1, 2, \ldots, k$.

41. (1997 AMO Q7) Let $1 < m, n \in \mathbb{Z}$. Prove that

$$\frac{1}{\sqrt[m]{n+1}} + \frac{1}{\sqrt[n]{m+1}} > 1.$$

42. (1998 AMO Q6) Prove that for any $n \in \mathbb{N}$,

$$(1998n)! \leqslant \Big(\frac{3995n+1}{2} \cdot \frac{3993n+1}{2} \cdot \frac{3991n+1}{2} \cdot \cdots \cdot \frac{n+1}{2}\Big)^n.$$

[§]CIS (Commonwealth of Independent States) MO, was previously the USSR MO.

43. (1999 AMO Q5) Let $1 < x \in \mathbb{R}$ and $1 < n \in \mathbb{N}$. Prove that

$$1 + \frac{x-1}{nx} < \sqrt[n]{x} < 1 + \frac{x-1}{n}.$$

44. (2001 AMO Q4) Prove the polynomial $4x^8 - 2x^7 + x^6 - 3x^4 + x^2 - x + 1$ has no real root.

45. (1990 APMO Q2) Let $0 < a_1, a_2, \ldots, a_n \in \mathbb{R}$ and let S_k be the sum of all products of $a_1, a_2, \ldots, a_n$ taken k at a time. Show that

$$S_k S_{n-k} \geqslant \binom{n}{k}^2 a_1 a_2 \cdots a_n,$$

for $k = 1, 2, \ldots, n-1$.

46. (1991 APMO Q3) Let $0 < a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{R}$ such that

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} b_k.$$

Show that

$$\sum_{k=1}^{n} \frac{(a_k)^2}{a_k + b_k} \geqslant \tfrac{1}{2} \sum_{k=1}^{n} a_k.$$

47. (1996 APMO) Let $m, n \in \mathbb{N}$ such that $n \leqslant m$. Prove that

$$2^n n! \leqslant \frac{(m+n)!}{(m-n)!} \leqslant (m^2 + m)^n.$$

48. (1996 APMO) Let a, b, c be the side-lengths of a triangle. Prove that

$$\sqrt{a+b-c} + \sqrt{b+c-a} + \sqrt{c+a-b} \leqslant \sqrt{a} + \sqrt{b} + \sqrt{c},$$

and determine when equality occurs.

49. (1998 APMO Q3) Let $0 < a, b, c \in \mathbb{R}$. Prove that

$$\Big(1 + \frac{a}{b}\Big)\Big(1 + \frac{b}{c}\Big)\Big(1 + \frac{c}{a}\Big) \geqslant 2\Big(1 + \frac{a+b+c}{\sqrt[3]{abc}}\Big).$$

50. (2002 APMO) Positive real numbers x, y, z satisfy $\dfrac{1}{x} + \dfrac{1}{y} + \dfrac{1}{z} = 1$. Prove that

$$\sqrt{x + yz} + \sqrt{y + zx} + \sqrt{z + xy} \geqslant \sqrt{xyz} + \sqrt{x} + \sqrt{y} + \sqrt{z}.$$

51. (1990 IMO Short List Q23) If $0 \leqslant w, x, y, z \in \mathbb{R}$ such that $wx + xy + yz + zw = 1$, prove

$$\frac{w^3}{x+y+z} + \frac{x^3}{y+z+w} + \frac{y^3}{z+w+x} + \frac{z^3}{w+x+y} \geqslant \frac{1}{3}.$$

52. (1993 IMO Short List Q24) For positive $a, b, c, d \in \mathbb{R}$, show that

$$\frac{a}{b+2c+3d} + \frac{b}{c+2d+3a} + \frac{c}{d+2a+3b} + \frac{d}{a+2b+3c} \geqslant \frac{2}{3}.$$

53. (1998 IMO Short List A1) If $0 < x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that $x_1 + x_2 + \cdots + x_n < 1$, prove

$$\frac{x_1 x_2 \cdots x_n (1 - x_1 - x_2 - \cdots - x_n)}{(x_1 + x_2 + \cdots + x_n)(1 - x_1)(1 - x_2) \cdots (1 - x_n)} \leqslant \frac{1}{n^{n+1}}.$$

54. (1964 IMO Q2) Let a, b, c be the side-lengths of a triangle. Prove that

$$a^2(b + c - a) + b^2(c + a - b) + c^2(a + b - c) \leqslant 3abc.$$

55. (1975 IMO Q1) Let $x_i, y_i \in \mathbb{R}$ for $i = 1, 2, \ldots, n$, such that $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n$ and $y_1 \geqslant y_2 \geqslant \cdots \geqslant y_n$. Prove that, if $z_1, z_2, \ldots, z_n$ is any permutation of $y_1, y_2, \ldots, y_n$, then

$$\sum_{i=1}^{n} (x_i - y_i)^2 \leqslant \sum_{i=1}^{n} (x_i - z_i)^2.$$

56. (Adapted from 1978 IMO Q5) Let $a_1, a_2, \ldots, a_n$ be a sequence of distinct positive integers. Prove that for all $n \in \mathbb{N}$,

$$\sum_{k=1}^{n} \frac{a_k}{k^2} \geqslant \sum_{k=1}^{n} \frac{1}{k}.$$

57. (1983 IMO Q6) Let a, b, c be the side-lengths of a triangle. Prove that

$$a^2 b(a - b) + b^2 c(b - c) + c^2 a(c - a) \geqslant 0.$$

When does equality hold?

58. (1995 IMO Q2) Let $0 < a, b, c \in \mathbb{R}$ such that $abc = 1$. Prove that

$$\frac{1}{a^3(b + c)} + \frac{1}{b^3(c + a)} + \frac{1}{c^3(a + b)} \geqslant \frac{3}{2}.$$

59. (2000 IMO) Let $0 < a, b, c \in \mathbb{R}$ such that $abc = 1$. Prove that

$$\left(a - 1 + \frac{1}{b}\right)\left(b - 1 + \frac{1}{c}\right)\left(c - 1 + \frac{1}{a}\right) \leqslant 1.$$

60. (2001 IMO Q2) For all positive $a, b, c \in \mathbb{R}$, prove that

$$\frac{a}{\sqrt{a^2 + 8bc}} + \frac{b}{\sqrt{b^2 + 8ca}} + \frac{c}{\sqrt{c^2 + 8ab}} \geqslant 1.$$

CHAPTER 12

Plane Geometry

12.1 Introduction

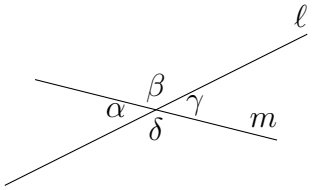
We adopt common conventions with regard to notation. **Points** are denoted by capitals $A, B, C, \ldots$. Given two points A, B there is exactly one **line** joining them denoted by AB. A **line** is a *straight line* that is infinite in extent in both directions. In other contexts, AB may denote the **line segment** of points between A and B, or the **ray** (or half-line) that starts at A and passes through B. The length of the line segment AB is usually also denoted by AB, but sometimes by $|AB|$. Given two rays AB and AC starting from the common point A, the angle they form is denoted by $\angle BAC$ or $\angle CAB$, or by just $\angle A$ if there is only one (interior) angle formed at the point A. Sometimes lines are denoted by lowercase letters $\ell, m, n, \ldots$. Lengths are sometimes denoted by lowercase Roman letters, e.g. $a, b, c, \ldots$, especially in triangles where a represents the length of the side opposite angle A, b represents the length of the side opposite angle B, etc. Sizes of angles are commonly represented by lowercase Greek letters $\alpha, \beta, \ldots$.

Ideally the theorems should be ordered so that the later ones follow from those proved earlier. Our ordering is close to such an ordering, but at times we compromise when a natural grouping of theorems dictates a different ordering.

12.2 Lines and angles

We take as an axiom that the angle at a point on a straight line is a constant regardless of the point or line. Such an angle is called a **straight angle** and its measure is 180°.

When two lines intersect, four angles are formed. Two such angles are called **vertically opposite** (or just **opposite**) if they are not formed on the same side of one of the lines. The straight angle axiom (postulate) implies the following theorem.



Theorem 12.2.1. *The opposite angles formed by intersecting straight lines are equal. In the diagram, $\alpha = \gamma$ and $\beta = \delta$.*

Proof. Let ℓ, m be intersecting straight lines, and let α, γ be opposite angles, with β an angle adjacent to both α and γ. Then $\alpha + \beta$ and $\gamma + \beta$ are straight angles. By the straight angle postulate,

$$\begin{aligned} \alpha + \beta &= \gamma + \beta \\ \alpha &= \gamma, \qquad \text{adding } -\beta \text{ to both sides.} \end{aligned}$$

□

12.3 Congruence of triangles

Definition 12.3.1. Two polygons are said to be **congruent**, if their corresponding sides and corresponding angles are equal.

When we say two *triangles* ABC and XYZ are *congruent* we mean that the correspondence of vertex A to X, B to Y and C to Z determines the congruence.

We denote that two triangles ABC and XYZ are *congruent* by writing $\triangle ABC \cong \triangle XYZ$.

Triangles may be determined to be congruent by rules known by the initialisms: SAS, SSS, ASA, and RHS. More precisely, these rules are as per the following theorem.

Theorem 12.3.2 (SAS, SSS, ASA, RHS Rules). *If, for two triangles,*

SAS: *two sides and the included angle of one triangle are equal to the two sides and the included angle of the other,*
or

SSS: *three sides of one triangle are equal to the three sides of the other,*
or

ASA: *two angles and the included side of one triangle are equal to the two angles and the included side of the other,*
or

RHS: *the hypotenuse and one other side of a right-angled triangle are equal to the hypotenuse and one side of the other right-angled triangle,*

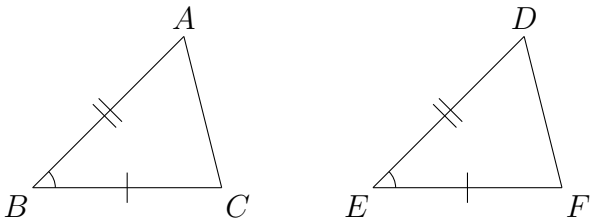
then the triangles are congruent.

Definition 12.3.3. A triangle is **isosceles** if two of its sides are equal. By convention, the common vertex of the two equal sides of an *isosceles* triangle is written between the other two vertices, i.e. to say $\triangle XYZ$ is isosceles we imply that $YX = YZ$.

Theorem 12.3.4. *If a triangle is* isosceles *then the angles opposite the equal sides are equal. Conversely, if two angles of a triangle are equal then the two sides opposite the equal angles are equal, so that the triangle is isosceles.*

To prove the two theorems of this section, it's convenient to do so in this order: *SAS Rule*, Theorem 12.3.4, *SSS Rule*, *ASA Rule*. For now we omit the proof of the *RHS Rule*.

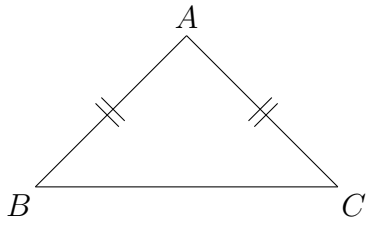
Proof of Theorem 12.3.2 SAS Rule.
In the triangles ABC and DEF we have side $AB = DE$, included angle $\angle ABC = \angle DEF$, and side $BC = EF$. We must show $\triangle ABC \cong \triangle DEF$.



Place triangle ABC over triangle DEF so that B falls on E and edge BC runs along line EF. Since $BC = EF$, C falls on F. Since $\angle ABC = \angle DEF$, line BA falls on ED, and since $AB = DE$, A fall on D. Since A falls on D and C falls on F, line segment AC falls on DF. Hence $\triangle ABC \cong \triangle DEF$. □

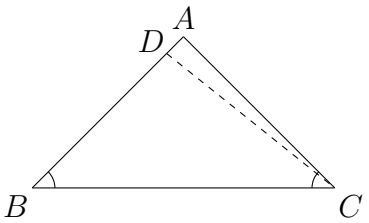
Proof of Theorem 12.3.4.
($\Longrightarrow$) Assume in triangle ABC that $AB = AC$. Then

$$\begin{aligned} AB = AC, &\quad \text{given} &\quad (12.3.1)\\ \angle BAC = \angle CAB, &\quad \text{same angle} &\\ AC = AB, &\quad \text{equivalent to (12.3.1)} &\\ \triangle ABC \cong \triangle ACB, &\quad \text{by SAS Rule} &\\ \therefore \angle ABC = \angle ACB & &\end{aligned}$$



So we have shown that an isosceles triangle has the angles opposite its equal sides equal.

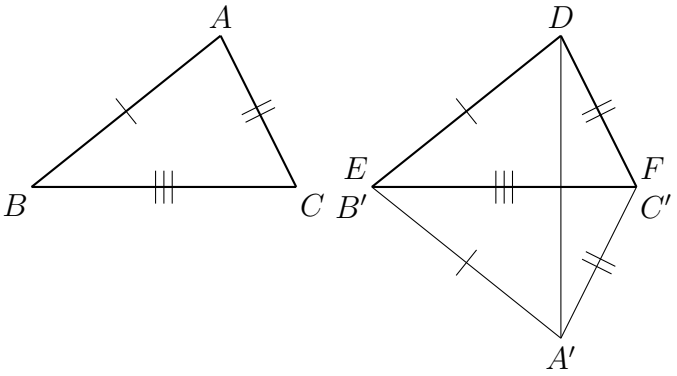
($\Longleftarrow$) Now assume in $\triangle ABC$ that $\angle ABC = \angle ACB$. Along the ray BA, construct (by compass) the point D such that $DB = AC$. Now we have



$$\begin{aligned} DB &= AC, && \text{by construction} \\ \angle ABC = \angle DBC &= \angle ACB, && \text{given} \\ BC &= CB, && \text{same line segment} \\ \triangle DBC &\cong \triangle ACB, && \text{by SAS Rule} \\ \therefore \angle DCB &= \angle ABC = \angle ACB \end{aligned}$$

Thus line DC coincides with line AC. Hence $D = A$, and $AB = DB = AC$. So we have shown that a triangle with two angles equal has the sides opposite the equal angles equal. □

Proof of Theorem 12.3.2 SSS Rule. Assume in triangles ABC and DEF that $AB = DE$, $BC = EF$ and $CA = FD$. Transport triangle ABC so that B falls on E and line BC runs along EF. Since $BC = EF$, C falls on F. Now let triangle ABC fall on the opposite side of line EF to triangle DEF so that A falls on A'. The transported copy of $\triangle ABC$ is $\triangle A'B'C'$ in the diagram.



By construction, $\triangle ABC \cong \triangle A'B'C'$, where $E = B'$ and $F = C'$. In particular, $A'E = AB = DE$ and $A'F = AC = DF$, so that triangles DEA' and DFA' are isosceles. So by Theorem 12.3.4, we have

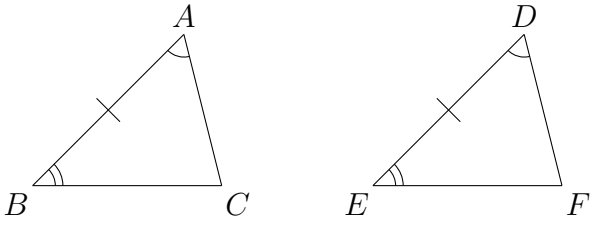
$$\angle EDA' = \angle EA'D \quad \text{and} \quad \angle FDA' = \angle FA'D.$$

Hence,

$$\angle EDF = \angle EDA' + \angle FDA' = \angle EA'D + \angle FA'D = \angle EA'F = \angle BAC.$$

So now $\triangle ABC \cong \triangle DEF$ by the SAS Rule. □

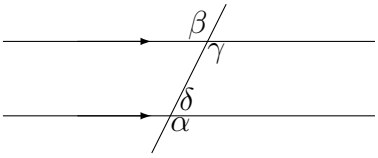
Proof of Theorem 12.3.2 ASA Rule. In the triangles ABC and DEF we have $\angle ABC = \angle DEF$, included side $AB = DE$ and $\angle BAC = \angle EDF$. Place $\triangle ABC$ over $\triangle DEF$ so that A falls on D and AB runs along DE. Since $AB = DE$, B falls on E. Also AC runs along DF because $\angle BAC = \angle EDF$. Similarly, BC runs along EF because $\angle ABC = \angle DEF$.



Thus the intersection point C of AC and BC must fall on the intersection point F of EF and DF. So $\triangle ABC$ is exactly superimposed over $\triangle DEF$, and hence $\triangle ABC \cong \triangle DEF$. □

12.4 Parallel lines

In the diagram the two horizontal lines are **parallel**. The line cutting the parallel lines is called a **transversal**. Angles α and β are called **alternate angles**, α and γ are **corresponding angles**, and angles α and δ are **supplementary angles**. Alternate angles and corresponding angles are equal, and pairs of supplementary angles sum to 180°.



12.5 Similarity of triangles

Definition 12.5.1. Two polygons are said to be **similar** (denoted by $\sim$), if

(i) corresponding sides are in the same proportion, and

(ii) corresponding angles are equal.

As with congruence, when we say two triangles ABC and XYZ are *similar* we mean that the correspondence of vertex A to X, B to Y and C to Z determines the similarity. We denote that two triangles ABC and XYZ are *similar* by writing $\triangle ABC \sim \triangle XYZ$.

Triangles may be determined to be similar by rules known by the initialisms: PAP, PPP, AA, and PPA. Each rule corresponds to a congruence rule, with side-length proportionality replacing equality. More precisely, these rules are as per the following theorem.

Theorem 12.5.2 (PAP, PPP, AA, PPA Rules). *If, for two triangles,*

PAP: *two sides of one of the triangles are in the same proportion to the two sides of the other, and the included angles between each pair of sides are equal,*
or

PPP: *three sides of one of the triangles are in the same proportion to the three sides of the other,*
or

AA: *two angles of one of the triangles equal two angles of the other,*
or

PPA: *two sides of one of the triangles are in the same proportion to the two sides of the other, and a corresponding non-included, non-acute angle of each triangle are equal,*

then the triangles are similar.

Theorem 12.5.3. *If a line joins the midpoints of two sides of a triangle then that line is parallel to the third side and its length is equal to one half of the length of the third side.*

Theorem 12.5.4. *A line parallel to one side of a triangle divides the other two sides in the same proportion.*

Theorem 12.5.5. *The bisector of one side of a triangle divides the opposite side in the same ratio as the other two sides.*

12.6 More triangle theorems

Theorem 12.6.1. *The sum of the interior angles of a triangle is* 180°.

Theorem 12.6.2. *An exterior angle of a triangle equals the sum of the two non-adjacent interior angles.*

12.7 Angles of a convex polygon

Theorem 12.7.1. *The sum of the interior angles of an n-sided convex polygon is $180(n-2)^\circ$.*

Proof. Let an interior point of the polygon be O. Construct line segments from the vertices of the polygon to O. The polygon is now divided into n triangles. The angle sum of the polygon is thus equal to the angle sum of the triangles *minus* the total of the angles around O, namely

$$180n^\circ - 360^\circ = 180(n-2)^\circ. \qquad \square$$

12.8 Quadrilaterals

Theorem 12.8.1. *The opposite sides and opposite interior angles of a parallelogram are equal.*

Theorem 12.8.2. *If a quadrilateral has opposite sides equal then it is a parallelogram.*

Theorem 12.8.3. *If a quadrilateral has opposite interior angles equal then it is a parallelogram.*

Theorem 12.8.4. *The diagonals of a parallelogram bisect each other.*

Theorem 12.8.5. *The diagonals of a rhombus are perpendicular.*

Also see the Theorem 12.12.10 on *cyclic quadrilaterals* in the *Circles* section.

12.9 Special Triangle Theorems

Theorem 12.9.1 (Pythagoras' Theorem). *In a right triangle the square of the hypotenuse is equal to the sum of the squares of the other two sides.*

Note that the next theorem is 'out of order'; it depends on Theorems 12.12.2 and 12.12.10.

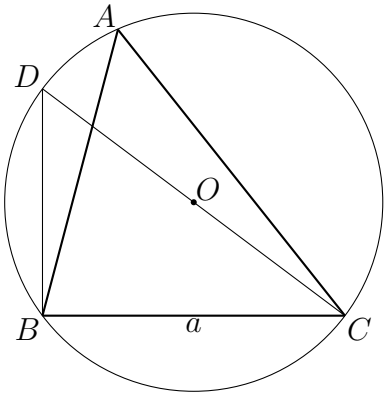
Theorem 12.9.2 (Sine Rule). *In a triangle ABC where a, b, c are the lengths of the sides opposite the vertices A, B, C, respectively,*

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R$$

where R is the circumradius of $\triangle ABC$.

Proof. Around the $\triangle ABC$ we draw its circumcircle, with circumcentre at O (initially assumed to be inside $\triangle ABC$), and radius R. Produce CO to meet the circumference at D, so that CD is a diameter; and then draw chord DB. Now $\angle CBD = 90^\circ$ since it is inscribed in a semicircle. So

$$\sin D = \frac{a}{CD} = \frac{a}{2R}.$$



But $\angle D = \angle A$, since both are inscribed in the arc BC. Thus we have $\sin A = \sin D = a/(2R)$, or equivalently

$$\frac{a}{\sin A} = 2R.$$

By symmetry, we also have

$$\frac{b}{\sin B} = 2R \quad \text{and} \quad \frac{c}{\sin C} = 2R.$$

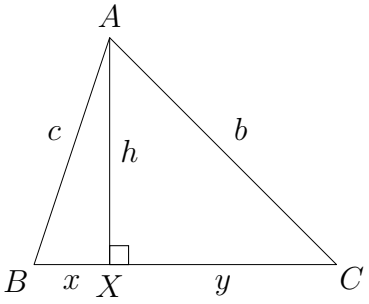
So we have proved the result for the case where O is inside $\triangle ABC$.

If O is outside $\triangle ABC$, producing CO to D as before, $CDBA$ is a cyclic quadrilateral, so that $\angle A$ and $\angle D$ are supplementary, whence $\sin A = \sin D$ and hence the result still follows. $\square$

Theorem 12.9.3 (Cosine Rule). *In a triangle ABC where a, b, c are the lengths of the sides opposite the vertices A, B, C, respectively,*

$$c^2 = a^2 + b^2 - 2ab\cos C$$

Proof. With $\triangle ABC$ as per the diagram, with altitude $h = AX$, $x = BX$, $y = XC$, we have

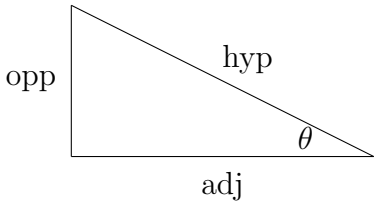


$$\begin{aligned} c^2 &= x^2 + h^2 \\ &= (a-y)^2 + h^2 \\ &= a^2 + y^2 + h^2 - 2ay \\ &= a^2 + b^2 - 2ab\cos C. \end{aligned}$$

□

12.10 Essential Trigonometry

Standard functions and their reciprocals:



$$\boxed{\sin\theta = \frac{\text{opp}}{\text{hyp}}} \qquad \boxed{\operatorname{cosec}\theta = \frac{1}{\sin\theta}}$$

$$\boxed{\cos\theta = \frac{\text{adj}}{\text{hyp}}} \qquad \boxed{\sec\theta = \frac{1}{\cos\theta}}$$

$$\boxed{\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{\sin\theta}{\cos\theta}} \qquad \boxed{\cot\theta = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta}}$$

Observe that each (function, reciprocal function) pair has one function whose name starts with `co` and one that doesn't.

Trig. versions of Pythagoras' Theorem:

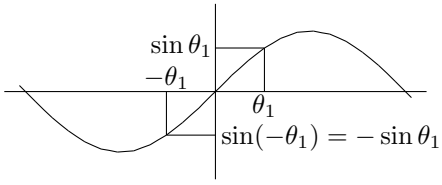
$$\boxed{\sin^2\theta + \cos^2\theta = 1}$$

$$\boxed{\tan^2\theta + 1 = \sec^2\theta}$$

$$\boxed{1 + \cot^2\theta = \operatorname{cosec}^2\theta}$$

The second equation is obtained from the first by dividing throughout by $\cos^2\theta$.
The third equation is obtained from first by dividing throughout by $\sin^2\theta$.

Oddness and evenness:



$$\boxed{\sin(-\theta) = -\sin\theta} \qquad \boxed{\sin \text{ is } \textit{odd}}$$

$$\boxed{\cos(-\theta) = \cos\theta} \qquad \boxed{\cos \text{ is } \textit{even}}$$

$$\boxed{\tan(-\theta) = -\tan\theta} \qquad \boxed{\tan \text{ is } \textit{odd}}$$

Complementary angles:

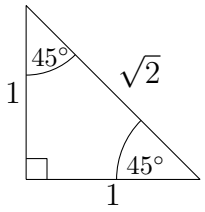
$$\boxed{\sin(90^\circ - \theta) = \cos\theta}$$

$$\boxed{\cos(90^\circ - \theta) = \sin\theta}$$

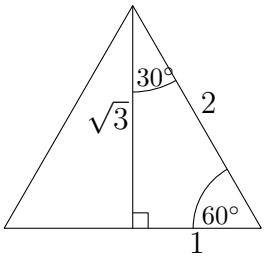
$$\boxed{\tan(90^\circ - \theta) = \cot\theta = \frac{1}{\tan\theta}}$$

Standard values:

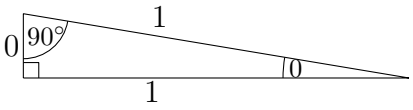
Isosceles right-angled triangle



Half equilateral triangle



Degenerate triangle



θ	$0°$	$30°$	$45°$	$60°$	$90°$
$\sin\theta$	$0 = \sqrt{\frac{0}{4}}$	$\frac{1}{2} = \sqrt{\frac{1}{4}}$	$\frac{1}{\sqrt{2}} = \sqrt{\frac{2}{4}}$	$\frac{\sqrt{3}}{2} = \sqrt{\frac{3}{4}}$	$1 = \sqrt{\frac{4}{4}}$
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
$\tan\theta$	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	∞

Sine Rule:

Theorem. *For $\triangle ABC$, with sides opposite angles $\angle A, \angle B, \angle C$ equal to a, b, c, respectively, and circumradius R,* $\boxed{\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R}$.

Periodicity: The *period* of sin and cos is $360°$, and of tan it is $180°$, i.e.

$$\boxed{\sin(\theta + 360°) = \sin\theta}$$

$$\boxed{\cos(\theta + 360°) = \cos\theta}$$

$$\boxed{\tan(\theta + 180°) = \tan\theta}$$

Symmetries: By sketching the graphs of $\sin\theta, \cos\theta, \tan\theta$ many symmetries are apparent, e.g.

$\boxed{\sin\theta = \sin(180° - \theta)}$ (symmetry about $\theta = 90°$)

$\boxed{\cos\theta = \cos(360° - \theta)}$ (symmetry about $\theta = 180°$)

Angle sum and difference identities:

$$\boxed{\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta}$$

$$\boxed{\sin(\alpha - \beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta}$$

$$\boxed{\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta}$$

$$\boxed{\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta}$$

$$\boxed{\tan(\alpha + \beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}}$$

$$\boxed{\tan(\alpha - \beta) = \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta}}$$

Double angle identities: These follow from the angle sum identities by putting $\beta = \alpha$.

$$\boxed{\sin(2\alpha) = 2\sin\alpha\cos\alpha}$$

$$\boxed{\begin{aligned}\cos(2\alpha) &= \cos^2\alpha - \sin^2\alpha \\ &= 1 - 2\sin^2\alpha \\ &= 2\cos^2\alpha - 1\end{aligned}}$$

$$\boxed{\tan(2\alpha) = \frac{2\tan\alpha}{1-\tan^2\alpha}}$$

12.11 Areas and perimeters

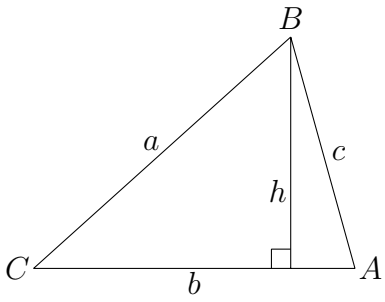
Notation 12.11.1. If a closed figure is denoted by $XY\ldots Z$ then its area is denoted by $|XY\ldots Z|$, i.e. by enclosing in vertical lines.

Theorem 12.11.2. *The area of a parallelogram is equal to bh where b is the length of its base and h is its height (the perpendicular distance from the base to the parallel side opposite).*

Theorem 12.11.3. *(a) The area of a triangle is equal to $\frac{1}{2}bh$ where b is the length of its base and h is its height (the perpendicular distance from the base to the vertex opposite).*

(b) Let the triangle be ABC, labelled in the standard way, and with vertex B opposite the base (which is thus labelled b). Then the area of $\triangle ABC$,

$$|ABC| = \tfrac{1}{2}ab\sin C.$$



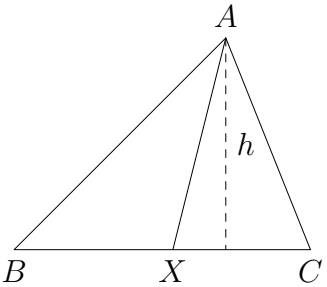
Proof of (b). Let the sides and height of $\triangle ABC$ be as labelled in the diagram. Then

$$\begin{aligned}\frac{h}{a} &= \sin C \\ \therefore h &= a\sin C \\ |ABC| &= \tfrac{1}{2}bh \\ &= \tfrac{1}{2}ba\sin C.\end{aligned}$$

□

Definition 12.11.4. In a triangle any side can be considered a **base**, and we refer to the vertex opposite such a side, as an **apex relative to** that chosen **base**. Similarly, an *altitude* dropped from an apex to its corresponding base, is an **altitude relative to** that **base**, e.g. in $\triangle ABC$, A is the *apex relative to base BC*, B is the *apex relative to base CA*, and the altitude emanating from A is the *altitude relative to base BC*.

Theorem 12.11.5. *If triangles $\triangle_1, \triangle_2$ (with areas $|\triangle_1|, |\triangle_2|$) have bases b_1, b_2 respectively along a common line and share an apex relative to bases b_1, b_2, then $|\triangle_1| : |\triangle_2| = b_1 : b_2$.*



Proof. Let $\triangle_1 = \triangle ABX$, $\triangle_2 = \triangle AXC$.
Then $\triangle_1, \triangle_2$ have common apex A relative to bases $b_1 = BX$, $b_2 = XC$ along a common line

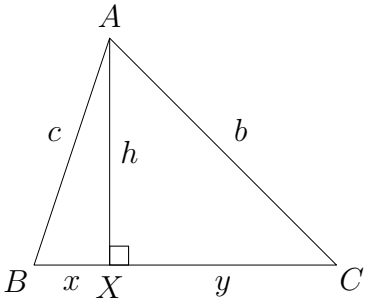
$$\begin{aligned}&\implies \triangle_1, \triangle_2 \text{ have common altitude } h \text{ dropped from } A \text{ to } BC \\ &\implies |\triangle_1| : |\triangle_2| = |ABX| : |AXC| \\ &\qquad\qquad = \tfrac{1}{2}b_1h : \tfrac{1}{2}b_2h \\ &\qquad\qquad = b_1 : b_2\end{aligned}$$

□

Theorem 12.11.6 (Heron's Theorem). *For $\triangle ABC$ with sides a, b, c and* semiperimeter *$s = (a+b+c)/2$, its area*

$$|ABC| = \sqrt{s(s-a)(s-b)(s-c)}.$$

Proof. With $\triangle ABC$ as per the diagram, with altitude $h = AX$, $x = BX$, $y = XC$, we have



$$\begin{aligned}
c^2 - x^2 &= h^2 \\
&= b^2 - (a-x)^2 \\
c^2 - b^2 &= x^2 - (a^2 - 2ax + x^2) \\
&= -a^2 + 2ax \\
2ax &= a^2 - b^2 + c^2 \\
x &= \frac{a^2 - b^2 + c^2}{2a} \\
h^2 &= c^2 - \frac{(a^2 - b^2 + c^2)^2}{4a^2} \\
|ABC|^2 &= \tfrac{1}{4}a^2h^2 \\
&= \tfrac{1}{4}a^2\Big(c^2 - \frac{(a^2 - b^2 + c^2)^2}{4a^2}\Big) \\
&= \tfrac{1}{16}\big(4a^2c^2 - (a^2 - b^2 + c^2)^2\big) \\
&= \tfrac{1}{16}\big((2ac)^2 - (a^2 - b^2 + c^2)^2\big) \\
&= \tfrac{1}{16}\big(2ac + (a^2 - b^2 + c^2)\big)\big(2ac - (a^2 - b^2 + c^2)\big) \\
&= \tfrac{1}{16}(a^2 + 2ac + c^2 - b^2)\big(b^2 - (a^2 - 2ac + c^2)\big) \\
&= \tfrac{1}{16}\big((a+c)^2 - b^2\big)\big(b^2 - (a-c)^2\big) \\
&= \tfrac{1}{16}(a+c+b)(a+c-b)(b+a-c)(b-a+c) \\
&= \tfrac{1}{16}\cdot 2s\cdot(2s-2b)(2s-2c)(2s-2a), \text{ where } 2s = a+b+c \\
&= s(s-b)(s-c)(s-a) \\
|ABC| &= \sqrt{s(s-b)(s-c)(s-a)}.
\end{aligned}$$

□

Theorem 12.11.7. *The area of a circle of radius r is πr^2 and its circumference is $2\pi r$.*

12.12 Circles

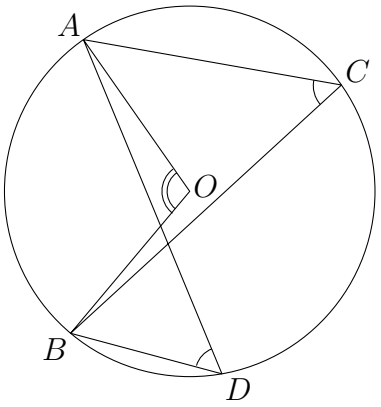
Theorem 12.12.1. *There is a unique circle through any triple of non-collinear points.*

Proof. Let the three non-collinear points be A, B, C. The points are distinct since two distinct points are sufficient to define a line, so that if any points are coincident then the points would be collinear, contradicting their non-collinearity. Form the perpendicular bisector of each of AB and BC. (Recall that the perpendicular bisector of two points is the locus of points that are equidistant from two given points.) These bisectors are non-parallel, since A, B, C are non-collinear. Hence the bisectors intersect. Let the point of intersection be O. Then $OA = OB$ since O lies on the bisector of AB. Similarly, $OB = OC$ since O lies on the bisector of BC. Thus

$$OA = OB = OC$$

so that O is equidistant from A, B and C. Hence we A, B and C lie on a circle with centre O and radius OA. Since O is the unique intersection of the perpendicular bisectors of AB and BC the circle through A, B and C is unique. □

Theorem 12.12.2. *If AB is an arc of a circle then angles subtended at the circumference opposite AB are equal and are equal to half the angle subtended at the centre, i.e. in the diagram $\angle ACB = \angle ADB = \frac{1}{2}\angle AOB$.*



Proof. Construct OC, forming isosceles triangles AOC and BOC. Let the equal base angles of $\triangle AOC$ be x and the equal base angles of $\triangle BOC$ be y. Then

$$\begin{aligned}\angle ACB &= x + y\\ \angle AOB &= 360^\circ - (180^\circ - 2x) - (180^\circ - 2y)\\ &= 2x + 2y = 2\angle ACB\\ \therefore \angle ACB &= \tfrac{1}{2}\angle AOB\end{aligned}$$

Suppose that C is in fact at D, and x and y are defined as before. Then $\angle ADB = y - x$ and

$$\begin{aligned}\angle AOB &= 180^\circ - 2x - (180^\circ - 2y)\\ &= 2(y - x) = 2\angle ACB\end{aligned}$$

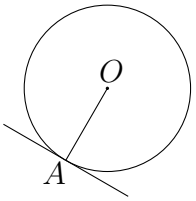
with the same conclusion as before. □

Theorem 12.12.3. *If AB is a semicircular arc of a circle and C is any point on the circumference of the circle then $\angle ACB$ is a right angle.*

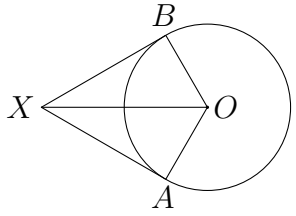
Theorem 12.12.4. *If A and B are points on the circumference of a circle with centre O and C is an exterior point of the circle such that BC is a tangent to the circle then $\angle ABC = \frac{1}{2}\angle AOB$.*

Theorem 12.12.5. *A line from the centre of a circle perpendicular to a chord bisects the chord and its arc.*

Theorem 12.12.6. *A line meeting a circle at a point A is tangent to the circle if and only if the radius to the point of contact with the line at A is perpendicular to the line.*



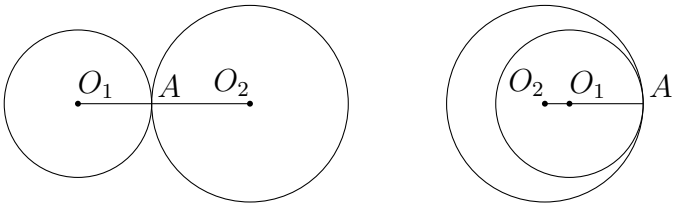
Theorem 12.12.7. *The two tangents drawn to a circle from an exterior point of the circle have the same length. In the diagram, $XA = XB$.*
Moreover, the line joining the centre of the circle and the exterior point bisects the angle between the two tangents. In the diagram, OX bisects $\angle AXB$.



Proof.

$OA = OB,$	(radii of same circle)
$90^\circ = \angle OAX = \angle OBX,$	by Theorem 12.12.6
OX is common	
$\therefore \triangle OAX \cong \triangle OBX,$	by the RHS Rule
$\therefore XA = XB$	
and $\angle OXA = \angle OXB,$	i.e. OX bisects $\angle AXB$. □

Theorem 12.12.8. *If two circles touch at a single point then this point and the centres of the circles are collinear. Below, O_1, O_2 and the point of contact A of the circles are collinear.*

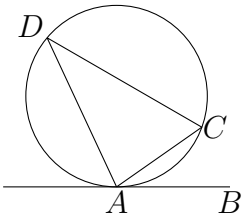


Theorem 12.12.9. *If two circles intersect at two points then the line through their centres. is the perpendicular bisector of their common chord.*

Theorem 12.12.10. *Opposite angles of a cyclic quadrilateral sum to* 180° *and if a pair of opposite angles of a quadrilateral sum to* 180° *then it is cyclic.*

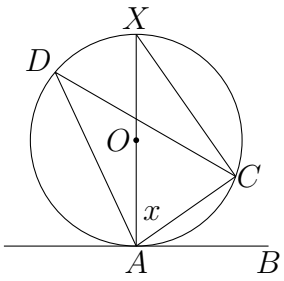
Theorem 12.12.11. *The centre of the circumcircle of a triangle is the intersection of the perpendicular bisectors of the sides of the triangle.*

Theorem 12.12.12 (Tangent-chord Theorem or Alternate Segment Theorem). *Let AC be a chord in a circle and let AB be a line meeting the circle at A. Then AB is tangent to the circle if and only if $\angle CAB = \angle ADB$ for any point D on the arc AC of the circle opposite B.*



Proof. Draw a diameter from A through the centre O (to X) and let $x = \angle XAC$. Then

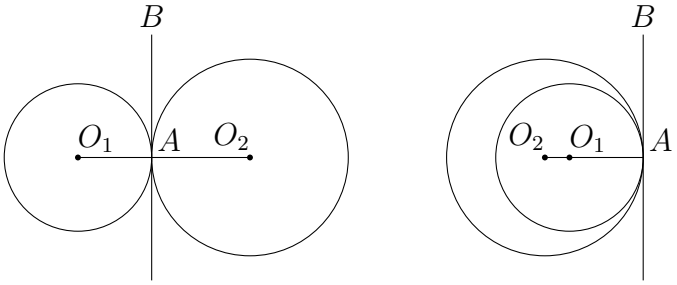
$$\begin{aligned}
\angle XAB &= 90^\circ, && \text{by Theorem 12.12.6, } OA \perp AB \\
\therefore \angle CAB &= \angle XAB - \angle XAC \\
&= 90^\circ - x \\
\angle ACX &= 90^\circ, && \text{by Theorem 12.12.3, since } AX \text{ is a diameter} \\
\therefore \angle AXC &= 180^\circ - \angle ACX - \angle XAC \\
&= 90^\circ - x \\
&= \angle CAB \\
\angle ADC &= \angle AXC, && \text{by Theorem 12.12.2, common arc: } AC \\
\therefore \angle ADC &= \angle CAB.
\end{aligned}$$



□

Remark. In the diagram of the proof above, if we think of X moving around the circumference of the circle, we always have $\angle AXC = \angle ADC$, by Theorem 12.12.2. As X moves around toward A, $\angle BAC$ can be thought of as the *limit* of $\angle AXC$ as $X \to A$, since the chord XA (extended) becomes the tangent AB when X and A become the one point.

Theorem 12.12.13. *If two circles are tangential at a point A then the tangent to one of the circles at A is also a tangent to the other circle.*



Theorem 12.12.14. *For any triangle ABC, the perpendicular bisectors of the three sides AB, BC and CA are concurrent, at some point O. Furthermore, $AO = BO = CO$, so that O is the centre of a circle K of radius $R = AO$ that passes through each of the vertices A, B, C of $\triangle ABC$.*
(K, O, R are respectively the ***circumcircle****,* ***circumcentre*** *and* ***circumradius*** *of $\triangle ABC$, and K is* ***circumscribed*** *about $\triangle ABC$.)*

Proof. Draw $\triangle ABC$ and let the midpoints of sides AB, BC and CA be D, E and F, respectively. Draw the perpendicular bisectors from sides AB and BC to meet at a point O, and join O to F (we must show OF is also a perpendicular bisector). Then

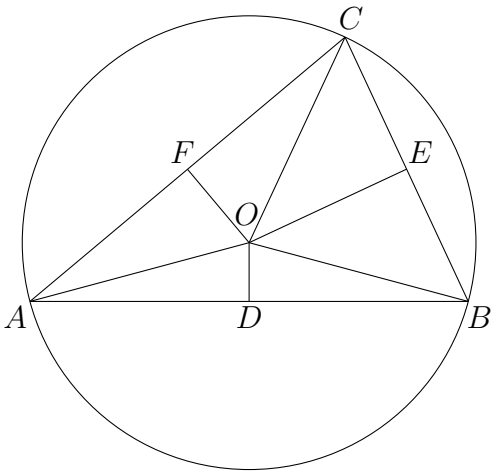
$$\begin{aligned} AD &= BD \\ 90^\circ = \angle ODA &= \angle ODB \\ OD &\text{ is common} \\ \therefore \triangle ODA &\cong \triangle ODB, && \text{by the SAS Rule} \\ \therefore AO &= BO \end{aligned}$$

Similarly, $\triangle OEB \cong \triangle OEC$

$$\therefore BO = CO.$$

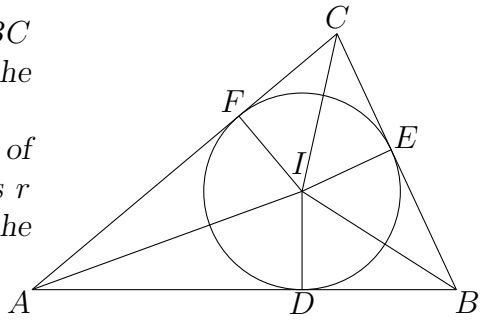
So now we have

$$\begin{aligned} AO &= CO \\ AF &= CF \\ OF &\text{ is common} \\ \therefore \triangle OFA &\cong \triangle OFB, && \text{by the SSS Rule} \\ \therefore \angle OFA = \angle OFC &= 90^\circ, && \text{since } \angle AFC \text{ is a straight angle } (180^\circ) \end{aligned}$$



Thus, OF is the perpendicular bisector of AC, and hence the perpendicular bisectors of $\triangle ABC$ concur at O. Also, we showed that $AO = BO = CO$, so that A, B and C lie on a circle centred at O with radius $R = AO$. □

Theorem 12.12.15. *The angle bisectors of any triangle ABC are* concurrent, *at some point I that is equidistant from the sides of $\triangle ABC$.*
(Let r be the common distance of I from the three sides of $\triangle ABC$. Then the circle $K(I, r)$ with centre I and radius r is ***inscribed*** *in $\triangle ABC$, and $K(I, r), I, r$ are respectively the* ***incircle****,* ***incentre*** *and* ***inradius*** *of $\triangle ABC$.)*



12.13 Glossary

Acute An angle is **acute** if it is smaller than a right angle.

Altitude The line through a vertex of a triangle that is perpendicular to the opposite side. A triangle has three **altitude**s; they are concurrent, meeting at the triangle's **orthocentre**.

Arc Any portion of the circumference of a circle.

Centroid The point at which the three medians of a triangle concur. The **centroid** trisects each of the medians, i.e. splits each median in the ratio $2 : 1$. More generally, the *centroid* of a figure is its **centre of mass**.

Cevian A line segment in a triangle joining a vertex and a point on the side opposite the vertex.

Chord A line segment whose endpoints lie on the circumference of a circle.

Circumcentre, circumcircle The three perpendicular bisectors of the sides of a triangle concur at the **circumcentre** of the triangle, which is the centre of the **circumcircle**, the circle that passes through the three vertices of the triangle.

Collinear This means *lying on the same straight line.* Several points are **collinear** if you can draw a single straight line through all of them.

Complementary angles A pair of angles whose sum is $90°$.

Concurrent This means *going through the same point.* Several lines are **concurrent** if they all intersect in the same point.

Congruent Two polygons are **congruent** if they have the same size and shape (i.e. if one were to shift and/or reflect one polygon the vertices of the two polygons could be made to line up exactly); in particular corresponding sides are of the same length.

Convex A set S of points on a line, plane or in space is **convex** if for any points A, B in S, all points on the line segment AB are in S. We say a polygon is *convex* if any line segment between points on the boundary of the polygon only intersects the interior of the polygon, i.e. all its interior angles are less than $180°$, e.g. any regular polygon is convex.

Cyclic A quadrilateral is **cyclic** if a circle may be drawn that passes through each of its four vertices.

Diameter A chord of a circle that passes through the circle's centre.

Edge A side of a geometrical figure, or more generally, a line segment that joins two vertices.

Equilateral A triangle is **equilateral** if all its sides are of equal length. An equilateral triangle necessarily has all its angle equal to $60°$.

Euler line The line in a triangle on which the *orthocentre, centroid* and *circumcentre* lie.

Hypotenuse The side opposite the right angle of a right triangle.

Incentre, incircle, inradius The three internal bisectors of the angles of a triangle concur at the **incentre** of the triangle, which is the centre of the **incircle**, the circle that touches each side of the triangle, i.e. each side of the triangle is a tangent to the incircle. The radius of the *incircle* is the triangle's **inradius**.

Isosceles A triangle is **isosceles** if two of its sides are of equal length, in which case, the two angles not included by the sides of equal length are equal.

Line In plane geometry, a **line** always means a *straight line* that is infinite in both directions.

Line segment A piece of a line of a definite length with two ends.

Locus (plural: **loci**) The line, curve or region traced out by a point satisfying certain conditions, e.g. if a point moves with fixed distance from a fixed point then its **locus** is a circle.

Median A line joining the vertex of a triangle to the midpoint of the opposite side. A triangle has three **median**s; they concur at the *centroid* of the triangle.

Medial triangle of a triangle ABC. Triangle formed by joining the midpoints of the sides of $\triangle ABC$.

Nine-point circle The feet of the three altitudes of a triangle ABC (i.e. the vertices of its *orthic triangle*), the midpoints of the sides of $\triangle ABC$ (i.e. the vertices of its *medial triangle*), and the midpoints of the line segments from the vertices of $\triangle ABC$ to the *orthocentre* of $\triangle ABC$, lie on the same circle; this circle is known as the *nine-point circle* of $\triangle ABC$. Its radius is $\frac{1}{2}R$, where R is the radius of the *circumcircle* of $\triangle ABC$. Its centre is the midpoint of the *Euler line* of $\triangle ABC$.

Obtuse An angle is *obtuse* if is larger than a right angle and smaller than a straight angle.

Orthogonal Same as *perpendicular*.

Orthocentre The common intersection point of the three altitudes of a triangle.

Orthic triangle of a triangle ABC. Triangle formed by joining the feet of the altitudes of $\triangle ABC$.

Parallel Two lines are **parallel** if they never meet.

Parallelogram A quadrilateral that has two pairs of parallel sides.

Pedal point, pedal triangle A **pedal point** is a point P inside a triangle ABC from which perpendiculars are dropped to the three sides of $\triangle ABC$. A triangle formed by joining the feet of the three perpendiculars dropped from a pedal point is called a **pedal triangle**. The **orthic triangle** is the *pedal triangle* formed when the *pedal point* P is the *orthocentre* of $\triangle ABC$. The **medial triangle** is the *pedal triangle* formed when the *pedal point* P is the *circumcentre* of $\triangle ABC$. In the case where P lies on the *circumcircle* of $\triangle ABC$, the feet Q, R, S of the perpendiculars to the sides of $\triangle ABC$ are collinear, so that the 'pedal triangle' formed is degenerate; the line through Q, R, S, in this case is known as a **simson**.

Perpendicular At right angles.

Plane figure A geometrical figure consisting of vertices and edges that can be drawn in the plane; a 2-dimensional object.

Polygon A plane figure whose edges are connected end to end in a loop. A **polygon** with n sides is sometimes called an n**-gon**. (Technically, a *gon* is an angle, but an n*-gon* has just as many sides as it has angles, so could just as easily have been called an n-lateral.) *Trigon* and *trilateral* are uncommon synonyms for *triangle*. 4-gons are generally referred to as *quadrilaterals* and sometimes as *quadrangles*. And we have *pentagon* (5-gon), *hexagon* (6-gon), *heptagon* (7-gon), *octagon* (8-gon), *nonagon* (9-gon), *decagon* (10-gon), *dodecagon* (12-gon), etc.

Point of Contact The common point of a tangent with a circle, or of two circles that are tangential.

Quadrangle, quadrilateral A polygon with 4 sides (and therefore 4 angles).

Radius (plural: **radii**) A line segment from the centre to the circumference of a circle.

Ray The part of a line that lies on one side of a point.

Reflex angle An angle that is larger than a straight angle but less than a full rotation, i.e. an angle of size between 180° and 360°.

Regular A polygon is **regular** if all its sides are equal and all its angles are equal.

Rhombus A parallelogram whose sides are all of equal length.

Right angle Half a straight angle. Its measure is 90°.

Right triangle, right-angled triangle A triangle with a right angle.

Secant A line that intersects a circle in two distinct points. A **chord** is just the segment of a secant that joins the two points of intersection with the circle.

Sector The area bounded by an arc of a circle and the two radii joining the arc.

Similar Two polygons are **similar** if angles at corresponding vertices are equal (if the two polygons are $ABC\ldots$ and $XYZ\ldots$ then A corresponds to X, B corresponds to Y, etc.), in which case corresponding sides are in the same proportion.

Simple A **simple** plane figure is one that does not cross itself.

Simson line, simson If P lies on the *circumcircle* of a triangle ABC then the feet Q, R, S of the perpendiculars drawn to the (extensions of the) sides of $\triangle ABC$ are *collinear*. The line through Q, R and S is the **Simson line** or **simson** of the point P with respect to triangle $\triangle ABC$. Also see *pedal point*.

Straight angle The angle at a point on (either side of) a straight line. Its measure is 180°.

Supplementary angles A pair of angles whose sum is 180°.

Tangent A line in the same plane as a circle that intersects (i.e. touches) the circle at exactly one point.

Tangential, touch A line and a circle, or two circles, are **tangential** (or **touch**) if they intersect at exactly one point.

Transversal A line that intersects two or more parallel lines.

Trapezium, trapezoid A quadrilateral that has one pair of opposite sides parallel.

Vertex (plural: **vertices**) A "corner" of a geometrical figure, i.e. a point at which edges meet.

12.14 Ceva's Theorem

Definition 12.14.1. A **cevian** of a triangle is a line segment joining a vertex of the triangle to a point on the opposite side. Thus, if X is a point on BC of $\triangle ABC$, then AX is a *cevian*. Medians, altitudes and angle bisectors are all examples of cevians.

Lemma 12.14.2 (Addendo). $k = \dfrac{a}{b} = \dfrac{c}{d}$ *where* $b, d, b+d \neq 0 \implies k = \dfrac{a+c}{b+d}$.

Proof. Assume $k = a/b = c/d$ and $b, d, b+d \neq 0$. Then

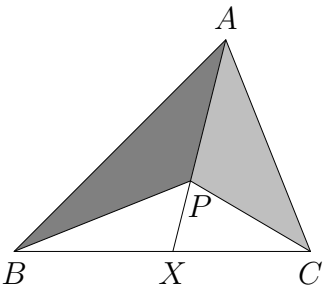
$$\begin{aligned} & a = kb \text{ and } c = kd \\ \implies \frac{a+c}{b+d} &= \frac{kb+kd}{b+d} \\ &= \frac{k(b+d)}{b+d} \\ &= k. \end{aligned}$$

□

Corollary 12.14.3. $k = \dfrac{a}{b} = \dfrac{c}{d}$ *where* $b, d, b-d \neq 0 \implies k = \dfrac{a-c}{b-d}$.

Proof. Observe that $c/d = (-c)/(-d)$. Then the result follows from Lemma 12.14.2 with c replaced by $-c$ and d replaced by $-d$. □

Lemma 12.14.4 (Ceva's Lemma). *If P is a point other than A on cevian AX of $\triangle ABC$ then $|ABP| : |CAP| = BX : XC$.*
In this case, we say $\triangle ABP$ and $\triangle CAP$ are above BX and XC.



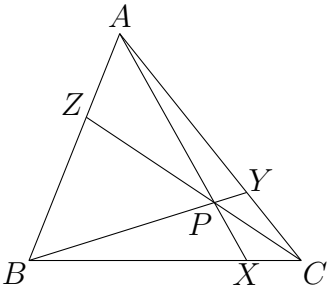
Proof. Since triangle pairs $\triangle ABX, \triangle AXC$ and $\triangle PBX, \triangle PXC$ have apices A and P, respectively, relative to bases BX and XC along a common line, by Theorem 12.11.5,

$$\begin{aligned} \frac{|ABX|}{|AXC|} &= \frac{BX}{XC}, \text{ and} \\ \frac{|PBX|}{|PXC|} &= \frac{BX}{XC} \\ \therefore \frac{|ABP|}{|CAP|} = \frac{|ABX| - |PBX|}{|AXC| - |PXC|} &= \frac{BX}{XC}, \text{ by Corollary 12.14.3, since } A \neq P \end{aligned}$$

i.e. $|ABP| : |CAP| = BX : XC$. □

Theorem 12.14.5 (Ceva's Theorem). *Let AX, BY, CZ be cevians of $\triangle ABC$. Then*

$$AX, BY, CZ \text{ concur} \iff \frac{BX}{XC}\frac{CY}{YA}\frac{AZ}{ZB} = 1.$$



Proof. ($\Longrightarrow$) Assume AX, BY, CZ of $\triangle ABC$ concur at a point P. Then in accordance with Lemma 12.14.4,

$\triangle ABP$ and $\triangle CAP$ are *above* BX and XC,
$\triangle BCP$ and $\triangle ABP$ are *above* CY and YA, and
$\triangle CAP$ and $\triangle BCP$ are *above* AZ and ZB.

Hence, by Lemma 12.14.4,

$$\begin{aligned}\frac{|ABP|}{|CAP|} &= \frac{BX}{XC},\\ \frac{|BCP|}{|ABP|} &= \frac{CY}{YA},\\ \frac{|CAP|}{|BCP|} &= \frac{AZ}{ZB}.\end{aligned}$$

So we have,

$$\begin{aligned}\frac{BX}{XC}\,\frac{CY}{YA}\,\frac{AZ}{ZB} &= \frac{|ABP|}{|CAP|}\cdot\frac{|BCP|}{|ABP|}\cdot\frac{|CAP|}{|BCP|}\\ &= 1.\end{aligned}$$

($\Longleftarrow$) Now suppose $\dfrac{BX}{XC}\,\dfrac{CY}{YA}\,\dfrac{AZ}{ZB} = 1$, and let P be the point of intersection of AX and BY. Let CP meet AB at Z'. Then by the forward argument we have

$$\frac{BX}{XC}\,\frac{CY}{YA}\,\frac{AZ'}{Z'B} = 1$$

and hence we have

$$\frac{AZ}{ZB} = \frac{AZ'}{Z'B}$$

so that both Z and Z' divide AB in the same ratio and must therefore be the same point. The result follows. □

12.15 The Euler Line

The **Euler line** of a triangle is the line segment joining its *circumcentre* and its *orthocentre*. What's particularly interesting is that the *centroid* also lies on this line:

Theorem 12.15.1. *Let the circumcentre, orthocentre and centroid of a triangle be O, H and G, respectively. Then O, H and G are collinear, and G divides the* Euler line *HO in the ratio $2:1$, i.e. $HG:GO = 2:1$.*

12.16 The Nine-point Circle

The Orthic Triangle. Let D, E, F be the feet of the altitudes from A, B, C, respectively, of $\triangle ABC$. Then $\triangle DEF$ is the **orthic triangle** of $\triangle ABC$.
The Medial Triangle. Let A', B', C' be the midpoints of sides BC, CA, AB, respectively, of $\triangle ABC$. Then $\triangle A'B'C'$ is the **medial triangle** of $\triangle ABC$.

Theorem 12.16.1 (Nine-point Circle). *The feet of the three altitudes of a triangle, the midpoints of the three sides, and the midpoints of the line segments from the vertices to the orthocentre, all lie on the circle with centre the centre of the Euler Line and with radius $\frac{1}{2}R$, where R is the circumradius, i.e. if in the notation above, $\triangle ABC$ has* orthic triangle *$\triangle DEF$,* medial triangle *$\triangle A'B'C'$, H is the orthocentre, and, further, K, L, M are the midpoints of line segments AH, BH, CH, respectively, and N is the midpoint of the Euler Line, then $\triangle DEF$, $\triangle A'B'C'$ and $\triangle KLM$ have the same circumcircle centred at N and of radius $\frac{1}{2}R$. Moreover, $\triangle A'B'C' \cong \triangle KLM$ and a half-turn (a rotation through $180°$ takes $\triangle A'B'C'$ onto $\triangle KLM$.*

12.17 The Radical Axis of Two Circles

Theorem 12.17.1 (Radical Axis of Two Circles). *The locus of points whose powers with respect to two non-concentric circles are equal is a line perpendicular to the line joining the centres of the circles. This line is called the **radical axis** of the two circles.*

12.18 Power of a Point

Definition 12.18.1. For a circle K of radius R and a point P of distance d from the centre of K, the **power** $\mathcal{P}(P, K)$ of P with respect to the circle K is

$$\mathcal{P}(P, K) = d^2 - R^2.$$

Theorem 12.18.2 (Power of a point). *If a line through a point P meets K a circle at points A and A' then the product*

$$PA \times PA'$$

*is the **power** of P with respect to the circle, i.e. if P is distance d from the centre of the circle K,*

$$PA \times PA' = d^2 - R^2.$$

Directed segment convention. *By adopting the convention that line segments are **directed**, the above equation accounts for the **power** being positive (PA and PA' are in the same direction), zero, or negative (PA and PA' are in opposite directions), according to whether P is outside, on, or inside the circle, respectively.*

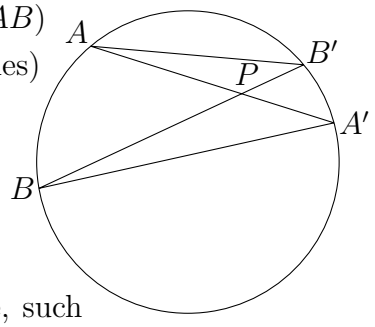
The above result follows from the next theorem.

Theorem 12.18.3 (Bowtie Theorem). *If two lines through a point P meet a circle at points A, A' and B, B', respectively, then*

$$PA \times PA' = PB \times PB'.$$

Proof. We have two main cases to consider: P inside the circle and P outside the circle. First suppose P is inside the circle.

$$\begin{aligned}
\angle AB'P = \angle AB'B = \angle AA'B = \angle BA'P, &\quad \text{(standing on same arc } AB\text{)}\\
\angle APB' = \angle BPA', &\quad \text{(vertically opposite angles)}\\
\therefore \triangle AB'P \sim \triangle BA'P, &\quad \text{by the AA Rule}\\
\therefore \frac{PA}{PB'} = \frac{PB}{PA'}&\\
\therefore PA \times PA' = PB \times PB'.&
\end{aligned}$$

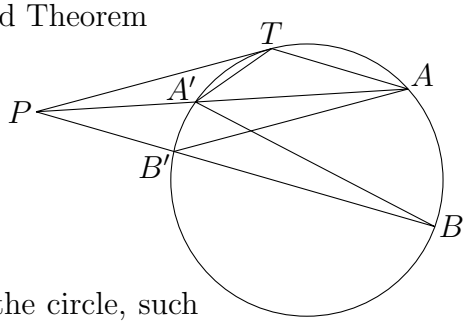


Now this result holds for any such pair of points B, B' on the circle, such that BB' passes through P. Consider the case where BB' is the diameter of the circle, let O, R be the centre and radius of the circle, and let $d = OP$. Then one of PB and PB' has length $R + d$ and the other $R - d$, i.e.

$$\begin{aligned}
PB \times PB' &= (R + d)(R - d)\\
&= R^2 - d^2.
\end{aligned}$$

Now suppose P is outside the circle, and let T be a point on the circle such that PT is tangent to the circle.

$$\begin{aligned}
\angle PTA' = \angle A'AT = \angle PAT &\quad \text{by the Tangent-chord Theorem}\\
\angle TPA' = \angle APT, &\quad \text{(same angle)}\\
\therefore \triangle TPA' \sim \triangle PAT, &\quad \text{by the AA Rule}\\
\therefore \frac{PT}{PA'} = \frac{PA}{PT}&\\
\therefore PA \times PA' = PT^2.&
\end{aligned}$$



Now this result holds for any other pair of points B, B' on the circle, such that BB' passes through P, i.e.

$$PB \times PB' = PT^2 = PA \times PA'.$$

In particular, for the case where BB' is the diameter of the circle, as before, let O, R be the centre and radius of the circle, and let $d = OP$. Then one of PB and PB' has length $d + R$ and the other $d - R$, i.e.

$$\begin{aligned}
PB \times PB' &= (d + R)(d - R)\\
&= d^2 - R^2.
\end{aligned}$$

In the case, that P lies on the circle, then exactly one of each pair of points A, A' and B, B' is P, say $P = A' = B'$, in which case, $PA' = PB' = 0$ and hence

$$PA \times PA' = 0 = PB \times PB'.$$

This case occurs precisely when $OP = d = R$, i.e. when

$$d^2 - R^2 = 0 = R^2 - d^2.$$

So finally by assigning a positive direction to one of the segments PA and PA', we see that when P is inside the circle, PA and PA' are oppositely directed so that $PA \times PA'$ is negative, i.e. with this convention,

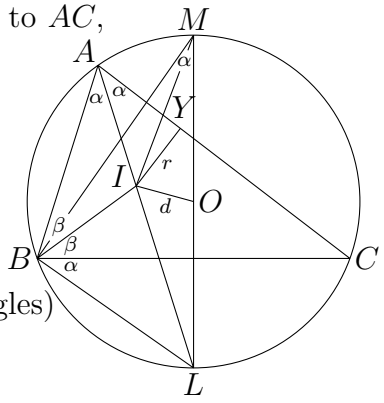
$$PA \times PA' = -(R^2 - d^2) = d^2 - R^2,$$

which is then the same formula in d and R as the case for P outside (or on) the circle. □

Theorem 12.18.4 (Euler). *Let O, R, I, r be the circumcentre, circumradius, incentre, and inradius of $\triangle ABC$ and let $d = OI$. Then*

$$d^2 = R^2 - 2rR.$$

Proof. Let K be the circumcircle of $\triangle ABC$. Produce AI to intersect K again at L. Produce LO to meet K again at M. Drop a perpendicular from I to AC, and let the foot of the perpendicular be Y, so that $IY = r$.
Let $\alpha = \frac{1}{2}\angle A$ and let $\beta = \frac{1}{2}\angle B$. Then



$$\begin{aligned}
\angle BML = \angle BAL &= \alpha, && \text{(angles standing on arc } BL\text{)}\\
\text{and } \angle LBC = \angle LAC &= \alpha, && \text{(angles standing on arc } LC\text{)}
\end{aligned}$$

Now,

$$\begin{aligned}
\angle BIL &= \alpha + \beta, && \text{(ext. angle is sum of int. opp. angles)}\\
&= \angle LBI\\
\therefore \triangle BLI &\text{ is isosceles}\\
\therefore LI &= LB.
\end{aligned}$$

Thus now we have,

$$\begin{aligned}
R^2 - d^2 &= LI \times IA, && \text{(power of point } I \text{ relative to } K\text{)}\\
&= LB \times IA\\
&= LM \cdot \frac{LB}{LM} \cdot \frac{IA}{IY} \cdot IY\\
&= LM \cdot \frac{\dfrac{LB}{LM}}{\dfrac{IY}{IA}} \cdot IY\\
&= LM \cdot \frac{\sin\alpha}{\sin\alpha} \cdot IY\\
&= 2R \cdot r\\
\therefore d^2 &= R^2 - 2rR.
\end{aligned}$$

□

Exercise Set 12.

1. Prove for any $\triangle ABC$, even if B or C is obtuse, that

$$a = b\cos C + c\cos B.$$

Thus use the Sine Rule to deduce the *addition formula:*

$$\sin(B + C) = \sin B\cos C + \sin C\cos B.$$

2. Prove for $\triangle ABC$,

$$a(\sin B - \sin C) + b(\sin C - \sin A) + c(\sin A - \sin B) = 0.$$

3. Prove for $\triangle ABC$, $|ABC| = \dfrac{abc}{4R}$.

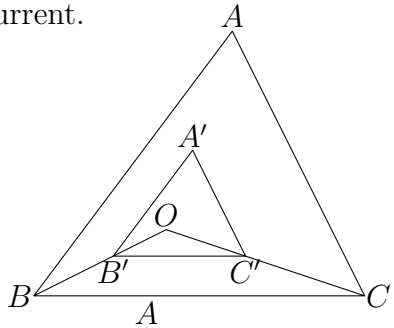
4. For $\triangle ABC$, let p and q be the radii of two circles through A, touching BC at B and C, respectively. Prove $pq = R^2$.

5. If X, Y, Z are the midpoints of sides BC, CA, AB, respectively, of $\triangle ABC$, prove the cevians AX, BY, CZ are concurrent.

The cevians AX, BY, CZ here, are the *medians* of $\triangle ABC$ and the point at which they concur is the *centroid* or *centre of gravity* of $\triangle ABC$.

6. Prove cevians perpendicular to the opposite sides are concurrent.

Such cevians of a triangle are its *altitudes* and the point at which they concur is the *orthocentre.*

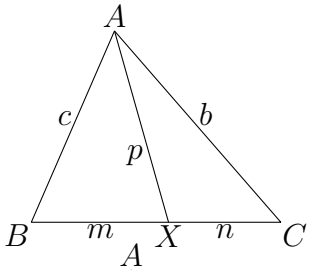


7. Let $\triangle ABC$ and $\triangle A'B'C'$ be non-congruent triangles whose corresponding sides are parallel. Prove the three lines AA', BB' and CC' (extended) are concurrent. Such triangles are said to be **homothetic**.

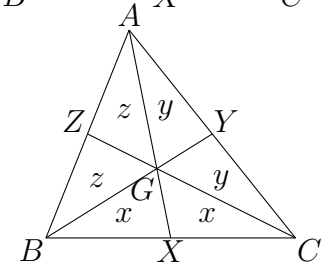
8. Let AX be a cevian of $\triangle ABC$ of length p dividing BC into segments $BX = m$ and $XC = n$. Prove

$$a(p^2 + mn) = b^2m + c^2n.$$

This result is known as **Stewart's Theorem.**

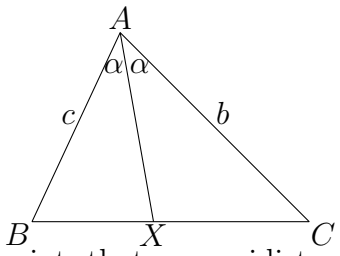


9. Prove that the medians of a triangle dissect the triangle into six smaller triangles of equal area.
Hint. Suppose in the diagram that x, y, z represent the areas of the smaller triangles they lie in. Start by showing the two triangles marked x have the same area, and similarly for y and z.



10. Prove the medians of a triangle divide each other in the ratio 2 : 1, i.e. the medians of a triangle *trisect* one another.

11. Prove that each (internal) angle bisector of a triangle divides the opposite side into segments proportional in length to the adjacent sides, e.g. if AX is the cevian of $\triangle ABC$ that bisects the angle at A internally, then $BX : XC = c : b$.



12. The angle bisector of the angle between two sides is the locus of points that are equidistant from the sides making the angle. One consequence of this is that any pair of internal angle bisectors of a triangle meet at a point that is equidistant from all three sides of the triangle, and hence in fact the three internal angle bisectors are concurrent.

    The point at which the angle bisectors of a triangle concur is the *incentre* I, the common (perpendicular) distance from I to the three sides is the *inradius* r, and the circle with centre I and radius r thus touches each side tangentially and is called the *incircle* of the triangle.

    Find an alternative proof that the (internal) angle bisectors of a triangle concur, using Ceva's Theorem and the result of the previous problem.

13. Prove the circumcentre and orthocentre of an obtuse-angled triangle lie outside the triangle.

14. Find the ratio of the area of a given triangle to that of a triangle whose sides have the same lengths as the medians of the original triangle.

15. Prove a triangle with two equal medians is isosceles.

16. Prove a triangle with two equal altitudes is isosceles.

17. Suppose that AX, of Exercise 8., is a median of $\triangle ABC$. Find the length of AX in terms of a, b, c.

18. If cevian AX of $\triangle ABC$ bisects $\angle A$, prove

$$AX^2 = bc\left(1 - \left(\frac{a}{b+c}\right)^2\right).$$

19. Find the length of the internal bisector of the right angle in a triangle with sides $3, 4, 5$.

20. Prove the product of two sides of a triangle is equal to the product of the circumdiameter and the altitude on the third side.

     For $\triangle ABC$ sides are labelled a, b, c opposite the vertices A, B, C respectively. Also,

$$R = \text{circumradius of } \triangle ABC$$
$$r = \text{inradius of } \triangle ABC$$
$$s = \frac{a+b+c}{2} = \text{semiperimeter of } \triangle ABC$$

21. Let I be the incentre of $\triangle ABC$, and let X, Y, Z be the feet of the perpendiculars dropped from I to sides BC, CA, AB, respectively, and let

$$x = AZ = AY,\ y = BX = BZ,\ z = CX = CY.$$

    Prove

$$x = s - a,\ y = s - b,\ z = s - c.$$

22. Prove $|ABC| = sr$.

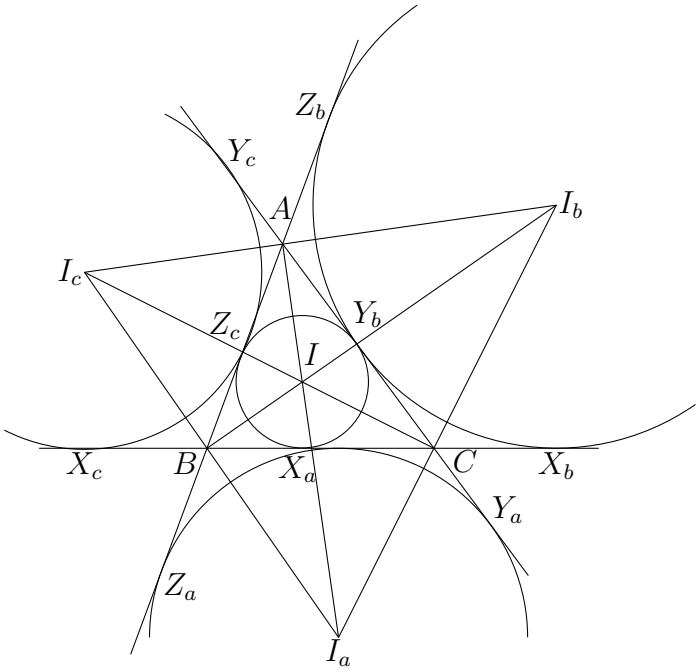
23. Prove the external bisectors of two angles of a triangle are concurrent with the internal bisector of the third angle.

24. If three circles with centres A, B, C all touch one another, show their radii are $s-a$, $s-b$, $s-c$, respectively.

25. Prove $abc = 4srR$.

26. If X, Y, Z are the feet of the perpendiculars dropped from the incentre I of $\triangle ABC$, as in Exercise 21., prove the cevians AX, BY, CZ concur. Their point of concurrence is called the **Gergonne point** of $\triangle ABC$.

If one produces the sides of a triangle beyond its vertices, external angles are formed at the vertices; so one can construct **external angle bisector**s. Each external angle bisector is perpendicular to the corresponding *internal angle bisector*. Let the intersections of the external angle bisectors of $\triangle ABC$ be I_a, I_b and I_c, labelled according to their being opposite vertices A, B and C, respectively.



27. Prove $\triangle ABC$ is the orthic triangle of $\triangle I_aI_bI_c$.
Note. If K, L, M are the feet of the altitudes of $\triangle DEF$ then $\triangle KLM$ is the **orthic triangle** of $\triangle DEF$.

28. Prove that each of I_a, I_b and I_c, are equidistant from the sides of $\triangle ABC$, and let these common distances be r_a, r_b and r_c, respectively.
Thus, r_a, r_b, r_c are **exradii** of $\triangle ABC$, i.e. radii of **excircles** of $\triangle ABC$ (circles touching sides of $\triangle ABC$ externally), with respective centres (**excentres**) I_a, I_b, I_c.
Thus prove $|ABC| = (s-a)r_a = (s-b)r_b = (s-c)r_c$.

29. Prove $\dfrac{1}{r_a} + \dfrac{1}{r_b} + \dfrac{1}{r_c} = \dfrac{1}{r}$.

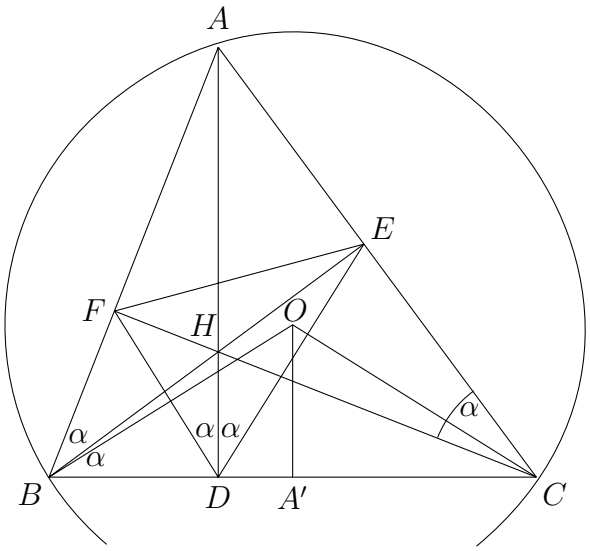
Part (ii) of the following theorem is known as the **Steiner-Lehmus Theorem**.

Theorem. *(i) If a triangle has two different angles then the smaller angle has the longer internal bisector.*

(ii) If a triangle has two equal-length (internal) angle bisectors then it is isosceles.

30. Let $\angle B = 12^\circ$ and $\angle C = 132^\circ$ of $\triangle ABC$. Without using trigonometric functions, compare the lengths of the external angle bisectors from B and C to their respective opposite sides (produced).

31. Prove the orthocentre of an acute-angled triangle is the incentre of its orthic triangle.
Hint. The diagram is of $\triangle ABC$ with circumcentre at O, feet of altitudes at D, E and F, and orthocentre at H. Also A' is the foot of the perpendicular dropped from O to BC. Start by showing the angles marked α in the diagram are all equal to $90^\circ - \angle A$, so that AD bisects $\angle EDF$.



32. Prove $\triangle AEF \sim \triangle DBF \sim \triangle DEC \sim \triangle ABC$.

33. Prove $\angle HAO = |\angle B - \angle C|$.

34. What is the minimum value that the power of a point can have, relative to a given circle of radius R? Which point has this power?

35. What is the locus of points of constant power relative to a given circle of radius R?

36. If the power of a point has the positive value t^2, interpret the length t geometrically.

37. Given PT and PU are tangents from P to two concentric circles, with T on the smaller circle, and PT meets the larger circle at Q, prove

$$PT^2 - PU^2 = QT^2.$$

38. Prove the circumradius of a triangle is at least twice the inradius.

39. Express in terms of the inradius r and circumradius R, the power of the incentre relative to the circumcircle of a triangle.

40. (AIMO 2013 Q8) A circle K meets equilateral triangle ABC in points D, E, F, G, H, I (in that order), where D, E are on CA, F, G are on AB and H, I are on BC.
Also, $AE = 4$, $ED = 26$, $DC = 2$, $FG = 14$ and a circle with diameter HI has area πb.
Find b.

41. Prove that the notation of directed segments enables us to express Stewart's Theorem in the following symmetrical form.

If P, A, B, C are four points of which the last three are *collinear*, then

$$PA^2 \cdot BC + PB^2 \cdot CA + PC^2 \cdot AB + BC \cdot CA \cdot AB = 0.$$

Directed segment convention. The product of two directed segments on one line is regarded as being positive or negative according as whether the directions are aligned or opposite.

42. A line through the centroid G of $\triangle ABC$ intersects the sides of the triangle in points X, Y, Z. Prove that
$$\frac{1}{GX} + \frac{1}{GY} + \frac{1}{GZ} = 0,$$
using the directed segment convention.

43. How far away is the horizon as seen from the top of a 2 km high mountain? (Assume the earth is a sphere of radius 6399 km.)

CHAPTER 13

Vectors

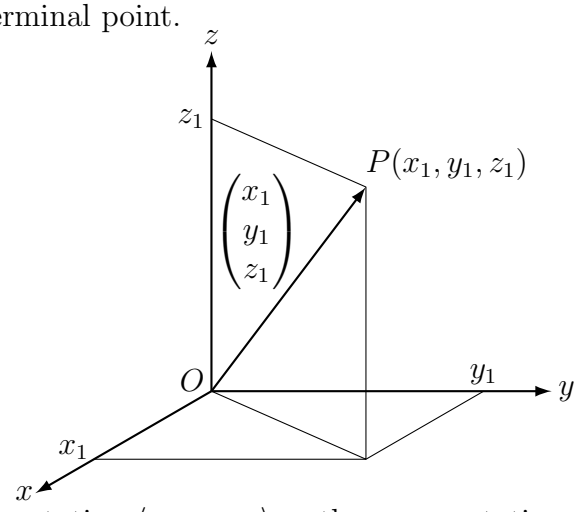
13.1 Introduction and notation

Definition 13.1.1. A **vector** is a quantity that has both *magnitude* and *direction*, e.g. wind described by speed and direction, such as, 20 km/h north-east, force, displacement. Essentially, a *vector* is a directed line segment. Geometrically, vectors are represented as *arrows;* the tail of the arrow is its initial point and the tip is its terminal point.

In $\mathbb{R}^3$ (3-space), the directed line segment from the origin $O(0,0,0)$ to the point $P(x_1, y_1, z_1)$ is the **position vector** of the point $P(x_1, y_1, z_1)$. The **vector**

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$$

denotes *any* directed line segment *parallel* (and pointing in the same direction) and having the same *length* as the *position vector* of the point (x_1, y_1, z_1). In this way, a *vector* has *magnitude* (length) and *direction,* but not *position.*



Many modern textbooks use the more compact notation $\langle x_1, y_1, z_1 \rangle$ or the same notation as for the point (x_1, y_1, z_1) for the vector

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix},$$

but the vertical notation has particular advantages when adding vectors or calculating *scalar* (*dot*) products (which we define later), since the components being added or multiplied are side-by-side.

For two points, A and B in space, the *vector* parallel to the directed line segment from A to B may be represented by $\overrightarrow{AB}$. The *position vector* of P above, is $\overrightarrow{OP}$. Another notation is to use a lower case letter with a tilde underneath; textbooks commonly instead embolden the letter; in these notes, we will do both together, e.g.

$$\underset{\sim}{\boldsymbol{u}} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

The above also demonstrates a convention; if a certain letter is used to represent a vector (in this case, u), then its x-, y- and z- components are that letter with subscripts 1, 2 and 3, respectively.

A lowercase letter without such adornments typically represents a quantity with magnitude, and except possibly for sign, not direction. Such non-vectors are called **scalar**s.


It will become important to distinguish *scalar* quantities from *vector*s, and hence important to remember the tildes!

Two vectors of the same length and direction (but not necessarily position) are **equivalent** (i.e. **equal**).

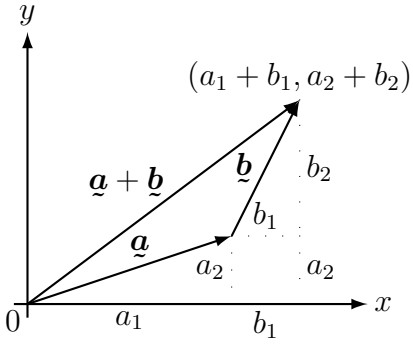
13.2 Vector addition

To add two vectors, we add their corresponding components:

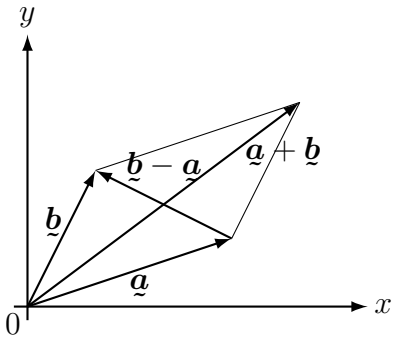
$$\underset{\sim}{a} + \underset{\sim}{b} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}.$$

We have demonstrated *vector addition* for $\mathbb{R}^2$ vectors, but the rule is similar with $\mathbb{R}^3$ vectors (three components), or more generally, with $\mathbb{R}^n$ vectors (with n components).

Geometrically, we can add vectors by placing them head to tail as shown. This is the **Triangle Law**.



The **Parallelogram Law** is another approach, where we place the tails of the vectors ($\underset{\sim}{a}, \underset{\sim}{b}$ here) at the origin. Then the vectors make two sides of a parallelogram. The fourth vertex of this parallelogram is at the tip of the sum vector ($\underset{\sim}{a} + \underset{\sim}{b}$ here). The other diagonal of the parallelogram gives the difference vector $\underset{\sim}{b} - \underset{\sim}{a}$. Also notice that the sum of vectors around a closed path is $\underset{\sim}{0}$ (the zero vector, which we will describe later).



13.3 Scalar multiplication

For $k \in \mathbb{R}$, $\underset{\sim}{a} \in \mathbb{R}^n$,

$$k\underset{\sim}{a} = k \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_n \end{pmatrix}$$

is a vector in the same direction as $\underset{\sim}{a}$, but of $|k|$ times the length of $\underset{\sim}{a}$, so that $k\underset{\sim}{a}$ is **parallel** to $\underset{\sim}{a}$ (written: $k\underset{\sim}{a} \parallel \underset{\sim}{a}$).

If $k < 0$ then $k\underset{\sim}{a}$ is oppositely directed to $\underset{\sim}{a}$ (and we can impart this extra information by saying $k\underset{\sim}{a}$ is **antiparallel** to $\underset{\sim}{a}$).

13.4 Length of a vector

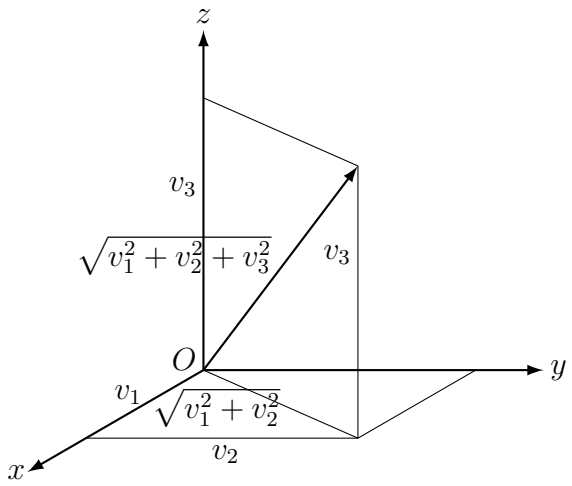
In general, if

$$\underset{\sim}{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

then the **length** of $\underset{\sim}{v}$ is

$$\|\underset{\sim}{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

The diagram shows the case for $n = 3$; and if you look in the (x, y)-plane also the $n = 2$ case, from which the extension to $n = 3$ is quite natural.



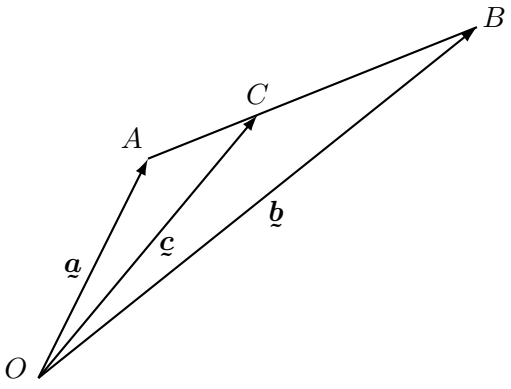
We write the length of a vector with "double bar" notation ($\|\cdot\|$). Most texts use "single bar" notation, so that it looks like *absolute value*. Sometimes, however, it is helpful to be able to distinguish the two notations, as with the next lemma.

Lemma 13.4.1. $\|k\underset{\sim}{v}\| = |k|\,\|\underset{\sim}{v}\|$.

The lemma says that the effect of multiplying a vector by a scalar k is to increase its length by a factor $|k|$.

Example 13.4.2. *Given C is a point on the line segment AB, twice as far from B as from A, write the position vector $\underset{\sim}{c} = \overrightarrow{OC}$ of C in terms of $\underset{\sim}{a} = \overrightarrow{OA}$ and $\underset{\sim}{b} = \overrightarrow{OB}$.*

Solution.



$$\begin{aligned} \underset{\sim}{a} + \overrightarrow{AB} &= \underset{\sim}{b} \\ \therefore \overrightarrow{AB} &= \underset{\sim}{b} - \underset{\sim}{a} \\ \overrightarrow{AC} &= \tfrac{1}{3}\overrightarrow{AB} \\ &= \tfrac{1}{3}(\underset{\sim}{b} - \underset{\sim}{a}) \\ \underset{\sim}{c} &= \underset{\sim}{a} + \overrightarrow{AC} \\ &= \underset{\sim}{a} + \tfrac{1}{3}(\underset{\sim}{b} - \underset{\sim}{a}) \\ &= \tfrac{2}{3}\underset{\sim}{a} + \tfrac{1}{3}\underset{\sim}{b}. \end{aligned}$$

Exercise 13.4.3. For triangle ABC, let the position vectors of the vertices be $\underset{\sim}{a}, \underset{\sim}{b}, \underset{\sim}{c}$, respectively. Find expressions in terms of these vectors for the midpoints of the sides, and the centroid G. Also, show that the medians trisect one another, in the sense that, if X is the midpoint of BC, then $AG : GX = 2 : 1$.

13.5 Properties

Suppose $\underset{\sim}{a}, \underset{\sim}{b}, \underset{\sim}{c} \in \mathbb{R}^n$ and $k, \ell \in \mathbb{R}$. Then the following properties hold.

1. $\underset{\sim}{a} + \underset{\sim}{b} = \underset{\sim}{b} + \underset{\sim}{a}$. (**commutativity of vector +**)

2. $(\underset{\sim}{a} + \underset{\sim}{b}) + \underset{\sim}{c} = \underset{\sim}{a} + (\underset{\sim}{b} + \underset{\sim}{c})$. (**associativity of vector +**)

3. There is a vector $\underset{\sim}{0}$ such that

$$\underset{\sim}{0} + \underset{\sim}{a} = \underset{\sim}{a} + \underset{\sim}{0} = \underset{\sim}{a}.$$

In $\mathbb{R}^n$,

$$\underset{\sim}{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and is called the **zero vector**, as opposed to the number 0 which is a **scalar**.

(existence of + **identity**)

4. For every $\underset{\sim}{a}$ there is a vector $\underset{\sim}{b}$ $(= -\underset{\sim}{a})$ such that

$$\underset{\sim}{b} + \underset{\sim}{a} = \underset{\sim}{0} = \underset{\sim}{a} + \underset{\sim}{b}.$$

Here

$$\underset{\sim}{b} = -\underset{\sim}{a} = \begin{pmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{pmatrix}.$$

(existence of + **inverses**)

5. $k(\underset{\sim}{a} + \underset{\sim}{b}) = k\underset{\sim}{a} + k\underset{\sim}{b}$. (**distribution**)

6. $(k + \ell)\underset{\sim}{a} = k\underset{\sim}{a} + \ell\underset{\sim}{a}$. (**distribution**)

7. $k(\ell\underset{\sim}{a}) = (k\ell)\underset{\sim}{a}$.

8. $1\underset{\sim}{a} = \underset{\sim}{a}$.

13.6 Special vectors

Definition 13.6.1. A **unit vector** is a vector of length 1. In general, a *unit vector* may be formed from any non-zero vector by dividing through by its length. We add a *caret* above the symbol for a vector to indicate "the unit vector formed from" that vector; thus

$$\hat{\underset{\sim}{a}} = \frac{1}{\|\underset{\sim}{a}\|}\underset{\sim}{a}.$$

The *unit vectors* in $\mathbb{R}^3$, in the x-, y- and z- directions, respectively, are

$$\underset{\sim}{\boldsymbol{i}} = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \underset{\sim}{\boldsymbol{j}} = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \underset{\sim}{\boldsymbol{k}} = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Lemma 13.6.2. $\hat{\underset{\sim}{\boldsymbol{v}}} = (c\underset{\sim}{\boldsymbol{v}})\hat{}$, *for any* $c > 0$.

The usefulness of the above lemma is that it tells us that we can clear ugly common factors first. This may greatly simplify the process of finding unit vectors.

Example 13.6.3. *1. Find the unit vector for* $\underset{\sim}{\boldsymbol{v}} = \begin{pmatrix}\frac{1}{12}\\ \frac{1}{15}\\ \frac{1}{20}\end{pmatrix}$.

Solution.

$$\hat{\underset{\sim}{\boldsymbol{v}}} = (60\underset{\sim}{\boldsymbol{v}})\hat{} = \begin{pmatrix}5\\4\\3\end{pmatrix}\hat{} = \frac{1}{\sqrt{50}}\begin{pmatrix}5\\4\\3\end{pmatrix} = \frac{1}{5\sqrt{2}}\begin{pmatrix}5\\4\\3\end{pmatrix}.$$

2. Find the unit vector for $\begin{pmatrix}\frac{1}{2\sqrt{2}}\\ -\frac{1}{2\sqrt{2}}\end{pmatrix}$.

Solution.

$$\begin{pmatrix}\frac{1}{2\sqrt{2}}\\ -\frac{1}{2\sqrt{2}}\end{pmatrix}\hat{} = \begin{pmatrix}1\\-1\end{pmatrix}\hat{} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

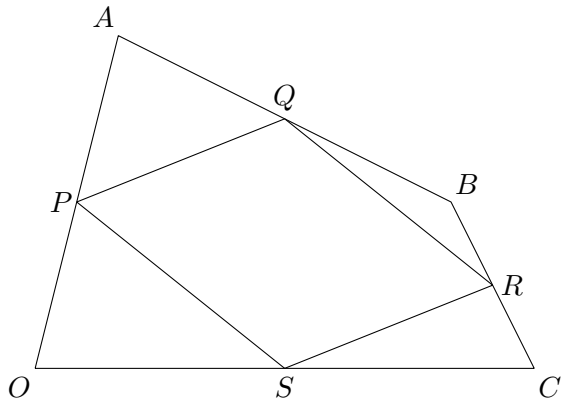
3. Find the unit vector for $\begin{pmatrix}123\\246\\246\end{pmatrix}$.

Solution.

$$\begin{pmatrix}123\\246\\246\end{pmatrix}\hat{} = \begin{pmatrix}1\\2\\2\end{pmatrix}\hat{} = \frac{1}{3}\begin{pmatrix}1\\2\\2\end{pmatrix}.$$

Example 13.6.4. *Prove that the midpoints of the sides of a quadrilateral form a parallelogram.*

Solution.



Let $\overrightarrow{OA} = \underset{\sim}{\boldsymbol{a}}$, $\overrightarrow{OB} = \underset{\sim}{\boldsymbol{b}}$, $\overrightarrow{OC} = \underset{\sim}{\boldsymbol{c}}$. *Then*

$$\begin{aligned}\overrightarrow{PA} &= \tfrac{1}{2}\underset{\sim}{\boldsymbol{a}}\\ \overrightarrow{AQ} &= \tfrac{1}{2}\overrightarrow{AB}\\ &= \tfrac{1}{2}(\underset{\sim}{\boldsymbol{b}} - \underset{\sim}{\boldsymbol{a}})\\ \overrightarrow{PQ} &= \overrightarrow{PA} + \overrightarrow{AQ}\\ &= \tfrac{1}{2}\underset{\sim}{\boldsymbol{a}} - \tfrac{1}{2}(\underset{\sim}{\boldsymbol{b}} - \underset{\sim}{\boldsymbol{a}})\\ &= \tfrac{1}{2}\underset{\sim}{\boldsymbol{b}}\end{aligned}$$

Similarly, $\overrightarrow{SR} = \frac{1}{2}\underset{\sim}{\boldsymbol{b}}$, *and so* $\overrightarrow{PQ} = \overrightarrow{SR}$.
Similarly, $\overrightarrow{PS} = \overrightarrow{QR}$.
Hence, $PQRS$ *is a parallelogram.*

13.7 The scalar product or dot product

Definition 13.7.1. The **scalar product** or **dot product** of two vectors $\underset{\sim}{a}, \underset{\sim}{b} \in \mathbb{R}^n$ is defined by

$$\begin{aligned}\underset{\sim}{a} \cdot \underset{\sim}{b} &= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \\ &= a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.\end{aligned}$$

Note that the result is a number, i.e. a *scalar* quantity.

13.7.1 Properties of the dot product

If $\underset{\sim}{a}, \underset{\sim}{b}, \underset{\sim}{c} \in \mathbb{R}^n$ and $k \in \mathbb{R}$, then

1. $\underset{\sim}{a} \cdot \underset{\sim}{a} = a_1^2 + a_2^2 + \cdots + a_n^2 = \|\underset{\sim}{a}\|^2$.

2. $\underset{\sim}{a} \cdot \underset{\sim}{b} = \underset{\sim}{b} \cdot \underset{\sim}{a}$. (**commutativity**)

3. $\underset{\sim}{a} \cdot (\underset{\sim}{b} + \underset{\sim}{c}) = \underset{\sim}{a} \cdot \underset{\sim}{b} + \underset{\sim}{a} \cdot \underset{\sim}{c}$. (**distribution of $\cdot$ over $+$ for vectors**)

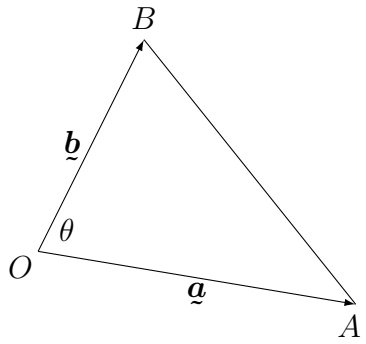
4. $(k\underset{\sim}{a}) \cdot \underset{\sim}{b} = k(\underset{\sim}{a} \cdot \underset{\sim}{b}) = \underset{\sim}{a} \cdot (k\underset{\sim}{b})$.

5. $\underset{\sim}{0} \cdot \underset{\sim}{a} = 0$.
Note that on the left there is the zero vector and on the right the zero scalar.

All these properties are straightforward to prove directly from the definition.

Theorem 13.7.1.1. *$\underset{\sim}{a} \cdot \underset{\sim}{b} = \|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta$, where θ is the angle between the vectors $\underset{\sim}{a}$ and $\underset{\sim}{b}$.*

Proof. Consider a triangle OAB.



By the Cosine rule,

$$\begin{aligned}|AB|^2 &= |OA|^2 + |OB|^2 - 2\,|OA|\,|OB| \cos\theta \\ \|\underset{\sim}{b} - \underset{\sim}{a}\|^2 &= \|\underset{\sim}{a}\|^2 + \|\underset{\sim}{b}\|^2 - 2\|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta \\ (\underset{\sim}{b} - \underset{\sim}{a}) \cdot (\underset{\sim}{b} - \underset{\sim}{a}) &= \underset{\sim}{a} \cdot \underset{\sim}{a} + \underset{\sim}{b} \cdot \underset{\sim}{b} - 2\|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta \\ \underset{\sim}{b} \cdot \underset{\sim}{b} - \underset{\sim}{a} \cdot \underset{\sim}{b} - \underset{\sim}{b} \cdot \underset{\sim}{a} + \underset{\sim}{a} \cdot \underset{\sim}{a} &= \underset{\sim}{a} \cdot \underset{\sim}{a} + \underset{\sim}{b} \cdot \underset{\sim}{b} - 2\|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta \\ -2\underset{\sim}{a} \cdot \underset{\sim}{b} &= -2\|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta \\ \underset{\sim}{a} \cdot \underset{\sim}{b} &= \|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\| \cos\theta.\end{aligned}$$

□

13.7.2 Further properties of the dot product

6. $|\underset{\sim}{a} \cdot \underset{\sim}{b}| = \|\underset{\sim}{a}\| \, \|\underset{\sim}{b}\|$ if and only if $\underset{\sim}{a} \parallel \underset{\sim}{b}$.

7. $\underset{\sim}{a} \cdot \underset{\sim}{b} = 0$ if and only if $\underset{\sim}{a} \perp \underset{\sim}{b}$.

CHAPTER 14

Analysis

14.1 Real Numbers

Our first notion of the real numbers, is that they are the numbers that can be located somewhere on the **number line**, but apart from that, real numbers may be combined, two at a time, via operations $+$ and $\cdot$ and have certain properties. Mathematicians, of course, like to collect together commonalities, distil those elements which are somehow basic, in the sense, they cannot be proved — these give rise to *axioms* — and organise what's left in a sequence of *theorems* that discover the properties of what it is that is under scrutiny. In this way, often one can predict structure that wasn't apparent before. And so, mathematicians came to define an **abelian group** and build from that a **field**.

Definition 14.1.1. An **abelian group** $(G, *)$ is a set G with a binary operation $*$ satisfying:

G1: $g, h \in G \implies g * h \in G$. (**closure**)

G2: $\forall g, h, k \in G, (g * h) * k = g * (h * k)$. (**associativity**)

G3: $\exists e \in G$ such that $\forall g \in G, e * g = g * e = g$. (identity)
An element with the property $e * g = g * e = g$ for all $g \in G$, is called an **identity** element of G (under $*$).
One can prove that when such an element e exists, then it is **unique**. So, let us write id_* for the unique element $e \in G$ satisfying $e * g = g * e = g$ for all $g \in G$.

G4: $\forall g \in G, \exists h \in G$ such that $g * h = h * g = \mathrm{id}_*$. (inverse)
Here id_* is the identity element determined in G3.
For $g \in G$, an element $h \in G$ s.t. $g * h = h * g = \mathrm{id}_*$ is called an **inverse** of g (in G, under $*$). In fact, when such an element h exists, one can show it is unique, and so it is *the* inverse of g (in G, under $*$), and we write g^{-1} for the unique element h s.t. $g * h = h * g = \mathrm{id}_*$.

G5: $\forall g, h \in G,\ g * h = h * g$. (**commutativity**)

The statements G1, ..., G5 are called the **abelian group axioms**.

When the operation $*$ is $+$ (addition), instead of id_+ we write 0 (**zero**), and when the operation $*$ is $\cdot$ (multiplication), instead of $\mathrm{id}_\cdot$ we write 1 (**one**).

Also, when the operation $*$ is $+$, the unique *additive inverse* of g is written as $-g$ (rather than g^{-1}, which would be confusing!), and while less confusing we'll usually use the *reciprocal* notation for the *multiplicative inverse* of g, i.e. $\frac{1}{g}$, when it exists.

Definition 14.1.2. A **field** $(F, +, \cdot)$ is a set F with binary operations $+$ and $\cdot$ satisfying:

F1: $(F, +)$ is an abelian group with identity 0,

F2: $(F \setminus \{0\}, \cdot)$ is an abelian group with identity 1, and $(F, \cdot)$ only fails to be an abelian group, in that 0 has no inverse, and

F3: $\forall x, y, z \in F,\ x \cdot (y + z) = x \cdot y + x \cdot z$. (**distributive law**)

Usually we write F instead of $(F, +, \cdot)$, where $+$ usually corresponds to ordinary *addition,* and $\cdot$ to ordinary *multiplication,* and we usually write xy rather than $x \cdot y$.

Remark 14.1.3. Implicit in axiom F2, is that $1 \neq 0$.

The Real Numbers $\mathbb{R}$ is a *field*.

It would be very boring if after developing such a theory of fields that there was just one example. In fact, the Rational Numbers $\mathbb{Q}$ and the Complex Numbers $\mathbb{C}$ are also fields, and there an infinite number of fields between $\mathbb{Q}$ and $\mathbb{C}$. All of these fields have an infinite number of elements, but aside from these there are also *finite fields*.

Let's write out all the rules we have identified for $\mathbb{R}$ again in expanded form, remembering that Rules 1–5 are F1 expanded, Rules 6–10 are F2 expanded and Rule 11 is F3.

14.2 Real Number Laws

For all $x, y, z \in \mathbb{R}$:

R1. $x + y \in \mathbb{R}$. (closure under $+$)

R2. $(x + y) + z = x + (y + z)$. (associativity under $+$)

R3. $0 + x = x + 0 = x$ and $0 \in \mathbb{R}$. (additive identity)

R4. There is an element $-x \in \mathbb{R}$ s.t. $x + (-x) = (-x) + x = 0$. (additive inverses)

R5. $x + y = y + x$. (commutativity under $+$)

R6. $x \cdot y \in \mathbb{R}$. (closure under $\cdot$)

R7. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$. (associativity under $\cdot$)

R8. $1 \cdot x = x \cdot 1 = x$ and $1 \in \mathbb{R}$. (multiplicative identity)

R9. If $x \neq 0$, there is an element $\frac{1}{x} \in \mathbb{R}$ s.t. $x \cdot \frac{1}{x} = \frac{1}{x} \cdot x = 1$. (multiplicative inverses)

R10. $x \cdot y = y \cdot x$. (commutativity under $\cdot$)

R11. $x \cdot (y + z) = x \cdot y + x \cdot z$. (distribution)

These field laws are not the end of the story. Another property of $\mathbb{R}$ is that it is an *ordered* field. Note that $\mathbb{C}$ is not an ordered field.

Definition 14.2.1. An **ordered field** is a field F with a binary relation $<$ satisfying:

OF1: $\forall x \in F,\ x \not< x$. (**irreflexivity**)

OF2: $\forall x, y, z \in F,\ x < y$ and $y < z \implies x < z$. (**transitivity**)

OF3: $\forall x, y \in F$, one of $x = y$, $x < y$, $y < x$ holds. (**weak trichotomy**)

OF4: $\forall x, y, z \in F,\ x < y \implies x + z < y + z$. (**compatibility w.r.t. addition**)

OF5: $\forall x, y, z \in F,\ x < y$ and $0 < z \implies xz < yz$. (**compatibility w.r.t. multiplication**)

In an ordered field F, we define $x > y$ to mean $y < x$. Similarly, $x \leqslant y$ means $x < y$ or $x = y$, and $x \geqslant y$ means $y \leqslant x$. An element $x \in F$ is called **positive** if $x > 0$, **negative** if $x < 0$, **nonnegative** if $x \geqslant 0$ and **nonpositive** if $x \leqslant 0$.

We express the 'compatibility with respect to addition' property by saying that adding any field element to an inequality **preserves** the inequality. Notice how the 'compatibility with respect to multiplication' property is a little different: multiplying an inequality by a *positive* field element **preserves** the inequality.

One can show that the axioms OF1–OF5 imply

OF3*: $\forall x, y \in F$, *exactly* one of $x = y$, $x < y$, $y < x$ holds. (**strong trichotomy**)

and that replacing OF1 and OF3 with OF3* one obtains an alternative Ordered Field axiom system.


Consider the sequence

$$3, 3.1, 3.14, 3.141, 3.1415, \ldots$$

containing decimal approximations of π to successively more decimal places. The **limit** of the sequence is π, but every number in the sequence is **rational**. The above sequence is an example of a **Cauchy sequence** which has the property that its elements get arbitrarily close to one another. If a set of numbers has the property that all *Cauchy sequences* have a limit, it is said to be **complete**; $\mathbb{R}$ is *complete* but $\mathbb{Q}$ is not.

We have only mentioned the operations $+$ and $\cdot$ on $\mathbb{R}$ so far. What about $-$ and $/$? We define **subtraction** as the addition of an *additive inverse:*

$$x - y := x + (-y).$$

Similarly, we define **division** as the multiplication by a *multiplicative inverse:*

$$x/y := x \cdot \frac{1}{y}.$$

Every field can be shown to have the following properties.

14.3 Properties of a field F.

1. $\forall x \in F$, $x \cdot 0 = 0$.
2. If $x, y \in F$ and $x \cdot y = 0$ then either $x = 0$ or $y = 0$.

We finish this section with an example of the consequences of the ordering rules.

Lemma 14.3.1. *If x and y are* positive *then*

$$x < y \iff x^2 < y^2.$$

Proof. We assume $x, y > 0$.
($\Longrightarrow$) Suppose $x < y$. Then

$$x^2 < xy, \qquad \text{by OF5, using } x > 0 \text{ and } x < y \tag{14.3.1}$$
$$\text{and } xy < y^2, \qquad \text{by OF5, using } y > 0 \text{ and } x < y \tag{14.3.2}$$
$$x^2 < y^2, \qquad \text{by OF3, using (14.3.1) and (14.3.2)}$$

($\Longleftarrow$) Now suppose $x^2 < y^2$. Then

$$x^2 - y^2 < 0$$
$$(x - y)(x + y) < 0$$
$$(x - y)(x + y)(x + y)^{-1} < 0, \qquad \text{since } x, y > 0 \implies x+y > 0, \text{ so that by Law 9., } (x+y)^{-1} \text{ exists, and then } 0 \cdot (x + y)^{-1} = 0 \text{ by Property 1.}$$
$$x - y < 0, \qquad \text{using Laws 9. and 8. to simplify}$$
$$x < y.$$

□

14.4 Absolute Value

Definition 14.4.1. For $x \in \mathbb{R}$, the **absolute value** of x, written $|x|$, is defined by

$$|x| = \begin{cases} x, & \text{if } x \geqslant 0, \\ -x, & \text{if } x < 0. \end{cases}$$

From this definition, it follows that in general

$$|x - a| = \begin{cases} x - a, & \text{if } x \geqslant a, \\ -x + a, & \text{if } x < a, \end{cases}$$

for any $a \in \mathbb{R}$.

A useful observation is that $|b - a|$ is the **distance** between $a, b \in \mathbb{R}$ on the number line. Also, for $a > 0$,

$$|x| < a \iff -a < x < a \iff (-a < x) \text{ and } (x < a)$$
$$|x| > a \iff (x < -a) \text{ or } (x > a).$$

Note, that without the condition $a > 0$, the above statements are somewhat vacuous, e.g. for $a < 0$, $|x| < a$ is equivalent to saying that x belongs to the empty set!

The *absolute value* function, satisfies the following properties:

$$|xy| = |x|\,|y|, \tag{14.4.1}$$
$$\big||x| - |y|\big| \leqslant |x + y| \leqslant |x| + |y|. \tag{14.4.2}$$

The inequality (14.4.2) is the extended version of the **Triangle Inequality**.

14.5 Subsets of the Real Numbers

Many subsets of $\mathbb{R}$ can be described as unions of *intervals*.

Definition 14.5.1. The *set* of all points x that satisfy the inequality $a < x < b$, for some constants $a, b \in \mathbb{R}$ is the **interval** between a and b which is represented by

$$(a, b).$$

The round brackets indicate that the endpoints a and b are *not* included, and (a, b) is said to be an **open** *interval*.

If the endpoints are included, i.e. we have $a \leqslant x \leqslant b$, then the *interval* is **closed** and we use square brackets:

$$[a, b].$$

An interval may also be half open and half closed (sometimes referred to as **clopen** intervals), e.g. the interval such that $a \leqslant x < b$, is represented as

$$[a, b).$$

In short, a square bracket indicates the point is included and a round bracket means it isn't.

Example 14.5.2. *1.* $(1,3) = \{x \mid 1 < x < 3\}$ *is an* open interval.

2. $[2,4] = \{x \mid 2 \leqslant x \leqslant 4\}$ *is a* closed interval.

3. $(3,5] = \{x \mid 3 < x \leqslant 5\}$ *and* $[3,5) = \{x \mid 3 \leqslant x < 5\}$ *are each* half-open *(or* clopen*)* intervals.

4. $[2,\infty) = \{x \mid x \geqslant 2\}$, $(2,\infty) = \{x \mid x > 2\}$, $(-\infty,2] = \{x \mid x \leqslant 2\}$, $(-\infty,2) = \{x \mid x < 2\}$ *are* infinite intervals.

5. $\{x \mid |x| \leqslant 1\} = [-1,1]$.

6. $\{x \mid |x| > 1\} = (-\infty,-1) \cup (1,\infty)$.

7. $\bigcup_{n=1}^{\infty} \left(n, n + \frac{1}{n}\right)$ *is an* infinite union of intervals *of decreasing size.*

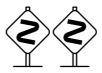
The set above is an example of an *open set,* a *set* of points with a "fuzzy" boundary (in the sense that all boundary points are missing). Adding in all missing boundary points gives a *closed set*. Infinite unions of *open sets* are again *open,* but infinite intersections of *open sets* need not be open. Similarly, infinite unions of *closed sets* need not be *closed,* but infinite intersections of *closed sets* are *closed*.

Note that ∞ *is* not a real number; *it represents something that is* larger *than any* real number. *Similarly,* $-\infty$ *is less than any* real number.

Infinite sets tend to have properties that seem paradoxical. The rationals $\mathbb{Q}$ is an infinite subset of $\mathbb{R}$ with the following properties:

(i) Between each pair of rationals there is an irrational.

(ii) Between each pair of irrationals there is a rational.



There are also different *infinities*. The size (*cardinality*) of $\mathbb{N}$ is what's called a *countable* infinity, denoted by $\aleph_0$ ("aleph null"). It turns out that the cardinality of each of $\mathbb{Z}$ and $\mathbb{Q}$ is also $\aleph_0$. This seems strange since $\mathbb{Z}$ has an infinite number of elements that are not in $\mathbb{N}$. Similarly, $\mathbb{Q}$ has an infinite number of elements that are not in $\mathbb{Z}$. On the other hand, $\mathbb{R}$ is a lot bigger than any of $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{Q}$. It is said to be *uncountable* and has cardinality denoted by $\aleph_1$ or c, and there are larger infinities! Things become somewhat counter-intuitive at the infinite level.

14.6 Functions

Definition 14.6.1. Given two sets X and Y, a **function** f is a rule that associates with each $x \in X$, an *unique* element $y \in Y$. In this case, we write $y = f(x)$, and say that y is the **value** or **image** of the function f at x. The set X is the **domain** of f; and the set Y is the **codomain** of f. In this case, we may write:

$$f : X \to Y$$

and say, f **maps** X **into** Y. The **image** of f, also called the **range** of f, is the set

$$\{f(x) \mid x \in X\}.$$

In general the *range* of f is a subset of Y. If the *range* of f equals Y, then we say that f **maps** X **onto** Y. We may write, $\mathrm{dom}(f)$ and $\mathrm{range}(f)$, for the *domain* of f and *range* of f, respectively. Also, for each $x \in X$ and $y \in Y$ such that $y = f(x)$, we say that f **maps** x **to** y, and we may write

$$f : X \to Y$$
$$x \mapsto y.$$

Note the slightly different shape of the right arrow for individual points: we write $f : X \to Y$ (map between sets) and $f : x \mapsto y$ (map between points).

Note that an element x can only be in the *domain* of f, if f has a *value* at x, i.e. if $f(x)$ is **defined**.

Example 14.6.2. *Real-valued functions of real variables are functions whose domain and codomain are both (subsets of)* $\mathbb{R}$*. Here are some examples.*

1. $f : \mathbb{R} \to \mathbb{R}$ *such that* $f(x) = 2x + 1$.
2. $f : [1, \infty) \to (0, \infty)$ *such that* $f(x) = \sqrt{x - 1}$.
3. $f : \mathbb{R} \to \mathbb{Z}$ *such that* $f(x) = \lfloor x \rfloor$*, where* $\lfloor x \rfloor$ *is the* ***floor*** *(or "integer part" of* x*), the* largest *integer that is* $\leqslant x$.

There is also the **ceiling** $\lceil x \rceil$ of x, which is the *smallest* integer that is $\geqslant x$. In general, for $x \in \mathbb{R}$, we have

$$\lfloor x \rfloor = x = \lceil x \rceil, \text{ if } x \in \mathbb{Z}, \text{ and}$$
$$\lfloor x \rfloor < x < \lceil x \rceil \text{ if } x \notin \mathbb{Z}.$$

For a real function f*, its* ***graph*** $\Gamma(f)$ *is the set of ordered pairs* $\big(x, f(x)\big)$ *for all* x *for which* f *is defined. Thus a real function's graph is a subset of* $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$*. Hence, in summary, we have*

$$\Gamma(f) = \{\big(x, f(x)\big) \mid x \in \mathrm{dom}(f)\} \subseteq \mathbb{R}^2.$$

This algebraic meaning of the graph $\Gamma(f)$ *of* f*, doesn't stop us from identifying a geometric interpretation, e.g. the graph* $\Gamma(f)$ *of* $f(x) = x^2, x \in \mathbb{R}$ *is a* parabola.

If f and g are two functions, then $f + g$, $f - g$, $f \cdot g$ and f/g are the functions defined **pointwise** (i.e. for each point x) by

$$(f + g)(x) = f(x) + g(x)$$
$$(f - g)(x) = f(x) - g(x)$$
$$(f \cdot g)(x) = f(x) \cdot g(x)$$
$$(f/g)(x) = f(x)/g(x)$$

on the domain $\mathrm{dom}(f) \cap \mathrm{dom}(g)$, except that $f(x)/g(x)$ is not defined whenever $g(x) = 0$, so that

$$\mathrm{dom}(f/g) = \big(\mathrm{dom}(f) \cap \mathrm{dom}(g)\big) \setminus \{x \mid g(x) = 0\}.$$

For two functions $g : X \to Y$ and $f : Y \to Z$, we can also define the **composition** $f \circ g$ of f and g with domain X by

$$(f \circ g)(x) = f\big(g(x)\big),$$

so that $f \circ g : X \to Z$.

More generally, if instead $f : U \to Z$ and U contains Y, the same definition for $f \circ g$ applies, but if $U \subset Y$, then we must find the largest **restriction** $g_{X'}$ defined on a subset X' of X such that $g_{X'} : X' \to U$. Then $f \circ g$ is defined to be the function $f \circ g_{X'} : X' \to Z$.

Note. When the domain of a real function f is not given explicitly, it is taken to be the the largest subset of $\mathbb{R}$ on which f can be defined.

Example 14.6.3. *1. $h(x) = \sin(2x^2+1)$ is a composite of $f(x) = \sin x$ and $g(x) = 2x^2+1$, with $\mathrm{dom}(h) = \mathrm{dom}(g) = \mathbb{R}$ and $\mathrm{range}(h) = \mathrm{range}(f) = [-1,1]$.*

Note that since $\mathrm{range}(g) \neq \mathrm{dom}(f)$, a little checking is necessary to see that $\mathrm{range}(h)$ is indeed all of $\mathrm{range}(f)$ here (observe that $[\pi/2, 3\pi/2] \subset \mathrm{range}(g)$ is enough to show this).


Note that once we have delved this far into Analysis, the trigonometric functions are defined with their arguments in *radians*. So we have $\sin(\pi/2) = 1$.

2. Let $f(x) = \sqrt{x}$ and $g(x) = 1 - x^2$. Then $\mathrm{dom}(f) = [0,\infty)$ and $\mathrm{dom}(g) = \mathbb{R}$. So $\mathrm{dom}(g \circ f) = [0,\infty)$, despite the fact that

$$(g \circ f)(x) = g\big(f(x)\big) = g(\sqrt{x}) = 1 - (\sqrt{x})^2 = 1 - x,$$

i.e. despite the fact that $h(x) = 1 - x$ can be defined with domain $\mathbb{R}$, we must respect where $g \circ f$ came from when determining its domain; its domain cannot be larger than $\mathrm{dom}(f)$.

On the other hand,

$$(f \circ g)(x) = f\big(g(x)\big) = f(1 - x^2) = \sqrt{1 - x^2},$$

is only defined for $1 - x^2 \geqslant 0$, i.e. for $-1 \leqslant x \leqslant 1$. So in this case, $\mathrm{dom}(g) = [-1,1]$ a proper subset of $\mathrm{dom}(g) = \mathbb{R}$. This possibility was foreshadowed in the dangerous bend that appeared before these examples. Putting that information in another way, we have that:

$$f \circ g \text{ is defined at } x \iff \begin{cases} g \text{ is defined at } x, \text{ and} \\ f \text{ is defined at } g(x). \end{cases}$$

14.7 Inverse functions

Definition 14.7.1. A function $f : X \to Y$ is said to be **one-to-one** (or **injective**) if for each $y \in \mathrm{range}(f)$ there is only one element $x \in X$ such that $f(x) = y$. Equivalently, f is *one-to-one* if and only if

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$


What we see here is a frequent strategy in mathematics: to prove *uniqueness* of something, take two of whatever it is that has the property and show that, in fact, they are the same.

In topology, for a function $f : X \to Y$ one defines

$$f^{-1}(y) = \{x \in X \mid f(x) = y\}.$$

So in the case, where y is not in the range of f, $f^{-1}(y)$ is the empty set, and in the case where there are many values of $x \in X$ such that $f(x) = y$, $f^{-1}(y)$ is a set containing many elements. To guarantee $f^{-1}(y) \neq \emptyset$, we need $\mathrm{range}(f) = Y$, i.e. f must be *onto*. If also, $f^{-1}(y)$ are always sets containing just one element, which is to say that f is *one-to-one*, then we may define a *function* from Y to X, by essentially stripping the curly braces from each singleton set $f^{-1}(y)$. It is customary, to use the same notation for such a function in Real Analysis. Thus we have the following definition.

Definition 14.7.2. For a function $f : X \to Y$ that is both *onto* and *one-to-one,* the **inverse function** $f^{-1} : Y \to X$ is defined by

$$f^{-1}(y) = x \iff f(x) = y.$$

A function that is *onto* is also said to be **surjective** (*sur* is *on* in French). A function that is both *onto* and *one-to-one* is thus both *surjective* and *injective* and so it won't surprise you too much that such a function is said to be *bijective,* and since we have just seen that these are precisely the functions that have *inverses,* such functions are also called *invertible* functions.

An **identity** function (on a set X) is a function $\mathrm{id} : X \to X$ such that $\mathrm{id}(x) = x$ for all $x \in X$, which is to say it returns what you give it unchanged.

Now, if $f : X \to Y$ is *invertible,* i.e. $f^{-1} : Y \to X$ exists, then $f^{-1} \circ f$ is the composite function that does

$$\begin{array}{ccccc} X & \xrightarrow{f} & Y & \xrightarrow{f^{-1}} & X \\ x & \mapsto & y & \longmapsto & x \end{array}$$

in one step, i.e. $f^{-1} \circ f : x \mapsto x$ for all $x \in X$, so that $f^{-1} \circ f$ is the *identity* function on X. Similarly, $f \circ f^{-1}$ is the composite function that does

$$\begin{array}{ccccc} Y & \xrightarrow{f^{-1}} & X & \xrightarrow{f} & Y \\ y & \longmapsto & x & \mapsto & y \end{array}$$

in one step, i.e. $f \circ f^{-1} : x \mapsto x$ for all $y \in Y$, so that $f \circ f^{-1}$ is the *identity* function on Y.

Conversely, if there is a function g such that $g \circ f : x \mapsto x$ for all $x \in X$ and $f \circ g : y \mapsto y$ for all $y \in Y$ then $g = f^{-1}$ and $f = g^{-1}$.

Example 14.7.3. *1. Let $f : [0, \infty) \to [0, \infty)$ such that $f(x) = x^n$ where $n \in \mathbb{N}$. Then f has an inverse and $f^{-1} = x^{\frac{1}{n}}$.*

2. Let $f : [0, \infty) \to (0, 1]$ such that

$$f(x) = \frac{1}{x^2 + 1}.$$

Then f is invertible with $f^{-1}(y) = \sqrt{\frac{1}{y} - 1}$.

3. If $f(x) = 3x - 2$ then $f^{-1}(y) = \frac{1}{3}(y + 2)$.

To find a formula for f^{-1} when it exists, replace $f(x)$ by y and, if possible, isolate y. Generally, an explicit formula for f^{-1} can only be found if a composition of inverse functions be applied in order to isolate y.

If f and g are the inverses of one another then their graphs $\Gamma(f)$ and $\Gamma(g)$ are reflections of one another in the line $y = x$, e.g. sketch $f(x) = x^2, x \in [0, \infty)$ and $g(x) = \sqrt{x}$.

14.8 Indices and Logarithms

Suppose that $x = p/q \in \mathbb{Q}$, where $p, q \in \mathbb{Z}$ and $q > 0$, and $a > 0$. Then we may define a function $\exp_a$ by

$$\exp_a(x) := a^{\frac{p}{q}} = (a^p)^{1/q} = \sqrt[q]{a^p},$$

where for $b = a^p > 0$, $b^{1/q} = \sqrt[q]{b}$ is the *positive real* q^{th} *root* of b.

We can *extend* this definition to have $x \in \mathbb{R}$ by *continuity* — essentially this means for any sequence of rational numbers $p_1/q_1, p_2/q_2, \ldots$, with each of $q_1, q_2, \ldots > 0$, that approach a given $x \in \mathbb{R}$, the sequence $a^{p_1/q_1}, a^{p_2/q_2}, \ldots$ approaches some real number y — a^x is then defined to be y.

Now, why do we need $a > 0$? ...Well, if $a = 0$ then $\exp_a(x)$ is undefined for nonpositive x; and we avoid negative values of a since otherwise we encounter conflicts like the following:

- $a^{\frac{1}{3}}$ exists if we interpret this as the real cube root of a (then the value of $a^{\frac{1}{3}}$ is *negative* for negative a);
- $a^{\frac{2}{6}}$ ought to be interpreted as $(a^2)^{\frac{1}{6}}$ which is the positive 6^{th} root of the positive number a^2 (i.e. $a^{\frac{2}{6}}$ would necessarily be *positive* for negative a); and
- $\frac{1}{3} = \frac{2}{6}$.

So, insisting that $a > 0$ ensures that $\exp_a$ is *well-defined*.

The **index laws** which hold whenever their component parts are defined, are:

1. $a^x a^y = a^{x+y}$,
2. $(a^x)^y = a^{xy}$, and
3. $a^x b^x = (ab)^x$.

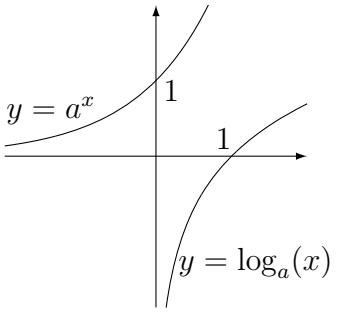
From these one may deduce the further rules:

4. $a^0 = 1$,
5. $a^1 = a$,
6. $a^{-x} = \dfrac{1}{a^x}$, and
7. $a^x / a^y = a^{x-y}$.

Now, the **logarithm** to **base** a of b, written $\log_a(b)$ is the power x that a must be raised to get b. For this definition to make sense we need $a > 0$ but $a \neq 1$ and $b > 0$. (Usually a is also an integer but this is not necessary.) More precisely then, for $1 \neq a > 0$ and $b > 0$,

$$\log_a(b) = x \iff a^x = b.$$

Note, that the $\log_a$ function is the inverse function of the $\exp_a$ function defined above.



 For $0 < a < 1$, $\log_a$ is a *decreasing* function. If we let $c = \frac{1}{a}$, then $c > 1$ and

$$\begin{aligned} \log_a(b) = x &\iff a^x = b \\ &\iff c^{-x} = \left(\frac{1}{a}\right)^{-x} = b \\ &\iff \log_c(b) = -x \\ &\iff -\log_c(b) = x \end{aligned}$$

i.e. $\log_a$ and $-\log_c$ are the same function, which is to say that we don't really get any new functions by considering $0 < a < 1$. So it is more usual to consider only the functions $\log_a$ for $a > 1$, which are all *increasing* functions. We needed to avoid $a = 1$, since

$$\exp_1(x) = 1^x = 1 \text{ for all } x,$$

i.e. $\exp_1$ is not *one-to-one* and hence is not invertible. Thus no function $\log_1$ can be defined.

What are commonly called the **log laws** follow from the first two *index laws* (by putting $b = a^x$ and $c = a^y$):

1. $\log_a(bc) = \log_a(b) + \log_a(c)$,
2. $\log_a(b^y) = y \log_a(b)$.

Properties equivalent to Rules 4–7. are similarly deduced:

$$\begin{aligned} \log_a(1) &= 0, \\ \log_a(a) &= 1, \\ \log_a\left(\frac{1}{b}\right) &= -\log_a(b), \\ \log_a\left(\frac{b}{c}\right) &= \log_a(b) - \log_a(c). \end{aligned}$$

Interpreting Index Law 2. differently, we obtain the **change of base rule**:

$$\log_b(c) = \frac{\log_a(c)}{\log_a(b)}.$$

Proof. We start with Index Law 2. and put $b = a^x$ and $b^y = c$. Then

$$\begin{aligned} (a^x)^y = b^y = c &= a^{xy} \\ \therefore \log_b(c) &= y \\ &= \frac{xy}{x} \\ &= \frac{\log_a(c)}{\log_a(b)}. \end{aligned}$$

□

With $c = a$, from the *change of base rule* it follows that:

$$\log_b(a) = \frac{1}{\log_a(b)}.$$

14.9 Limits

Limits are the essence of Analysis. Many properties that are "intuitively obvious" require careful application of limits to be proved formally. Newton and Leibniz knew they were on the right track when they defined *derivatives* in terms of their understanding of limits, but it wasn't until Cauchy came up with the formal definition of limit that all their notions could be proved to be correct.

Suppose f is a function defined on some open *neighbourhood* of $x = a$, i.e. f is defined at every point of an open interval containing the point $x = a$, except possibly a itself. Then we write $f(x) \to L$ as $x \to a$, or, equivalently,

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

and say, the function f has **limit** L at $x = a$, to mean that the value of $f(x)$ approaches L as x approaches a.

Intuitively, we can make $f(x)$ as close as we like to L, simply by choosing x close enough to a. This is the essential idea captured in Cauchy's formal definition:

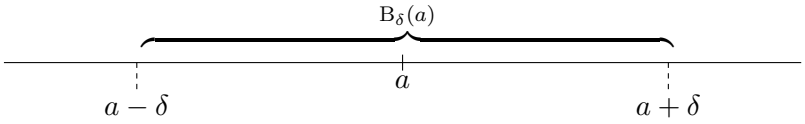
Definition 14.9.1. The function f has **limit** L at $x = a$ if

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0 \text{ such that } (0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon).$$

It's helpful to introduce *ball* notation here to simplify the statement.

If $a \in \mathbb{R}, \delta > 0$ then the δ-**neighbourhood** (abbreviated to δ-**nhd** of a, or the δ-**ball** centred at a is

$$\mathrm{B}_\delta(a) = (a - \delta, a + \delta) = \{x \in \mathbb{R} : |x - a| < \delta\}.$$



Also, the **punctured** δ-**neighbourhood** of a, or **punctured** δ-**ball** centred at a is

$$\mathrm{B}^*_\delta(a) = \mathrm{B}_\delta(a) \setminus \{a\} = \{x \in \mathbb{R} : 0 < |x - a| < \delta\} = (a - \delta, a) \cup (a, a + \delta).$$

Using the ball notation, we can write, f has limit L at $x = a$ if

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0 \text{ such that } f\big(\mathrm{B}^*_\delta(a)\big) \subseteq \mathrm{B}_\varepsilon(L)$$

which says, for all positive ε, as small as we like, there is a positive δ such that the punctured δ-neighbourhood of a maps under f entirely within an ε-neighbourhood of L.

By using the formal definition of *limit,* one can prove the following Limit Rules.

Algebra of Limits Laws. Suppose $f(x) \to L$ and $g(x) \to M$ as $x \to a$. Then:

1. $(f + g)(x) \to L + M$,
2. $(f - g)(x) \to L - M$,
3. $(f \cdot g)(x) \to L \cdot M$,
4. $\left(\frac{f}{g}\right)(x) \to \frac{L}{M}$, so long as $M \neq 0$.

Example. *1.* $\lim_{x\to 3} x^2 + 2x + 4 = 19$.

For polynomials there are no surprises. Limits can be evaluated by direct substitution.

2. Evaluate $\lim_{x\to 2} \dfrac{x^2-4}{x-2}$.

Solution.

$$\begin{aligned}\lim_{x\to 2} \frac{x^2-4}{x-2} &= \lim_{x\to 2} \frac{(x-2)(x+2)}{x-2} \\ &= \lim_{x\to 2}(x+2) \\ &= 4.\end{aligned}$$

□

Note that everywhere except at $x = 2$, we have:

$$\frac{(x-2)(x+2)}{x-2} = x + 2,$$

i.e. the expression $x - 2$ can be cancelled everywhere except at $x = 2$, where it is zero. However, we can get as close as we like to $2 + 2 = 4$ by getting sufficiently close to $x = 2$. Using the formal definition, one can prove this. This legitimises the cancelling of expressions such as the $(x - 2)$ here, when evaluating limits.

3. $\lim_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$.

This is a standard limit. One can show that for small θ that $\sin\theta \approx \theta$, and that the error in making this approximation is bounded by $\frac{1}{3}|\theta|^3$.

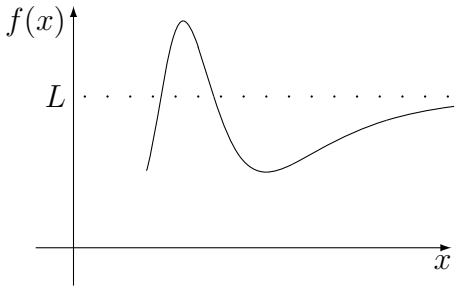
Note that as mentioned earlier, once one delves into Analysis, the trigonometric functions are defined to have arguments in radians.

14.10 Limits at Infinity

Given a function f defined for all large values of x, we write $f(x) \to L$ as $x \to \infty$, or equivalently,

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

to indicate that we can make the values of $f(x)$ as close to L as we like by making x sufficiently large.



As before there is a formal definition involving ε:

Definition 14.10.1. The function f has **limit** L **at infinity** if

$$\forall\, \varepsilon > 0\ \exists\, K \text{ such that } (x > K \implies |f(x) - L| < \varepsilon).$$

Now we said before that ∞ is not a real number. So it might surprise you that in certain circumstances we like to write:

$$\lim_{x\to a} f(x) = \infty.$$

Doesn't this mean that the limit doesn't exist? The answer is: Yes, but in a certain "regular" way, that we can use, and so it is convenient to have a short way to express it. For finite a, this notation is usually used for *one-sided* limits, i.e. instead of a one has either a^+ or a^- (where x approaches a from the *right* or *left*, respectively).

For now we will content ourselves with defining *infinite limits at infinity:*

Definition 14.10.2. The function f has **limit ∞ at infinity** if

$$\forall\, A > 0\ \exists\, K \text{ such that } (x > K \implies f(x) > A).$$

In this case, we write: $f(x) \to \infty$ as $x \to \infty$ or, equivalently: $\lim_{x\to\infty} f(x) = \infty$.

There are similar definitions for limits as $x \to -\infty$ and with limit $-\infty$. Many of the *Algebra of Limits Laws* still work with each of L, M or a equal to either ∞ or $-\infty$ except where the "limit" would be $\infty - \infty$ or ∞/∞; in these cases, *Algebra of Limits* is inconclusive, though some other approach may provide a conclusive result. We give some examples of the possibilities.

Example 14.10.3.

1. *If $f(x) \to L$, where L is finite, and $g(x) \to \infty$ as $x \to a$ then $(f+g)(x) \to \infty$ and $\left(\frac{f}{g}\right)(x) \to 0$ as $x \to a$.*
2. *If $f(x) \to \infty$ and $g(x) \to \infty$ as $x \to a$ then $(f-g)(x)$ may approach any finite or infinite limit as $x \to a$, e.g. consider the limits as $x \to \infty$ for $f(x) = x^2$ with each of the following possibilities for $g(x)$:*

   (i) $g(x) = x$ *(ii)* $g(x) = 2x^2$ *(iii)* $g(x) = x^2 - c$, $c \in \mathbb{R}$

3. *If $f(x) \to \infty$ and $g(x) \to \infty$ as $x \to a$ then $\left(\frac{f}{g}\right)(x)$ may approach any finite or infinite limit as $x \to a$, e.g. consider the limits as $x \to \infty$ for $f(x) = x^2$ with each of the following possibilities for $g(x)$:*

   (i) $g(x) = x$; *here* $\left(\frac{f}{g}\right)(x) \to \infty$ *as* $x \to \infty$;

   (ii) $g(x) = x^3$; *here* $\left(\frac{f}{g}\right)(x) \to 0$ *as* $x \to \infty$;

   (iii) $g(x) = 2x^2 - 1$; *here* $\left(\frac{f}{g}\right)(x) \to \frac{1}{2}$ *as* $x \to \infty$.

14.11 Continuous Functions

One would think the **continuous** functions ought to be those functions whose graphs one could draw without lifting one's pen from the page, and yes *continuous functions* have that property, but we define *continuous* in terms of *limits:*

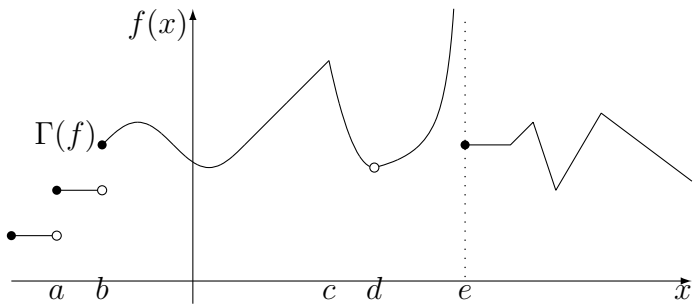
Definition 14.11.1. We say a function f is **continuous** *at a point* $x = a$ if

(i) the limit of f at $x = a$ exists, i.e. $\lim_{x\to a} f(x) = L$, say, and

(ii) $f(a) = L$.

We say a function f is **continuous** *in an open interval* I if f is *continuous* at every point in I. If $I = \mathbb{R}$ then we simply say f is *continuous.*

Consider the graph $\Gamma(f)$, below.



The function f has a **jump discontinuity** at each of $x = a$ and $x = b$. There is a hole in the graph at $x = d$, which can be removed by adding d to the domain of f and defining $f(d)$ to "fill in the hole", i.e. by defining $f(d) = \lim_{x\to d} f(x)$; the point at d is called a **removable singularity** of f. Redefining f in this way at d, makes f *continuous* at d. At $x = e$ there is an **essential singularity** of f since $f(x) \to \infty$ as x approaches e from below (written $x \to e^-$), whereas $f(x)$ approaches a finite number as x approaches e from above, i.e. $\lim_{x\to e^+} f(x)$ is finite. This exemplifies one way a function can fail to be continuous, i.e. when *left* and *right* limits exist at a point but are different.

Example 14.11.2. *1. Every polynomial function is continuous.*

2. The trigonometric functions sin *and* cos *are continuous.*

3. The floor and ceiling functions are discontinuous at each integer point; each discontinuity is a jump discontinuity.

If f and g are continuous at $x = a$, then from the Algebra of Limits Laws we can deduce that $f + g$, $f - g$ and $f \cdot g$ are continuous at $x = a$, and $\frac{f}{g}$ is continuous at $x = a$, if $g(a) \neq 0$.

Also, the composite function $f \circ g$ is continuous at a, if g is continuous at a and f is continuous at $g(a)$.

14.12 Bolzano's Theorem

Theorem 14.12.1 (Bolzano). *If f is continuous on $[a, b]$, and $f(a)$ and $f(b)$ have different signs then $\exists c \in (a, b)$ such that $f(c) = 0$.*

Example 14.12.2. *1. Every odd degree polynomial $p(x)$ has a zero, since*

$$\lim_{x\to\infty} p(x) \text{ and } \lim_{x\to-\infty} p(x)$$

are both infinite but of opposite sign.

2. For any $n \in \mathbb{N}$ such that $n > 1$, the real n^{th} root of 2 lies strictly between 0 and 2, since the polynomial $p(x) = x^n - 2$ is continuous on $[0, 2]$, and $p(0) = -2$ and $p(2) = 2^n - 2 > 0$.

14.13 Intermediate Value Theorem

Theorem 14.13.1 (Intermediate Value Theorem). *If f is continuous on $[a, b]$, $f(a) \neq f(b)$ and M lies between $f(a)$ and $f(b)$, then $\exists c \in (a, b)$ such that $f(c) = M$.*

The *Intermediate Value Theorem* follows from and generalises *Bolzano's Theorem.*

14.14 Extreme Value Theorem

Theorem 14.14.1 (Extreme Value Theorem). *If f is continuous on $[a, b]$ then $\exists c_1, c_2 \in [a, b]$ such that $f(c_1) \leqslant f(x) \leqslant f(c_2)\, \forall x \in [a, b]$.*

From this last result it follows that a function that is continuous on a closed interval attains both its maximum and minimum on that interval. The statement is false for open intervals, e.g. $f(x) = \frac{1}{x}$ with domain $(0, 1)$ has neither a minimum nor maximum in $(0, 1)$.

14.15 Sequences and Series

Definition 14.15.1. If for each $n \in \mathbb{N}$, $a_n \in \mathbb{R}$ then the infinite list of terms

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

which may be abbreviated to (a_n) or more explicitly $(a_n)_{n=1}^{\infty}$, is called a **sequence**.
Thus, the terms of a *sequence* are just the values $a_n = f(n)$ of a *function* f whose domain is $\mathbb{N}$.
The sequence (a_n) **converges** *(is convergent)* if $\exists L \in \mathbb{R}$ such that

$$\lim_{n\to\infty} a_n \left(= \lim_{n\to\infty} f(n)\right) = L.$$

We also write, in this case, $a_n \to L$ as $n \to \infty$.


There is a formal definition for the **limit** of a *sequence* that is similar to the formal definition of *limit* for a function:

A *sequence* (a_n) has **limit** L if

$$\forall\, \varepsilon > 0\, \exists\, N \in \mathbb{N} \text{ such that} (n > N \implies |a_n - L| < \varepsilon).$$

A sequence that is *not convergent,* is said to be **divergent**.

Example 14.15.2. *1. As* $n \to \infty$, $a_n = r^n \begin{cases} \to 0, & \text{if } |r| < 1 \\ \to 1, & \text{if } r = 1 \\ \text{diverges}, & \text{otherwise.} \end{cases}$

2. As $n \to \infty$, $a_n = \frac{1}{n^p} \begin{cases} \to 0, & \text{if } p > 0, \\ \to 1, & \text{if } p = 0, \\ \text{diverges}, & \text{if } p < 0. \end{cases}$

3. $a_n = \left(1 + \frac{x}{n}\right)^n \to e^x$ *as* $n \to \infty$.

A sequence (a_n) is said to be **increasing** if $a_{n+1} \geqslant a_n\ \forall n \in \mathbb{N}$, and to be **decreasing** if $a_{n+1} \leqslant a_n\ \forall n \in \mathbb{N}$; (a_n) is **monotone** if it is either *increasing* or *decreasing.*

Example 14.15.3. *1. If* $a_n = \frac{1}{n}$, *then* (a_n) *is* decreasing.

2. If $a_n = n^2 + n$, *then* (a_n) *is* increasing.

3. If $a_n = (-1)^n$, *then* (a_n) *is* alternating *(i.e. its sign alternates, for increasing* n*);* (a_n) *is* not *monotone.*

Definition 14.15.4. If for each $n \in \mathbb{N}$, $a_n \in \mathbb{R}$ then an infinite *sum* of terms

$$a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

which may be abbreviated to $\sum a_n$ or more explicitly $\sum_{n=1}^{\infty} a_n$, is called a **series**.

Of primary interest is when a *series* has a *finite* sum, i.e. *converges*. The *convergence* of a *series* is defined in terms of its **sequence of partial sums**, namely the sequence (s_n) where s_n is the n^{th} **partial sum** of $\sum a_n$, given by,

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

If s is finite and $s_n \to s$ as $n \to \infty$ then we say that $\sum a_n$ **converges** to s and write $\sum_{n=1}^{\infty} a_n = s$. If (s_n) diverges then $\sum a_n$ is said to **diverge**.

Theorem 14.15.5. *If $\sum a_n$ converges then $a_n \to 0$.*

Proof. Suppose the $\sum a_n$ converges then the partial sums s_n and s_{n-1} converge to the same limit L, say, as $n \to \infty$.
Hence $a_n = s_n - s_{n-1} \to L - L = 0$, as $n \to \infty$. □

The above theorem is usually used in its *contrapositive* form:

Theorem 14.15.6. *If $a_n \not\to 0$ then $\sum a_n$ does not converge.*

Example 14.15.7. *1. $\sum \frac{n^2}{2n^2+n}$ diverges since $a_n = \frac{n^2}{2n^2+n} \to \frac{1}{1}2 \neq 0$.*

> Testing whether $a_n \to 0$ is such an easy check, that it should be the *first* thing one checks for. Of course, if $a_n \to 0$ *no* conclusion may be made, unless the series has some other property, e.g. it's *geometric* (then it converges) or it's an *alternating series* (then it converges).

*2. If $a_n = ar^n$ then $\sum a_n = \sum_{n=1}^{\infty} ar^{n-1}$ is a **geometric series**, for which the n^{th}* partial sum *is*

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

which converges *to $\frac{a}{1-r} \iff |r| < 1$.*

*3. $\sum \frac{1}{n}$ is the **harmonic series**. It* diverges *since the $(2^{m+1})st$* partial sum

$$\begin{aligned} & 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \cdots + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^m+1} + \cdots + \frac{1}{2^{m+1}}\right) \\ \geqslant\ & 1 + \frac{1}{2} + \frac{1}{2} \qquad + \frac{1}{2} \qquad\qquad + \cdots + \frac{1}{2} \\ =\ & 1 + \frac{m+1}{2} \to \infty \text{ as } m \to \infty. \end{aligned}$$

*4. $\sum \frac{1}{n^p}$ (a p-**series**) converges $\iff p > 1$.*

> The *harmonic series* is the case $p = 1$. Essentially this result says that if the terms of a *p-series* grow more slowly (i.e. $p > 1$) than the *harmonic series* then it converges. Otherwise, the terms grow at least as fast as the *harmonic series* and the series diverges.

14.16 Comparison Tests

Theorem 14.16.1 (Comparison Test). *If for some $N \in \mathbb{N}$, $0 < a_n < b_n \ \forall n > N$, then*

(i) $\sum b_n$ *converges* $\implies \sum a_n$ *converges; and*

(ii) $\sum a_n$ *diverges* $\implies \sum b_n$ *diverges.*

Theorem 14.16.2 (Limit Comparison Test). *If* $\lim_{n\to\infty} \frac{a_n}{b_n}$ *exists, then* $\sum a_n$ *and* $\sum b_n$ *either* both converge *or* both diverge.

Example 14.16.3.

1. $\sum \frac{1}{2n!}$ *converges since* $\frac{2^n}{2n!} \to 0$ *and* $\sum \frac{1}{2^n}$ *is a convergent* geometric series.
2. $\sum \frac{1}{\sqrt{n(n+1)}}$ *diverges since* $\frac{n}{\sqrt{n(n+1)}} \to 1$ *and* $\sum \frac{1}{n}$ *(the* harmonic series*) diverges.*

14.17 The Ratio Test

Theorem 14.17.1 (Ratio Test). *Suppose* $\left|\frac{a_{n+1}}{a_n}\right| \to \rho$. *Then*

$$\sum a_n \begin{cases} \text{converges}, & \text{if } \rho < 1 \\ \text{diverges}, & \text{if } \rho > 1. \end{cases}$$

If $\rho = 1$ *the test is inconclusive.*

For a *geometric series* $\sum ar^{n-1}$, $\rho = r$. Essentially, the *Ratio Test* determines whether $\sum a_n$ behaves like a *geometric series.* The comparison is too weak if the ratio is 1, and so then no conclusion is possible, unless $\sum a_n$ *is* a *geometric series;* in which case, one can conclude immediately that $\sum a_n$ diverges. The lesson here is: don't use the *Ratio Test* if the series is a *geometric series,* since that knowledge already *conclusively* tells you whether or not the series converges.

Example 14.17.2. 1. $\sum \frac{n!}{n^n}$ *converges since*

$$\left|\frac{a_{n+1}}{a_n}\right| = \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n \to \frac{1}{e} < 1 \text{ as } n \to \infty.$$

2. $\sum \frac{n!}{2^n}$ *diverges since*

$$\left|\frac{a_{n+1}}{a_n}\right| = \frac{(n+1)!}{2^{n+1}} \cdot \frac{2^n}{n!} = \frac{n+1}{2} \to \infty \text{ as } n \to \infty.$$

We noted earlier that the *first* thing we should check is whether the n^{th} term a_n goes to 0. Here, $a_n \to \infty$. So, already we have that $\sum a_n$ diverges! Using the *Ratio Test* here is wasted effort. *Always* check the easy tests first.

14.18 Functional Equations

A **functional equation** is an equation involving an unknown function f and variables such as x and y.

In order to find a *solution* of a *functional equation* one must usually find *explicit* expressions $f(x)$.

Example 14.18.1. *1.* $f(xy) = y^k f(x)$

Solution. *Substituting $x = 1$, gives*

$$f(y) = y^k f(1) \ \forall y \in \mathbb{R}.$$

Letting $c = f(1)$, and changing the dummy variable to x, we have that solutions of 1. must be of form

$$f(x) = cx^k \text{ for some constant } c \ \forall x \in \mathbb{R}.$$

Now we check that all such f are solutions. Assume $f(x) = cx^k$ for some constant c. Then

$$\begin{aligned} f(xy) &= c(xy)^k \\ &= y^k c x^k \\ &= y^k f(x), \end{aligned}$$

as required.

2. $f(x + y) = f(y)$

Solution. *Substituting $y = 0$, we have*

$$f(x) = f(0).$$

Letting $c = f(0)$, we have that solutions of 2. must be of form

$$f(x) = c \text{ for some constant } c \ \forall x \in \mathbb{R}.$$

Now assume $f(x) = c$ for some constant c. Then

$$\begin{aligned} f(x + y) &= c \\ &= f(y). \end{aligned}$$

Thus the solutions of 2. are the constant functions: $f(x) = c$, c constant, for all $x \in \mathbb{R}$.

3. $f(x + y) = f(x) + f(y)$ *(Cauchy's equation)*

Solution. *Letting $y = x$ in 3. we obtain $f(2x) = f(x) + f(x) = 2f(x)$. Then letting $y = 2x$ in 3. we get*

$$f(3x) = f(x) + f(2x) = 3f(x). \tag{14.18.1}$$

Thus, it is apparent that by induction we can show that

$$f(nx) = nf(x) \ \forall n \in \mathbb{N}. \tag{14.18.2}$$

Indeed, defining the proposition

$$P(n) : f(nx) = nf(x),$$

we have $P(1) : f(x) = f(x)$, trivially, and $P(k) \implies P(k+1)$ is immediate, since (14.18.1) can be generalised, by replacing 2 by k and 3 by $k + 1$. Thus (14.18.2) follows.

Substituting $x = 1$, in (14.18.2), we have

$$f(n) = nf(1)\ \forall n \in \mathbb{N}.$$

Set $c = f(1)$. Then we have

$$f(n) = nc\ \forall n \in \mathbb{N}. \tag{14.18.3}$$

Now substituting $x = \frac{m}{n}$ (so that $nx = m$), where $m, n \in \mathbb{N}$, in (14.18.2), we have

$$\begin{aligned} nf(x) &= f(nx) \\ &= f(m) = mc, && \text{by (14.18.3)} \\ \therefore f(x) &= \frac{m}{n}c = xc\ \forall x \in \mathbb{Q}^+ && (14.18.4) \end{aligned}$$

Substituting $x = y = 0$ in 3. we have

$$\begin{aligned} f(0) &= f(0) + f(0) \\ \therefore 0 &= f(0). \end{aligned}$$

Thus, now substituting $y = -x$ in 3. we have

$$\begin{aligned} 0 &= f(0) \\ &= f(x + -x) \\ &= f(x) + f(-x) \\ \therefore f(-x) &= -f(x) \\ &= -(xc) = (-x)c \\ \therefore f(x) &= \frac{m}{n}c = xc, c \text{ constant } \forall x \in \mathbb{Q} \end{aligned}$$

If f is also continuous *for all $x \in \mathbb{R}$, then, by considering rational sequences that converge to each irrational $x \in \mathbb{R}$ (and taking limits), we have*

$$f(x) = xc, c \text{ constant } \forall x \in \mathbb{R}.$$

4. $f(x + y) = f(x)f(y)$

Solution. *Letting $x = y = 0$ in 4. we have*

$$\begin{aligned} f(0) &= f(0) \cdot f(0) \\ f(0)\big(1 - f(0)\big) &= 0 \\ \therefore f(0) &= 0 \text{ or } 1. \end{aligned}$$

Let $y = 0$ in 4. and suppose $f(0) = 0$. Then $f(x) = f(x)f(0) = 0$, i.e. $f(x) = 0$ for all $x \in \mathbb{R}$. Checking, we find $f(x) = 0$ satisfies 4.
Now let $f(0) = 1$.
Letting $y = x$ in 4. we obtain $f(2x) = f(x)f(x) = f(x)^2$. Then letting $y = 2x$ in 4. we get

$$f(3x) = f(x)f(2x) = f(x)^3.$$

Thus, by a similar induction to 3. we can deduce

$$f(nx) = f(x)^n\ \forall n \in \mathbb{N}. \tag{14.18.5}$$

Substituting $x = 1$, in (14.18.5), we have

$$f(n) = f(1)^n \ \forall n \in \mathbb{N}.$$

Set $a = f(1)$. Then we have

$$f(n) = a^n \ \forall n \in \mathbb{N}. \tag{14.18.6}$$

Now substituting $x = \frac{m}{n}$ (so that $nx = m$), where $m, n \in \mathbb{N}$, in (14.18.2), we have

$$\begin{aligned} f(x)^n &= f(nx) \\ &= f(m) = a^m, \qquad \text{by (14.18.6)} \end{aligned}$$

At this point, we would like to take the n^{th} root, but this is only well-defined if a^m (and therefore a) is positive.
So assume from now on that $a > 0$. Then

$$f(x) = a^{\frac{m}{n}} = a^x \ \forall x \in \mathbb{Q}^+.$$

Now substitute $y = -x$ in 4. we have

$$\begin{aligned} 1 &= f(0) \\ &= f(x + -x) \\ &= f(x) + f(-x) \\ \therefore f(-x) &= \frac{1}{f(x)} \\ &= \frac{1}{a^x} = a^{-x} \\ \therefore f(x) &= a^x, a \text{ a positive constant } \forall x \in \mathbb{Q} \end{aligned}$$

If f is also continuous *for all $x \in \mathbb{R}$, then, by considering rational sequences that converge to each irrational $x \in \mathbb{R}$ (and taking limits), we have*

$$f(x) = a^x, a \text{ a positive constant } \forall x \in \mathbb{R}.$$

Thus we see that if f is required to be continuous and have domain $\mathbb{R}$, then either

$$f(x) = 0 \ \forall x \in \mathbb{R} \quad \text{or} \quad f(x) = a^x, \ a > 0, \forall x \in \mathbb{R}.$$

Earlier we wrote $\exp_a$ for the function $f : x \mapsto a^x$, $a > 0$. All such functions can be written in terms of $\exp = \exp_e$,

$$\exp_a(x) = a^x = (e^{\ln a})^x = e^{cx} = \exp(cx), \ \text{where } c = \ln a.$$

5. $f(xy) = f(x) + f(y)$

Solution. *If f is defined at 0, then letting $x = y = 0$ in 5. we have*

$$\begin{aligned} f(0) &= f(0) + f(0) \\ 0 &= f(0). \end{aligned}$$

Now let $y = 0$ in 5. Then

$$0 = f(0) = f(x) + f(0).$$

Thus if 0 is in the domain of f, then

$$f(x) = 0 \ \forall x \in \mathbb{R}.$$

Checking, we see that this does indeed satisfy 5.

To get other solutions, we need $0 \notin \text{dom}(f)$. *If* $\text{dom}(f) = (0, \infty)$ *and* f *is continuous on this domain, then*

$$f(x) = c \log x, \text{ for constant } c \in \mathbb{R}.$$

If $\text{dom}(f) = (-\infty, 0)$ *with* f *is continuous on this domain, then*

$$f(x) = c \log |x|, \text{ for constant } c \in \mathbb{R}$$

is a family of solutions.

6. $f(xy) = f(x)f(y)$

Solution. *Let* $y = 0$ *in 6. Then*

$$\begin{aligned} f(0) &= f(x)f(0) \\ f(0)\big(1 - f(x)\big) &= 0 \\ \therefore f(0) = 0 \text{ or } f(x) &= 1 \ \forall x \in \mathbb{R}. \end{aligned}$$

Let $y = 1$ *in 6. Then*

$$\begin{aligned} f(x) &= f(x)f(1) \\ f(x)\big(1 - f(1)\big) &= 0 \\ \therefore f(1) = 1 \text{ or } f(x) &= 0 \ \forall x \in \mathbb{R}. \end{aligned}$$

Suppose that for some x, *say* $x = a$, *that* $f(a) \neq 0$ *and* $f(a) \neq 1$. *Then for* $y = a^n$, $n \in \mathbb{N}$, *we can show by induction that*

$$f(a^n) = f(a)^n \ \forall n \in \mathbb{N}.$$

Now suppose $x = e^n > 0$ *and suppose* $f(e) = b > 0$ *and* $b \neq 1$. *Then*

$$f(x) = f(e^n) = f(e)^n = b^n = (e^{\ln b})^n = (e^n)^{\ln b} = x^{\ln b}.$$

Thus, if we let $c = \ln b$ *then we have solutions of form*

$$f(x) = x^c, x = e^n > 0, n \in \mathbb{N}.$$

Arguments similar to those in 3.–5. with f *assumed continuous lead to*

$$f(x) = x^c, x > 0.$$

Checking, we find that each of the solutions we have found,

$$f(x) = 0 \ \forall x \in \mathbb{R}, f(x) = 1 \ \forall x \in \mathbb{R}, f(x) = x^c, c \in \mathbb{R} \ \forall x \in (0, \infty),$$

satisfy 6. The last of these can be extended to include 0 *in* $\text{dom}(f)$ *if* $c > 0$.

7. $f\big(\frac{x+y}{2}\big) = \frac{1}{2}\big(f(x) + f(y)\big)$ *(Jensen's equation)*

Solution. *Let* $f(0) = b$ *and* $f(1) = a + b$ *for some* $a, b \in \mathbb{R}$. *Then put* $y = 0$ *and consider* $x = n \in \mathbb{N}$, *and then arguments similar to those in 3. with continuity of* f *assumed, lead to:*

$$f(x) = ax + b, a, b \in \mathbb{R}, \ \forall x \in \mathbb{R}.$$

Exercise Set 14.

1. Determine all monotonic involutions.

   
   Note that **monotonic** is synonymous with *monotone*. An **involution** is a function f that satisfies
   $$f\big(f(x)\big) = x,$$
   for all x, i.e. a function that is its own inverse.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that the n^{th} iterate f^n has a unique fixed point x_0. Prove that f has a unique fixed point x_0.

3. (1994 AMOC Senior Contest Q5) Let $f : \mathbb{R} \to \mathbb{R}$ is a function such that for all $x \in \mathbb{R}$,

   (i) $f(x) \neq 0$, and
   (ii) $f(x+2) = f(x-1)f(x+5)$.

   Prove that f is **periodic**, i.e. there exists $P > 0$ such that $f(x+P) = f(x)$ for all x.

4. If $f : [0,1] \to [0,1]$ is continuous, prove that $f(x) = x$ for some x.

5. Prove there are 2 points on opposite sides of the Earth with exactly the same temperature.

6. Prove that $e^x = x^2$ has a real solution.

7. Find all $x \in \mathbb{R}$ such that $x + \sqrt{x-5} > 3$.

8. Solve $|x-a| + |x-b| = k$ for all real a, b, k.

9. Show that there is a solution $f(x) = 0$ between 0 and 1, given $f(x) = 4x^2 + x - 2$.

10. Find all $x \in \mathbb{R}$ such that $\sqrt[3]{x+2} + \sqrt[3]{x+3} + \sqrt[3]{x+4} = 0$.

11. Determine all real solutions of the following system of simultaneous equations:
    $$a^2 + b^2 = 6c,\ b^2 + c^2 = 6a,\ c^2 + a^2 = 6b.$$

12. Given a continuous function $f : [0,1] \to \mathbb{R}$ with $f(0) = f(1)$, for what d does there exist an x such that $f(x) = f(x+d)$.

13. Suppose $x + y + z = 5$ and $xy + yz + zx = 3$. Show that x, y, z lie between -1 and $\frac{13}{3}$.

14. Let $n \in \mathbb{N}$. Determine how many $x \in \mathbb{R}$ with $1 \leqslant x < n$ satisfy $x^3 = \lfloor x \rfloor^3 + \{x\}^3$, where $\{x\} = x - \lfloor x \rfloor$.

15. Find all $x \in \mathbb{R}$ such that $(x+2010)(x+2011)(x+2012)(x+2013) + 1 = 0$.

16. Let $f(x) = 5^x$. Find all $x \in \mathbb{R}$ such that $f\big(x + f(2008)\big) = 2008 - x$.

17. Find all functions $f : \mathbb{Q} \to \mathbb{R}$ such that

    (a) $f(x+y) = f(x) + f(y)$
    (b) $f(x+y) = f(x)f(y)$
    (c) $f(xy) = f(x) + f(y)$

18. Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f\big(x - f(y)\big) = 1 - x - y$ for all $x, y \in \mathbb{R}$.

CHAPTER 15

Invariants

Sometimes a problem can be easily solved once one has an **invariant**. Generally, an *invariant* is a *property* that is satisfied by a class of mathematical objects that remains unchanged when transformations of a certain type are applied to the objects.

In a problem that involves a number of steps, if one can show a certain property is satisfied at every step, often by induction, then one can often deduce a solution to a problem by checking satisfaction of the property at the last step.

The idea is best demonstrated with some examples.

1. (2006 Pólya) The first 1998 positive integers are listed on the blackboard. Two randomly selected numbers are erased and their difference is written on the board. The step is repeated until only one number is left on the board.

   Is the remaining number odd or even?

   Solution. Observe that we start with 999 odd numbers and 999 even numbers. Let us consider what cases can occur at each step.

   Case 1: Two even numbers are replaced. The replacement number is also even. So the number of even numbers has reduced by 1, and the number of odd numbers remains the same.

   Case 2: Two odd numbers are replaced. The replacement number is even. So the number of even numbers has increased by 1, and the number of odd numbers has reduced by 2.

   Case 3: One odd number and one even number are replaced. The replacement number is odd. So the number of even numbers has decreased by 1, and the number of odd numbers is the same as it was previously.

   Observe that in all cases, the number of odd numbers remains odd after each step, and of course the total number of numbers reduces by 1 at each step. So an *invariant* for each step of this problem is:

   The number of odd numbers is odd.

   Thus, after 1997 steps there will be one number left, and since after each step the number of odd numbers is odd, that last number remaining is odd.

2. (2009 AMO Q3) The polynomials $x^2 + x$ and $x^2 + 2$ are written on a white board. Sue is allowed to write on this board the sum, the difference or the product of any two polynomials already on the board. She repeats this process as many times as she likes.

   Can Sue ever write the polynomial x on the board?

   Solution. Define a polynomial $p(x)$ to have Property P if

   $$6 \,|\, p(2).$$

   Observe that for

   $$f(x) = x^2 + x \text{ and } g(x) = x^2 + 2,$$

   we have

   $$6 \,|\, f(2) \text{ and } 6 \,|\, g(2),$$

   since $f(2) = 6 = g(2)$. So each of the initial polynomials has Property P.

Lemma. *If f, g are on the board and $6 \mid f(2)$ and $6 \mid g(2)$ then $6 \mid (f+g)(2)$, $6 \mid (f-g)(2)$, and $6 \mid (f \cdot g)(2)$.*

Proof. This follows immediately from the definitions of $f+g$, $f-g$ and $f \cdot g$:

$$(f+g)(x) = f(x) + g(x)$$
$$(f-g)(x) = f(x) - g(x)$$
$$(f \cdot g)(x) = f(x) \cdot g(x)$$

So, for example, since $6 \mid f(2)$ and $6 \mid g(2)$, we have $6 \mid \big(f(2) + g(2)\big) = (f+g)(2)$. □

Put another way, the Lemma says:

> If two chosen polynomials on the board have Property P then the resultant polynomial has Property P.

Thus, it follows, by induction, that at every stage, every polynomial on the board has Property P, i.e. P is an *invariant*.
Suppose $h(x) = x$ is written on the board at some stage. Then it must have Property P.
But $h(2) = 2$ and $6 \nmid 2$. So $h(x)$ does not have Property P (a contradiction).
Thus Sue can never write x on the board.

3. (2009 AMO trial) In a latest theory of particle physics there are three fundamental particles. When two particles of different types collide they are replaced by a particle of the third type. Two particles of the same type never collide. Prove that if an experiment begins with an equal number of particles of each type it cannot end with just one particle remaining.

   Solution. We are done if we can find a property that is invariant over each step of the experiment. Observe that the numbers of two types of particles decreases by 1 and the other type increases by 1, so that the total number of particles decreases by 1 at each step. So the experiment proceeds for only a finite number of steps, terminating when there is only one type of particle remaining.

   Since at each step, the numbers of each type of particles changes by 1, the parity (i.e. what it is modulo 2) of the numbers of each type of particle changes at each step. If (ℓ, m, n) represents that there are ℓ, m, n of the three types of particle, respectively, then the initial state is (N, N, N) for some $N \in \mathbb{N}$. Thus initially, the numbers (and hence the parities of those numbers) of each particle type are the same, and since all the particle types swap parity at each step, we have an invariant property, namely that: the numbers of particles of each type have the same parity.

   At each step, the state parity toggles between (even, even, even) and (odd, odd, odd). Since $(0, 0, 1)$ (and any permutation) is of neither parity type, the experiment can never finish with just one particle.

Exercise Set 15.

1. (2012 TT Northern Autumn JO Q3) A 10×10 table is filled out according to the rules of the 'Minesweeper' game: each cell either contains a mine or a number that shows how many mines are in neighbouring cells, where cells are neighbours if they have a common edge or vertex.

   If all mines are removed from the table and then new mines are placed in all previously mine-free cells, with the remaining cells to be filled out with the numbers according to the 'Minesweeper' game rule as above, can the sum of all numbers in the table increase?

CHAPTER 16

Graph Theory

Graph Theory was invented in 1736, when Leonhard Euler solved the *Königsberg Bridge Problem* (see Exercise 19). In older texts, the diagram that Euler used to solve the problem was referred to as a *graph*; the more modern term (that we use) is *multigraph*. And what we refer to as a *graph*, in older texts was referred to as a **simple graph**.
Since the terminology is extensive, we start with a core set of definitions.

16.1 Introductory definitions

multigraph, G, consists of a pair V, E, where E is a set of unordered pairs of elements of V.

The elements of V are called **vertices** (singular: **vertex**).
The elements of E are called **edge**s.
An unordered pair of vertices of V, that is *not* an element of E is called a **non-edge**.

A *multigraph* is realised as a diagram by representing the *vertices* as points, with a line segment joining pairs of vertices u, v if and only if $\{u, v\}$ is an *edge*.

A *multigraph* may have more than one edge joining a pair of its vertices, but all its edges join distinct vertices, i.e. a *multigraph* has no *loops* (see below).

If $e = \{u, v\}$ (which may be abbreviated to uv or vu) is an *edge* of a graph G, then u and v are said to be **adjacent**, and u, v are called **endpoints** (or **ends**, or **terminals**) of e.

incident. If $e = uv$ is an *edge* then vertices u, v are said to be **incident** with e, and also e is said to be **incident** with each of u and v.

loop. An edge whose endpoints are *not* distinct.

pseudograph, G, consists of a *vertex* set V and *edge* set E, where E may contain *loops*.

graph. A *multigraph* (i.e. it has no *loops*) that has at most one *edge* joining any pair of vertices.

degree, of a vertex v, written $\partial(v)$, is the number of edges it is incident with.

order, of a graph G, is the cardinality of its *vertex* set.

size, of a graph G, is the cardinality of its *edge* set.

(p, q) **graph.** A graph of *order* p and *size* q, i.e. a graph with p *vertices* and q *edges*.

16.1.1 Convention when drawing graphs

Vertices are indicated by filled in dots. So, if two edges cross and there is *no* filled in dot at the crossing point of those edges, then the edges do not intersect and there is not a vertex at the crossing point of those edges.

subgraph. A *graph* H is a **subgraph** of a *graph* G, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$.

If e is an *edge* of G, then $G - e$ is the *subgraph* of G whose vertex set is $V(G)$ and whose edge set is $E(G) \setminus \{e\}$.

If v is a *vertex* of G, then $G - v$ is the *subgraph* of G whose vertex set is $V(G) \setminus \{v\}$ and whose edge set consists of all the edges of G except those incident with v.

complement. The **complement** $\overline{G}$ of a graph G is the graph with vertex set $V(G)$ and edge set, the set of *non-edges* of G.

The union of G and its complement $\overline{G}$ is the complete graph with $|V|$ vertices.

digraph, directed graph, D, consists of a pair V, E, where E is a set of *ordered* pairs of elements of V. As with a *graph*, V is the *vertex set* of D. The elements of E are called **arc**s (or **directed edge**s). An *arc* $e = (u, v)$ of D is represented by a line segment joining the pair of vertices u, v with an *arrow* in the direction from u to v.

even vertex. A *vertex* of *even degree.*

odd vertex. A *vertex* of *odd degree.*

leaf. A vertex of degree 1.

isomorphic. Two graphs G_1 and G_2 are **isomorphic** if there is a one-to-one map from $V(G_1)$ to $V(G_2)$ that preserves adjacency.

walk. A u-v **walk** is an alternating sequence $u = u_1, e_1, u_2, e_2, u_3, \ldots, u_n = v$ of vertices and edges of a graph such that vertices u_i and u_{i+1} are incident with edge e_i for $1 \leqslant i < n$.

closed walk. A u-v **walk** for which $u = v$.

trail. A u-v **trail** is a u-v *walk* that does not pass through the same *edge* twice.

path. A u-v **path** is a u-v *walk* that does not pass through the same *vertex* twice, except that possibly $u = v$.

circuit. A u-v *trail* with at least three vertices, for which $u = v$.

cycle. A *circuit* that does not repeat a *vertex,* except for the first and last vertices.

Euler trail, Euler tour. A *trail* of a multigraph G that passes through every edge of G *exactly once.*

Euler circuit. An *Euler trail* of a multigraph G that is a *circuit.*

Euler multigraph. A multigraph G for which an *Euler trail* exists.

Hamiltonian path. A u-v path that passes through each vertex of a multigraph exactly once, except that possibly $u = v$.

Hamiltonian circuit. A Hamiltonian path that is a circuit.

Hamiltonian multigraph. A multigraph G for which an *Hamiltonian path* exists.

regular. A graph is r**-regular** if every vertex is of *degree* r.

complete graph, K_n, is an $(n-1)$-*regular* graph, i.e. every pair of its vertices are adjacent.

connected. Two *vertices* u, v in a graph G are **connected** if $u = v$ or a u-v *path* exists in G.

A *graph* G is *connected* if every pair of vertices of G is *connected*.

component. A *subgraph* H of a *graph* G that is maximal with respect to the property of being *connected*, i.e. H is not contained in any connected subgraph of G having more vertices or edges.

cut-vertex, of a *connected graph* G is a *vertex* v of G, such that $G - v$ is *not connected*.

bridge, of a *connected graph* G is an *edge* e of G, such that $G - e$ is *not connected*.

tree. A *connected graph* that has no *cycles*.

forest. A *graph* that has no *cycles*, i.e. a *graph* whose *components* are *trees*.

spanning tree, of a *connected graph* G is a *subgraph* H of G that is a tree and such that $V(H) = V(G)$.

planar. A *graph* G is **planar** if one can draw G in the plane in such a way that no edges cross.

bipartite graph. A graph G for which it is possible to find a *partition* of its vertex set $V(G)$ into two sets X, Y such that every edge in $E(G)$ joins a vertex in X to a vertex in Y.

colouring, of a graph G is an an assignment of *colours* (i.e. labels) to the vertices of G such that each pair of adjacent vertices are assigned *different* colours.

An n**-colouring** is a *colouring* of a graph G using n *colours*.

chromatic number, χ**.** The **chromatic number** $\chi(G)$ of a graph G is the *minimum* value n for which an n-*colouring* of G exists.

length, of a *path* P of a graph (resp. digraph) is the number of *edges* (resp. *arcs*) in P.

distance, $d(u, v)$ is the *length* of the *shortest* u-v *path*.

diameter, $\operatorname{diam}(G) = \max_{u,v \in V(G)} d(u, v)$.

clique. A *subgraph* of a graph that is complete.

tournament. A *digraph* for which each pair of vertices u, v exactly one of (u, v) or (v, u) is an arc.

Exercise Set 16.

1. Prove:

   Lemma (Handshaking Lemma). *For a (p, q)-graph whose vertices have degrees $d_1, d_2, \ldots, d_p$ (called a **degree sequence**),*

   $$\sum_{i=1}^{p} d_i = 2q.$$

   i.e. the sum of the degrees is twice the number of edges.

2. Using the previous result, why must the number of odd-degree vertices be even?

3. Show there is no graph with vertices of degrees 2, 3, 3, 4, 4, and 5.

4. Show there is no graph with vertices of degrees 2, 3, 4, 4, and 5.

5. Show there is no graph with vertices of degrees 1, 3, 3, and 3.

6. Give an example of a graph that

   (i) has no vertex of odd degree.
   (ii) has no vertex of even degree.
   (iii) has 4 components, 6 vertices and 3 edges.
   (iv) is 3-regular but is not complete.

7. (a) Let $m, n \in \mathbb{Z}$ such that $1 \leqslant m \leqslant n$. Give an example of a graph with n vertices and m components.
   (b) Is it possible for a graph to have more components than vertices? Explain.

8. Show that a tree of order n has size $n - 1$.

9. Suppose you and your partner attend a party with 3 other couples. Several handshakes take place. No one shakes hand with themselves or their partner, and no one shakes hand with the same person more than once. After the handshaking is completed, you ask each person including your partner, how many hands they had shaken, and each person gave a different answer.

   (a) How many hands did you shake?
   (b) How many hands did your partner shake?

10. Show that in any group of at least two people, there are always two people with the same number of friends.

11. Let G be a graph of order n. Show that if every vertex of G has degree at least $\frac{1}{2}(n-1)$ then G is connected.

12. Out of 6 boys, exactly 2 were known to have stolen apples. But whom? Albert dobbed in Billy and Charlie. Billy said that Edward and Donald stole the apples. Charlie dobbed in Albert and Billy. Donald said that it was Edward and Fred. Edward said that Fred and Billy were the thieves. Fred couldn't be found. Four of the boys questioned had nominated one thief correctly, but lied about the other. The fifth had lied outright.
    Who stole the apples?

13. Prove that among 6 people there are either 3 mutual friends or 3 mutual strangers.

14. Is it possible that at a party with 777 attendees, each attendee knows exactly 77 other attendees?

15. Prove the distance relationship:

$$d(u, v) \leqslant d(u, w) + d(v, w), \text{ for all vertices } u, v, w.$$

16. Prove that any graph with n vertices and n edges has a cycle.

17. Several lines divide the plane into regions. Prove that each region can be coloured black or white so that any pair of adjacent regions are coloured differently.

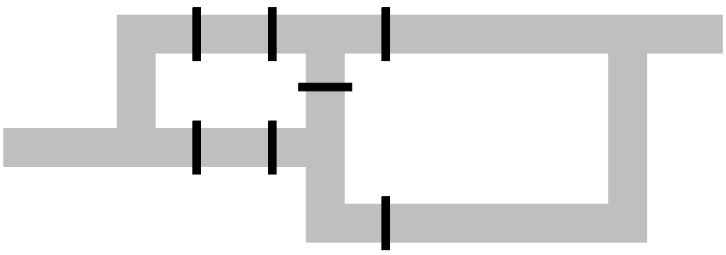
18. Prove that a connected graph is bipartite if and only if the graph has no odd cycle.

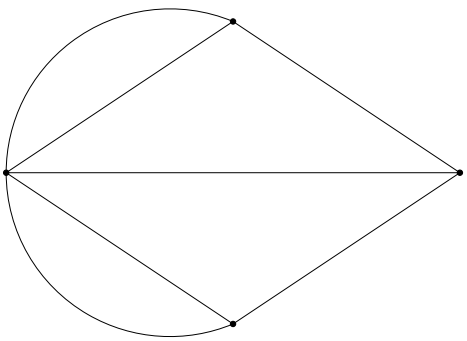
19. Prove:

> **Theorem (Euler circuit).** *A multigraph has an Euler circuit if every vertex has even degree.*
>
> In fact, the converse of the *Euler circuit theorem* is also true.

The above result was motivated by the **Königsberg bridge problem**:

> In the town of Königsberg in the 18th century there were seven bridges across the river Pregel. They connected two islands in the river with each other and with opposite banks. The townsfolk had long amused themselves with the problem:
>
> > Is it possible to cross the seven bridges in a continuous walk without recrossing any of them?





Leonhard Euler proved the above theorem in 1736, and hence solved the *Königsberg bridge problem*. Find Euler's solution.

The problem is equivalent to determining whether there is an Euler path for the following graph (each bridge is represented by an edge of the graph and the islands and banks of the river Pregel are represented by vertices of the graph).

APPENDIX A

Greek alphabet

uppercase	**lowercase**	**name**
A	α	alpha
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ε	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	mu
N	ν	nu
Ξ	ξ	xi
O	o	omicron
Π	π	pi
P	ρ	rho
Σ	σ	sigma
T	τ	tau
Y	υ	upsilon
Φ	φ	phi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Index