1AC: Algebra 1
Lecture Notes

UNIVERSITY OF BIRMINGHAM

Simon Goodwin
s.m.goodwin@bham.ac.uk
Office: Watson 107
School of Mathematics
University of Birmingham

Semester 1, 2023–24

Contents

A bit of a warning

These notes should be the final version, though some changes may still be made. As some edits have been made recently it is likely that there are some typos and possibly some errors. Please let me know if you spot any typos, or anything that you think may be a mistake by emailing s.m.goodwin@bham.ac.uk. The version of these notes available via the 1AC Canvas pages will be updated with any changes.

Chapter 0

Notation

We begin with a little bit of notation about sets, which is also covered in other courses. *You can look through this quickly now, then look back to it if you need to later.*

Definition 0.1. A *set* is a collection of objects.
We write:

- $a \in A$ to mean a is an element of the set A;
- $a \notin A$ to mean a is not an element of A; and
- $A \subseteq B$ to mean that A is a subset of B; this means that all elements of A are elements of B.

Now we define some sets of numbers that you are familiar with.

Definition 0.2. (a) We write $\mathbb{N}$ for the set of *natural numbers*:

$$\mathbb{N} = \{1, 2, 3, \dots\}.$$

Note that, for this module, 0 is not a natural number, though sometimes we want to consider the set consisting of 0 and the natural numbers, and we use the notation

$$\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}.$$

You should be warned that there is no universal agreement about whether 0 should be considered as a natural number, and you will find that some books do include 0 in the set of natural numbers.

(b) We write $\mathbb{Z}$ for the set of *integers*:

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}.$$

(c) We write $\mathbb{R}$ for the set of *real numbers*. These are numbers that can be written using a decimal expansion.

(d) We write $\mathbb{Q}$ for the set of *rational numbers*. These are the real numbers that can be written as a fraction, so

$$\mathbb{Q} = \left\{ \frac{a}{b} \in \mathbb{R} : a \in \mathbb{Z} \text{ and } b \in \mathbb{N} \right\}.$$

(e) We write $\mathbb{C}$ for the set of *complex numbers*. These are expressions of the form $a+bi$, where i is a square root of -1, so

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}.$$

You may not have learned about the complex numbers yet, but don't worry if not, as we'll only use them in some examples that are not part of the syllabus.

We demonstrate some notation for sets by example. Sometimes we give a set by listing its elements. For example

$$\{1, 3, 6\}$$

is a set with three elements. We used similar notation for $\mathbb{N}$ and $\mathbb{Z}$ above. Another example, the set of positive real numbers, is neatly described by the notation

$$\{x \in \mathbb{R} : x > 0\}.$$

The colon : can be read as "such that", so the set above is the set of $x \in \mathbb{R}$ such that $x > 0$. We used similar notation to describe $\mathbb{Q}$. Sometimes a vertical line | is written rather than a colon :. Another example is

$$\{x \in \mathbb{Z} : x^2 < 5\} = \{-2, -1, 0, 1, 2\}.$$

You shouldn't write a colon instead of "such that" outside of sets though.

One piece of notation that we'll use and may be a bit different from what you've done before is that we often use a dot to denote multiplication of integers. For example we write $3 \cdot 5$ to denote 3 multiplied by 5, which is of course equal 15, so we have $3 \cdot 5 = 15$.

Chapter 1

Introduction to proofs and prime numbers

Before we start, I'd like to make some comments about the material in this course. It is likely to be quite different from the mathematics that you have seen before university. An emphasis in this course is put on setting out pure mathematics well and writing proofs, and this is the most important skill for you to develop during this course. Consequently, the course may seem challenging to begin with, but with some perseverance you will be able to grasp the topic and I hope you will enjoy it!

In this first chapter, we introduce some ideas about definitions, proofs and counterexamples, with some very interesting topics from the theory of prime numbers.

1.1 What is a prime number?

You all should know what a prime number is. As we are going to be proving theorems about prime numbers, we need to make sure that a prime number means the same thing to all of us. For example, we need to decide whether 1 is a prime number. Therefore, we need a *definition* of a prime number.

Definition 1.1. A natural number $n \in \mathbb{N}$ is a *prime number* if $n \neq 1$ and the only positive factors of n are 1 and n.

We immediately see that there is a problem with this definition, as we don't yet know what we mean by factors. So we better define this now.

Definition 1.2. Let $a, b \in \mathbb{Z}$. We say that a *is a factor of* b if there exists $z \in \mathbb{Z}$ such that $b = az$.
We write $a \mid b$ to mean that a is a factor of b, and $a \nmid b$ to mean that a is not factor of b.

Sometimes we say a *divides* b or b *is divisible by* a to mean the same thing as a is a factor of b.

For $a \neq 0$, we remark that saying that a is a factor of b is equivalent to $\frac{b}{a} \in \mathbb{Z}$; you may find it helpful to think of it in this way.

Examples 1.3. (a) $7 \mid 21$, because $21 = 7 \cdot 3$. (The dot here is a shorthand for multiplication.)

(b) $4 \nmid 19$, because if $19 = 4z$, then $z = \frac{19}{4}$, which is not an integer.

(c) Let $n \in \mathbb{N}$. Then $n \mid n$, because $n = n \cdot 1$.

Definition 1.1 may seem very formal, but it is important that we have an agreement of exactly what it means for a natural number to be prime. From now we do not argue about what a prime number is – the above definition gives the answer. In particular, 1 is not a prime number because the definition says that it is not. From the definition we can work out the first few prime numbers:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots.$$

Note that from Definitions 1.2 and 1.1, if $n \in \mathbb{N}$ with $n \neq 1$ is not prime, then there exist $a, b \in \mathbb{N}$ with $1 < a, b < n$ such that $n = ab$. We often refer to natural numbers greater than 1 that are not prime as composite; this is stated in the following definition.

Definition 1.4. A natural number $n \in \mathbb{N}$ is called a *composite number* if $n \neq 1$ and there exist $a, b \in \mathbb{N}$ with $1 < a, b < n$ such that $n = ab$.

1.2 Mersenne numbers and primes

We are going to look at an interesting sequence of numbers called the Mersenne numbers. They are named after a French monk and scholar, Father Marin Mersenne (1585–1647), who studied them. A *Mersenne number* is a number of the form $2^n - 1$ for some $n \in \mathbb{N}$. In the table below I have listed the first 10 Mersenne numbers.

n	1	2	3	4	5	6	7	8	9	10
$2^n - 1$	1	3	7	15	31	63	127	255	511	1023

Let us observe that for each value of n in the table:

- if n is prime, then $2^n - 1$ is prime; and
- if n is composite, then $2^n - 1$ is composite.

For example, 7 is prime and 127 is prime, and 9 is composite and $511 = 7 \cdot 73$ is composite. It is, therefore, tempting to guess that this is always the case. In mathematics, a guess based on some evidence is called a *conjecture*. So we have the following two conjectures.

Conjecture 1.5. *Let $n \in \mathbb{N}$. Suppose that n is prime. Then $2^n - 1$ is prime.*

Conjecture 1.6. *Let $n \in \mathbb{N}$. Suppose that n is composite. Then $2^n - 1$ is composite.*

It turns out that Conjecture 1.5 is not true. To check that it is not true we only need to find one value of n for which n is prime and $2^n - 1$ is not prime. If we consider the prime 11, then we see that

$$2^{11} - 1 = 2047 = 23 \cdot 89$$

is composite; remember that the dot is our shorthand for multiplication. This means that $n = 11$ is a *counterexample* to Conjecture 1.5, so that the conjecture is not true.

We would like to decide if Conjecture 1.6 is true. If we continued the table to include all n up to 30, then we find out that 23 and 29 are other counterexamples to Conjecture 1.5:

$$2^{23} - 1 = 8{,}388{,}607 = 47 \cdot 178{,}481 \qquad 2^{29} - 1 = 536{,}870{,}911 = 2{,}089 \cdot 256{,}999.$$

However, there is no natural number $n \leq 30$ such that n is not prime and $2^n - 1$ is prime. So the evidence suggests that the statement is true. In order to be absolutely sure we need to give a *proof*.

Proof of Conjecture 1.6. Since n is composite, there exists $a, b \in \mathbb{N}$ with $1 < a, b < n$ such that $n = ab$. Consider the identity

$$t^m - 1 = (t-1)(1 + t + t^2 + \cdots + t^{m-1}),$$

for $m \in \mathbb{N}$. We apply this with $t = 2^b$ and $m = a$ to get

$$\begin{aligned} 2^n - 1 &= 2^{ab} - 1 \\ &= (2^b)^a - 1 \\ &= (2^b - 1)(1 + 2^b + (2^b)^2 + \cdots + (2^b)^{(a-1)}) \\ &= (2^b - 1)(1 + 2^b + 2^{2b} + \cdots + 2^{(a-1)b}). \end{aligned}$$

Let $x = 2^b - 1$ and $y = 1 + 2^b + 2^{2b} + \cdots + 2^{(a-1)b}$, then $2^n - 1 = xy$. Since $1 < b < n$, we have $1 < x < 2^n - 1$, which also implies that $1 < y < 2^n - 1$. Hence, $2^n - 1$ is composite. □

Now that we have a proof of Conjecture 1.6, we know beyond any doubt that it is true and we can call it a *theorem*; which we state below.

Theorem 1.6. *Let $n \in \mathbb{N}$. Suppose that n is composite. Then $2^n - 1$ is composite.*

We can use the proof of Conjecture 1.6 to find factors of large numbers of the form $2^n - 1$. For example, to find factors of $32767 = 2^{15} - 1$ we can write

$$\begin{aligned} 2^{15} - 1 &= (2^5 - 1)(1 + 2^5 + 2^{10}) \\ &= (32 - 1)(1 + 32 + 1024) \\ &= 31 \cdot 1057. \end{aligned}$$

Alternatively, we can write

$$\begin{aligned} 2^{15} - 1 &= (2^3 - 1)(1 + 2^3 + 2^6 + 2^9 + 2^{12}) \\ &= (8 - 1)(1 + 8 + 64 + 512 + 4096) \\ &= 7 \cdot 4681. \end{aligned}$$

From these factorizations, we see that 7 must divide 1057, and we obtain

$$32767 = 7 \cdot 31 \cdot 151.$$

We can check that 151 is prime, so we have factorized 32767 as a product of prime numbers.

Although we know that not all numbers of the form 2^p-1 with p prime are prime, these numbers are still very interesting. Mersenne numbers that are prime are called *Mersenne primes*. At present, 51 Mersenne primes have been found and it is unknown whether there are infinitely many. The largest known prime number is the Mersenne prime $2^{82,589,933}-1$. It has 24,862,048 digits and was found by the Great Internet Mersenne Prime Search, see https://en.wikipedia.org/wiki/Great_Internet_Mersenne_Prime_Search.
It was discovered in December 2018, and if you were to write out this number on a long reel of paper taking 1cm for each digit, then it would reach from Birmingham to London and carry on all the way to Brighton!

You'll be able to learn more about Mersenne primes in the number theory course 3NT, which you can take in your third year. You may wonder how $2^{82,589,933}-1$ was proved to be prime. Given the size of this number, a naïve method of checking that it's prime would require about $2^{41 \text{ million}}$ steps. However, using a billion computers, each doing a billion operations per second, and doing this for 100 billion years would do about 2^{200} steps which wouldn't get very far! In the third year course 3NT Number Theory you'll be able to learn about the Lucas–Lehmer primality test, which can be used to check whether a Mersenne number 2^n-1 is prime in just n steps, and is the test that was used to check that $2^{82,589,933}-1$ is prime. Proving that it works involves some more advanced topics that you can learn in the number theory course. For now you can read a bit more about it at https://en.wikipedia.org/wiki/Lucas-Lehmer_primality_test

1.3 The infinitude of primes

Above we have seen that there are very large prime numbers, and you may suspect that there are infinitely many prime numbers. Below we state this as a theorem and prove it, so we know beyond any doubt that it is true. This proof was first given by Euclid in around 350BC, and is one of the most famous proofs in mathematics.

Theorem 1.7. *There are infinitely many prime numbers.*

Proof. Suppose for a contradiction that there are not infinitely many prime numbers. Then there are a finite number of primes and we can write down the finite list of prime numbers

$$p_1, p_2, \ldots, p_r.$$

Let

$$n = p_1 p_2 \ldots p_r + 1.$$

Then $n > p_i$ for all $i = 1, 2, \ldots, r$. Since our list gives all of the prime numbers, this means that n is not a prime number. Therefore, there is a factor d of n with $1 < d < n$. We choose $1 < d < n$ to be a factor of n that is as small as possible. Then d must be a prime number, because if c is a factor of d with $1 \leq c < d$, then c is also a factor of n that is smaller than d so must be equal to 1. Hence, $d = p_i$ for some $i = 1, 2, \ldots, r$.

Therefore, we have $\frac{n}{p_i} \in \mathbb{Z}$. But also we have

$$\begin{aligned} \frac{n}{p_i} &= \frac{p_1 p_2 \ldots p_r + 1}{p_i} \\ &= p_1 p_2 \ldots p_{i-1} p_{i+1} \ldots p_r + \frac{1}{p_i} \notin \mathbb{Z}. \end{aligned}$$

Thus, we have a contradiction, because $\frac{n}{p_i}$ cannot both be an integer and not an integer.

The only conclusion that we can draw is that our initial assumption that there are only finitely many prime numbers must be wrong. Therefore, there are infinitely many prime numbers as required. □

This proof above is an example of a *proof by contradiction*. We'll see more of these types of proofs later.

1.4 The distribution of the primes

We now know that there are infinitely many prime numbers, so we can ask how they are distributed in all the natural numbers. Calculations suggest that the prime numbers are more sparsely distributed as we look at larger numbers. For example, there are 25 primes between 1 and 100, 16 primes between 1000 and 1100 and only 6 primes between 1,000,000 and 1,000,100. To demonstrate this thinning out we prove that we can find an arbitrarily large gap between prime numbers. For the proof recall that for $n \in \mathbb{N}$, we define n *factorial* by

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

Theorem 1.8. *Let $n \in \mathbb{N}$. There exists a sequence of n consecutive natural numbers containing no primes.*

Proof. Let $m = n + 1$. We will show that none of the $n = m - 1$ consecutive integers

$$m! + 2, m! + 3, m! + 4, \ldots, m! + m$$

are prime. First consider,

$$\begin{aligned} m! + 2 &= (1 \cdot 2 \cdot 3 \cdot 4 \cdots m) + 2 \\ &= (2 \cdot 1 \cdot 3 \cdot 4 \cdots m) + 2 \\ &= 2 \cdot (1 \cdot 3 \cdot 4 \cdots m + 1). \end{aligned}$$

Therefore, 2 is a factor of $m! + 2$, so $m! + 2$ is not prime. Similarly,

$$\begin{aligned} m! + 3 &= (1 \cdot 2 \cdot 3 \cdot 4 \cdots m) + 3 \\ &= (3 \cdot 1 \cdot 2 \cdot 4 \cdots m) + 3 \\ &= 3 \cdot (1 \cdot 2 \cdot 4 \cdots m + 1). \end{aligned}$$

Therefore, 3 is a factor of $m! + 3$, so $m! + 3$ is not prime. In general, consider $m! + i$ where $2 \leq i \leq m$. We have

$$\begin{aligned} m! + i &= (1 \cdot 2 \cdots (i-1) \cdot i \cdot (i+1) \cdots m) + i \\ &= (i \cdot 1 \cdot 2 \cdots (i-1) \cdot (i+1) \cdots m) + i \\ &= i \cdot (1 \cdot 2 \cdots (i-1) \cdot (i+1) \cdots m + 1). \end{aligned}$$

Therefore, i is a factor of $m! + i$, so $m! + i$ is not prime.

Hence, we have found a sequence of n consecutive natural numbers containing no prime numbers. □

We finish this chapter by briefly mentioning some more advanced statements about the distribution of the primes.

The *prime number theorem* is a remarkable theorem that tells us approximately how frequently primes occur as we look at larger numbers. Roughly it says that if we pick $n \in \mathbb{N}$ at random near a large number N, then the probability that n is prime is about

$$\frac{1}{\log_e(N)}.$$

It was conjectured by Gauss in 1793 and proved by Hadamard and de la Valée Pousin in 1896. You may ask why does e show up here! You can find out more at
https://en.wikipedia.org/wiki/Prime_number_theorem

A much deeper question about the distribution of primes is whether there is a pattern to how prime numbers lie in all the natural numbers. This is perhaps the biggest open problem in number theory. The *Riemann hypothesis*, which was proposed by B. Riemann in 1859, is a conjecture which implies a lot about the distribution of the primes. Proving that the Riemann hypothesis is true is one of the seven Millennium problems proposed by the Clay Mathematics Institute and there is a $1,000,000 prize for its solution. You can find out more at
https://en.wikipedia.org/wiki/Riemann_hypothesis

Another really interesting problem about prime numbers is the *twin primes conjecture*. This conjecture predicts that there are infinitely many pairs $(p, p+2)$ where both p and $p+2$ are prime. So the conjecture is saying that although we know that the primes become more sparsely distributed as we look at larger numbers and we can find arbitrarily large gaps, there are still infinitely many prime twins that are as close together as possible. You can find out more at
https://en.wikipedia.org/wiki/Twin_prime.
The twin primes conjecture remains an open problem, though there has been some major progress recently. In 2013, Yitang Zhang announced a proof that for some integer N less than 70 million, there are infinitely many pairs of primes that differ by at most N. Subsequently, there has been a flurry of research activity by many mathematicians around the world and the bound for N has been reduced to 246.

1.5 Summary of Chapter 1

At the end of each chapter of these notes, I will summarize the material in the chapter by giving a list of learning aims for the chapter. These aims are more specific than the learning outcomes for the module and serve the same purpose of informing you what you should be able to do to demonstrate that you have understood the chapter. As the main goal of this chapter is to give some interesting proofs the list here is quite short.

By the end of this chapter you should be able to:

- understand and recall the definitions of a factor and of a prime number; and
- appreciate the methods in the proofs given in this chapter and be able to apply similar arguments to prove related statements.

Don't worry too much if the proofs in this chapter seems a little challenging at the moment, once we've covered a few more you will get better at them.

Chapter 2

The integers

In this chapter, we make a structured study of the properties of the integers. One of ours goals is to prove the fundamental theorem of arithmetic, which roughly says that any natural number can be factorized uniquely as a product of prime factors. A precise statement is given in Theorem 2.23. *If you are ever asked to state the fundamental theorem of arithmetic then you should write down the precise statement in Theorem 2.23.*

Before we can state and prove the fundamental theorem of arithmetic we have to cover some material on factors and prime numbers. Of particular importance is the Euclidean algorithm, which is given in Section 2.4 and Bézout's Lemma, which is Theorem 2.14. Later in Section 2.9 we give some rather nice consequences of the fundamental theorem of arithmetic.

2.1 Factors of integers

Recall that we defined factors of integers in Definition 1.2. Below we give some elementary lemmas about factors of numbers. (A lemma is a name for a "little theorem"; often lemmas are used in the proofs of theorems.) We demonstrate the first of these lemmas with a couple of examples; we give a last reminder that we use a dot to denote multiplication as is done in these examples.

Examples 2.1. (a) We have that $13 \mid 26$, because $26 = 2 \cdot 13$, and $13 \mid 78$ because $78 = 13 \cdot 6$.
Then $13 \mid 104 = 26 + 78$, because $104 = 13 \cdot 8 = 13 \cdot (2 + 6)$.

(b) We have that $7 \mid (-28)$, because $-28 = 7 \cdot (-4)$, and $7 \mid 91$ because $91 = 7 \cdot 13$.
Then $7 \mid 63 = -28 + 91$, because $63 = 7 \cdot 9 = 7 \cdot (-4 + 13)$.

These examples may seem a bit trivial, but the important point is that it gives us an idea of how to prove the following lemma.

Lemma 2.2. *Let $a, b, c \in \mathbb{Z}$. Suppose that $a \mid b$ and $a \mid c$. Then $a \mid (b + c)$.*

Proof. Since $a \mid b$, there exists $x \in \mathbb{Z}$ such that

$$b = ax. \tag{2.1}$$

Since $a \mid c$, there exists $y \in \mathbb{Z}$ such that

$$c = ay. \tag{2.2}$$

Adding (2.1) and (2.2) gives

$$b + c = ax + ay = a(x + y).$$

Let $z = x + y$. Then $z \in \mathbb{Z}$, because $x, y \in \mathbb{Z}$ and

$$b + c = az.$$

Therefore, $a \mid (b + c)$. □

You should notice how central the definition of being a factor is to the proof of Lemma 2.2. The proof starts by using the definition to write down what the hypothesis says. Then it ends with a sentence saying that the conclusion holds in terms of the definition. The important point that I'm hoping to make here is that we have to use definitions properly in proofs.

The next lemma collects together some more general properties of factors. The proof of this lemma is an exercise. You can use the proof of Lemma 2.2 as a guide on how to write your proof.

Lemma 2.3. *Let $a, b, c, k, l \in \mathbb{Z}$.*

(a) *Suppose that $a \mid b$ and $a \mid c$. Then $a \mid (kb + lc)$.*

(b) *Suppose that $a \mid b$ and $b \mid c$. Then $a \mid c$.*

(c) *Suppose that $a \mid b$ and $b \mid a$. Then $a = \pm b$.*

2.2 The division theorem

The division theorem should be very familiar to you. It essentially says that when we divide an integer by a positive integer we obtain a quotient and remainder. It may seem that we are being very formal here, but our statement is very clear and the proof confirms beyond doubt something we have believed for a long time.

Theorem 2.4. *Let $a \in \mathbb{Z}$ and $d \in \mathbb{N}$. Then there exist unique $q, r \in \mathbb{Z}$ such that*

$$a = qd + r \text{ and } 0 \leq r < d.$$

Proof. Let q be the largest integer such that $q \leq \frac{a}{d}$, and let $r = a - qd$. Then $a = qd + r$ and $r \geq 0$.
If $r \geq d$, then $\frac{r}{d} \geq 1$, so $q + 1 \leq q + \frac{r}{d} = \frac{a}{d}$. But q was chosen to be maximal such that $q \leq \frac{a}{d}$, so $r < d$.

So far, we have proved that there exist $q, r \in \mathbb{Z}$ such that

$$a = qd + r \text{ and } 0 \leq r < d.$$

We also need to prove uniqueness. Suppose that there are $q, q', r, r' \in \mathbb{Z}$ such that

$$a = qd + r = q'd + r'$$

and $0 \leq r, r' < d$. Then

$$r - r' = d(q' - q).$$

If $q' - q \neq 0$, then $|q - q'| \geq 1$, which implies that $|r - r'| = d|q - q'| \geq d$. But this is not possible, because $0 \leq r, r' < d$. Therefore, we must have $q' - q = 0$, so $q = q'$ and $r = r'$ too. This proves uniqueness. □

We say that q is the *quotient* and r is the *remainder* when a is divided by d.

We note that $r = 0$ precisely when d is a factor of a. In other words d is a factor of a if and only if the remainder when a is divided by d is zero.

Example 2.5. Let $a = 137$ and $d = 11$. Then by performing long division we see that

$$137 = 12 \cdot 11 + 5.$$

So $q = 12$ and $r = 5$ in this case.

2.3 Highest common factors

An important notion for us is the highest common factor of two integers, which is defined below.

Definition 2.6. Let $a, b \in \mathbb{Z}$.

(a) A *common factor* of a and b is an integer c such that $c \mid a$ and $c \mid b$.

(b) The *highest common factor* of a and b is the largest integer h that is a common factor of a and b and is denoted $h = \mathrm{hcf}(a, b)$; unless $a = b = 0$ and by convention we define $\mathrm{hcf}(0, 0) = 0$.

Note that if $a = b = 0$, then $\mathrm{hcf}(a, b)$ then any integer is a common factor of a and b, so there is no largest integer that is a common factor of a and b. This is why we need to make the convention that $\mathrm{hcf}(0, 0) = 0$. Although this makes the definition above look a bit more complicated, you won't really need to worry about it.

Also we note that clearly, $\mathrm{hcf}(a, b) = \mathrm{hcf}(b, a)$ and $\mathrm{hcf}(a, b) = \mathrm{hcf}(-a, b)$.

The highest common factor of a and b is sometimes called the *greatest common divisor* of a and b, and denoted $\gcd(a, b)$; in these notes we will always say highest common factor.

Example 2.7.

(a) Let $a = 28$ and $b = 42$. By working out all the factors of a and b we calculate that the common factors of a and b are

$$-14, -7, -2, -1, 1, 2, 7, 14.$$

Therefore, the highest common factor of a and b is

$$\mathrm{hcf}(a, b) = 14.$$

(b) Let $a = 27$ and $b = 20$. By working out all the factors of a and b we calculate that the only common factors of a and b are -1 and 1, so $\mathrm{hcf}(a, b) = 1$.

(c) Let $a \in \mathbb{Z}$, $b \in \mathbb{N}$ and suppose that $b \mid a$. Then $\mathrm{hcf}(a, b) = b$.

Below we define what it means for two integers to be coprime to each other.

Definition 2.8. Let $a, b \in \mathbb{Z}$. We say that a *is coprime to* b if $\mathrm{hcf}(a, b) = 1$.

Note that saying a is coprime to b is equivalent to saying b is coprime to a. We sometimes say that a *and* b *are coprime to each other* instead of a is coprime to b; and sometimes just a *and* b *are coprime*.

As usual we demonstrate this definition with some examples.

Examples 2.9. (a) Let $a = 168$ and $b = 205$. Then a is coprime to b.

(b) Let $p, q \in \mathbb{N}$ be distinct primes. Then p is coprime to q.

2.4 The Euclidean algorithm

Example 2.7 shows that it is easy to calculate $\mathrm{hcf}(a, b)$ when a and b are small, by working out all the factors of a and b. When a and b are large this becomes impractical. There is a more efficient way to calculate the highest common factors called the Euclidean algorithm. The following lemma is key for the Euclidean algorithm.

Lemma 2.10. *Let $a, b, q, r \in \mathbb{Z}$. Suppose that $a = qb + r$. Then*

$$\mathrm{hcf}(a, b) = \mathrm{hcf}(b, r).$$

Proof. Let c be a common factor of a and b. Then by Lemma 2.3(a), c is a factor of $r = a - qb$, and thus a common factor of b and r.

Now let c be a common factor of b and r. Then by Lemma 2.3(a), c is a factor of $a = qb + r$, and thus a common factor of a and b.

Therefore, a and b have the same common factors as b and r, and hence the same highest common factor. □

Before formally stating the Euclidean algorithm we demonstrate it with an example.

Example 2.11. We are going to use Lemma 2.10 to calculate the highest common factor of 1989 and 1508.

First, using Theorem 2.4, we write

$$1989 = 1 \cdot 1508 + 481$$

and use Lemma 2.10 to deduce that

$$\mathrm{hcf}(1989, 1508) = \mathrm{hcf}(1508, 481).$$

Next we write

$$1508 = 3 \cdot 481 + 65$$

and use Lemma 2.10 again to deduce that

$$\mathrm{hcf}(1508, 481) = \mathrm{hcf}(481, 65).$$

For the third step we write

$$481 = 7 \cdot 65 + 26$$

and deduce that

$$\mathrm{hcf}(481, 65) = \mathrm{hcf}(65, 26).$$

For the fourth step we write

$$65 = 2 \cdot 26 + 13$$

and deduce that

$$\mathrm{hcf}(65, 26) = \mathrm{hcf}(26, 13).$$

In the fifth step we write

$$26 = 2 \cdot 13 + 0,$$

so $13 \mid 26$, and thus

$$\mathrm{hcf}(26, 13) = 13.$$

Putting all this together we obtain

$$\mathrm{hcf}(1989, 1508) = \mathrm{hcf}(1508, 481) = \mathrm{hcf}(481, 65) = \mathrm{hcf}(65, 26) = \mathrm{hcf}(26, 13) = 13.$$

So we have calculated $\mathrm{hcf}(1989, 1508) = 13$.

We now give a formal description of the Euclidean algorithm.

Algorithm 2.12 (Euclidean Algorithm)**.**
Input: $a, b \in \mathbb{N}$ with $a \geq b$, and set $a_0 = a$, $a_1 = b$.

1**st step**: Find $q_1, a_2 \in \mathbb{Z}$ with

$$a_0 = a_1 q_1 + a_2 \quad \text{and} \quad 0 \leq a_2 < a_1.$$

If $a_2 = 0$, then we output $\mathrm{hcf}(a, b) = a_1$ and stop.
If $a_2 \neq 0$, then we proceed to the 2nd step.

2**nd step**: Find $q_2, a_3 \in \mathbb{Z}$ with

$$a_1 = a_2 q_2 + a_3 \quad \text{and} \quad 0 \leq a_3 < a_2.$$

If $a_3 = 0$, then we output $\mathrm{hcf}(a, b) = a_2$ and stop.
If $a_3 \neq 0$, then we proceed to the 3rd step.

$\vdots \qquad \vdots \qquad \vdots$

k**th step**: Find $q_k, a_{k+1} \in \mathbb{Z}$ with

$$a_{k-1} = a_k q_k + a_{k+1} \quad \text{and} \quad 0 \leq a_{k+1} < a_k.$$

If $a_{k+1} = 0$, then we output $\mathrm{hcf}(a, b) = a_k$ and stop.
If $a_{k+1} \neq 0$, then we proceed to the $(k+1)$th step.

We make two comments about this algorithm. First we note that in the kth step we can find the required $q_k, a_{k+1} \in \mathbb{Z}$ using Theorem 2.4. Second, we note that as $a_0 > a_1 > a_2 > \ldots$, we must eventually get $a_{k+1} = 0$ so that the algorithm does terminate.

We give another example of the use of the Euclidean algorithm in Example 2.15. Below we prove that the Euclidean algorithm does correctly calculate highest common factors. The idea of the proof is given by Example 2.11.

Theorem 2.13. *Let $a, b \in \mathbb{N}$ with $a > b$, and let h be the output of Algorithm 2.12. Then $h = \mathrm{hcf}(a, b)$.*

Proof. Suppose the algorithm terminates on the mth step, Then $h = a_m$. We have

$$a_{m-1} = a_m q_m,$$

so $a_m \mid a_{m-1}$ and $\mathrm{hcf}(a_{m-1}, a_m) = a_m$.

Consider the kth step for $1 \leq k < m$. We have

$$a_{k-1} = a_k q_k + a_{k+1},$$

so

$$\mathrm{hcf}(a_{k-1}, a_k) = \mathrm{hcf}(a_k, a_{k+1})$$

by Lemma 2.10.

We obtain the sequence of equalities:

$$\begin{aligned} h = a_m &= \mathrm{hcf}(a_{m-1}, a_m) \\ &= \mathrm{hcf}(a_{m-2}, a_{m-1}) \\ &\;\;\vdots \qquad\quad \vdots \\ &= \mathrm{hcf}(a_1, a_2) \\ &= \mathrm{hcf}(a_0, a_1) \\ &= \mathrm{hcf}(a, b). \end{aligned}$$

□

2.5 Bézout's lemma and the extended Euclidean algorithm

The next theorem gives an important property of highest common factors; it is often called Bézout's lemma.

Theorem 2.14 (Bézout's lemma)**.** *Let $a, b \in \mathbb{Z}$ and let $h = \mathrm{hcf}(a, b)$. Then there exist $x, y \in \mathbb{Z}$ such that*

$$h = xa + yb.$$

In other words h can be expressed as an integral linear combination *of a and b*

Before giving a proof, we demonstrate how we prove it by an example. The idea is to reverse the Euclidean algorithm.

Example 2.15. We use the Euclidean algorithm to calculate $\mathrm{hcf}(2681, 931)$.
In the first step we write

$$2681 = 2 \cdot 931 + 819 \tag{2.3}$$

In the second step we write

$$931 = 1 \cdot 819 + 112. \tag{2.4}$$

In the third step we write

$$819 = 7 \cdot 112 + 35 \tag{2.5}$$

In the fourth step we write

$$112 = 3 \cdot 35 + 7. \tag{2.6}$$

In the fifth step we write

$$35 = 5 \cdot 7.$$

Therefore, the Euclidean algorithm tells us that

$$\mathrm{hcf}(2681, 931) = 7.$$

Now we reverse the algorithm. First we rearrange (2.6) to obtain

$$7 = 112 - 3 \cdot 35.$$

Second we use (2.5) to substitute for 35, and obtain

$$\begin{aligned} 7 &= 112 - 3 \cdot (819 - 7 \cdot 112) \\ &= -3 \cdot 819 + 22 \cdot 112. \end{aligned}$$

Third we use (2.4) to substitute for 112, and obtain

$$\begin{aligned} 7 &= -3 \cdot 819 + 22 \cdot (931 - 819) \\ &= 22 \cdot 931 - 25 \cdot 819 \end{aligned}$$

Fourth we use (2.3) to substitute for 819, and obtain

$$\begin{aligned} 7 &= 22 \cdot 931 - 25 \cdot (2681 - 2 \cdot 931) \\ &= -25 \cdot 2681 + 72 \cdot 931. \end{aligned}$$

So we have written $7 = \mathrm{hcf}(2681, 931)$ in the form $x2681 + y931$ with $x, y \in \mathbb{Z}$, where $x = -25$ and $y = 72$.

Now we extract the method used in the example above to prove Theorem 2.14.

Proof of Theorem 2.14. We first note that it suffices to consider $a, b > 0$. Also we assume without loss of generality that $a \geq b$, otherwise we can swap a and b.

Let $a_0, a_1, a_2, \ldots, a_m$ and $q_1, q_2, \ldots, q_m$ be the sequences of natural numbers produced by the Euclidean algorithm, so $a_m = \mathrm{hcf}(a, b)$.

For each k we have the equation

$$a_k = a_{k-2} - q_{k-1}a_{k-1},$$

which we refer to as equation (k).

From equation (m) we have

$$h = a_m = a_{m-2} - q_{m-1}a_{m-1}.$$

Now using equation $(m-1)$ we can substitute for a_{m-1} to obtain

$$\begin{aligned} h &= a_{m-2} - q_{m-1}(a_{m-3} - q_{m-2}a_{m-2}) \\ &= -q_{m-1}a_{m-3} + (q_{m-1}q_{m-2} + 1)a_{m-2}. \end{aligned}$$

So we have written h in the form

$$h = x_{m-3}a_{m-3} + y_{m-2}a_{m-2},$$

where $x_{m-3}, y_{m-2} \in \mathbb{Z}$. Next we can use equation $(m-2)$ and substitute for a_{m-2} to write

$$h = x_{m-4}a_{m-4} + y_{m-3}a_{m-3},$$

where $x_{m-4}, y_{m-3} \in \mathbb{Z}$. Continuing in this way we eventually obtain an expression

$$h = x_0a_0 + y_1a_1,$$

where $x_0, y_1 \in \mathbb{Z}$. But by definition $a_0 = a$ and $a_1 = b$, so for $x = x_0$ and $y = y_1$ we have

$$h = xa + yb,$$

with $x, y \in \mathbb{Z}$ as required. □

The method that we used in Example 2.15 and have described in the proof of Theorem 2.14 to find x and y such that $\mathrm{hcf}(a, b) = xa + yb$ is known as the *extended Euclidean algorithm*. We do not state this algorithm formally in these notes as it is easiest to understand through examples.

Now we give an alternative proof of Theorem 2.14, which is in a sense more direct. It is a little bit complicated, so you may want to omit it on a first reading, but it is a really nice proof. For the proof we require the fact that a nonempty subset S of $\mathbb{N}$ contains a least element, i.e. there exists $n \in S$ such that $n \leq m$ for all $m \in S$.

Alternative proof of Theorem 2.14. Assume that $a, b \neq 0$ otherwise the result is trivial. Consider the set

$$S = \{ua + vb \in \mathbb{Z} : u, v \in \mathbb{Z}\}$$

of all integral linear combinations of a and b.

First we note that $S \cap \mathbb{N}$ is nonempty, as either a or $-a$ is in $S \cap \mathbb{N}$. Therefore, there is a least element of $S \cap \mathbb{N}$, which we denote by $h = xa + yb$, where $x, y \in \mathbb{Z}$. We are going to show that $h = \mathrm{hcf}(a, b)$.

First we show that $h \mid a$. By Theorem 2.4, there exist $q, r \in \mathbb{Z}$ with $a = qh + r$ and $0 \leq r < h$. Then

$$\begin{aligned} r &= a - qh \\ &= a - q(xa + yb) \\ &= (1 - qx)a - qyb \end{aligned}$$

Therefore, $r \in S$, because $1 - qx, -qy \in \mathbb{Z}$. If $r \neq 0$, then $r \in S \cap \mathbb{N}$, which is not possible as $r < h$ and h is the least element of $S \cap \mathbb{N}$. Hence, $r = 0$ and $a = qh$, so $h \mid a$.

Similarly, we can prove that $h \mid b$. Therefore, $h \mid a$ and $h \mid b$, so h is a common factor of a and b.

Next we prove that $h \geq c$ for any common factor $c \in \mathbb{N}$ of a and b. Suppose that c is a common factor of a and b. Then $c \mid h$, by Lemma 2.3(a). In particular, this means that $c \leq h$. □

In Examples 2.7(a) we saw that the common factors of 28 and 42 are $\pm 1, \pm 2, \pm 7$ and ± 14, so $\text{hcf}(28, 42) = 14$. So in this case each common factor is a factor of the highest common factor. This statement is true in general; it is a consequence of Theorem 2.14 and is very useful later. We call it a *corollary*, which is a theorem that follows easily from another theorem.

Corollary 2.16. *Let $a, b, c \in \mathbb{Z}$ and let $h = \text{hcf}(a, b)$. Suppose that c is a common factor of a and b. Then $c \mid h$.*

Proof. By Theorem 2.14, there exist $x, y \in \mathbb{Z}$ such that $h = xa + yb$. Then $c \mid h$ by Lemma 2.3(a). □

2.6 Least common multiples

Before we move on towards the fundamental theorem of arithmetic, we briefly consider least common multiples. This is related to highest common factors, and we define the least common multiple of two integers next.

Definition 2.17. Let $a, b \in \mathbb{Z}$.

(a) A *common multiple* of a and b is an integer m such that $a \mid m$ and $b \mid m$.

(b) The *least common multiple* of a and b is the smallest $l \in \mathbb{N}$ that is a common multiple of a and b and is denoted $l = \text{lcm}(a, b)$; unless one of a or b is equal to 0 and by convention we define $\text{lcm}(a, 0) = 0 = \text{lcm}(0, b)$.

We require the convention that $\text{lcm}(a, 0) = 0$, as a and 0 have no positive common multiples; similarly, we require the convention $\text{lcm}(0, b) = 0$.

The following lemma gives a relationship between highest common factors and least common multiples. The proof is left as an exercise, and you will require Bézout's lemma (Theorem 2.14) for the proof.

Lemma 2.18. *Let $a, b \in \mathbb{N}$. Then $\text{lcm}(a, b)\text{hcf}(a, b) = ab$.*

2.7 Primes and products

The next theorem is the key result that we require for our proof of the fundamental theorem of arithmetic in the next section.

Theorem 2.19. *Let $a, b \in \mathbb{Z}$ and $p \in \mathbb{N}$ be prime. Suppose that $p \mid ab$. Then $p \mid a$ or $p \mid b$.*

Proof. Let $h = \text{hcf}(p, b)$. Since p is prime, the only positive factors of p are 1 and p. Therefore, h must be either 1 or p. We consider these two cases separately.
Case 1: $h = p$. Then $p \mid b$.
Case 2: $h = 1$. Then by Theorem 2.14 there exist $x, y \in \mathbb{Z}$ such that

$$1 = xp + yb. \tag{2.7}$$

Multiplying (2.7) by a we obtain

$$a = axp + ayb = (ax)p + y(ab).$$

Now $p \mid p$, and $p \mid ab$. Therefore, $p \mid a$ by Lemma 2.3(a).

In both cases we have shown that $p \mid a$ or $p \mid b$, which proves the theorem. □

The corollary below is proved by repeated use of Theorem 2.19.

Corollary 2.20. *Let $a_1, a_2, \ldots, a_n \in \mathbb{Z}$ and let $p \in \mathbb{N}$ be prime. Suppose that $p \mid a_1 a_2 \ldots a_n$. Then $p \mid a_i$ for some $i = 1, 2, \ldots, n$.*

Proof. We have $p \mid (a_1 a_2 \ldots a_{n-1})a_n$ so by Theorem 2.19, either $p \mid a_1 a_2 \ldots a_{n-1}$ or $p \mid a_n$. If $p \mid a_n$, then we are done. Otherwise, using Theorem 2.19 again we see that $p \mid a_1 a_2 \ldots a_{n-2}$ or $p \mid a_{n-1}$. Continuing in this way we will eventually see that $p \mid a_i$ for some $i = 1, 2, \ldots, n$. □

2.8 The fundamental theorem of arithmetic

The purpose of this section is to state and prove the fundamental theorem of arithmetic, see Theorem 2.23. This theorem is really important even though it may not seem very exciting at the moment as you've probably known it is true for some time. In Section 2.9, we'll prove a couple of interesting consequences, and we'll also see its importance again in Chapter 3.

First we give an example showing how we can calculate a prime factorization of a natural number. This example gives the idea for how we prove Proposition 2.22, which says that any natural number can be factorized as a product of primes. A *proposition* is just another name for a theorem that we don't think is important enough to call a theorem.

Example 2.21. We calculate a prime factorization of 8658. First we take out the factor 2 to obtain $8658 = 2 \cdot 4329$. Next we see that 4329 is divisible by 3 and we have $4329 = 3 \cdot 1443$. Now we see that 3 is still a factor of 1443 and we have $1443 = 3 \cdot 481$. Finally, we see that $481 = 13 \cdot 37$, and 13 and 37 are prime. Hence, we obtain the prime factorization

$$8658 = 2 \cdot 3 \cdot 3 \cdot 13 \cdot 37.$$

The idea of the proof of Proposition 2.22 is that if $n \in \mathbb{N}$ is not prime, then we can find a prime factor p of n and continue by applying the same process to the quotient $\frac{n}{p}$.

Proposition 2.22. *Let $n \in \mathbb{N}$ with $n > 1$. Then there exist prime numbers $p_1, p_2, \ldots, p_k$ such that*

$$n = p_1 p_2 \cdots p_k.$$

Proof. If n is prime, then we take $k = 1$ and $p_1 = n$.
So suppose that n is not prime. Let d be a factor of n with $1 < d < n$ and d as small as possible. Then d must be prime, because if c is a factor of d with $1 \leq c < d$, then c is a factor of n that is smaller than d so must be equal to 1.
We set $p_1 = d$ and let $n_2 \in \mathbb{N}$ be such that $n = p_1 n_2$.
If n_2 is prime, then we can take $k = 2$ and $p_2 = n_2$ and we are done.
Otherwise, we can apply the argument above to n_2 in place of n and find a prime number p_2 and a natural number n_3 such that $n_2 = p_2 n_3$. Then $n = p_1 p_2 n_3$.
Continuing in this way, we get a sequence of prime numbers $p_1, p_2, p_3, \ldots$ and natural numbers $n = n_1 > n_2 > n_3 > \ldots$. Eventually, for some $k \in \mathbb{N}$ we must have that n_k is prime. Then we take $p_k = n_k$ and we have

$$n = p_1 p_2 \cdots p_k$$

is a factorization of n as a product of primes. □

At the start of this chapter, we said that the fundamental theorem of arithmetic roughly says that any natural number can be factorized uniquely as a product of primes. We have just proved that a natural number can be factorized as a product of primes, so now we need to work out what it means for this factorization to be unique. A first guess might be the following statement.
Let $n \in \mathbb{N}$. Then:

(a) *there exist prime numbers $p_1, p_2, \ldots, p_k$ such that $n = p_1 p_2 \cdots p_k$.*

(b) *if $q_1, q_2, \ldots, q_l$ are prime numbers such that $n = q_1 q_2 \cdots q_l$, then $l = k$ and $q_i = p_i$ for all $i = 1, \ldots, k$.*

If we think about this a little bit, then we can find a problem with this statement. Namely that there is nothing stopping us from reordering the prime factors. For example, consider the case $n = 6$. We have $6 = 2 \cdot 3$ and $6 = 3 \cdot 2$, so we can take $r = s = 2$, $p_1 = 2$, $p_2 = 3$, $q_1 = 3$ and $q_2 = 2$. Then $p_1 \neq q_1$. So $n = 6$ gives a counterexample to the statement above.

To deal with this problem we have to make sure we order the prime factors. This is done in our statement of the fundamental theorem of arithmetic below.

Theorem 2.23 (Fundamental theorem of arithmetic)**.** *Let $n \in \mathbb{N}$ with $n > 1$. Then:*

(a) *there exist prime numbers $p_1 \leq p_2 \leq \cdots \leq p_k$ such that*

$$n = p_1 p_2 \cdots p_k.$$

(b) *if* $q_1 \leq q_2 \leq \cdots \leq q_l$ *are prime numbers such that* $n = q_1 q_2 \cdots q_l$*, then*

$$k = l \text{ and } q_i = p_i \text{ for all } i = 1, 2, \ldots, k.$$

Proof. (a) This is just Proposition 2.22.
(b) We have $p_1 \mid n$ and $n = q_1 q_2 \cdots q_l$, so $p_1 \mid q_i$ for some $i = 1, 2, \ldots, l$ by Corollary 2.20. Since q_i is prime, the only factors of q_i are 1 and q_i, and thus we must have $p_1 = q_i$. In particular, $q_1 \leq p_1$. Similarly, we can show that $q_1 = p_j$ for some $j = 1, 2, \ldots, k$, so in particular $p_1 \leq q_1$. Hence $p_1 = q_1$, and so $p_2 p_3 \cdots p_k = q_2 q_3 \cdots q_l$.

Suppose that $k \leq l$. Continuing to argue as above we can show that

$$p_1 = q_1, \quad p_2 = q_2, \quad \ldots, \quad p_k = q_k.$$

Therefore,

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_k.$$

Now suppose that $k < l$. Then we have $1 = q_{k+1} q_{k+2} \cdots q_l$, which is impossible. Therefore, we must have $k = l$. So we have proved that $l = k$ and $q_i = p_i$ for all $i = 1, 2, \ldots, k$, as required.

If $k \geq l$, then we can prove $l = k$ and $q_i = p_i$ for all $i = 1, 2, \ldots, k$ similarly. □

We note that in a prime factorization of $n \in \mathbb{N}$ given by Theorem 2.23, some of the p_is may be equal. If we collect these equal primes together we get a factorization of n of the form

$$n = p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k}$$

where $p_1 < p_2 < \cdots < p_k$ are primes and $s_1, s_2, \ldots, s_k \in \mathbb{N}$. Part (b) of the fundamental theorem of arithmetic, then tells us that if $q_1 < q_2 < \cdots < q_l$ are primes and $t_1, t_2, \ldots, t_l \in \mathbb{N}$ such that $n = q_1^{t_1} q_2^{t_2} \cdots q_l^{t_l}$, then

$$k = l \quad \text{and} \quad q_i = p_i, s_i = t_i \text{ for all } i = 1, \ldots, k.$$

Sometimes it is more convenient for us to use this formulation.

Another convention that we sometimes use is to consider the prime factorization of 1, which is the *empty product*, so $1 = p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k}$ with $k = 0$. This may sound a bit strange at first, but it is convenient to do this. It is similar to saying that $a^0 = 1$ for any $a \in \mathbb{R}$.

We end this section by remarking that the fundamental theorem of arithmetic is not an obvious statement. You may think that it is, because you have probably believed it for a long time, though never seen a proof. Let me try to convince you that it is really not obvious. It may take a bit of time for this to sink in, and you're encouraged to ask if you doesn't make sense straightaway.

I'll let you know that 1487, 1559, 1789 and 1873 are all prime numbers. Now consider the question: Is $1559 \cdot 1789 = 1487 \cdot 1873$? Suppose that you're not allowed to use the fundamental theorem of arithmetic and you have to answer this. How would you do this? I imagine that you would calculate that $1559 \cdot 1789 = 2789051 \neq 2785151 = 1487 \cdot 1873$.

Now let p_1, p_2, q_1, q_2 be primes with $p_1 \leq p_2$, $q_1 \leq q_2$ and $\{p_1, p_2\} \neq \{q_1, q_2\}$ and consider the question: Is $p_1 p_2 = q_1 q_2$? If you are given specific values, then you would check this by calculating the value of both products – but you can't do this for general p_1, p_2, q_1, q_2 and it seems that you're a bit stuck. So we need to have proved the fundamental theorem of arithmetic to know that $p_1 p_2 \neq q_1 q_2$.

2.9 Some consequences of the fundamental theorem

As mentioned at the start of Section 2.8 we are going to demonstrate the importance of Theorem 2.23 with a couple of nice consequences.

Our first consequence of Theorem 2.23 is Theorem 2.25, which says that square root of a natural number that is not a perfect square is irrational. First we prove a special case of this, which gives us an idea how to prove the general theorem. For the statement, we recall that $x \in \mathbb{R}$ is *irrational* if $x \notin \mathbb{Q}$.

Proposition 2.24. *$\sqrt{2}$ is irrational.*

Proof. Suppose for a contradiction that $\sqrt{2} \in \mathbb{Q}$. Then there exists $b, c \in \mathbb{N}$ such that

$$\sqrt{2} = \frac{b}{c}.$$

By Theorem 2.23 there are factorizations

$$b = q_1^{t_1} q_2^{t_2} \cdots q_l^{t_l}$$

and

$$c = r_1^{u_1} r_2^{u_2} \cdots r_m^{u_m}$$

where $q_1 < q_2 < \cdots < q_l, r_1 < r_2 < \cdots < r_m$ are primes and $t_1, t_2, \ldots, t_l, u_1, u_2, \ldots, u_m \in \mathbb{N}$. By cancelling common factors we assume that $q_i \neq r_j$ for any $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$. We note that $c \neq 1$ (so that $m \geq 1$), because $\sqrt{2} \notin \mathbb{N}$.

We have

$$2 = \frac{b^2}{c^2} \quad \text{so} \quad 2c^2 = b^2.$$

Therefore,

$$2r_1^{2u_1} r_2^{2u_2} \cdots r_m^{2u_m} = q_1^{2t_1} q_2^{2t_2}, \cdots q_l^{2t_l}.$$

which gives two prime factorizations of $2c^2$.

By Theorem 2.23, we have $r_1 = q_j$ for some $j = 1, 2, \ldots, m$. But we assumed that $q_i \neq r_j$ for any $i = 1, 2, \ldots, l$, $j = 1, 2, \ldots, m$. This is not possible so we have the required contradiction. □

For the statement of Theorem 2.25, we recall that $a \in \mathbb{N}$ is a *perfect square* if there exists $b \in \mathbb{N}$ such that $a = b^2$. For the proof we just need to add some extra bits to the proof of Proposition 2.24.

Theorem 2.25. *Let $a \in \mathbb{N}$. Suppose that a is not a perfect square. Then $\sqrt{a}$ is irrational.*

Proof. Suppose for a contradiction that $\sqrt{a}$ is rational. Then

$$\sqrt{a} = \frac{b}{c}$$

for some $b, c \in \mathbb{N}$.

By Theorem 2.23 there are factorizations

$$a = p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k},$$

$$b = q_1^{t_1} q_2^{t_2} \cdots q_l^{t_l}$$

and

$$c = r_1^{u_1} r_2^{u_2} \cdots r_m^{u_m},$$

where $p_1 < p_2 < \cdots < p_k, q_1 < q_2 < \cdots < q_l, r_1 < r_2 < \cdots < r_m$ are primes and $s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_l, u_1, u_2, \ldots, u_m \in \mathbb{N}$. By cancelling common factors we assume that $q_i \neq r_j$ for any $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$. We note that $c \neq 1$ (so that $m \geq 1$), because a is not a perfect square.

We have

$$a = \frac{b^2}{c^2} \quad \text{so} \quad ac^2 = b^2.$$

Therefore,

$$p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k} r_1^{2u_1} r_2^{2u_2} \cdots r_m^{2u_m} = q_1^{2t_1} q_2^{2t_2} \cdots q_l^{2t_l},$$

which gives two prime factorizations of ac^2.

By Theorem 2.23, we have $r_1 = q_j$ for some $j = 1, 2, \ldots, m$. But we know that $q_i \neq r_j$ for any $i = 1, 2, \ldots, l$, $j = 1, 2, \ldots, m$, so we have the required contradiction. □

An alternative proof of Proposition 2.24 is given below, which you may have seen before. In a sense this proof is more elementary and is shorter but it does not generalize as easily.

Alternative proof of Proposition 2.24. Suppose that $\sqrt{2} \in \mathbb{Q}$. Then there exists $a, b \in \mathbb{N}$ such that

$$\sqrt{2} = \frac{a}{b}.$$

If a and b have any common factors, then we can cancel them. So we can assume that a and b have no common factors.

Now

$$2 = \frac{a^2}{b^2} \quad \text{so} \quad 2b^2 = a^2.$$

Therefore, a^2 is even, which in turn means that a must be even. So we can write $a = 2c$ for some $c \in \mathbb{Z}$. From this we see that

$$2 = \frac{a^2}{b^2} = \frac{4c^2}{b^2} \quad \text{so} \quad 2 = \frac{b^2}{c^2}.$$

Arguing exactly as before, we see that b must be even. But this means that 2 is a factor of both a and b, and we assumed that a and b do not have any common factors, which is a contradiction. □

Now we give our second consequence of Theorem 2.23. For the statement we require the definition of a perfect nth power for $n \in \mathbb{N}$, which generalizes that of a perfect square. We say that $a \in \mathbb{N}$ is a *perfect nth power* if there exists $b \in \mathbb{N}$ such that $a = b^n$. For example, $81 = 3^4$ is a perfect 4th power and $64 = 2^6$ is a perfect 6th power.

Theorem 2.26. *Let $a, b, n \in \mathbb{N}$. Suppose that a is coprime to b and ab is a perfect nth power. Then both a and b are perfect nth powers.*

Proof. Let $c \in \mathbb{N}$ be such that $ab = c^n$. By Theorem 2.23, we have factorizations

$$c = p_1^{s_1} p_2^{s_2} \cdots p_k^{s_k},$$
$$a = q_1^{t_1} q_2^{t_2} \cdots q_l^{t_l}$$

and

$$b = r_1^{u_1} r_2^{u_2} \cdots r_m^{u_m},$$

where $p_1 < p_2 < \cdots < p_k, q_1 < q_2 < \cdots < q_l, r_1 < r_2 < \cdots < r_m$ are primes and $s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_l, u_1, u_2, \ldots, u_m \in \mathbb{N}$. Since a is coprime to b we have that $q_i \neq r_j$ for any $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$.

The equation $ab = c^n$ gives

$$q_1^{t_1} q_2^{t_2} \cdots q_l^{t_l} r_1^{u_1} r_2^{u_2} \cdots r_m^{u_m} = p_1^{ns_1} p_2^{ns_2} \cdots p_k^{ns_k}.$$

By Theorem 2.23, each q_i is equal to some p_j, and the corresponding powers t_i and ns_j must be equal. Similarly, each r_i is equal to some p_j and then u_i is equal to ns_j. We conclude that n is a factor of each of the powers t_i and u_i, say $t_i = nv_i$ and $u_i = nw_i$. Therefore,

$$a = q_1^{nv_1} q_2^{nv_2} \cdots q_l^{nv_l} = (q_1^{v_1} q_2^{v_2} \cdots q_l^{v_l})^n$$

and

$$b = r_1^{nw_1} r_2^{nw_2} \cdots r_m^{nw_m} = (r_1^{w_1} r_2^{w_2} \cdots r_m^{w_m})^n$$

are perfect nth powers. □

Theorem 2.26 may not seem that exciting, but in Example 2.28 below, we give a quite spectacular consequence. Before moving on this example, we note that for applications like this it is useful to have a more general version of Theorem 2.26 which allows a and b to be integers. We state this more general theorem below and outline how to deduce it. You are not required to know this proof fully. It will be helpful if you have an understanding of the main ideas, but no need to worry about this as long as you understand how to apply it, as the details are a bit fiddly to follow and check.

Theorem 2.27. *Let $a, b, c \in \mathbb{Z}$ and $n \in \mathbb{N}$. Suppose that a is coprime to b and $ab = c^n$. Then:*

(a) *If n is odd, there exists $r, s \in \mathbb{Z}$ such that $a = r^n$ and $b = s^n$.*

(b) *If n is even, there exists $r, s \in \mathbb{Z}$ such that $a = r^n$ and $b = s^n$, or $-a = r^n$ and $-b = s^n$.*

Outline of proof. In the case where $a, b \in \mathbb{N}$ we must have $c \in \mathbb{N}$ too and the result is exactly Theorem 2.26

If $a = 0$, then for $\mathrm{hcf}(a, b) = 1$ we must have $b = \pm 1$. So in this case we can see that the conclusions are satisfied. We can similarly deal with the case $b = 0$.

We are left to consider cases where $a, b \neq 0$ and at least one of a or b is negative. We define $a' \in \mathbb{N}$ by $a' = a$ if $a > 0$ and $a' = -a$ if $a < 0$. We similarly define $b', c' \in \mathbb{N}$. Then we observe that $a'b' = (c')^n$ and we can apply Theorem 2.26 to find $r', s' \in \mathbb{N}$ such that $a' = (r')^n$ and $b' = (s')^n$. We now define $r \in \mathbb{Z}$ by $r = r'$ if $a > 0$ and $r = -r'$ is $a < 0$, and similarly define $s \in \mathbb{Z}$. This has all been set up so that $a = r^n$ if $a > 0$ or n is odd, and so that $-a = r^n$ if $a < 0$ and n is even. Similarly, we have $b = s^n$ if $b > 0$ or n is odd, and so that $-b = s^n$ if $b < 0$ and n is even. □

Example 2.28. We are going to show that

There is no nonzero even square that is one more than a cube.

In other words we are going to show that the equation

$$4x^2 = y^3 + 1 \tag{2.8}$$

has no solutions with $x, y \in \mathbb{Z}$ and $x \neq 0$.

First we rewrite the equation as $y^3 = 4x^2 - 1$, and then factorize to get

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

Both $2x+1$ and $2x-1$ are both odd, so $\mathrm{hcf}(2x+1, 2x-1)$ is odd. Also, by Lemma 2.3(a), $\mathrm{hcf}(2x+1, 2x-1)$ is a factor of $2 = (2x+1) - (2x-1)$. It follows that $\mathrm{hcf}(2x+1, 2x-1)$ must be equal to 1, so $2x + 1$ is coprime to $2x - 1$.

Now using Theorem 2.27, we see that both $2x + 1$ and $2x - 1$ are perfect cubes. However, from the list of cubes

$$0, \pm 1, \pm 8, \pm 27, \pm 64, \pm 125, \ldots$$

we see that the only cubes that differ by 2 are 1 and -1. Therefore, we must have $x = 0$, which shows that (2.8) has no solutions with $x, y \in \mathbb{Z}$ and $x \neq 0$.

The equation (2.8) is an example of a *Diophantine equation*. In general a Diophantine equation is a polynomial equation in which the solutions are required to be integers. Solving Diophantine equations is a fascinating area of mathematics, and in general they are very difficult to solve. For example, the Diophantine equation

$$y^2 = x^3 + k$$

has not been completely solved for all values of $k \in \mathbb{N}$. These equations are called *Mordell equations* and it is known that they only have finitely many solutions.

A particularly famous example of a Diophantine equation is

$$x^n + y^n = z^n,$$

for $n \in \mathbb{N}$, which is the subject of *Fermat's last theorem*. It was proved by Andrew Wiles in 1995 that it has no nonzero integer solutions for $n \geq 3$.

2.10 Summary of Chapter 2

By the end of this chapter you should be able to:

- prove elementary lemmas and properties about factors;
- understand and apply the division theorem;
- understand and recall the definitions of common factors, highest common factors and coprime integers and prove elementary properties;
- understand Bézout's lemma and apply it to prove related statements;
- apply the Euclidean algorithm to find the highest common factor of $a, b \in \mathbb{Z}$, and the extended Euclidean algorithm to find $x, y \in \mathbb{Z}$ such that $\mathrm{hcf}(a, b) = xa + yb$;
- understand and recall the proof of the theorem about primes and products;
- understand the fundamental theorem of arithmetic and its proof, and be able to prove consequences; and
- apply the material in the chapter to solve problems and prove related statements.

Chapter 3

Modular arithmetic

In this chapter we introduce the notion of congruence of integers modulo a fixed natural number, and use this to develop the theory of modular arithmetic. This is an important area of algebra, which is a very useful method for studying the integers. Modular arithmetic is important in many areas of mathematics and computer science. Later in the chapter, in Section 3.9 we explain a particularly striking application to the theory of cryptography, which we depend on all the time when making financial transactions on the internet.

3.1 Congruence modulo n

We start by giving the main definition for this chapter.

Definition 3.1. Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. We write

$$a \equiv b \bmod n$$

and say that *a is congruent to b modulo n* if

$$n \mid a - b.$$

We write $a \not\equiv b \bmod n$ if a is not congruent to b modulo n.

Note that $a \equiv b \bmod n$ is equivalent to saying that there exists $x \in \mathbb{Z}$ such that

$$a = b + nx.$$

As usual some examples will help us to understand the definition.

Examples 3.2. (a) $43 \equiv 7 \bmod 9$, because $9 \mid 36 = 43 - 7$.

(b) $11 \equiv -28 \bmod 13$, because $13 \mid 39 = 11 - (-28)$.

(c) $31 \not\equiv 15 \bmod 7$, because $7 \nmid 31 - 15 = 16$.

(d) Let $a \in \mathbb{Z}$. Then:

- a is even if and only if $a \equiv 0 \bmod 2$; and

– a is odd if and only if $a \equiv 1 \bmod 2$.

(e) It is 4pm now, so in 269 hours it will be 9pm.
This is because $269 \equiv 5 \bmod 24$.

(f) It is Tuesday today, so in 100 days time it will be Thursday.
This is because $100 \equiv 2 \bmod 7$ and Thursday is two days after Tuesday.

The examples (d), (e) and (f) show that we are already familiar with certain cases of congruences.

We begin our study of congruences by relating them to remainders in the following lemma and corollary.

Lemma 3.3. *Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Then $a \equiv b \bmod n$ if and only if a and b leave the same remainder when divided by n.*

Before we give the proof, we explain the statement briefly by recapping the meaning of the phrase "if and only if". It means that we have to prove two things:

- if $a \equiv b \bmod n$, then a and b leave the same remainder when divided by n; and
- if a and b leave the same remainder when divided by n, then $a \equiv b \bmod n$.

Proof. Using Theorem 2.4, we can write $a = qn + r$ and $b = q'n + r'$, where $q, q', r, r' \in \mathbb{Z}$ and $0 \leq r, r' < n$.

Suppose that $a \equiv b \bmod n$. Then $n \mid a - b$, so by Lemma 2.3(a)

$$n \mid (a - b) - (q - q')n = r - r'.$$

Also $-n < r - r' < n$. This forces $r - r' = 0$, so that $r = r'$, which says that a and b leave the same remainder when divided by n.

Now suppose that a and b leave the same remainder when divided by n. Then $r = r'$, and $a - b = (q - q')n$. Thus $n \mid a - b$ and $a \equiv b \bmod n$. □

Corollary 3.4. *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then there exists unique $b \in \mathbb{Z}$ with $0 \leq b < n$ such that $a \equiv b \bmod n$.*

Proof. By Lemma 3.3, we take b to be the remainder when a is divided by n. □

Next we give some elementary properties of congruences.

Lemma 3.5. *Let $n \in \mathbb{N}$ and $a, b, c \in \mathbb{Z}$. Then:*

(a) $a \equiv a \bmod n$; *(Reflexive property)*

(b) *if $a \equiv b \bmod n$, then $b \equiv a \bmod n$; and* *(Symmetric property)*

(c) *if $a \equiv b \bmod n$ and $b \equiv c \bmod n$, then $a \equiv c \bmod n$.* *(Transitive property)*

Proof. (a) We have $n \mid 0 = a - a$, so $a \equiv a \bmod n$.
(b) Since, $a \equiv b \bmod n$ we have $n \mid a - b$. Then $n \mid b - a = -(a - b)$. Hence, $b \equiv a \bmod n$.
(c) Since, $a \equiv b \bmod n$ and $b \equiv c \bmod n$ we have $n \mid a - b$ and $n \mid b - c$. Then $n \mid a - c = (a - b) + (b - c)$. Hence, $a \equiv c \bmod n$. □

On the right in the statement of Lemma 3.5, we have given names to the properties satisfied by congruence modulo n. These are the properties required for a relation to be an *equivalence relation* – you will learn about equivalence relations in 1AC Combinatorics 1.

3.2 Arithmetic with congruences

In Lemma 3.6 below we show that congruence interacts well with arithmetic operations. The proof of (b) is an exercise.

Lemma 3.6. *Let $n, m \in \mathbb{N}$ and $a, b, a', b' \in \mathbb{Z}$. Suppose that $a \equiv b \bmod n$ and that $a' \equiv b' \bmod n$. Then:*

(a) $a + a' \equiv b + b' \bmod n$;

(b) $aa' \equiv bb' \bmod n$; *and*

(c) $a^m \equiv b^m \bmod n$.

Proof. (a) Since $a \equiv b \bmod n$ and $a' \equiv b' \bmod n$, there exist $x, x' \in \mathbb{Z}$ such that

$$a = b + nx \tag{3.1}$$

and

$$a' = b' + nx'. \tag{3.2}$$

Adding (3.1) and (3.2) gives

$$a + a' = b + b' + n(x + x').$$

We have $x + x' \in \mathbb{Z}$, because $x, x' \in \mathbb{Z}$. Therefore, $a + a' \equiv b + b' \bmod n$.

(b) is an exercise.

(c) Using (b), for the case $a' = a$ and $b' = b$ we obtain $a^2 \equiv b^2 \bmod n$.

Now we can apply Lemma 3.6(b), for the case $a' = a^2$ and $b' = b^2$, and we obtain $a^3 \equiv b^3 \bmod n$.

Continuing in this way we eventually get $a^m \equiv b^m \bmod n$. □

The properties given in Lemmas 3.5 and 3.6 allow us to manipulate expressions with congruences in a similar way to how we manipulate expressions with equals signs, as we'll see when we work with them.

We can use Lemmas 3.6 to work out remainders when we do divisions of large numbers, as demonstrated in the following examples.

Examples 3.7. (a) We are going to find the remainder when $107 \cdot 122 + 73$ is divided by 11. So by Lemma 3.3 we have to find $r \in \mathbb{Z}$ with $0 \leq r < 11$ such that $107 \cdot 122 + 73 \equiv r \bmod 11$.

First we see that $107 \equiv 8 \bmod 11$, and $122 \equiv 1 \bmod 11$. Therefore, by Lemma 3.6(b),

$$\begin{aligned} 107 \cdot 122 &\equiv 8 \cdot 1 \bmod 11 \\ &\equiv 8 \bmod 11. \end{aligned}$$

Next we see that $73 \equiv 7 \bmod 11$, so, by Lemma 3.6(a),

$$\begin{aligned} 107 \cdot 254 + 73 &\equiv 8 + 7 \bmod 11 \\ &\equiv 15 \bmod 11 \\ &\equiv 4 \bmod 11. \end{aligned}$$

Therefore, the remainder when $107 \cdot 122 + 73$ is divided by 11 is 4.

Note that we could have also worked this out by first calculating $107 \cdot 122 + 73 = 13127$ and then working out the remainder when 13127 is divided by 11, but this would have been more work. Actually with a calculator we can do this pretty quickly. However, if we want to multiply very large numbers together and work out remainders, then it is infeasible to use a calculator in this way. We will see in Section 3.9 that on occasions such calculations need to be carried out.

(b) We are going to find the remainder when 14^{24} is divided by 9. First we note that $14 \equiv 5 \bmod 9$, so $14^{24} \equiv 5^{24} \bmod 9$, by Lemma 3.6(c). Next we calculate

$$\begin{aligned} 5^2 &\equiv 25 \bmod 9 \\ &\equiv 7 \bmod 9 \end{aligned}$$

$$\begin{aligned} 5^4 &\equiv 7^2 \bmod 9 \\ &\equiv 49 \bmod 9 \\ &\equiv 4 \bmod 9 \end{aligned}$$

$$\begin{aligned} 5^8 &\equiv 4^2 \bmod 9 \\ &\equiv 16 \bmod 9 \\ &\equiv 7 \bmod 9 \end{aligned}$$

$$\begin{aligned} 5^{16} &\equiv 7^2 \bmod 9 \\ &\equiv 4 \bmod 9. \end{aligned}$$

Therefore, using Lemma 3.6, we get

$$\begin{aligned} 5^{24} &\equiv 5^{16} \cdot 5^8 \bmod 9 \\ &\equiv 7 \cdot 4 \bmod 9 \\ &\equiv 28 \bmod 9 \\ &\equiv 1 \bmod 9 \end{aligned}$$

Thus $14^{24} \equiv 1 \bmod 9$, so the remainder when 14^{24} is divided by 9 is 1.

(c) We are going to find the remainder when 27^{67} is divided by 7. We start with the congruence $27 \equiv -1 \bmod 7$. Then we can calculate

$$\begin{aligned} 27^{67} &\equiv (-1)^{67} \bmod 7 \\ &\equiv -1 \bmod 7 \\ &\equiv 6 \bmod 7. \end{aligned}$$

So the remainder is 6. The trick here is to use the negative number -1, it would have been more work if we had started by writing $27 \equiv 6 \bmod 7$.

In the examples below we give some nice applications of congruences. First we give a test to see whether an integer can be a perfect square, and then we give an easy way to determine if an integer is divisible by 3.

Examples 3.8. (a) We are going to show that 59778 is not a perfect square.

Let $a \in \mathbb{Z}$. By Corollary 3.4, there exists $b \in \{0, 1, 2, \ldots, 9\}$ such that $a \equiv b \bmod 10$. Then by Lemma 3.6(c), we have $a^2 \equiv b^2 \bmod 10$. Now we can make the following table, where the third row gives $c \in \{0, 1, 2, \ldots, 9\}$ such that $b^2 \equiv c \bmod 10$, so that we have $a^2 \equiv c \bmod 10$.

b	0	1	2	3	4	5	6	7	8	9
b^2	0	1	4	9	16	25	36	49	64	81
c	0	1	4	9	6	5	6	9	4	1

Therefore, a^2 is congruent to one of 0, 1, 4, 5, 6 or 9 modulo 10. Since $59778 \equiv 8 \bmod 10$ it cannot be a perfect square.

In fact we have shown that any integer whose last digit is 2, 3, 7 or 8 is not a perfect square.

(b) Let $a \in \mathbb{N}$ with digits $a_r a_{r-1} \ldots a_2 a_1 a_0$. So

$$a = a_0 + 10a_1 + 10^2 a_2 + \cdots + 10^{r-1} a_{r-1} + 10^r a_r.$$

We are going to show that $3 \mid a$ if and only if $3 \mid a_0 + a_1 + a_2 + \cdots + a_{r-1} + a_r$.

First we see that $10 \equiv 1 \bmod 3$, so by Lemma 3.6(c) we have $10^s \equiv 1 \bmod 3$ for all $s \in \mathbb{N}$. Therefore, using Lemma 3.6, we get

$$a \equiv a_0 + a_1 + a_2 + \cdots + a_{r-1} + a_r \bmod 3.$$

We have $3 \mid a$ if and only if $a \equiv 0 \bmod 3$. Thus $3 \mid a$ if and only if

$$a_0 + a_1 + a_2 + \cdots + a_{r-1} + a_r \equiv 0 \bmod 3.$$

if and only if

$$3 \mid a_0 + a_1 + a_2 + \cdots + a_{r-1} + a_r.$$

3.3 Linear congruence equations

A *linear congruence equation* is an equation of the form

$$ax \equiv b \bmod n$$

where $n \in \mathbb{N}$, $a, b \in \mathbb{Z}$ and we are trying to solve for x. We begin by looking at some examples.

Examples 3.9. (a) We are going to find all $x \in \mathbb{Z}$ such that

$$3x \equiv 6 \bmod 12. \tag{3.3}$$

If $x \equiv 2 \bmod 12$, then x is a solution to (3.3). We need to check whether there are any more solutions.

To do this we can we use the fact that any $x \in \mathbb{Z}$ is congruent modulo 12 to an element of the set $\{0, 1, 2, \ldots, 11\}$, so it suffices to consider only elements of this set. Then we can form the following table, where the last row gives $y \in \{0, 1, 2, \ldots, 11\}$ such that $3x \equiv y \bmod 12$.

x	0	1	2	3	4	5	6	7	8	9	10	11
$3x$	0	3	6	9	12	15	18	21	24	27	30	33
y	0	3	6	9	0	3	6	9	0	3	6	9

Thus $x = 2$, $x = 6$ and $x = 10$ are solutions to (3.3).

Hence, the solutions to (3.3) are given by

$$x \equiv 2 \bmod 12, \quad x \equiv 6 \bmod 12, \quad \text{or} \quad x \equiv 10 \bmod 12.$$

This shows that we cannot just cancel the 3 in (3.3).

We note that an alternative way to approach this example is to say that $3x \equiv 6 \bmod 12$ if and only if there exists $k \in \mathbb{Z}$ such that $3x = 6+12k$. Now we can divide by 3 to say that this occurs if and only if $x = 2+4k$. Therefore, the solutions of $3x \equiv 6 \bmod 12$ are given by $x \equiv 2 \bmod 4$, which is the same as $x \equiv 2 \bmod 12$, $x \equiv 6 \bmod 12$ or $x \equiv 10 \bmod 12$.

(b) We are going to find all $x \in \mathbb{Z}$ such that

$$2x \equiv 8 \bmod 9. \tag{3.4}$$

If $x \equiv 4 \bmod 9$, then x is a solution to (3.4). Again, we need to check whether there are any more solutions.

To do this we can we use the fact that any $x \in \mathbb{Z}$ is congruent modulo 9 to an element of the set $\{0, 1, 2, \ldots, 8\}$, so it suffices to consider only elements of this set. Then we can form the following table, where the last row gives $y \in \{0, 1, 2, \ldots, 8\}$ such that $2x \equiv y \bmod 9$.

x	0	1	2	3	4	5	6	7	8
$2x$	0	2	4	6	8	10	12	14	16
y	0	2	4	6	8	1	3	5	7

Hence, the solutions to (3.4) are given by $x \equiv 4 \bmod 9$. So in this case we can cancel the 2 in (3.4).

We'll see that the key difference in the examples is that we can make a cancellation when a is coprime to n. In Corollary 3.11, we prove that this is the case in general, and show that a linear congruence equation has a unique solution modulo n under this coprime assumption. After that we move on towards a general method for solving linear congruence equations given in Algorithm 3.12.

First we prove the following important theorem about when we can find "multiplicative inverses" for congruences.

Theorem 3.10. *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Suppose that a is coprime to n. Then there exists $z \in \mathbb{Z}$ such that*

$$az \equiv 1 \bmod n.$$

Proof. Since a is coprime to n, there exist $z, y \in \mathbb{Z}$ such that

$$1 = az + ny,$$

by Theorem 2.14. Thus $az = 1 - ny$ and hence

$$az \equiv 1 \bmod n.$$

□

The statement of the next corollary may look a little complicated to start with, but all it is saying is that the linear congruence equation (3.5) has "a unique solution modulo n" under the assumption that a is coprime to n.

Corollary 3.11. *Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Suppose that a is coprime to n. Consider the linear congruence equation*

$$ax \equiv b \bmod n. \tag{3.5}$$

There exists $s \in \mathbb{Z}$ such that the solutions of (3.5) are given by $x \equiv s \bmod n$

Proof. Since a is coprime to n, there exists $z \in \mathbb{Z}$ such that

$$az \equiv 1 \bmod n,$$

by Theorem 3.10.

Let $s \in \mathbb{Z}$ with $s \equiv zb \bmod n$, and let $x \in \mathbb{Z}$. Suppose that $x \equiv s \bmod n$. Then

$$\begin{aligned} ax &\equiv as \bmod n \\ &\equiv a(zb) \bmod n \\ &\equiv (az)b \bmod n \\ &\equiv b \bmod n. \end{aligned}$$

Now suppose, $ax \equiv b \bmod n$. Then

$$\begin{aligned} z(ax) &\equiv zb \bmod n \\ (az)x &\equiv zb \bmod n \\ x &\equiv zb \bmod n. \end{aligned}$$

Thus $x \equiv s \bmod n$.

Above we have shown that x is a solution of (3.5) if and only if $x \equiv s \bmod n$, so this gives the solutions. □

Summing up Corollary 3.11 tells us that we can solve the linear congruence equation (3.5) for a unique x modulo n when a is coprime to n. To do this we have to find z such that $az \equiv 1 \bmod n$ and then the solutions are given by $x \equiv zb \bmod n$. We can find such z using the extended Euclidean algorithm, so this gives us a method of solving (3.5). We note that when a and n are small it is often easy to find z by inspection rather than using the extended Euclidean algorithm, or to just spot a solution $s \in \{0, 1, \ldots, n-1\}$ of (3.5) and then use Corollary 3.11 to say that all solutions are given by $x \equiv s \bmod n$.

We move on to consider the case where a is not coprime to n, where we essentially give a reduction to the coprime case. We are quite brief here and the details of this reduction left as an exercise. We outline a method of how to solve general linear congruence equations. in Algorithm 3.12.

Algorithm 3.12. Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$.

To solve the linear congruence equation

$$ax \equiv b \bmod n \tag{3.6}$$

we proceed as follows.

(1) Calculate $h = \mathrm{hcf}(a, n)$.
(This can be done with the Euclidean algorithm or by inspection.)

(2) If $h \nmid b$, then there are no solutions.
Otherwise, if $h \mid b$, then consider

$$a'x \equiv b' \bmod n' \tag{3.7}$$

where $a' = \frac{a}{h}$, $b' = \frac{b}{h}$ and $n' = \frac{n}{h}$.

(3) Find $s \in \mathbb{Z}$ such that $a's \equiv b' \bmod n'$.

The solutions of (3.6) are given by $x \equiv s \bmod n'$.

To verify that the algorithm works we should justify the implicit assertions made in steps (2) and (3). Step (3) is justified by Corollary 3.11 once we know that a' is coprime to n'. The final part of the next lemma shows that a' is coprime to n', and the rest of the lemma justifies Step (2) by there are no solutions to (3.6), and if $h \mid b$, then the solutions to (3.6) are the same as the solutions to (3.7). The proof of this lemma is left as an exercise.

Lemma 3.12. *Let $a, b, x \in \mathbb{Z}$, let $n \in \mathbb{N}$ and let $h = \mathrm{hcf}(a, n)$.*

(a) *Suppose that $ax \equiv b \bmod n$. Then $h \mid b$.*

(b) *Suppose that $h \mid b$, let $a' = \frac{a}{h}$, $b' = \frac{b}{h}$ and $n' = \frac{n}{h}$.*

(i) *$ax \equiv b \bmod n$ if and only if $a'x \equiv b' \bmod n'$; and*

(ii) *a' is coprime to n'.*

We next discuss how Step (3) in Algorithm 3.12 can be achieved. One way to do this is to find $z \in \mathbb{Z}$ such that $a'z \equiv 1 \bmod n'$, and then let $s = zb$; we can find such z using the extended Euclidean algorithm, or just by inspection. For small values of a' and n', we can also just look for $s \in \{0, 1, \ldots, n' - 1\}$ such that $a's \equiv b' \bmod n'$; we note that this will take a long time for larger values of a' and n'.

At first Algorithm 3.12 may look a bit daunting, and it is easier to understand through examples. We demonstrate how to use it solve linear congruence equation in Examples 3.13 below. Note that when you are asked to solve such linear congruence equations, it is not necessary for you to refer to the steps in Algorithm 3.12 explicitly

Examples 3.13. (a) Consider the linear congruence equation

$$4x \equiv 7 \bmod 11. \tag{3.8}$$

We see that 4 is coprime to 11, so we can proceed to step (3) in Algorithm 3.12.

We look for $z \in \mathbb{Z}$ such that $4z \equiv 1 \bmod 11$. We could find this using the extended Euclidean algorithm (though we don't include the calculation here) to find that $-11 + 3 \cdot 4 = 1$, and therefore see that we can take $z = 3$. Therefore, we can multiply (3.8) by 3 to obtain

$$12x \equiv 21 \bmod 11.$$

We have $12 \equiv 1 \bmod 11$, so $12x \equiv x \bmod 11$, and also $21 \equiv 10 \bmod 11$. Therefore, we obtain

$$x \equiv 10 \bmod 11.$$

This gives the solutions to (3.8).

Given that we are working with small numbers, using the extended Euclidean algorithm to find z is not really necessary, and it would be ok to do this in another way. We could spot that $z = 3$ satisfies $3z \equiv 1 \bmod 11$ by inspection, and then proceed as above. Alternatively it would be fine to just look at each $s \in \{0, 1, 2, \ldots, 10\}$ and find that $s = 10$ satisfies $4s \equiv 7 \bmod 11$. Of course, whichever way we proceed we end up finding that the solutions are given by $x \equiv 10 \bmod 11$.

(b) Consider the linear congruence equation

$$3x \equiv 7 \bmod 12. \tag{3.9}$$

We see that $\mathrm{hcf}(3, 12) = 3$ and $3 \nmid 7$. So from step (2) in Algorithm 3.12, there are no solutions of (3.9).

(c) Consider the linear congruence equation

$$4x \equiv 6 \bmod 14. \tag{3.10}$$

We see that $\mathrm{hcf}(4, 14) = 2$ and that $2 \mid 6$. Thus as in step (2) of Algorithm 3.12 we consider instead

$$2x \equiv 3 \bmod 7. \tag{3.11}$$

In step (3) we can just look for $s \in \{0, 1, \ldots, 6\}$ such that $2s \equiv 3 \bmod 7$, and we find that $s = 5$ does the job. Therefore, we obtain

$$x \equiv 5 \bmod 7$$

gives the solutions to (3.10).

We remark that for step (3) we could have instead looked for $z \in \mathbb{Z}$ such that $2z \equiv 1 \bmod 7$. By inspection we find $z = 4$ does the job, or we could have used the extended Euclidean algorithm to calculate that $-7 + 4 \cdot 2 = 1$. Then we can multiply (3.11) by 4 to obtain

$$8x \equiv 12 \bmod 7.$$

We have $8 \equiv 1 \bmod 7$, so $8x \equiv x \bmod 7$, and also $12 \equiv 5 \bmod 7$. Therefore, we obtain

$$x \equiv 5 \bmod 7.$$

This gives the solutions to (3.10).

As we can see here this second method seems much longer in this case, so it is easier to just use the first method. However, when the numbers involved are larger this first method is not efficient, as there are lots of possibilities to check. We'll do an example in the lectures, which will demonstrate this.

We end this section with a corollary about cancelling in congruences, which will be useful later.

Corollary 3.14. *Let $n \in \mathbb{N}$ and $a, b, c \in \mathbb{Z}$. Suppose that c is coprime to n and*

$$ac \equiv bc \bmod n.$$

Then

$$a \equiv b \bmod n.$$

Proof. Since c is coprime to n, there exists $z \in \mathbb{Z}$ such that

$$cz \equiv 1 \bmod n$$

by Theorem 3.10. Then, using Lemma 3.6, we have

$$\begin{aligned} a &\equiv acz \bmod n \\ &\equiv bcz \bmod n \\ &\equiv b \bmod n. \end{aligned}$$

Therefore,

$$a \equiv b \bmod n.$$

□

3.4 Simultaneous congruences and the Chinese remainder theorem

Think of a natural number x less than 30? Work out

- the remainder a when x is divided by 2;
- the remainder b when x is divided by 3; and
- the remainder c when x is divided by 5.

It may seem surprising at first that we can determine x uniquely from a, b and c. This is sometimes called the "30 riddle", and is based on an ancient Chinese puzzle. We will see a case of this in Examples 3.18(a) below.

More generally, in this section, we look at the theory of systems of simultaneous congruences. The important result is the Chinese remainder theorem, which is Theorem 3.17. As a special case, this explains why x is uniquely determined by a, b and c, as above.

We consider a pair of simultaneous congruences

$$\begin{aligned} x &\equiv a \bmod n \\ x &\equiv b \bmod m, \end{aligned} \tag{3.12}$$

where $a, b \in \mathbb{Z}$, $n, m \in \mathbb{N}$ and we are trying to solve for x. We look at some examples.

Examples 3.15. (a) We are going to look for $x \in \mathbb{Z}$ such that

$$\begin{aligned} x &\equiv 1 \bmod 3 \\ x &\equiv 2 \bmod 5. \end{aligned} \tag{3.13}$$

Let $x \in \mathbb{Z}$ be a solution of (3.13). Then $x \equiv 1 \bmod 3$ so $x = 1 + 3y$ for some $y \in \mathbb{Z}$. Since $x \equiv 2 \bmod 5$, we obtain

$$\begin{aligned} 1 + 3y &\equiv 2 \bmod 5 \\ 3y &\equiv 1 \bmod 5. \end{aligned} \tag{3.14}$$

Now we can solve this linear congruence equation in y using the methods from the previous section. In this case it is quickest to spot that $y = 2$ is a solution. Thus as 3 is coprime to 5 the solutions of (3.14) given by

$$y \equiv 2 \bmod 5.$$

Therefore, $y = 2 + 5z$ for some $z \in \mathbb{Z}$, so

$$\begin{aligned} x &= 1 + 3(2 + 5z) \\ &= 7 + 15z. \end{aligned}$$

So $x \equiv 7 \bmod 15$.

Now let $x \in \mathbb{Z}$ with $x \equiv 7 \bmod 15$. Then reversing the arguments above shows that x is a solution of (3.13). Alternatively we can check this directly, by saying: since $3 \mid 15$, we have $x \equiv 7 \bmod 3$, so that $x \equiv 1 \bmod 3$, and similarly, we can show that $x \equiv 2 \bmod 5$.

It follows that the solutions of (3.13) are given by $x \equiv 7 \bmod 15$.

We note that an alternative method is to first list all integers x with $0 \leq x \leq 14$ and $x \equiv 1 \bmod 3$. These are:

$$1, 4, 7, 10, 13$$

Next we list x with $0 \leq x \leq 14$ and $x \equiv 2 \bmod 5$. These are:

$$2, 7, 12.$$

We observe that 7 is the only number on both lists, so $x = 7$ is a solution to the simultaneous congruences above.

Now let $y \in \mathbb{Z}$ and let z be the unique element of $\{0, 1, \ldots, 14\}$ such that $y \equiv z \bmod n$. Then we have $y \equiv z \bmod 3$ because $3 \mid 15$; and similarly $y \equiv z \bmod 5$, because $5 \mid 15$. Hence, $x = y$ is a solution of (3.13) if and only if $x = z$ is a solution of (3.13).

It follows that the solutions of (3.13) are given by $x \equiv 7 \bmod 15$.

It may seem that this latter method is easier, but we note that this does not work efficiently when the numbers involved are larger. Also if you use this method, then you should justify that you have indeed found all the solutions.

(b) We are going to find all $x \in \mathbb{Z}$ such that

$$\begin{aligned} x &\equiv 4 \bmod 9 \\ x &\equiv 7 \bmod 11. \end{aligned} \tag{3.15}$$

Let $x \in \mathbb{Z}$ be a solution of (3.15). Then $x \equiv 7 \bmod 11$ so $x = 7 + 11y$ for some $y \in \mathbb{Z}$. Since $x \equiv 4 \bmod 9$, we obtain

$$\begin{aligned} 7 + 11y &\equiv 4 \bmod 9 \\ 11y &\equiv -3 \bmod 9 \\ 2y &\equiv 6 \bmod 9. \end{aligned} \tag{3.16}$$

To obtain the last congruence above, we use that $11y \equiv 2y \bmod 9$ and $-3 \equiv 6 \bmod 9$. We solve this linear congruence equation for y using the methods in the previous section. As 2 is coprime to 9, we can just spot that $y = 3$ is a solution, so that

$$y \equiv 3 \bmod 9$$

gives all solutions of (3.16). Therefore, $y = 3 + 9z$ for some $z \in \mathbb{Z}$. Thus,

$$\begin{aligned} x &= 7 + 11(3 + 9z) \\ &= 40 + 99z. \end{aligned}$$

So $x \equiv 40 \bmod 99$.

Let $x \in \mathbb{Z}$ with $x \equiv 40 \bmod 99$. Then reversing the arguments above, we can deduce that x is a solution of (3.15).

It follows that the solutions of (3.15) are given by $x \equiv 40 \bmod 99$.

We note that in the example above we chose to let $x = 7+11y$ rather than $x = 4+9y$. It is a good exercise for you to try to do it by first letting $x = 4+9y$. You'll find that the linear congruence equation that you end up having to solve is a bit more difficult. This is why we chose to solve starting with $x = 7 + 11y$.

Now we are going to give an alternative way to solve (3.15); you can skip this at first if you prefer, as it is fine to just understand the other method. Using the extended Euclidean algorithm along with the fact that 9 is coprime to 11 we can find $k, l \in \mathbb{Z}$ such that $9k + 11l = 1$. In this case we obtain $k = 5$ and $l = -4$, and $5 \cdot 9 + (-4) \cdot 11 = 1$.

Consequently, we have

$$\begin{aligned} 45 = 5 \cdot 9 &\equiv 1 \bmod 11 \\ -44 = (-4) \cdot 11 &\equiv 1 \bmod 9 \end{aligned}$$

Also we have

$$\begin{aligned} -44 &\equiv 0 \bmod 11 \\ 45 &\equiv 0 \bmod 9. \end{aligned}$$

Now consider $x = 45 \cdot 7 + (-44) \cdot 4 = 139$. From the congruences above we obtain

$$\begin{aligned} 45 \cdot 7 + (-44) \cdot 4 &\equiv 0 \cdot 7 + 1 \cdot 4 \bmod 9 \\ &\equiv 4 \bmod 9. \end{aligned}$$

and

$$\begin{aligned} 45 \cdot 7 + (-44) \cdot 4 &\equiv 1 \cdot 7 + 0 \cdot 4 \bmod 11 \\ &\equiv 7 \bmod 11. \end{aligned}$$

Hence, $x = 139$ is a solution of (3.15).

This method of solution does not yet guarantee that all other solutions of (3.15) are given by $x \equiv 139 \bmod 99$. This does, however, follow from the Chinese remainder theorem below. Therefore, since $139 \equiv 40 \bmod 99$, the solutions of (3.15) are given by $x \equiv 40 \bmod 99$.

In both of these examples above, we have considered pairs of simultaneous congruences of the form (3.12), where n is coprime to m. Generalizing what we did in these examples above, we can obtain methods for solving such pairs of simultaneous congruences.

We move on to consider general systems of simultaneous congruences of the form (3.17) in the Chinese remainder theorem (Theorem 3.17) below. The Chinese remainder theorem tells us about solutions to systems of simultaneous congruences, under a coprimeness assumption.

We'll need the following lemma for the proof of the Chinese remainder theorem; the proof is an exercise.

Lemma 3.16. *Let $a, b, c \in \mathbb{Z}$.*

(a) *Suppose that a is coprime to b, and that $a \mid c$ and $b \mid c$. Then $ab \mid c$.*

(b) *Suppose that a is coprime to c and that b is coprime to c. Then ab is coprime to c.*

We now move on to the statement and proof of the Chinese remainder theorem. The proof is a bit more advanced than most of the proofs in the course and is a little brief in places, but is included in these printed notes for completeness. This proof is not part of the syllabus and is not examinable, so you may want to omit reading it carefully at first. The idea of the proof is to solve the congruences two at a time using the second method from (b) in Examples 3.15.

Theorem 3.17 (The Chinese remainder theorem)**.** *Let $n_1, n_2, \ldots, n_k \in \mathbb{N}$ and $a_1, a_2, \ldots, a_k \in \mathbb{Z}$. Suppose that $\mathrm{hcf}(n_i, n_j) = 1$ for $i \neq j$. Consider the system of simultaneous congruences*

$$\begin{aligned} x &\equiv a_1 \bmod n_1 \\ x &\equiv a_2 \bmod n_2 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ x &\equiv a_k \bmod n_k. \end{aligned} \tag{3.17}$$

There exists $s \in \mathbb{Z}$ such that the solutions of (3.17) are given by $x \equiv s \bmod n_1 n_2 \cdots n_k$.

Proof. We begin by considering the case $k = 2$ and we let $n_1 = n$, $n_2 = m$, $a_1 = a$ and $a_2 = b$. Since n is coprime to m there exists $k, l \in \mathbb{Z}$ such that $kn + ml = 1$ by Theorem 2.14. From this equation we obtain the congruences

$$\begin{aligned} kn &\equiv 1 \bmod m \\ lm &\equiv 1 \bmod n. \end{aligned}$$

We also clearly have the congruences

$$lm \equiv 0 \bmod m$$
$$kn \equiv 0 \bmod n.$$

Let $s = knb + lma$. We see that

$$\begin{aligned} s = knb + lma &\equiv 0b + 1a \bmod n \\ &\equiv a \bmod n, \end{aligned}$$

and

$$\begin{aligned} s = knb + lma &\equiv 1b + 0a \bmod m \\ &\equiv b \bmod m. \end{aligned}$$

Hence, $x = s$ is a solution of $x \equiv a \bmod n$ and $x \equiv b \bmod m$.

Now let $x \in \mathbb{Z}$. Suppose that $x \equiv s \bmod nm$. Then $x \equiv s \bmod n$ and $x \equiv s \bmod m$, and thus $x = r$ is also a solution of (3.17).

Now suppose that $x \in \mathbb{Z}$ is also a solution of (3.17). Then $x \equiv s \bmod n$ and $x \equiv s \bmod m$. So $n \mid x - s$ and $m \mid x - s$. Therefore, $nm \mid x - s$ by Lemma 3.16(a), and hence $x \equiv s \bmod nm$. This shows that the solutions of (3.17) in this case where $k = 2$, as given by $x \equiv s \bmod nm$.

In the case $k = 2$, we have proved that the pair of simultaneous congruences

$$x \equiv a_1 \bmod n_1$$
$$x \equiv a_2 \bmod n_2.$$

is equivalent to the single congruence

$$x \equiv c \bmod n_1 n_2,$$

where $c = kn_1a_2 + l_2a_1$. Therefore, solving (3.17) is equivalent to solving

$$\begin{aligned} x &\equiv c \bmod n_1 n_2 \\ x &\equiv a_3 \bmod n_3 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ x &\equiv a_k \bmod n_k \end{aligned}$$

so we have reduced the number of congruence equations by one. Also we have that $\mathrm{hcf}(n_1n_2, n_j) = 1$ for all $j = 3, 4, \ldots, k$ by Lemma 3.16(b).

Continuing in this way we can reduce to having a single congruence of the form

$$x \equiv s \bmod n_1 n_2 \cdots n_k,$$

for some $s \in \mathbb{Z}$. This proves the theorem. □

We note that the proof of the Chinese remainder theorem gives a method for solving a system of simultaneous congruences. This method involves repeatedly solving pairs of simultaneous congruences. You can solve these pairs of congruences in different ways as we saw in Examples 3.15.

We give a couple of examples, where we solve systems of three simultaneous congruences by considering pairs of congruences in turn. The first example is a case of the 30 riddle from the start of this section.

Examples 3.18. (a) We are going to look for $x \in \mathbb{Z}$ such that

$$\begin{aligned} x &\equiv 1 \bmod 2 \\ x &\equiv 2 \bmod 3 \\ x &\equiv 3 \bmod 5. \end{aligned} \tag{3.18}$$

It is straightforward to solve the first two congruences. Let x be a solution of the first pair. Then $x = 2 + 3y$ for some $y \in \mathbb{Z}$. We substitute this in to the first congruence to obtain

$$\begin{aligned} 2 + 3y &\equiv 1 \bmod 2 \\ 3y &\equiv -1 \bmod 2 \\ y &\equiv 1 \bmod 2 \end{aligned}$$

Thus we have $y = 1 + 2z$ for some $z \in \mathbb{Z}$, and so $x = 2 + 3(1 + 2z) = 5 + 6z$. (We note that usually at this stage we would need to solve a linear congruence equation, but we have got lucky here.)

Hence, any solution of the first pair congruences satisfies

$$x \equiv 5 \bmod 6,$$

and we can check that any such x is indeed a solution, so the first pair of congruences is equivalent to this single congruence.

Now we have to solve the pair of congruences

$$\begin{aligned} x &\equiv 5 \bmod 6 \\ x &\equiv 3 \bmod 5. \end{aligned}$$

We let x be a solution and say that $x = 5 + 6u$ for some $u \in \mathbb{Z}$, and thus

$$\begin{aligned} 5 + 6u &\equiv 3 \bmod 5 \\ 6u &\equiv -2 \bmod 5 \\ u &\equiv 3 \bmod 5. \end{aligned}$$

Thus we have $u = 3 + 5v$ for some $v \in \mathbb{Z}$, and so $x = 5 + 6(3 + 5v) = 23 + 30v$. (We note that usually at this stage we would need to solve a linear congruence equation, but we have got lucky here again.)

Therefore, any solution of (3.18) satisfies

$$x \equiv 23 \bmod 30.$$

Further, we can check that any such x is indeed a solution, so that $x \equiv 23 \bmod 30$ gives all the solutions of (3.18).

(b) We are going to solve of simultaneous congruences

$$\begin{aligned} x &\equiv 3 \bmod 5 \\ x &\equiv 4 \bmod 7 \\ x &\equiv 8 \bmod 11. \end{aligned} \tag{3.19}$$

We first solve the first pair of congruences. We let x be a solution of the first pair, and say that $x = 4 + 7y$ for some $y \in \mathbb{Z}$. Then we substitute this in to the first congruence to obtain

$$\begin{aligned} 4 + 7y &\equiv 3 \bmod 5 \\ 7y &\equiv -1 \bmod 5 \\ 2y &\equiv 4 \bmod 5 \end{aligned}$$

We can solve this linear congruence for y quickly, as 2 is coprime to 5. We can just spot that $y = 2$ is a solution by dividing by 2. Thus the solutions are given by

$$y \equiv 2 \bmod 5.$$

Thus we have $y = 2 + 5z$ for some $z \in \mathbb{Z}$, and so $x = 4 + 7(2 + 5z) = 18 + 35z$. Hence, any solution of the first pair congruences satisfies

$$x \equiv 18 \bmod 35,$$

and we can check that any such x is indeed a solution, so the first pair of congruences is equivalent to this single congruence.

Now we have to solve the pair of congruences

$$\begin{aligned} x &\equiv 18 \bmod 35 \\ x &\equiv 8 \bmod 11. \end{aligned}$$

We let x be a solution and say that $x = 18 + 35u$ for some $u \in \mathbb{Z}$, and thus

$$\begin{aligned} 18 + 35u &\equiv 8 \bmod 11 \\ 35u &\equiv -10 \bmod 11 \\ 2u &\equiv 1 \bmod 11. \end{aligned}$$

Above we have used that $35 \equiv 2 \bmod 11$, so that $35u \equiv 2u \bmod 11$, and also that $-10 \equiv 1 \bmod 11$. We observe that $6 \cdot 2 = 12 \equiv 1 \bmod 11$, so $u = 2$ is a solution to the above linear congruence equation for u. Therefore, the solutions are given by

$$u \equiv 6 \bmod 11.$$

Therefore,$u = 6 + 11v$ for some $v \in \mathbb{Z}$, and so $x = 18 + 35(6 + 11v) = 228 + 385v$. Therefore, any solution of (3.19) satisfies

$$x \equiv 228 \bmod 385.$$

Further, we can check that any such x is indeed a solution, so that $x \equiv 228 \bmod 385$ gives all the solutions of (3.19).

We give a quick comment about the method used in (b). Also we note that we chose to write $x = 4 + 7y$ rather than $x = 3 + 5y$ in solving the first pair as this makes the subsequent calculation easier; similarly we chose to write $x = 18 + 35u$ rather than $x = 8 + 11u$ later on.

Let's finish this section by looking at the "210 riddle", which is a step up from the 30 riddle that we saw at the start of the section.

Think of a natural number x less than 210? Work out

- the remainder a when x is divided by 2;
- the remainder b when x is divided by 3;
- the remainder c when x is divided by 5; and
- the remainder d when x is divided by 7.

It may seem surprising at first that we can determine x uniquely from a, b, c and d. But now we know that this is the case, thanks to the Chinese remainder theorem. In fact if we work through the proof we see that x is the natural number less than 210, which is congruent to

$$105a + 70b + 126c + 120d.$$

You should think through why this works, and then maybe you want to try it out on your friends.

You can also work out the formula that you require for the 30 riddle, and this is left as an exercise.

3.5 Congruence classes

In the next section, we're going to define the ring of integers modulo n, which is a "number system" a bit like the integers and is based on congruence modulo n. First we need to introduce congruence classes, so the next definition is key to our development of modular arithmetic.

Definition 3.19. Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. We define the *congruence class of a modulo n* to be

$$[a]_n = \{x \in \mathbb{Z} : x \equiv a \bmod n\}.$$

In words, $[a]_n$ is the set of integers that are congruent to a modulo n.

We demonstrate this definition with some examples.

Examples 3.20. (a) $[2]_6 = \{\dots, -10, -4, 2, 8, 14, \dots\}$ and
$[13]_6 = \{\dots, 1, 7, 13, 19, 25, \dots\}$.

(b) $[5]_{11} = \{\dots, -17, -6, 5, 16, 27, \dots\}$ and
$[-17]_{11} = \{\dots, -39, -28, -17, -6, 5, \dots\}$.
So $[5]_{11} = [-17]_{11}$.

(c) $[0]_2$ is the set of even integers, and
$[1]_2$ is the set of odd integers.

(d) Let $a \in \mathbb{Z}$. Then $[a]_1 = \mathbb{Z}$.

The next lemma gives some useful properties of congruences classes.

Lemma 3.21. *Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Then*

(a) $[a]_n = [b]_n$ *if and only if* $a \equiv b \bmod n$.

(b) $[a]_n = [b]_n$ *or* $[a]_n \cap [b]_n = \varnothing$.

(c) *there are exactly n congruence classes modulo n, namely*

$$[0]_n, [1]_n, [2]_n, \ldots, [n-2]_n \text{ and } [n-1]_n.$$

Proof. (a) Suppose that $[a]_n = [b]_n$. We know that $a \in [a]_n$ so we deduce that $a \in [b]_n$. Therefore, $a \equiv b \bmod n$.

Conversely suppose that $a \equiv b \bmod n$.
Let $c \in [a]_n$. Then we have $c \equiv a \bmod n$. Since $a \equiv b$, we deduce that $c \equiv b \bmod n$, so that $c \in [b]_n$. Therefore, $[a]_n \subseteq [b]_n$.
Now let $d \in [b]_n$. Then we have $d \equiv b \bmod n$. Since $a \equiv b \bmod n$, we have $b \equiv a \bmod n$ and thus also that $d \equiv a \bmod n$, so that $d \in [a]_n$. Therefore, $[b]_n \subseteq [a]_n$.
Hence, $[a]_n = [b]_n$.
(b) Suppose that $[a]_n \cap [b]_n \neq \varnothing$, and let $c \in [a]_n \cap [b]_n$. Then $c \equiv a \bmod n$ and $c \equiv b \bmod n$. Thus $a \equiv c \bmod n$ and $c \equiv b \bmod n$, so that $a \equiv b \bmod n$. Therefore, $[a]_n = [b]_n$ by (a)
(c) This follows from (a) and Corollary 3.4, because any $a \in \mathbb{Z}$ is congruent modulo n to a unique element of $\{0, 1, \ldots, n-1\}$. □

We quickly note here that Lemma 3.21 can also be deduced from the theory of equivalence relations and equivalence classes. As mentioned earlier, you'll cover this equivalence relations in 1AC Combinatorics 1.

3.6 The ring of integers modulo n

The properties of congruences that we saw earlier can be put together nicely by defining an addition and multiplication on the set of congruence classes modulo n, which we denote by $\mathbb{Z}_n$. In Definition 3.22 we define addition and multiplication on $\mathbb{Z}_n$, so it is a "number system" a bit like the integers $\mathbb{Z}$. We call $\mathbb{Z}_n$ with this addition and multiplication the ring of integers modulo n. In Section 3.7, we'll see that $\mathbb{Z}_n$ shares a lot of properties with $\mathbb{Z}$.

Lets's dive in with the definition of the ring of integers modulo n. Then we'll have some examples to help us to understand it.

Definition 3.22. Let $n \in \mathbb{N}$. We define the *set of congruence classes modulo n* to be

$$\mathbb{Z}_n = \{[a]_n : a \in \mathbb{Z}\}.$$

We define an addition $+$ and multiplication $\cdot$ on $\mathbb{Z}_n$ as follows. Let $x, y \in \mathbb{Z}_n$ and choose $x_0, y_0 \in \mathbb{Z}$ such that

$$x = [x_0]_n \quad \text{and} \quad y = [y_0]_n.$$

Define

$$x + y = [x_0 + y_0]_n$$

and

$$x \cdot y = [x_0 y_0]_n.$$

The set $\mathbb{Z}_n$ with the addition $+$ and multiplication $\cdot$ is called *the ring of integers modulo* n.

Note that $\mathbb{Z}_n$ is a set of subsets of $\mathbb{Z}$, which may seem a bit weird to get your head around at first, but once we've worked with it for a bit, it will get better. We have

$$\mathbb{Z}_n = \{[0]_n, [1]_n, [2]_n, \ldots, [n-1]_n\},$$

and, in practice, we can think of the elements $[a]_n$ of $\mathbb{Z}_n$ just to be some symbols and we have rules for adding and multiplying them. Let's look at a couple of examples to help us understand the definition of $\mathbb{Z}_n$.

Examples 3.23. (a) We consider the case $n = 2$. We have

$$\mathbb{Z}_2 = \{[0]_2, [1]_2\}.$$

So $\mathbb{Z}_2$ is the set containing the set of even numbers and the set of even numbers. For now we denote $[0]_2 = \text{even}$ and $[1]_2 = \text{odd}$, so $\mathbb{Z}_2 = \{\text{even}, \text{odd}\}$.

The addition on $\mathbb{Z}_2$ is given by the addition table below.

$+$	even	odd
even	even	odd
odd	odd	even

So one thing this table is saying is

$$\text{even} + \text{odd} = \text{odd},$$

which is just the familiar fact that if we add an even number to an odd number, then we get an odd number.

The multiplication on $\mathbb{Z}_2$ is given by the multiplication table below.

$\cdot$	even	odd
even	even	even
odd	even	odd

One thing this table is saying is that

$$\text{even} \cdot \text{odd} = \text{even},$$

which is just saying that if we multiply an even and odd number then, as we know well, we get an even number.

Now putting the addition and multiplication table in our original notation for $\mathbb{Z}_2$ we have.

$+$	$[0]_2$	$[1]_2$
$[0]_2$	$[0]_2$	$[1]_2$
$[1]_2$	$[1]_2$	$[0]_2$

$\cdot$	$[0]_2$	$[1]_2$
$[0]_2$	$[0]_2$	$[0]_2$
$[1]_2$	$[0]_2$	$[1]_2$

(b) Now we consider the case $n = 6$, which is large enough to give us a better feeling about the definition on $\mathbb{Z}_n$. We have

$$\mathbb{Z}_6 = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}.$$

We can work out the addition table below. A couple of examples of the calculations required are:

- $[4]_6 + [1]_6 = [5]_6$; and
- $[4]_6 + [5]_6 = [9]_6 = [3]_6$.

For the second sum above, we have the equality $[9]_6 = [3]_6$, because $9 \equiv 3 \bmod 6$. In general for $a, b \in \{0, 1, 2, 3, 4, 5\}$ we work out $[a]_6 + [b]_6 = [c]_6$, where $c \in \{0, 1, 2, 3, 4, 5\}$ with $c \equiv a + b \bmod 6$ to get the table below.

$+$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[0]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[1]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$	$[0]_6$
$[2]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$	$[0]_6$	$[1]_6$
$[3]_6$	$[3]_6$	$[4]_6$	$[5]_6$	$[0]_6$	$[1]_6$	$[2]_6$
$[4]_6$	$[4]_6$	$[5]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$
$[5]_6$	$[5]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$

Similarly, we can work out the multiplication table below. A couple of examples of the calculations required are:

- $[2]_6 \cdot [4]_6 = [8]_6 = [2]_6$; and
- $[5]_6 \cdot [3]_6 = [15]_6 = [3]_6$.

Above we have the equalities: $[8]_6 = [2]_6$, because $8 \equiv 2 \bmod 6$; and $[15]_6 = [3]_6$, because $15 \equiv 3 \bmod 6$. In general for $a, b \in \{0, 1, 2, 3, 4, 5\}$ we work out $[a]_6 \cdot [b]_6 = [c]_6$, where $c \in \{0, 1, 2, 3, 4, 5\}$ with $c \equiv ab \bmod 6$ to get the table below.

$\cdot$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$
$[1]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[2]_6$	$[0]_6$	$[2]_6$	$[4]_6$	$[0]_6$	$[2]_6$	$[4]_6$
$[3]_6$	$[0]_6$	$[3]_6$	$[0]_6$	$[3]_6$	$[0]_6$	$[3]_6$
$[4]_6$	$[0]_6$	$[4]_6$	$[2]_6$	$[0]_6$	$[4]_6$	$[2]_6$
$[5]_6$	$[0]_6$	$[5]_6$	$[4]_6$	$[3]_6$	$[2]_6$	$[1]_6$

We note that in the $n = 6$ example above, we worked out $[a]_6 + [b]_6 = [c]_6$, where $c \in \{0, 1, 2, 3, 4, 5\}$ with $c \equiv a + b \bmod 6$. In fact we could define addition on $\mathbb{Z}_n = \{[0]_n, [1]_n, [2]_n, \ldots, [n-1]_n\}$ by saying that for $a, b \in \{0, 1, 2, \ldots, n-1\}$, we define $[a]_n + [b]_n = [c]_n$, where $c \in \{0, 1, 2, \ldots, n-1\}$ with $c \equiv a + b \bmod n$; and we could define multiplication similarly. This may seem simpler and it is not difficult for us to show that the definition is equivalent, using Lemma 3.6. However, it turns out to be more convenient to work with Definition 3.22 as we'll see in the next section when we prove some properties of $\mathbb{Z}_n$. We are left with a potential problem to think about though, which we explain next.

Are $+$ and $\cdot$ well defined on $\mathbb{Z}_n$?

You may have noticed that there are many different ways to work out a sum or products in $\mathbb{Z}_6$ in the example above.

For instance, let $x = [3]_6$ and $y = [4]_6$. It is also possible write $x = [15]_6$ and $y = [-2]_6$. Then to work out $x + y$ we have the choice of working out either $[3]_6 + [4]_6 = [7]_6$, or working out $[15]_6 + [-2]_6 = [13]_6$. We have $[7]_6 = [1]_6 = [13]_6$, so in the end the calculation didn't depend on the choice.

Let's consider another instance, let $x = [13]_6$ and $y = [5]_6$, where it is possible to write $x = [1]_6$ and $y = [-1]_6$. Then to calculate $x \cdot y$ we can either calculate $[13]_6 \cdot [5]_6 = [65]_6$, or $[1]_6 \cdot [-1]_6 = [-1]_6$. We have $[65]_6 = [5]_6 = [-1]_6$, so in the end the calculation didn't depend on the choice.

Now let's consider this idea generally. Let $n \in \mathbb{N}$. There is a potential ambiguity in the definition of the addition on $\mathbb{Z}_n$. Let $x, y \in \mathbb{Z}_n$ and suppose that we wish to calculate $x + y$. Using the rule in Definition 3.22, we choose $x_0, y_0 \in \mathbb{Z}$ such that $x = [x_0]_n$ and $y = [y_0]_n$ and then get the answer

$$x + y = [x_0 + y_0]_n.$$

But what would happen if instead we picked different $x_0', y_0' \in \mathbb{Z}$ such that $x = [x_0']_n$ and $y = [y_0']_n$ then we would get the answer

$$x + y = [x_0' + y_0']_n.$$

Obviously, there would be a problem if

$$[x_0 + y_0]_n \neq [x_0' + y_0']_n.$$

It turns out that this cannot happen, and we explain why below.

Since, $[x_0]_n = [x_0']_n$, we have $x_0 \equiv x_0' \bmod n$, and similarly $y_0 \equiv y_0' \bmod n$. Therefore, by Lemma 3.6, we have $x_0 + y_0 \equiv x_0' + y_0' \bmod n$, so that $[x_0 + y_0]_n = [x_0' + y_0']_n$. So the two possible definitions of $x + y$ are equal. We express this by saying that $+$ is *well defined* on $\mathbb{Z}_n$.

In general if we define something, which involves some choices, then we say that it is *well defined*, if it does not depend on those choices. We can show that $\cdot$ is well defined using a similar argument to above, and doing this is left as an exercise.

3.7 Properties of $\mathbb{Z}_n$

Below we will see that addition and multiplication in $\mathbb{Z}_n$ satisfy a number of familiar properties of addition and multiplication in $\mathbb{Z}$. Before we do this, we give a list of some properties for $\mathbb{Z}$: all of these properties should be very familiar to you. On the right are the names for these properties.

(A0) For all $x, y \in \mathbb{Z}$, $x + y \in \mathbb{Z}$
(closure under addition)

(A1) For all $x, y, z \in \mathbb{Z}$, $(x + y) + z = x + (y + z)$.
(associative law of addition)

(A2) There exists $0 \in \mathbb{Z}$ such that for all $x \in \mathbb{Z}$, $x + 0 = x = 0 + x$.
(existence of additive identity)

(A3) For all $x \in \mathbb{Z}$, there exists $-x \in \mathbb{Z}$ such that $x + (-x) = 0 = (-x) + x$.
(existence of additive inverses)

(A4) For all $x, y \in \mathbb{Z}$, $x + y = y + x$.
(commutative law of addition)

(M0) For all $x, y \in \mathbb{Z}$, $x \cdot y \in \mathbb{Z}$
(closure under multiplication)

(M1) For all $x, y, z \in \mathbb{Z}$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
(associative law of multiplication)

(M2) There exists $1 \in \mathbb{Z}$ such that for all $x \in \mathbb{Z}$, $x \cdot 1 = x = 1 \cdot x$.
(existence of multiplicative identity)

(M4) For all $x, y \in \mathbb{Z}$, $x \cdot y = y \cdot x$.

(commutative law of multiplication)

(D) For all $x, y, z \in \mathbb{Z}$, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.

(distributive law)

The way that (A2) and (M2) are phrased may seem a little odd at first, they are just saying that there are special elements in $\mathbb{Z}$, which we denote by 0 and 1; and these are of course just the integers 0 and 1. Similarly, the element $-x \in \mathbb{Z}$ in (A3) is the integer that the notation suggests. Also don't worry that (M3) is missing, this is not a typo and there is a reason for this, which you'll see if you study rings in the course 2AC Algebra 2, which you can take next year.

We'll see that the addition and multiplication on $\mathbb{Z}_n$ satisfy (essentially) the same list of properties as those above for $\mathbb{Z}$. In the following lemma we show that addition on $\mathbb{Z}_n$ is associative and also that the distributive law holds in $\mathbb{Z}_n$.

Lemma 3.24. *Let $x, y, z \in \mathbb{Z}_n$. Then:*

(a) $(x + y) + z = x + (y + z)$. *In other words $+$ is associative.*

(b) $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$. *In other words $+$ is distributive over $\cdot$.*

Proof. (a) Let $x_0, y_0, z_0 \in \mathbb{Z}$ such that $x = [x_0]_n$, $y = [y_0]_n$ and $z = [z_0]_n$. By the rule for $+$, we have

$$x + y = [x_0 + y_0]_n.$$

Applying the rule again gives

$$(x + y) + z = [(x_0 + y_0) + z_0]_n. \tag{3.20}$$

Similarly, we get

$$x + (y + z) = [x_0 + (y_0 + z_0)]_n. \tag{3.21}$$

We know that addition of integers is associative, so $(x_0 + y_0) + z_0 = x_0 + (y_0 + z_0)$. Therefore, (3.20) and (3.21) give

$$(x + y) + z = x + (y + z).$$

(b) Let $x_0, y_0, z_0 \in \mathbb{Z}$ such that $x = [x_0]_n$, $y = [y_0]_n$ and $z = [z_0]_n$. By the rule for $+$, we have

$$y + z = [y_0 + z_0]_n.$$

Applying the rule for $\cdot$ gives

$$x \cdot (y + z) = [x_0(y_0 + z_0)]_n. \tag{3.22}$$

Similarly we can show that

$$(x \cdot y) + (x \cdot z) = [x_0 y_0 + x_0 z_0]_n. \tag{3.23}$$

We know that the distributive law holds for $\mathbb{Z}$, so $x_0(y_0 + z_0) = x_0 y_0 + x_0 z_0$. Therefore, (3.22) and (3.23) give

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z).$$

□

We can prove in a similar way that all of the properties in the following list are satisfied.

(A0) For all $x, y \in \mathbb{Z}_n$, $x + y \in \mathbb{Z}_n$
(closure under addition)

(A1) For all $x, y, z \in \mathbb{Z}_n$, $(x + y) + z = x + (y + z)$.
(associative law of addition)

(A2) There exists $[0]_n \in \mathbb{Z}_n$ such that for all $x \in \mathbb{Z}_n$, $x + [0]_n = x = [0]_n + x$.
(existence of additive identity)

(A3) For all $x \in \mathbb{Z}_n$, there exists $-x \in \mathbb{Z}_n$ such that $x + (-x) = [0]_n = (-x) + x$.
(existence of additive inverses)

(A4) For all $x, y \in \mathbb{Z}_n$, $x + y = y + x$.
(commutative law of addition)

(M0) For all $x, y \in \mathbb{Z}_n$, $x \cdot y \in \mathbb{Z}_n$
(closure under multiplication)

(M1) For all $x, y, z \in \mathbb{Z}_n$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
(associative law of multiplication)

(M2) There exists $[1]_n \in \mathbb{Z}_n$ such that for all $x \in \mathbb{Z}_n$, $x \cdot [1]_n = x = [1]_n \cdot x$.
(existence of multiplicative identity)

(M4) For all $x, y \in \mathbb{Z}_n$, $x \cdot y = y \cdot x$.
(commutative law of multiplication)

(D) For all $x, y, z \in \mathbb{Z}_n$, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.
(distributive law)

In mathematical language this list of properties tells us that $\mathbb{Z}_n$ (with the addition and multiplication defined in Definition 3.22) is a *commutative ring with one*. Also as these properties are satisfied by $\mathbb{Z}$, we have that $\mathbb{Z}$ is another example of a commutative ring with one.

There are many other important examples of rings in mathematics, and you'll be able to learn more about rings in 2AC Algebra 2. In that course we'll see how many familiar properties of the integers hold more generally for other rings. The theory of rings is an important area of mathematics, with motivation and applications throughout mathematics and the physical sciences. The use of rings in number theory and algebraic geometry led to a major development of their theory throughout the 20th century, and remain amongst the most important areas of mathematics research today. Indeed much of my own research regards the structure and representation theory of certain rings.

Don't worry if this last section seems a bit abstract at the moment. For now you should just get an idea of what terms like "associative", "commutative", "additive inverse" and "multiplicative identity" mean.

3.8 Fermat's little theorem

To end this chapter we build on the material we've developed and cover a couple of very nice applications of the theory, namely Fermat's little theorem (in this section) and RSA cryptography (in the next section). We will use the notation given in the following definition.

Definition 3.25. Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. We write $a \pmod n$ to denote the unique element r of $\{0, 1, 2, \ldots, n-1\}$ such that $a \equiv r \bmod n$.

For example, $19 \pmod 4 = 3$, $23 \pmod 6 = 5$, $-7 \pmod 9 = 2$ and $-15 \pmod 1 1 = 7$. It is important that you understand the difference between $a \pmod n = r$ and $a \equiv b \bmod n$, so you should think this through here: we have that $a \equiv b \bmod n$ means $n \mid a-b$; whereas $a \pmod n = r$ means that r is the unique element of $\{0, 1, 2, \ldots, n-1\}$ such that $a \equiv r \bmod n$.

We going to prove a cool theorem called Fermat's little theorem. Before, we prove the theorem we demonstrate it with an example.

Example 3.26. Let $p = 7$ and let $a = 3$. In the table below we look at the values of $3b \pmod 7$ for all $b = 1, 2, \ldots, 6$.

b	1	2	3	4	5	6
$3b$	3	6	9	12	15	18
$3b \pmod 7$	3	6	2	5	1	4

We can see that the bottom row gives a rearrangement of $1, 2, 3, 4, 5, 6$. Therefore, we see that

$$(1 \cdot 3)(2 \cdot 3)(3 \cdot 3)(4 \cdot 3)(5 \cdot 3)(6 \cdot 3) \equiv 6! \bmod 7,$$

by Lemma 3.6. Therefore, we have

$$6! \cdot 3^6 \equiv 6! \bmod 7.$$

Now we see that $7 \nmid 6! = 720$. Therefore, as 7 is prime, it is coprime to 6!. Thus by Corollary 3.14, we obtain

$$3^6 \equiv 1 \bmod 7.$$

We now use the idea in the example above to prove Fermat's little theorem.

Theorem 3.27. *Let $p \in \mathbb{N}$ be prime and let $a \in \mathbb{Z}$. Suppose that $p \nmid a$. Then*

$$a^{p-1} \equiv 1 \bmod p.$$

Proof. Since p is prime and $p \nmid a$, we have that a is coprime to p.

For $b = 1, 2, \ldots, p-1$, let $d_b = ab \pmod p$.
Suppose that $d_b = d_c$ for $b, c \in \{0, 1, 2, \ldots, p-1\}$. Then we have we have $ab \equiv ac \bmod p$ and thus $b \equiv c \bmod p$ by Corollary 3.14. Therefore, $b = c$, because $b, c \in \{0, 1, \ldots, p-1\}$.

It follows that $d_1, d_2, \ldots, d_{p-1}$ is a rearrangement of $1, 2, \ldots, p-1$. Thus $d_1 d_2 \ldots d_{p-1} = (p-1)!$. Also using Lemma 3.6, we have

$$\begin{aligned}(p-1)!\, a^{p-1} &= (1a)(2a)\ldots((p-1)a) \\ &\equiv d_1 d_2 \ldots d_{p-1} \bmod p\end{aligned}$$

Hence,

$$(p-1)!\, a^{p-1} \equiv (p-1)! \bmod p.$$

Now $p \nmid (p-1)!$, by Corollary 2.20, and thus p is coprime to $(p-1)!$. Therefore, by Corollary 3.14, we obtain

$$a^{p-1} \equiv 1 \bmod p.$$

□

We now give the following corollary of Theorem 3.27.

Corollary 3.28. *Let $p \in \mathbb{N}$ be prime and let $a \in \mathbb{Z}$. Then*

$$a^p \equiv a \bmod p.$$

Proof. We consider two cases.
Case 1: $a \equiv 0 \bmod p$. Then $a^p \equiv 0 \bmod p$, so $a^p \equiv a \bmod p$.
Case 2: $a \not\equiv 0 \bmod p$. Then $a^{p-1} \equiv 1 \bmod p$ by Theorem 3.27. So $a^p \equiv a \bmod p$. □

Another way of stating the corollary above is to say that for any integer a and a prime p, we have that

$$p \mid a^p - a.$$

This is a really striking statement!

Next we prove a theorem that is similar to Fermat's last theorem. We'll need this in the next section when we look at RSA public key cryptography.

Theorem 3.29. *Let $p, q \in \mathbb{N}$ be distinct primes, $k \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then*

$$a^{k(p-1)(q-1)+1} \equiv a \bmod pq.$$

Proof. First we show that $a^{k(p-1)(q-1)+1} \equiv a \bmod p$.
If $a \equiv 0 \bmod p$, then this is clear.
If $a \not\equiv 0 \bmod p$, then $p \nmid a$, so $a^{p-1} \equiv 1 \bmod p$, by Theorem 3.27 and, therefore, $a^{k(p-1)(q-1)} \equiv 1 \bmod p$. Hence,

$$a^{k(p-1)(q-1)+1} \equiv a \bmod p.$$

Similarly, we can show that

$$a^{k(p-1)(q-1)+1} \equiv a \bmod q.$$

Therefore,

$$p \mid a^{k(p-1)(q-1)+1} - a \quad \text{and} \quad q \mid a^{k(p-1)(q-1)+1} - a.$$

Since, $p \neq q$, we have that p is coprime to q. Therefore,

$$pq \mid a^{k(p-1)(q-1)+1} - a,$$

by Lemma 3.16(a). Hence,

$$a^{k(p-1)(q-1)+1} \equiv a \bmod pq.$$

□

3.9 The RSA cryptosystem

The material in the section is very interesting, and there is an exercise on this on the fifth problem sheet. However, this is not a part of the syllabus that could be assessed on the exam.

People have always had the need to communicate in a secret way, so that their enemies are not able to understand what they are saying. Consequently many *cryptosystems* for encoding communications have been designed. Typically a cryptosystem works as explained below.

Alice wants to send a secret message to Bob. They proceed as follows.

- Alice converts the message into a sequence of numbers $\mathbf{m} = (m_1, m_2, \ldots, m_r)$ called the *plaintext*.
- Alice enciphers the plaintext by performing some operation on the string of numbers to obtain a different sequence of natural numbers $\mathbf{c} = (c_1, c_2, \ldots, c_r)$ called the *ciphertext* and sends it to Bob.
- Bob knows how to invert the operation that Alice performed, so he is able decipher the ciphertext to calculate $\mathbf{m}$ from $\mathbf{c}$.

Often the process of enciphering and deciphering involves knowledge of a *key*.

The general description of a cryptosystem given above is unlikely to make that much sense, until we have seen an example.

A symmetric key cryptosystem

Alice wants to send Bob a message using a simple cryptosystem known as a *Caesar shift*. In advance they have agreed on a key, which is used to encrypt and decrypt the message. The key k is an integer between 0 and 25; in this example we take $k = 18$. Alice wants to send the message

EAT MY SHOES

She first converts each letter in the alphabet to a natural number between 0 and 25, where $A \mapsto 1, B \mapsto 2, \ldots Y \mapsto 25, Z \mapsto 0$ to obtain the plaintext

$$\mathbf{m} = (5, 1, 20, 13, 25, 19, 8, 15, 5, 19).$$

Next for each of the entries m_i in $\mathbf{m}$ she calculates

$$c_i = m_i + 18 \pmod{26}$$

to obtain the ciphertext

$$\mathbf{c} = (23, 19, 12, 5, 17, 11, 0, 7, 23, 11).$$

Then Alice sends $\mathbf{c}$ to Bob. Since Bob knows the key is 18, he is able to calculate

$$m_i = c_i - 18 \pmod{26}$$

and recover the plaintext $\mathbf{m}$.

This Caesar shift is not very secure because if someone is able to guess what the key is, then they can break the code. In fact it would be very easy to guess the key using some "frequency analysis" if the message was longer. Also an enemy would only need to try 26 possible keys before managing to decipher the message.

The Caesar shift is an example of a *symmetric key cryptosystem*. This is one in which two parties agree on a secret key in advance of communication. There are symmetric key cryptosystems that are secure if Alice and Bob are able to secretly communicate the key between themselves. However, this is likely to be problem, because they don't yet have a way to communicate securely. This difficulty makes symmetric key cryptosystems impractical for the amount of information that needs to be encoded nowadays for secure internet transactions.

Below we describe the RSA cryptosystem, which is a *public key cryptosystem*. Public key cryptosystems involve a *public key* used to encode, and a *private key* used to decode. Therefore, they avoid the problem of having to communicate the key used for encryption and decryption.

The RSA cryptosystem

The RSA public key cryptosystem is used for many of the secure transactions that we make on the internet, so we are utterly dependent on it. The security is based on the belief that it is very difficult to factorize large numbers into a product of primes, which we discuss before moving on to the RSA cryptosystem.

Suppose you wanted to factorize 6557, then you could get a calculator out, and you would work out quite quickly that $6557 = 79 \cdot 83$. However, if you wanted to factorize 9,088,109 then it would take you quite a long time to work out that $9,088,109 = 2969 \cdot 3061$. As we see below the security of the RSA cryptosystem depends on factorizing a number with about 400 digits into the product of two primes. It is estimated that this would take thousands of years using the most powerful computers. So for practical purposes it is completely infeasible.

Before explaining the RSA cryptosystem, we give a little history. The idea of an asymmetric cryptosystem is attributed to Diffie and Hellman, who published the idea in 1976. Subsequently, the RSA cryptosystem was devised by Rivest, Shamir and Adleman, and published in 1977. More recently in 1997 it was revealed that Clifford Cocks had also devised the system in 1973 whilst working at GCHQ, but this work remained classified for 24 years.

The RSA cryptosystem works as follow, when Alice wants to send a message to Bob.

Encryption

First Bob needs a *public key*. To get a public key Bob finds two large prime numbers p and q with $p \neq q$ and sets $N = pq$, and he also chooses $e \in \mathbb{N}$ such that $0 < e < (p-1)(q-1)$ and e is coprime to $(p-1)(q-1)$. Bob's public key is the pair (N, e). He makes his public key available to everyone.

Alice wants to send a message to Bob. First Alice converts her message so that the

plaintext is a sequence of natural numbers

$$\mathbf{m} = (m_1, m_2, \ldots, m_r),$$

where $0 \leq m_i < N$ for $i = 1, 2, \ldots, r$. We don't go into details, but this can be done in a similar way to the assignment $A \mapsto 1, B \mapsto 2, \ldots Y \mapsto 25, Z \mapsto 0$, except that each m_i corresponds to a sequence of letters. To encode this Alice calculates

$$c_i = m_i^e \pmod{N}$$

for $i = 1, 2, \ldots, r$. Then the ciphertext is

$$\mathbf{c} = (c_1, c_2, \ldots, c_r).$$

Decryption

Bob needs his *private key* to decrypt the ciphertext. To calculate the private key, he uses the extended Euclidean algorithm to find $x, y \in \mathbb{Z}$ such that

$$x(p-1)(q-1) + ye = 1.$$

Then the private key is $d = y \pmod{(p-1)(q-1)}$. Thus $d \in \mathbb{N}$ is the unique natural number that satisfies $0 < d < (p-1)(q-1)$ and $ed \equiv 1 \bmod (p-1)(q-1)$.

When Bob receives the ciphertext he calculates $c_i^d \pmod{N}$ for $i = 1, 2, \ldots, r$. Now $ed = k(p-1)(q-1)+1$ for some $k \in \mathbb{N}$ and $m_i^{k(p-1)(q-1)+1} \equiv m_i \bmod N$, by Theorem 3.29. Therefore,

$$\begin{aligned} c_i^d \pmod{N} &= m_i^{ed} \pmod{N} \\ &= m_i^{k(p-1)(q-1)+1} \pmod{N} \\ &= m_i. \end{aligned}$$

So Bob has recovered the plaintext.

Security

The security of the communication using the RSA cryptosystem depends on the fact that an enemy who intercepts the message is not able to decode it. Suppose an enemy, called Eve, wants to intercept and decode the message. Eve knows what $N = pq$ and e are, but wants to know what d is. At present the only known way of finding d is to find $(p-1)(q-1)$ first and then to calculate d as explained above. Now

$$(p-1)(q-1) = pq - p - q + 1$$

so if Eve knows $(p-1)(q-1)$, then she can work out what $p+q$ is. Then from knowing $p+q$ and pq she can work out what p and q are.

It follows that to find the secret key d for the RSA cryptosystem with public key (N, e), Eve needs to be able to find the prime numbers p and q such that $N = pq$. The only known way to decode messages is to find d, so to decrypt messages Eve needs to be able to factorize a large number into a product of primes. At present the prime numbers p and q used for an RSA public key typically have about 200 digits. As discussed above it is infeasible to factorize a number with 400 digits into a product of primes, so it is infeasible to break the RSA cryptosystem. Thus it is effectively impossible for Eve to decode Alice's message to Bob.

Summary

We summarize the protocol for a secret message to be sent by Alice to Bob using the RSA cryptosystem.

- Bob creates a public key (N, e), where $N = pq$ is the product of primes p and q and $e \in \mathbb{N}$ such that $0 < e < (p-1)(q-1)$ and e is coprime to $(p-1)(q-1)$.
- Alice encodes the plaintext $\mathbf{m} = (m_1, m_2, \ldots, m_r)$ by setting $c_i = m_i^e \pmod N$ to obtain the ciphertext $\mathbf{c} = (c_1, c_2, \ldots, c_r)$.
- Bob calculates the private key d, which is the unique natural number such that $0 < d < (p-1)(q-1)$ and $ed \equiv 1 \bmod (p-1)(q-1)$ using the extended Euclidean algorithm.
- Bob calculates $c_i^d \pmod N = m_i$, to decrypt the ciphertext and recover the plaintext.

To end this section we give an example of using the RSA cryptosystem. We use much smaller primes than those used in practice.

Example 3.30. Let $p = 29$ and $q = 37$, so we have $N = 1073$, and we let $e = 11$. So the public key is $(1073, 11)$.

Next we find the private key. First we calculate $(p-1)(q-1) = 1008$. Then we use the extended Euclidean algorithm to find $x, y \in \mathbb{Z}$ such that $1008x + 11y = 1$. First we write

$$1008 = 91 \cdot 11 + 7$$

Second we write

$$11 = 7 + 4.$$

Third we write

$$7 = 4 + 3.$$

Fourth we write

$$4 = 3 + 1.$$

Then we reverse these steps to get

$$\begin{aligned} 1 &= 4 - 3 \\ &= 4 - (7 - 4) \\ &= -7 + 2 \cdot 4 \\ &= -7 + 2 \cdot (11 - 7) \\ &= 2 \cdot 11 - 3 \cdot 7 \\ &= 2 \cdot 11 - 3 \cdot (1008 - 91 \cdot 11) \\ &= -3 \cdot 1008 + 275 \cdot 11. \end{aligned}$$

Therefore, we have $275 \cdot 11 \equiv 1 \bmod 1008$. Thus the private key is $d = 275$.

Now suppose we want to encode the plaintext

$$\mathbf{m} = (134, 529, 406).$$

We calculate

$$\begin{aligned} 134^{11} \pmod{1073} &= 251, \\ 529^{11} \pmod{1073} &= 545, \\ 406^{11} \pmod{1073} &= 406. \end{aligned}$$

So we obtain the ciphertext

$$\mathbf{c} = (251, 545, 406).$$

Decoding involves the calculations

$$\begin{aligned} 251^{275} \pmod{1073} &= 134, \\ 545^{275} \pmod{1073} &= 529, \\ 406^{275} \pmod{1073} &= 406. \end{aligned}$$

Note that I did these calculations using a modular arithmetic calculator like the one you can find on http://users.wpi.edu/~martin/mod.html.

The theory of cryptography is a really interesting branch of pure mathematics. An excellent book that you can read to find out more is:

- S. Singh, *The Code Book: The Secret History of Codes and Code-breaking*, Fourth Estate Ltd., 2002.

Cryptography is also discussed in Chapter 15 of the recommended textbook for this course by M. W. Liebeck. There is also a lot of information on wikipedia and there are many other references.

Another interesting problem related to the RSA cryptosystem, is the need to find very large primes. There is some really nice mathematics behind this, and you can read more about it in Chapter 14 of Liebeck's book.

3.10 Summary of Chapter 3

By the end of this chapter you should be able to:

- understand and recall the definition of congruence modulo n;
- prove elementary lemmas and properties about congruences and arithmetic of congruences;
- perform calculations with congruences;
- understand Theorem 3.10 and its proof, and apply it to prove related statements;
- solve linear congruence equations;
- solve systems of simultaneous congruences;
- explain the construction of $\mathbb{Z}_n$ and make calculations in $\mathbb{Z}_n$;
- prove properties of $\mathbb{Z}_n$;
- understand and recall the definition of $a \pmod{n}$;
- understand and apply Fermat's little theorem; and
- apply the material in the chapter to solve problems and prove related statements.

Chapter 4

Permutations

In this chapter we are going to take a change in direction and study permutations, which are bijective functions of sets. These are particularly nice and useful functions in mathematics. They're really important in group theory, as we'll see in the next chapter.

4.1 Functions

We will use some results about functions that you should have covered in 1RA. There is a brief recap on functions, which includes everything that we'll need in Appendix A. Before we go on we recall here some of the key things about functions that we'll want. We start off with the definitions of identity functions and inverse functions.

Definition 4.1. Let A and B be sets.

(a) The *identity function on* A is the function $\mathrm{id}_A : A \to A$ defined by $\mathrm{id}_A(x) = x$.

(b) Let $f : A \to B$ be a bijection. The *inverse of* f is the function $f^{-1} : B \to A$ defined by

$$f^{-1}(x) \text{ is the unique element } y \in A \text{ such that } f(y) = x.$$

Next we state a proposition about functions that collects a few useful properties of functions. All parts of the proposition are covered within Lemmas A.9, A.6, A.11 and A.14.

Proposition 4.2. *Let A, B, C and D be sets and let $f : A \to B$, $g : B \to C$ and $h : C \to D$ be functions.*

(a) *Suppose that f and g are bijections. Then $g \circ f : A \to C$ is a bijection.*

(b) $(h \circ g) \circ f = h \circ (g \circ f)$.

(c) $f \circ \mathrm{id}_A = f$ *and* $\mathrm{id}_B \circ f = f$.

(d) *Suppose that f is a bijection then $f^{-1} : B \to A$ is a bijection, and $f^{-1} \circ f = \mathrm{id}_A$ and $f \circ f^{-1} = \mathrm{id}_B$.*

4.2 Permutations

We now introduce permutations in the main definition of this chapter. It is conventional to use the Greek letter Ω (pronounced omega) for a set when working with permutations.

Definition 4.3. Let Ω be a set. A *permutation* of Ω is a bijection $\Omega \to \Omega$.
We define

$$\mathrm{Sym}(\Omega) = \{f : f \text{ is a permutation of } \Omega\}.$$

So $\mathrm{Sym}(\Omega)$ is the set of all permutations of Ω.

We give some examples.

Examples 4.4. Let $\Omega = \{1, 2, 3, 4\}$.

(a) Define $f : \Omega \to \Omega$ by

$$f(x) = \begin{cases} x+1 & \text{if } x \neq 4 \\ 1 & \text{if } x = 4. \end{cases}$$

Then f is a permutation of Ω.

(b) Define $g : \Omega \to \Omega$ by

$$g(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ x+1 & \text{if } x \text{ is odd.} \end{cases}$$

Then g is *not* a permutation of Ω, because g is not injective (or surjective).

(c) Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$. Then f is a permutation of $\mathbb{R}$.

4.3 Two-row notation

For the rest of this chapter we are only interested in permutations of finite sets. In fact we mainly just consider sets of the form $\Omega = \{1, 2, \ldots, n\}$, where $n \in \mathbb{N}$. In this case we just write S_n for $\mathrm{Sym}(\Omega)$, and we just write id rather than id_Ω.

Below we give a convenient notation for representing permutations.

Definition 4.5. Let $n \in \mathbb{N}$ and $f \in S_n = \mathrm{Sym}(\{1, 2, \ldots, n\})$.
The *two-row notation* for f is the symbol

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ f(1) & f(2) & \ldots & f(n) \end{pmatrix}.$$

To help us understand this we give some examples.

Examples 4.6. (a) Let $f \in S_4$ be as in Examples 4.4(a). Then the two-row notation for f is

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}.$$

(b) Define $g \in S_5$ by

$$g(1) = 3,\ g(2) = 2,\ g(3) = 5,\ g(4) = 4,\ g(5) = 1.$$

Then the two-row notation for g is

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 4 & 1 \end{pmatrix}.$$

(b) We can list all 6 elements of S_3:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

As we can see in the examples the second row in the two-row notation for $f \in S_n$ is a rearrangement of $1, 2, \ldots, n$. From this we can work out that the number of permutations of $\{1, 2, \ldots, n\}$ is $n!$. Therefore, we have $|S_n| = n!$.

4.4 Composition

By Proposition 4.2(a) we know that the composition of two permutations is a permutation. Therefore, if $f, g \in \mathrm{Sym}(\Omega)$, then $g \circ f \in \mathrm{Sym}(\Omega)$, where Ω is a set. An alternative way of saying this is to say that $\mathrm{Sym}(\Omega)$ is *closed under composition*.

In the example below we show how to work out the composition of two permutations using two-row notation.

Example 4.7. Let $f, g \in S_5$ with two-row notation

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \end{pmatrix}.$$

and

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 4 & 3 & 2 \end{pmatrix}.$$

Then we have

$$g \circ f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix}$$

To work this out we can write

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ f(1) & f(2) & f(3) & f(4) & f(5) \\ g(f(1)) & g(f(2)) & g(f(3)) & g(f(4)) & g(f(5)) \end{pmatrix}.$$

We obtain the bottom two rows by rearranging the columns of g so that the top row of g is the same as the bottom row of f. Then we remove the middle row. It's not necessary for you to write the middle row, if you can do the calculation without it.

Also we can calculate

$$f \circ g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix}.$$

We observe that $g \circ f \neq f \circ g$, so composition of permutations is not commutative.

4.5 Inversion

Let f be a permutation of a set Ω. Then f is a bijection of Ω, so f has an inverse f^{-1} and f^{-1} is a bijection, by Proposition 4.2(d). Therefore, $f^{-1} \in \mathrm{Sym}(\Omega)$. We demonstrate how to work out the inverse of a permutation in two-row notation in the example below.

Example 4.8. Let $f \in S_4$ with two-row notation

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}.$$

Then we have

$$f^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

To work this out we can swap the rows of f and write

$$\begin{pmatrix} 2 & 4 & 3 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} = \begin{pmatrix} f(1) & f(2) & f(3) & f(4) \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

then rearrange the columns so that the top row is 1234; though you can, and may prefer, to do it directly.

4.6 Powers of a permutation

We now define powers of permutations in a very similar way to how we define powers of real numbers, just that we are using composition rather than multiplication.

Definition 4.9. Let Ω be a set, $f \in \mathrm{Sym}(\Omega)$, and $r \in \mathbb{Z}$.
We define f^r as follows.

- For $r = 0$, we set $f^0 = \mathrm{id}_\Omega$.
- For $r > 0$, we set $f^r = f \circ f \circ \cdots \circ f$, where there are r factors all equal to f.
- For $r < 0$, we let $s = -r$, so $s > 0$ and then set $f^r = (f^{-1})^s$.

So we have

$$f^1 = f,\ f^2 = f \circ f,\ f^3 = f \circ f \circ f,\ f^4 = f \circ f \circ f \circ f, \ldots,$$

and

$$f^{-2} = f^{-1} \circ f^{-1},\ f^{-3} = f^{-1} \circ f^{-1} \circ f^{-1}, \ldots.$$

We give an example of taking powers.

Example 4.10. Let $f \in S_4$ with two-row notation

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}.$$

Then we have

$$f^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix},$$

$$f^3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix},$$

and

$$f^4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

So $f^4 = \mathrm{id}$.

Also we can work out that $f^{-1} = f^3$. Can you explain why?

The next lemma says that powers of permutation have similar properties to powers of real numbers. We omit the proof, as this can be done similarly to how it would be proved for powers of numbers.

Lemma 4.11. *Let Ω be a set, $f \in \mathrm{Sym}(\Omega)$, and $r, s \in \mathbb{Z}$. Then*

(a) $f^{r+s} = f^r \circ f^s$; *and*

(b) $f^{rs} = (f^r)^s$.

4.7 Cycles

In this section we define cycles. This leads to an alternative convenient way of thinking about permutations that we develop in the next section.

Definition 4.12. Let Ω be a set, and $f \in \mathrm{Sym}(\Omega)$.
We say that f is a *cycle* of length k if there exist distinct elements $a_1, a_2, \ldots, a_k \in \Omega$ such that

$$f(a_1) = a_2,\ f(a_2) = a_3,\ \ldots,\ f(a_{k-1}) = a_k,\ f(a_k) = a_1,$$

and $f(a) = a$ for all $a \in \Omega \setminus \{a_1, a_2, \ldots, a_k\}$.

We use the notation

$$f = (a_1\ a_2\ \ldots\ a_k)$$

Often we say *k-cycle* instead of cycle of length k.

It is best to understand this definition through some examples.

Example 4.13. .

(a) Let $f \in S_5$ with two-row notation

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 3 & 1 & 4 \end{pmatrix}.$$

Then f is a 4-cycle, because

$$f(1) = 2,\ f(2) = 5,\ f(5) = 4,\ f(4) = 1,$$

and $f(3) = 3$. So

$$f = (1\,2\,5\,4).$$

(b) Let $g \in S_5$ with two-row notation

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 2 & 4 & 3 \end{pmatrix}.$$

Then g is a 3-cycle and

$$g = (2\,5\,3).$$

4.8 Cycle decomposition and cycle notation

In Theorem 4.16 below we state and then sketch a proof that any permutation can be written as a product of disjoint cycles. Before this we demonstrate it with an example.

Example 4.14. We are going to express

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 3 & 2 & 5 & 1 & 6 & 9 & 4 & 7 \end{pmatrix} \in S_9$$

as a product of disjoint cycles.

First we look at the sequence

$$1,\ f(1),\ f^2(1), \cdots = 1, 8, 4, 5, 1, 8, \ldots.$$

This gives us our first cycle

$$(1\,8\,4\,5).$$

Next we look at

$$2,\ f(2),\ f^2(2), \cdots = 2, 3, 2, 3, 2, 3, \ldots.$$

This gives our second cycle

$$(2\,3).$$

Next we look at

$$6,\ f(6),\ f^2(6), \cdots = 6, 6, 6, \ldots.$$

So 6 is a fixed point of 6, and we view it as a cycle of length 1. So we have our third cycle

$$(6).$$

Looking at

$$7,\ f(7),\ f^2(7), \cdots = 7, 9, 7, 9, 7, 9 \ldots.$$

This gives our last cycle

$$(7\,9).$$

Thus we have decomposed f as a product of cycles:

$$f = (1\,8\,4\,5) \circ (2\,3) \circ (6) \circ (7\,9).$$

The cycle shape of f is the symbol $(4, 2, 2, 1)$, which tells us the length of the cycles in this decomposition of f.

Before stating Theorem 4.16 we need to say what a "product of disjoint cycles" means.

Definition 4.15. Let Ω be a set, and let $c_1, c_2, \ldots, c_m \in \mathrm{Sym}(\Omega)$ be cycles:

$$c_i = (a_{i,1}\, a_{i,2}\, \ldots\, a_{i,k_i}).$$

(a) The *product of the cycles* $c_1, c_2, \ldots, c_m$ is just their composition

$$c_1 \circ c_2 \circ \cdots \circ c_m.$$

(b) We say that the cycles $c_1, c_2, \ldots, c_m$ are *disjoint* if

$$a_{i,j} \neq a_{k,l}$$

whenever $i \neq k$. So this means that no two cycles contain an entry in common.

We do not include all of the details of the proof of Theorem 4.16 below, so we only call it a sketch proof; in particular, we don't explain why the cycles are disjoint in the sketch proof. It will be helpful for your understanding to fill in the gaps.

Theorem 4.16. *Let Ω be a finite set and $f \in \mathrm{Sym}(\Omega)$. Then f can be written a product of disjoint cycles.*

Sketch proof. Since Ω is finite, we assume for this proof that $\Omega = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$.
We construct the cycle decomposition as follows.
First we form the sequence

$$a_1 = 1,\ a_2 = f(a_1)\ a_3 = f(a_2),\ \ldots.$$

Since Ω is finite, we can show that $a_{k+1} = a_1$ for some $k \in \mathbb{N}$, and we can choose k be to be minimal. Then we let c_1 be the cycle of length k:

$$c_1 = (a_1\, a_2\, \ldots\, a_k).$$

If $k = n$, then we see that $f = c_1$. So we have written f as a product of disjoint cycles.
So suppose $k \neq n$, then we can pick $i \in \{1, 2, \ldots, n\} \setminus \{a_1, a_2, \ldots, a_k\}$ to be minimal. We define another cycle

$$c_2 = (a'_1\, a'_2\, \ldots\, a'_{k'}),$$

where

$$a'_1 = i,\ a'_2 = f(a'_1)\ a'_3 = f(a'_2),\ \ldots.$$

We can check that c_1 and c_2 are disjoint.
If $k + k' = n$, then we see that $f = c_1 \circ c_2$. So we have written f as a product of disjoint cycles.
Continuing in this way, we will eventually have written f as a product of disjoint cycles. □

Armed with Theorem 4.16, we can now define the cycle notation and cycle shape of a permutation.

Definition 4.17. Let Ω be a finite set and $f \in \mathrm{Sym}(\Omega)$.

(a) The *cycle notation* of f is the decomposition of f as a product of disjoint cycles:

$$f = c_1 \circ c_2 \circ \cdots \circ c_m.$$

(b) The *cycle shape* of f is the sequence $(r_1, r_2, \ldots, r_m)$ giving the lengths of the cycles in the cycle notation of f ordered so that $r_1 \geq r_2 \geq \cdots \geq r_m$.

We give some more examples of cycle decompositions.

Examples 4.18. (a) Let

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 3 & 6 & 1 & 4 \end{pmatrix}.$$

Then the cycle notation of f is

$$f = (1\,2\,5) \circ (3) \circ (4\,6).$$

So the cycle shape of f is $(3, 2, 1)$.

(b) Let

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 2 & 1 & 5 & 3 & 6 \end{pmatrix}.$$

Then the cycle notation of g is

$$g = (1\,4\,5\,3) \circ (2) \circ (6).$$

So the cycle shape of g is $(4, 1, 1)$.

The cycle notation for a permutation is not unique. First any of the cycles can begin with any element in it and secondly the disjoint cycles can be rearranged. For instance, with f as in Example 4.18, we could write the cycle notation as

$$f = (3) \circ (6\,4) \circ (5\,1\,2).$$

The order of the disjoint cycles can be changed because *disjoint cycles commute*; but beware that in general cycles do not commute.

Before we move on, we explain some notational conventions that we use when writing permutations in cycle notation.

- When we write permutations in cycle notation, we omit the symbol $\circ$ for composition. This is really just to be lazy and save ourselves space and time when writing them out.
- We omit 1-cycles when writing out permutations. This is justified as a 1-cycle is just the identity permutation – you should think about this and make sure that you understand why.

So for example f and g as in Example 4.18 would be written as

$$f = (125)(46), \quad \text{and} \quad g = (1453).$$

We've made the notation more compact by removing some space too.

In this shorter notation, we can write down all $6 = 3!$ the elements of S_3 in cycle notation. They are id, (12), (13), (23), (123) and (132).

4.9 Calculating in cycle notation

Here we look at calculating compositions and inverses of permutations in cycle notation. We just do this by looking at one extended example. The main idea of how to do this is by talking to yourself, as we'll see.

Example 4.19. Let

$$f = (1\,2\,4) \circ (3\,6\,5), \quad g = (1\,3\,2\,6) \circ (4\,5), \quad h = (1\,5\,3) \circ (2) \circ (4\,6) \quad \in S_6.$$

As explained at the end of the previous section, when we write permutations in cycle notation, we omit the composition symbols and 1-cycles. With this notation, we get

$$f = (124)(365), \quad g = (1326)(45), \quad h = (153)(46) \quad \in S_6.$$

We're going to work out $g \circ f$. To do this we first write out the cycle notation of h and g next to each other to denote their composition.

$$g \circ f = (1326)(45)(124)(365)$$

Then going along the cycles **from right to left** we can determine $g \circ f$ by saying.
Well f sends 1 to 2 and g sends 2 to 6, so $g \circ f$ sends 1 to 6.
Next we consider 6, and say that f sends 6 to 5 and g sends 5 to 4, so $g \circ f$ sends 6 to 4.
Next we consider 4, and say that f sends 4 to 1 and g sends 1 to 3, so $g \circ f$ sends 4 to 3.
Next we consider 3, and say that f sends 3 to 6 and g sends 6 to 1, so $g \circ f$ sends 3 to 1.
Thus we get that (1643) is a cycle in $g \circ f$.
Now we consider 2, and say that f sends 2 to 4 and g sends 4 to 5, so $g \circ f$ sends 2 to 5.
Next we consider 5, and say that f sends 5 to 3 and g sends 3 to 2, so $g \circ f$ sends 5 to 2.
Thus we get that (25) is a cycle in $g \circ f$.
Hence,

$$g \circ f = (1643)(25).$$

Let's do $h \circ g$ too, to help us to get used to this. First we write out the cycle notation of h and g next to each other to denote their composition.

$$h \circ g = (153)(46)(1326)(45)$$

Then going along the cycles **from right to left** we say:
1 goes to 3 goes to 1.
So (1) is a cycle in $h \circ g$.
2 goes to 6 goes to 4.
4 goes to 5 goes to 3.
3 goes to 2.
So (243) is a cycle in $h \circ g$.
5 goes to 4 goes to 6.
6 goes to 1 goes to 5.
So (56) is a cycle in $h \circ g$.
Hence,

$$h \circ g = (243)(56).$$

As a last example on composing in cycle notation we'll do $g \circ h$. First write them next to each other to denote their composition.

$$g \circ h = (1326)(45)(153)(46).$$

Then going along the cycles **from right to left** we say:
1 goes to 5 goes to 4.
4 goes to 6 goes to 1.
So (14) is a cycle in $g \circ h$.
2 goes to 6.
6 goes to 4 goes to 5.
5 goes to 3 goes to 2.
So (265) is a cycle in $g \circ h$.
3 goes to 1 goes to 3.
So (3) is a cycle in $g \circ h$.
Hence,

$$g \circ h = (14)(265).$$

In these calculations you may find it a bit unnatural that we have to read the cycles from right to left – it was written in bold to make sure you noticed. This is because when we write a composition like $f \circ g$ it means do g and then f, so we are going from right to left. Sometimes functions are "written on the right" to make this more natural, but we choose not to do that here, though you'll possibly see this in some books and in future courses.

The last thing we'll do in this example is to work out f^{-1}, g^{-1} and h^{-1}.
To work out f^{-1}. We say:
Well 1 is the image of 4 under f, so f^{-1} sends 1 to 4.
Next we say that 4 is the image of 2 under f, so f^{-1} sends 4 to 2.
Next we say that 2 is the image of 1 under f, so f^{-1} sends 2 to 1.
Thus (142) is a cycle in f^{-1}. Similarly we obtain that (356) is a cycle in f^{-1}. Hence,

$$f^{-1} = (142)(356).$$

Note that $(142) = (421)$, because we can change which element we write first in the cycle, and similarly $(356) = (563)$. Therefore, $f^{-1} = (421)(563)$. So that we obtain f^{-1} by reversing the order of the elements in the cycles.
In fact this method of reversing the order of the elements in the cycles work for finding the inverse of any permutation, you should convince yourself of this. In particular, we obtain

$$g^{-1} = (6231)(54) = (1623)(45) \quad \text{and} \quad h^{-1} = (351)(64) = (135)(46).$$

4.10 The order of a permutation

In this section we define the order of a permutation, and then move on to eventually give a formula to calculate the order of a permutation from its cycle shape.

In the definition below we use powers of permutations as defined in Definition 4.9.

Definition 4.20. Let Ω be a finite set and $g \in \mathrm{Sym}(\Omega)$. The *order of* g is the smallest $s \in \mathbb{N}$ such that $g^s = \mathrm{id}$. We write $o(g)$ for the order of g.

So the order of g is the number of times that you have to repeat g before every element of $\{1, 2, \ldots, n\}$ gets sent back to itself. We note that the definition above assumes that there does exist $r \in \mathbb{N}$ such that $g^r = \mathrm{id}$, and this may not be immediately obvious. However, it follows from Lemma 4.22 later that this is indeed the case.

Let's get a better understanding of the order of permutations in some examples.

Examples 4.21. (a) Let $g = (123) \in S_3$. We can calculate that $g^2 = (132)$ and $g^3 = \mathrm{id}$. Thus $o(g) = 3$.

More generally let $r, n \in \mathbb{N}$ with $r \leq n$ and let $g \in S_n$ be an r-cycle. If g is repeated r times, then each element of the cycle gets sent all the way round and back to each self. Therefore, we have $o(g) = r$.

(b) Let $g = (12)(345)$. We can calculate that

$$g^2 = (354);$$

$$g^3 = (12);$$

$$g^4 = (345);$$

$$g^5 = (12); \text{ and}$$

$$g^6 = \mathrm{id}.$$

Therefore, $o(g) = 6$. We note that $o(g) = 6 = \mathrm{lcm}(2, 3)$, so that the order of g is the least common multiple of the parts of the cycle shape of g. To understand this, we observe from the calculations above that we g has to be applied a multiple of 2 times to "kill off" (12) and also a multiple of 3 times to "kill off" (345).

More generally, let $n, r_1, r_2 \in \mathbb{N}$ with $r_1 + r_2 \leq n$, and let $g = c_1 \circ c_2 \in S_n$, where c_1 and c_2 are disjoint cycles of length r_1 and r_2 respectively. Then for $r \in \mathbb{N}$, we see that $g^r = c_1^r \circ c_2^r$, and that this is the identity if and only if both $c_1^r = \mathrm{id}$ and $c_2^r = \mathrm{id}$. From (a) and Lemma 4.24, we see that $c_1^r = \mathrm{id}$ if and only if $r_1 \mid r$, and that $c_2^r = \mathrm{id}$ if and only if $r_2 \mid r$. Hence, we deduce that the smallest $r \in \mathbb{N}$ such that $g^r = \mathrm{id}$ is the least common multiple of r_1 and r_2.

In (b) of the examples we saw that the order of a permutation that is the product of two disjoint cycles is the least common multiple of the parts of its cycle shape; we recall that least common multiple of two natural numbers is defined in Definition 2.17 and the cycle shape is defined in Definition 4.17. In general we can give a similar expression for the order of any permutation and we do so in Lemma 4.22.

For the statement of this lemma we need to say what we mean by the least common multiple of a list of integers, which is done by generalizing Definition 2.17 as follows. For $a_1, a_2, \ldots, a_m \in \mathbb{N}$, the *least common multiple of* $a_1, a_2, \ldots, a_m$ is the smallest $l \in \mathbb{N}$ such that $a_i \mid l$ for all $i = 1, 2, \ldots, m$; and we denote it by $l = \mathrm{lcm}(a_1, a_2, \ldots, a_m)$.

Now we can state and prove the lemma giving the formula for the order of a permutation in terms of its cycle shape. The idea of the proof is to generalize what we said for the product of two disjoint cycles in Examples 4.21(b).

Lemma 4.22. *Let $n \in \mathbb{N}$ and let $g \in S_n$ with cycle shape $(r_1, r_2, \ldots, r_m)$. Then $o(g) = \operatorname{lcm}(r_1, r_2, \ldots, r_m)$.*

Proof. Let $g = c_1 \circ c_2 \circ \cdots \circ c_m$ be the cycle notation of g, where c_i is a cycle of length r_i. Then for $r \in \mathbb{N}$ we have $g^r = c_1^r \circ c_2^r \circ \cdots \circ c_m^r$, and we have $g^r = \mathrm{id}$ if and only if $c_i^r = \mathrm{id}$ for each i. Moreover, we have $c_i^r = \mathrm{id}$ if and only if $r_i \mid r$, by Examples 4.21(a). Hence, we see that the smallest $r \in \mathbb{N}$ such that $g^r = \mathrm{id}$ is the least common multiple of $r_1, r_2, \ldots, r_m$. □

We next give a quick example, where we use the formula in Lemma 4.22 to determine the order of some permutations.

Example 4.23. Let

$$f = (1234)(567)(89), \quad g = (12345)(67)(89), \quad h = (12345)(789) \quad \in S_9$$

Then we have

$$o(f) = \operatorname{lcm}(4, 3, 2) = 12, \quad o(g) = \operatorname{lcm}(5, 2, 2) = 10, \quad o(h) = \operatorname{lcm}(5, 3) = 15.$$

We end this section by stating a useful lemma about the order of a permutation..

Lemma 4.24. *Let Ω be a finite set, $s \in \mathbb{Z}$ and $g \in \mathrm{Sym}(\Omega)$. Then $g^s = \mathrm{id}_\Omega$ if and only if $o(g) \mid s$.*

Proof. Let $m = o(g)$. Using the division theorem (Theorem 2.4), we can write $s = qm + r$, where $q, r \in \mathbb{Z}$ and $0 \leq r < m$. Then we have

$$\begin{aligned} g^s &= g^{qm+r} \\ &= (g^m)^q g^r \\ &= g^r. \end{aligned}$$

Above we used Lemma 4.11. Also we used that $g^m = \mathrm{id}$, because $m = o(g)$. Since $0 \leq r < m$, we have that $g^r = \mathrm{id}_\Omega$ if and only if $r = 0$. Thus $g^s = \mathrm{id}_\Omega$ if and only if $m \mid s$. □

4.11 The sign of a permutation

We will not cover the material in this section in the lecture videos, and it is not part of the syllabus so is not examinable.

We're going to cover a more subtle aspect of theory of permutations, which may take a bit more time to grasp. This is the sign of a permutation, which is defined in Definition 4.25 below. Before we get on to the definition it will help to demonstrate the idea by first considering permutations in S_3.

Let x_1, x_2 and x_3 be three variables (by this we just mean they are symbols that we can write polynomials in). We let permutations in S_3 act on these variables in the same way that they act on the numbers 1, 2 and 3. By this we mean for $f \in S_3$ and $i \in \{1, 2, 3\}$, we say that f sends x_i to $x_{f(i)}$. This can be extended to polynomials in x_1, x_2 and x_3 and given a polynomial $M = m(x_1, x_2, x_3)$ we define $f(M) = m(x_{f(1)}, x_{f(2)}, x_{f(3)})$.

There's quite a lot being defined here, so lets see a couple of examples: for $f = (123)$ and $M = x_1x_2 - x_2x_3^2$, we have $f(M) = x_2x_3 - x_3x_1^2 = x_2x_3 + x_1^2x_3$; and for $g = (23)$ and $N = x_1^2x_3 + x_1x_2x_3$, we have $g(N) = x_1^2x_2 + x_1x_3x_2 = x_1^2x_2 + x_1x_2x_3$.

We are particularly interested in the polynomial

$$\Delta_3 = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).$$

Let's see what happens when we apply each of the elements of S_3 to Δ_3. Clearly we have

$$\mathrm{id}(\Delta_3) = \Delta_3.$$

Next we consider $f = (12)$, and we calculate

$$\begin{aligned} f(\Delta_3) &= (x_2 - x_1)(x_2 - x_3)(x_1 - x_3) \\ &= (x_2 - x_1)(x_1 - x_3)(x_2 - x_3) \\ &= -(x_1 - x_2)(x_2 - x_3)(x_1 - x_3) \\ &= -\Delta_3. \end{aligned}$$

We got from the first line to the second line, by rearranging the factors and from the second line to the third by using $x_2 - x_1 = -(x_1 - x_2)$.
Let's also consider $f = (123)$, and we calculate

$$\begin{aligned} f(\Delta_3) &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) \\ &= (x_2 - x_1)(x_3 - x_1)(x_2 - x_3) \\ &= (x_1 - x_2)(x_2 - x_3)(x_1 - x_3) \\ &= \Delta_3. \end{aligned}$$

We can do all of the permutations in a similar way and we can summarize what we find in the table below.

f	id	(12)	(13)	(23)	(123)	(132)
$f(\Delta_3)$	Δ_3	$-\Delta_3$	$-\Delta_3$	$-\Delta_3$	Δ_3	Δ_3.

So we see that $f(\Delta_3)$ is always equal to either Δ_3 or $-\Delta_3$. To see why this occurs note that when we apply $f \in S_3$ to Δ_3, we obtain

$$f(\Delta_3) = (x_{f(1)} - x_{f(2)})(x_{f(1)} - x_{f(3)})(x_{f(2)} - x_{f(3)}).$$

Then we observe that the factors $x_i - x_j$, for $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$ have been permuted, but some of the factors have been reversed from $x_i - x_j$ to $x_j - x_i = -(x_i - x_j)$.

Now let's consider general $n \in \mathbb{N}$. We can consider the polynomial

$$\Delta_n = \prod_{1 \le i < j \le n} (x_i - x_j) \tag{4.1}$$

in the variables $x_1, x_2, \ldots, x_n$. The symbol $\prod$ here means the product of all the terms; similarly to how we use the symbol $\sum$ to denote a sum. Then for $f \in S_n$ we can define

$$f(\Delta_n) = \prod_{1 \le i < j \le n} (x_{f(i)} - x_{f(j)}) \tag{4.2}$$

For similar reasons to those given above for the case $n = 3$, we always have $f(\Delta_n) = \pm\Delta_n$. This allows us define the sign and parity of an element of S_n.

Definition 4.25. Let $n \in \mathbb{N}$ and let $f \in S_n$.

We define the *sign* of f to be the number $\operatorname{sgn}(f) \in \{1, -1\}$ such that $f(\Delta_n) = \operatorname{sgn}(f)\Delta_n$, where Δ_n is defined in (4.1) and $f(\Delta_n)$ is defined in (4.2).

We define the *parity* of f by saying that f is *even* if $\operatorname{sgn}(f) = 1$ and f is *odd* if $\operatorname{sgn}(f) = -1$.

Now that we have the definition of the sign of a permutation, we move on to consider how to calculate it examples. The first step is the next lemma about the sign of the composition of permutations.

Lemma 4.26. *Let $n \in \mathbb{N}$ and let $f, g \in S_n$. Then $\operatorname{sgn}(f \circ g) = \operatorname{sgn}(f)\operatorname{sgn}(g)$.*

Proof. We calculate $(f \circ g)(\Delta_n)$. On the one hand we get

$$(f \circ g)(\Delta_n) = \operatorname{sgn}(f \circ g)\Delta_n,$$

and other the other hand we get

$$\begin{aligned}(f \circ g)(\Delta_n) &= f(g(\Delta_n)) \\ &= f(\operatorname{sgn}(g)(\Delta_n)) \\ &= \operatorname{sgn}(g)f(\Delta_n)) \\ &= \operatorname{sgn}(g)\operatorname{sgn}(f)\Delta(n).\end{aligned}$$

Hence, $\operatorname{sgn}(f \circ g) = \operatorname{sgn}(f)\operatorname{sgn}(g)$. □

We move on to determine the sign of a 2-cycle; often we refer to a 2-cycle as a *transposition*.

Lemma 4.27. *Let $n \in \mathbb{N}$ and let $f = (kl) \in S_n$ be a transposition, where $k, l \in \{1, 2, \ldots, n\}$. Then $\operatorname{sgn}(f) = -1$.*

Proof. We may assume that $k < l$. We consider $f(\Delta_n) = \prod_{1 \le i < j \le n}(x_{f(i)} - x_{f(j)})$. The factors $x_i - x_j$ for $i < j$ all occur in $f(\Delta_n)$ with some of them reversed. The factors that are reversed are

$$x_k - x_{k+1}, x_k - x_{k+2}, \ldots, x_k - x_l,$$

$$x_{k+1} - x_l, x_{k+1} - x_l, x_{l-1} - x_l.$$

So there are $(l - k) + (l - k) - 1 = 2(l - k) - 1$ such factors, which is an odd number. Hence, we obtain that $f(\Delta_n) = -\Delta_n$, so that $\operatorname{sgn}(f) = -1$. □

Using the previous two lemmas, we are now in a position to determine the sign of any cycle.

Lemma 4.28. *Let $n, k \in \mathbb{N}$ with $k \le n$, and let $f \in S_n$ be a k-cycle. Then $\operatorname{sgn}(f) = (-1)^{k-1}$.*

Proof. We have $f = (a_1 a_2 \ldots a_k)$ for some a_i, and we observe that we can write f as a product of $k - 1$ transpositions:

$$f = (a_1 a_2) \circ (a_2 a_3) \circ \cdots \circ (a_{k-2} a_{k-1}) \circ (a_{k-1} a_k).$$

Now the lemma follows from Lemmas 4.26 and 4.27. □

We are now in a position to assemble the ingredients from the previous three lemmas to obtain a general formula for the sign of a permutation in terms of its cycle shape; we recall that the cycle shape of a permutation is the sequence of numbers giving the lengths of the cycles in its cycle notation, as defined in Definition 4.17.

Proposition 4.29. *Let $n \in \mathbb{N}$ and let $f \in S_n$ with cycle shape $(r_1, r_2, \ldots, r_m)$. Then* $\operatorname{sgn}(f) = (-1)^{r_1-1}(-1)^{r_2-1}\ldots(-1)^{r_m-1}$.

Proof. This follows immediately from Lemma 4.26 and 4.28. □

We give some examples where we work out the signs of some permutations using Proposition 4.29.

Examples 4.30. (a) Let $f = (16)(279)(3845) \in S_9$.
Then $\operatorname{sgn}(f) = (-1)^1(-1)^2(-1)^3 = 1$, so f is even.

(b) Let $g = (13)(27)(4865) \in S_8$.
Then $\operatorname{sgn}(g) = (-1)^1(-1)^1(-1)^3 = -1$, so g is odd.

We now state and prove a corollary, which gives an alternative interpretation of the parity of a permutation.

Corollary 4.31. *Let $f \in S_n$. Then f can be written as a product of transpositions. Moreover*

- *the parity of f is even if there are an even number of transpositions in this product; and*
- *the parity of f is odd if there are an odd number of transpositions in this product.*

Proof. The fact that f can be written as a product of transpositions follows from Theorem 4.16 and the proof of Lemma 4.28.
The statement about the parity then follows from Lemmas 4.26 and 4.27. □

We note that for $f \in S_n$, there may be many different ways to write f as a product of transpositions. However, it is implicit in the statement of Corollary 4.31 that the parity of the number of transpositions does not depend on how we write f as a product of transpositions.

You may wonder why we have made quite a lot of fuss about the sign of a permutation, as it may not seem that useful straightaway. However, you should rest assured that this is something important that you're likely to encounter in your further studies. For instance it is needed to work with determinants. Also it is important in group theory, which we'll see in the next chapter.

Finally, we note that although we have worked throughout this section just with permutations in S_n, we can also define the parity of a permutation of a finite set Ω. Possibly the easiest way to do this is to fix a bijection $h : \Omega \to \{1, 2, \ldots, n\}$, where n is the number of elements Ω. Then we define the parity of $g \in \operatorname{Sym}(\Omega)$ to be the same as the parity of $h \circ g \circ h^{-1} \in S_n$.

4.12 Summary of Chapter 4

By the end of this chapter you should be able to:

- calculate the two-row notation of a permutation;
- calculate compositions, inverses and powers of permutations in two-row notation;
- calculate the cycle notation and cycle shape of a permutation;
- calculate compositions, inverses and powers of permutations in cycle notation;
- apply the formula for the order of permutations in cycle notation to determine the order of permutations; and
- apply the material in the chapter to solve problems and prove related statements.

Chapter 5

Groups

The last chapter of these notes gives an introduction to group theory. This is a really interesting area of mathematics, which gives us a language to study symmetry. Group theory crops up all over mathematics and also has applications in physics and chemistry. A lot of important research on group theory has be done at the University of Birmingham, and breakthroughs in this area are being made by current members of staff.

We'll see that group theory brings together a few of the topics that you have learned in your degree so far. You'll be able to learn more about group theory in courses later in your degree.

5.1 Permutation groups and symmetry groups

We start off by studying groups of permutations, this will help us to see how group theory gives a means to study symmetry. Later we'll define groups abstractly, and see that permutation groups are indeed groups. We start by saying what we mean by a permutation group.

Let Ω be a set. A subset G of $\mathrm{Sym}(\Omega)$ is called a *permutation group* on Ω if it satisfies the following conditions.

(PG0) For all $g, h \in G$, we have $g \circ h \in G$.

(PG2) $\mathrm{id}_\Omega \in G$.

(PG3) For all $g \in G$, we have $g^{-1} \in G$.

The labels of the conditions on the left are used, so that we can refer to them more easily. There isn't a typo and (PG1) is supposed to be missing – we'll see why this is later. The conditions can be expressed in words as follows.
(PG0) says that G is closed under composition.
(PG2) says that G contains the identity.
(PG3) says that G is closed under taking inverses.

Given a permutation group G that has a finite number of elements, the *order of* G is defined to be the number of elements of G and is denoted by $|G|$.

We note that $G = \mathrm{Sym}(\Omega)$ is a permutation group. The conditions (PG0), (PG2) and (PG3) are given by parts (a), (c) and (d) Proposition 4.2. We refer to $\mathrm{Sym}(\Omega)$ as

the *symmetric group on* Ω. In particular, for the case $\Omega = \{1, 2, \ldots, n\}$, we see that S_n is a permutation group. We refer to S_n as the *symmetric group of degree* n. The order of S_n is $|S_n| = n!$.

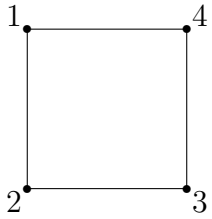
We next explain how the sign of a permutation gives rise to permutation groups called an alternating groups; as the sign of a permutation is not part of the examinable syllabus neither is this definition, but it is of interest, so we have included it here. In Definition 4.25, we saw how to define the sign $\mathrm{sgn}(f) \in \{1, -1\}$ of a permutation $f \in S_n$. We recall from Corollary 4.31 that any $f \in S_n$ can be written as a product of transpositions and that

$$\mathrm{sgn}(f) = \begin{cases} 1 & \text{if there are an even number of transpositions in this product;} \\ -1 & \text{if there are an odd number of transpositions in this product.} \end{cases}$$

We define the *alternating group of degree* n to be $A_n = \{f \in S_n : \mathrm{sgn}(f) = 1\}$. We leave it as an exercise to prove that A_n is a permutation group; the key ingredient that you require is Lemma 4.26. We note that for $n \geq 2$. The function $f \mapsto f \circ (1\ 2)$ gives a bijection from A_n to $\{f \in S_n : \mathrm{sgn}(f) = -1\}$. From this we deduce that $|A_n| = \frac{n!}{2}$.

Next we give some examples, where we see how permutation groups show up naturally when considering symmetry in geometry; also we see that permutation groups show up in puzzles and in chemistry.

Examples 5.1. (a) Consider a square in the plane with vertices labelled anticlockwise by $\Omega = \{1, 2, 3, 4\}$.



We define an *isometry* of the square to be a bijection of the square to itself that preserves distances. We can make our life easier by saying that an isometry is given by a permutation of the vertices of the square that preserves distances; so we can view isometries of the square as elements of S_4. So the symmetry group of the square is the set of $g \in S_4$ such g gives an isometry of the square. The group is denoted by D_8 and referred to as the *dihedral group of order 8*. The elements of D_8 are the symmetries of the square in the sense that you are familiar with.
There are 8 elements of D_8: the identity, three rotations and four reflections (usually we view the identity as a rotation, so say that there are four rotations and four reflections). These are given by the following elements of S_4.

- id = do nothing.
- $\rho_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = (1\,2\,3\,4)$
  $=$ a rotation through $\frac{\pi}{2}$ radians
- $\rho_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (1\,3)(2\,4)$
  $=$ a rotation through π radians

- $\rho_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} = (1\,4\,3\,2)$
  $=$ a rotation through $\frac{3\pi}{2}$ radians
- $\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = (1\,4)(2\,3)$
  $=$ a reflection in the vertical axis
- $\sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} = (2\,4)$
  $=$ a reflection in the 1–3 diagonal
- $\sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (1\,2)(3\,4)$
  $=$ a reflection in the horizontal axis
- $\sigma_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = (1\,3)$
  $=$ a reflection in the 2–4 diagonal

We can check that D_8 satisfies (PG0), (PG2) and (PG3), so that D_8 is a permutation group.

The set of the rotational symmetries of the square is $\{\mathrm{id}, \rho_1, \rho_2, \rho_3\}$. We let $\rho = \rho_1$ and then we can calculate that $\rho^2 = \rho_2$, $\rho^3 = \rho_3$ and $\rho^4 = \mathrm{id}$. Then we can check that $\{\mathrm{id}, \rho, \rho^2, \rho^3\}$ is a permutation group. We refer to this group as the *cyclic group of order* 4 and denote it by C_4.

We can calculate the multiplication table of G, where the entry in the row labelled α and column labelled β is $\alpha \circ \beta$.

	id	ρ_1	ρ_2	ρ_3	σ_1	σ_2	σ_3	σ_4
id	id	ρ_1	ρ_2	ρ_3	σ_1	σ_2	σ_3	σ_4
ρ_1	ρ_1	ρ_2	ρ_3	id	σ_2	σ_3	σ_4	σ_1
ρ_2	ρ_2	ρ_3	id	ρ_1	σ_3	σ_4	σ_1	σ_2
ρ_3	ρ_3	id	ρ_1	ρ_2	σ_4	σ_1	σ_2	σ_3
σ_1	σ_1	σ_4	σ_3	σ_2	id	ρ_3	ρ_2	ρ_1
σ_2	σ_2	σ_1	σ_4	σ_3	ρ_1	id	ρ_3	ρ_2
σ_3	σ_3	σ_2	σ_1	σ_4	ρ_2	ρ_1	id	ρ_3
σ_4	σ_4	σ_3	σ_2	σ_1	ρ_3	ρ_2	ρ_1	id

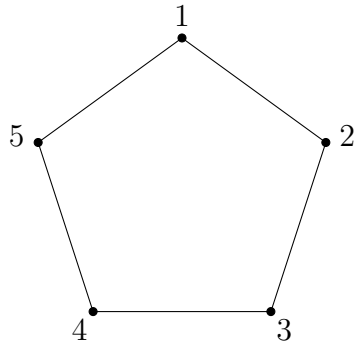
We go on to explain more about D_8, which is helpful if we want to work with D_8. We begin by modifying our notation and let $\rho = \rho_1$ as above and also let $\sigma = \sigma_1$. Then using the multiplication table above we can calculate that

- $\mathrm{id} = \rho^0$
- $\rho_1 = \rho$

- $\rho_2 = \rho^2$
- $\rho_3 = \rho^3$
- $\sigma_1 = \sigma$
- $\sigma_2 = \sigma \circ \rho^3$
- $\sigma_3 = \sigma \circ \rho^2$
- $\sigma_4 = \sigma \circ \rho$

Therefore, the elements of D_8 can be written uniquely in the form $\sigma^j \circ \rho^i$, where $i \in \{0, 1, 2, 3\}$ and $j \in \{0, 1\}$. Also we can check that $\rho \circ \sigma = \sigma \circ \rho^{-1} = \sigma \circ \rho^3$, by calculating $(1234)(14)(23) = (24) = (14)(23)(1432)$. Further, we can calculate that $\rho^4 = \mathrm{id}$ and $\sigma^2 = \mathrm{id}$. From this we can deduce that $\sigma \circ \rho^i = \rho^{-i} \circ \sigma = \rho^{4-i} \circ \sigma$ for $i = 1, 2, 3$.
You may want to think about what all this means geometrically, and this may help you to understand what is going on.

(b) For $n \in \mathbb{N}$, we can consider the isometries of a regular n-gon. For example for $n = 5$ we have a pentagon, and we can label the vertices by $\{1, 2, 3, 4, 5\}$ anticlockwise.



and write out all the isometries of the pentagon as permutations in S_5. This is left as an exercise, but we note that there are 10 of them and the description is to similar to what we had for the square above: there are five rotations (including the identity) and five reflections. The symmetry group of the pentagon is denoted by D_{10} and called the *dihedral group of order 10*. The group of rotations of the pentagon is denoted by C_5 and called the *cyclic group of order 5*.

In general, we define $D_{2n} = \{g \in S_n : g \text{ gives an isometry of the regular } n\text{-gon}\}$, and can show that D_{2n} is a permutation group called the *dihedral group of order* $2n$. We can describe its elements as follows. Let $\rho \in D_{2n}$ be a rotation of $\frac{2\pi}{n}$ and σ be a reflection. Then we can show that each element of D_{2n} can be written uniquely in the form $\sigma^j \circ \rho^i$, where $i \in \{0, 1, 2, \ldots, n-1\}$ and $j \in \{0, 1\}$. Also we can show that $\rho \circ \sigma = \sigma \circ \rho^{-1}$, and that $\rho^n = \mathrm{id}$ and $\sigma^2 = \mathrm{id}$. So $D_{2n} = \{\sigma^j \circ \rho^i : 0 \leq i < n, 0 \leq j \leq 1\}$, and we have $\sigma \circ \rho = \rho^{-1} \circ \sigma$, $\rho^n = \mathrm{id}$ and $\sigma^2 = \mathrm{id}$. Also the rotations in D_{2n} form a permutation group $C_n = \{\rho^i : 0 \leq i < n\}$, which is referred to as the *cyclic group of order* n.

(c) We can consider the isometry group of shapes in three dimensions too. For example, we could consider the Platonic solids, namely the tetrahedron, the hexahedron, the

octahedron, the dodecahedron and the icosahedron. The names of the Platonic solids give the number of faces, with a Greek numerical prefix. Note that we usually say cube rather than hexahedron. The Platonic solids are the convex regular polyhedra: the word convex here means that the straight line between any two points in the polyhedron lies entirely inside the polyhedron; and regular means that all of the faces are congruent regular polygons and each vertex belongs to the same number of edges.

You can read more about Platonic solids and find diagrams in Chapter 9 of the recommended text book for this course by M. W. Liebeck. In particular, this includes that proof that the Platonic solids are the only convex regular polyhedra, which is a really nice piece of mathematics. Alternatively another first place to find more about the Platonic solids, including diagrams, is https://en.wikipedia.org/wiki/Platonic_solid.

We can describe the symmetry groups of each of the Platonic solids, though we don't go into detail here and just give the number of elements of the isometry group of each Platonic solids.

- The symmetry group of the tetrahedron has 24 elements, and the rotation group has 12 elements.
- The symmetry group of the cube has 48 elements, and the rotation group has 24 elements.
- The symmetry group of the octahedron has 48 elements, and the rotation group has 24 elements.
- The symmetry group of the dodecahedron has 120 elements, and the rotation group has 60 elements.
- The symmetry group of the icosahedron has 120 elements, and the rotation group has 60 elements.

For example, we could work out the number of rotations of the dodecahedron by arguing as follows. If we fix a face, then there are 12 choices for where this face can be sent (because it can be sent to any of the other faces), and there are 5 choices for how this face is placed (corresponding to the 5 rotations of the regular pentagon).

A couple of observations from the above are given below.

- The group of rotations always has half the number of elements of the symmetry group. This was also the case for symmetries of regular n-gons in 2 dimensions.
- The isometry group of the cube has the same number of elements as that of the octahedron; as is also the case for the isometry groups of the dodecahedron and the icosahedron. There is a reason for this, and it is that these permutation groups are essentially the same: if we view the isometry group of the octahedron as permutations on the 6 vertices, it turns out to be the same as the isometry group of the cube viewed as permutations of the faces; and similarly for the icosahedron and the dodecahedron.

(d) Permutation groups can be used to study and help to solve certain puzzles, for example the Rubik's cube. There is quite a lot in the literature about this and a first place you

could find out more is
https://en.wikipedia.org/wiki/Rubik's_Cube_group.

(e) Permutation groups are used to study the symmetries of molecules, which gives applications in chemistry. We don't go into this here, but it is an interesting topic that you may want to look up.

5.2 Groups

The permutation groups that we have studied in the previous section consist of a set (of permutations) and a way that we can "combine them" (by composition). You have encountered many other examples of sets in mathematics with a way to combine the elements, for example by addition or multiplication. We're going to move on to consider abstract groups, which gives us a unified way of considering lots of examples, which share some key properties.

First we require the definition of a binary operation

Definition 5.2. Let A be a set. A *binary operation* on A is a function

$$* : A \times A \to A.$$

For $a, b \in A$, we write $a * b$ for the image of $(a, b) \in A \times A$ under $*$ (instead of $*(a, b)$).

Informally a binary operation $*$ on a set A is a way of combining two elements of A. This may seem a bit abstract at the moment, but will make more sense in context below.

Now we're ready to give the definition of group.

Definition 5.3. A *group* is a set G along with a binary operation $*$ satisfying the following axioms.

(G0) For all $g, h \in G$, $g * h \in G$.
(closure)

(G1) For all $g, h, k \in G$, $(g * h) * k = g * (h * k)$.
(associative law)

(G2) There exists $e \in G$ such that for all $g \in G$, $g * e = g = e * g$.
(existence of identity)

(G3) For all $g \in G$ there exists $g^{-1} \in G$ such that $g * g^{-1} = e = g^{-1} * g$.
(existence of inverses)

On the right we include the names of the axioms that must be satisfied by a group.

Sometimes we write $(G, *)$ instead of just G for a group to show what the binary operation is.

We give some remarks about the definition of a group below. These remarks are helpful for working with groups, but we go through them quickly here, as we want to get on to saying more about examples of groups. You shouldn't spend long looking at these now, as they'll make more sense once you've seen some examples of groups. In fact, I

would suggest that you only look at the first two bullet points below at first, and come back to look at the others later. The short proofs given in some of these remarks are not examinable, but just included in these notes for completeness, and you should focus more on just understanding the statements.

- It is important to remember that the binary operation is part of the definition of a group, and the axioms are too. When we speak about a group G, we implicitly understand that there is a binary operation; and we are not just thinking of G as a set. As mentioned above , we sometimes we write $(G, *)$ rather than just G to specify the binary operation, or we say that "G is a group under $*$" or something similar to clarify which binary operation we are considering.

- **You should learn the definition of a group and remember that this includes the axioms. Stating this definition will be a question on the exam.**

- The binary operation $*$ is often called *multiplication*, but we will see in the examples in Section 5.3 that it can be other things. Sometimes it may be the case that we use a different notation for the binary operation, for example sometimes the binary operation is addition and so we denote it by $+$ and sometimes it really is multiplication and we denote it by $\cdot$ or just by juxtaposition.

- There is a "generalized associativity" for the binary operation $*$ in a group. We don't go into any detail about this here, but just note this it tells us that we can unambiguously write a product $g_1 * g_2 * \cdots * g_n$, where $g_1, g_2, \ldots, g_n$ are elements of a group G. You'll see a similar statement for addition and multiplication of complex numbers in the module 1VGLA, and it can be proved in exactly the same way for any group.

- The element e from axiom (G2) is a special element of the group G called the *identity of* G. We can prove that this element is unique as follows:
  Suppose that $e' \in G$ is another identity element in G, then $e = ee' = e'$.

- Given $g \in G$ the element g^{-1} from axiom (G3) is unique, which justifies the notation. We refer to g^{-1} as the *inverse of* g. The proof of the uniqueness goes as follows:
  Suppose that $h, h' \in G$ satisfy $gh = e$ and $h'g = e$. Then $h' = h'e = h'gh = eh = h$.

- The axiom (G0) is not strictly necessary, as the definition of a binary operation ensures that it is automatically satisfied. It is useful to have it there to help us remember to check that $*$ really is a binary operation on G.

Now we move on to the definition of an abelian group.

Definition 5.4. Let G be a group. We say that G is an *abelian group* if the following additional axiom is satisfied.

(G4) For all $g, h \in G$, $g * h = h * g$. (commutative law)

A group that is not abelian, is called a *nonabelian group*.

Before moving on to see some examples, we define of the order of a finite group. Let G be a finite group, by which we mean G has finitely many elements, then the *order of* G is defined to be the number of elements of G, and is denoted by $|G|$.

5.3 Examples of groups

Later we'll see that permutation groups are indeed groups, and this provides us with many examples of groups. First we'll give other examples of groups from familiar number systems.

Examples 5.5. The following are all examples of groups. The axioms are familiar properties, so we do not include details here; in fact they are all abelian groups. You should go through them to convince yourself that the axioms do hold.

- $\mathbb{Z}$ is a group under addition: the identity is 0 and the inverse of $x \in \mathbb{Z}$ is $-x$.
- $\mathbb{Q}$ is a group under addition.
- $\mathbb{Q} \setminus \{0\}$ is a group under multiplication: the identity is 1, and the inverse of $x \in \mathbb{Q} \setminus \{0\}$ is $\frac{1}{x}$.
- $\mathbb{R}$ is a group under addition.
- $\mathbb{R} \setminus \{0\}$ is a group under multiplication.

We note that $\mathbb{Z} \setminus \{0\}$ is not a group under multiplication, and you should think about why this is.

If you know about complex numbers, then you can add the following two groups to the list.

- $\mathbb{C}$ is a group under addition.
- $\mathbb{C} \setminus \{0\}$ is a group under multiplication.

If you've not already learnt about complex numbers, then you'll cover them in the module 1VGLA.

We move on to show that we get some groups from modular arithmetic. First we state that $\mathbb{Z}_n$ with the binary operation addition gives a group. We do not include a proof, as the proposition is a consequence of the properties of $\mathbb{Z}_n$ given in Section 3.7.

Proposition 5.6. *Let $n \in \mathbb{N}$. Then $\mathbb{Z}_n$ is a group under addition.*

We note that $\mathbb{Z}_n$ is in fact an abelian group under addition, as axiom (G4) is given by the property (A4) of $\mathbb{Z}_n$ given in Section 3.7.

Multiplication is also a binary operation on $\mathbb{Z}_n$, so we can ask whether $\mathbb{Z}_n$ with multiplication gives a group. We immediately see that this is not possible if we include $[0]_n$, as this has no multiplicative inverse. In general, it is not enough to just omit $[0]_n$ as some nonzero elements of $\mathbb{Z}_n$ do not have multiplicative inverses; for example, there is no a multiplicative inverse of $[2]_4 \in \mathbb{Z}_4$. However, it is enough when n is prime.

Proposition 5.7. *Let $p \in \mathbb{N}$ be prime. Then $\mathbb{Z}_p \setminus \{[0]_p\}$ is a group under multiplication.*

Proof. We need to check the axioms:
(G0). Let $x, y \in \mathbb{Z}_p \setminus \{[0]_p\}$, and let $x_0, y_0 \in \mathbb{Z}$ such that $x = [x_0]_p$ and $y = [y_0]_p$.
Then $p \nmid x_0$ and $p \nmid y_0$.
Thus $p \nmid x_0 y_0$ by Theorem 2.19.
Therefore, $x \cdot y = [x_0 y_0]_p \neq [0]_p$, so $x \cdot y \in \mathbb{Z}_p \setminus \{[0]_p\}$.
Axioms (G1) and (G2) are properties given in Section 3.7, where for (G2) we have $e = [1]_p$.
(G3). Let $x \in \mathbb{Z}_p \setminus \{[0]_p\}$, and let $x_0 \in \mathbb{Z}$ such that $x = [x_0]_p$. Then $p \nmid x_0$, so x_0 is coprime to p. Therefore, by Theorem 3.10 there exists $y_0 \in \mathbb{Z}$ such that

$$x_0 y_0 \equiv 1 \bmod p.$$

So for $x^{-1} = [y_0]_p$, we have $x \cdot x^{-1} = [x_0 y_0]_p = [1]_p = x^{-1} \cdot x = [1]_p$.
We note that as $x_0 y_0 \equiv 1 \bmod p$, we have $p \nmid x_0 y_0$ and thus $p \nmid y_0$, so that $y \in \mathbb{Z}_p \setminus \{[0]_p\}$.
Hence, (G3) is true. □

We note that $\mathbb{Z}_p \setminus \{[0]_p\}$ is in fact an abelian group under multiplication, as axiom (G4) is given by the property (M4) of $\mathbb{Z}_p$ given in Section 3.7.

For $n \in \mathbb{N}$, we define $U(\mathbb{Z}_n) = \{[a]_n \in \mathbb{Z}_n : \text{hcf}(a, n) = 1\}$, i.e. $U(\mathbb{Z}_n)$ consists of the congruence classes of the integers that are coprime to n. We can prove that $U(\mathbb{Z}_n)$ is a group under multiplication by generalizing the arguments in the proof of Proposition 5.7, and leave this as an exercise.

We'll next verify that the permutations of a set give a group under composition. It turns out that this is a consequence of Proposition 4.2.

Proposition 5.8. *Let Ω be a set. Then $\text{Sym}(\Omega)$ is a group under composition.*

Proof. We need to check the axioms, which we do in turn.
(G0). Let $f, g \in \text{Sym}(\Omega)$. Then by Proposition 4.2(a), we have $f \circ g \in \text{Sym}(\Omega)$. Thus (G0) is true.
(G1). Let $f, g, h \in \text{Sym}(\Omega)$. Then by Proposition 4.2(b), we have $(f \circ g) \circ h = f \circ (g \circ h)$. Thus (G1) is true.
(G2). Let $f \in \text{Sym}(\Omega)$. Then $f \circ \text{id}_\Omega = f = \text{id}_\Omega \circ f$, by Proposition 4.2(c), and clearly we have $\text{id}_\Omega \in \text{Sym}(\Omega)$. Thus (G2) is true, where $e = \text{id}_\Omega$.
(G3). Let $f \in \text{Sym}(\Omega)$. Then $f^{-1} \in \text{Sym}(\Omega)$ and $f \circ f^{-1} = \text{id}_\Omega = f^{-1} \circ f$ by Proposition 4.2(d). Hence, (G3) is true. □

In fact we can generalize the proof above to show that any permutation group is indeed a group. Rather than going in to the details, we just note that the conditions (PG0), (PG2) and (PG3) are precisely what we need to make this work. Therefore, the permutation groups that we saw in Section 5.1 give lots of examples of groups. We leave it as an exercise to show that a permutation group on Ω is a subgroup of $\text{Sym}(\Omega)$, which implies that it is a group; we'll learn about subgroups in Section 5.5.

There is a theorem in group theory, which we do not cover in this course, called Cayley's theorem giving a converse of the statement above. This theorem says that any group can be viewed as a permutation group on some set.

We discuss one more important class of examples of groups, namely those given by matrices. If you have not already done so, then you will learn about matrices in 1VGLA,

you can omit this next example for now if you have not learnt about matrices. Here we restrict to matrices with entries in $\mathbb{R}$, though once you have learnt about fields, you'll be able to see that we could just as well work with matrices over any field.

We use the notation $\mathrm{GL}_n(\mathbb{R})$ for the set of all $n \times n$ invertible matrices with entries in $\mathbb{R}$. Equivalently $\mathrm{GL}_n(\mathbb{R})$ is the set of all $n \times n$ matrices with nonzero determinant. Multiplication of matrices gives a binary operation on $\mathrm{GL}_n(\mathbb{R})$ and we can check that axioms (G0)–(G3) hold. Therefore, $\mathrm{GL}_n(\mathbb{R})$ is a group under multiplication and it is referred to as the *general linear group of degree n over $\mathbb{R}$*. We do not go into the details of checking the axioms, as you will cover all of this in 1VGLA.

Notation: We have seen several examples of groups, and that the binary operation in these groups can be different things. It can be addition, multiplication or composition of functions, and can be other things in other examples. This means that the notation that we use can vary in different examples, as we explain below. This may sound a bit confusing at first, but don't worry about it, as it shouldn't cause any problem and it should make sense more when you're working in specific examples.

In general the binary operation tends to have more of a multiplicative flavour, so we usually choose to denote it simply by juxtaposition; thus for g and h in a group G, we denote their product by gh. This is the notation that we use in the remainder of this chapter. There will be exceptions to this, for example the binary operation in an abelian group is often denoted by $+$. We'll make sure that we explain if we're using a different notation for the binary operation.

A remark about the notation used for the identity element of a group is also helpful here. In general we use e for the identity element of a group G. However, in many examples of groups the identity element is something that we already have a name for, and then we continue to use that. For example, when we are considering a multiplicative group then the identity element is usually written as 1, and in an additive group the identity element is usually written as 0.

5.4 Orders of elements of groups

In this short section, we are going to define and discuss the order of an element of a group. Before doing this we have to explain how to take powers of elements in a group, which is similar to taking powers of a number. This is all very similar to what we covered about powers and orders of permutations, but it doesn't do any harm to go through it again here. Remember that as explained above we will denote the binary operation in a group simply by juxtaposition; thus for g and h in a group G, we denote their product by gh.

Let G be a group, $g \in G$ and $r \in \mathbb{Z}$. We define g^r as follows.

- For $r = 0$, we set $g^0 = e$.
- For $r > 0$, we set $g^r = gg \cdots g$, where there are r factors all equal to g.
- For $r < 0$, we let $s = -r$, so $s > 0$ and then set $g^r = (g^{-1})^s$.

We have the familiar elementary properties of powers given in the lemma below. It can be proved in exactly the same way as it would be proved for powers of numbers.

Lemma 5.9. *Let G be a group, let $g \in G$, and let $r, s \in \mathbb{Z}$. Then*

(a) $g^r g^s = g^{r+s}$

(b) $(g^r)^s = g^{rs}$, *in particular* $g^{-r} = (g^r)^{-1}$.

Next we give the definition of the order of element of a group.

Definition 5.10. Let G be a group and $g \in G$. If there exists $m \in \mathbb{N}$ such that $g^m = e$, then we say that g has *finite order*. The least such m is called the *order* of g and is denoted $o(g)$.

We note that the identity element e has order 1 in any group G, and is the only element of G with order 1.

We saw how to work out the orders of elements of S_n in Section 4.10. A good exercise it to determine the orders of elements of $\mathbb{Z}_n$ under addition.

We next have a lemma about orders. Part (b) of this lemma is proved in exactly the same way as we proved Lemma 4.24, but there is no harm including the details again here.

Lemma 5.11. *Let G be a group, $g \in G$ and $a \in \mathbb{Z}$.*

(a) *Suppose that G is finite. Then g has finite order.*

(b) *Suppose that g has finite order with $o(g) = m$. Then $g^a = e$ if and only if $m \mid a$.*

Proof. (a) Let $|G| = n$, and consider the elements $e = g^0, g = g^1, g^2, g^3, \ldots, g^n$ of G. These elements cannot all be distinct, so there exist $k, l \in \mathbb{N}$ such that $k < l$ and $g^k = g^l$. Then we have $g^{l-k} = e$ and $l - k \in \mathbb{N}$. Hence, g has finite order.

(b) Using the division theorem we can write $a = qm + r$, where $q, r \in \mathbb{Z}$ and $0 \leq r < m$. Then we have $g^a = g^{qm+r} = (g^m)^q g^r = e^q g^r = e g^r = g^r$.

Suppose that $g^a = e$, then we have $g^r = e$. But $0 \leq r < m$ and $m = o(g)$ so m is the smallest natural number such that $g^m = e$. Thus $r = 0$, and we have $a = qm$, so that $m \mid a$.

Conversely, suppose that $m \mid a$. Then we have $r = 0$, so that $g^a = g^r = e$. □

We include some examples of orders of elements in the groups $\mathbb{Z}$ and $\mathbb{Z}_n$ under addition and in the group $\mathbb{Z}_5 \backslash \{[0]$ under multiplication. The numbering coincides with the previous lemma as these examples were added to the notes late.

Examples 5.11. (a) We consider $\mathbb{Z}$ as a group under addition and $1 \in \mathbb{Z}$. The notation here takes a bit of time to get our head around as the binary operation is addition and the identity element is 0. Since the binary operation is $+$, the nth power of 1 is $1+1+\cdots+1$ where there are n summands, so that the nth power of 1 is n for $n \in \mathbb{N}$. Since $n \neq 0$ for $n \in \mathbb{N}$ we have that 1 has infinite order in the group $\mathbb{Z}$ under addition.

Similarly any $a \in \mathbb{Z}$ has infinite order.

(b) Let $n \in \mathbb{N}$. We consider $\mathbb{Z}_n$ as a group under addition and $[1]_n \in \mathbb{Z}_n$. Similarly to the case of $\mathbb{Z}$ considered above we have that the mth power of $[1]_n$ is $[m]_n$ for $m \in \mathbb{N}$, and we recall that the identity element is $[0]_n$. Since the smallest $m \in \mathbb{N}$ such that $[m]_n = [0]_n$ is $m = n$, we see that $o([1]_n) = n$ in the group $\mathbb{Z}_n$ under addition.

More generally we can show that $o([a]_n) = \frac{n}{\text{hcf}(a,n)}$ for any $[a]_n \in \mathbb{Z}_n$.

(c) We consider $\mathbb{Z}_5 \setminus \{[0]_5\}$ as a group under multiplication. We recall that the identity element is $[1]_5$.
We have $o([1]_5) = 1$.
We calculate $([2]_5)^2 = [4]_5$, $([2]_5)^3 = [8]_5 = [3]_5$, $([2]_5)^4 = [16]_5 = [1]_5$, so that $o([2]_5) = 4$.
We calculate $([3]_5)^2 = [9]_5 = [4]_5$, $([3]_5)^3 = [27]_5 = [2]_5$, $([3]_5)^4 = [81]_5 = [1]_5$, so that $o([3]_5) = 4$.
We calculate $([4]_5)^2 = [16]_5$, so that $o([4]_5) = 2$.

5.5 Subgroups and cyclic groups

We move on to consider subgroups of a group. We begin with the definition of a subgroup.

Definition 5.12. Let G be a group and let H be a subset of G. We say that H is a *subgroup* of G if it is a group with the same binary operation as G.

We write $H \leq G$ to mean that H is a subgroup of G.

To check whether a subset H of a group G is a subgroup, we need to check that the axioms of a group hold for H. We'll go through the axioms to see what this involves.

(G0) We need to check that for all $h, k \in H$, we have $hk \in H$.

(G1) This axioms holds, because it holds in G.

(G2) We need to check that $e \in H$.

(G3) Given $h \in H$, we need to check that $h^{-1} \in H$.

This leads us immediately to a test to determine whether a subset of a group G is a subgroup. It is called the subgroup test and we state it below.

Lemma 5.13 (The subgroup test)**.** *Let G be a group and let H be a subset of G. Then H is a subgroup of G provided*

(SG1) *$e \in H$; and*

(SG2) *for all $h, k \in H$, we have $hk \in H$*

(SG3) *for all $h \in H$, we have $h^{-1} \in H$.*

By comparing our definition of a permutation group, with the definition of a subgroup. We can observe that, for a set Ω, a permutation group on Ω is a subgroup of $\text{Sym}(\Omega)$ and we leave it as an exercise to give the details for showing this. Conversely, we can see that a subgroup of $\text{Sym}(\Omega)$ is a permutation group on Ω. So permutation groups are the "same thing" as subgroups of symmetric groups. This gives us lots of examples of subgroups.

We include quick example of using the subgroup test to show that a certain subset H of S_6 is indeed a subgroup (and thus H is a permutation group on $\{1, 2, 3, 4, 5\}$). The numbering of the example coincides with that of the previous lemma as this example was added to the notes late.

Example 5.13. Let $G = S_6$ and $H = \{g \in S_6 : g(3) = 3\}$. We show that H is a subgroup is a subgroup of G using the subgroup test.

(SG1) e is the identity function, so $e(3) = 3$ and $e \in H$.

(SG2) Let $g, h \in H$.
Then $g(3) = 3 = h(3)$.
So $(g \circ h)(3) = g(h(3)) = g(3) = 3$.
Thus $g \circ h \in H$.

(SG3) Let $h \in H$.
Then $h(3) = 3$.
So $h^{-1}(h(3)) = h^{-1}(3)$ and so $h^{-1}(3) = 3$.
Thus $h^{-1} \in H$.

We move on to consider a special type of subgroup of a group.

Definition 5.14. Let G be a group and let $g \in G$. Let $\langle g \rangle = \{g^m : m \in \mathbb{Z}\}$. We refer to $\langle g \rangle$ as the *subgroup generated by* g.

We have implicitly said that $\langle g \rangle$ is a subgroup of G in the definition above, but we should really check that this is the case, and this is part of the following lemma. The proof is left as an exercise. You can prove (a) by applying the subgroup test.

Lemma 5.15. *Let G be a group and let $g \in G$. Then*

(a) $\langle g \rangle = \{g^m : m \in \mathbb{Z}\}$ *is a subgroup of G.*

(b) *If g has finite order, then $\langle g \rangle$ is finite and $|\langle g \rangle| = o(g)$.*

Some groups are generated by a single element. These groups, which are defined next, are in a sense easy groups, but there are fundamental in understanding groups more generally.

Definition 5.16. A group G is called a *cyclic group* if $G = \langle g \rangle$ for some $g \in G$.

Let's look at some examples of cyclic groups and of subgroups.

Examples 5.17. (a) The additive group of $\mathbb{Z}$ is a cyclic group. It is generated by $1 \in \mathbb{Z}$, which is of infinite order. Remember that $m = 1 + 1 + \cdots + 1$, where there are m summands, is the mth power of 1 in the additive group of $\mathbb{Z}$, because we use additive notation. The negative integers are the negative powers of 1.

Next we consider subgroups of $\mathbb{Z}$ as a group under addition.
We can show that for any $m \in \mathbb{N}$, we have that $m\mathbb{Z} = \{ma : a \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$ by using the subgroup test. Moreover, it is possible to prove that any subgroup of $\mathbb{Z}$ is equal to $m\mathbb{Z}$ for some $m \in \mathbb{Z}$, but we do not go into that here.

(b) Let $n \in \mathbb{N}$. Then the additive group of $\mathbb{Z}_n$ is cyclic of order n. It is generated by $[1]_n$, which has order n.

We can find all the subgroups of $\mathbb{Z}_n = \{[a]_n : a \in \mathbb{Z}\}$. Let $m \in \mathbb{N}$ be a factor of n and define $m\mathbb{Z}_n = \{[ma]_n : a \in \mathbb{Z}\}$. Using the subgroup test we can show that $m\mathbb{Z}_n$ is a

subgroup of $\mathbb{Z}_n$. Also we leave it as an exercise to check that the order of $m\mathbb{Z}_n$ is $\frac{n}{m}$. So we can observe that the order of any subgroup of $\mathbb{Z}_n$ divides $|\mathbb{Z}_n| = n$.

(c) In Examples 5.11(c) we showed that in the group $\mathbb{Z}_5 \setminus \{[0]_5\}$ under multiplication, the order of $[2]_5$ is 4. It follows that the cyclic subgroup generated by $[2]_5$ has order 4 and thus must be equal to $\mathbb{Z}_5 \setminus \{[0]_5\}$. Therefore, we have that $\mathbb{Z}_5 \setminus \{[0]_5\}$ is a cyclic group.

(d) The group of rotations C_n of a regular n-gon is a cyclic group. In the notation of Example 5.1, we have $C_n = \langle \rho \rangle$, where ρ is the rotation by $\frac{2\pi}{n}$ radians.

We have that C_n is a subgroup of D_{2n} and that D_{2n} is a subgroup of S_n.

(e) We have seen that $\mathrm{GL}_2(\mathbb{R})$ is a group. Also matrices in $\mathrm{GL}_2(\mathbb{R})$ can be used to represent linear transformations of $\mathbb{R}^2$; you may not have seen this yet, but it will be covered in 1VGLA. In this way we can represent the symmetries of a square (with centre at the origin) as a subgroup of $\mathrm{GL}_2(\mathbb{R})$. In this way the group of symmetries of the square is

$$\left\{ \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix} \right\},$$

and this is a subgroup of $\mathrm{GL}_2(\mathbb{R})$. We could write down the matrices representing the elements of D_{2n} is a similar way for any n, but leave this as an exercise.

We can also represent the symmetries of the Platonic solids (which we covered in Examples 5.1) by matrices in $\mathrm{GL}_3(\mathbb{R})$. This is quite straightforward to do for the cube (and the octahedron), but more complicated for the other Platonic solids. We don't go into this here.

(f) Let G be a group. The trivial group $\{e\}$ is a subgroup of G, and G is a subgroup of itself.
You should think about why these are subgroups. Once you've thought about it for long enough you should hopefully see that it is trivial.

A remark about motivation for the theory of (abstract) groups can be made here. In Examples 5.17(d) above, we saw the symmetries of a square can be given by matrices, whereas earlier on we viewed them as permutations of the vertices. So these are just two ways of writing down the same thing, and we shouldn't really think about them as being different. By studying groups abstractly, we can think of groups without reference to an object that it acts on as symmetries, and this abstraction turns out to be very powerful. It may allow us to see when two groups are essentially the same even though they may not look the same at first. As another example the symmetry group of the dodecahedron and the icosahedron are "the same"! I'll leave you to ponder this, but you can ask if you're interested. You'll also be able to learn more about these ideas in future courses. In particular, you'll be able to learn what it means for two groups to be *isomorphic*. The term isomorphic translates to "same shape" and is used to mean that two groups are abstractly the same, though the notation for their elements may be different.

5.6 Lagrange's theorem and consequences

We move on to state Lagrange's theorem, which is a highlight of this chapter on group theory. In some examples, we have seen that the orders of subgroups of a finite group G

divide the order of G, and Lagrange's theorem tells us that this is always true. We only give a sketch of the proof here, as this involves the language of cosets, which we do not cover in this course.

Theorem 5.18 (Lagrange's theorem)**.** *Let G be a finite group and let H be a subgroup of G. Then $|H|$ is a factor of $|G|$.*

Sketch of proof. Let $g \in G$. We define $gH = \{gh : h \in H\}$, which is called a coset of H in G.

We can prove that $|gH| = |H|$ by noting that the function $H \to gH$ given by $h \mapsto gh$ is a bijection.

Also we can see that $G = \bigcup_{g \in G} gH$, because $g \in gH$.

Moreover, for $x, y \in G$ we can show that $xH = yH$ or $xH \cap yH = \varnothing$.

Therefore, we can choose $g_1, g_2, \ldots, g_r$, where $r = |\{gH : g \in G\}|$, such that $G = g_1H \cup g_2H \cup \ldots \cup g_rH$ and $g_iH \cap g_jH = \varnothing$ for $i \neq j$.

Therefore,

$$\begin{aligned} |G| &= \sum_{i=1}^{r} |g_iH| \\ &= \sum_{i=1}^{r} |H| \\ &= r|H|. \end{aligned}$$

Hence, $|H|$ is a factor of $|G|$. □

Let G be a finite group and let $g \in G$. The subgroup $\langle g \rangle$ of G generated by g is defined in Definition 5.14, and we have that it is a subgroup with $|\langle g \rangle| = o(g)$ by Lemma 5.15. We thus obtain the following corollary as a consequence of Lagrange's theorem.

Corollary 5.19. *Let G be a finite group and let $g \in G$. Then $o(g)$ is a factor of $|G|$. In particular, $g^{|G|} = e$.*

Proof. We apply Lagrange's theorem to the subgroup $\langle g \rangle$ of G, and we obtain that $o(g) = |\langle g \rangle|$ is a factor of $|G|$.

Let $m = o(g)$ and $|G| = ml$, where $l \in \mathbb{N}$. Then $g^{|G|} = g^{ml} = (g^m)^l = e^l = e$. □

We also have the following rather nice corollary of Lagrange's theorem. We leave the proof as an exercise.

Corollary 5.20. *Let $p \in \mathbb{N}$ be a prime and let G be a finite group of order p. Then G is cyclic.*

Now we move on to show that Fermat's little theorem can be deduced as a consequence of Lagrange' theorem for the multiplicative group $\mathbb{Z}_p \setminus \{[0]_p\}$. As mentioned earlier this is a really nice theorem, and we see how Lagrange's theorem leads to a neat proof. We know that $\mathbb{Z}_p \setminus \{[0]_p\}$ is a group under multiplication by Proposition 5.7.

Theorem 5.21 (Fermat's little theorem)**.** *Let $p \in \mathbb{N}$ be a prime, and let $a \in \mathbb{Z}$. Suppose that $p \nmid a$. Then $a^{p-1} \equiv 1 \bmod p$.*

Proof. We have $[a]_p \in \mathbb{Z}_p \setminus \{[0]_p\}$ and $|\mathbb{Z}_p \setminus \{[0]_p\}| = p - 1$. Therefore, by Corollary 5.19, we have $([a]_p)^{p-1} = [1]_p$, remember that $[1]_p$ is the identity element in $\mathbb{Z}_p \setminus \{[0]_p\}$. Therefore, $[a^{p-1}]_p = [1]_p$, so that $a^{p-1} \equiv 1 \bmod p$. □

After showing $\mathbb{Z}_p \setminus \{[0]_p\}$ is a group under multiplication, we mentioned that in general $U(\mathbb{Z}_n)$ is a group under multiplication. This can be used to prove a more general theorem than Fermat's little theorem, which is known as *Euler's theorem*. This theorem states that for any $n \in \mathbb{N}$ and any $a \in \mathbb{Z}$, which is coprime to n, we have $a^{\phi(n)} \equiv 1 \bmod n$, where $\phi(n) = \{x \in \mathbb{N} : 0 < x < n \text{ and } x \text{ is coprime to } n\}$. It would be a good exercise for you to think about how you would prove this.

5.7 Groups and polynomial equations

We are going to briefly describe a particularly important and wonderful application of group theory. Below, we use the letter x to denote a variable and the letters a, b, c, d, e for scalars, which are real or complex numbers and $a \neq 0$.

Linear equations

A linear equation is of the form

$$ax + b = 0.$$

It is easy to solve a linear equation, we just rearrange to get

$$x = -\frac{b}{a}.$$

Quadratic equations

We know that there is a formula to solve a quadratic equation. The solutions of

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

are given by

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

Cubic equations

A cubic equation is of the form

$$ax^3 + bx^2 + cx + d = 0.$$

If we are interested in solving cubic equations, then we can "complete the cube", which gives a "reduction" so that we only need to consider equations of the form

$$x^3 + bx + c.$$

Then there is a formula giving a solution:

$$x = \sqrt[3]{-\frac{1}{2}c + \sqrt{\frac{c^2}{4} + \frac{b^3}{27}}} + \sqrt[3]{-\frac{1}{2}c - \sqrt{\frac{c^2}{4} + \frac{b^3}{27}}}.$$

Once, we have one solution we can factorize and use the formula for quadratic equations to find the other solutions.

Quartic equations

A quartic equation is of the form

$$ax^4 + bx^3 + cx^2 + dx + e = 0.$$

There is a formula for solving quartic equations, but it is too complicated to write down here. You can find out more on
http://en.wikipedia.org/wiki/Quartic_equation#Solving_a_quartic_equation.

Quintic equations

For quintic equations

$$ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0,$$

there is no formula giving the solutions!
Big surprise!
Here is a brief outline of why not, saying where group theory comes in.
To any polynomial equation we can associate a group called the *Galois group*. This Galois group gives "the symmetries of the roots of the polynomial".
It is possible to show that there is a formula giving solutions of the polynomial equation if and only if the Galois group has an "uncomplicated structure".
When the polynomial is of degree 4 or less, then this Galois group has an "uncomplicated structure".
There are quintic polynomials that have a "complicated" Galois group, so there is no formula giving their solutions.

This is only a very brief glimpse at a fascinating area of mathematics called Galois theory, where group theory is really important.

5.8 Summary of Chapter 5

By the end of this chapter you should be able to:

- understand the definition of permutation groups;
- determine the elements of isometry groups as permutation groups, and be able to calculate in them;
- understand and recall the definition of a group–**this includes the axioms**;
- understand some examples of groups, and calculate and check axioms in examples;
- prove that axioms hold for groups from modular arithmetic;
- understand the definition of the order of an element of a group and be able to calculate it in examples;
- understand and recall the definition of a subgroup and apply the subgroup test in examples;
- understand and recall the definition of cyclic groups, and determine whether examples of groups are cyclic;
- understand and apply Lagrange's theorem and its consequence about orders of elements in a group; and
- apply the material in the chapter to solve problems and prove related statements.

Appendix A

Functions

As we have used functions in Chapter 4 where we studied permutations, we include a recap on functions in this appendix. You learned about functions in 1RA and should have covered all of the material here. As this appendix is quite brief in places you may benefit by looking in other places for more details. It is convenient to have this appendix here, as we can refer to it in Chapter 4.

A.1 Functions

We begin with the definition of a function.

Definition A.1. A *function* f consists of three things:

(a) a set $A = \mathrm{dom}(f)$ called the *domain of* f;

(b) a set $B = \mathrm{codom}(f)$ called the *codomain of* f; and

(c) a rule that assigns to each element $a \in A$ a unique element $f(a) \in B$.

We write $f : A \to B$ to mean that f is a function with domain A and codomain B, and say that f is a function from A to B.
Given $a \in A$, we say that $f(a)$ is the *image of* a *under* f.
The *image of* f is defined to be

$$\mathrm{im}(f) = \{b \in B : \text{there exists } a \in A \text{ such that } b = f(a)\}.$$

We give some examples of functions.

Examples A.2. Let $A = \{2, 4, 6, 8\}$, $B = \{1, 2, 3, 4, 5\}$, $C = \{-2, -1, 0, 1, 2\}$ and $D = \{0, 1, 4\}$.

(a) Define $f : A \to B$ by

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

(b) Define $g : B \to C$ by

$$g(x) = x - 3.$$

(c) Define $h : C \to D$ by

$$h(x) = x^2.$$

(d) Define $k : C \to B$ by

$$k(x) = x + 3.$$

We now define what it means for two functions to be equal.

Definition A.3. Let $f : A \to B$ and $g : C \to D$ be functions. We say that f *is equal to* g and write $f = g$ if the following three conditions hold:

(a) $A = C$;

(b) $B = D$; and

(c) $f(a) = g(a)$ for all $a \in A$.

We stress that the definition says that for functions to be equal they have to have the same domain and codomain. It is not enough for them to just have the same rule defining them.

A.2 Composition of functions

Below we define the composition of two functions, which just means doing one function after the other.

Definition A.4. Let $f : A \to B$ and $g : B \to C$ be functions. The *composition of* f *and* g is the function $g \circ f : A \to C$ defined by

$$(g \circ f)(x) = g(f(x))).$$

We demonstrate composition of functions in the next example.

Example A.5. For f, g, h as in Examples A.2, we have

$$(g \circ f)(x) = \frac{x}{2} - 2 \quad \text{for } x \in A. \tag{A.1}$$

and

$$(h \circ g)(x) = (x - 3)^2 \quad \text{for } x \in B. \tag{A.2}$$

The following lemma says that composition of functions is associative.

Lemma A.6. *Let $f : A \to B$, $g : B \to C$ and $h : C \to D$ be functions. Then*

$$(f \circ g) \circ h = f \circ (g \circ h).$$

Proof. Let $a \in A$. Then we have

$$((f \circ g) \circ h)(a) = (f \circ g)(h(a)) = f(g(h(a))).$$

Similarly,

$$(f \circ (g \circ h))(a) = f(g(h(a))).$$

Hence,

$$((f \circ g) \circ h)(a) = (f \circ (g \circ h))(a).$$

This holds for all $a \in A$, so

$$(f \circ g) \circ h = f \circ (g \circ h).$$

□

A.3 Injections, surjections and bijections

We give the definition of an injections, surjections and bijections.

Definition A.7. Let $f : A \to B$ be a function. We say that:

(a) f is an *injection* if

for all $a, a' \in A$, if $f(a) = f(a')$, then $a = a'$.

An injection is sometimes called an injective function or a one-to-one function.

(b) f is a *surjection* if

for all $b \in B$, there exists $a \in A$ such that $f(a) = b$;
equivalently, $\mathrm{im}(f) = B$.

A surjection is sometimes called a surjective function or an onto function.

(c) f is a *bijection* if it is both an injection and a surjection.

A bijection is sometimes called a bijective function.

We demonstrate these concepts with some examples.

Example A.8. In Examples A.2, the functions f, g and k are injective, but h is not injective, because $h(1) = h(-1)$. The composition $g \circ f$ given in (A.1) is injective.

The functions g, h and k are surjective, but f is not surjective, because there is no $a \in A$ such that $f(a) = 1$. The composition $h \circ g$ given in (A.2) is surjective.

Therefore, the functions g and k are bijective, but f and h are not bijective.

The next lemma tells us that compositions of injective functions are injective, and similarly for surjective and bijective functions.

Lemma A.9. *Let $f : A \to B$ and $g : B \to C$ be functions.*

(a) *Suppose that f and g are injections. Then $g \circ f$ is an injection.*

(b) *Suppose that f and g are surjections. Then $g \circ f$ is a surjection.*

(c) *Suppose that f and g are bijections. Then $g \circ f$ is a bijection.*

Proof. (a) Let $a, a' \in A$ such that $(g \circ f)(a) = (g \circ f)(a')$.
Then $g(f(a)) = g(f(a'))$, so $f(a) = f(a')$, because g is injective.
Thus $a = a'$, because f is injective.
Hence, $g \circ f$ is injective.
(b) Let $c \in C$.
Since g is surjective, there exists $b \in B$ such that $g(b) = c$.
Since f is surjective, there exists $a \in A$ such that $f(a) = b$.
Therefore,

$$\begin{aligned}(g \circ f)(a) &= g(f(a)) \\ &= g(b) \\ &= c.\end{aligned}$$

Hence, $g \circ f$ is surjective.
(c) This follows immediately from (a) and (b). □

A.4 Identity functions and inverse functions

In this section we define identity functions and inverse functions.

Definition A.10. Let A be a set. The *identity function on A* is the function $\mathrm{id}_A : A \to A$ defined by $\mathrm{id}_A(x) = x$.

Next we give an elementary lemma about identity functions.

Lemma A.11. *Let $f : A \to B$ be a function. Then*

(a) $f \circ \mathrm{id}_A = f$*; and*

(b) $\mathrm{id}_B \circ f = f$.

Proof. (a) Let $a \in A$. Then

$$\begin{aligned}(f \circ \mathrm{id}_A)(a) &= f(\mathrm{id}_A(a)) \\ &= f(a)\end{aligned}$$

This holds for all $a \in A$, so $f = f \circ \mathrm{id}_A$.
(b) A similar argument proves that $\mathrm{id}_B \circ f = f$. □

Definition A.12. Let $f : A \to B$ be a bijection. The *inverse of f* is the function $f^{-1} : B \to A$ defined by

$$f^{-1}(x) \text{ is the unique element } y \in A \text{ such that } f(y) = x.$$

To justify this definition, we need the following two facts:

- there exists $y \in A$ such that $f(y) = x$, because f is surjective; and
- y is unique because f is injective.

In the next example with very quickly demonstrate an inverse function.

Example A.13. In Examples A.2, k is the inverse of g.

We have the following lemma giving some properties of inverses. The proof is left as an exercise.

Lemma A.14. *Let $f : A \to B$ be a bijection. Then*

(a) *for all $a \in A$, we have $f^{-1}(f(a)) = a$, so $f^{-1} \circ f = \mathrm{id}_A$;*

(b) *for all $b \in B$, we have $f(f^{-1}(b)) = b$, so $f \circ f^{-1} = \mathrm{id}_B$;*

(c) *f^{-1} is a bijection; and*

(d) *$(f^{-1})^{-1} = f$.*

Our next lemma is about inverses of compositions.

Lemma A.15. *Let $f : A \to B$ and $g : B \to C$ be bijections. Then*

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

Proof. Note that by Lemma A.9, $g \circ f$ is a bijection so $(g \circ f)^{-1}$ is defined.
For $x \in C$, we have

$$\begin{aligned}(g \circ f)((f^{-1} \circ g^{-1})(x)) &= (g \circ f)(f^{-1}(g^{-1}(x))) \\ &= g(f(f^{-1}(g^{-1}(x)))) \\ &= g(g^{-1}(x)) \\ &= x.\end{aligned}$$

Therefore, the definition of $(g \circ f)^{-1}$ says that

$$(g \circ f)^{-1}(x) = (f^{-1} \circ g^{-1})(x),$$

because $y = (f^{-1} \circ g^{-1})(x)$ satisfies $(g \circ f)(y) = x$. This holds for all $x \in C$, so

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

□

Note that when we take the inverse of the composition $g \circ f$ we have to swap the order. This is similar to what happens when we take inverses of matrices, i.e. if A and B are invertible $n \times n$ matrices, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. This is no coincidence, as matrices correspond to certain functions between vector spaces.