2AC: Algebra 2

School of Mathematics

UNIVERSITY OF
BIRMINGHAM

Simon Goodwin
s.m.goodwin@bham.ac.uk
Office: Watson 107

Spring term 2016

Contents

Chapter -1

Module information

Below is some practical information about this course. **There's some important stuff here, so you should read it carefully at some point.** All information and resources about the course 2AC Algebra 2 are also on the 2AC Canvas page.

-1.1 Learning arrangements

There will be 21 lectures for this course:

- Thursday 9am in Aston Webb WG5; and
- Friday 4pm in Watson LRA, except in week 11.

In addition, there will be 4 examples class:

- Friday 3pm in weeks 3, 5, 7 and 9 of term in Watson LRA.

In addition you have the following opportunities to get help with the course.

- The Peer Assisted Study Scheme PASS.
- The drop-in sessions with postgraduate teaching assistants in the maths learning centre.
- Through the maths centre drop-in support in the main library. You can find about this from https://intranet.birmingham.ac.uk/as/libraryservices/library/skills/asc/mathematicalsupport.aspx.
- During my office hours, Monday 12:00–13:30 and Thursday 10:00–11:30.

In the lectures, I'll present all of the material of the course. Sometimes it may be difficult to keep up with the pace of the lectures, so **it's really important that you spend some time looking through the notes afterwards to make sure you understand them**. You will also find it very helpful to read the typed notes before the lecture. I will write on the board in lectures, and strongly recommend that you take notes, as I will present the material slightly differently and give different examples to those in the typed notes, as well as giving more details in some of the proofs.

As explained below you will be given four sets of formative exercises during the term. and two sets of assessed exercises. The examples class gives you the opportunity to ask questions about the exercises that you are finding most challenging. There will be postgraduate teaching assistants present in the examples classes and I will be there too. You can also use the examples class to ask questions more generally about the course, and to discuss your marked work and get further feedback on it.

The course 2AC Algebra 2 is 10 credits, which means that you should expect to do a total of approximately 100 hours of work for the course. As you will be able to work out, this means that you should spend quite a lot of time working on the module outside of the lectures and examples classes, either individually or in collaboration with others taking the module.

-1.2 Exercise sheets, assessed exercises, feedback and plagiarism

In the Thursday lecture in weeks 2, 4, 6 and 8 of term you will receive exercise sheets for 2AC Algebra 2. Exercise sheet 2 will include Assessed exercises 1, and Exercise sheet 4 will include Assessed exercises 2. The exercise sheets will also be on the 2AC Canvas page. Most of the exercises are set as formative exercises, and will be taken from the exercises given at the end of each of the chapters in the notes. The assessed exercises will be in a similar format to exam questions for 2AC.

Your solutions to selected exercises should be handed in on Tuesdays in weeks 4, 6, 8 and 10 of term. The deadlines for handing in work are 3pm on:

- Tuesday 2 February 2016;
- Tuesday 16 February 2016;
- Tuesday 1 March 2016; and
- Tuesday 15 March 2016.

The work that you hand in will be marked and there will also be comments on your work giving **feedback**. In addition, you can discuss your marked work with me after the lectures, in my office hours or in the examples class for more **feedback**. A **feedback** sheet, which contains model solutions, comments on common mistakes and advice on how to improve will be put on the 2AC Canvas page.

In my opinion working through the exercises is the most effective part of the learning process, so it is very important that you make a serious effort at all the exercises. When you work through the exercises you will attain a deeper understanding of the material.

You are encouraged to work with others on the course, when you are completing the exercises. However, **you should make sure that your work is not considered to plagiarized** as set out by the university regulations, which you have signed at the start of the year. Therefore, for the assessed exercises you should make sure that after understanding how to do the exercises you write out your solutions by yourself and in your own words.

-1.3 Any questions?

If you have any questions about the module, then you are encouraged to ask me after the lecture, email your question to s.m.goodwin@bham.ac.uk, or come to my office Watson 107 during my office hours:

- Monday 12:00–13:30 and Thursday 10:00–11:30.

If your question is urgent, or my office hours are not convenient, then you can try to find me at another time and see if I am available, or you can email me to make an appointment.

-1.4 Course materials

All course materials will be available on the 2AC Canvas page.

- Typed lecture notes will be on the 2AC Canvas page, and printed copies will be handed out during term.
- Exercise sheets will be handed out during term as explained in Section -1.2, and will also be put on the 2AC Canvas page.
- Feedback sheets will be put on the 2AC Canvas page.

You will be able to make another set of notes during the lectures, and I recommend that you do this as explained in Section -1.1. Also as mentioned in that section you will benefit from reading the typed notes before the lectures, and it is very important that you study your notes after the lectures.

In preparing these notes, I have used the books listed below, especially the book by Peter Cameron. Also I used the notes from the course *203a Polynomials and Rings* written by Richard Kaye; that course is in a sense the predecessor to *2AC Algebra 2*.

-1.5 Books

Pretty much all of the material from the course is covered in

- Peter J. Cameron, *Introduction to Algebra*, Second Edition, Oxford University Press.

This is the recommended book for the course. There is electronic access to this book via findit@bham. The material of this course is mainly contained in Chapters 2 and 3 of this book. You may also find Chapter 1 useful for revising topics that you have covered in earlier courses, and the later chapters will be useful for other courses that you will be able to take.

There are a lot of other good books that cover the material in this course, many of which are in the library; for example,

- Marlow Anderson and Todd Feil, *A First Course in Abstract Algebra*, Second Edition, Chapman and Hall.

is a book that I have used in the preparation of the course. Some of the other books will treat groups before rings, which makes them less suitable, though you should be able to get round this with a small amount of effort. A book that I have used for the course is

- Niels Lauritzen, *Concrete abstract algebra*, Third Edition, Chapman and Hall.

-1.6 Assessment

2AC (Combinatorics 2 and Algebra 2 combined) is worth 20 credits. The assessment is divided up as follows:

- 80% from an examination in the summer;
- 20% from assessed exercises completed during term.

This term there will be 2 sets of assessed exercises, and you completed 2 sets of assessed exercises in Combinatorics 2 last term.

- The best 3 out of 4 of these assessment sheets (for Combinatorics 2 and Algebra 2) will contribute $6.\dot{6}\%$ each towards the final mark for the module giving the 20%.

-1.7 Syllabus

The syllabus below contains a list of topics that will be covered in the course.

- **Rings:** binary operations; axioms for a ring; rings with identity; commutative rings; fields; integral domains; examples of rings; arithmetic properties of rings; subrings.
- **Homomorphisms and ideals:** homomorphisms and isomorphisms; kernel and image of a homomorphism; ideals; principal ideals; sum and intersection of ideals; kernel of a homomorphism is an ideal; homomorphism is injective if and only if kernel is 0; cosets; quotient rings; natural homomorphism on to quotient ring; the isomorphism theorem; principal ideal domains; maximal ideals; quotients by maximal ideals.
- **Fields:** field of fractions of an integral domain; field extensions; minimal polynomials; finite fields.
- **Groups:** axioms for a group; examples of groups including groups of permutations and automorphisms; orders of groups; orders of elements of groups; subgroups; cosets; Lagrange's theorem; cyclic groups.

-1.8 Learning outcomes

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

- understand ring theory, including ideals, homomorphisms and quotient rings, and apply the theory and calculate in examples;
- understand the basic theory of fields, and make calculations in examples;
- understand basic group theory, and make calculations in examples;
- write proofs or provide counterexamples for statements concerning the material in the course;
- apply the material in the course to solve weakly-posed problems and prove related statements.

The learning outcomes above are statements of what "a learner is expected to know, understand and be able to demonstrate after completion of a process of learning". These are important as they describe what you are expected to be able to do in order to demonstrate that you have understood the course. The assessment of the course is based on these outcomes, so **if you can do the things on this list, then you should do well in the assessment.** Also at the end of each chapter, I have provided a list of more specific learning aims in a summary.

A bit of a warning

This is the final version of the lecture notes, and hopefully most typos have been spotted and corrected. However, it is likely that there will be some typos and some bits of the notes that are not set out as well as they could be. Please let me know if you spot any typos, or anything that you think may be a mistake.

Chapter 0

Recap of some material from 1AC

This course builds on the material you have learned in 1AC, and we recap some of this here. When we use these things in the lectures, we'll recap them briefly, but you will benefit from revising it yourself before. The recollection below is brief and no proofs or examples are included, so you'll have to look at your notes for 1AC to revise this material properly. One of the reasons for including this recap, is that it will be helpful to refer to some of this material later on in the notes.

I've just covered certain topics below, which are some of the important things that you will need to be on top of. However, we will use other material from 1AC, so you certainly shouldn't assume that you don't need to know this.

0.0 Notation

Before we start we recall the notation we use for certain number systems.

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

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

Note that here 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

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

(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

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

0.1 Equivalence relations

In 1AC Combinatorics 1 you learned about equivalence relations. We'll use this material, so below the main definitions and results about equivalence relations are recalled.

Definition 0.1. Let A be a set. A *relation* on A is a subset

$$R \subseteq A \times A.$$

For $a, b \in A$, we write aRb to mean $(a, b) \in R$.

Informally a relation R on a set A is a way of comparing two elements of A. Usually a relation is defined by giving a rule for when two things are related.

Next we define equivalence relations. We usually use the symbol $\sim$ rather than R to denote an equivalence relation.

Definition 0.2. Let $\sim$ be a relation on a set A. We say that $\sim$ is an *equivalence relation* if it satisfies the three properties:

(ER1) for all $a \in A$, $a \sim a$. (Reflexive property)
(ER2) for all $a, b \in A$, if $a \sim b$, then $b \sim a$. (Symmetric property)
(ER3) for all $a, b, c \in A$, if $a \sim b$ and $b \sim c$, then $a \sim c$. (Transitive property)

We now define equivalence classes.

Definition 0.3. Let A be a set, $\sim$ an equivalence relation on A and $a \in A$.

(a) The *equivalence class of* a is defined to be

$$[a]_\sim = \{x \in A : x \sim a\}.$$

(b) The *set of equivalence classes of* $\sim$ is defined to be

$$A/\!\sim\, = \{[a]_\sim : a \in A\}.$$

Note that $A/\!\sim$ is a set of subsets of A.

We move on to the definition of a partition.

Definition 0.4. Let A be a set. A *partition of* A is a set P of non-empty subsets of A such that:

(P1) for all $x \in A$ there exists $B \in P$ such that $x \in B$; and
(P2) for all $B, C \in P$, either $B = C$ or $B \cap C = \varnothing$.

We end this recap on equivalence relations with the statement of a theorem and proposition, which give an important connection between equivalence relations and partitions.

Theorem 0.5. *Let A be a set, $\sim$ an equivalence relation on A, and $a, b \in A$. Then the following hold:*

(a) $a \in [a]_\sim$;
(b) $[a]_\sim = [b]_\sim$ *if and only if* $a \sim b$;
(c) $[a]_\sim = [b]_\sim$ *or* $[a]_\sim \cap [b]_\sim = \varnothing$;
(d) $A/\sim$ *is a partition of* A.

Proposition 0.6. *Let A be a set and let $P \subseteq \mathcal{P}(A)$ be a partition. Define $\sim$ on A by*

$$a \sim b \text{ means there exists } B \in P \text{ such that } a, b \in B.$$

Then $\sim$ is an equivalence relation on A.

A consequence of Theorem 0.5 and Proposition 0.6 is that equivalence relations and partitions are essentially the same thing. From an equivalence relation you can construct a partition and from a partition you can construct an equivalence relation. This sets up a correspondence between equivalence relations and partitions. You may need to think about this for a bit to understand what is meant here, or you can ask.

0.2 The ring of integers modulo n

We're going to use the notion of congruence modulo n and the ring of integers modulo n throughout this course. We recap this material on congruence modulo n below.

We start by giving the definition of congruence modulo n, which is a relation on $\mathbb{Z}$.

Definition 0.7. 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

$$a = b + nx.$$

The next lemma relates congruence modulo n to remainders when dividing by n.

Lemma 0.8. *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.*

We have the following corollary.

Corollary 0.9. *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 $a \equiv b \bmod n$.*

Next we give some elementary properties of congruences.

Lemma 0.10. *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)*

The properties of congruence modulo n given in Lemma 0.10 are precisely those required for an equivalence relation in Definition 0.2. We thus obtain the following corollary.

Corollary 0.11. *Let $\sim$ be the relation on $\mathbb{Z}$ defined by $a \sim b$ means $a \equiv b \bmod n$. Then $\sim$ is an equivalence relation.*

Next we recall a lemma about arithmetic properties of congruence modulo n.

Lemma 0.12. *Let $m, n \in \mathbb{N}$ and $a, b, a', b' \in \mathbb{Z}$. Suppose that $a \equiv b \bmod n$ and $a' \equiv b' \bmod n$. Then:*

(a) $a + a' \equiv b + b' \bmod n$;
(b) $aa' \equiv bb' \bmod n$; *and*
(c) $a^m \equiv b^m \bmod n$.

We're going to want the following important theorem about multiplicative inverses for congruences, which we recall next. This theorem was proved in Chapter 3 of 1AC Algebra 1, and the crucial ingredient for the proof was Bézout's lemma.

Theorem 0.13. *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*

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

We move on to recall the material on congruence classes starting with their definition.

Definition 0.14. 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.

The next proposition gives an alternative description of congruence classes.

Proposition 0.15. *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Then $[a]_n$ is the set of $x \in \mathbb{Z}$ such that a and x leave the same remainder when divided by n.*

Combining Corollary 0.11 and Theorem 0.5 we obtain the following corollary.

Corollary 0.16. *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*

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

Alternatively, we note that it is fairly easy to deduce the Corollary 0.16 directly from Proposition 0.15.

Next we recall the definition of the ring of integers modulo n, which we get by defining an addition and multiplication on the set of congruence classes modulo n.

Definition 0.17. 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 we have

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

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.

We next recap that addition and multiplication on $\mathbb{Z}_n$ are well defined. There is a potential ambiguity in the definition of the binary operation $+$ 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 0.17, 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 0.12, 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.

Last in this section we recall that the addition and multiplication on $\mathbb{Z}_n$ satisfy a number of arithmetic properties, which we list below. We note that these are all analogues of properties that are satisfied by $\mathbb{Z}$. As we'll recall in §1.2.2 the below list of properties tells us that $\mathbb{Z}_n$ is a *commutative ring with one*.

(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)

0.3 Polynomials

In 1AC Algebra 1 Chapter 4, we learnt about polynomials, and saw that they had many properties in common with the integers. We'll be seeing polynomials quite a lot in this course, so below is a recap of some of the definitions and results from 1AC Algebra 1. We'll just consider polynomials with coefficients in $\mathbb{F} = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ here. Later we'll work more generally with polynomials with coefficients in any ring.

We use the letter X to denote an indeterminate, this is just a formal symbol used in a polynomial. We use the capital letter X to help us to distinguish it from elements of $\mathbb{F}$; elements of $\mathbb{F}$ will be denoted by lower case letters.

We start with the definition of a polynomial over $\mathbb{F}$.

Definition 0.18. A *polynomial* over $\mathbb{F}$ in the indeterminate X is an expression of the form

$$f(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_i X^i + \ldots,$$

where $a_i \in \mathbb{F}$ for all $i \in \mathbb{N}_0$, and there exists $n \in \mathbb{N}_0$, such that $a_i = 0$ for all $i > n$.

We write $\mathbb{F}[X]$ for the set of all polynomials over $\mathbb{F}$.

From the Definition 0.18, we can write any element of $\mathbb{F}[X]$ as an expression of the form

$$f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0,$$

where $n \in \mathbb{N}_0$ and $a_0, a_1, \ldots, a_n \in \mathbb{F}$. When we write $f(X)$ in this form, we may not necessarily be assuming that $a_n \neq 0$, though this will often be the case. Also when we write $f(X)$ as above we are implicitly saying that $a_i = 0$ for $i > n$.

We proceed with a number of basic definitions about polynomials.

Definition 0.19. Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{F}[X]$ and $g(X) = b_mX^m + b_{m-1}X^{m-1} + \cdots + b_1X + b_0 \in \mathbb{F}[X]$. We say $f(X)$ and $g(X)$ are *equal* and write $f(X) = g(X)$ if $a_i = b_i$ for all $i \in \mathbb{N}_0$.

Definition 0.20. Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{F}[X]$. We say that $f(X)$ is the *zero polynomial* and write $f(X) = 0$ if $a_i = 0$ for all $i \in \mathbb{N}_0$.

Definition 0.21. Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{F}[X]$, with $a_n \neq 0$.

(a) The *degree* of $f(X)$, written $\deg f(X)$, is n.
(b) The *leading term* of $f(X)$ is a_nX^n.
(c) The *leading coefficient* is $f(X)$ is a_n.
(d) We say that $f(X)$ is *monic* if $a_n = 1$.

Polynomials of low degrees have special names:

- polynomials of degree 0 are called *constant polynomials*;
- polynomials of degree 1 are called *linear polynomials*;
- polynomials of degree 2 are called *quadratic polynomials*;
- polynomials of degree 3 are called *cubic polynomials*;
- polynomials of degree 4 are called *quartic polynomials*; and
- polynomials of degree 5 are called *quintic polynomials*.

Next we define addition and multiplication of polynomials.

Definition 0.22. Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in \mathbb{F}[X]$ and $g(X) = b_mX^m + b_{m-1}X^{m-1} + \cdots + b_1X + b_0 \in \mathbb{F}[X]$. We define the *sum* of $f(X)$ and $g(X)$ by

$$f(X) + g(X) = c_kX^k + c_{k-1}X^{k-1} + \cdots + c_1X + c_0,$$

where $k = \max\{n, m\}$ and $c_i = a_i + b_i$; and the *product of $f(X)$ and $g(X)$* by

$$f(X)g(X) = d_{n+m}X^{n+m} + d_{n+m-1}X^{n+m-1} + \cdots + d_1X + d_0,$$

where $d_i = a_0b_i + a_1b_{i-1} + \cdots + a_{i-1}b_1 + a_ib_0$.

In §1.2.5, we'll cover the fact that polynomials over $\mathbb{F}$ form a ring with the addition and multiplication defined in Definition 0.22; in fact we'll do this more generally for polynomials over any ring.

The next lemma tell us about how degree interacts with addition and multiplication.

Lemma 0.23. *Let $f(X), g(X) \in \mathbb{F}[X]$ with $f(X), g(X) \neq 0$. Then*

(a) $f(X) + g(X) = 0$ *or* $\deg(f(X) + g(X)) \leq \max\{\deg f(X), \deg g(X)\}$.
(b) $f(X)g(X) \neq 0$ *and* $\deg(f(X)g(X)) = \deg f(X) + \deg g(X)$.

Although it is straightforward to prove (b) in the lemma above, this turns out to be a very useful property that is used a lot when studying polynomials.

The next definition formalizes the notions of evaluating polynomials and roots of polynomials.

Definition 0.24. Let $f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \in \mathbb{F}[X]$ and let $c \in \mathbb{F}$.

(a) The *value of $f(X)$ at c* is

$$f(c) = a_n c^n + a_{n-1} c^{n-1} + \cdots + a_1 c + a_0 \in \mathbb{F}.$$

(b) We say that c is a *root of $f(X)$* if $f(c) = 0$.

It is also helpful for us to recall what it means for one polynomial to be a factor of another.

Definition 0.25. Let $f(X), g(X) \in \mathbb{F}[X]$. We say that $f(X)$ is a factor of $g(X)$ and write $f(X) \mid g(X)$ if there exists $s(X) \in \mathbb{F}[X]$ such that $g(X) = f(X)s(X)$.

We'll want to talk about long division of polynomials in the course, so we'll recall the division theorem for polynomials over $\mathbb{F}$ here.

Theorem 0.26 (The division theorem for polynomials)**.** *Let $f(X), g(X) \in \mathbb{F}[X]$ with $g(X) \neq 0$. Then there exist unique $q(X), r(X) \in \mathbb{F}[X]$ such that*

$$f(X) = q(X)g(X) + r(X), \ \textit{and } r(X) = 0 \textit{ or } \deg r(X) < \deg g(X).$$

The remainder theorem, which is proved using the division theorem is recalled next.

Corollary 0.27 (The remainder theorem)**.** *Let $f(X) \in \mathbb{F}[X]$ and let $c \in \mathbb{F}$. Then there exists unique $q(X) \in \mathbb{F}[X]$ such that*

$$f(X) = q(X)(X - c) + f(c).$$

Next we state a corollary of the remainder theorem, which is sometimes called the factor theorem.

Corollary 0.28 (The factor theorem)**.** *Let $f(X) \in \mathbb{F}[X]$ and let $c \in \mathbb{F}$. Then $X - c$ is a factor of $f(X)$ if and only if c is a root of $f(X)$.*

We note that later in §1.2.5, we'll give a more general version of the division theorem, and then observe that more general versions of the remainder theorem and the factor theorem hold too.

We'll also want the definition of an irreducible polynomial which we give next.

Definition 0.29. Let $f(X) \in \mathbb{F}[X]$ with $\deg f(X) > 0$.

(a) We say that $f(X)$ is *reducible* if there exist $g(X), h(X) \in \mathbb{F}[X]$ such that $\deg g(X), \deg h(X) > 0$ and $f(X) = g(X)h(X)$.

(b) We say that $f(X)$ is *irreducible* if it is not reducible.

To end this section we recall that we did quite a bit more about polynomials in 1AC Algebra 1, especially about factorization of polynomials, but we won't worry much about that for the time being. In fact this theory of factorization can be developed in a more general framework, which is an excellent example of the power of abstraction. The final chapter of these notes, which will be added later and is not part of 2AC covers this theory of factorization in rings.

0.4 Permutations

In this course, we will use the theory of permutations, which was introduced in 1AC Algebra 1. Below we recall the material that we require starting with the definition of a permutation.

Definition 0.30. 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.

Next we recall the two-row notation of a permutation.

Definition 0.31. Let $n \in \mathbb{N}$, let $\Omega = \{1, 2, \dots, n\}$ and $f \in \mathrm{Sym}(\Omega)$.
The *two-row notation* for f is the symbol

$$\begin{pmatrix} 1 & 2 & \dots & n \\ f(1) & f(2) & \dots & f(n) \end{pmatrix}.$$

The next two lemmas tell us about composition of permutations and inverses of permutations.

Lemma 0.32. *Let Ω be a set, and $f, g \in \mathrm{Sym}(\Omega)$. Then:*

(a) $g \circ f \in \mathrm{Sym}(\Omega)$*; and*
(b) $f^{-1} \in \mathrm{Sym}(\Omega)$.

We also learned about how to compose and invert permutations in two-row notation.
The next two definitions are about cycles and products of cycles.

Definition 0.33. 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 $a_1, a_2, \dots, a_k \in \Omega$ such that

$$f(a_1) = a_2,\ f(a_2) = a_3,\ \dots,\ f(a_{k-1}) = a_k,\ f(a_k) = a_1,$$

and $f(a) = a$ for all $a \in \Omega \setminus \{a_1, a_2, \dots, a_k\}$. We use the notation

$$f = (a_1\, a_2\, \dots\, a_k)$$

Often we say *k-cycle* instead of cycle of length k.

Definition 0.34. Let $\Omega = \{1, 2, \dots, n\}$, and let $c_1, c_2, \dots, c_r \in \mathrm{Sym}(\Omega)$ be cycles:

$$c_i = (a_{i,1}\, a_{i,2}\, \dots\, a_{i,k_r}).$$

(a) The *product of the cycles* $c_1, c_2, \dots, c_r$ is just their composition

$$c_1 \circ c_2 \circ \dots \circ c_r.$$

(b) We say that the cycles $c_1, c_2, \dots, c_r$ 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.

The following theorem is important in the theory of permutations. It allows us to define the cycle notation of a permutation in the subsequent definition, which turns out to be a really convenient notation for permutations.

Theorem 0.35. *Let $n \in \mathbb{N}$, $\Omega = \{1, 2, \dots, n\}$ and $f \in \mathrm{Sym}(\Omega)$. Then f can be written a product of disjoint cycles.*

Definition 0.36. Let $n \in \mathbb{N}$, $\Omega = \{1, 2, \dots, n\}$ and $f \in \mathrm{Sym}(\Omega$.

(a) The *cycle notation* of f is the decomposition of f as a product of disjoint cycles:

$$f = c_1 \circ c_2 \circ \cdots \circ c_r.$$

(b) The *cycle type* of f is the symbol

$$1^{m_1} 2^{m_2} \dots n^{m_n},$$

where m_i is the number of cycles of length i, and we do not write i^{m_i} if $m_i = 0$.

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. The order of the disjoint cycles can be changed because *disjoint cycles commute.* But beware that in general cycles do not commute.

We also learned about how to compose and invert permutations in cycle notation. Later in these notes, we'll recap how to do this composition and inversion in cycle notation, and we note that our notation will change slightly.

Chapter 1

Rings

In 1AC Algebra 1, we started the study of rings and looked at a number of examples of rings. At the start of this course, we are going to cover this material again in more detail and we will develop the theory of rings further.

Rings are abstract number systems where we can add and multiply elements, and a number of properties are satisfied. Our main example of a ring is the ring of integers $\mathbb{Z}$, and it may be helpful for you to keep this example in mind throughout the course. There are many other important examples of rings in mathematics, which we will encounter in this course.

The theory of rings is central to 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. Nowadays ring theory and applications of ring theory remain amongst the most important areas of mathematics research. Indeed much of my own research regards the structure and representation theory of certain rings.

1.1 Rings

We begin with the definition of a ring. This will hopefully won't look too abstract now, as you already saw it in 1AC Algebra 1. Remember that we can think of a ring as an algebraic structure where we have a set with an addition and a multiplication, which satisfy a load of axioms.

Before we give the definition of a ring we recall the definition of a binary operation. Informally a binary operation $*$ on a set A is a way of combining two elements of A.

Definition 1.1. 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 $*(a, b)$).

Below is the definition of a ring. A crucial part of the definition are the nine axioms, and their names of the axioms are given on the right.

Definition 1.2. A *ring* is a set R, along with

- a binary operation $+$ on R called addition; and
- a binary operation $\cdot$ on R called multiplication

satisfying the following axioms.

(A0) For all $x, y \in R$, $x + y \in R$
(closure under addition)

(A1) For all $x, y, z \in R$, $(x + y) + z = x + (y + z)$
(associative law of addition)

(A2) There is a special element $0 \in R$ such that for all $x \in R$, $x + 0 = x = 0 + x$
(existence of additive identity)

(A3) For all $x \in R$ there exists $-x \in R$ such that $x + (-x) = 0 = (-x) + x$
(existence of additive inverses)

(A4) For all $x, y \in R$, $x + y = y + x$
(commutative law of addition)

(M0) For all $x, y \in R$, $x \cdot y \in R$
(closure under multiplication)

(M1) For all $x, y, z \in R$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
(associative law of multiplication)

(Dl) For all $x, y, z \in R$, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$
(left distributive law)

(Dr) For all $x, y, z \in R$, $(x + y) \cdot z = (x \cdot z) + (y \cdot z)$
(right distributive law)

We'll be working with rings a lot during this course, so you should make an effort to remember these axioms. It is really important that you remember that the definition of a ring includes all of the axioms. Once we've been working with them for a while, they should stick in your head. Recalling the axioms of a ring is sometimes part of a question on the exam for this course.

As mentioned above the integers $\mathbb{Z}$ is our prototypical example of a ring. We'll see many more examples in Section 1.2, including the ring of integers modulo n and the ring of polynomials over a ring. It is worth keeping these examples in mind as we go through a number of definitions about rings in the rest of this section.

Before these definitions we have some remarks on the definition of a ring. Don't worry too much if these remarks seem a bit vague to start with. That shouldn't cause any problems at first, and hopefully once we're further in, they'll make more sense. As usual if you are a bit confused about anything, then you can ask me, or maybe discuss it with someone else from the lectures and try to figure out what is meant together.

- The special element $0 \in R$ from axiom (A2) is called the *zero of* R (or the *additive identity of* R). We'll see later in Lemma 1.25(a) that this element is unique with the property in this axiom.
- We often refer to nonzero elements of a ring R, by which we mean all elements of R except for 0.

- We'll see later in Lemma 1.25(c) that for $x \in R$ the element $-x$ in (A3) is unique, so that the notation used is not ambiguous. For $x, y \in R$ we usually write $x - y$ instead of $x + (-y)$.
- It is important to remember that the addition and multiplication are part of the definition of a ring. When we speak about a ring R, we implicitly understand that there is an addition and multiplication for R; and we are not just thinking of R as a set.
- Occasionally we may use a different notation for the addition and multiplication or zero in a ring, but when we do this we'll be careful to explain this. Also the notation for the zero in a ring may be different in some examples.
- Usually we will omit the $\cdot$ in the notation for multiplication in a ring R and just use juxtaposition, so we write xy instead of $x \cdot y$ for $x, y \in R$.
- The axioms (A0) and (M0) are not strictly necessary, as the definition of a binary operation ensures that they are automatically satisfied. It is useful to have them there to help us remember to check that addition and multiplication are really binary operations on R.
- Given the commutativity of addition given by (A4), we note that it is enough in (A2) to just have $x + 0 = x$, and in (A3) to just have $x + (-x) = 0$

The next definition is about rings in which there is a multiplicative identity.

Definition 1.3. Let R be a ring. We say that R is a *ring with one* if the following additional axiom holds.

(M2) There is a special element $1 \in R$ such that $1 \neq 0$ and for all $x \in R$, $x \cdot 1 = x = 1 \cdot x$ (existence of multiplicative identity).

The special element $1 \in R$ from axiom (M2) is called the *one of* R (or the *multiplicative identity of* R). We'll see later in Lemma 1.25(b) that this element is unique with the property in the axiom. We also note that in some examples of ring, we may use a different notation for the one of R.

Most of the rings that we'll be interested in later in the course have a one, but there are some examples, which we'll consider of rings without a one. In some books that you may look in the existence of a multiplicative identity is part of the axioms of a ring, and there isn't really much consistency across the literature. So it's worth being aware that in such books things will look a bit different.

Next we give the definition of a commutative ring.

Definition 1.4. Let R be a ring. We say that R is a *commutative ring* if the following additional axiom holds.

(M4) For all $x, y \in R$, $x \cdot y = y \cdot x$ (commutative law of multiplication)

We note that in a commutative ring we can replace the two distributivity axioms (Dl) and (Dr) with just (Dl). You should think about this to convince yourself why this is true.

Most of the examples of rings we consider in this course are commutative rings. A ring that is not a commutative ring is often called a *noncommutative ring*.

We also speak of a *commutative ring with one* to mean a ring with one, which satisfies the commutative law of multiplication.

In the remainder of this section, we consider zero divisors and units in rings, which lead to the definition of the integral domains and fields. When we look at examples of rings in Section 1.2, we'll consider these concepts, which will help us to understand the definitions.

First we give the definition of a zero divisor in a commutative ring.

Definition 1.5. Let R be a commutative ring, and let $a \in R$. We say that a is a *zero divisor* if $a \neq 0$ and there exists $b \in R$ such that $b \neq 0$ and $ab = 0$.

We have just given the definition of a zero divisor in a commutative ring, as we won't consider zero divisors in noncommutative rings in this course. For noncommutative rings the situation is more complicated, and we have to consider "left zero divisors" and "right zero divisors".

We proceed to the definition of an integral domain.

Definition 1.6. Let R be a commutative ring with one. We say that R is an *integral domain*, if there are no zero divisors in R.

Next we give the definition of units in a ring with one.

Definition 1.7. Let R be a ring with one.

(a) Let $a \in R$. We say that a is a *unit* if there exists $a^{-1} \in R$ such that $aa^{-1} = 1 = a^{-1}a$.
(b) We define $U(R) = \{a \in R : a \text{ is a unit}\}$ to be the *set of units in* R.

Another way of stating the definition of a unit is to say that $a \in R$ is a unit if it has a multiplicative inverse in R. Suppose that $a \in R$ is a unit. We'll see later in Lemma 1.25(d) that the multiplicative inverse is unique, and this justifies using the notation a^{-1}. Sometimes we write $\frac{1}{a}$ for a^{-1} and $\frac{a}{b}$ for ab^{-1}.

Next we define fields.

Definition 1.8. Let R be a commutative ring with one. Then we say that R is a *field* if the following additional axiom holds

(M3) for all $x \in R \setminus \{0\}$, there exists $x^{-1} \in R$ such that $x \cdot x^{-1} = 1 = x^{-1} \cdot x$
(existence of multiplicative inverses)

We note that an alternative way to state the axiom (M3) is to say that any nonzero element of R is a unit.

We usually write $\mathbb{F}$ rather than R for a field. We have that $U(\mathbb{F}) = \mathbb{F} \setminus \{0\}$ for a field and we use the notation $\mathbb{F}^\times$ for $\mathbb{F} \setminus \{0\}$ rather than $U(\mathbb{F})$.

You have seen the definition of a field in 1VGLA and in 2LALP, though it may look a bit different here because it is given in terms rings. Another way to phrase the definition of a field is to say that it is a set with an addition and multiplication that satisfies the axioms (A0–(A4), (M0)–(M4), and (Dl); we recall that (Dr) is not necessary if we have (M4). This is closer to what you have seen elsewhere.

We end this section by proving a lemma about zero divisors and units. Then use this lemma to deduce that a field is an integral domain in Corollary 1.10.

Lemma 1.9. *Let R be a commutative ring with one, and let $a \in R$ with $a \neq 0$. Suppose that a is a unit. Then a is not a zero divisor.*

Proof. Let $b \in R$ and suppose that $ab = 0$. Then we have $0 = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1b = b$. Hence, a is not a zero divisor as required. □

Corollary 1.10. *Let $\mathbb{F}$ be a field. Then $\mathbb{F}$ is an integral domain.*

Proof. Let $a \in \mathbb{F}^\times$. Then a is a unit and thus not a zero divisor by Lemma 1.9. Hence, $\mathbb{F}$ has no zero divisors and thus is an integral domain. □

1.2 Examples of rings

Now let's look lots of examples of rings, so we can see that they show up all over mathematics. We saw most of these examples already in 1AC Algebra 1. For some of the examples, we'll also consider whether they are integral domains and whether they are fields.

1.2.1 The ring of integers

The integers $\mathbb{Z}$ form a ring with the usual additional and multiplication. In fact $\mathbb{Z}$ is a commutative ring with one. The axioms for a ring are familiar properties about the integers.

As mentioned earlier, the ring of integers $\mathbb{Z}$ is our prototypical example a ring, and we'll frequently come back to consider this example. We learnt a lot about the ring of integers in 1AC Algebra 1.

Let $a, b \in \mathbb{Z}$ and suppose that $ab = 0$. Then we know that $a = 0$ or $b = 0$. Therefore, $\mathbb{Z}$ has no zero divisors and is an integral domain. (In fact the name integral domain comes from the idea that an integral domain is an algebraic structure similar to the integers.)

Now suppose that $ab = 1$. We know that this implies $a = b = \pm 1$. Therefore, we see that the only units in $\mathbb{Z}$ are ± 1, so $U(\mathbb{Z}) = \{\pm 1\}$. Thus $\mathbb{Z}$ is not a field.

1.2.2 The ring of integers modulo n

Let $n \in \mathbb{N}$. In Section 0.2 we recalled the definition of the ring of integers modulo n, which is denoted $\mathbb{Z}_n$. At the end of that section we listed a number of properties satisfied by $\mathbb{Z}_n$, and we can now observe that these are precisely the properties we require in order to say that $\mathbb{Z}_n$ is a commutative ring with one. We state this formally below.

Theorem 1.11. *Let $n \in \mathbb{N}$. Then $\mathbb{Z}_n$ is a commutative ring with 1.*

We move on to consider which elements of $\mathbb{Z}_n$ are zero divisors and which are units. For this we assume that $n > 1$, so that $[0]_n \neq [1]_n$.

We can show that $[a]_n$ is a zero divisor in $\mathbb{Z}_n$ if and only if a is not coprime to n. Also we can show that $[a]_n$ is a unit in $\mathbb{Z}_n$ if and only of a is coprime to n. The proofs of the assertions are left as exercises. We note that using Lemma 1.9 it suffices to just show that $[a]_n$ is a zero divisor in $\mathbb{Z}_n$ if a is not coprime to n and that $[a]_n$ is a unit in $\mathbb{Z}_n$ if and only of a is coprime to n. You'll want to use Theorem 0.13 for the statement about units.

Thus we see that if n is prime, then any nonzero element of $\mathbb{Z}_n$ is a unit. Moreover, if n is not prime, then $\mathbb{Z}_n$ contains zero divisors. We summarize this in the following proposition.

Proposition 1.12. *Let $n \in \mathbb{N}$ with $n > 1$. Then $\mathbb{Z}_n$ is an integral domain if and only if n is prime, and in this case $\mathbb{Z}_n$ is in fact a field.*

In case $n = p$ is prime, the above proposition tells that $\mathbb{Z}_p$ is a field. Consequently, we often use the notation $\mathbb{F}_p$ instead of $\mathbb{Z}_p$ to emphasize that it is a field.

We remark that when we encounter the ring of integers modulo n again in future, we sometimes use the shorter notation a instead of $[a]_n$ for elements of $\mathbb{Z}_n$, for example we will write 3 instead of $[3]_n$. This is potentially really confusing, but in practice it doesn't tend to cause any problems.

1.2.3 The fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$

Let $\mathbb{F}$ be one of $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Then $\mathbb{F}$ with the usual addition and multiplication is a field. The axioms for a field are familiar properties about $\mathbb{F}$.

1.2.4 The zero ring

This example may seem a bit strange at first, but stick with it, and it should make sense.

We're going to consider $R = \{0\}$ to be a set with one element. Then we want to define an addition and multiplication on R. Well let's think about what we have to do and what we can do.

- To define addition we only have to say what $0 + 0$ is, and we only have one choice for this, namely 0. So we define addition by saying $0 + 0 = 0$.
- Similarly, we have to define multiplication by saying $0 \cdot 0 = 0$.

Now to verify that R really is a ring, we have to check the axioms. Now we have to think for a bit before deciding that all the axioms hold trivially.

To do this let's first consider say the axiom (M1), we have to show

$$\text{for all } x, y, z \in R, x \cdot (y \cdot z) = (x \cdot y) \cdot z.$$

Well the only possibility for x, y and z is that they are all 0, and in this case both $x \cdot (y \cdot z)$ and $(x \cdot y) \cdot z$ are also equal to 0. Therefore, this axiom holds trivially.

Now we can apply the same logic to all the other axioms from Definition 1.2 to deduce that R is indeed a ring.

We leave it as an (easy) exercise to determine whether R is commutative. We note that in (M2) we insist that $1 \neq 0$, so we see that R does not have a one.

1.2.5 Polynomial rings

In Section 0.3, we recalled some material about polynomials over $\mathbb{F} = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Here we are going to see that much of this can applied to polynomials with coefficients in any ring.

Let R be a ring and let X be an indeterminate. The set of polynomials over R denoted $R[X]$ is defined in exactly the same way as in Definition 0.18 except that we allow coefficients from R rather than $\mathbb{F}$. Similarly, addition and multiplication of polynomials in $R[X]$ can be defined using the formulas in Definition 0.22. We note here that Definitions 0.19, 0.20, 0.21, 0.24, 0.25 and 0.29 have versions for $R[X]$, and we will use these without saying any more about it.

In the theorem below we state that $R[X]$ is a ring.

Theorem 1.13. *Let R be a ring. Then $R[X]$ is a ring. Moreover,*

(a) *if R has a one, then $R[X]$ has a one; and*
(b) *if R is commutative, then $R[X]$ is commutative.*

To prove this proposition we need to check all the axioms from Definition 1.2, and to check (M2) holds for (a) and that (M4) holds for (b). The zero of $R[X]$ is the zero polynomial, and we note that checking (A0), (A2) and (M0) is trivial. Also if R has a one, then the one of $R[X]$ is the constant polynomial with constant term 1, and checking (M2) is straightforward.

Checking all of the other axioms is a rather laborious task, so we will only check a couple of the axioms here and we'll do another one in the lectures. The other axioms can be verified in a similar manner and you should check some of the other axioms as exercises.

First we'll deal with (A3).

Proof of (A3). Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in R[X]$.
Define $-f(X) = -a_nX^n - a_{n-1}X^{n-1} - \cdots - a_1X - a_0 \in R[X]$.
We show that $f(X) + (-f(X)) = 0$, where 0 is the zero polynomial, and note that it follows from our proof that also $-f(X)+f(X) = 0$, because addition in R is commutative. We have

$$\begin{aligned} f(X) + (-f(X)) &= a_nX^n + \cdots + a_0 + (-a_nX^n - \cdots - a_0) \\ &= (a_n - a_n)X^n + (a_{n-1} - a_{n-1})X^{n-1} + \cdots + (a_1 - a_1)X + (a_0 - a_0) \\ &= 0X^n + 0X^{n-1} + \cdots + 0X + 0, \end{aligned}$$

which is the zero polynomial. Hence, $f(X) + (-f(X)) = 0$ as required. □

Next we're going to check that axiom (Dl) holds. For this, and also for some of the other axioms, it is convenient to write polynomials in summation notation so we will write a polynomial $f(X) \in R[X]$ in the form $f(X) = \sum_{i=0}^{\infty} a_iX^i$, where $a_i \in R$ for each $i \in \mathbb{N}_0$ and there exists $n \in \mathbb{N}_0$ such that $a_i = 0$ for all $i > n$. It may be a bit tricky to get your head around this proof at first with this summation notation.

Proof of (Dl). Let $f(X) = \sum_{i=0}^{\infty} a_i X^i$, $g(X) = \sum_{i=0}^{\infty} b_i X^i$ and $h(X) = \sum_{i=0}^{\infty} c_i X^i$ be polynomials in $R[X]$. We aim to show that $f(X)(g(X)+h(X)) = f(X)g(X)+f(X)h(X)$. Using the definitions of multiplication and addition we have

$$\begin{aligned} f(X)(g(X)+h(X)) &= \left(\sum_{i=0}^{\infty} a_i X^i\right)\left(\sum_{i=0}^{\infty} b_i X^i + \sum_{i=0}^{\infty} c_i X^i\right) \\ &= \left(\sum_{i=0}^{\infty} a_i X^i\right)\left(\sum_{i=0}^{\infty} (b_i + c_i) X^i\right) \\ &= \sum_{i=0}^{\infty} d_i X^i, \end{aligned}$$

where $d_i = \sum_{j=0}^{i} a_j(b_{i-j} + c_{i-j})$. Also we have

$$\begin{aligned} f(X)g(X)+f(X)h(X) &= \left(\sum_{i=0}^{\infty} a_i X^i\right)\left(\sum_{i=0}^{\infty} b_i X^i\right) + \left(\sum_{i=0}^{\infty} a_i X^i\right)\left(\sum_{i=0}^{\infty} c_i X^i\right) \\ &= \sum_{i=0}^{\infty} r_i X^i + \sum_{i=0}^{\infty} s_i X^i \\ &= \sum_{i=0}^{\infty} (r_i + s_i) X^i, \end{aligned}$$

where $r_i = \sum_{j=0}^{i} a_j b_{i-j}$ and $r_i = \sum_{j=0}^{i} a_j c_{i-j}$. Since $d_i = r_i + s_i$ for each i, we deduce that $f(X)(g(X)+h(X)) = f(X)g(X)+f(X)h(X)$ as required. □

We move on to focus on the case where R is an integral domain. In this case, we can prove the following lemma in exactly the same way as Lemma 0.23(b).

Lemma 1.14. *Let R be an integral domain and let $f(X), g(X) \in \mathbb{F}[X]$ with $f(X), g(X) \neq 0$. Then $f(X)g(X) \neq 0$ and $\deg(f(X)g(X)) = \deg f(X) + \deg g(X)$.*

From this we immediately deduce the following corollary.

Corollary 1.15. *Let R be an integral domain. Then $R[X]$ is an integral domain.*

Let R be an integral domain, and suppose $f(X), g(X) \in R[X]$ with $f(X)g(X) = 1$. Then by Lemma 1.14, we have $\deg f(X) + \deg g(X) = 0$. For this to be true, we must have $\deg f(X) = 0 = \deg g(X)$. Therefore, $f(X) = a$ and $g(X) = b$ are constant polynomials, for some $a, b \in R$. Moreover, we have $ab = 1$, so that a is a unit in R and $b = a^{-1}$. Conversely, if $f(X) = a$ where $a \in U(R)$, then clearly the constant polynomial $f(X) = a$ is a unit in $R[X]$. So the set of units in $R[X]$ is the set of constant polynomials with constant term a unit of R, i.e. $U(R[X]) = \{f(X) \in R[X] : f(X) = a \text{ for some } a \in U(R)\}$. Therefore, $R[X]$ is not a field.

As a special case, we note that if $R = \mathbb{F}$ is a field, then $U(\mathbb{F}[X])$ is the set of nonzero constant polynomials.

Next we give a version of the division theorem, which holds for polynomials over a commutative ring with one. This will be useful for us later in the course. We do not include the proof as it can be proved in essentially the same way as Theorem 0.26.

Theorem 1.16 (The division theorem for polynomials). *Let R be a commutative ring with one and let $f(X), g(X) \in R[X]$. Suppose that $g(X) \neq 0$ and the leading coefficient of $g(X)$ is a unit in R. Then there exist unique $q(X), r(X) \in R[X]$ such that*

$$f(X) = q(X)g(X) + r(X), \text{ and } r(X) = 0 \text{ or } \deg r(X) < \deg g(X).$$

With this division theorem in hand, we can deduce the remainder theorem, and factor theorem for polynomials over integral domains, which we state below. These can be proved in exactly the same way as the versions in Section 0.3 for polynomials over $\mathbb{F}$, so we omit the proofs.

Corollary 1.17 (The remainder theorem). *Let R be a commutative ring with one, let $f(X) \in R[X]$ and let $c \in R$. Then there exists unique $q(X) \in R[X]$ such that*

$$f(X) = q(X)(X - c) + f(c).$$

Corollary 1.18 (The factor theorem). *Let R be a commutative ring with one, let $f(X) \in R[X]$ and let $c \in R$. Then $X - c$ is a factor of $f(X)$ if and only if c is a root of $f(X)$.*

1.2.6 The Gaussian integers and extensions of $\mathbb{Z}$ and $\mathbb{Q}$ in $\mathbb{C}$

The title above may look a bit daunting at first, but don't worry it is not as bad as you may think, and we'll only look at some specific examples.

First we consider the *Gaussian integers*, which are defined to be

$$\mathbb{Z}[i] = \{a + bi \in \mathbb{C} : a, b \in \mathbb{Z}\} \subseteq \mathbb{C},$$

where as usual $i = \sqrt{-1}$ satisfies $i^2 + 1 = 0$. We have addition and multiplication of $\mathbb{Z}[i]$ given by considering $\mathbb{Z}[i]$ as a subset of $\mathbb{C}$. The following proposition says that $\mathbb{Z}[i]$ is a commutative ring with 1.

Proposition 1.19. *$\mathbb{Z}[i]$ is a commutative ring with 1.*

Proof. Note that $0 = 0 + 0i \in \mathbb{Z}[i]$ and $1 = 1 + 0i \in \mathbb{Z}[i]$, these are the zero and one in $\mathbb{Z}[i]$. Consequently we see that the axioms (A2) and (M2) hold for $\mathbb{Z}[i]$.

To check that $\mathbb{Z}[i]$ is a commutative ring with 1 we now just need to check that (A0), (A3) and (M0) hold, as the remaining axioms hold in $\mathbb{C}$ and therefore also in $\mathbb{Z}[i] \subseteq \mathbb{C}$. In order to verify (A0), (A3) and (M0), let $a + bi, c + di \in \mathbb{Z}[i]$.
Then $(a + bi) + (c + di) = (a + c) + (b + d)i \in \mathbb{Z}[i]$, so (A0) holds.
We have $(a + bi) + (-a - bi) = 0$, and $-a - bi \in \mathbb{Z}[i]$, so (A3) holds,
Also $(a + bi)(c + di) = ac + (ad + bc)i + bdi^2 = (ac - bd) + (ad + bc)i \in \mathbb{Z}[i]$, so (M0) holds. □

The proof above that $\mathbb{Z}[i]$ is a ring used the fact that the axioms hold in a bigger ring containing $\mathbb{Z}[i]$, namely $\mathbb{C}$, and then noting that we just have to check (A0), (A2), (A3) and (M0). In Section 1.4, we define subrings, and will see that the proof fits into a bigger picture of showing that the certain subsets of rings are subrings.

Since $\mathbb{Z}[i]$ in a subset of $\mathbb{C}$, we see that it has no zero divisors. Therefore, $\mathbb{Z}[i]$ is an integral domain.

Next we work out the units in $\mathbb{Z}[i]$. To do this let us recall that the modulus of a complex number $a+bi$ is defined to be $|a+bi| = \sqrt{a^2+b^2}$. We also recall that for $z, w \in \mathbb{C}$ we have $|zw| = |z||w|$. Now let $z = a+bi \in \mathbb{Z}[i]$. Then we clearly have $|z|^2 \in \mathbb{Z}$. Note that we also have $|1| = 1$.

Suppose that z is a unit in $\mathbb{Z}[i]$ and let $w \in \mathbb{Z}[i]$ with $zw = 1$. Then $|zw| = 1$, so $|z||w| = 1$ and $|z|^2|w|^2 = 1$. Now since, $|z|^2, |w|^2 \in \mathbb{Z}$ and $|z|^2|w|^2 = 1$, we must have $|z|^2 = 1$. Therefore, we have $a^2+b^2 = 1$. We note that this is only possible if $a = \pm 1$ and $b = 0$, or $a = 0$ and $b = \pm 1$. Therefore, we obtain $U(\mathbb{Z}[i]) = \{\pm 1, \pm i\}$. From this we deduce that $\mathbb{Z}[i]$ is not a field.

We can get similar rings by taking any monic irreducible polynomial $f(X) \in \mathbb{Z}[X]$, and then *adjoining* a root of this polynomial to $\mathbb{Z}$. For example, if we took $f(X) = X^2 - 2$, then we would consider $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \in \mathbb{C} : a, b \in \mathbb{Z}\}$. In general let $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0 \in \mathbb{Z}[X]$ be monic and irreducible, and let $\alpha \in \mathbb{C}$ be a root of $f(X)$. Then we can consider

$$\mathbb{Z}[\alpha] = \{b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1} : b_0, b_1, \ldots, b_{n-1} \in \mathbb{Z}\} \subseteq \mathbb{C},$$

and we can show that $\mathbb{Z}[\alpha]$ is a ring. To multiply elements in $\mathbb{Z}[\alpha]$ we multiply out brackets and then use the fact that $f(\alpha) = 0$ to make substitutions to give an element of the form $b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$. We do not go into the details here, as this is better understood through examples.

Such rings are very important in number theory. Their theory can be built up in a similar way to how we built up the theory of the integers in 1AC Algebra 1. These sorts of rings will show up quite a bit in the rest of this course. In fact, in the last chapter of these notes, which is not part of 2AC and will be added later, you can read that an analogue of the fundamental theorem of arithmetic holds for both $\mathbb{Z}[i]$ and for $\mathbb{Z}[\sqrt{2}]$.

Similarly given an monic irreducible polynomial $f(X) \in \mathbb{Q}[X]$, we can form a ring $\mathbb{Q}[\alpha]$ by *adjoining* a root $\alpha \in \mathbb{C}$ of $f(X)$ to $\mathbb{Q}$. The ring $\mathbb{Q}[\alpha]$ can be described very similarly to above, so we omit the details here. In fact it turns out that $\mathbb{Q}[\alpha]$ is a field, and we leave the details of this as an exercise.

1.2.7 Extensions of commutative rings with one

We move on to look at some more abstract examples of rings, which are similar (but more abstract) to those in §1.2.6.

For a first example, we define $\mathbb{Z}[\beta] = \{a + b\beta : a, b \in \mathbb{Z}\}$, where β is a symbol that satisfies $\beta^2 = 1$. (Note that here β is just a formal symbol here, so β is not an element of $\mathbb{C}$.)

Addition on $\mathbb{Z}[\beta]$ is defined by

$$(a + b\beta) + (c + d\beta) = (a + c) + (b + d)\beta \in \mathbb{Z}[\beta]$$

and multiplication on $\mathbb{Z}[\beta]$ is defined by

$$(a + b\beta)(c + d\beta) = (ac + bd) + (ad + bc)\beta,$$

where $a, b, c, d \in \mathbb{Z}$; it is defined, so that we can calculate by multiplying out brackets and using the relation $\beta^2 = 1$, so

$$\begin{aligned}(a + b\beta)(c + d\beta) &= ac + (ad + bc)\beta + bd\beta^2 \\ &= (ac + bd) + (ad + bc)\beta.\end{aligned}$$

Then $\mathbb{Z}[\beta]$ is a commutative ring with one, where the zero is $0 = 0 + 0\beta \in \mathbb{Z}[\beta]$ and the one is $1 = 1 + 0\beta \in \mathbb{Z}[\beta]$.
We note that we have not checked the axioms for a ring here, but that we should really do this: it is a bit of work to do this and it ultimately follows from the fact that the axioms hold for $\mathbb{Z}$.

We can calculate that $(\beta - 1)(\beta + 1) = 0$, so that $\beta - 1$ and $\beta + 1$ are zero divisors in $\mathbb{Z}[\beta]$. Therefore, $\mathbb{Z}[\beta]$ is not an integral domain.

Now let R be any commutative ring with one, and let $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in R[X]$ be a monic polynomial, and let β be a symbol that satisfies of $f(\beta) = 0$. Then we define

$$R[\beta] = \{b_0 + b_1\beta + \cdots + b_{n-1}\beta^{n-1} : b_0, b_1, \ldots, b_{n-1} \in R\},$$

where addition is defined in the natural way and multiplication is defined by multiplying out brackets and then using $f(\beta) = 0$ to make substitutions to give an element of the form $b_0+b_1\beta+\cdots+b_{n-1}\beta^{n-1}$. We do not give the details here, as this is better understood through examples. We can prove that $R[\beta]$ is a commutative ring with one, though we omit the proof here.

A specific example is $\mathbb{Z}_2[\omega]$, where ω is a symbol that satisfies $\omega^2 + \omega + 1 = 0$. Note here we are writing 0 for $[0]_2$ and 1 for $[1]_2$. Then we can calculate the multiplication table of $\mathbb{Z}_2[\omega]$ to be

$\cdot$	0	1	ω	$1+\omega$
0	0	0	0	0
1	0	1	ω	$1+\omega$
ω	0	ω	$1+\omega$	1
$1+\omega$	0	$1+\omega$	1	ω

From this multiplication table, we observe that every nonzero element of $\mathbb{Z}_2[\omega]$ is a unit. Thus $\mathbb{Z}_2[\omega]$ is a field.

1.2.8 Matrix rings

Let R be a ring and let $M_2(R)$ denote the set of 2×2 matrices with entries in R. So

$$M_2(R) = \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} : a_{11}, a_{12}, a_{21}, a_{22} \in R \right\}.$$

In 1VGLA you learnt about matrix addition and multiplication, which are defined by

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix}$$

and

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \end{pmatrix}.$$

(Actually in 1VGLA you only saw these definitions for matrices with entries in $\mathbb{R}$ or $\mathbb{C}$, but they make sense over any ring R.)

With these definitions of addition and multiplication $M_2(R)$ is a ring. We do not include a proof of this assertion, as usual it necessary to check the axioms. You have seen that some of the axioms hold in 1VGLA, for example associativity of multiplication, and all of the other axioms can be checked similarly. The zero 2×2 matrix in $M_2(R)$ is

$$0_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and this is the zero in $M_2(R)$. Also if R has a one, then $M_2(R)$ has a one which is the identity 2×2 matrix

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in M_2(R).$$

We note that $M_2(R)$ is a noncommutative ring, as long as R is not the zero ring. We leave showing this as an exercise, which you can do by finding $A, B \in M_2(R)$ such that $AB \neq BA$.

Similarly, for any $n \in \mathbb{N}$, we can consider the set $M_n(R)$ of $n \times n$ matrices with entries in $\mathbb{F}$ with matrix addition and multiplication. Then $M_n(R))$ is a ring as stated in the lemma below.

Lemma 1.20. *Let R be a ring and let $n \in \mathbb{N}$. Then $M_2(R)$ is a ring.*

1.2.9 Direct products of rings

Given rings R and S we can form another ring by taking their direct product as defined next.

Definition 1.21. Let R and S be rings. We define addition and multiplication on

$$R \times S = \{(r, s) : r \in R, s \in S\}$$

by

$$(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$$

and

$$(r_1, s_1)(r_2, s_2) = (r_1 r_2, s_1 s_2).$$

Then $R \times S$ is called the *direct product* of R and S.

We should verify that $R \times S$ is actually a ring, which is stated in the next lemma.

Lemma 1.22. *Let R and S be rings. Then $R \times S$ is a ring.*

In order to prove this lemma, we have to check that the axioms hold. We do not include the details here: it is a consequence of the axioms holding in R and S, and is a straightforward exercise. We note that the zero of $R \times S$ is $(0, 0)$, where the first entry is the zero of R and the second entry is the zero of S. Moreover, we note that if R and S are rings with one, then $R \times S$ is a ring with one, where the one of $R \times S$ is $(1, 1)$, where the first entry is the one of R and the second entry is the one of S.

More generally, if $R_1, R_2, \dots, R_m$ are rings then we can recursively define the direct product $R_1 \times R_2 \times \cdots \times R_m$, and this is a ring.

1.2.10 The ring of even integers

Let $2\mathbb{Z}$ denote the set of all even integers with the usual addition and multiplication. In Section 1.4 on subrings later we will see that $2\mathbb{Z}$ is a ring. We can show that $2\mathbb{Z}$ does not have a one, which is an immediate consequence of the following easy lemma – you should make sure that you understand why.

Lemma 1.23. *There is no even integer $e \in 2\mathbb{Z}$ such that $ex = x$ for all $x \in 2\mathbb{Z}$.*

Proof. Suppose that $e \in \mathbb{Z}$ satisfies the property that $ex = x$ for all $x \in 2\mathbb{Z}$. Then, in particular, this is satisfied for $x = 2$, so we obtain $2e = 2$. But this implies that $e = 1$, which is not an element of $2\mathbb{Z}$. Hence, there is no element of $2\mathbb{Z}$ satisfying the property. □

1.2.11 An example of not a ring

Often in understanding a concept in mathematics, it is useful to have some examples that are not that concept. This helps us to grasp the essence of the definition. In this spirit we give an example of a set with an addition and multiplication below, which is not a ring. There are many other examples of "non-rings", and you'll see some more in this course. From these examples, you should be convinced that a ring really isn't just a set with an addition and multiplication – you also need to prove that the set is closed under these operations and that the axioms are satisfied.

Let $R = \mathbb{R}^3$ be the set of column vectors of length 3 with entries in $\mathbb{R}$. That is

$$R = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \ : \ a_1, a_2, a_3 \in \mathbb{R} \right\}.$$

There is (vector) addition on R and we can define multiplication on R to be the cross product of vectors, which is denoted by $\times$, and defined by

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}$$

Of course, there are other ways of writing the cross product, which you are familiar with and may be easier to work with. We could check that axioms (A0)–(A4) and also (M0), (Dl) and (Dr) all hold for R. However, we can check that (M1) is not true for R. To do this we need a counterexample to that axiom.

Counterexample to (M1). We need to find vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ such that $\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) \neq (\mathbf{x} \times \mathbf{y}) \times \mathbf{z}$. Let us consider

$$\mathbf{x} = \mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{y} = \mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{z} = \mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Then we know that $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = -\mathbf{j}$, whilst $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0}$, where $\mathbf{0}$ is the zero vector. So this is indeed a counterexample to (M1).

1.2.12 The quaternions

The quaternions are obtained from the real numbers by adjoining three different square roots of -1 and specifying the multiplication between them. This is similar to how we define the complex numbers from the real numbers, as we see in the definition below. It may look a bit abstract at first, but once you've worked with them a bit, you'll get used to them. Despite this abstract nature, you may be interested to know that their origin is in the work of W. R. Hamilton on mechanics, and they now are used in 3 dimensional computer graphics.

We define the quaternions to be the set

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\},$$

where i, j and k are symbols that satisfy

$$i^2 = j^2 = k^2 = ijk = -1. \tag{1.1}$$

(Note that here i, j and k are just formal symbols that we can work with as explained below.)

Addition on $\mathbb{H}$ is defined by

$$(a + bi + cj + dk) + (a' + b'i + c'j + d'k) = (a + a') + (b + b')i + (c + c')j + (d + d')k.$$

To define multiplication we note that from (1.1) we can deduce that

$$ij = k, jk = i, ki = j \text{ and } ji = -k, kj, = -i, ik = -j.$$

Now multiplication can be defined by multiplying out brackets to obtain

$$\begin{aligned}(a + bi + cj + dk)(a' + b'i + c'j + d'k) = (aa' - bb' - cc' - dd') +& \\ (ab' + ba' + cd' - dc')i + (ac' + ca' + db' - bd')j + (ad' + da' + bc' - cb')k.&\end{aligned}$$

Then we can show that $\mathbb{H}$ is a ring with one, but we omit checking the axioms here, which would be quite a lot of work. We note that $\mathbb{H}$ is not commutative, and leave it as an exercise to show this.

1.3 Elementary arithmetic properties of rings

A lot of familiar properties about addition and multiplication of integers can be proved to be true for all rings by deducing them from the axioms of the rings. In this section we'll see several such properties. We'll prove some of them and some others will be given as exercises. Later on we'll frequently use this properties safe in the knowledge that they have been proved, and it's quite likely that we'll use them without referring back to them. Actually, we already used some of these properties in the proof of Lemma 1.9, and maybe we didn't even notice that we used them – but it is important that we do prove these properties, so that we know that we are on a firm footing going forward.

In 1VGLA and 2LALP, you have already seen similar statements proved about vector spaces. We'll go through them all in these notes, as this may help you to understand the ideas better, though we'll go through them quite quickly in the lectures.

First, we consider what happens when we add many elements of a ring together. We can add two elements of a ring together, but when we add more elements together there are different orders to add them depending on how they are bracketed.

For example, when we have three elements a, b and c in a ring R, which we want to add together, then they can be bracketed both as $(a+b)+c$ or as $a+(b+c)$; in this case the associative law of addition ensures that these are equal, so that we can denote both by $a+b+c$.

The next lemma says that the sum of any number of elements does not depend on the order in which they are added, and also the similar statement holds for multiplication. This may seem like a fairly obvious statement, but we have to prove it. The idea of the proof is quite simple, and just involves applying the associative law of addition many times.

To illustrate the proof let's consider a case of adding 4 elements a, b, c and d in a ring R together with different bracketing. Two possible ways to bracket this sum are as $a+((b+c)+d)$ or as $(a+b)+(c+d)$. Now we use the associative law a couple times to see that these two sums give the same result.

$$\begin{aligned} a+((b+c)+d) &= a+(b+(c+d)) \\ &= (a+b)+(c+d). \end{aligned}$$

Below is the general statement for the sum or product of any number of elements.

Lemma 1.24. *Let R be a ring, and let $a_1, a_2, \dots a_n \in R$. Then*

(a) *the sum $a_1+a_2+\cdots+a_n$ is independent on the bracketing we use to calculate it; and*

(b) *the product $a_1 a_2 \cdots a_n$ is independent on the bracketing we use to calculate it.*

Proof. (a) We work by induction on n. The cases where $n=1$ or $n=2$ being trivial, and the case where $n=3$ reduces to the associative law of addition.

So let's assume by induction that the statement holds for sums of fewer than n terms. So by this inductive hypothesis we can write the sum of m elements $b_1, b_2, \dots, b_m \in R$ as $b_1+b_2+\cdots+b_m$ whenever $m<n$.

Now consider a bracketing of the sum of $a_1, a_2, \dots, a_n$. In the last stage of calculating this sum, we will be adding together a sum of $a_1, \dots, a_k$ to a sum of $a_{k+1}, \dots, a_n$ where

$0 < k < n$. By the inductive hypothesis we can write this unambiguously as $(a_1 + \cdots + a_k) + (a_{k+1} + \cdots + a_n)$, because the sum of the first k terms does not depend on the bracketing and similarly for the last $n - k$ terms. Given another bracketing of the sum of $a_1, a_2, \ldots, a_n$, we can write this as $(a_1 + \cdots + a_l) + (a_{l+1} + \cdots + a_n)$ for some $0 < l < n$.

If $k = l$, then these expressions are equal, and we are done. So suppose that $k \neq l$. Then without loss of generality we may assume that $k < l$. Then using the inductive hypothesis, we can write

$$(a_1 + \cdots + a_k) + (a_{k+1} + \cdots + a_n) = (a_1 + \cdots + a_k) + ((a_{k+1} + \cdots + a_l) + (a_{l+1} + \cdots + a_n))$$

and

$$(a_1 + \cdots + a_l) + (a_{l+1} + \cdots + a_n) = ((a_1 + \cdots + a_k) + (a_{k+1} + \cdots + a_l)) + (a_{l+1} + \cdots + a_n).$$

Now these two expressions on the right hand side are equal by the associative law, the first is of the form $x + (y + z)$ and the second $(x + y) + z$, where $x = a_1 + \cdots + a_k$, $y = a_{k+1} + \cdots + a_l$ and $z = a_{l+1} + \cdots + a_n$. This completes the induction step.

(b) This can be proved using a very similar argument. □

A consequence of the lemma is that we can unambiguously write $a_1 + a_2 + \cdots + a_n$ for the sum of these elements, and similarly for $a_1 a_2 \cdots a_n$.

In the next lemma we consider uniqueness of various elements in a ring.

Lemma 1.25. *Let R be a ring.*

(a) *Let $z_1, z_2 \in R$. Suppose that for all $x \in R$, we have*

$$x + z_1 = x = z_1 + x \quad \textit{and} \quad x + z_2 = x = z_2 + x.$$

Then $z_1 = z_2$.
In other words the zero of R is unique.

(b) *Let $e_1, e_2 \in R$. Suppose that for all $x \in R$, we have*

$$e_1 x = x = x e_1 \quad \textit{and} \quad e_2 x = x = x e_2.$$

Then $e_1 = e_2$.
In other words if R is a ring with one, then the one of R is unique.

(c) *Let $a, b_1, b_2 \in R$. Suppose that*

$$a + b_1 = 0 = b_1 + a \quad \textit{and} \quad a + b_2 = 0 = b_2 + a.$$

Then $b_1 = b_2$.
In other words the additive inverse of a in R is unique.

(d) *Let $a, b_1, b_2 \in R$. Suppose that R has a one and*

$$a b_1 = 1 = b_1 a \quad \textit{and} \quad a b_2 = 1 = b_2 a.$$

Then $b_1 = b_2$.
In other words if a is a unit, then the multiplicative inverse of a in R is unique.

Proof. (a) By the assumptions we have

$$z_1 = z_1 + z_2 = z_2.$$

Here we are first applying the assumption on z_2 in the case $x = z_1$, and then the assumption on z_1 in the case $x = z_2$.

(b) The proof is similar to (a), so we leave this as an exercise.

(c) We have

$$(b_1 + a) + b_2 = 0 + b_2 = b_2$$

by the assumption on b_1 and by (A2). Similarly, we also we have

$$b_1 + (a + b_2) = b_1.$$

Therefore, using (A1) we obtain $b_1 = b_2$.

(d) The proof is similar to (c), so we leave it as an exercise. □

As mentioned earlier, the uniqueness of the additive inverse of a ensured by Lemma 1.25(c) justifies the notation $-a$; similarly, the uniqueness of the multiplicative inverse of a unit $a \in R$ ensured by Lemma 1.25(d) justifies the notation a^{-1}.

Next we move on to a particularly useful properties in rings called the cancellation laws. Note that the assumption that R is an integral domain is necessary for the multiplicative cancellation law.

Lemma 1.26. *Let R be a ring and let $a, b, c \in R$.*

(a) *Suppose that $a + b = a + c$. Then $b = c$.*
(b) *Suppose that R be an integral domain, $a \neq 0$ and $ac = ab$. Then $b = c$.*

Proof. To prove (a) we add $-a$ to the left of both sides of $a + b = a + c$. We obtain

$$\begin{aligned} -a + (a + b) &= -a + (a + c) \\ (-a + a) + b &= (-a + a) + c \quad \text{by (A1)} \\ 0 + b &= 0 + c \quad \text{by (A3)} \\ b &= c \quad \text{by (A2)} \end{aligned}$$ □

For (b) we first note that adding $-ac$ to both sides of $ab = ac$ and applying (A3) we get $ab - ac = 0$. Therefore, $a(b - c) = 0$ by (Dl). Since $a \neq 0$ and R is an integral domain, we must have $b - c = 0$. Thus $b = c$.

The last lemma in this section collects a few more useful properties that are satisfied in rings. We only prove (a) and (d) here, and (b), (c) and (e) are left as exercises one of which we'll do in the lectures.

Lemma 1.27. *Let R be a ring and let $a, b \in R$. Then*

(a) *$a0 = 0 = 0a$;*
(b) *$-(-a) = a$;*
(c) *$-(a + b) = -b - a$;*
(d) *$(-a)b = -ab = a(-b)$; and*

(e) $(-a)(-b) = ab$

Proof of (a) *and* (d). (a) We have $a(0+0) = a0 + a0$ by (Dl) and also $0 + 0 = 0$ by (A2). Therefore, we obtain $a0 = a0 + a0$. We also have $a0 = a0 + 0$ by (A2), so we get $a0 + 0 = a0 + a0$. From this we obtain $a0 = 0$ from the cancellation law.

We can prove that $0a = 0$ in a similar way.

(d) We have

$$\begin{aligned}(-a)b + ab &= ((-a)+a)b \quad \text{by (Dr)}\\ &= 0b \quad \text{by (A3)}\\ &= 0 \quad \text{by (a)}\end{aligned}$$

Similarly, we have $ab + (-a)b = 0$. Now we can deduce $(-a)b = -ab$ from the uniqueness of additive inverses given by Lemma 1.25(c).

The proof that $-ab = a(-b)$ is similar. □

In case R is a ring with one, we note that a consequence of Lemma 1.27(d) and (M2) is that $(-1)a = -a$ for any $a \in R$.

We end this section with a bit of notation that will be used later.
Let R be a ring, let $n \in \mathbb{Z}$ and $a \in R$.
We define $na \in R$ as follows:
if $n = 0$, then $na = 0$;
if $n > 0$, then $na = a + a + \cdots + a$, where there are n summands.
if $n < 0$, then $na = (-n)(-a)$.
Also for $n > 0$, we define $a^n = a \cdot a \cdot \cdots \cdot a$, where there are n terms in the product.
If R has a one, then we also define $a^0 = 1$,
and if a is a unit, then for $n < 0$ we define $a^n = (a^{-1})^{-n}$.

In the definition of na, we should be a bit careful when elements of $\mathbb{Z}$ can be viewed as elements of R, as the notation na could also mean the product of n and a in R. However, using the distributive law we can check that these two elements of R are in fact equal. We don't include the details here, but if you're interested feel free to ask.

1.4 Subrings

In §1.2.6, we considered the ring of Gaussian integers $\mathbb{Z}[i]$ and proved in Proposition 1.19 that it is a ring. To prove that it is a ring we used the fact that it is a subset of the ring $\mathbb{C}$, so we could deduce that all the axioms hold in $\mathbb{Z}[i]$ with the possible exception of (A0), (A2), (A3) and (M0). This is an instance of a more general phenomenon, where certain subsets of a ring are rings themselves with the addition and multiplication from the ring. Such subsets are called subrings and we define them next.

Definition 1.28. Let R be a ring and let S be a subset of R. We say that S is a *subring* of R if it is a ring with the addition and multiplication from R.

We sometimes write $S \leq R$ to mean that S is a subring of R.

We know some examples of subrings, for instance:

- $\mathbb{Z}$ is a subring of $\mathbb{Q}$;

- $\mathbb{Q}$ is a subring of $\mathbb{R}$; and
- $\mathbb{R}$ is a subring of $\mathbb{C}$.

To check whether a subset S of a ring R is a subring, we need to check that the axioms of a ring hold for S. We'll go through the axioms to see what this involves.

(A0) We need to check that for all $x, y \in S$, we have $x + y \in S$. In other words that S is closed under addition.
(A1) The associative law of addition holds, because it holds in R.
(A2) We need to check that $0 \in S$.
(A3) Given $x \in S$, we need to check that $-x \in S$.
(A4) The commutative law of addition holds, because it holds in R.
(M0) We need to check that for all $x, y \in S$, we have $xy \in S$. In other words that S is closed under multiplication.
(M1) The associative law of multiplication holds, because it holds in R.
(Dl, Dr) These distributive laws hold, because they hold in R.

We can summarize what is required for a subset of a ring to be a subring in the following lemma. We call this the first subring test. The proof is immediate from the discussion above.

Lemma 1.29 (First subring test)**.** *Let R be a ring and let S be a subset of R. Then S is a subring of R provided*

(SR1) $0 \in S$*; and*
(SR2) *for all $x, y \in S$, we have $-x \in S$, $x + y \in S$ and $xy \in S$.*

Another way of phrasing the first subring test is to say that a subset S of R is a subring provided it contains 0, and is closed under additive inverses, addition and multiplication.

It turns out that we can do a bit better than the first subring test and have a test, which allows us to reduce the amount of work required. We state this second subring test and prove that it is valid in the next lemma.

Lemma 1.30 (Second subring test)**.** *Let R be a ring and let S be a subset of R. Then S is a subring of R provided*

(SR1) $0 \in S$*; and*
(SR2′) *for all $x, y \in S$, we have $x - y \in S$ and $xy \in S$.*

Proof. We use the first subring test and check that if S satisfies the conditions of the second subring test, then it satisfies the first subring test, so that it is indeed a subring.
Certainly (SR1) holds.
Now let $x, y \in S$.
We have $0 \in S$, we obtain $0 - x = -x \in S$ by (SR2′).
Then also $-y \in S$, so $x - (-y) = x + y \in S$ again by (SR2′).
Finally, we also have $xy \in S$ by (SR2′).
Therefore, (SR2) also holds. □

Another way of phrasing the second subring test is to say that a subset S of R is a subring provided it contains 0, and is closed under subtraction and multiplication.

Note also that it is reasonably straightforward to check that if a subset S of R does not satisfy the conditions of the first or second subring test, then it is not a subring.

Another remark here is that in the second subring test, we could replace (SR1) with the weaker condition that $S \neq \varnothing$, and this is how the test is often stated. In essentially all cases where you want to show that $S \subseteq R$ is a subring, you would check that $S \neq \varnothing$ by showing that $0 \in S$, so it is not really worth making this change.

As always it will be helpful for us to look at some examples of subrings, to help us to understand Definition 1.28, and Lemmas 1.29 and 1.30.

Examples 1.31. (a) In §1.2.6, we showed that the Gaussian integers $\mathbb{Z}[i]$ are a subring of $\mathbb{C}$. There we essentially used the first subring test.

More generally, let $n \in \mathbb{Z}$ which is not a perfect square, and let $\mathbb{Z}[\sqrt{n}] = \{a + b\sqrt{n} : a, b \in \mathbb{Z}$. Then using similar arguments to those for $\mathbb{Z}[i]$, we can prove that $\mathbb{Z}[\sqrt{n}]$ is a subring of $\mathbb{C}$ using the first subring test. Note that in case $n > 0$, we have that $\mathbb{Z}[\sqrt{n}]$ is in fact a subring of $\mathbb{R}$.

Even more generally, let $f(X) \in \mathbb{Z}[X]$ be a monic polynomial, which is irreducible in $\mathbb{Q}[X]$, and let $\alpha \in \mathbb{C}$ be a root of $f(X)$. Then the ring $\mathbb{Z}[\alpha]$ defined in §1.2.6 is a subring of $\mathbb{C}$. This can be proved similarly using one of the subring tests.

(b) As promised in §1.2.10, we show that $2\mathbb{Z} = \{x \in \mathbb{Z} : x = 2y \text{ for some } y \in \mathbb{Z}\}$ is a subring of $\mathbb{Z}$. We do this using the second subring test.
First we note that $0 \in 2\mathbb{Z}$, so that $2\mathbb{Z}$ is nonempty.
Now let $a, b \in 2\mathbb{Z}$, then $a = 2c$ and $b = 2d$ for some $c, d \in \mathbb{Z}$.
We have $a - b = 2c - 2d = 2(c - d) \in 2\mathbb{Z}$, and $ab = (2c)(2d) = 4cd = 2(2cd) \in 2\mathbb{Z}$.
Therefore, $2\mathbb{Z}$ satisfies the conditions of the second subring test, and therefore is a subring of $\mathbb{Z}$.

More generally, let $n \in \mathbb{N}$ and $n\mathbb{Z} = \{x \in \mathbb{Z} : x = ny \text{ for some } y \in \mathbb{Z}\}$. Then we can apply the second subring test with entirely similar arguments to the case $n = 2$ to show that $n\mathbb{Z}$ is a subring of $\mathbb{Z}$. In fact, it is possible to prove that any subring of $\mathbb{Z}$ is equal to $n\mathbb{Z}$ for some $n \in \mathbb{N}$, which we leave as an exercise.

(c) Let R be a ring and S a subring of R. Then $S[X]$ is a subring of $R[X]$. We leave checking this as an exercise, which you can do with the first or second subring test.

(d) Let R be a ring. Denote by $R[X^2]$ the subset of $R[X]$ of all polynomials such that the coefficient of X^i for odd i is zero; in other words $R[X^2]$ is the set of polynomials in X^2. So elements of $R[X^2]$ are polynomials of the form

$$f(X) = a_{2m}X^{2m} + a_{2m-2}X^{2m-2} + \cdots + a_2X^2 + a_0,$$

where $a_0, a_2, \ldots, a_{2m} \in R$. We are going to show that $R[X^2]$ is a subring of $R[X]$ using the second subring test.

First we note that $0 \in R[X^2]$ so that $R[X^2]$ is nonempty.

Now let $f(X), g(X) \in R[X^2]$ and write $f(X) = a_{2m}X^{2m} + a_{2m-2}X^{2m-2} + \cdots + a_2X^2 + a_0$ and $g(X) = b_{2m}X^{2m} + b_{2m-2}X^{2m-2} + \cdots + b_2X^2 + b_0$. Then we have that $f(X) - g(X)$ is equal to

$$(a_{2m} - b_{2m})X^{2n} + (a_{2m-2} - b_{2m-2})X^{2m-2} + \cdots + (a_2 - b_2)X^2 + (a_0 - b_0) \in R[X^2].$$

Also we see that $f(X)g(X)$ is the sum of terms of the form $a_{2i}b_{2j}X^{2i+2j}$, so that $f(X)g(X)$ only involves even powers of X. Hence, $f(X)g(X) \in R[X^2]$.

Therefore, $R[X^2]$ satisfies the second subring test and is thus a subring of $R[X]$.

Now consider the subset S of $R[X]$ of all polynomials such that the coefficient of X^i for even i is zero; so S is the set of polynomials with only odd powers of X. We leave it as an exercise to apply one of the subring tests, to determine whether S is a subring of $R[X]$.

(e) Let R be a ring. The zero ring $\{0\}$ is a subring of R.
You should think about how to check this using one of the subring tests. Once you've thought about it for long enough you should hopefully see that it is trivial.

(f) Let R be a ring. Then R is a subring of itself.
As in the previous example, you should think about how to check this using one of the subring tests, and should see that it is trivial.

As we have seen in these examples, to show that something is a ring it is often a good idea to do this by first seeing it as at a subset of a known ring, and then use one of the subring tests. This is usually quicker than checking all of the axioms for a ring.

We end this section with a lemma saying that the intersection of two subrings is a subring.

Lemma 1.32. *Let R be a ring and let S and T be subrings of R. Then $S \cap T$ is a subring of R.*

Proof. We use the first subring test, to show that $S \cap T$ is a subring of R.
First, we note that $0 \in S$ and $0 \in T$, so $0 \in S \cap T$.
Now let $a, b \in S \cap T$. Then $a, b \in S$ and $a, b \in T$.
Thus $a + b \in S$, because S is a subring, and similarly $a + b \in T$. Hence, $a + b \in S \cap T$.
Also $-a \in S$, because S is a subring, and similarly $-a \in T$. Therefore, $-a \in S \cap T$.
Last, $ab \in S$, because S is a subring of R, and similarly $ab \in T$. Hence, $ab \in S \cap T$.

Therefore, $S \cap T$ satisfies the first subring test, and so is a subring of R. □

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 that were given in Section -1.8 and serve the same purpose of informing you what you should be able to do to demonstrate that you have understood the chapter.

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

- state the definition of a ring, including all axioms;
- check whether the axioms of a ring hold in examples;
- state the definition of zero divisors and of units, and determines them in examples;
- state the definition of an integral domain and of a field, and determine whether examples of rings are integral domains and/or fields;
- understand some examples of rings and be able to calculate in them;
- apply elementary arithmetic properties of rings; and
- state the definition of a subring and state, prove and apply the subring tests.

1.6 Exercises for Chapter 1

At the end of each of the chapters of the notes there are a set of exercises. More exercises may be added to the version of the notes on canvas, and some extra hints may be added too. Some of these exercises will be used for the formative exercise sheets that will be given out during term. Model solutions to most of the exercises will be added to the end of the notes at some point during the term.

Q1.1. Which of the following are rings with the usual addition and multiplication.

(a) $\mathbb{N}$.
(b) $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.
(c) $\mathbb{R}_{>0} = \{x \in \mathbb{R} : x > 0\}$.
(d) $3\mathbb{Z} = \{x \in \mathbb{Z} : x = 3y \text{ for some } y \in \mathbb{Z}\}$.
(e) $\{a + b\sqrt[3]{2} : a, b \in \mathbb{Z}\}$.
(f) $\{f(X) \in \mathbb{R}[X] : f(X) \text{ has constant term } 0\}$.
(g) $\{f(X) \in \mathbb{R}[X] : f(2) = 0\}$.
(h) $\{f(X) \in \mathbb{R}[X] : f(X) = 0 \text{ or } \deg f(X) \leq 5\}$.

You should give justification of your solutions, but you do not need to give detailed proofs.

Q1.2. Let $n \in \mathbb{N}$ and let $[a]_n \in \mathbb{Z}_n$ with $[a]_n \neq [0]_n$.

(a) Suppose that a is not coprime to n. Prove that $[a]_n$ is a zero divisor in $\mathbb{Z}_n$.
(b) Suppose that a is coprime to n. Prove that $[a]_n$ is a unit in $\mathbb{Z}_n$.

Q1.3. Let R be a ring.

(a) Prove that axiom (A1) holds for $R[X]$.
(b) Prove that axiom (M2) holds for $R[X]$.

Q1.4. Let R be an integral domain and let $f(X) \in R[X]$. Prove that $f(X)$ has at most $\deg f(X)$ roots in R.

Q1.5. Let $R = \mathbb{Z}_2[\beta] = \{a + b\beta : a, b \in \mathbb{Z}_2\}$, where β is a symbol that satisfies $\beta^2 = 1$.

(a) Calculate the multiplication table of R.
(b) What are the zero divisors in R?
(c) What are the units in R?

Q1.6. Let $R = \mathbb{Q}[\epsilon] = \{a + b\epsilon : a, b \in \mathbb{Q}\}$, where ϵ is a symbol that satisfies $\epsilon^2 = 0$.

(a) Show that $a + b\epsilon \in \mathbb{Q}[\epsilon]$ is a zero divisor if and only if $a = 0$.

(b) Show that $a + b\epsilon \in \mathbb{Q}[\epsilon]$ is a unit if and only if $a \neq 0$.

Q1.7. Show that $M_2(\mathbb{R})$ is a noncommutative ring.
By Proposition 1.8 we have that $M_2(\mathbb{R})$ is a ring, so you just have to show that it is not commutative.

Q1.8. Let $\mathbb{H}$ be the quaternions as defined in §1.2.12. For $z = a + bi + cj + dk \in \mathbb{H}$, define the conjugate of z to be $\overline{z} = a - bi - cj - dk \in \mathbb{H}$.

(a) Let $z = a + bi + cj + dk \in \mathbb{H}$. Prove that $\overline{z}z = z\overline{z} = a^2 + b^2 + c^2 + d^2 \in \mathbb{R}$.
(b) Deduce that all nonzero elements of $\mathbb{H}$ are units.
(c) Is $\mathbb{H}$ a field?

Q1.9. Let A be a nonempty set. Recall that the power set $\mathcal{P}(A)$ of A is by definition the set of all subsets of A, that is $\mathcal{P}(A) = \{B : B \subseteq A\}$.

(a) Define addition and multiplication on $\mathcal{P}(A)$ by $B + C := B \cup C$ and $B \cdot C := B \cap C$. Is $\mathcal{P}(A)$ a ring with this addition and multiplication?
(b) Define an addition and multiplication on $\mathcal{P}(A)$ by $B + C := (B \cup C) \setminus (B \cap C)$ and $B \cdot C := B \cap C$. Is $\mathcal{P}(A)$ a ring with this addition and multiplication?

You should justify your solutions. In this exercise it is ok to use Venn diagrams to show that axioms hold.

Q1.10. Prove Lemma 1.25(b).

Lemma. *Let R be a ring and let $e_1, e_2 \in R$. Suppose that for all $x \in R$, we have $e_1 x = x = x e_1$and $e_2 x = x = x e_2$. Then $e_1 = e_2$.*

Q1.11. Prove Lemma 1.27(b) and (e).

Lemma. *Let R be a ring and let $a, b \in R$. Then*

(b) $-(-a) = a$*; and*
(e) $(-a)(-b) = ab$

Q1.12. Let R be a commutative ring, let $n \in \mathbb{N}$ and $a, b \in R$.
Verify that the binomial theorem holds in R. That is $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.
You do not need to give a detailed proof, rather explain why the usual proof works in R. Above we are using the notation given at the send of Section 1.3

Q1.13. Let R be a integral domain with a finite number of elements. Prove that R is a field.

Hint: Let $a \in R \setminus \{0\}$, and consider the set $aR = \{ar : r \in R\}$. Using Lemma 1.26(b) show that $|aR| = R$, and deduce that there exists $r \in R$ such that $ar = 1$.

Q1.14. Let S be a subring of $\mathbb{Z}$ with $S \neq \{0\}$. Prove that $S = m\mathbb{Z} = \{mx : x \in \mathbb{Z}\}$ for some $m \in \mathbb{N}_0$.

Hint: Below are some of the steps that you may want to take, though there is some more that you will have to write.
First show that $S \cap \mathbb{N} = \varnothing$.
Let $m \in S \cap \mathbb{N}$ be minimal.
Show that for any $n \in \mathbb{Z}$ we have $nm \in \mathbb{Z}$.
Let $a \in S$ and use the division theorem to write $a = mq+r$, where $q, r \in \mathbb{Z}$ and $0 \leq r < m$.
Show that $r \in S$ and deduce that $r = 0$.

Q1.15. Let $\omega = \frac{-1+i\sqrt{3}}{2} \in \mathbb{C}$ be a primitive cube root of 1 and let $\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}$.

(a) Show that $\omega^2 + \omega + 1 = 0$.
(b) Prove that $\mathbb{Z}[\omega]$ is a subring of $\mathbb{C}$.
(c) Determine $U(\mathbb{Z}[\omega])$.

Q1.16. Let R be a ring, S a subring of R, and X an indeterminate. Prove that $S[X]$ is a subring of $R[X]$.

Q1.17. Let $R = M_2(\mathbb{R})$. Which of the following are subrings of R.

(a) The set $S = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$ of upper triangular matrices.
(b) The set $S = \left\{ \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} : a \in \mathbb{R} \right\}$ of strictly upper triangular matrices.
(c) The set $S = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$ of symmetric matrices.
(d) The set $S = \left\{ \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix} : b \in \mathbb{R} \right\}$ of skew-symmetric matrices.
(e) The set $S = \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$ of matrices of trace 0.
(f) The set $S = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a, b \in \mathbb{R} \right\}$ of diagonal matrices.

You should justify your answers.
You may also want to think about which have a one and which are commutative.

Q1.18. Let R and S be rings. Let R' be a subring of R and let S' be a subring of S. Prove that $R' \times S'$ is a subring of $R \times S$.

Q1.19. Let $\mathbb{H}$ be the quaternions as defined in §1.2.12, and define $\mathbb{Z}[i,j,k] = \{a + bi + cj + dk : a, b, c, d \in \mathbb{Z}\}$.

(a) Prove that $\mathbb{Z}[i,j,k]$ is a subring of $\mathbb{H}$.
(b) Determine $U(\mathbb{Z}[i,j,k])$.

The rest of this question is more challenging.
Let $H = \{a + bi + cj + dk : a, b, c, d \in \mathbb{Z} \text{ or } a - \frac{1}{2}, b - \frac{1}{2}, c - \frac{1}{2}, d - \frac{1}{2} \in \mathbb{Z}\}$
(so for example $-\frac{1}{2} + \frac{1}{2}i + \frac{3}{3}j + \frac{1}{2}k \in H$, but $1 + \frac{5}{2}i + 2j - \frac{1}{2}k \notin H$).

(c) Prove that H is a subring of $\mathbb{H}$.
(d) Determine $U(H)$.

Q1.20. Let R be an integral domain and let S be a subring of R containing the one of R. Prove that S is an integral domain.

Q1.21. Let R be a ring in which $x^2 = x$ for all $x \in R$.

(a) Prove that $x + x = 0$ for all $x \in R$.
(b) Prove that R is commutative.

Q1.22. Let $p \in \mathbb{N}$ be prime, and let X and Y be indeterminates. We consider the polynomial ring $\mathbb{Z}_p[X,Y]$, it is a commutative ring by applying Theorem 1.6 twice. Elements of $\mathbb{Z}_p[X,Y]$ are polynomials in X and Y.

Investigate $(X+Y)^p$ in $\mathbb{Z}_p[X,Y]$.

Work out the value of $(X+Y)^p$ in $\mathbb{Z}_p[X,Y]$ for some small values of p.
Make a conjecture of a general formula for $(X+Y)^p$.
Prove your conjecture.

Q1.23. Let $p \in \mathbb{N}$ be an odd prime.

(a) Investigate the value of $[(p-1)!]_p = [1 \cdot 2 \cdot \dots \cdot (p-2) \cdot (p-1)]_p$ in $\mathbb{Z}_p$.
(b) Investigate the value of $\left[\left(\left(\frac{p-1}{2}\right)!\right)^2\right]_p = \left[\left(1 \cdot 2 \cdot \dots \cdot \left(\frac{p-3}{2}\right) \cdot \left(\frac{p-1}{2}\right)\right)^2\right]_p$ in $\mathbb{Z}_p$.

Calculate $[(p-1)!]_p$ and $\left[\left(\left(\frac{p-1}{2}\right)!\right)^2\right]_p$ for small p, and make conjectures. Then try to prove your conjectures.

Chapter 2

Homomorphisms, ideals and quotient rings

In this chapter, we further develop the theory of rings. We introduce homomorphisms, which are maps between rings that preserve the structure of the rings. So homomorphisms give us a way of "comparing" two rings, for example saying whether they are "essentially the same", or if "parts of them are essentially the same". This naturally leads on to the theory of ideals and quotient rings, and consequently the isomorphism theorem for rings, which is one of the most important theorems in this course. In the other sections of the chapter, we cover principal ideal domains and maximal ideals, and in the final section we briefly discuss the Chinese remainder theorem for commutative rings which generalizes the Chinese remainder theorem for the integers that we saw in 1AC Algebra 1.

As mentioned earlier there is a health warning, that there are likely to be some typos in the notes, as they have recently been edited. In particular, note that in a previous version of these notes a different notation for principal ideals (as defined in Definition 2.13) was used: (a) was used instead of $\langle a \rangle$ – I've tried to ensure that all round brackets have been changed to pointy brackets, but be warned that I may have missed some.

Please let me know if you spot any. Any updates and edits will be made to the version on the 2AC Canvas page.

2.1 Homomorphisms

We begin with the definition of a homomorphism between two rings, which as mentioned above can be used to compare these rings. The word homomorphism comes from Greek and means "similar shape". The idea is that a homomorphism is a function which preserves the addition and multiplication of the rings.

Definition 2.1. Let R and S be rings, and let $\theta : R \to S$ be a function. We say that θ is a *homomorphism* provided it satisfies:

(H1) for all $a, b \in R$, we have $\theta(a + b) = \theta(a) + \theta(b)$; and
(H2) for all $a, b \in R$, we have $\theta(ab) = \theta(a)\theta(b)$.

We give a quick remark about this definition. The rings R and S both have their own addition, but we use the same notation for the addition in R and S; a similar comment relates to the multiplication. This shouldn't cause any confusion, so we won't worry about it.

Another quick comment here is that in the recommended book for this course, by Peter Cameron, homomorphisms are "written on the right", and this is the case in some other books. This won't cause you any problems, but I thought that there would be no harm warning you about it, in case anything seems confusing from comparing what is there and what is here.

Let's look at some examples of homomorphisms to help us understand them, and so that you can see that we already know some.

Examples 2.2. (a) Define $\theta : \mathbb{Z} \to \mathbb{Z}_n$ by $\theta(a) = [a]_n$. Then θ is a homomorphism. To check this we need to verify that $\theta(a+b) = \theta(a) + \theta(b)$ and $\theta(ab) = \theta(a)\theta(b)$ for all $a, b \in \mathbb{Z}$.
Well $\theta(a+b) = [a+b]_n$, which is equal to $[a]_n + [b]_n = \theta(a) + \theta(b)$ from the definition of addition in $\mathbb{Z}_n$.
Similarly, $\theta(ab) = [ab]_n$, which is equal to $[a]_n[b]_n = \theta(a)\theta(b)$ from the definition of multiplication in $\mathbb{Z}_n$.

(b) Let R be a commutative ring, let S be a subring of R, and let $c \in R$. We define the function $\epsilon_c : S[X] \to R$ by $\epsilon_c(f(X)) = f(c)$. We recall that for a polynomial $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0 \in S[X]$, the *value of* $f(X)$ *at* c is defined to be $f(c) = a_nc^n + a_{n-1}c^{n-1} + \cdots + a_1c + a_0$ as in Definition 0.24.

Now we claim that ϵ_c is homomorphism. To check this we need to observe that for $f(X), g(X) \in S[X]$ we have $\epsilon_c(f(X) + g(X)) = f(c) + g(c) = \epsilon_c(f(X)) + \epsilon_c(g(X))$ and $\epsilon_c(f(X)g(X)) = f(c)g(c) = \epsilon_c(f(X))\epsilon_c(g(X))$. We give the details below, where we see that it is essentially a consequence of the definition of polynomial addition and multiplication. You can choose to skip reading this if you're happy with it, as the details may be a bit confusing at first.

To check this we use summation notation for $f(X)$ and $g(X)$ to write $f(X) = \sum_{i=0}^{\infty} a_iX^i$, where $a_i \in S$ and there exists $n \in \mathbb{N}_0$ such that $a_i = 0$ for $i > n$, and similarly write $g(X) = \sum_{i=0}^{\infty} b_iX^i$. Then we have $f(X) + g(X) = \sum_{i=0}^{\infty}(a_i + b_i)X^i$, and so $\epsilon_c(f(X) + g(X)) = \sum_{i=0}^{\infty}(a_i + b_i)c^i = \sum_{i=0}^{\infty} a_ic^i + \sum_{i=0}^{\infty} b_ic^i = f(c) + g(c)$ as required. Also we have $f(X)g(X) = \sum_{i=0}^{\infty} d_iX^i$, where $d_i = \sum_{j=0}^{i} a_jb_{i-j}$. Therefore, $\epsilon_c(f(X)g(X)) = \sum_{i=0}^{\infty} d_ic^i$. Now consider $f(c)g(c) = (\sum_{i=0}^{\infty} a_ic^i)(\sum_{j=0}^{\infty} b_jc^j)$, we can multiply out the brackets using the distributive law and we get a sum of terms of the form $a_ib_jc^{i+j}$. Thinking about it for a short while, we see that the sum of coefficients of c^i is $d_i = \sum_{j=0}^{i} a_jb_{i-j}$; this is because polynomial multiplication is defined by multiplying out brackets. Thus $\epsilon_c(f(X)g(X)) = f(c)g(c)$ as required.

(c) Let R and S be rings.
(i) Recall that the identity map $\mathrm{id}_R : R \to R$ is defined by $\mathrm{id}_R(a) = a$ for each $a \in R$. It is trivial to check that id_R is a homomorphism from R to itself, and is left as an exercise.
(ii) The zero map $\zeta : R \to S$ is defined by $\zeta(a) = 0$ for each $a \in R$. It is trivial to check that ζ is a homomorphism, and is left as an exercise.

(d) Let R be a ring with one. For $n \in \mathbb{Z}$ we defined $n1 \in R$ at the end of Section 1.3. Thus we can define a function $\phi : \mathbb{Z} \to R$ by $\phi(n) = n1$. We can prove that ϕ is a homomorphism. The proof is straightforward (though a little technical), so we omit the details here.

Next we give a couple of easy, but useful, consequences of the definition of a homomorphism.

Lemma 2.3. *Let R and S be rings, let $\theta : R \to S$ be a homomorphism and let $a, b \in R$. Then*

(a) $\theta(0) = 0$*;*
(b) $\theta(-a) = -\theta(a)$*; and*
(c) $\theta(a - b) = \theta(a) - \theta(b)$*.*

Proof. (a) We have $\theta(0) = \theta(0) + \theta(0) - \theta(0) = \theta(0 + 0) - \theta(0) = \theta(0) - \theta(0) = 0$.
(b) We have $\theta(-a) = \theta(-a) + \theta(a) - \theta(a) = \theta(-a + a) - \theta(a) = \theta(0) - \theta(a) = 0 - \theta(a) = -\theta(a)$.
(c) We have $\theta(a - b) = \theta(a + (-b)) = \theta(a) + \theta(-b) = \theta(a) + (-\theta(b)) = \theta(a) - \theta(b)$. □

When we have a homomorphism that is a bijection, then the two rings are essentially the same, we'll explain this more later. In this case we call it an isomorphism, which we define next.

Definition 2.4. Let R and S be rings, and let $\theta : R \to S$ be a homomorphism. We say that θ is an *isomorphism* if it is a bijection. If there is an isomorphism from R to S, then we say that *R is isomorphic to S*, and we use the notation $R \cong S$ to mean R is isomorphic to S.

Recall that a bijection has an inverse, so it is natural to consider the inverse of an isomorphism.

Lemma 2.5. *Let R and S be rings, and let $\theta : R \to S$ be an isomorphism. Then $\theta^{-1} : S \to R$ is a homomorphism, and therefore an isomorphism.*

Proof. Let $c, d \in S$. Consider $\theta^{-1}(c + d)$ and $\theta^{-1}(c) + \theta^{-1}(d)$.
We have $\theta(\theta^{-1}(c + d)) = c + d$, by the definition of inverses.
Also we have $\theta(\theta^{-1}(c) + \theta^{-1}(d)) = \theta(\theta^{-1}(c)) + \theta(\theta^{-1}(d)) = c + d$.
Thus $\theta(\theta^{-1}(c + d)) = \theta(\theta^{-1}(c) + \theta^{-1}(d))$.
Therefore, since θ is injective, we have $\theta^{-1}(c + d) = \theta^{-1}(c) + \theta^{-1}(d)$.
Similarly, we can prove that $\theta^{-1}(cd) = \theta^{-1}(c)\theta^{-1}(d)$.
Hence, θ^{-1} is a homomorphism.
Also θ^{-1} is a bijection, and therefore θ^{-1} is an isomorphism. □

Let's try to elaborate a bit on what it means for two rings R and S to be isomorphic, so let $\theta : R \to S$ be an isomorphism. Then we can think of θ as just giving different names to the elements of R, without changing how they are added or multiplied. So the rings R and S are essentially the same, it's just that we give the elements different names. For example, there are different numeral systems that are used for the integers: we usually use arabic numerals and represent integers in a decimal expansion, but we could also represent them in a binary expansion (you should agree that this is just giving the integers different names and doesn't change the integers in any way); we could also give integers names by roman numerals, or by one of the Chinese numeral systems, or one of the many other numeral systems used in the world. If this doesn't make complete sense yet, then don't worry too much as it should make more sense once you've has some chance to digest the ideas a bit more.

Now lets move on to consider a couple of examples of isomorphisms.

Example 2.6. (a) Let $\gamma : \mathbb{C} \to \mathbb{C}$ be the function of complex conjugation, so by definition for $z = x + iy$, where $x, y \in \mathbb{R}$, we have $\gamma(z) = x - iy = \overline{z}$. We know for $z, w \in \mathbb{C}$, we have $\overline{z+w} = \overline{z} + \overline{w}$ and $\overline{zw} = (\overline{z})(\overline{w})$, so that γ is a homomorphism. Also it is clear that γ is a bijection, and therefore is an isomorphism.

(b) Let $R = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in M_2(\mathbb{Z}) \ : \ a, b \in \mathbb{Z} \right\}$, which is a subring of $M_2(\mathbb{Z})$ (we leave it as an exercise to check this using one of the subring tests). We define $\theta : \mathbb{Z}[i] \to R$ by $\theta(a + ib) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ for $a, b \in \mathbb{Z}$. We can check that θ is a homomorphism of rings as follows. We have

$$\begin{aligned} \theta((a+ib)+(c+id)) &= \theta((a+c)+i(b+d)) \\ &= \begin{pmatrix} a+c & b+d \\ -(b+d) & a+c \end{pmatrix} \\ &= \begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} c & d \\ -d & c \end{pmatrix} \\ &= \theta(a+ib)+\theta(c+id). \end{aligned}$$

Also we have

$$\begin{aligned} \theta((a+ib)(c+id)) &= \theta((ac-bd)+i(ad+bc)) \\ &= \begin{pmatrix} ac-bd & ad+bc \\ -(ad+bc) & ac-bd \end{pmatrix} \\ &= \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} c & d \\ -d & c \end{pmatrix} \\ &= \theta(a+ib)\theta(c+id). \end{aligned}$$

Clearly θ is a bijection and therefore it is an isomorphism of rings.

For this example let's briefly go back to the idea of isomorphism as something that just gives the elements of a ring new names. By an abuse of notation we could write $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in R$ and $i = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in R$. Then all elements of R can be written in the form $a1 + bi$, where $a, b \in \mathbb{Z}$ (so they look exactly the same as elements of $\mathbb{Z}[i]$), and we could just write $a + bi$ instead of $a1 + bi$. Whether we write $a + bi$ or $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ doesn't make any difference to the ring, it is just a different way for writing it down.

We move on to define the kernel and image of homomorphism. These are useful subrings for studying a homomorphism and key to the isomorphism theorem for rings, which is coming up.

Definition 2.7. Let R and S be rings and let $\theta : R \to S$ be a homomorphism. The *kernel of* θ is defined to be

$$\ker\theta = \{r \in R : \theta(r) = 0\}.$$

The image of θ is defined to be

$$\operatorname{im}\theta = \{s \in S : s = \theta(r) \text{ for some } r \in R\};$$

a shorter notation that we can use is $\operatorname{im}\theta = \{\theta(r) : r \in R\}$.

Let's give some examples, where we determine the kernel and image of a homomorphism.

Examples 2.8. (a) Let $n \in \mathbb{N}$ and let $\theta : \mathbb{Z} \to \mathbb{Z}_n$ be the homomorphism defined by $\theta(a) = [a]_n$ considered in Examples 2.2(a).

We have that θ is surjective (because any element of $\mathbb{Z}_n$ is of the form $[a]_n = \theta(a)$ for some $a \in \mathbb{Z}$), so $\operatorname{im}\theta = \mathbb{Z}_n$.

We have $\theta(a) = [a]_n = 0$ if and only if $a \equiv 0 \bmod n$. Therefore, $a \in \ker\theta$ if and only if $n \mid a$, so we have $\ker\theta = n\mathbb{Z} = \{nx : x \in \mathbb{Z}\}$.

(b) Define $\theta : \mathbb{R}[X] \to M_2(\mathbb{R})$ by $\theta(a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0) = \begin{pmatrix} a_0 & a_1 \\ 0 & a_0 \end{pmatrix}$. Then we can check that θ is a homomorphism.
To do this we let $f(X) = a_nX^n + \cdots + a_1X + a_0, g(X) = b_mX^m + \cdots + b_1X + b_0 \in \mathbb{R}[X]$. Then we have $f(X) + g(X) = (a_0 + b_0) + (a_1 + b_1)X +$ higher terms, so we have

$$\begin{aligned}\theta(f(X) + g(X)) &= \begin{pmatrix} a_0 + b_0 & a_1 + b_1 \\ 0 & a_0 + b_0 \end{pmatrix} \\ &= \begin{pmatrix} a_0 & a_1 \\ 0 & a_0 \end{pmatrix} + \begin{pmatrix} b_0 & b_1 \\ 0 & b_0 \end{pmatrix} \\ &= \theta(f(X)) + \theta(g(X)).\end{aligned}$$

Also $f(X)g(X) = a_0b_0 + (a_0b_1 + a_1b_0) +$ higher terms, so we have

$$\begin{aligned}\theta(f(X)g(X)) &= \begin{pmatrix} a_0b_0 & a_0b_1 + a_1b_0 \\ 0 & a_0b_0 \end{pmatrix} \\ &= \begin{pmatrix} a_0 & a_1 \\ 0 & a_0 \end{pmatrix}\begin{pmatrix} b_0 & b_1 \\ 0 & b_0 \end{pmatrix} \\ &= \theta(f(X))\theta(g(X)).\end{aligned}$$

Clearly we have

$$\operatorname{im}\theta = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \ : \ a, b \in \mathbb{R} \right\},$$

which we can check is a subring of $M_2(\mathbb{R})$.

We show that

$$\ker\theta = \{h(X) \in \mathbb{R}[X] : X^2 \text{ is a factor of } h(X)\}.$$

Let $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$.
Suppose that $f(X) \in \ker\theta$. Then $\theta(f(X)) = 0$, so that $a_0, a_1 = 0$. Therefore, $f(X) = X^2(a_nX^{n-2} + a_{n-1}X^{n-3} + \cdots + a_3X + a_2)$ and X^2 is a factor of $f(X)$. Thus $\ker\theta \subseteq \{h(X) \in \mathbb{R}[X] : X^2 \text{ is a factor of } f(X)\}$.
Now suppose that X^2 is a factor of $f(X)$. Then $f(X) = X^2g(X)$ for some $g(X) \in \mathbb{R}[X]$. Let $g(X) = b_mX^m + \cdots + b_1X + b_0$. Then $f(X) = b_mX^{m+2} + \cdots + b_1X^3 + b_0X^2$, so that $\theta(f(X)) = 0$ from the definition of θ. Thus $\{h(X) \in \mathbb{R}[X] : X^2 \text{ is a factor of } h(X)\} \subseteq \ker\theta$.
Hence, $\ker\theta = \{h(X) \in \mathbb{R}[X] : X^2 \text{ is a factor of } h(X)\}$ as claimed.
We note that an alternative way to write this down is $\ker\theta = \{X^2g(X) : g(X) \in \mathbb{R}[X]\}$, and it is straightforward to check that $\ker\theta$ is a subring of $\mathbb{R}[X]$.

(c) Consider the homomorphism $\theta = \epsilon_{\sqrt{2}} : \mathbb{Q}[X] \to \mathbb{R}$ defined by $\theta(f(X)) = f(\sqrt{2})$; this is a special case of Examples 2.2(b).

First we consider $\operatorname{im}\theta$. Let $f(X) = \sum_{k=0}^{\infty} a_kX^k$ where $a_k \in \mathbb{Q}$ and there is $n \in \mathbb{N}_0$ such that $a_k = 0$ for $k > n$. Then we have

$$\theta(f(X) = f(\sqrt{2}) = (a_0 + 2a_2 + 4a_4 - \dots) + (a_1 + 2a_3 + 4a_5 - \dots)\sqrt{2},$$

from which we can see that $\operatorname{im}\theta = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\} = \mathbb{Q}[\sqrt{2}]$.

Next we are going to show that $\ker\theta = \{(X^2 - 2)g(X) : g(X) \in \mathbb{Q}[X]\}$.

Let $f(X) \in \mathbb{Q}[X]$. Then using the division theorem (Theorem 1.16), we can write $f(X) = (X^2 - 2)q(X) + aX + b$, where $q(X) \in \mathbb{Q}[X]$ and $a, b \in \mathbb{Q}$. Then we have

$$\begin{aligned} \theta(f(X)) &= f(\sqrt{2}) \\ &= \left(\left(\sqrt{2}\right)^2 - 2\right) q(\sqrt{2}) + a\sqrt{2} + b \\ &= a\sqrt{2} + b. \end{aligned}$$

Thus, as $\sqrt{2}$ is irrational we see that $\theta(f(X)) = 0$ if and only if $a = b = 0$.
Hence, we have $f(X) \in \ker\theta$ if and only if $f(X) = (X^2 - 2)q(X)$ for some $q(X) \in \mathbb{Q}[X]$.
Therefore $\ker\theta = \{(X^2 - 2)g(X) : g(X) \in \mathbb{Q}[X]\}$ as claimed.

In the examples above we can check that the kernel and the image of the ring homomorphisms are subrings. Actually, we already said this holds in general before Definition 2.7 without justification, but the next lemma does justify it. We note that we use some of the properties of homomorphisms given by Lemma 2.3 in the proof below.

Lemma 2.9. *Let R and S be rings and let $\theta : R \to S$ be a homomorphism. Then:*

(a) $\ker\theta$ *is a subring of* R*; and*
(b) $\operatorname{im}\theta$ *is a subring of* S*.*

Proof. (a) We apply the second subring test.
First we note that $\theta(0) = 0$, so $0 \in \ker\theta$.
Now let $a, b, \in \ker\theta$. Then we have $\theta(a) = \theta(b) = 0$. Thus $\theta(a - b) = \theta(a) - \theta(b) = 0$, and $\theta(ab) = \theta(a)\theta(b) = 0$. Therefore, $a - b \in \ker\theta$ and $ab \in \ker\theta$.

(b) Again we apply the second subring test.
First we note that $\theta(0) = 0$, so $0 \in \operatorname{im}\theta$.

Now let $c, d \in \operatorname{im}\theta$, and let $a, b \in R$ such that $\theta(a) = c$ and $\theta(b) = d$. Then we have $\theta(a-b) = \theta(a) - \theta(b) = c - d$, so $c - d \in \operatorname{im}\theta$. Also we have $\theta(ab) = \theta(a)\theta(b) = cd$, so $cd \in \operatorname{im}\theta$. □

We finish the section with a proposition showing that the kernel of homomorphism tells us whether the homomorphism is injective; and also that the image tells us if it is surjective.

Proposition 2.10. *Let R and S be rings, let $\theta : R \to S$ be a homomorphism and let $a, b \in R$. Then*

(a) *$\theta(a) = \theta(b)$ if and only if $a - b \in \ker\theta$, and therefore θ is injective if and only if $\ker\theta = \{0\}$.*

(b) *θ is surjective if and only if $\operatorname{im}\theta = S$.*

Proof. (a) Suppose that $\theta(a) = \theta(b)$. Then $\theta(a-b) = \theta(a) - \theta(b) = 0$, and $a - b \in \ker\theta$. Conversely, suppose that $a - b \in \ker\theta$. Then $\theta(a) - \theta(b) = \theta(a-b) = 0$, and $\theta(a) = \theta(b)$.

Now suppose that θ is injective and let $a \in \ker\theta$. We have $\theta(0) = 0 = \theta(a)$, and thus $a = 0$, because θ is injective. It follows that $\ker\theta = \{0\}$.

Conversely, suppose that $\ker\theta = \{0\}$ and that $\theta(a) = \theta(b)$. Then $a - b \in \ker\theta$, so $a - b = 0$, and $a = b$. Thus θ is injective.

(b) This is clear from the definition of $\operatorname{im}\theta$. □

2.2 Ideals

In Lemma 2.9 we saw that the kernel of a homomorphism $\theta : R \to S$ is a subring of R. In fact, the kernel satisfies an extra condition making it a special kind of subring known as an ideal, which we define next. Ideals are also required for the construction of quotient rings, which we do in the next section.

Definition 2.11. Let R be a ring and let I be a subset of R. We say that I is an *ideal of R* if

(I1) I is a subring of R; and

(I2) for all $a \in I, r \in R$, we have $ar \in I$ and $ra \in I$.

We sometimes write $I \trianglelefteq R$ to mean that I is an ideal of R.

Before we continue we give a couple of quick remarks about ideals. You can safely skip these on a first reading.

- In case R is a commutative ring, we only require that $ra \in I$ for all $a \in I$ and $r \in I$, because $ar = ra$.
- For noncommutative rings, we can define *left ideals of R* by reducing the second condition just to $ra \in I$ for all $r \in R$ and $a \in I$; and we can define *right ideals of R* similarly. We won't use these in this course, so we won't mention them again.

Next we state and prove the ideal test, which gives us a quick way of determining whether a subset of a ring is an ideal.

Lemma 2.12 (The ideal test)**.** *Let R be a ring and let I be a subset of R. Then I is an ideal of R provided*

(IT1) $0 \in I$;
(IT2) *for all* $a, b \in I$ *we have* $a - b \in I$; *and*
(IT3) *for all* $a \in I$, $r \in R$, *we have* $ra \in I$ *and* $ar \in I$.

Proof. Assume that I satisfies the three conditions (IT1)–(IT3).
We have that $0 \in I$ by (IT1).
Now let $a, b \in I$. Then $a - b \in I$ by (IT1) and $ab \in I$ by (IT3) in the special case that $r = b \in I$.
Therefore, I is a subring of R by the second subring test.
It just remains to observe that (I2) from Definition 2.11 is the same as (IT3). □

We remark that in some books the ideal test may be stated with the condition that $0 \in I$ replaced by the weaker condition that $I \neq \varnothing$. We choose not to do this here, as in almost all cases where you want to show that $I \subseteq R$ is an ideal, you would check that $I \neq \varnothing$ by showing that $0 \in I$.

Also we note that if R is commutative then it suffices to check $ra \in I$ in (IT3) as we have $ar = ra$.

Before proceeding to some general examples of ideal we introduce a special type of ideal in a commutative ring in the next definition.

Definition 2.13. Let R be a commutative ring and let $a \in R$. The *principal ideal of R generated by a* is defined to be $\langle a \rangle = \{ar : r \in R\} \subseteq R$.

We sometimes use the alternative notation aR for $\langle a \rangle$, and this notation suggestive of what the principal ideal is. By the definition above we have $\langle a \rangle = \{ar : r \in R\}$, which means that $\langle a \rangle$ is the set of all elements of the form ar for some $r \in R$. One advantage of the notation aR is that it reminds us which ring we are considering the principal ideal in, whereas with the notation $\langle a \rangle$ we have to be a bit careful about which ring it is a principal ideal in (but this shouldn't cause any problems, so don't worry much about it).

In the next lemma we verify that principal ideals really are ideals, as the name suggests.

Lemma 2.14. *Let R be a commutative ring and let $a \in R$. Then $\langle a \rangle$ is an ideal of R.*

Proof. We use the ideal test.
First we observe that $0 = 0a \in \langle a \rangle$.
Now let $ab, ac \in \langle a \rangle$ and $r \in R$.
We have $ab - ac = a(b - c) \in \langle a \rangle$.
Also $r(ab) = a(rb) \in \langle a \rangle$.
Also we remember that as R is commutative, we don't also need to check that $(ab)r \in \langle a \rangle$).
Therefore, $\langle a \rangle$ is an ideal of R by the ideal test. □

We now move on to give some examples of ideals.

Examples 2.15. (a) Let $n \in \mathbb{Z}$. In Examples 1.31(b), we encountered $n\mathbb{Z} = \{nx : x \in \mathbb{Z}\}$ and saw that it is a subring of $\mathbb{Z}$. In fact we can show that it $n\mathbb{Z} = \langle n \rangle$ is the principal ideal of $\mathbb{Z}$ generated by a.

Actually, we remarked in 1.31(b) that any subring of $\mathbb{Z}$ is of form $n\mathbb{Z}$ for some $n \in \mathbb{N}_0$. A consequence is that any subring of $\mathbb{Z}$ is in fact an ideal, and therefore any ideal of $\mathbb{Z}$ is of the form $\langle n \rangle = n\mathbb{Z}$ for some $n \in \mathbb{N}_0$.

To help make sure, we understand the definition of a principal ideal, let's be explicit about what they are in $\mathbb{Z}$. The notation $\{nx : x \in \mathbb{Z}\}$ means that $\langle n \rangle = n\mathbb{Z}$ is the set consisting of all integers of the form nx for some $x \in \mathbb{Z}$. So for example, we have $\langle 2 \rangle = \{2x :| \; x \in \mathbb{Z}\} = \{\dots, -4, -2, 0, 2, 4, \dots\}$, and $\langle 15 \rangle = \{15x :| \; x \in \mathbb{Z}\} = \{\dots, -30, -15, 0, 15, 30, \dots\}$.

(b) Let R be a commutative ring, and let A be a subset of R.
Consider $I = \{f(X) \in R[X] : f(a) = 0 \text{ for all } a \in A\}$.
We will show that I is an ideal of $R[X]$ using the ideal test.
First recall that $0 \in R[X]$ is the zero polynomial, and the value of the zero polynomial on any element of R is $0 \in R$. Therefore, $0 \in I$.
Now let $f(X), g(X) \in I$ and $h(X) \in R[X]$, and let $a \in A$. So we have $f(a) = 0$ and $g(a) = 0$. Then the value of $f(X) - g(X)$ at a is $f(a) - g(a) = 0 - 0 = 0$. Also the value of $h(X)f(X)$ at a is $h(a)f(a) = h(a)0 = 0$. This holds for any $a \in A$, so we have $f(X) - g(X) \in I$ and $h(X)f(X) \in I$ (as $R[X]$ is commutative we also have $f(X)h(X) \in I$).
Therefore, I is an ideal of $R[X]$ by the ideal test.

(c) Let $I = \{f(X) \in \mathbb{Z}[X] \mid f(0) \text{ is even}\} \subseteq \mathbb{Z}[X]$. So for $f(X) = a_nX^n + \cdots + a_1X + a_0 \in \mathbb{Z}[X]$ we have $f(X) \in I$ if and only if a_0 is even. We leave it as an exercise to use the ideal test to check that I is an ideal of $\mathbb{Z}[X]$.

We note that any element of I can be written in the form $2a + Xh(X)$ for some $a \in \mathbb{Z}$ and $h(X) \in \mathbb{Z}[X]$. Thus we see that $I \subseteq \{2g(X) + Xh(X) : g(X), h(X) \in \mathbb{Z}[X]\}$. Also we see that $\{2g(X) + Xh(X) : g(X), h(X) \in \mathbb{Z}[X]\} \subseteq I$, because for $g(X), h(X) \in \mathbb{Z}[X]$ the value of $2g(X) + Xh(X)$ at 0 is $2g(0) + 0h(0) = 2g(0)$ is even. Hence, we have that $I = \{2g(X) + Xh(X) : g(X), h(X) \in \mathbb{Z}[X]\}$.

(d) Let R be a ring.
(i) The zero ring $\{0\}$ is an ideal of R.
(ii) R is an ideal of itself.
For these two examples, you should think about it for a little while and hopefully you will see that checking they are indeed ideals is trivial.

(e) Let $\mathbb{F}$ be a field and let I be an ideal of $\mathbb{F}$. Suppose that $a \in I$ with $a \neq 0$. Since $\mathbb{F}$ is a field, a has a multiplicative inverse a^{-1}. For any $b \in \mathbb{F}$, we have $b = b(a^{-1}a) = (ba^{-1})a \in I$, so that $\mathbb{F} \subseteq I$, and thus $I = \mathbb{F}$. It follows that the only ideals of $\mathbb{F}$ are the zero ideal $\{0\}$ and $\mathbb{F}$ itself.

More generally, the argument above shows that if an ideal I of a ring R contains a unit, then $I = R$.

It can be proved that a commutative ring with one R is a field if and only if the only ideals of R are $\{0\}$ and R. This is left as an exercise.

As promised at the start of this section, we now prove that the kernel of a homomorphism is an ideal.

Proposition 2.16. *Let R and S be rings, and let $\theta : R \to S$ be a homomorphism. Then $\ker\theta$ is an ideal of R.*

Proof. We know that $\ker\theta$ is a subring by Lemma 2.9. Let $a \in \ker\theta$ and $r \in R$. Then we have $\theta(ra) = \theta(r)\theta(a) = \theta(r)0 = 0$, so $ra \in \ker\theta$. Similarly, $ar \in \ker\theta$. Therefore, $\ker\theta$ is an ideal of R. □

In particular, we note that the kernels that we saw in Examples 2.8 are ideals. In fact they are all principal ideals: in (a) $\ker\theta = \langle n\rangle$; in (b) $\ker\theta = \langle X^2\rangle$; and in (c) $\ker\theta = \langle X^2 - 2\rangle$.

The next lemma shows how we can use ideals to get new ideals.

Lemma 2.17. *Let R be a ring and let I and J be ideals of R. Then*

(a) *$I \cap J$ is an ideal of R; and*
(b) *$I + J = \{a + b \in R : a \in I, b \in J\}$ is an ideal of R.*

Proof. (a) We know that I and J are subrings of R, and therefore $I \cap J$ is a subring of R by Lemma 1.32.
Let $a \in I \cap J$ and $r \in R$. Then we have $a \in I$, so $ar, ra \in I$, and we have $a \in J$, so $ar, ra \in J$. Therefore, $ar, ra \in I \cap J$. Hence, $I \cap J$ is an ideal of R.
(b) We prove that $I + J$ is an ideal using the ideal test.
First we note that $0 = 0 + 0 \in I + J$.
Now let $a + b, c + d \in I + J$, where $a, c \in I$, $b, d \in J$ and $r \in R$. Then we have $(a + b) - (c + d) = (a - c) + (b - d) \in I + J$, also $r(a + b) = ra + rb \in I + J$ and $(a + b)r = ar + br \in I + J$, because $a - c, ra, ar \in I$ and $b - d, rb, br \in J$.
Therefore, $I + J$ is an ideal of R by the ideal test. □

Let R be a ring and let $I_1, I_2, \ldots, I_m$ be ideals of R. Then by using (b) in the lemma above and induction we can prove that

$$I_1 + I_2 + \cdots + I_m = \{a_1 + a_2 + \cdots + a_m \in R : a_j \in I_j \text{ for } j = 1, 2, \ldots, m\}$$

is an ideal of R.

For R a commutative ring, and $a_1, a_2, \ldots, a_m \in R$, we have $\langle a_1\rangle + \langle a_2\rangle + \cdots + \langle a_m\rangle \trianglelefteq R$ and we denote it by $\langle a_1, a_2, \ldots, a_m\rangle$. So we have

$$\langle a_1, a_2, \ldots, a_m\rangle = \{r_1a_1 + r_2a_2 + \cdots + r_ma_m : r_1, r_2, \ldots, r_m \in R\}.$$

We refer to $\langle a_1, a_2, \ldots, a_m\rangle$ as the ideal of R generated by $a_1, a_2, \ldots, a_m$. For example, the ideal of $\mathbb{Z}[X]$ in Examples 2.15(c) is $\langle 2, X\rangle$.

We end this section with a useful lemma about principal ideals.

Lemma 2.18. *Let R be a commutative ring with one, let I be an ideal of R and let $a, b \in R$.*

(a) *Suppose that $a \in I$. Then $\langle a\rangle \subseteq I$.*
(b) *$\langle a\rangle \subseteq \langle b\rangle$ if and only if $a = bx$ for some $x \in R$.*
(c) *Suppose that R is an integral domain. Then $\langle a\rangle = \langle b\rangle$ if and only if $a = bu$ for some unit $u \in R$.*

Proof. (a) Since $a \in I$, we have $ar \in I$ for all $r \in R$. Hence, $\langle a\rangle \subseteq I$.
(b) Suppose that $\langle a\rangle \subseteq \langle b\rangle$. Then $a = a1 \in \langle b\rangle$ so $a = bx$ for some $x \in R$.
Conversely suppose that $a = bx$ for some $x \in R$, and let $ar \in \langle a\rangle$, where $r \in R$. Then $ar = bxr \in \langle b\rangle$. Thus $\langle a\rangle \subseteq \langle b\rangle$.
(c) is left as an exercise. □

2.3 Quotient rings

In this section we give the construction of the quotient of a ring by an ideal. The concept of quotient rings is crucial in abstract algebra, though possibly a bit difficult to grasp at first. The construction of $\mathbb{Z}_n$, the ring of integers modulo n, given in 1AC Algebra 1 and recapped in Section 0.2 is an example of this construction, and we first consider this case in Example 2.20 below. The main idea in the construction of a quotient ring is that we use an ideal to define an equivalence relation on our ring, and then define an addition and multiplication on the set of equivalence classes. These equivalence classes are the cosets, which we define next.

Definition 2.19. Let R be a ring, let I be an ideal of R and let $a \in R$.
The *coset of I with respect to a* is defined to be $a + I = \{a + x : x \in I\}$.
We say that a is a *coset representative* of the coset $a + I$.
The *set of cosets of I in R* is defined to be $R/I = \{a + I : a \in R\}$.

We note that $a + I$ is a subset of R, and R/I is a set of subsets of R. It will be of relevance later that cosets can have many different coset representatives in general. In Corollary 2.24(b), we'll see when two coset representatives give the same coset.

Sometimes we use an alternative notation for cosets and write $\overline{a}$ as a shorthand for $a+I$. We will say when we do this, and should be a bit careful when we use this notation as it means different things in different contexts. Other notation that is sometimes used in the literature is to write $[a]$ or $[a]_I$ for $a + I$.

We consider the ideal $\langle n\rangle = n\mathbb{Z}$ of $\mathbb{Z}$, for $n \in \mathbb{N}$, in the following example and relate cosets to congruence classes modulo n. We go on to demonstrate the construction of a quotient ring in this example via the ring of integers modulo n.

Example 2.20. Let $n \in \mathbb{N}$. Then we know that $\langle n\rangle$ is an ideal of $\mathbb{Z}$ by Lemma 2.14. For $a \in \mathbb{Z}$, the coset $a + \langle n\rangle$ is equal to $\{a + ny : y \in \mathbb{Z}\}$. We also have that the congruence class $[a]_n$ of a modulo n is equal to $\{x \in \mathbb{Z} : x \equiv a \bmod n\} = \{a+ny : y \in \mathbb{Z}\}$. Therefore, $a + \langle n\rangle = [a]_n$.

Now we see that we have $R/I = \mathbb{Z}/\langle n\rangle = \{a + \langle n\rangle : a \in \mathbb{Z}\} = \{[a]_n : a \in \mathbb{Z}\} = \mathbb{Z}_n$. From Section 0.2, we know that $\mathbb{Z}_n$ is a ring with addition and multiplication defined by $[a]_n + [b]_n = [a + b]_n$ and $[a]_n[b]_n = [ab]_n$, for $a, b \in \mathbb{Z}$. In the notation of cosets this can be written as $(a + \langle n\rangle) + (b + \langle n\rangle) = a + b + \langle n\rangle$ and $(a + \langle n\rangle)(b + \langle n\rangle) = ab + \langle n\rangle$.

Now write $R = \mathbb{Z}$ and $I = \langle n\rangle$. Then above we have said that R/I is an ring with addition and multiplication defined by $(a+I)+(b+I) = a+b+I$ and $(a+I)(b+I) = ab+I$ for $a, b \in R$.

Without further ado, let's move on the general definition of a quotient ring.

Definition 2.21. Let R be a ring and let I be an ideal of R. We define addition and multiplication on R/I by

$$(a + I) + (b + I) = a + b + I$$

and

$$(a + I)(b + I) = ab + I$$

for $a, b \in R$. The set R/I with this addition and multiplication is called the *quotient ring of R by I*.

The definition above should make you a bit nervous as the definitions of addition and multiplication involve a choice of a and b representing the cosets involved; though we haven't mentioned this explicitly. Below we will, as part of Theorem 2.25, show that the addition and multiplication is well defined. Also although we say implicitly that R/I is a ring in the definition, so we're going to need to check the axioms, and this is part of Theorem 2.25 too.

It is worth remarking that the notion of a quotient ring R/I is possibly a bit difficult to grasp at first. As for the ring of integers modulo n (which we have seen is a quotient ring) the elements of are subsets of R, but we also want to think about them as objects that we can add and multiply. This may take a bit of time to get used to, but once we've worked with a few quotient rings, we'll get used to idea that we can think of $a + I$ as a "symbol that we can calculate with", it sometimes help to use the shorthand $\overline{a} = a + I$ for this.

Before proceeding to Theorem 2.25 it will be helpful to consider another quotient ring.

Example 2.22. Let $I = \langle X \rangle$ be the principal ideal of $R = \mathbb{R}[X]$ generated by X. We recall that this means $I = \{Xf(X) : f(X) \in \mathbb{R}[X]\}$ or in other words I is the ideal of polynomials with constant term 0.

We're going to determine the elements of the quotient ring R/I.
Let $f(X) = a_nX^n + \dots a_1X + a_0 \in R$. Then $f(X) + I = \{f(X) + Xg(X) : g(X) \in R\}$. Now let $f(X) + Xg(X) \in f(X) + I$. Then we have

$$f(X) + Xg(X) = a_0 + X(a_nX^{n-1} + \dots + a_1 + g(X)) \in a_0 + I,$$

because $X(a_nX^{n-1} + \dots + a_1 + g(X)) \in I$. This shows that $f(X) + I \subseteq a_0 + I$. Using a similar argument we can show that $a_0 + I \subseteq f(X) + I$. Therefore, we have $f(X) + I = a_0 + I$.

It follows from the discussion in the previous paragraph that $R/I = \{a + I : a \in \mathbb{R}\}$. Moreover, we see that for $a, b \in \mathbb{R}$ with $a \neq b$ we have $a + I \neq b + I$ otherwise we would have $a = b + Xg(X)$ for some $g(X) \in R$, which is impossible. So when we write $R/I = \{a + I : a \in \mathbb{R}\}$ this gives the elements of R/I without repeats.
Alternatively, we could say $a + I = \{g(X) \in R : g(0) = a\}$; and from this deduce that for $a \neq b$, we have $a + I \neq b + I$.

Also for $a, b \in \mathbb{R}$, we can calculate $(a+I)+(b+I) = a+b+I$ and $(a+I)(b+I) = ab+I$.

Now from what we've seen above and a bit of thought, it is not difficult to prove that R/I is a ring, because we can deduce that the axioms hold from the fact that they hold in $\mathbb{R}$. In fact R/I is isomorphic to $\mathbb{R}$, but we won't worry about that right now.

In the example above, it is rather easy to show that R/I is a ring, it is a little bit more difficult to show that R/I is a ring for general R and I in Theorem 2.25 below. For our proof of Theorem 2.25 it is important for the cosets of I in R to form a partition of R. We prove this in the following proposition by showing that they are the equivalence classes of a certain equivalence relation.

Proposition 2.23. *Let R be a ring and I an ideal of R. Define the relation $\sim$ on R by $x \sim y$ means $x - y \in I$. Then $\sim$ is an equivalence relation on R. Moreover, for $a \in R$ the equivalence class $[a]_\sim$ is equal to the coset $a + I$.*

Proof. We have to show that $\sim$ is reflexive, symmetric and transitive. Let $a, b, c \in R$. We have $a - a = 0 \in I$, because I is an ideal, so $a \sim a$. Thus $\sim$ is reflexive.
Now suppose $a \sim b$, so that $a - b \in I$. Then $b - a = -(a - b) \in I$, because I is an ideal, so $b \sim a$. Thus $\sim$ is symmetric.
Suppose that $a \sim b$ and $b \sim c$, so that $a-b \in I$ and $b-c \in I$. Then $a-c = (a-b)+(b-c) \in I$, because I is an ideal, so $a \sim c$. Thus $\sim$ is transitive.
Hence, $\sim$ is an equivalence relation.

We have

$$\begin{aligned}[a]_\sim &= \{x \in R : x \sim a\}\\ &= \{x \in R : x - a \in I\}\\ &= \{x \in R : x = a + y \text{ for some } y \in I\}\\ &= \{a + y : y \in I\}\\ &= a + I.\end{aligned}$$

Hence, $[a]_\sim = a + I$. □

The equivalence relation $\sim$ in the proposition above is sometimes called *congruence modulo I*; so we would say a is congruent to b modulo I if $a - b \in I$.

We also remark that it is sufficient for I to just be a subring for the proof of Proposition 2.23 to work. However, we won't ever want to use this, and we do need I to be an ideal rather than just a subring to form a quotient (this is shown in one of the exercises).

We'll use the following corollary of Proposition 2.23 in the proof of Theorem 2.25. Actually it is mainly (b) of the corollary that we require and we will use this frequently in the rest of this chapter too; but it is nice to state all the properties in the corollary.

Corollary 2.24. *Let R be a ring, I an ideal of R and $a, b \in R$. Then*

(a) *$a \in a + I$;*
(b) *$a + I = b + I$ if and only if $a - b \in I$ if and only if $a = b + x$ for some $x \in I$;*
(c) *$a + I = b + I$ or $(a + I) \cap (b + I) = \varnothing$.*
(d) *R/I is a partition of R.*

Proof. This follows directly from Theorem 0.5 for the equivalence relation $\sim$ from Proposition 2.23, and that $[a]_\sim = a + I$ for $a \in R$ which is also part of Proposition 2.23. □

Now let's move on to showing that R/I really is a ring.

Theorem 2.25. *Let R be a ring and I an ideal of R. Then R/I with the addition and multiplication from Definition 2.21 is indeed a ring. Moreover,*

(a) *if R has a one, then R/I has a one; and*
(b) *if R is commutative, then R/I is commutative.*

Proof. First we have to show that addition and multiplication are well defined. The definitions $(a + I) + (b + I) = a + b + I$ and $(a + I)(b + I) = ab + I$ involve the implicit choice of representatives $a, b \in R$ of the cosets. We need to check that if we choose $c, d \in R$ such that $a + I = c + I$ and $b + I = d + I$, then $(a + b) + I = (c + d) + I$ and

$cd + I = ab + I$; this shows that addition and multiplication don't depend on the choice of coset representatives.

Since $a + I = c + I$, we have $a - c \in I$ so $a = c + x$ for some $x \in I$, by Corollary 2.24(b). Similarly, $b - d \in I$, so $b = d + y$ for some $y \in I$.
Therefore, $a+b = c+d+x+y$ and $x+y \in I$, because I is an ideal. Thus $(a+b)-(c+d) \in I$ and $a + b + I = c + d + I$ by Corollary 2.24(b).
Also $ab = (c + x)(d + y) = cd + cy + xd + xy$. Since $x, y \in I$ and I is an ideal of R, we have $cy \in I$, $xd \in I$ and $xy \in I$. Therefore, we have $cy + xd + xy \in I$, so that $ab - cd \in I$ and $ab + I = cd + I$ by Corollary 2.24(b).
Hence, we have proved that addition and multiplication in R/I are well defined.

Now we have to prove that R/I is a ring, which means that we have to check the axioms for a ring. We essentially deduce that they hold in R/I because they hold in R, and we do not include all the details here, rather we just check some of them, and say that others can be proved similarly.
It is clear that (A0) and (M0) hold from the definition of addition and multiplication in R/I.
We'll prove that (A1) holds. To do this let $a, b, c \in I$. Then we have

$$\begin{aligned}((a + I) + (b + I)) + (c + I) &= (a + b + I) + (c + I)\\ &= (a + b) + c + I\\ &= a + (b + c) + I\\ &= (a + I) + (b + c + I)\\ &= (a + I) + ((b + I) + (c + I)),\end{aligned}$$

which proves (A1) holds in R/I. In the above sequence of equalities the third equality uses the associative law for R, and the others are using the definition of addition in R/I.
For (A2) we note that $0 + I$ is the zero of R/I, and we do not include the details.
For (A3) we note that the additive inverse of $a + I \in R/I$ is $-a + I$, and we do not include the details.
We can prove (A4), (M1), (Dr) and (Dl) in a similar way to how we proved (A1) so we do not include the details.
We note that if R has a one $1 \in R$, then we can check that $1 + I \in R/I$ is a one in R/I, which proves (a).
Finally, we note that if (M4) holds for R, then we can prove that it holds for R/I, which proves (b). □

It is a good exercise for you to fill in some of the details omitted in the proof above, by verifying the axioms do hold in R/I.

Before we look at more examples of quotient ring, we reiterate the point made above that Corollary 2.24(b) is really important when working with quotient rings and we will use this frequently. So we restate the equivalence that we need here to help us remember it.
Let R be a ring, I be an ideal of R and $a, b \in R$. Then
$a + I = b + I$ if and only if $a = b + x$ for some $x \in I$

One way to think of this, which may possibly be helpful, is to think of elements of R/I as elements of R, but also there is a relation saying that elements of I are now thought of as being equal to zero.

Let's look at a couple of examples now to help us to understand quotient rings.

Examples 2.26. (a) Let $R = \mathbb{Z}[i]$ be the Gaussian integers, let $n \in \mathbb{N}$ be a prime, and let $I = \langle n \rangle$ be the principal ideal of R generated by n. So $I = \{na + nbi : a, b \in \mathbb{Z}\}$. We consider the quotient ring R/I. Let $a, b, c, d \in \mathbb{Z}$ then from Corollary 2.24(b) we have that $a + bi + I = c + di + I$ if and only if $a = c + nx$ and $b = d + ny$ for some $x, y \in \mathbb{Z}$; and this is true if and only if $a \equiv c \bmod n$ and $b \equiv d \bmod n$. It follows that each element of R/I can be written in the form $a + bi + I$, where $a, b \in \{0, 1, \dots, n-1\}$, Also if $a, b, c, d \in \{0, 1, \dots, n-1\}$ with $a+bi+I = c+di+I$, then $a = c$ and $b = d$. Therefore, we have $R/I = \{a + bi + I : a, b \in \{0, 1, \dots, n-1\}\}$ and this gives the elements of R/I without repeats.

We can then make calculations in R/I in a similar way to how we make calculations in $\mathbb{Z}_n$ – "that is we add and multiply and then reduce modulo n". For example, take $n = 5$, then we can calculate

$$(1 + 3i + I) + (2 + 4i + I) = 3 + 7i + I = 3 + 2i + I;$$

$$(4 + i + I)(3 + i + I) = 11 + 7i + I = 1 + 2i + I;$$

and

$$(2 + 4i + I)(1 + 3i + I) = -10 + 10i + I = 0 + I.$$

We finish this example by considering the shorthand notation that we can use for elements of R/I; you may want to skip reading through this at first, as it may be a bit confusing at first, and it is not needed for now.

In this notation we write $\overline{a + bi}$ instead of $a + bi + \langle n \rangle = a + bi + I$, which is similar to notation that we used for $\mathbb{Z}_n$ in 1AC. In this notation we have

$$\overline{4i} + \overline{2 + 2i} = \overline{2 + 6i} = \overline{2 + i}.$$

and

$$(\overline{2 + 3i})(\overline{1 + i}) = \overline{-1 + 5i} = \overline{4}.$$

So we see that this is very similar to working in $\mathbb{Z}_n$. You shouldn't be concerned about this idea of using different notation, and there is nothing mysterious going on: it just gives a different way to write down the elements of $\mathbb{Z}[i]/\langle n \rangle$, which may be easier to work with.

(b) We consider the case $R = \mathbb{R}[X]$ and $I = \langle X^2 \rangle$. (We saw $I = \ker \theta$ earlier in Examples 2.8.)

We start off by trying to think what the elements in R/I look like. Well let $f(X) \in \mathbb{R}[X]$. Then we can write $f(X)$ as $f(X) = a + bX + X^2 g(X)$ for some $g(X) \in \mathbb{R}[X]$. So by Corollary 2.24(b) we have $f(X) + I = a + bX + I$, because $X^2 g(X) \in I$. Moreover, for $a, b, c, d \in \mathbb{R}$, we have $a + bX + I = c + dX + I$ if and only if $a + bX = c + dX + X^2 g(X)$ for some $g(X) \in \mathbb{R}[X]$. This is only possible if $g(X) = 0$, $a = c$ and $b = d$. Therefore, we see that $R/I = \{a + bX + I : a, b, \in \mathbb{R}\}$ and this gives the elements of R/I without repeats.

Now let's look at how we add and multiply elements of $R/I = \{a+bX+I : a, b, \in \mathbb{R}\}$. Let $a, b, c, d \in \mathbb{R}$. Then we have

$$(a + bX + I) + (c + dX + I) = (a + c) + (b + d)X + I$$

and

$$\begin{aligned}(a+bX+I)(c+dX+I) &= (ac)+(ad+bc)X+bdX^2+I\\ &= (ac)+(ad+bc)X+I,\end{aligned}$$

where we have used Corollary 2.24(b) and the fact that $bdX^2 \in I$ for the last equality.

So we have seen that the elements of R/I can be written in a nice way as $a+bX+I$ for $a, b \in \mathbb{R}$; we could refer to such an expression for an element of R/I as a *normal form.* Further, we have rules for adding and multiplying elements in this normal form together as given above: roughly speaking we add and multiply them as polynomials and then use the relation that $X^2+I = 0+I$

The last part of this example may seem a little confusing at first, but it is just trying to emphasize that this quotient ring is something that is quite easy to work with. You may want to skim through it on a first reading, and not worry if it you don't understand it all, as you'll understand it better later when we've seen more examples.

We can use the alternative notation for elements of R/I, and write $\overline{a+bX}$ as a shorthand for $a+bX+I$, for $a, b \in \mathbb{R}$. There is nothing mysterious going on, all we're saying is that we write $\overline{a+bX}$ for $a+bX+I$, as it is a bit shorter to do this. We introduce one more piece of new notation and let $\alpha = \overline{X}$, i.e. we just write α instead of $\overline{X}$.

Now for $a, b \in R$ we have $\overline{a} = a+I$ and $\overline{b} = b+I$. Thus we have

$$\overline{a+bX} = \overline{a}+\overline{b}\alpha,$$

because

$$a+bX+I = (a+I)+(b+I)(X+I).$$

Thus elements of R/I are expressions of the form $\overline{a}+\overline{b}\alpha$, where $a, b \in \mathbb{R}$, and addition and multiplication are defined by

$$(\overline{a}+\overline{b}\alpha)+(\overline{c}+\overline{d}\alpha) = (\overline{a+c})+(\overline{b+d})\alpha$$

and

$$(\overline{a}+\overline{b}\alpha)(\overline{c}+\overline{d}\alpha) = (\overline{ac})+(\overline{ad+bc})\alpha.$$

This gives us a way to work with this quotient ring.

If we abuse notation, then we could take things one step further and not include the bars and just write $a+b\alpha$ instead of $\overline{a}+\overline{b}\alpha$. This is an abuse of notation and can be confusing, so we won't do this yet, as we need to careful about what is what. But we do note that if we did this, then R/I would look just like one of the examples of rings that we considered in §1.2.7. In fact viewing the rings from §1.2.7 in this way, leads to the best way to prove that they really are rings, as this is given by Theorem 2.25.

Next we have a proposition about the canonical homomorphism from a ring to a quotient ring; the word canonical is used quite frequently in mathematics though less frequently in everyday English, it is used to mean that it is defined in some sort of standard or natural way. The proposition is a consequence of the definition of addition and multiplication in R/I, so the proof is pretty short.

Proposition 2.27. *Let R be a ring and I an ideal of R. Then the function $\pi : R \to R/I$ defined by $\pi(a) = a+I$ is a homomorphism. Moreover, $\ker \pi = I$ and $\operatorname{im} \pi = R/I$*

Proof. Let $a, b \in R$. Then $\pi(a+b) = a+b+I = (a+I)+(b+I) = \pi(a)+\pi(b)$ and $\pi(ab) = ab+I = (a+I)(b+I) = \pi(a)\pi(b)$. Hence, π is a homomorphism.

Clearly, we have $\operatorname{im}\pi = R/I$, as any element of R/I is of the form $a+I = \pi(a)$ for some $a \in R$.

Now let $a \in \ker\pi$. Then $\pi(a) = a+I = 0+I$, we recall that $0+I$ is the zero in R/I. Therefore, $a = a-0 \in I$. So $\ker\pi \subseteq I$. Conversely if $a \in I$, then $\pi(a) = a+I = 0+I$. Thus $I \subseteq \ker\pi$. Hence, $\ker\pi = I$. □

The homomorphism $\pi : R \to R/I$ from the proposition is called the *canonical homomorphism* from R to R/I. We have already seen in Proposition 2.16 that the kernel of a homomorphism is an ideal, and Proposition 2.27 gives a sort of converse to this by saying that any ideal is in fact the kernel of a homomorphism.

We'll have one final remark at the end of this section about how we can think of quotient rings. We can view a quotient ring R/I as "part of" the ring R, in a similar way to how we view a subring S of R as a "part of" R. We think of S as being the part where we only consider some of the elements of R, whereas we think of R/I as being the part where we consider the elements of I to be "equivalent to 0". So in a sense the elements of R/I can be thought of as an approximation of elements of R. For example in $\mathbb{Z}/\langle n\rangle = \mathbb{Z}_n$, the elements are approximations to integers where we just see the remainder when divided by n, and in $\mathbb{R}[X]/\langle X^2\rangle$ we take the approximation of a polynomial where we just see the constant term and the X term. You may think of this as looking at elements of a ring R, and then squinting so that you only see part of them. Don't worry much if this does not fully make sense yet; the concept of quotient rings is a bit difficult to grasp at first.

2.4 The isomorphism theorem

In this section, we prove the isomorphism theorem. This is one of the most important theorems of the course. It shows that a homomorphism from a ring R to a ring S gives an isomorphism between a quotient ring of R and a subring of S; this explains the vague statement in the introduction to this chapter that a homomorphism says that "parts of the rings are essentially the same".

Although the isomorphism theorem may be a bit tricky to understand at first, let's dive in and state and prove the theorem. Afterwards we'll look at some examples to help us understand the theorem. Remember that some things in maths take a little while to sink in, so you might not understand this proof fully at first, but stick with it and later on you'll maybe wonder why you found it difficult in the first place. The proof of the theorem is completely natural in the sense that once you understand the statement and what you need to prove, then there is only one sensible way to do it.

For the statement of the isomorphism theorem we recall from Proposition 2.16 that the kernel of a homomorphism $\theta : R \to S$ is an ideal of R, so we can define the quotient ring $R/\ker\theta$. Also we recall from Lemma 2.9 that $\operatorname{im}\theta$ is a subring of S.

Theorem 2.28 (The isomorphism theorem)**.** *Let R and S be rings and let $\theta : R \to S$ be a homomorphism. Then $R/\ker\theta \cong \operatorname{im}\theta$.*

Proof. For the proof we let $I = \ker\theta$.

We define $\overline{\theta} : R/I \to \operatorname{im}\theta$ by $\overline{\theta}(a+I) = \theta(a)$. We are going to show that $\overline{\theta}$ is an isomorphism, but first we need to check that it is well defined.

To check that θ is well defined, we let $a, b \in R$ and suppose that $a+I = b+I$. Then we have $a - b \in I = \ker\theta$ by Proposition 2.23, and thus $\theta(a) = \theta(b)$ by Proposition 2.10(a). Therefore, the value of $\overline{\theta}(a+I)$ does not depend on the choice of coset representative, and thus $\overline{\theta}$ is well defined.

We move on to prove that $\overline{\theta}$ is a homomorphism. So let $a+I, b+I \in R/I$. Then we have

$$\begin{aligned}\overline{\theta}((a+I)+(b+I)) &= \overline{\theta}(a+b+I)\\ &= \theta(a+b)\\ &= \theta(a)+\theta(b)\\ &= \overline{\theta}(a+I)+\overline{\theta}(b+I).\end{aligned}$$

Similarly,

$$\begin{aligned}\overline{\theta}((a+I)(b+I)) &= \overline{\theta}(ab+I)\\ &= \theta(ab)\\ &= \theta(a)\theta(b)\\ &= \overline{\theta}(a+I)\overline{\theta}(b+I).\end{aligned}$$

Therefore, $\overline{\theta}$ is a homomorphism.

We are left to show that $\overline{\theta}$ is a bijection. We note that $\overline{\theta}$ is clearly surjective from the definitions of $\overline{\theta}$ and $\operatorname{im}\theta$. Now suppose that $\overline{\theta}(a+I) = \overline{\theta}(b+I)$. Then $\theta(a) = \theta(b)$, so that $a - b \in \ker\theta = I$ by Proposition 2.10(a). Thus, $a+I = b+I$ by Proposition 2.23. So $\overline{\theta}$ is injective and therefore bijective.

Putting this all together, we have proved that $\overline{\theta}$ is an isomorphism as required. □

As promised above, we'll give a couple of examples now to help us understand the isomorphism theorem.

Examples 2.29. (a)(i) Consider the homomorphism $\theta = \epsilon_i : \mathbb{R}[X] \to \mathbb{C}$ defined by $\epsilon_i(f(X)) = f(i)$, this homomorphism was introduced in Examples 2.2(b).

First we observe that $\theta(a+bX) = a+bi$, and deduce that $\operatorname{im}\theta = \{a+bi : a, b \in \mathbb{R}\} = \mathbb{C}$.

Now consider $\ker\theta$, we are going to show that $\ker\theta = \langle X^2+1\rangle$ is the principal ideal generated by X^2+1. Let $I = \langle X^2+1\rangle$.

Suppose that $f(X) \in \ker\theta$, so that $\theta(f(X)) = 0$. By the division theorem for polynomials we can write $f(X) = (X^2+1)q(X) + a + bX$ for some $q(X) \in \mathbb{R}[X]$ and $a, b \in \mathbb{R}$. Then we have

$$\begin{aligned}\theta(f(X)) = f(i) &= (i^2+1)q(i) + a + bi\\ &= a + bi.\end{aligned}$$

Hence, $a + bi = 0$, so $a = 0 = b = 0$. Therefore, $f(X) = (X^2+1)q(X)$ and $f(X) \in I$. Thus $\ker\theta \subseteq I$.

Now let $f(X) \in I$. Then $f(X) = (X^2+1)g(X)$ for some $g(X) \in \mathbb{R}[X]$. So $f(i) = (i^2+1)g(i) = 0$. Thus $I \subseteq \ker\theta$.
Hence, $\ker\theta = I$.

Therefore, from the isomorphism theorem we obtain $\mathbb{R}[X]/I \cong \mathbb{C}$.

To try to understand this isomorphism, let's see what the quotient ring $\mathbb{R}[X]/I$ looks like. By the division theorem for polynomials we can write any $f(X) \in \mathbb{R}[X]$ in the form $f(X) = (X^2+1)q(X) + a + bX$ for some $q(X) \in \mathbb{R}[X]$ and $a, b \in \mathbb{R}$. Therefore, we have $f(X) + I = a + bX + I$. Moreover, if for $a, b, a', b' \in \mathbb{R}$, we have $a + bX + I = a' + b'X + I$, then $a - a' + (b - b')X \in I$, so that $X^2 + 1 \mid (a - a') + (b - b')X$, which forces $a = a'$ and $b = b'$. Therefore, we can write $\mathbb{R}[X]/I = \{a + bX + I : a, b \in I\}$ and this gives the elements without repeats.

Let's use the alternative notation for $\mathbb{R}[X]/I$ and write $\overline{a+bX} = a + bX + I$, and also we'll let $\alpha = \overline{X}$. (This is similar to what we did in Examples 2.26(b).) Then we have

$$\begin{aligned} a + bX + I &= (a + I) + (b + I)(X + I) \\ &= \overline{a} + \overline{b}\alpha, \end{aligned}$$

so that the elements of R/I are of the form $\overline{a} + \overline{b}\alpha$, where $a, b \in \mathbb{R}$.

Let $a, b, c, d \in \mathbb{R}$. Then we quite easily calculate that

$$(\overline{a} + \overline{b}\alpha) + (\overline{c} + \overline{d}\alpha) = (\overline{a+c}) + (\overline{b+d})\alpha.$$

Further we can calculate that

$$\begin{aligned} \alpha^2 &= (X + I)^2 \\ &= X^2 + I \\ &= -1 + I \\ &= \overline{-1}. \end{aligned}$$

In the above we get the equality $X^2 + I = -1 + I$ using Corollary 2.24(b), and that $X^2 = -1 + (X^2 + 1)$ and $X^2 + 1 \in I$. From this we deduce that

$$\begin{aligned} (\overline{a} + \overline{b}\alpha)(\overline{c} + \overline{d}\alpha) &= \overline{ac} + \overline{ad + bc}\alpha + \overline{bd}\alpha^2 \\ &= \overline{ac} + (\overline{ad + bc})\alpha + \overline{bd}\,\overline{-1} \\ &= \overline{ac} + (\overline{ad + bc})\alpha - \overline{bd} \\ &= (\overline{ac - bd}) + (\overline{ad + bc})\alpha. \end{aligned}$$

So we see that in this notation $\mathbb{R}[X]/I$ looks more or less the same as $\mathbb{C}$, we are just writing $\overline{a} + \overline{b}\alpha$ for elements of $\mathbb{R}[X]/I$ in place of the notation $a + bi$, which we use in $\mathbb{C}$. Indeed the isomorphism $\overline{\theta} : \mathbb{R}[X]/I \to \mathbb{C}$ from the proof of the isomorphism theorem is simply the function defined by $\overline{\theta}(\overline{a} + \overline{b}\alpha) = a + bi$.

If we abuse notation, then we could have chosen another notation for $\mathbb{R}[X]/I$, which would make it look just like $\mathbb{C}$. We just remark about this at the end of this example, as it may sound a bit confusing at first. You can safely skip reading this and should do so it is doesn't make sense.

By abusing notation we could write a for $a + I$ (so we're not bothering with the bar now, and we shouldn't really do this as we do not mean that a is equal to $a + I$, it's just that we're being lazy), and we could write i for $X + I$ (rather than α as we did previously). Then the elements of $\mathbb{R}[X]/I$ are of the form $a + bi$ for $a, b \in R$, and the rules for adding and multiplying them is exactly the same as in $\mathbb{C}$. The abuse of notation is a bit like when we may sometimes write a instead of $[a]_n$ when working in $\mathbb{Z}_n$. The main point is that we should only do this if we're careful and we understand that we are abusing notation.

(ii) Similarly to in (i) we can also consider $\epsilon_i : \mathbb{Z}[X] \to \mathbb{C}$. In this case we obtain that $\operatorname{im} \epsilon_i = \mathbb{Z}[i]$ and that $\ker \epsilon_i = \langle X^2 + 1\rangle$ (where $\langle X^2 + 1\rangle$ now denotes the principal ideal in $\mathbb{Z}[X]$). So we obtain an isomorphism $\mathbb{Z}[X]/\langle X^2 + 1\rangle \cong \mathbb{Z}[i]$.

(b) Recall the homomorphism $\theta : \mathbb{R}[X] \to M_2(\mathbb{R})$ defined by

$$\theta(a_nX^n + \cdots + a_1X + a_0) = \begin{pmatrix} a_0 & a_1 \\ 0 & a_0 \end{pmatrix}$$

from Example 2.8. We worked out there that $\ker \theta = \langle X^2\rangle$ is the principal ideal generated by X^2, though we did not phrase it like that there. Also we saw that

$$\operatorname{im} \theta = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \ : \ a, b \in \mathbb{R} \right\},$$

Let $I = \ker \theta$. We have $\mathbb{R}[X]/I \cong \operatorname{im} \theta$ by the isomorphism theorem.

In Examples 2.26(b), we already looked at $\mathbb{R}[X]$ and there we saw that $\mathbb{R}[X]/I = \{a+bX+I : a, b \in \mathbb{R}\}$ and there are no repeated elements in this set. Then we introduced the notation where $\overline{a} + \overline{b}\alpha = a + bX + I$, and we have that $\alpha^2 = 0$. The isomorphism $\overline{\theta} : R[X]/I \to \operatorname{im} \theta$ from the isomorphism theorem is given by $\overline{\theta}(\overline{a} + \overline{b}\alpha) = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$.

Let us also give some notation for elements of $\operatorname{im} \theta$. We write $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then we can calculate that $\beta^2 = 0$. So all elements of $\operatorname{im} \theta$ can be written in the $a + b\beta$ for $a, b \in \mathbb{R}$, and we have $\beta^2 = 0$.

In this notation $\overline{\theta}$ is given by $\overline{\theta}(\overline{a} + \overline{b}\alpha) = a + b\beta$, which we can see is indeed an isomorphism.

(c) Let's have one more example and consider the homomorphism $\theta = \epsilon_{\sqrt{2}} : \mathbb{Z}[X] \to \mathbb{R}$. We'll give less details here.

We can show that $\operatorname{im} \theta = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\} = \mathbb{Z}[\sqrt{2}]$. Then we can also prove that $\ker \theta = \langle X^2 - 2\rangle$. Next we write $I = \ker \theta$ and use the division theorem to show that $\mathbb{Z}[X]/I = \{a + bX + I : a, b \in \mathbb{Z}\}$ and that this gives the elements of $\mathbb{Z}[X]/I$ without repeats.

The isomorphism $\overline{\theta}$ from the isomorphism theorem is given by $\overline{\theta}(a+bX+I) = a+b\sqrt{2}$.

Writing $\alpha = X + I$ and $\overline{a} = a + I$ for $a \in \mathbb{Z}$, we see that the elements of $\mathbb{Z}[X]/I$ are of the form $\overline{a} + \overline{b}\alpha$ and we have $\alpha^2 = \overline{2}$. In this notation we have $\overline{\theta}(\overline{a} + \overline{b}\alpha) = a + b\sqrt{2}$.

2.5 Principal ideal domains

We now have a short section where we give the definition of principal ideal domains, and give two important examples of principal ideal domains.

We start with the definition of a principal ideal domain.

Definition 2.30. Let R be an integral domain. We say that R is a *principal ideal domain* if for any ideal I of R, there exists $a \in R$ such that $I = \langle a \rangle = \{ar : r \in R\}$ is the principal ideal generated by a.

We often use the term PID as an abbreviation of principal ideal domain.

You should note that in the definition above it is implicit that a principal ideal domain is an integral domain.

The following two propositions give our main examples of principal ideal domains. First we consider the ring of integers.

Proposition 2.31. *$\mathbb{Z}$ is a principal ideal domain.*

Proof. Let I be an ideal of $\mathbb{Z}$. We need to show $I = \langle m \rangle$ for some $m \in \mathbb{Z}$.
First we consider the case where $I = \{0\}$. Then we have $I = \langle 0 \rangle$.
So we may assume that $I \neq \{0\}$, and let $a \in I$ be nonzero. Then also $-a \in I$, so there is a positive element of I.
We let $m \in I \cap \mathbb{N}$ be the smallest positive element of I.
Then for all $a \in \mathbb{Z}$ we have $ma \in I$, because I is an ideal of $\mathbb{Z}$. Therefore, $\langle m \rangle \subseteq I$.
Now let $a \in I$, then by the division theorem there exist $q, r \in \mathbb{Z}$ with $a = qm + r$ and $0 \leq r < m$. Then $r = a - qm \in I$. Also m was the chosen to be the smallest element of I, so we must have $r = 0$ and $a = qm \in \langle m \rangle$. Therefore, $I \subseteq \langle m \rangle$.
Hence, $I = \langle m \rangle$. □

We note there was an exercise to show that any subring of $\mathbb{Z}$ is equal to $\langle m \rangle = m\mathbb{Z}$ for some $m \in \mathbb{Z}$. This exercise implies that $\mathbb{Z}$ is a principal ideal domain, as any ideal is a subring but we chose to include a proof here, as the proof for ideals is slightly simpler.

We move on to consider rings of polynomials over fields.

Proposition 2.32. *Let $\mathbb{F}$ be a field. The polynomial ring $\mathbb{F}[X]$ is a principal ideal domain.*

Proof. Let I be an ideal of $\mathbb{F}[X]$. We need to show $I = \langle m(X) \rangle$ for some $m(X) \in \mathbb{F}[X]$.
First we consider the case where $I = \{0\}$. Then we have $I = (0)$.
So we may assume that $I \neq \{0\}$.
We let $m(X)$ be a polynomial of minimal degree in I.
Then for all $f(X) \in \mathbb{F}[X]$ we have $m(X)f(X) \in I$, because I is an ideal of $\mathbb{F}[X]$. Therefore, $\langle m(X) \rangle \subseteq I$.
Now let $f(X) \in I$, then by the division theorem for polynomials (Theorem 0.26) there exist $q(X), r(X) \in \mathbb{F}[X]$ with $f(X) = q(X)m(X) + r(X)$, and $r(X) = 0$ or $\deg r(X) < \deg m(X)$. Then $r(X) = f(X) - q(X)m(X) \in I$. Also $m(X)$ was the chosen to be a polynomial of minimal degree in I, so we must have $r(X) = 0$ and $f(X) = q(X)m(X) \in \langle m(X) \rangle$. Therefore, $I \subseteq \langle m(X) \rangle$.
Hence, $I = \langle m(X) \rangle$. □

We recall that Lemma 2.18 gave us some useful properties of principal ideals. As a consequence of (b) of that lemma and the fact that $\mathbb{F}[X]$ is a principal ideal domain, we can deduce that any nonzero ideal of $\mathbb{F}[X]$ can be written in the form $\langle m(X) \rangle$ where

$m(X)$ is a monic polynomial – you should think about why this is true. Similarly, any nonzero ideal of $\mathbb{Z}$ can be written in the form $\langle m\rangle$ where $m \in \mathbb{N}$.

We end this section by noting that there are plenty of examples of integral domains that are not principal ideal domains. We don't mention any here and just say that there are some exercises giving examples.

2.6 Maximal ideals

We begin with the definition of a maximal ideal.

Definition 2.33. Let R be a ring and I an ideal of R. We say that I is a *maximal ideal* if $I \neq R$, and for any ideal J of R with $I \subseteq J \subseteq R$, we have $J = I$ or $J = R$.

In other words I is a maximal ideal if there are no ideals properly between I and R. Alternatively, I is not maximal if there exists an ideal J of R with $I \subseteq J \subseteq R$ such that $J \neq I$ and $J \neq R$.

In the following two propositions we give the maximal ideals in $\mathbb{Z}$ and in $\mathbb{F}[X]$ (where $\mathbb{F}$ is a field). In these propositions we show which principal ideals are maximal, but note that this gives all of the maximal ideals, as we have proved that $\mathbb{Z}$ and $\mathbb{F}[X]$ are principal ideal domains in the previous section.

Proposition 2.34. *Let $n \in \mathbb{N}$ with $n \neq 1$. The principal ideal $\langle n\rangle$ is maximal if and only if n is prime.*

Proof. First suppose that $\langle n\rangle$ is not maximal and let J be an ideal of $\mathbb{Z}$ with $\langle n\rangle \subsetneqq J \subsetneqq \mathbb{Z}$. Then since $\mathbb{Z}$ is a principal ideal domain (by Proposition 2.31), we have $J = \langle m\rangle$ for some $m \in \mathbb{Z}$. Then $\langle n\rangle \subsetneqq \langle m\rangle$ so $n = mx$ for some $x \in \mathbb{Z}$ by Lemma 2.18(a), and x is not a unit by Lemma 2.18(b). Therefore, n is not prime.

Now suppose that n is not prime, then we have $n = mx$ where $m, x \in \mathbb{N}$ with $1 < m, x < n$. Then we have $\langle n\rangle \subsetneqq \langle m\rangle$ by Lemma 2.18, so $\langle n\rangle$ is not maximal. □

We recall that $\mathbb{Z}/\langle n\rangle = \mathbb{Z}_n$, and we also know that $\mathbb{Z}_n$ is a field if and only if n is prime. So we see that $\langle n\rangle$ is maximal if and only if $\mathbb{Z}/\langle n\rangle$ is a field; this is a special case of Theorem 2.36 below. Recall also that, for $p \in \mathbb{N}$, we sometimes write $\mathbb{F}_p$ instead of $\mathbb{Z}_p$ to commemorate that it is a field.

We move on to consider maximal ideals in $\mathbb{F}[X]$.

Proposition 2.35. *Let $\mathbb{F}$ be a field and let $m(X) \in \mathbb{F}[X]$ be a nonconstant polynomial. Then $\langle m(X)\rangle$ is a maximal ideal of $\mathbb{F}[X]$ if and only if $m(X)$ is an irreducible polynomial.*

We omit the proof of this proposition as it is very similar to the proof of Proposition 2.34 above. It is an exercise to adapt that proof to prove Proposition 2.35.

We now go on to consider the quotient ring of a commutative ring by a maximal ideal, and show in the next theorem that this is a field. In fact this theorem also gives the converse.

Theorem 2.36. *Let R be a commutative ring with one and let I be an ideal of R. Then R/I is a field if and only if I is maximal.*

Proof. We have that R/I is a commutative ring with one.

First we suppose that I is maximal and aim to show that R/I is a field. So we just need to show that every nonzero element of R/I is a unit. Let $a \in R$ such that $a \notin I$, so that $a + I$ is a nonzero element of R/I. We need to show that there exists $r \in R$ such that $(a + I)(r + I) = 1 + I$.

We consider $J = \langle a \rangle + I = \{ra + x : r \in R, x \in I\}$, which is an ideal of R by Lemma 2.17 and clearly we have that $I \subseteq J$. Since $a \notin I$ and $a = 1a + 0 \in J$, we have $I \subsetneq J$. Therefore, we have $J = R$ by maximality of I. Thus, there exist $r \in R$ and $x \in I$ such that $ra + x = 1$. Hence, we have $(a + I)(r + I) = ar + I = 1 + I$ as required.

Now suppose that R/I is a field, and let J be an ideal of R that properly contains I. Let $x \in J \setminus I$. Then we have $x + I$ is a nonzero element of R/I, and thus $x + I$ has a multiplicative inverse in R/I, because R/I is a field. In other words there exists $y \in R$ such that $(x+I)(y+I) = 1+I$. Therefore, we have $xy+I = 1+I$, so that $1-xy \in I \subseteq J$. Since $x \in J$, we also have $xy \in J$ and thus $1 = (1 - xy) + xy \in J$. Hence, for any $r \in R$, we have $r = r1 \in J$, so that $J = R$. This shows that I is a maximal ideal. □

Next we go on to give a couple of examples of fields obtained in this way. Before this we introduce some notation that we will use from now on.

Examples 2.37. (a) Let $m(X) = X^3 - 2 \in \mathbb{Q}[X]$ and $I = \langle m(X) \rangle$ the principal ideal in $\mathbb{Q}[X]$ generated by $m(X)$. If $m(X)$ was reducible in $\mathbb{Q}[X]$, then it must have a linear factor, and therefore a root in $\mathbb{Q}$. But we know that the roots of $m(X)$ are $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$, where $\omega = e^{\frac{2\pi i}{3}}$, and none of these are in $\mathbb{Q}$. Therefore, $m(X)$ is irreducible and $\mathbb{Q}[X]/I$ is a field.

Let $\alpha = \sqrt[3]{2}$ and consider the homomorphism $\theta = \epsilon_\alpha : \mathbb{Q}[X] \to \mathbb{R}$, which we recall is defined by $\theta(f(X)) = f(\alpha)$.
We'll show that $\ker\theta = I$. To do this note that $\ker\theta$ is an ideal of $\mathbb{Q}[X]$, so $\ker\theta = \langle g(X) \rangle$ for some $g(X) \in \mathbb{Q}[X]$ by Proposition 2.32, and we can assume that $g(X)$ is monic. We have $m(X) \in \ker\theta$, because $m(\alpha) = 0$. Therefore, $g(X)$ is a factor of $m(X)$, because $m(X) \in \ker\theta$. But $m(X)$ is irreducible, so we deduce that $g(X)$ is equal to either 1 or $m(X)$, and $m(X) \neq 1$, because $1 \notin \ker\theta$. Thus $g(X) = m(X)$ and $\ker\theta = \langle m(X) \rangle = I$.

It is straightforward to show that $\operatorname{im}\theta = \{a + b\alpha + c\alpha^2 : a, b, c \in \mathbb{Q}\}$ using arguments similar to those in Examples 2.29(a). We write $\mathbb{Q}(\alpha) = \{a + b\alpha + c\alpha^2 : a, b, c \in \mathbb{Q}\}$. By the isomorphism theorem we have $\mathbb{Q}[X]/I \cong \mathbb{Q}(\alpha)$ so that $\mathbb{Q}(\alpha)$ is a field. In fact $\mathbb{Q}(\alpha)$ is the smallest subfield of $\mathbb{C}$ containing α. The isomorphism $\overline{\theta} : \mathbb{Q}[X]/I \to \mathbb{Q}(\alpha)$ given by the isomorphism theorem sends $X + I$ to α.

(b) In this example we work with the finite field $\mathbb{F}_p$ for $p = 3$, and we use some different notation, which we explain first. We use the notation a rather than $[a]_p$ for elements of $\mathbb{F}_p$, so that $\mathbb{F}_p = \{0, 1, 2, \dots, p - 1\}$. But we do also allow ourselves to write other integers as elements of $\mathbb{F}_p$, for example we may consider -1 to be an element of $\mathbb{F}_p$, and this is equal to $p - 1$. This could potentially be confusing, but it doesn't tend to cause any problems, and it saves time and space. I expect you'll get used to it quickly, but you should be careful at first.

Consider $\mathbb{F}_3[X]$ and the polynomial $f(X) = X^2 + 1 \in \mathbb{F}_3[X]$. Let $I = \langle f(X) \rangle$ be the principal ideal of $\mathbb{F}_3[X]$ generated by $f(X)$. We can check that $f(X)$ is irreducible in $\mathbb{F}_3[X]$. If $f(X)$ was reducible, then it would have a linear factor, and thus a root by the factor theorem. But we can calculate that $f(0) = 1$, $f(1) = 2$ and $f(2) = 5 = 2$, so that $f(X)$ has no roots in $\mathbb{F}_3$. Therefore, $I = \langle f(X) \rangle$ is a maximal ideal of $\mathbb{F}_3[X]$, and so $F_3[X]/I$ is a field.

Using a similar arguments to those in Examples 2.29, we can show that any element of $\mathbb{F}_3[X]/I$ can be written uniquely in the form $a + bX + I$, where $a, b \in \mathbb{F}_3$. So that $\mathbb{F}_3[X]/I = \{a + bX + I : a, b \in \mathbb{F}_3\}$, and $\mathbb{F}_3[X]/I$ has $9 = 3^2$ elements.

We can use the notation where we write $\overline{a} = a + I$ for $a \in \mathbb{F}_p$ and $\alpha = X + I$ for elements of $\mathbb{F}_3[X]/I$. Then the elements of $\mathbb{F}_3[X]/I$ can be written uniquely in the form $\overline{a} + \overline{b}\alpha$ for $a, b \in \mathbb{F}_3$ (we have to be careful with the notation here, as we are writing a for an element of $\mathbb{F}_3$ and $\overline{a}$ for an element of $\mathbb{F}_3[X]/I$). Also we have $\alpha^2 = \overline{-1}$, because $(X+I)^2 = X^2+I = -1+I$, as $X^2+1 \in I$. So we have that $\mathbb{F}_3[X]/I = \{\overline{a}+\overline{b}\alpha : a, b \in \mathbb{F}_3\}$, where $\alpha^2 = -1$.

We can make calculations in $\mathbb{F}_3[X]/I$ using this notation. For example, we can work out that

$$\begin{aligned}
(\overline{2} + \alpha)(\overline{1} + \overline{2}\alpha) &= \overline{2} + (\overline{4} + \overline{1})\alpha + \overline{2}\alpha^2 \\
&= \overline{3} + \overline{5}\alpha + (\overline{2})(\overline{-1}) \\
&= \overline{3 - 2} + \overline{2}\alpha \\
&= \overline{1} + \overline{\alpha}
\end{aligned}$$

and

$$\begin{aligned}
(\overline{1} + \alpha)(\overline{1} - \alpha) &= \overline{1} + (\overline{1} - \overline{1})\alpha - \overline{1}\alpha^2 \\
&= \overline{1} - (\overline{-1}) \\
&= \overline{2}.
\end{aligned}$$

We can observe that $\mathbb{F}_3[X]$ looks like "the complex numbers, but with $\mathbb{F}_3$ in place of $\mathbb{R}$".
(c) We give another example similar to that in (b) above, and we use similar notation for $\mathbb{F}_p$ as we did there, this time for the case $p = 5$.

Let $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$ and let $f(X) = X^3 + X + 1 \in \mathbb{F}_5[X]$. If $f(X)$ were reducible in $\mathbb{F}_5[X]$, then it would have a root in $\mathbb{F}_5$ because it has degree 3. However, we can calculate that $f(0) = 1$, $f(1) = 3$, $f(2) = 11 = 1$, $f(3) = 31 = 1$ and $f(4) = 69 = 4$. Therefore, $f(X)$ is irreducible, and $\mathbb{F}_5[X]/\langle f(X) \rangle$ is a field.

Let $I = \langle f(X) \rangle$. Then any element of $\mathbb{F}_5[X]/I$ can be written uniquely in the form $a + bX + cX^2 + I$ for $a, b, c \in \mathbb{F}_5$. Alternatively, using the notation $\overline{a} = a + I$ for $a \in \mathbb{F}_5$, and $\alpha = X + I$, we see that the elements of $\mathbb{F}_5[X]/I$ are written uniquely in the form $\overline{a} + \overline{b}\alpha + \overline{c}\alpha^2$. Also we have $\alpha^3 + \alpha + \overline{1} = 0$, and this allows us to calculate in $\mathbb{F}_5[X]/I$.

Once you get comfortable working with examples like this you may be comfortable to write $a + b\alpha + c\alpha^2$ for elements of the quotient ring, i.e. to omit the bars. This is an abuse of notation, but if we're careful and remember what we're doing then it is ok to do this.

2.7 The Chinese remainder theorem

We'll cover this section fairly quickly in the lectures, and it is not examinable. The material is really nice though.

In this section we're going to prove the Chinese remainder theorem for rings with one. This generalizes the Chinese remainder theorem for $\mathbb{Z}$, which was in 1AC Algebra 1.

Let's jump to it and state the Chinese remainder theorem for two ideals.

Theorem 2.38 (The Chinese remainder theorem (for two ideals))**.** *Let R be a ring with one, and let I and J be ideals of R. Suppose that $I+J+R$. Then $R/(I\cap J)\cong R/I\times R/J$.*

Before we begin the proof of this theorem, let's make sure that we understand the statement. So we have two ideals I and J, and then their intersection $I\cap J$ is also an ideal of R by Lemma 2.17. Therefore, we can form the quotient ring $R/(I\cap J)$, which is the ring on the left of the isomorphism in the theorem. Also we can form the quotient rings R/I and R/J and then take their direct product $R/I\times R/J$, which is the ring on the right of the isomorphism in the theorem.

Ok, now let's do the proof.

Proof. We define $\theta : R\to R/I\times R/J$ by $\theta(a)=(a+I,a+J)$.

We first show that θ is a homomorphism. To do this let $a,b\in R$. Then we have

$$\begin{aligned}\theta(a+b)&=(a+b+I,a+b+J)\\&=(a+I,a+I)+(b+I,b+J)\\&=\theta(a)+\theta(b),\end{aligned}$$

and similarly $\theta(ab)=\theta(a)\theta(b)$.

We are going to prove that $\ker\theta=I\cap J$ and that $\operatorname{im}\theta=R/I\times R/J$, and then we get $R/(I\cap J)\cong R/I\times R/J$ by the isomorphism theorem (Theorem 2.28).

Let $a\in R$. Then we have $a\in\ker\theta$ if and only if $(a+I,a+J)=(0+I,0+J)$. This occurs if and only if $a\in I$ and $a\in J$. Thus $a\in\ker\theta$ if and only if $a\in I\cap J$, so that $\ker\theta=I\cap J$.

Let $a,b\in R$. To show that $(a+I,b+J)\in\operatorname{im}\theta$, we have to find $x\in R$ such that $\theta(x)=(a+I,b+J)$. Since I is coprime to J, we have $I+J=R$. Therefore, we may write $1=r+s$, where $r\in I$ and $s\in J$. Then we have $r+I=0+I$ and $s+I=1+I$, and similarly $s+J=0+J$ and $r+J=1+J$. Now let $x=as+br$, then we have $x+I=a+I$, because $s+I=1+I$ and $r+I=0+I$. Similarly, $x+J=b+J$. Thus $\theta(x)=(a+I,b+J)$. Therefore, any element of $R/I\times R/J$ is in $\operatorname{im}\theta$, so that $\operatorname{im}\theta=R/I\times R/J$. □

Now we're going to look at this theorem in the case $R=\mathbb{Z}$. To do this we'll require the following results about ideals in $\mathbb{Z}$. We leave the proof of the lemma below as an exercise. Recall that $\operatorname{hcf}(a,b)$ denote the *highest common factor of a and b* and is by definition the largest $h\in\mathbb{N}$ such that $h\mid a$ and $h\mid b$. We say that a is coprime to b if $\operatorname{hcf}(a,b)=1$. Also $\operatorname{lcm}(a,b)$ denotes the *lowest common multiple of a and b* and is by definition the smallest $l\in\mathbb{N}$ such that $a\mid l$ and $b\mid l$.

Lemma 2.39. *Let $a,b\in\mathbb{N}$.*

(a) $\operatorname{lcm}(a,b)=\frac{ab}{\operatorname{hcf}(a,b)}$.
(b) $\langle a\rangle\cap\langle b\rangle=(\operatorname{lcm}(a,b))$.
(c) $\langle a\rangle+\langle b\rangle=\langle\operatorname{hcf}(a,b)\rangle$.

(d) $\langle a\rangle + \langle b\rangle = \mathbb{Z}$ *if and only if* a *is coprime to* b.

Example 2.40. In this example, we will aim to understand the Chinese remainder when $R = \mathbb{Z}$. Let $n, m \in \mathbb{N}$ and suppose that n is coprime to m, so that $\operatorname{hcf}(n, m) = 1$. Therefore, $\langle n\rangle + \langle m\rangle = \langle 1\rangle = \mathbb{Z}$ and $\langle n\rangle \cap \langle m\rangle = \langle nm\rangle$ by Lemma 2.39. Recall also that $Z_n = \mathbb{Z}/\langle n\rangle$ and $\mathbb{Z}_m = \mathbb{Z}/\langle m\rangle$.

Thus by Theorem 2.38 there is an isomorphism $\mathbb{Z}_{nm} \cong \mathbb{Z}_n \times \mathbb{Z}_m$ given by $[x]_{nm} \mapsto ([x]_n, [x]_m)$. This means that given $a, b \in \mathbb{Z}$, we can find $x \in \mathbb{Z}$ such that $[x]_n = [a]_n$ and $[x]_m = [b]_m$. Moreover if $y \in \mathbb{Z}$ satisfies $[y]_n = [a]_n$ and $[y]_m = [b]_m$, then we have $[y]_{nm} = [x]_{nm}$.

Restating what we have just written in terms of congruences, we obtain the corollary below, which is the Chinese remainder theorem for integers in the case of two congruences.

Corollary. *Let* $n, m \in \mathbb{N}$*, and let* $a, b \in \mathbb{Z}$*. Suppose that* $\operatorname{hcf}(n, m) = 1$*. Consider the pair of simultaneous congruences*

$$\begin{aligned} x &\equiv a \bmod n \\ x &\equiv b \bmod m. \end{aligned} \tag{2.1}$$

(a) *There is a solution* $x = s \in \mathbb{Z}$ *of* (2.1)*.*

(b) *Let* $r \in \mathbb{Z}$*. Then* $x = r$ *is a solution of* (2.1) *if and only if* $r \equiv s \bmod nm$*.*

Hence, the solutions of (2.1) *are given by* $x \equiv s \bmod nm$*.*

We now move on towards the general version Chinese remainder theorem. This is a bit more advanced and we'll only mention it briefly in the lectures. You may just want to aim to understand the statement and also how it could be applied, as in the example at the end.

We require the following lemma for the proof of the Chinese remainder theorem.

Lemma 2.41. *Let* R *be a ring with one and* I*,* J *and* K *ideals of* R*. Suppose that* $I + J = R$ *and* $I + K = R$*. Then* $I + (J \cap K) = R$*.*

Proof. There exist $a, b \in I$, $c \in J$ and $d \in K$ such that $1 = a + c$ and $1 = b + d$. Therefore, $1 = (a + c)(b + d) = (ab + ad + cb) + cd \in I + (J \cap K)$, because $(ab + ad + cb) \in I$ and $cd \in J \cap K$. Therefore, for any $r \in R$, we have $r = r1 \in I + (J \cap K)$. Hence, $R = I + (J \cap K)$. □

For the statement the Chinese remainder theorem below, we have to consider the intersection of an arbitrary number of ideals $I_1, I_2, \ldots, I_k$ of a commutative ring R. By repeated application of Lemma 2.17 we see that $I_1 \cap I_2 \cap \cdots \cap I_k$ is an ideal of R. Also in the statement we have the direct product $R/I_1 \times R/I_2 \times \cdots \times R/I_k$, which can be defined using Definition 1.21 recursively. The main idea in the proof of the Chinese remainder theorem is to repeatedly apply Theorem 2.38.

Theorem 2.42 (The Chinese remainder theorem)**.** *Let* R *be a ring with one and let* $I_1, I_2, \ldots, I_k$ *be ideals. Suppose that* $I_i + I_j = R$ *for each* $i \neq j$*. Then*

$$R/(I_1 \cap I_2 \cap \cdots \cap I_k) \cong R/I_1 \times R/I_2 \times \cdots \times R/I_k.$$

Proof. By repeated use of Lemma 2.41,we have that $I_k + J = R$ for $J = I_1 \cap I_2 \cap \cdots \cap I_{k-1}$. Therefore, by Theorem 2.38 we have $R/(J \cap I_k) \cong R/J \times R/I_k$.

Now we can repeat this argument to show that $R/J = R/(I_1 \cap I_2 \cap \cdots \cap I_{k-1}) \cong R/(I_1 \cap I_2 \cap \cdots \cap I_{k-2}) \times R/I_{k-1}$.

Continuing in this way we obtain $R/(I_1 \cap I_2 \cap \cdots \cap I_k) \cong R/I_1 \times R/I_2 \times \cdots \times R/I_k$, as required. □

We note that in the case $R = \mathbb{Z}$, we recover the Chinese remainder theorem for integers that we saw in 1AC Algebra 1. Well actually to do this we need to convert to a statement about congruences, but this can be done in the way we did it in Example 2.40 above.

The way that we can apply the Chinese remainder theorem in examples is slightly obscured by the inductive proof that is given here. To illuminate what is going on let's cut through the induction and see what is really going on in the proof, so that we get a better idea. We define

$$\theta : R \to R/I_1 \times R/I_2 \times \cdots \times R/I_k$$

by $\theta(a) = (a + I_1, a + I_2, \ldots, a + I_k)$. Then we can check that this is a homomorphism and that $\ker \theta = I_1 \cap I_2 \cap \cdots \cap I_k$. So we are just left to prove that θ is surjective. To do this we try to find for $i = 1, 2, \ldots, k$ some $r_i \in R$ such that $r_i + I_i = 1 + I_i$ and $r_i + I_j = 0 + I_j$ for $j \neq i$. Then given $a_1, a_2, \ldots, a_k \in R$, we see that $\theta(a_1 r_1 + a_2 r_2 + \cdots + a_k r_k) = (a_1 + I_1, a_2 + I_2, \ldots, a_k + I_k)$, and we are done. To find these r_i we write $1 = r_i + s_i$, where $s_i \in I_i$, so that $r_i + I_i = 1 + I_i$, and where $r_i \in I_j$ for all $j \neq i$, this is possible by repeated use of Lemma 2.41. This gives the desired r_i.

We illustrate this construction in the following example, where we consider the Chinese remainder theorem in the case of a polynomial ring.

Example 2.43. We consider a case of the Chinese remainder theorem for $R = \mathbb{C}[X]$.

Let $I_0 = \langle X \rangle$, $I_1 = \langle X - 1 \rangle$ and $I_{-1} = \langle X + 1 \rangle$. Then we have $1 = X - (X - 1) \in I_0 + I_1$, so $f(X) = f(X)X - f(X)(X - 1) \in I_0 + I_1$ for any $f(X) \in \mathbb{C}[X]$. Thus $I_0 + I_1 = R$. Similarly, we can show that $I_0 + I_{-1} = R$ by writing $1 = -X + (X + 1) \in I_0 \cap I_{-1}$, and that $I_1 + I_{-1} = R$ by writing $1 = -\frac{1}{2}(X - 1) + \frac{1}{2}(X + 1)$.

We have $I_0 \cap I_1 \cap I_{-1} = \langle X^3 - X \rangle$ is the principal ideal generated by $X^3 - X$. Then we can show that $R/(I_0 \cap I_1 \cap I_{-1}) = \{a + bX + cX^2 : a, b, c \in \mathbb{C}\}$. We can show that for $f(X) \in \mathbb{C}[X]$, we have that $f(X) + I_0 = f(0) + I_0$, and similarly $f(X) + I_1 = f(1) + I_1$ and $f(X) + I_{-1} = f(-1) + I_{-1}$. Therefore, the Chinese remainder theorem tells us that for any $r_0, r_1, r_{-1} \in \mathbb{C}$, we can find a quadratic polynomial $f(X) = a + bX + cX^2$ such that $f(0) = r_0$, $f(1) = r_1$ and $f(-1) = r_{-1}$.

In order to find $f(X)$ we proceed as suggested in the discussion before this example, and we define

$$g_0(X) = (X - 1)(X + 1) \quad \text{and} \quad f_0(X) = \frac{g_0(X)}{g_0(0)} = -(X - 1)(X + 1)$$

$$g_1(X) = X(X + 1) \quad \text{and} \quad f_1(X) = \frac{g_1(X)}{g_1(1)} = \frac{1}{2}X(X + 1)$$

$$g_{-1}(X) = X(X - 1) \quad \text{and} \quad f_{-1}(X) = \frac{g_{-1}(X)}{g_{-1}(-1)} = -\frac{1}{2}X(X - 1).$$

Then we can calculate that $f_0(0) = 1$, $f_0(1) = 0$ and $f_0(-1) = 0$, and similarly that $f_1(1) = 1$, $f_1(0) = 0$, $f_1(-1) = 0$, $f_{-1}(-1) = 1$, $f_{-1}(0) = 0$ and $f_{-1}(1) = 0$. Hence, we can take

$$f(X) = r_0 f_0(X) + r_1 f_1(X) + r_{-1} f_{-1}(X).$$

You may have seen something similar to this construction before.

We note that the fact that we can replace $\mathbb{C}$ in this example with any field in which $1 \neq -1$. Also you should think about how this may generalize to any number of ideals generated by linear polynomials of the form $X - a$ for $a \in \mathbb{C}$.

2.8 Correspondence of subrings and ideals

We're not going to cover this section in the lectures. and it is not part of the syllabus for 2AC, so is not examinable. However, if you read through it, then it will help with your understanding of quotient rings.

Given a ring R and an ideal I we can form the quotient ring R/I. Then it is a natural to ask if there is some relationship between the subrings in R and the subrings in R/I. We can also consider the similar question about ideals of R/I. It turns out that there is a close connection, which is stated in the next theorem, which is called the correspondence theorem. This theorem may take a bit of time to sink in, so don't worry too much if it doesn't click straightaway.

Theorem 2.44 (The correspondence theorem)**.** *Let R be a ring and I an ideal of R. There is a one to one to correspondence between the set of subrings of R that contain I and the set of subrings of R/I. This correspondence is given by sending a subring S of R to the subring S/I of R/I. Under this correspondence ideals of R containing I correspond to ideals of R/I.*

Proof. Let S be a subring of R containing I. To start with we have to understand what S/I in the statement of the theorem is. Since I is contained in S, we have that I is an ideal of S. Therefore, the quotient ring S/I is defined and $S/I = \{s + I : s \in S\}$.

We have that S/I is a subset of R/I. Next we note that S/I is a subring of R/I, because S is a ring and the addition and multiplication on S/I are the same as the addition and multiplication on R/I.

Now suppose that T is a subring of R/I, then we let $\widetilde{T} = \{t \in R : t + I \in T\}$. Then it is straightforward to check that $\widetilde{T}$ is a subring of R, for example by using the second subring test. Also we have $I \subseteq \widetilde{T}$ because $x + I = 0 + I \in T$ for all $x \in I$.

For S a subring of R containing I, we have $\widetilde{(S/I)} = \{r \in R : r + I \in S/I\}$, which is clearly equal to S. Also for T a subring of R/I, we have $\widetilde{T}/I = \{r + I : r \in R \text{ and } r + I \in T\}$, which is clearly equal to T. Therefore, we have the desired correspondence of subrings.

To check the statement about ideals, we just need to observe that if J is an ideal of R containing I, then J/I is an ideal of R/I, and that if K is an ideal of R/I, then $\widetilde{K}$ is an ideal of R/I. This is fairly straightforward to check, and left as an exercise. □

We next give a consequence of the isomorphism theorem and the correspondence theorem.

Corollary 2.45. *Let R be a ring and let I and J be ideals of R with $I \subseteq J$. Then*

$$R/J \cong (R/I)/(J/I).$$

Before giving the proof, we briefly explain what is the ring on the right hand side above. We have that R/I is a ring and that J/I is an ideal of R/I so we can form the quotient ring $(R/I)/(J/I)$.

Proof. Define $\theta : R/I \to R/J$ by $\theta(r + I) = r + J$. Then θ is a homomorphism, and we have $\ker\theta = J/I$ and $\operatorname{im}\theta = R/J$. The result now follows from the isomorphism theorem. □

We end this short section with two examples to help us to understand the correspondence theorem.

Examples 2.46. (a) In this example, we consider the case $R = \mathbb{Z}$, and $I = \langle n\rangle$ where $n \in \mathbb{N}$. Recall that we have $\mathbb{Z}/\langle n\rangle = \mathbb{Z}_n$. By the correspondence theorem, we know that the ideals of $\mathbb{Z}_n$ correspond to the ideals of $\mathbb{Z}$ containing $\langle n\rangle$. Now any ideal of $\mathbb{Z}$ is of the form $\langle m\rangle$ for some $m \in \mathbb{N}_0$, because $\mathbb{Z}$ is a principal ideal domain. Also we have $\langle n\rangle \subseteq \langle m\rangle$ if and only if m is a factor of n, by Lemma 2.18. Let $m \in \mathbb{N}$ with $m \mid n$. Then, through the correspondence theorem, we have that $\langle m\rangle$ corresponds to the ideal $\langle m\rangle/\langle n\rangle$ of $\mathbb{Z}/\langle n\rangle$. We see that this is the principal ideal of $\mathbb{Z}/\langle n\rangle$ generated by $m + \langle n\rangle$. In the notation of $\mathbb{Z}_n$ this is just the principal ideal $\langle [m]_n\rangle$ generated by $[m]_n$.

(b) Let $R = \mathbb{F}[X]$, where $\mathbb{F}$ is a field. Let $m(X) \in \mathbb{F}[X]$ and consider the quotient $\mathbb{F}[X]/\langle m(X)\rangle$. We can describe the ideals of $\mathbb{F}[X]/\langle m(X)\rangle$ using the same arguments as in (a), and they are of the form $\langle f(X)\rangle/\langle m(X)\rangle = \langle f(X) + \langle m(X)\rangle\rangle$, where $f(X) \in \mathbb{F}[X]$ is a factor of $m(X)$. In other words the ideals of $\mathbb{F}[X]/\langle m(X)\rangle$ are precisely the principal ideals generated by polynomials that are factors of $m(X)$.

2.9 Summary of Chapter 2

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

- state the definition of a homomorphism of rings, check whether examples of functions between rings are homomorphisms, and prove elementary properties about homomorphisms;
- state the definition of the kernel and image of a homomorphism, prove that they are subrings, determine them in examples, and prove that a homomorphism is injective if and only if its kernel is the zero subring;
- state the definition of an ideal, state, prove and apply the ideal test, prove that the kernel of a homomorphism is an ideal, and prove that the sum and intersection of ideals are ideals;
- explain the construction of a quotient ring and prove that this gives a well defined ring, and understand and calculate in examples of quotient rings;
- state, prove and apply the isomorphism theorem for rings;
- state the definition of a principal ideal domain, and show that $\mathbb{Z}$ and $\mathbb{F}[X]$ are principal ideal domains;
- state the definition of a maximal ideal, determine whether examples of ideals are maximal;
- state and prove the necessary and sufficient condition for an ideal of $\mathbb{Z}$ or of $\mathbb{F}[X]$ to be a maximal ideal; and
- state, prove and apply that the quotient ring by a maximal ideal is a field and state and apply its converse.

2.10 Exercises for Chapter 2

These exercises may be edited and some more exercises may be added here later. Also some of the exercises are more challenging than others, so you should ask if you have questions about any of them.

Q2.1. Which of the following functions between rings are homomorphisms.

(a) $\theta : M_2(\mathbb{R}) \to \mathbb{R}$ defined by $\theta(A) = \det A$, where $\det A$ denotes the determinant of the matrix A.
(b) $\theta : \mathbb{Z}[i] \to \mathbb{Z}$ defined by $\theta(a + bi) = a$.
(c) $\theta : \mathbb{R}[X] \to \mathbb{R}[X]$ defined by

$$\theta(a_nX^n + \cdots + a_2X^2 + a_1X + a_0) = na_nX^{n-1} + \cdots + 2a_2X + a_1.$$

You should justify your answers.

Q2.2. Recall that $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$.
Define $\theta : \mathbb{Z}[\sqrt{2}] \to M_2(\mathbb{Z})$ by $\theta(a + b\sqrt{2}) = \begin{pmatrix} a & 2b \\ b & a \end{pmatrix}$.
Prove that θ is a homomorphism.

Q2.3. Let R, S and T be rings and let $\theta : R \to S$ and $\phi : S \to T$ be homomorphisms. Prove that $\phi \circ \theta : R \to T$ is a homomorphism.

Q2.4. Let $m, n \in \mathbb{N}$ and suppose that $m \mid n$. Define $\theta : \mathbb{Z}_n \to \mathbb{Z}_m$ by $\theta([a]_n) = [a]_m$.

(a) Show that θ is well defined.
(b) Show that θ is a homomorphism.
(c) Determine $\ker \theta$.
(d) Determine $\operatorname{im} \theta$.

Q2.5. Define $\theta : \mathbb{R}[X] \to M_3(\mathbb{R})$ by

$$\theta(a_nX^n + \cdots + a_2X^2 + a_1X + a_0) = \begin{pmatrix} a_0 & a_1 & a_2 \\ 0 & a_0 & a_1 \\ 0 & 0 & a_0 \end{pmatrix}$$

.

(a) Show that θ is a homomorphism.
(b) What is $\operatorname{im} \theta$?
(c) Show that $\ker \theta = \langle X^3 \rangle$ is the principal ideal generated by X^3.

In (a) you'll need to check that $\theta(f(X)g(X)) = \theta(f(X))\theta(g(X))$ for $f(X), g(X) \in \mathbb{R}[X]$. Note that you only need to work out the coefficients of 1, X and X^2 in $f(X)g(X)$ to calculate $\theta(f(X)g(X))$

Q2.6. Let R be a commutative ring with one, let $I \trianglelefteq R$ and let $a \in I$. Suppose that a is a unit. Prove that $I = R$.

Q2.7. Let R be a commutative ring with one and suppose that the only ideals of R are $\{0\}$ and R. Prove that R is a field.

Q2.8. Let $I = \{f(X) \in \mathbb{Z}[X] \mid f(0) \text{ is even}\}$ be the ideal of $\mathbb{Z}[X]$ given in Examples 2.15(c). Prove that $I = \langle 2, X\rangle$.

Q2.9. Let R be an integral domain, let $a, b \in R$ and let $\langle a\rangle$ and $\langle b\rangle$ be the principal ideals generated by a and b. Prove that $\langle a\rangle = \langle b\rangle$ if and only if there exists a unit $u \in R$ such that $a = ub$.

Q2.10. Let $a, b \in \mathbb{N}$. Let $\langle a\rangle = a\mathbb{Z}$ and $\langle b\rangle = b\mathbb{Z}$ be the principal ideals generated by a and b.

(a) Investigate the ideal $\langle a\rangle + \langle b\rangle$.
(b) Investigate the ideal $\langle a\rangle \cap \langle b\rangle$.

Pick some small values of a and b and try to work out $\langle a\rangle + \langle b\rangle$ and $\langle a\rangle \cap \langle b\rangle$ for these cases. Once you have done a few, you should be able to make conjectures about what $\langle a\rangle + \langle b\rangle$ and $\langle a\rangle \cap \langle b\rangle$ are in general. Then you can try to prove your conjectures.

Q2.11. Let $R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$ (you are not required to prove that R is a ring) and $I = \{2a + (1 + \sqrt{-5})b : a, b \in \mathbb{Z}\} \subseteq R$.

(a) Prove that I is an ideal of R.
(b) Prove that $I = \langle 2, 1 + \sqrt{-5}\rangle$ is the ideal generated by 2 and $1 + \sqrt{-5}$.
(c) Prove that I is not equal to the principal ideal $\langle x\rangle$ for any $x \in R$.

Hint: For (a) you should use the ideal test.
The only tricky part is to verify the third condition in the ideal test.
For this let $x = 2a + (1 + \sqrt{-5})b \in I$ and $y = c + d\sqrt{-5} \in R$, where $a, b, c, d \in \mathbb{Z}$. Then you have to calculate xy and write it in the form $m + n(1 + \sqrt{-5})$, where $m, n \in \mathbb{Z}$, and check that m is even.
Part (b) is difficult so you may want to ask for a hint for this.

Q2.12. Let R be a ring, I an ideal of R and $a, b \in R$.
By definition we have $a + I = \{a + x : x \in I\}$, $b + I = \{b + x : x \in I\}$,
$a + b + I = \{a + b + x : x \in I\}$ and $ab + I = \{ab + x : x \in I\}$.

(a) Prove that $a + b + I = \{c + d : c \in a + I, d \in b + I\}$.
(b) Prove that $\{cd : c \in a + I, d \in b + I\} \subseteq ab + I$.
(c) Give an example of R, I, a and b such that $\{cd : c \in a + I, d \in b + I\}$ is a proper subset of $ab + I$

Q2.13. Let $R = \mathbb{Z}[i]$ and $S = \mathbb{Z}$.
For $r \in R$, define $r + S = \{r + s : s \in S\}$.

(a) Let $a = 1$ and $b = 0$. Show that $a + S = b + S$.
(b) Let $c = i$. Show that $ac + S \neq bc + S$.

Q2.14. Let R be a ring and I an ideal of R. Prove that the axiom (Dl) holds in the quotient ring R/I.

Q2.15. Let $R = \mathbb{Z}[\sqrt{2}]$ and let $I = \langle 2\sqrt{2} \rangle$ be the principal ideal generated by $2\sqrt{2}$.

(a) By calculating $2\sqrt{2}(y + x\sqrt{2})$ show that elements of I are of the form $4x + 2\sqrt{2}y$ for some $x, y \in \mathbb{Z}$, and deduce that $I = \{4x + 2\sqrt{2}y : x, y \in \mathbb{Z}\}$.
(b) Let $a + b\sqrt{2} \in R$. Show that there exist $c \in \{0, 1, 2, 3\}$ and $d \in \{0, 1\}$ such that $a + b\sqrt{2} + I = c + d\sqrt{2} + I$.
(c) Let $a, c \in \{0, 1, 2, 3\}$ and $b, d \in \{0, 1\}$ and suppose that $a + b\sqrt{2} + I = c + d\sqrt{2} + I$. Prove that $a = c$ and $b = d$.
(d) Deduce that R/I has exactly 8 elements.
(e) Calculate $(3 + \sqrt{2} + I)(2 + \sqrt{2} + I)$ writing your answer in the form $a + b\sqrt{2} + I$, where $a \in \{0, 1, 2, 3\}$ and $b \in \{0, 1\}$.
(f) Define $\theta : \mathbb{Z}[X] \to R/I$ by $\theta(f(X)) = f(\sqrt{2}) + I$.
Show that $\ker\theta = \{(X^2 - 2)g(X) + 2aX + 4b : g(X) \in \mathbb{Z}[X] \text{ and } a, b \in \mathbb{Z}\}$.

Hint: For (b) use the division theorem to write $a = 4x + c$ *and* $b = 2y + d$, *where* $x, y, c, d \in \mathbb{Z}$ *with* $0 \leq c < 4$ *and* $0 \leq d < 2$.

Q2.16. Let R and S be rings, and let $R \times S$ be the direct product of R and S.
(a) Let $I = \{(r, 0) : r \in R\}$. Define $\theta : R \times S \to S$ by $\theta(r, s) = s$.

(i) Prove that I is an ideal of R.
(ii) Show that θ is a homomorphism.
(iii) Show that $\ker\theta = I$.
(iv) Show that $\operatorname{im}\theta = S$.
(v) Deduce that $(R \times S)/I \cong S$.

(b) Now suppose that $S = R$.
Let $T = \{(r, r) : r \in R\}$. Define $\theta : R \to R \times R$ by $\theta(r) = (r, r)$.

(i) Show that T is a subring of $R \times R$.
(ii) Is T an ideal of $R \times R$? Justify your answer.
(iii) Show that θ is a homomorphism.
(iv) Show that $\operatorname{im}\theta = T$.
(v) Show that $\ker\theta = \{0\}$.
(vi) Deduce that $R \cong T$.

Q2.17. (a) Let $I = \{f(X) = \mathbb{C}[X] : f(0) = f(1) = f(-1) = 0\}$. Then I is an ideal of $\mathbb{C}[X]$ by Examples 2.13(c).

(i) Show that $f(X) \in I$ if and only if $X(X-1)(X+1) \mid f(X)$.
(ii) Deduce that $I = \langle X^3 - X\rangle$ is the principal ideal generated by $X^3 - X$.
(iii) Show that all elements of $\mathbb{C}[X]/I$ can be written in the form $aX^2 + bX + c + I$, where $a, b, c \in \mathbb{C}$.
(iv) For $a_2, a_1, a_0, b_2, b_1, b_0 \in \mathbb{C}$, show that $a_2X^2 + a_1X + a_0 + I = b_2X^2 + b_1X + b_0 + I$ if and only if $a_2 = b_2$, $a_1 = b_1$ and $a_0 = b_0$.

(b) Define $\theta : \mathbb{C}[X] \to \mathbb{C}^3$ by $\theta(f(X)) = (f(0), f(1), f(-1))$.
Here $\mathbb{C}^3 = \mathbb{C} \times \mathbb{C} \times C$ is the direct product of three copies of $\mathbb{C}$.

(i) Show that θ is a homomorphism.
You do not need to write a lot for this.
(ii) Show that $\ker\theta = I$.
(iii) Show that $\operatorname{im}\theta = \mathbb{C}^3$.
(iv) Deduce that $\mathbb{C}[X]/I \cong \mathbb{C}^3$.

Q2.18. Let R be a commutative ring with one and I an ideal of R. Suppose that R/I is a field. Prove that I is a maximal ideal.

Hint: Let J be an ideal of R properly containing I, and let $a \in J$ with $a \notin I$, so that $a + I \neq 0 + I$. Now use the fact that R/I is a field to find an element $b \in R$ such that $ab + I = 1 + I$. Use this to deduce that $1 \in J$, and then that $J = R$.

Q2.19. Let $f(X) = X^2 - 3 \in \mathbb{Q}[X]$, let $I = \langle f(X))\rangle$ be the principal ideal in $\mathbb{Q}[X]$ generated by $f(X)$, and let $\mathbb{F} = \mathbb{Q}[\sqrt{3}] = \{a + b\sqrt{3} : a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$.

(a) Explain why $f(X)$ is irreducible in $\mathbb{Q}[X]$.
(b) Prove that $\mathbb{Q}[X]/I \cong \mathbb{F}$.
(c) Deduce that $\mathbb{F}$ is a field.

In the next few questions we work in $\mathbb{F}_p = \mathbb{Z}_p$ for some prime $p \in \mathbb{N}$. We use the notation, where we just write a instead of $[a]_p$ for an element of $\mathbb{F}_p$. This could possibly be quite confusing, but usually it doesn't cause any problems.

Q2.20. Let $I = \langle X^3 + 2 \rangle$ be the principal ideal of $\mathbb{F}_7[X]$ generated by $X^3 + 2$.

(a) Show that $X^3 + 2$ is irreducible in $\mathbb{F}_7[X]$.
(b) Deduce that the quotient $\mathbb{F}_7[X]/I$ is a field.
(c) Use the notation $\overline{a} = a + I$ for $a \in \mathbb{F}_7$ and $\alpha = X + I$. Show that any element of $\mathbb{F}_7[X]/I$ can be written uniquely in the form $\overline{a} + \overline{b}\alpha + \overline{c}\alpha^2$ for $a, b, c \in \mathbb{F}_7$, and that $\alpha^3 = \overline{5}$.
As we've done in some of the examples, you can abuse notation and omit the bar and write a rather than $\overline{a}$ – you should be careful if you do this, and make sure you understand that this is an abuse of notation.

Q2.21. Define $\theta : \mathbb{Z}[\sqrt{2}] \to \mathbb{F}_7$ by $\theta(a + b\sqrt{2}) = a + 3b$.
Let $I = \langle 3 - \sqrt{2} \rangle$ be the principal ideal generated by $3 - \sqrt{2}$.

(a) Show that θ is a homomorphism.
(b) Show that $\operatorname{im} \theta = \mathbb{F}_7$ and deduce that $\mathbb{Z}[\sqrt{2}]/\ker \theta \cong \mathbb{F}_7$
(c) Show that $\ker \theta = I$.
(d) Deduce that $\mathbb{Z}[\sqrt{2}]/I \cong \mathbb{F}_7$ and that I is a maximal ideal of $\mathbb{Z}[\sqrt{2}]$.

Hint: For (c) first show that $3 - \sqrt{2} \in \ker \theta$. Next show that any element of $\mathbb{Z}[\sqrt{2}]/I$ can be written in the form $a + I$, where $a \in \{0, 1, 2, 3, 4, 5, 6\}$ so that $\mathbb{Z}[\sqrt{2}]/I$ has 7 elements; it may be useful to observe that $(3-\sqrt{2})(3+\sqrt{2}) = 7$ for this. Then deduce that $\ker \theta = I$.

Q2.22. Consider the polynomial $f(X) = X^2 + X + 1 \in \mathbb{F}_5[X]$, and let $I = \langle f(X) \rangle$ be the principal ideal of $\mathbb{F}_5[X]$ generated by $f(X)$.

(a) Show that $f(X)$ is irreducible.
(b) Deduce that $\mathbb{F}_5[X]/I$ is a field.
(c) Write $\mathbb{F}$ for $\mathbb{F}_5[X]/I$.
You may assume that the elements of $\mathbb{F}$ can be written uniquely in the form $a+bX+I$ where $a, b \in \mathbb{F}_5$ (this can be proved similarly to how we have done similar examples in lectures).
Use the notation $\overline{a} + \overline{b}\alpha$ for $a + bX + I$, as we have done in lectures for other examples (or you could just write $a + b\alpha$ for $a + bX + I$ if you comfortable to do so, but remember that we should be careful with this notation).
   (i) How many elements does $\mathbb{F}$ have?
   (ii) Show that $\alpha^2 + \alpha + \overline{1} = \overline{0}$ in $\mathbb{F}$.
   (iii) Calculate $(\overline{2} + \overline{3}\alpha)(\overline{1} + \overline{4}\alpha)$ in $\mathbb{F}$ writing your answer in the form $\overline{a} + \overline{b}\alpha$ with $a, b \in \mathbb{F}_5$.
   (iv) Find $x \in \mathbb{F}$ such that $x^2 = \overline{2}$.

Q2.23. Let $\mathbb{F}_9 = \mathbb{F}_3[X]/\langle X^2+1\rangle$ as considered in Examples 2.37, and write the elements of $\mathbb{F}_9$ in the form $a + b\alpha$, where $a, b \in \mathbb{F}_3$ and α satisfies $\alpha^2 = -1$.
Let $g(X) = X^2 + cX + d \in \mathbb{F}_3[X]$ be a quadratic polynomial.
Show that $g(X)$ has roots in $\mathbb{F}_9$ and give a formula for them.

You should proceed by trying to find the roots of $g(X)$ in the same way as you would do for polynomials with coefficients in $\mathbb{C}$. You can do this by completing the square, but you'll have to think about how to do all of this over $\mathbb{F}_3$.

Q2.24. Consider the polynomials $f_1(X) = X^3 + X + 1$ and $f_2(X) = X^3 + X^2 + 1$ in $\mathbb{F}_2[X]$. Let $I_1 = \langle f_1(X)\rangle$ and $I_2 = \langle f_2(X)\rangle$.

(a) Show that $f_1(X)$ and $f_2(X)$ are irreducible in $\mathbb{F}_2[X]$.
(b) Deduce that $\mathbb{F}_2[X]/I_1$ and $\mathbb{F}_2[X]/I_2$ are fields.
(c) Prove that $\mathbb{F}_2[X]/I_1$ is isomorphic to $\mathbb{F}_2[X]/I_2$.

Q2.25. Let $D(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is differentiable}\}$ be the set of functions from $\mathbb{R}$ to $\mathbb{R}$ which are differentiable at all points in $\mathbb{R}$. Then $D(\mathbb{R})$ is a ring with addition and multiplication defined pointwise, that is $(f + g)(a) = f(a) + g(a)$ and $(fg)(a) = f(a)g(a)$ for $f, g \in D(\mathbb{R})$ and $a \in \mathbb{R}$. Then $D(\mathbb{R})$ is a commutative ring – you do not have to check this, but you should think about it and convince yourself that it is true.

Define $\theta : \mathbb{R}[X] \to D(\mathbb{R})$ by $\theta(p(X))(a) = p(a)$ for $p(X) \in \mathbb{R}[X]$ and $a \in \mathbb{R}$, so $\theta(p(X)$ is the function defined by evaluating $p(X)$. This may seem a bit confusing at first: we are viewing polynomials as expressions in X not as functions and θ maps a polynomial $p(X)$ to the function $a \mapsto p(a)$ that it defines.

Let $I = \{f \in D(\mathbb{R}) : f(0) = 0 \text{ and } f'(0) = 0\}$, here f' denotes the derivative of f.

(a) Show that θ is a homomorphism.
(b) Prove that I is an ideal of $D(\mathbb{R})$.
(c) Let $\phi : \mathbb{R}[X] \to D(\mathbb{R})/I$ be the homomorphism defined by $\phi(p(X)) = \theta(p(X)) + I$.
  (i) Prove that ϕ is surjective.
  (ii) Show that $\ker\phi = \langle X^2\rangle$.
  (iii) Deduce that $R[X]/\langle X^2\rangle \cong D(\mathbb{R})/I$.

Q2.26. Let $\mathbb{F}$ be a field, let $R = \mathbb{F}[X, Y]$ be the polynomial ring in two indeterminates X and Y and let $S = \mathbb{F}[X]$.

Define $\theta : R \to S$ by $\theta(f(X, Y)) = f(X, X)$, so $\theta(f(X, Y))$ is obtained from $f(X, Y)$ by substituting $Y = X$, for example $\theta(Y^2+2XY-Y+3X+1) = X^2+2X^2-X+3X+1 = 3X^2 + 2X + 1$.

Let $I = \langle X - Y)\rangle$ be the principal ideal of $\mathbb{F}[X, Y]$ generated by $X - Y \in \mathbb{F}[X, Y]$.

(a) Show that θ is a homomorphism.
(b) Show that $I \subseteq \ker\theta$.
(c) For $f(X, Y) \in R$, show that $f(X, Y) + I = f(X, X) + I$.

(d) Deduce that $\ker\theta = I$.
(e) Deduce that $R/I \cong S$.
(f) Give an isomorphism $\phi : S \to R/I$.

For (c) first show that $Y + I = X + I$. Then show that $Y^n + I = X^n + I$ for any $n \in \mathbb{N}$ by factorizing $X^n - Y^n$ as $X - Y$ multiplied by something else. Then deduce that $X^mY^n + I = X^{m+n} + I$ for any $m, n \in \mathbb{N}_0$ and finally $f(X,Y) + I = f(X,X) + I$ for any $f(X,Y) \in \mathbb{F}[X,Y]$.

Some exercises on the Chinese remainder theorem

Q2.27. Let $a, b \in \mathbb{N}$. Recall that $\text{hcf}(a,b)$ denotes the highest common factor of a and b and is by definition the largest $h \in \mathbb{N}$ such that $h \mid a$ and $h \mid b$. Also $\text{lcm}(a,b)$ denotes the lowest common multiple of a and b and is by definition the smallest $l \in \mathbb{N}$ such that $a \mid l$ and $b \mid l$.

(a) Prove that $\text{lcm}(a,b) = \frac{ab}{\text{hcf}(a,b)}$.
(b) Prove that $\langle a\rangle \cap \langle b\rangle = \langle \text{lcm}(a,b)\rangle$
(c) Prove that $\langle a\rangle + \langle b\rangle = \langle \text{hcf}(a,b)\rangle$.

Q2.28. Find $a, b \in \mathbb{Z}$ such that the pair of simultaneous congruences

$$\begin{aligned} x &\equiv a \bmod 4 \\ x &\equiv b \bmod 6 \end{aligned}$$

does not have a solution.
Explain why this is not a counterexample to the Chinese remainder theorem.

Q2.29. Find $f(X) \in \mathbb{C}[X]$ such that $f(-1) = 2$, $f(i) = i$ and $f(-i) = 1$.

Some exercises on the correspondence of subrings and ideals

Q2.30. Let R be a ring and let I be an ideal of R.

(a) Let J be an ideal of R containing I. Show that J/I is an ideal of R/I.
(b) Let K be an ideal of R/I. Prove that $\tilde{K} = \{x \in R : x + I \in K\}$ is an ideal of R.

Q2.31. Consider $\mathbb{Z}_8 = \mathbb{Z}/8\mathbb{Z}$.

(a) Determine all ideals of $\mathbb{Z}/8\mathbb{Z}$.
(b) For each ideal K of $\mathbb{Z}/8\mathbb{Z}$ find an ideal J of $\mathbb{Z}$ such that $J/I = K$.

Q 2.32. Let R be a ring and let I and J be ideals of R with $I \subseteq J$. Prove that $(R/I)/(J/I) \cong R/J$.
Note that we use the correspondence of ideals here to view J/I as an ideal of R/I.

Chapter 3

Fields

In this chapter, we look at some aspects of the theory of fields. This chapter is much shorter than the previous chapter and we just look at four important topics: fields of fractions; the characteristic of a field; field extensions; and finite fields.

This part of the notes has been written recently, so has less discussion than other parts of the notes and some bits could possibly be explained a little more. Also there are likely to be some typos in this chapter, and the formatting may look a little messy in places. It is possible that we will diverge slightly from these notes in the lectures, but will follow them pretty closely. If it turns out that lots of changes are made to the notes for this chapter, then they may be reprinted, so you can get a new copy.

Please let me know if you spot any typos or errors, and if there are any parts of the notes that you think could be improved.

Before we begin this section we define one term that we will use frequently. Let $\mathbb{F}$ and $\mathbb{K}$ be fields and suppose that $\mathbb{F}$ is a subset of $\mathbb{K}$. Then we say that $\mathbb{F}$ is a subfield of $\mathbb{K}$. So we see that a subfield of $\mathbb{K}$, which is also a field.

3.1 The field of fractions of an integral domain

We explain how to construct the field of fractions of an integral domain R. By definition this field is the smallest field containing R as a subring, and is obtained via an easy but slightly technical construction. The example that you should keep in mind is how we can construct the rational numbers $\mathbb{Q}$ from the integers $\mathbb{Z}$.

Before giving the formal construction of the field of fractions, we give an idea of what we aim to achieve. The field of fractions $Q(R)$ of an integral domain should consist of symbols of the form $\frac{a}{b}$, where $a, b \in R$ and $b \neq 0$. Then we should add and multiply these by the familiar rules: $\frac{a}{b} + \frac{c}{d} = \frac{ad+cb}{bd}$ and $(\frac{a}{b})(\frac{c}{d}) = \frac{ac}{bd}$. The first issue that we have to deal with is that we need to be more precise about symbols of the form $\frac{a}{b}$. Then we have the additional caveat that we want to say $\frac{a}{b} = \frac{c}{d}$ if $ad = bc$. We'll see how to get around this below where we see that using an equivalence relation on pairs of elements of R is the way forward.

Now we move on to the formal construction. Let R be an integral domain. We let $Q_0(R) = \{(a, b) \in R \times R : b \neq 0\}$. We define a relation $\sim$ on $Q_0(R)$ by $(a, b) \sim (c, d)$ means $ad = cb$. In the next lemma we show that $\sim$ is an equivalence relation, which is fairly straightforward, but there is no harm in including the details.

Lemma 3.1. *$\sim$ is an equivalence relation on $Q_0(R)$.*

Proof. We need to check conditions (ER1), (ER2) and (ER3) for being an equivalence relation from Definition 0.2. To do this, let $a, b, c, d, e, f \in R$ with $b, d, f \neq 0$.

(ER1) We have $ab = ab$, so $(a, b) \sim (a, b)$.
(ER2) Suppose that $(a, b) \sim (c, d)$, so that $ad = cb$. Then $cb = ad$, so that $(c, d) \sim (a, b)$.
(ER3) Suppose that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, so that $ad = cb$ and $cf = ed$. Then we have that $(af)d = (ad)f = (cb)f = b(cf) = b(ed) = (be)d$, and thus $af = be$ because $d \neq 0$ and R is an integral domain. Hence, $(a, b) \sim (e, f)$. □

Recall that $Q_0(R)/\sim\ = \{[(a, b)]_\sim : (a, b) \in Q_0(R)\}$ denotes the set of equivalence classes of $Q_0(R)$ under $\sim$. In the next definition we give an addition and multiplication on this set of equivalence classes, which gives the field of fractions of R.

Definition 3.2. We let $Q(R) = Q_0(R)/\sim$ and define addition and multiplication on $Q(R)$ as follows.
Let $a, b, c, d \in R$ with $b, d \neq 0$. We define

$$[(a, b)]_\sim + [(c, d)]_\sim = [(ad + bc, bd)]_\sim$$

and

$$[(a, b)]_\sim \cdot [(c, d)]_\sim = [(ac, bd)]_\sim.$$

We refer to $Q(R)$ with this addition and multiplication as *the field of fractions of R*.
Usually we use notation, where we write $\frac{a}{b}$ for $[(a, b)]_\sim \in Q(R)$

It is suggested in the definition above that $Q(R)$ is a field, which is indeed true, and we prove this below in the main theorem of this section. In the proof we frequently use that:
$[(a, b)]_\sim = [(c, d)]_\sim$ if and only if $(a, b) \sim (c, d)$ if and only $ad = cb$.

Theorem 3.3. *Let R be an integral domain. Then $Q(R)$ with the addition and multiplication defined in Definition 3.2 is a field. Moreover, the function $\iota : R \to Q(R)$ defined by $\iota(a) = [(a, 1)]_\sim$ is an injective homomorphism.*

Proof. There is quite a lot to check here.

First we need to check that the addition and multiplication on $Q(R)$ are well defined. We'll just check that addition is well defined, and leave checking the multiplication is well defined as an exercise. Let $a, b, a', b', c, d, c', d' \in R$ with $b, b',, d, d' \neq 0$ and suppose that $[(a, b)]_\sim = [(a', b')]_\sim$ and that $[(c, d)]_\sim = [(c', d')]_\sim$. Then we have $ab' = a'b$ and $cd' = c'd$. We want to check that $[(ad + bc, bd)]_\sim = [(a'd' + b'c', b'd')]_\sim$ so that addition on $Q(R)$ is well-defined. Well we have

$$\begin{aligned}(ad + bc)b'd' &= ab'dd' + bb'cd' \\ &= a'bdd' + bb'c'd \\ &= (a'd' + b'c')bd,\end{aligned}$$

which shows that $(ad+bc, bd) \sim (a'd'+b'c', b'd')$ so that $[(ad+bc, bd)]_\sim = [(a'd'+b'c', b'd')]_\sim$ as required.

Next we need to check that the axioms of a field hold. We only include some of the details.

Let $a, b, c, d \in R$ with $b, d \neq 0$.

Checking (A0) and (M0) requires us to note that $bd \neq 0$ so that $[(a,b)]_\sim + [(c,d)]_\sim = [(ad+bc, bd)]_\sim \in Q(R)$ and $[(a,b)]_\sim \cdot [(c,d)]_\sim = [(ac,bd)]_\sim \in Q(R)$.

We can check the axioms (A1)–(A4) about addition hold by deducing them from the corresponding axioms for R. For example, to prove (A4) we note that $ad + bc = bc + ad$, so that $[(a,b)]_\sim + [(c,d)]_\sim = [(c,d)]_\sim + [(a,b)]_\sim$. We also note that $[(0,1)]_\sim$ is the zero in $Q(R)$ and that $-[(a,b)]_\sim = [(-a,b)]_\sim$.

We leave checking axioms (M1), (M2), (M4) and (D) as exercises, we just note that the one of $Q(R)$ is $[(1,1)]_\sim$. Thus we have that $Q(R)$ is a commutative ring with one.

This leaves us to verify that (M3) holds, so that $Q(R)$ is indeed a field. Let $[(a,b)]_\sim$ be a nonzero element of $Q(R)$. Then we must have $a \neq 0$ because $[(0,b)]_\sim = [(0,1)]_\sim$ is the zero of $Q(R)$. Now we can check that the multiplicative inverse of $[(a,b)]_\sim$ is $[(b,a)]_\sim$, because $[(a,b)]_\sim \cdot [(b,a)]_\sim = [(ab,ba)]_\sim = [(1,1)]_\sim$.

Finally, we have to check that ι is indeed an injective homomorphism. It is clear from the definition of addition and multiplication in $Q(R)$ that ι is a homomorphism. To show that it is injective, we just have to prove that $\ker \iota = \{0\}$. Let $a \in \ker \iota$. Then we have $[(a,1)]_\sim = [(0,1)]_\sim$, so that $a = a1 = 0 \cdot 1 = 0$. Thus $\ker \iota = \{0\}$ as required. □

We can use the injective homomorphism ι to identify R as a subring of $Q(R)$, and we will do this. By this we mean that we really think of $a \in R$ as being the same as $\frac{a}{1} = [(a,1)]_\sim$, even though strictly they are not equal.

Before giving some examples, we state the following useful lemma, which helps us to say what the field of fractions of an integral domain is, when it lies in a known field. We omit the proof, which is not part of the syllabus, and just note that it involves making a few checks.

Lemma 3.4. *Let R be an integral domain and let $\mathbb{F}$ be a field. Suppose that $R \subseteq \mathbb{F}$. Then the function $\theta : Q(R) \to \mathbb{F}$ defined by $\theta([(a,b)]_\sim) = ab^{-1}$ is a well defined homomorphism. Moreover, θ is injective so $Q(R) \cong \operatorname{im}\theta$ is a subfield of $\mathbb{F}$.*

In the situation of Lemma 3.4, we will always identify $Q(R)$ with a subfield of $\mathbb{F}$ through the given isomorphism. We quickly note that this ensure that there is no possible ambiguity in using the notation $\frac{a}{b}$ for $[(a,b)]_\sim \in Q(R)$, as we have $\frac{a}{b} = ab^{-1}$.

What we mean when we say we identify $Q(R)$ with a subfield of $\mathbb{F}$ may not be clear to start with. All we're saying here is that as these two fields are isomorphic we're really going to think of them as being the same even though they are not exactly the same. This may sound a bit confusing to start with, but it is convenient to do so and shouldn't cause you any problems. Please ask if you're not sure about this.

Examples 3.5. (a) We have $Q(\mathbb{Z}) = \mathbb{Q}$. The construction above corresponds exactly to how we define $\mathbb{Q}$ from $\mathbb{Z}$, so this is what we should take as our definition of the rational numbers.

(b) We have $Q(\mathbb{Z}[i]) = \{a + bi \in \mathbb{C} : a, b \in \mathbb{Q}\}$, where we're using Lemma 3.4 to identify $Q(\mathbb{Z}[i])$ with a subfield of $\mathbb{C}$. We leave checking this as an exercise.

(c) Let $\mathbb{F}$ be a field. Then using Lemma 3.4, we can identify $Q(\mathbb{F})$ with a subfield of $\mathbb{F}$ and in fact we have $Q(\mathbb{F}) = \mathbb{F}$. You should think about this for a while, and convince yourself that it is true.

(d) Let $\mathbb{F}$ be a field. Then the polynomial ring $\mathbb{F}[X]$ is also an integral domain. We have $Q(\mathbb{F}[X])$ consists of so called *rational functions* over $\mathbb{F}$, which by definition are expressions of the form $\frac{f(X)}{g(X)}$, where $f(X), g(X) \in \mathbb{F}[X]$ and $g(X)$ is not the zero polynomial.

3.2 The characteristic of a field

In this short section we consider the characteristic of a field. We start with the definition and then move on to show that the characteristic is always zero or a prime number. We move on to determine the characteristic of some fields, and then end by stating a theorem about the smallest subfield of a field. We note that most of what we do at the start of this section applies more generally to integral domains, but we won't worry about that here.

For the definition of characteristic we recall that for $m \in \mathbb{N}$ and a an element of a field, $ma \in R$ is defined to be $ma = a + a + \cdots + a$, where there are m summands; for use later in this section we recall also that, by definition, $(-m)a = -ma$ and $0a = 0$.

Definition 3.6. Let $\mathbb{F}$ be a field. The *characteristic of* $\mathbb{F}$ is the smallest $m \in \mathbb{N}$ such that $m1 = 0$ in $\mathbb{F}$, if such m exists, and is equal to 0 if $m1 \neq 0$ for all $m \in \mathbb{N}$. The characteristic of $\mathbb{F}$ is denoted by $\operatorname{char} \mathbb{F}$.

We make the following observation before looking further at the characteristic of a field.
Let $\mathbb{F}$ be a field, $a, b \in \mathbb{F}$ and $m, n \in \mathbb{N}$. Then we have $(ma)(nb) = (mn)(ab)$.
It is a straightforward exercise to prove this using the distributive law, though it is a little technical and it is best done by induction. This is what is needed for the proof of the next proposition.

Proposition 3.7. *Let $\mathbb{F}$ be a field. Then $\operatorname{char} \mathbb{F} = 0$ or $\operatorname{char} \mathbb{F} = p$, where $p \in \mathbb{N}$ is prime.*

Proof. If there is no $m \in \mathbb{N}$ such that $m1 = 0$, then $\operatorname{char} \mathbb{F} = 0$. So we assume that there exists $m \in \mathbb{N}$ such that $m1 = 0$.

Let $m, n \in \mathbb{N}$ and suppose that $(mn)1 = 0$. We have that $(mn)1 = (m1)(n1)$, so as $\mathbb{F}$ is a field we have that $m1 = 0$ or $n1 = 0$. Thus the smallest $m \in \mathbb{N}$ such that $m1 = 0$ must be prime. □

Let's make sure that we have understood the definition of the characteristic of a field with some examples.

Examples 3.8. (a) The characteristic of $\mathbb{Q}$ is 0, because for $m \in \mathbb{N}$, we have $m1 = m \neq 0$. Similarly, we have that $\operatorname{char} \mathbb{R} = 0$ and $\operatorname{char} \mathbb{C} = 0$.
In fact any subfield of $\mathbb{C}$ has characteristic 0.

(b) Let p be a prime. Recall that we write $\mathbb{F}_p$ for $\mathbb{Z}_p$ to honour the fact that it is a field, and we write the elements of $\mathbb{F}_p$ simply as $\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$. In $\mathbb{F}_p$ we have that $p1 = 0$, so that $\operatorname{char} \mathbb{F}_p = p$. In the notation of congruence classes this corresponds to saying $[p]_p = [0]_p$.

We end this section by stating a theorem saying that the smallest subfield of any field must be either isomorphic to $\mathbb{Q}$ or $\mathbb{F}_p$. (We just break off to say what we mean by a subfield here, as this may not have been defined earlier. Given a field $\mathbb{F}$, we say that $\mathbb{E} \subseteq \mathbb{F}$ is a *subfield* if it is a subring that is a field.)

We explain what this smallest subfield is before the statement of Theorem 3.9. We can consider the subring S of $\mathbb{F}$ consisting of elements of the form $m1$, where $m \in \mathbb{Z}$; we can check that S is a subring using one of the subring tests. We have that S is an integral domain since $1 \in S$, and S is contained in $\mathbb{F}$ so has no zero divisors. Then we let $\mathbb{E} = Q(S)$ to be the field of fractions of S, so that $\mathbb{E} = \{\frac{m1}{n1} : m, n \in \mathbb{Z}, n1 \neq 0\}$. We refer to $\mathbb{E}$ as the *prime subfield* of $\mathbb{F}$.

The proof of Theorem 3.9 is only sketched, so we just give the main steps but do not include all the details. It is not particularly difficult, but there is quite a bit to check, and is not part of the syllabus.

Theorem 3.9. *Let $\mathbb{F}$ be a field and let $\mathbb{E}$ be the prime subfield of $\mathbb{F}$. Then $\mathbb{E}$ is the smallest subfield of $\mathbb{F}$. Moreover,*

(a) *if* $\operatorname{char} \mathbb{F} = 0$, *then* $\mathbb{E} \cong \mathbb{Q}$.

(b) *if* $\operatorname{char} \mathbb{F} = p$, *where* $p \in \mathbb{N}$ *is a prime, then* $\mathbb{E} \cong \mathbb{F}_p$.

Sketch proof. Let S be as above, so that $\mathbb{E} = Q(S)$. Let $p = \operatorname{char} \mathbb{F}$, so that $p = 0$ or p is a prime number.

Let $\mathbb{F}'$ be a subfield of $\mathbb{F}$. Then we must have $1 \in \mathbb{F}'$, and using the fact that $\mathbb{F}'$ is closed under addition and subtraction we deduce that $S \subseteq \mathbb{F}'$. Since $\mathbb{F}'$ is a field it is also closed under taking multiplicative inverses, so that $\mathbb{E} = Q(S) \subseteq \mathbb{F}'$. Hence, $\mathbb{E}$ is the smallest subfield of $\mathbb{F}$.

Consider the homomorphism $\theta : \mathbb{Z} \to \mathbb{F}$ defined by $\theta(m) = m1$. The image of θ is S, and the kernel is the principal ideal $\langle p \rangle$ in $\mathbb{Z}$ generated by p. By the isomorphism theorem we have that $\mathbb{Z}/\langle p \rangle \cong S$.

If $p = 0$, then we get $E = Q(S) \cong \mathbb{Q}$, whereas if p is a prime number then we have $S \cong \mathbb{Z}/\langle p \rangle = \mathbb{F}_p$, so that $Q(S) \cong Q(\mathbb{F}_p) = \mathbb{F}_p$. □

3.3 Extension fields and minimal polynomials

We're going to study the theory of field extensions, and start with the definition of what we mean by a field extension.

Definition 3.10. Let $\mathbb{F}$ and $\mathbb{K}$ be fields with $\mathbb{F} \subseteq \mathbb{K}$. We say that $\mathbb{K}$ is *extension field* of $\mathbb{F}$, or that $\mathbb{F} \subseteq \mathbb{K}$ is a *field extension*.

Before we move on we list some familiar examples of field extensions.

- $\mathbb{Q} \subseteq \mathbb{R}$.
- $\mathbb{R} \subseteq \mathbb{C}$.
- $\mathbb{Q} \subseteq \mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$, though we should strictly check that $\mathbb{Q}(i)$ is indeed a field for this example.
- $\mathbb{Q} \subseteq \mathbb{Q}(\alpha) = \{a + b\alpha + c\alpha^2 : a, b, c \in \mathbb{Q}\}$, where $\alpha = \sqrt[3]{2}$. We saw that $\mathbb{Q}(\alpha)$ is field in Examples 2.37(a).

All of these examples are of subfields of $\mathbb{C}$ and therefore fields of characteristic 0. In this section we will just consider examples that are fields of characteristic 0, but in the next section we will see how the theory applies in the case of fields of positive characteristic.

You may notice that we used round brackets for $\mathbb{Q}(i)$ above as opposed to the square brackets in $\mathbb{Q}[i]$ in §1.2.6. There is a difference in the meaning of these notations, but it turns out that they are interchangeable in this case. This will be explained more below.

You have learned about vector spaces in 2LALP: Linear Algebra and Lemma 3.11 below allows us to apply the theory of vector spaces to study field extensions. The only parts of the theory of vector spaces that we will use here is that if a vector space is finite dimensional, then it has a basis, and that the number of elements in a basis does not depend on the choice of basis. Although you have mainly considered vector spaces over $\mathbb{R}$ or $\mathbb{C}$ in 2LALP, you should be aware that the theory goes through for vector spaces over any field.

Before stating Lemma 3.11, we explain how to get an addition and scalar multiplication on $\mathbb{K}$ so that we can view it as a vector space over $\mathbb{F}$.

Let $u, v \in \mathbb{K}$ and let $a \in \mathbb{F}$. Then we can add u and v in $\mathbb{K}$ to get $u + v$ and this gives the addition in $\mathbb{K}$ as a vector space. Also we can multiply a and v in $\mathbb{K}$ to get av and this gives the scalar multiplication.

Next we show that this really does give $\mathbb{K}$ the structure of a vector space over $\mathbb{F}$.

Lemma 3.11. *Let $\mathbb{F}$ and $\mathbb{K}$ be fields with $\mathbb{F} \subseteq \mathbb{K}$. Then $\mathbb{K}$ is a vector space over $\mathbb{F}$ with addition and scalar multiplication given by the addition and multiplication in $\mathbb{K}$.*

Proof. We just have to check that the axioms of a vector space hold for $\mathbb{K}$ viewed as a vector space over $\mathbb{F}$. However, we observe that each axiom of a vector space follows immediately from one of the axioms of a ring. We leave it as an exercise for you to check this by "matching up" each axiom of a vector space with an axiom of a ring. □

Viewing $\mathbb{K}$ as a vector space over $\mathbb{F}$ we can consider whether it is finite dimensional. This leads to the following definition.

Definition 3.12. Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension. We say that $\mathbb{F} \subseteq \mathbb{K}$ is a *finite extension* if $\mathbb{K}$ is finite dimensional as a vector space over $\mathbb{F}$.

Going back to the examples of field extensions above, we can observe that $\mathbb{R} \subseteq \mathbb{C}$ is a finite extension, and $\{1, i\}$ is a basis of $\mathbb{C}$ over $\mathbb{R}$, so that $[\mathbb{C} : \mathbb{R}]$. Also $\mathbb{Q} \subseteq \mathbb{Q}(i)$ and $\mathbb{Q} \subseteq \mathbb{Q}(\alpha)$ are finite extensions – it is a good exercise to check this by giving basis of $\mathbb{Q}(i)$ as a vector space over $\mathbb{Q}$ and a basis of $\mathbb{Q}(\alpha)$ as a vector space over $\mathbb{Q}$. The field extension $\mathbb{Q} \subseteq \mathbb{R}$ is not finite, we won't go in to the details here, but just note that it follows from the facts that $\mathbb{Q}$ is countable and that $\mathbb{R}$ is uncountable.

Next we consider a field extension generated by a single element. Before doing this we need to check that the definition we want to make is sensible.

Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension and let $\alpha \in \mathbb{K}$.

We can consider the set $\mathbb{F}[\alpha]$ of all elements of $\mathbb{K}$ of the form $f(\alpha)$, where $f(X) \in \mathbb{F}[X]$. We can check using one of the subring tests, that $\mathbb{F}[\alpha]$ is a subring of $\mathbb{K}$.

Then we define $\mathbb{F}(\alpha) = Q(\mathbb{F}[\alpha])$ to be the field of fractions of $\mathbb{F}[\alpha]$. So $\mathbb{F}(\alpha)$ consists of all elements of $\mathbb{K}$ of the form $\frac{f(\alpha)}{g(\alpha)}$, where $f(X), g(X) \in \mathbb{F}[X]$ and $g(\alpha) \neq 0$. Also $\mathbb{F}(\alpha)$ is a field and we can see that it is the smallest subfield of $\mathbb{K}$ containing $\mathbb{F}$ and α.

The preceding discussion allows us to make the following definition.

Definition 3.13. Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension and let $\alpha \in \mathbb{K}$. We define:

(a) $\mathbb{F}[\alpha]$ to be the subring of $\mathbb{K}$ consisting of all elements of the form $f(\alpha)$, where $f(X) \in \mathbb{F}[X]$; and

(b) $\mathbb{F}(\alpha) = Q(\mathbb{F}[\alpha])$ to be the subfield of $\mathbb{K}$ consisting of all elements of the form $\frac{f(\alpha)}{g(\alpha)}$, where $f(X), g(X) \in \mathbb{F}[X]$ and $g(\alpha) \neq 0$.

We refer to $\mathbb{F} \subseteq \mathbb{F}(\alpha)$ as a *simple field extension* and we say $\mathbb{F}(\alpha)$ is obtained by *adjoining α to* $\mathbb{F}$.

As an example of a simple field extension we consider $\mathbb{Q}(i)$. In this case we actually have that $\mathbb{Q}(i) = \mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$, because if $f(X), g(X) \in \mathbb{Q}[X]$ with $g(i) \neq 0$. Then we have $f(i) = a + bi$ and $g(i) = c + di$ for some $a, b, c, d \in \mathbb{Q}$ and

$$\begin{aligned} \frac{f(i)}{g(i)} &= \frac{a+bi}{c+di} \\ &= \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} \in \mathbb{Q}[i]. \end{aligned}$$

For $\alpha = \sqrt[3]{2}$, we have seen in Examples 2.37(a) that $\mathbb{Q}[\alpha]$ is actually a field, and it follows that $\mathbb{Q}[\alpha] = \mathbb{Q}(\alpha)$ – you should think about this to understand why it is true.

In fact later we'll see that for $\alpha \in \mathbb{C}$ we have $\mathbb{Q}[\alpha] = \mathbb{Q}(\alpha)$ if and only if α is an algebraic number. We recall that $\alpha \in \mathbb{C}$ is an *algebraic number* if there exists a nonzero polynomial $f(X) \in \mathbb{Q}[X]$ such that $f(\alpha) = 0$.

More generally, we can define an algebraic element of a field extension.

Definition 3.14. Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension. We say that α is *algebraic* over $\mathbb{F}$ if there exists a nonzero polynomial $f(X) \in \mathbb{F}[X]$ such that $f(\alpha) = 0$.

So we see that saying $\alpha \in \mathbb{C}$ is an algebraic number is the same as saying that α is algebraic over $\mathbb{Q}$.

Our next theorem tells us all about finite simple field extensions. The proof of this theorem is not examinable, but you should certainly go through it, as it is really nice piece of mathematics

Theorem 3.15. *Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension and let $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$.*

(a) *There is a unique monic polynomial $m_\alpha(X) \in \mathbb{F}[X]$ of minimal degree such that $m_\alpha(\alpha) = 0$.*

(b) *$m_\alpha(X)$ is irreducible.*

(c) *$\mathbb{F}(\alpha) \cong \mathbb{F}[X]/\langle m_\alpha(X) \rangle$.*

(d) *$\mathbb{F}(\alpha) = \mathbb{F}[\alpha]$.*

(e) *$\mathbb{F} \subseteq \mathbb{F}(\alpha)$ is a finite extension and the dimension of $\mathbb{F}(\alpha)$ as a vector space over $\mathbb{F}$ is equal to the degree of $m_\alpha(X)$.*

Proof. (a) Consider the homomorphism $\epsilon_\alpha : \mathbb{F}[X] \to \mathbb{F}[\alpha]$ defined by $\epsilon_\alpha(f(X)) = f(\alpha)$. Since α is algebraic over $\mathbb{F}$, there exists $f(X) \in \mathbb{F}[X]$ such that $f(\alpha) = 0$, so $f(X) \in \ker \epsilon_\alpha$. Thus $\ker \epsilon_\alpha \neq \{0\}$. Since $\mathbb{F}[X]$ is a principal ideal domain by Proposition 2.32, we have that $\ker \epsilon_\alpha = \langle m(X) \rangle$ for some $m(X) \in \mathbb{F}[X]$. Moreover, we can assume that $m(X)$ is monic as discussed after the proof of Proposition 2.32

Since $m(X) \in \ker \epsilon_\alpha$, we have that $m(\alpha) = 0$. Now suppose $h(X) \in \mathbb{F}[X]$ is monic and satisfies $h(\alpha) = 0$. Then we have $h(X) \in \ker \epsilon_\alpha = \langle m(X) \rangle$, so that $h(X) = g(X)m(X)$ for some $g(X) \in \mathbb{F}[X]$. Thus either $\deg h(X) > \deg m(X)$ or $h(X) = m(X)$. Hence, $m(X)$ is the unique polynomial of minimal degree such that $m(\alpha)$, which proves (a) with $m_\alpha(X) = m(X)$.

(b) Suppose that $m_\alpha(X) = f(X)g(X)$ where $f(X), g(X) \in \mathbb{F}[X]$. Then we have $0 = m_\alpha(\alpha) = f(\alpha)g(\alpha)$, so that $f(\alpha) = 0$ or $g(\alpha) = 0$. Since $m_\alpha(X)$ is of minimal degree such that $m_\alpha(X) = 0$, we must have that one of $f(X)$ or $g(X)$ is a constant polynomial. Hence, $m_\alpha(X)$ is irreducible.

(c) We have that $\operatorname{im} \epsilon_\alpha = \mathbb{F}[\alpha]$, and we have seen that $\ker \epsilon_\alpha = \langle m_\alpha(X) \rangle$. Therefore, we have $\mathbb{F}[\alpha]s \cong \mathbb{F}[X]/\langle m_\alpha(X) \rangle$. by the isomorphism theorem.

(d) Since $m_\alpha(X)$ is irreducible, we have that $\langle m_\alpha(X) \rangle$ is maximal by Proposition 2.35. Thus $\mathbb{F}[X]/\langle m_\alpha(X) \rangle$ is a field by Theorem 2.36. So $\mathbb{F}[\alpha]$ is a field as it is isomorphic to $\mathbb{F}[X]/\langle m_\alpha(X) \rangle$. Hence, $\mathbb{F}(\alpha) = Q(\mathbb{F}[\alpha]) = \mathbb{F}[\alpha]$.

(e) Let $d = \deg m_\alpha(X)$. Using the division theorem for polynomials we can show that each element of $\mathbb{F}[X]/\langle m_\alpha(X) \rangle$ can be written uniquely in the form $a_0 + a_1X + \cdots + a_{d-1}X^{d-1} + \langle m_\alpha(X) \rangle$, where $a_0, a_1, \ldots, a_{d-1}$. Therefore, $\{1 + I, X + I, \ldots, X^{d-1} + I\}$ is a basis for $\mathbb{F}[X]/\langle m_\alpha(X) \rangle$. It follows that $\{1, \alpha, \ldots, \alpha^{d-1}\}$ is a basis for $\mathbb{F}[\alpha] = \mathbb{F}(\alpha)$. Hence, the dimension of $\mathbb{F}(\alpha)$ as a vector space over $\mathbb{F}$ is d. □

The following corollary gives a useful equivalent condition for an element in a field extension to be algebraic.

Corollary 3.16. *Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension and let $\alpha \in \mathbb{K}$. Then α is algebraic over $\mathbb{F}$ if and only if $\mathbb{F}(\alpha)$ is a finite extension of $\mathbb{F}$.*

Proof. First suppose that $\mathbb{F} \subseteq \mathbb{F}(\alpha)$ is a finite extension, so that $\mathbb{K}$ is a finite dimensional as a vector space over $\mathbb{F}$. Let d be the dimension of $\mathbb{K}$ as a vector space over $\mathbb{F}$. Consider $\{1, \alpha, \alpha^2, \ldots, \alpha^d\} \subseteq \mathbb{K}$. Since this set has $d+1 > d$ elements it must be linearly dependent over $\mathbb{F}$. Therefore, there exist $a_0, a_1, \ldots, a_d \in \mathbb{F}$ not all zero such that $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_d\alpha^d = 0$. Thus $f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_dX^d \in \mathbb{F}[X]$ satisfies $f(\alpha) = 0$.

Conversely suppose that α is algebraic over $\mathbb{F}$. Then $\mathbb{F} \subseteq \mathbb{F}(\alpha)$ is a finite extension by Theorem 3.15(e). □

The polynomial $m_\alpha(X)$ from Theorem 3.15(b) is important and gets a special name.

Definition 3.17. Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension and let $\alpha \in \mathbb{K}$ be algebraic over $\mathbb{F}$. The monic polynomial $m_\alpha(X) \in \mathbb{F}[X]$ of least degree such that $m_\alpha(\alpha) = 0$ is called the *minimal polynomial of α over $\mathbb{F}$*.

Let $\mathbb{F}$ and α be as in Definition 3.17, and suppose $m(X) \in \mathbb{F}[X]$ is a polynomial such that $m(\alpha) = 0$. Then it follows from the proof of Theorem 3.15 that $m_\alpha(X)$ is a factor of $m(X)$. Thus if $m(X)$ is monic and irreducible, then $m(X) = m_\alpha(X)$. Therefore, if we can find a monic irreducible polynomial $m(X) \in \mathbb{F}[X]$ such that $m(\alpha) = 0$, then $m(X) = m_\alpha(X)$ is the minimal polynomial of α, and this turns out to be helpful when trying to work out minimal polynomials.

In the examples below we just consider some algebraic numbers $\alpha \in \mathbb{C}$, and determine their minimal polynomial over $\mathbb{C}$.

Examples 3.18. (a) We consider $\alpha = i$. We know that $i^2 + 1 = 0$. Also $X^2 + 1$ is irreducible in $\mathbb{Q}[X]$, because it were reducible in $\mathbb{Q}[X]$, then it would have a root in $\mathbb{Q}$ and this is not the case. Therefore $m_i(X) = X^2 + 1$.

More generally, we can consider a primitive nth root of unity $\zeta_n = e^{\frac{2\pi}{n}} \in \mathbb{C}$. It turns out that the minimal polynomial of ζ_n is

$$m_{\zeta_n}(X) = \prod_{1 \le m < n, \mathrm{hcf}(m,n)=1} (X - \zeta_n^m).$$

We should explain the notation used above, it means we take the product of the linear factors $X - \zeta_n^m$ over all $m \in \{1, 2, \dots, n-1\}$ that are coprime to n. It is beyond the scope of this course to prove that this really in the minimal polynomial of ζ_n, so we omit this; we note that it is not really obvious that the coefficients are in $\mathbb{Q}$.

As a particular example, we consider $\omega = \zeta_3$. In this case we find that $m_\omega(X) = X^2 + X + 1$, and it is easy to check that this quadratic polynomial is irreducible in $\mathbb{Q}[X]$, as it has no roots there, so that it is the minimal polynomial of ω.

We have $\zeta_4 = i$, and have seen what $m_i(X)$ is above.

For ζ_5, we have that $m_{\zeta_5}(X) = X^4 + X^3 + X^2 + X + 1$, and we can show that this is irreducible is an elementary way. We do have to check it has no quadratic factors, which is a bit of work, and we may go through this in the lectures. The best to do this is to factorize it fully over $\mathbb{C}$ and then check that no pair of linear factors can be multiplied together to get a quadratic polynomial with coefficients in $\mathbb{Q}$.

(b) For $\alpha = \sqrt{2}$, we have $m_{\sqrt{2}}(X) = X^2 - 2$. This can be verified by noting that $\sqrt{2}$ is certainly a root of $X^2 - 2$, and that $X^2 - 2$ is irreducible in $\mathbb{Q}[X]$ (because it has no roots in $\mathbb{Q}$).

More generally, for $\alpha = \sqrt{n}$, where $n \in \mathbb{Z}$ is not a perfect square, we have $m_{\sqrt{n}}(X) = X^2 - n$.

(c) We consider $\alpha = \sqrt[3]{2}$ and will show that $m_{\sqrt[3]{2}} = X^3 - 2$. If $X^3 - 2$ is reducible in $\mathbb{Q}[X]$, then it would have a root in $\mathbb{Q}$, because it has degree 3. But we know that the roots are $\sqrt[3]{2}$, $\sqrt[3]{2}\omega$ and $\sqrt[3]{2}\omega^2$, none of which are in $\mathbb{Q}$. Also $\sqrt[3]{2}$ is a root of $X^3 - 2$. Hence, $X^3 - 2$ is the minimal polynomial of $\sqrt[3]{2}$.

(d) We consider $\alpha = i + \sqrt{2}$, and we have

$$\begin{aligned} m_\alpha(X) &= (X - i - \sqrt{2})(X - i + \sqrt{2})(X + i - \sqrt{2})(X + i + \sqrt{2}) \\ &= (X^2 - 2iX - 3)(X^2 + 2iX - 3) \\ &= X^4 - 2X^2 + 9. \end{aligned}$$

By definition of $m_\alpha(X)$, we have that α is a root. So we just need to check that $m_\alpha(X)$ is irreducible. From the factorization above we can determine all of the possible factors of $m_\alpha(X)$, and then check that none of these have coefficients in $\mathbb{Q}$. We may go through this in more detail in the lectures.

Following these examples we give a quick note of warning. Above we argued that some quadratic and cubic polynomials are irreducible, by saying that if they are reducible, then they must have a root. This only works in for polynomials of degree 2 or 3, because if they are factorized as the product of two nonconstant polynomials, then one of these factors must be a linear factor. For higher ranks, there can be reducible polynomials in $\mathbb{Q}[X]$, which have no roots in $\mathbb{Q}$. For example, $X^4 - X^2 - 2 = (X^2 + 1)(X^2 - 2)$ is reducible in $\mathbb{Q}[X]$ but has no roots in $\mathbb{Q}$.

We end this section by explaining how to construct simple field extensions. Suppose that $m(X) \in \mathbb{Q}$ is an irreducible polynomial. Then by the fundamental theorem of algebra, there is a root $\alpha \in \mathbb{C}$ of $m(X)$. Thus we can form the field $\mathbb{Q}(\alpha) \subseteq \mathbb{C}$, which is an extension of $\mathbb{Q}$ and contains a root of $m(X)$.

Now suppose that $\mathbb{F}$ is a field and that $m(X) \in \mathbb{F}[X]$ is an irreducible polynomial of degree $d \in \mathbb{N}$. Then we may wonder if it is possible to find an extension of $\mathbb{F}$, which contains a root of $m(X)$. We have already seen how to do this in some cases in Examples 2.37, and we'll go through it in general here.

Let $I = \langle m(X) \rangle$ be the principal ideal of $\mathbb{F}[X]$ generated by $m(X)$. Since $m(X)$ is irreducible, I is maximal by Proposition 2.35 and $\mathbb{F}[X]/I$ is a field by Theorem 2.36.

Let $\mathbb{K} = \mathbb{F}[X]/I$. We'll see that $\mathbb{K}$ is an extension of $\mathbb{F}$, and contains a root of $m(X)$. However, $\mathbb{F}[X]/I$ is given as a factor ring, so it is not that convenient to work with, and things are made easier if we make a nice choice of coset representatives and a good notation.

By the division theorem for polynomials we can write any $f(X) \in \mathbb{F}[X]$ uniquely in the form $f(X) = q(X)m(X) + r(X)$, where $q(X), r(X) \in \mathbb{F}[X]$, and $r(X) = 0$ or $\deg r(X) < d = \deg m(X)$. This implies that each coset in $\mathbb{K} = \mathbb{F}[X]/I$ has a unique representative $f(X)$ satisfying $f(X) = 0$ or $\deg f(X) < d$. Thus every element of $\mathbb{K}$ can be expressed uniquely in the form

$$a_0 + a_1 X + \cdots + a_{d-1} X^{d-1} + I,$$

where $a_0, a_1, \ldots, a_{d-1} \in \mathbb{F}$.

Next we'll explain how we can view $\mathbb{K}$ as an extension of $\mathbb{F}$. We note that the function $\iota : \mathbb{F} \to \mathbb{K}$ defined by $\iota(a) = a + I$ is an injective homomorphism. We use this to identify $\mathbb{F}$ as a subfield of $\mathbb{K}$. By this we mean that for $a \in \mathbb{F}$, we will think of $a + I$ as being the same as a; this is an abuse of notation, but as long as we're careful and remember what we're doing it won't do us any harm.

Now we let $\alpha = X + I \in \mathbb{K}$. Then we have that the elements of $\mathbb{K}$ are of the form

$$a_0 + a_1 \alpha + \cdots + a_{d-1} \alpha^{d-1},$$

where $a_0, a_1, \ldots, a_{d-1} \in \mathbb{F}$.

Also for $f \in \mathbb{F}[X]$, we have that $f(\alpha) = f(X) + I$, and therefore $m(\alpha) = 0$. Hence, α is a root of $m(X)$ in $\mathbb{K}$.

Summarizing what we have done above gives the following theorem.

Theorem 3.19. *Let $\mathbb{F}$ be a field and let $m(X) \in \mathbb{F}[X]$ be irreducible. Then there exists an simple extension $\mathbb{F}(\alpha)$ of $\mathbb{F}$, where α is a root of $m(X)$.*

We end this section by briefly discussing simple field extensions that are not finite. We say that α is *transcendental* over $\mathbb{F}$ if it is not algebraic. So we have α is transcendental over $\mathbb{F}$ if and only if $\mathbb{F} \subseteq \mathbb{F}(\alpha)$ is not a finite extension.

We say that $\alpha \in \mathbb{C}$ is a *transcendental number* if it is not an algebraic number, which is the same as saying that it is transcendental over $\mathbb{Q}$. You may have heard that π and e transcendental numbers. In fact, it can be shown that the set of algebraic numbers is countable, whereas $\mathbb{C}$ is uncountable, which means (in a sense) that almost all elements of $\mathbb{C}$ are transcendental, but we won't go in to that here.

Next we consider $\mathbb{F} \subseteq \mathbb{F}(X)$ where we recall that $\mathbb{F}(X) = Q(\mathbb{F}[X])$ is the field of fractions of the polynomial ring $\mathbb{F}[X]$. In this case we can take $\alpha = X \in \mathbb{F}(X)$, so that $\mathbb{F}(\alpha) = \mathbb{F}(X)$. We can show that $\mathbb{F} \subseteq \mathbb{F}(X)$ is not a finite extension, so that $\alpha = X$ is transcendental over $\mathbb{F}$.

In fact we can say what a simple field extension by a transcendental element looks like. Indeed if $\mathbb{F} \subseteq \mathbb{K}$ is a field extension and $\alpha \in \mathbb{K}$ is transcendental over $\mathbb{F}$, then we have $\mathbb{F}(\alpha) \cong \mathbb{F}(X)$. We do not go in to how this is proved, and leave that as an exercise. We note, in particular, that this implies that $\mathbb{Q}(\pi) \cong \mathbb{Q}(X)$ and that $\mathbb{Q}(e) \cong \mathbb{Q}(X)$.

3.4 Finite fields

In this section we discuss the theory of finite fields, much of which was developed by Évariste Galois.

In Examples 2.37(b), we have seen a field with $9 = 3^2$ elements, and in Examples 2.37(c), we have seen a field with $125 = 5^3$ elements. More generally, let p be a prime and let $m(X) \in \mathbb{F}_p[X]$ be an irreducible polynomial of degree $d \in \mathbb{N}$. In the discussion preceding Theorem 3.19, we showed that the quotient $\mathbb{F}_p[X]/\langle m(X)\rangle$ is a field. Moreover, we explained how we can write the elements of $\mathbb{F}_p[X]/\langle m(X)\rangle$ uniquely in the form $a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1}$, where $a_0, a_1, \dots, a_{d-1} \in \mathbb{F}_p$ and $\alpha = X + I$ (so that α satisfies $m(\alpha) = 0$). Thus $\mathbb{F}_p[X]/\langle m(X)\rangle$ has p^d elements.

In fact it turns out that any finite field has order p^d, where $p, d \in \mathbb{N}$ and p is prime, as we prove in the next lemma.

Lemma 3.20. *Let $\mathbb{F}$ be a finite field. Then $|\mathbb{F}| = p^d$, where $p, d \in \mathbb{N}$ and p is prime.*

Proof. Since $\mathbb{F}$ is a finite field, the characteristic of $\mathbb{F}$ is a positive; otherwise the prime subfield of $\mathbb{F}$ is isomorphic to $\mathbb{Q}$ by Theorem 3.9. Therefore, the prime subfield of $\mathbb{F}$ is isomorphic to $\mathbb{F}_p$, and we will use this to identify $\mathbb{F}_p$ with the prime subfield of $\mathbb{F}$.

Thus $\mathbb{F}_p \subseteq \mathbb{F}$ is a field extension and $\mathbb{F}$ is a vector space over $\mathbb{F}_p$ by Lemma 3.11. Since $\mathbb{F}$ is finite, this extension must be finite, so we can choose a basis $\{v_1, \dots, v_d\}$ of $\mathbb{F}$ as a vector space over $\mathbb{F}_p$, where d is the dimension of $\mathbb{F}$ as a vector space over $\mathbb{F}_p$. Then the elements of $\mathbb{F}$ are of the form $a_1v_1 + a_2v_2 + \cdots + a_dv_d$, where $a_1, a_2, \dots, a_d \in \mathbb{F}_p$, so there are exactly p^d of them. □

We now state our main theorem about finite fields, which tells us that there is a field of p^d for any prime power, and that this field is unique up to isomorphism. The proof is beyond the syllabus, and is omitted. We note however that for (a) it suffices to show that there is an irreducible polynomial $m(X) \in \mathbb{F}_p[X]$ of degree d, and then we can use the construction given in the proof of Theorem 3.19 to find a field of order p^d.

Theorem 3.21. *Let $p, d \in \mathbb{N}$ with p prime. Then*

(a) *there exists a field of order p^d; and*

(b) *any two fields of order p^d are isomorphic.*

We look at another example of a finite field.

Example 3.22. Let $m(X) = X^2 + 2 \in \mathbb{F}_5[X]$, let $I = \langle m(X) \rangle$ and $\mathbb{F} = \mathbb{F}_5[X]/I$.

We can check that $m(X)$ is irreducible in $\mathbb{F}_5[X]$ as follows. If $m(X)$ were irreducible, then it would have a root in $\mathbb{F}_5$, but we can check that no element of $\mathbb{F}_5$ is a root of $m(X)$.

Therefore, I is a maximal ideal by Proposition 2.35, and $\mathbb{F}$ is a field by Theorem 2.36.

Using the notation used in the previous subsection we identify $\mathbb{F}_5$ as a subfield of $\mathbb{F}$ (so by an abuse of notation we write a rather than $a+I$ for $a \in \mathbb{F}_5$) and we let $\alpha = X+I$. Then $\mathbb{F} = \{a + b\alpha : a, b \in \mathbb{F}_5\}$, and α satisfies $\alpha^2 + 2 = 0$, so that $\alpha^2 = 3$. The order of $\mathbb{F}$ is 25.

This may all be very nice and indeed the theory of finite fields is a beautiful area of mathematics, but it is also worth noting that the finite fields have a variety of important applications in coding theory and cryptography, as well as many other ares of mathematics and computer science. A good first place to look to find out a bit more is wikipedia and the links there
http://en.wikipedia.org/wiki/Finite_field#Applications.

3.5 Summary of Chapter 3

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

- understand and explain the construction of the field of fractions of an integral domain;
- define the characteristic of a field, prove that it is equal to zero or a prime number and determine the characteristic of examples of fields;
- state the definition a field extension, a finite field extension and a simple field extension;
- state the definition of an algebraic number and its minimal polynomial, prove that the minimal polynomial is irreducible and determine the minimal polynomial in examples; and
- calculate in examples of finite fields.

3.6 Exercises for Chapter 3

There are now quite a few exercises here. Some are perhaps quite challenging, so you may want to ask for some hints.

Q3.1. Let R be an integral domain. Prove that axiom (Dl) holds for the field of fractions $Q(R)$ of R.

Q3.2. (a) Let $\mathbb{Z}[i]$ be the ring of Gaussian integers. Prove that $Q(\mathbb{Z}[i]) = \mathbb{Q}[i]$. *In this question you should identify $Q(\mathbb{Z}[i])$ with a subfield of $\mathbb{C}$ using Lemma 3.4.* (b) Let $\mathbb{Z}[\sqrt[3]{2}] = \{a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2 : a, b, c \in \mathbb{Z}$. What is $Q(\mathbb{Z}[\sqrt[3]{2}])$?

Q3.3. Let $\mathbb{F}$ be a field and let R and S be subrings of $\mathbb{F}$ with $S \subseteq R$.

(a) Show that $Q(S) \subseteq Q(R)$.
You should identify $Q(S)$ and $Q(R)$ with subfields of $\mathbb{F}$.

(b) Give an example of R and S such that $S \subsetneqq R$ and $Q(S) = Q(R)$.

Q3.4. Let R be a commutative ring with one and let S be a subset of R. Suppose that S is closed under multiplication, i.e. for all $a, b \in S$, we have $ab \in S$. Define $\sim$ on $R \times S$ by $(a, b) \sim (c, d)$ means $ad = bc$.

(a) Show that $\sim$ is an equivalence relation on $R \times S$.

Let R_S denote the set of equivalence classes of $\sim$ in $R \times S$.

(b) Define an addition and multiplication on R_S, and show that they are well defined and give R_S the structure of a commutative ring with one.

(c) Show that $S = R \setminus \{0\}$ is closed under multiplication if and only if R is an integral domain.
Note that in this case we have $R_S = Q(R)$.

(d) Let $p \in \mathbb{N}$ be prime, and let $S = \{p^n : n \in \mathbb{N}\} \subseteq \mathbb{Z}$. Show that S is closed under multiplication, and give a description of R_S.

Q3.5. Let $\mathbb{F}$ be a field and let $\theta : \mathbb{Z} \to \mathbb{F}$, be the homomorphism defined by $\theta(m) = m1$. Prove that $\ker\theta$ is the principal ideal generated by $\operatorname{char}\mathbb{F}$.

Q3.6. Let $\mathbb{K}$ be a field and let $\mathbb{F}$ be a subfield of $\mathbb{K}$. Show that $\operatorname{char}\mathbb{F} = \operatorname{char}\mathbb{K}$.

Q3.7. Let $\mathbb{F}$ be a field with $\operatorname{char}\mathbb{F} = p$ where $p \in \mathbb{N}$ is prime.

(a) Prove that the function $\theta : \mathbb{F} \to \mathbb{F}$ is a homomorphism.
(b) Prove that $\ker\theta = \{0\}$.
(c) Suppose that $\mathbb{F}$ is finite. Show that θ is surjective.
(d) Give an example of $\mathbb{F}$ such that θ is not surjective.

Q3.8. Let $c \in \mathbb{C}$ and consider the homomorphism $\epsilon_c : \mathbb{Q}[X] \to \mathbb{C}$ defined by $\epsilon_c(f(X)) = f(c)$.
Show that ϵ_c is injective or that $\ker\epsilon_c = \langle m_c(X)\rangle$ for some monic irreducible polynomial $m_c(X) \in \mathbb{Q}[X]$. Deduce that $\operatorname{im}\epsilon_c$ is a field if ϵ_c is not injective.
This question is essentially asking you to cover some of the parts in Theorem 3.15 for the case $\mathbb{F} = \mathbb{Q}$.

Q3.9. Show that each of the following elements of $\mathbb{C}$ are algebraic over $\mathbb{Q}$ and determine their minimal polynomial.
(a) $\sqrt[3]{5}$.
(b) $e^{\frac{2\pi i}{6}}$.
(c) $i + \sqrt{3}$.

Q3.10. (a) Suppose $\alpha \in \mathbb{R}$ satisfies $\alpha^4 - \alpha^2 - 2 = 0$. Determine the minimal polynomial of α.
(b) Find $m_\alpha(X)$ when $\alpha \in \mathbb{R}$ is an irrational number satisfying $\alpha^3 + 3\alpha^2 - 2 = 0$.

Q3.11. Let $\mathbb{F} \subseteq \mathbb{K}$ be a field extension, and let $\alpha \in \mathbb{K}$. Suppose that α is transcendental over $\mathbb{F}$. Prove that $\mathbb{F}(\alpha) \cong \mathbb{F}(X)$.
Recall that $\mathbb{F}(X)$ is the field of rational functions over $\mathbb{F}$ in the indeterminate X.

Q3.12. Let $\mathbb{K}$ be an extension field of $\mathbb{C}$, and let $\alpha \in \mathbb{K}$. Prove that α is transcendental over $\mathbb{C}$.
Hint: You may want to use the fundamental theorem of algebra.

Q3.13. Prove that $\mathbb{Q} \subseteq \mathbb{R}$ is not a finite extension.

Q3.14. Let $\mathbb{F}$ be a finite field with q elements, and let $m(X) \in \mathbb{F}[X]$ be an irreducible polynomial of degree d. Prove that $\mathbb{F}[X]/\langle m(X)\rangle$ is a field with q^d elements.

Q3.15. Find an irreducible cubic polynomial $f(X) \in \mathbb{F}_3[X]$ and use it to construct a field of order 27.

Q3.16. Consider the polynomials $f_1(X) = X^2 + 1$ and $f_2(X) = X^2 + 2X + 2$ in $\mathbb{F}_3[X]$. Let $I_1 = \langle f_1(X)\rangle$ and $I_2 = \langle f_2(X)\rangle$.
(a) Show that $f_1(X)$ and $f_2(X)$ are irreducible in $\mathbb{F}_3[X]$.
(b) Deduce that $\mathbb{F}_2[X]/I_1$ and $\mathbb{F}_2[X]/I_2$ are fields of order 9.
(c) Find an isomorphism between $\mathbb{F}_2[X]/I_1$ and $\mathbb{F}_2[X]/I_2$.

Chapter 4

Groups

In this chapter, we're going to change direction a bit and start the study of group theory. Groups are sets which a single binary operation to combine elements which must satisfy some axioms. So the theory of groups has a slightly different flavour to that of rings as the conditions on a group are less restrictive. We'll see that groups show up all over mathematics, and that there are lots of examples of them. In particular, we see that groups give a mathematical language for the study of symmetry.

Before we move on to groups, we have a short section with a bit of a reminder about permutations. These are important in group theory, so we just want to remind ourselves about how to calculate with permutations.

4.1 Permutations

We start off by recapping some things about permutations.

Let Ω be a set.

- We recall that a *permutation of* Ω is by definition a bijection from Ω to itself.
- We write $\mathrm{Sym}(\Omega)$ for the set of all permutations of Ω, see Definition 0.30.
- We refer to $\mathrm{Sym}(\Omega)$ as the *symmetric group on* Ω.
- The *identity function on* Ω is the function $\mathrm{id}_\Omega : \Omega \to \Omega$ defined by $\mathrm{id}_\Omega(x) = x$. We note that id_Ω is clearly a bijection, so $\mathrm{id}_\Omega \in \mathrm{Sym}(\Omega)$. From now on we denote id_Ω by e.
- Let $f \in \mathrm{Sym}(\Omega)$. The *inverse of* f is the function $f^{-1} : \Omega \to \Omega$ defined by

  $$f^{-1}(x) \text{ is the unique element } y \in \Omega \text{ such that } f(y) = x.$$

  We have that f^{-1} is a bijection, so $f^{-1} \in \mathrm{Sym}(\Omega)$.
- In the case where $\Omega = \{1, 2, \dots, n\}$, we usually write S_n instead of $\mathrm{Sym}(\Omega)$.
- We refer to S_n as *the symmetric group of degree* n.
- We know that the number of permutations of $\{1, 2, \dots, n\}$ is $n!$ so that S_n has $n!$ elements.
- We have the two row notation and cycle notation for permutations in S_n, as covered in Section 0.4. There is a remark at the end of section explaining that cycle notation for a permutation is not unique, but that understand this non-uniqueness so it doesn't cause any problem; you may want to have a quick look at this.

In the next example we are going to do some calculations in S_n using cycle notation. Before we get on to this we explain some notational conventions that we use for S_n.

WARNING! We're going to do things a bit differently here, which may be confusing at first, but stick with it and you'll get used to it quickly.

When we work with $S_n = \mathrm{Sym}(\{1, 2, \dots, n\})$, we write functions "on the right", and we usually use small Greek letters for elements of S_n. So for $\sigma \in S_n$ and $j \in \{1, 2, \dots, n\}$, we write $j\sigma$ for the image of j under σ, instead of $\sigma(j)$. As mentioned above this may seem a bit peculiar to start with, but you'll get used to it, and it does make calculations easier. So the two row notation of σ is written as

$$\begin{pmatrix} 1 & 2 & \dots & n \\ 1\sigma & 2\sigma & \dots & n\sigma \end{pmatrix}.$$

For example, for $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} \in S_4$, we have $1\sigma = 3$, $2\sigma = 1$, $3\sigma = 2$ and $4\sigma = 4$. This means that we compose permutations from "left to right". Also we'll omit the symbol $\circ$ when we write compositions. So for $\sigma, \tau \in S_n$, we write $\sigma\tau$ to mean the permutation we obtain when we apply σ and then τ, that is $j(\sigma\tau) = (j\sigma)\tau$, for $j \in \{1, 2, \dots, n\}$. For example for $\sigma \in S_4$ as above and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \in S_4$ we have $\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}$.

In particular, we omit the composition symbol $\circ$ when writing permutations in cycle notation. Also we usually omit writing 1-cycles in the cycle notation for an element of S_n as a 1-cycle is equal to the identity. We'll usually use cycle notation for elements of S_n from now on.

Also we recall how we define powers of permutations Let $\sigma \in S_n$ and $r \in \mathbb{Z}$. We define σ^r as follows.

- For $r = 0$, we set $\sigma^0 = e$.
- For $r > 0$, we set $\sigma^r = \sigma\sigma \cdots \sigma$, where there are r factors all equal to σ.
- For $r < 0$, we let $s = -r$, so $s > 0$ and then set $\sigma^r = (\sigma^{-1})^s$.

All this notation may sound a little bewildering at first, but once you've seen it in action in some examples, then it should be fine. Note that the convention of whether to write permutations on the right or on the left varies between authors, hopefully you'll get an idea of some motivation why we write them on the right in the comments after the examples.

Examples 4.1. (a) First we'll do an example to remind ourselves of working out cycle notation. Let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 4 & 1 & 8 & 3 & 9 & 2 & 7 & 6 \end{pmatrix} \in S_9$$

written in two-row notation.

First we look at the sequence

$$1,\ 1\sigma,\ 1\sigma^2, \dots = 1, 5, 3, 1, 5, 3, \dots.$$

This give us our first cycle

$$(153)$$

Recall that σ^2 means $\sigma\sigma$, and more generally for $m \in \mathbb{N}$, we write σ^m for σ composed with itself m times.
Next we look at

$$2,\ 2\sigma,\ 2\sigma^2, \cdots = 2, 4, 8, 7, 2, 4, 8, 7 \ldots.$$

This gives our second cycle

$$(2487).$$

Next we look at

$$6,\ 6\sigma,\ 6\sigma^2, \cdots = 6, 9, 6, 9 \ldots.$$

This gives our third and final cycle

$$(69).$$

Thus we have decomposed σ as a product of cycles:

$$f = (153)(2487)(69).$$

Remember that we do not write the symbol $\circ$ between the cycles.

(b) We're going to see examples of how to compose and invert permutations in cycle notation. We did some examples like this in 1AC, but we'll do more here as we're doing things a bit differently. Let

$$\rho = (124)(365), \quad \sigma = (1326)(45), \quad \tau = (153)(46) \quad \in S_6.$$

Remember that we omit 1-cycles in the cycle notation, so we have not included the 1-cycle (2) in τ.
We work out $\rho\sigma$ by saying.
Well ρ sends 1 to 2 and σ sends 2 to 6, so $\rho\sigma$ sends 1 to 6.
Next we consider 6, and say that ρ sends 6 to 5 and σ sends 5 to 4, so $\rho\sigma$ sends 6 to 4.
Next we consider 4, and say that ρ sends 4 to 1 and σ sends 1 to 3, so $\rho\sigma$ sends 4 to 3.
Next we consider 3, and say that ρ sends 3 to 6 and σ sends 6 to 1, so $\rho\sigma$ sends 3 to 1.
Thus we get that (1643) is a cycle in $\rho\sigma$.
Now we consider 2, and say that ρ sends 2 to 4 and σ sends 4 to 5, so $\rho\sigma$ sends 2 to 5.
Next we consider 5, and say that ρ sends 5 to 3 and σ sends 3 to 2, so $\rho\sigma$ sends 3 to 1.
Thus we get that (25) is a cycle in $\rho\sigma$.
Hence,

$$\rho\sigma = (1643)(25).$$

Let's do $\sigma\tau$ too to help us to get used to this. First we can write out the cycle notation of σ and τ next to each other to denote their composition.

$$\sigma\tau = (1326)(45)(153)(46)$$

Then going from left to right we say:
1 goes to 3 goes to 1.
So (1) is a cycle in $\sigma\tau$.
2 goes to 6 goes to 4.

4 goes to 5 goes to 3.
3 goes to 2.
So (243) is a cycle in $\sigma\tau$.
5 goes to 4 goes to 6.
6 goes to 1 goes to 5.
So (56) is a cycle in $\sigma\tau$.
Hence,

$$\sigma\tau = (243)(56),$$

where we omit the 1-cycle (1).
As a last example on composing in cycle notation we'll do $\tau\sigma$. First write them next to each other to denote their composition.

$$\tau\sigma = (153)(46)(1326)(45).$$

Then going from left to right we say:
1 goes to 5 goes to 4.
4 goes to 6 goes to 1.
So (14) is a cycle in $\sigma\tau$.
2 goes to 6.
6 goes to 4 goes to 5.
5 goes to 3 goes to 2.
So (265) is a cycle in $\sigma\tau$.
3 goes to 1 goes to 3.
So (3) is a cycle in $\sigma\tau$.
Hence,

$$\sigma\tau = (14)(265).$$

Note that $\tau\sigma \neq \sigma\tau$, so that composition of permutations is not commutative.

The last thing we'll do in this example is to work out ρ^{-1}, σ^{-1} and τ^{-1}.
To work out ρ^{-1}. We say:
Well 1 is the image of 4 under ρ, so ρ^{-1} sends 1 to 4.
Next we say that 4 is the image of 2 under ρ, so ρ^{-1} sends 4 to 2.
Next we say that 2 is the image of 1 under ρ, so ρ^{-1} sends 2 to 1.
Thus (142) is a cycle in ρ^{-1}. Similarly we obtain that (356) is a cycle in ρ^{-1}. Hence,

$$\rho^{-1} = (142)(356).$$

Note that $(142) = (421)$, because we can change which element we write first in the cycle, and similarly $(356) = (563)$. Therefore, $\rho^{-1} = (421)(563)$. So that we obtain ρ^{-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 permutation, you should convince yourself of this. In particular, we obtain

$$\sigma^{-1} = (6231)(54) = (1623)(45) \quad \text{and} \quad \tau^{-1} = (351)(64) = (135)(46).$$

From composing permutations in (b) of this example you will possibly get some idea of why we write the permutations on the right and compose from left to right. We read English from left to right so when we write permutations in cycle notation next to each other, it is more natural for us to read them from left to right.

4.2 The definition of a group

We begin with the definition of a group. For the definition recall that binary operations are defined in Definition 1.1.

Definition 4.2. 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)

We give some remarks about the definition of a group below. Many of the remarks are similar to things we have seen about rings. You shouldn't spend too long looking at these now, as they'll make more sense once you've seen some examples of groups.

- 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.
- 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. It is possible that this definition will be on the exam.**
- The binary operation $*$ is often called *multiplication*, but we will see in the examples in Section 4.3 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, analogous to that from the addition and multiplication in rings from Lemma 1.24, which can be proved in exactly the same way. Therefore, 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.
- 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 it 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.

- Let $g, h \in G$. Then $(gh)^{-1} = h^{-1}g^{-1}$. This can be proved as follows.
  We have $(gh)(h^{-1}g^{-1}) = g(hh^{-1})g^{-1} = geg^{-1} = gg^{-1} = e$, and similarly $(h^{-1}g^{-1})(gh) = e$. Therefore, $(gh)^{-1} = h^{-1}g^{-1}$.
- Let $g \in G$. Then $(g^{-1})^{-1} = g$. This can be proved as follows.
  We have $gg^{-1} = 1 = g^{-1}g$, and this shows that g is the inverse of g^{-1}. Therefore, $(g^{-1})^{-1} = g$.
- Let $g, x, y \in G$. If $gx = gy$, then $x = y$. This can be proved as follows.
  Suppose $gx = gy$. Then multiplying on the left by g^{-1} gives $g^{-1}gx = g^{-1}gy$. We have $g^{-1}gx = ex = x$ and similarly $g^{-1}gy = ey = y$. Therefore, $x = y$.
  Also we can prove similarly that if $xg = yg$, then $x = y$.

Now we move on to the definition of an abelian group.

Definition 4.3. 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*.

We now give the definition of the order of a finite group.

Definition 4.4. Let G be a group. Suppose G has a finite number of elements, then the *order of* G is defined to be the number of elements of G, and is denoted by $|G|$.

In case G is a group with infinitely many elements, we allow ourselves to write $|G| = \infty$, as this is a convenient convention.

It is a good idea for you to compare axioms (G0)–(G3) for a group with axioms (A0)–(A3) for a ring, and to compare axioms (G4) and (A4). You should observe that they are exactly the same but with $*$ and $+$ interchanged and e and 0 interchanged. Also it means that some of the proofs about properties of addition in a ring from Section 1.3 can immediately be adapted to proofs about groups.

4.3 Examples of groups

Let's have a look at some examples of groups. This should give us a feeling of how groups show up all over the place.

4.3.1 Additive group of a ring

Let R be a ring. Then as observed at the end of the previous section, the axioms (A0)–(A4) for R tell us that R with the binary operation $+$ is an abelian group. We could write this as $(R, +)$ is an abelian group, or that R is a group under addition. This is known as the *additive group of R*.

Therefore, we get lots of examples of groups just by taking any ring and considering it as a group under addition. For example, $\mathbb{Z}$ is a group under addition, and $\mathbb{Z}_n$ is a group under addition, for $n \in \mathbb{N}$.

4.3.2 Multiplicative group of units of a ring with one

For this example we recall that units in a ring with one were defined in Definition 1.7, and that $U(R) = \{a \in R : a \text{ is a unit}\}$ We claim that $U(R)$ is a group with the binary operation $\cdot$ is a group, where $\cdot$ is the multiplication in R, which we state and prove in the lemma below. This allows us to call $U(R)$ *the group of units of* R.

Lemma 4.5. *Let R be a ring with one. Then $(U(R), \cdot)$ is a group.*

Proof. We need to check the axioms, which we do in turn.
(G0). Let $u, v \in U(R)$. Then we have $(uv)(v^{-1}u^{-1}) = u(vv^{-1})u^{-1} = u1u^{-1} = uu^{-1} = 1$. Similarly, we have $(v^{-1}u^{-1})(uv) = 1$. Therefore, $uv \in U(R)$.

(G1). This follows from axiom (M1), which holds in R.

(G2). This follows from axiom (M2), which holds in R. So we have $e = 1$.

(G3). By definition of $U(R)$, we have that for $u \in U(R)$, there exists $u^{-1} \in U(R)$ such that $uu^{-1} = 1 = u^{-1}u$. □

We note that when R is commutative then $U(R)$ is an abelian group. When R is noncommutative it is possible for $U(R)$ to be either abelian or nonabelian, but we won't go in to that here.

As a particular example, let $\mathbb{F}$ be a field. Then $U(\mathbb{F}) = \mathbb{F} \setminus \{0\}$ is a group under multiplication. We recall that we often use the notation $\mathbb{F}^\times = U(\mathbb{F}) = \mathbb{F} \setminus \{0\}$.

4.3.3 Groups of matrices

We consider a particular case of a multiplicative group of units in a ring. Let $\mathbb{F}$ be a field, $n \in \mathbb{N}$ and $R = M_n(\mathbb{F})$ be the ring of $n \times n$ matrices over $\mathbb{F}$. We have that $U(R)$ is the set of invertible matrices $A \in R$, i.e. the set of matrices that have an inverse $A^{-1} \in R$, and this inverse satisfies $AA^{-1} = 1 = A^{-1}A$, where 1 denotes the identity matrix. Therefore, $U(R)$ is precisely the set of invertible $n \times n$ matrices with entries in $\mathbb{F}$, and it is a group under multiplication as proved above. We use the notation $\mathrm{GL}_n(\mathbb{F}) = \{A \in M_n(\mathbb{F}) : A \text{ is invertible}\}$ for this group, and refer to it as the *general linear group of degree n over $\mathbb{F}$*.

We just give some quick comments about determinants of matrices. We are familiar with the determinant of a matrix with entries in $\mathbb{R}$ or $\mathbb{C}$, and know how to calculate it. In fact, we can define the determinant $\det(A)$ of $A \in M_n(\mathbb{F})$ when $\mathbb{F}$ is any field. Also we can check that the formula $(\operatorname{adj} A)A = \det(A)1 = A(\operatorname{adj} A)$, where 1 is the identity matrix and $(\operatorname{adj} A)$ is the adjugate matrix of A, is also valid over any field $\mathbb{F}$. Thus we see that A is invertible if and only if $\det(A) \neq 0$. Thus $\mathrm{GL}_n(\mathbb{F}) = \{A \in M_n(\mathbb{F}) : \det(A) \neq 0\}$. Don't worry if this doesn't sink in straightaway, for what we'll do with $\mathrm{GL}_n(\mathbb{F})$ in this course you'll be fine to just think about the cases $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

4.3.4 Groups of permutations

We state a proposition about functions that collects a few useful properties of functions. These should all be familiar to you from other courses that you have taken; if not then you can look back at Appendix B in your 1AC notes, where this is covered.

Proposition. *Let A, B and C 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 $f \circ f^{-1} = \mathrm{id}_B$.*

Let Ω be a set. We recall from Section 4.1 that $\mathrm{Sym}(\Omega)$ is set of all permutations of Ω and we call it the symmetric group on Ω. Composition of functions gives a binary operation on $\mathrm{Sym}(\Omega)$, and as proved below $\mathrm{Sym}(\Omega)$ is a group under composition. This is essentially of the proposition above.

Proposition 4.6. *Let Ω be a set. Then $(\mathrm{Sym}(\Omega), \circ)$ is a group under composition of functions.*

Proof. We need to check the axioms, which we do in turn.
(G0). Let $f, g \in \mathrm{Sym}(\Omega)$. Then by (a) of the above proposition, we have $f \circ g \in \mathrm{Sym}(\Omega)$. Thus (G0) is true.
(G1). Let $f, g, h \in \mathrm{Sym}(\Omega)$. Then by (b) of the proposition above, we have $(f \circ g) \circ h = f \circ (g \circ h)$. Thus (G1) is true.
(G2). Let $f \in \mathrm{Sym}(\Omega)$. Then $f \circ \mathrm{id}_\Omega = f = \mathrm{id}_\Omega \circ f$, by (c) of the proposition above, and clearly we have $\mathrm{id}_\Omega \in \mathrm{Sym}(\Omega)$. Thus (G2) is true, where $e = \mathrm{id}_\Omega$.
(G3). Let $f \in \mathrm{Sym}(\Omega)$. Then $f^{-1} \in \mathrm{Sym}(\Omega)$ and $f \circ f^{-1} = \mathrm{id}_\Omega = f^{-1} \circ f$ by (d) of the proposition above. Hence, (G3) is true. □

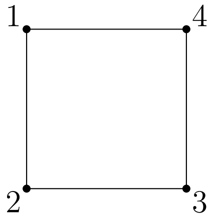
We recall also that in the case where $\Omega = \{1, 2, \dots, n\}$, we usually write S_n instead of $\mathrm{Sym}(\Omega)$, and refer to S_n as *the symmetric group of degree n*. We have $|S_n| = n!$.

From some of the calculations we have done in Examples 4.1, we see that S_6 is not abelian. In fact, we have that S_n is nonabelian for all $n \geq 3$.

4.3.5 Groups from symmetry in geometry

We look at how groups arise from symmetries of geometric shapes. We just do this by looking at the example of the symmetries of a square. Note that we can do this more generally for other shapes too.

Example 4.7. Consider a square in the plane



with vertices labelled by $\Omega = \{1, 2, 3, 4\}$. A symmetry of this square can be thought of as a permutation of the vertices of the square that preserves distances, so we can use elements of S_n to denote these symmetries. We'll see later that these symmetries give a group under composition. This called is denoted by D_8 and referred to as the *dihedral group of order 8.*

There are 8 symmetries of the square, so $|D_8| = 8$. These symmetries are the identity, three rotations and four reflections. These are given by the following elements of S_n, which are given in both two-row and cycle notation.

- $e =$ do nothing.
- $\begin{aligned}\rho_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = (1\,2\,3\,2) \\ &= \text{an anticlockwise rotation through } \tfrac{\pi}{2} \text{ radians}\end{aligned}$
- $\begin{aligned}\rho_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (1\,3)(2\,4) \\ &= \text{an anticlockwise rotation through } \pi \text{ radians}\end{aligned}$
- $\begin{aligned}\rho_3 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} = (1\,4\,3\,2) \\ &= \text{an anticlockwise rotation through } \tfrac{3\pi}{2} \text{ radians}\end{aligned}$
- $\begin{aligned}\sigma_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (1\,2)(3\,4) \\ &= \text{a reflection in the vertical axis}\end{aligned}$
- $\begin{aligned}\sigma_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = (1\,4)(2\,3) \\ &= \text{a reflection in the horizontal axis}\end{aligned}$
- $\begin{aligned}\sigma_3 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} = (2\,4) \\ &= \text{a reflection in the 1–3 diagonal}\end{aligned}$
- $\begin{aligned}\sigma_4 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = (1\,3) \\ &= \text{a reflection in the 2–4 diagonal}\end{aligned}$

Note that in σ_3 and σ_4, we have omitted writing the 1-cycles in the cycle notation, as mentioned after Examples 4.1.

We can look at symmetries over other shapes in 2-dimensions and also in 3-dimensions. We'll possibly give some other examples in the lectures.

4.3.6 Groups of automorphisms

We looked at homomorphisms and isomorphisms of rings in Section 2.1. Next we say what an automorphism of a ring is.

Let R be a ring. An isomorphism $\theta : R \to R$ is called an *automorphism.* We define

$$\mathrm{Aut}(R) = \{\theta : \theta \text{ is an automorphism of } R\}$$

and refer to $\operatorname{Aut}(R)$ as the *automorphism group of R*.

The elements of $\operatorname{Aut}(R)$ are isomorphisms of R, so that $\operatorname{Aut}(R)$ is a subset of $\operatorname{Sym}(R)$. The best way to prove that $\operatorname{Aut}(R)$ is a group is by showing that it is a subgroup of $\operatorname{Sym}(R)$. This is left as an exercise that you'll be able to do after we have covered subgroups.

Let's have a couple of quick examples of automorphism groups.

Let $R = \mathbb{Z}$ and let θ be an automorphism of $\mathbb{Z}$.
We have $\theta(1) = \theta(1^2) = \theta(1)^2$, so that $\theta(1)$ is equal to either 0 or 1. Since θ is an automorphism, we cannot have $\theta(1) = 0 = \theta(0)$, otherwise θ is not injective. So we must have $\theta(1) = 1$. We leave it as an exercise to prove that for $n \in \mathbb{Z}$, we have $\theta(n) = n\theta(1)$. Therefore, we have $\theta(n) = n$ for all $n \in \mathbb{Z}$, so that $\theta = \mathrm{id}_R$.

Hence, $\operatorname{Aut}(R) = \{\mathrm{id}_R\}$.

Let $R = \mathbb{Z}[i]$ be the ring of Gaussian integers, and let θ be an automorphism of R. As in (a) we have $\theta(n) = n$ for all $n \in \mathbb{Z}$. Therefore, we have $\theta(a + bi) = \theta(a) + \theta(b)\theta(i) = a + b\theta(i)$ for all $a, b \in \mathbb{Z}$. Also we have $\theta(i)^2 = \theta(i^2) = \theta(-1) = -1$. Thus $\theta(i)$ must be equal to either i or $-i$, because these are the only square roots of -1 in R. In the case $\theta(i) = i$ we get $\theta = \mathrm{id}_R$. In the case $\theta(i) = -i$, we get $\theta(a + ib) = a - ib$, so that θ is complex conjugation, and we'll denote this by γ.

Hence, $\operatorname{Aut}(R) = \{\mathrm{id}_R, \gamma\}$, where $\gamma : R \to R$ is complex conjugation.

A particularly interesting class of automorphism groups, are the automorphism groups of fields. This is a central topic in Galois theory, which you will be able to learn about next year. We have covered the some of theory of fields in Chapter 3, and this is used when determining the automorphism groups of fields.

4.3.7 The trivial group

Let's give one more example of a group. This is the group with one element $G = \{e\}$, and is called the *trivial group*. We must have $e * e = e$, and this determines the multiplication on G. The axioms for a group hold trivially; you should think about this for long enough to convince yourself.

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.

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 the multiplicative group of units $U(R)$ of a ring R with one, then the identity element is $1 \in R$, and we write 1 rather than e; and in the additive group of a ring R the identity is $e = 0$.

4.4 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.

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 below. It can be proved in exactly the same way as it would be proved for powers of numbers.

For $r, s \in \mathbb{Z}$, we have

(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 4.8. Let G be a group and $g \in G$.

(a) 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 $o(g)$.
(b) If no such m exists, then g is said to be of *infinite order*. We allow ourselves to write $o(g) = \infty$, as this is a convenient convention.

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 next have a lemma about orders.

Lemma 4.9. *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 $e = g^0, g = g^1, g^2, g^3, \dots, g^n$ 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 $g^a = g^{qm+r} = (g^m)^q g^r = e^q g^r = eg^r = g^r$.

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 end this section by giving some examples of the orders of elements of some groups.

We have that $[1]_n \in \mathbb{Z}_n$ has order n in $\mathbb{Z}_n$. It is good exercise to work out the orders of the other elements of $\mathbb{Z}_n$.

Given an element $\sigma \in S_n$ written in cycle notation, we can prove that the order of σ in the least common multiple of the lengths of all the cycles in the cycle notation, though we do not include a proof here. For example $(12)(345) \in S_5$ has order 6, also $(123)(4567) \in S_7$ has order 12 and $(12)(345)(6789X) \in S_{10}$ has order 30 (here we have used the symbol X rather than writing 10, as writing (12)(345)(678910) would be confusing).

4.5 Subgroups and cyclic groups

We move on to consider subgroups of a group. You should observe that much of the exposition here is very similar to what we saw for subrings earlier in Section 1.4. We begin with the definition of a subgroup.

Definition 4.10. 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 sometimes 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 the first subgroup test, which we state below.

Lemma 4.11 (First 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 $hk \in H$ and $h^{-1} \in H$.*

We can do a bit better than this, and reduce the amount that we have to check. This gives the second subgroup test, which we state in the next lemma.

Lemma 4.12 (Second 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 $hk^{-1} \in H$.*

The proof of the second subgroup test is similar to the proof of the second subring test, so we leave it as an exercise.

A useful remark here is that for a subgroup H of a group G and an element $g \in H$, we have $g^m \in H$ for all $m \in \mathbb{Z}$; or in other words H contains all of the powers of g. To see this first note that H contains $gg = g^2$ and $g^2 g = g^3$, and we can continue in this way to see that $g^m \in H$ for all $m \in \mathbb{N}$. Also we must have $e = g^0 \in H$, and the inverse of g namely $g^{-1} \in H$. Then $g^{-1}g^{-1} = g^{-2}$ and continuing in this way we can show that $g^{-m} \in H$ for all $m \in \mathbb{N}$.

Next we define the subgroup containing just the powers of a given a given element.

Definition 4.13. 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, which is part of the following lemma. The proof is left as an exercise. For (b), you'll want to show that $\langle g \rangle = \{1, g, g^2, \dots, g^{o(g)-1}\}$.

Lemma 4.14. *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*; and*
(b) $|\langle g \rangle| = o(g)$.

Moreover, if H is a subgroup of G and $g \in H$. Then $\langle g \rangle \subseteq H$.

Groups that are generated by a single element are in a sense easy groups, but they are fundamental in understanding groups more generally. We define them next.

Definition 4.15. A group G is called a *cyclic group* if $G = \langle g \rangle$ for some $g \in G$.

The next lemma gives an easy characterization of cyclic groups.

Lemma 4.16. *Let G be a finite group of order n. Then G is cyclic if and only if there exists $g \in G$ of order n.*

Proof. Let $g \in G$ and suppose that g has order n. Then $|\langle g \rangle| = n = |G|$. So we must have $G = \langle g \rangle$.
Suppose that G is cyclic and let $g \in G$ with $G = \langle g \rangle$. Then $o(g) = |\langle g \rangle| = |G| = n$. □

Let's look at some examples of cyclic groups and of subgroups.

Examples 4.17. (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 $\mathbb{Z}$. Moreover, we can prove that any subgroup of $\mathbb{Z}$ is equal to $m\mathbb{Z}$ for some $m \in \mathbb{Z}$: this can be proved in more or less the same way that we prove that any ideal of $\mathbb{Z}$ is principal in Proposition 2.31.

(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 $\mathbb{Z}$ and define $m\mathbb{Z}_n = \{[ma]_n : a \in \mathbb{Z}\}$. Then we can show that $m\mathbb{Z}_n$ is a subgroup of $\mathbb{Z}_n$. Moreover, any subgroup of $\mathbb{Z}_n$ is equal to $m\mathbb{Z}_n$ for some $m \in \mathbb{N}$ such that $m \mid n$. We leave it as an exercise to check this.

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) Let $n \in \mathbb{N}$ and let $Z_n = \{z \in \mathbb{C} : z^n = 1\}$. Then we can check that Z_n is a subgroup of the multiplicative group $\mathbb{C}^\times$ of $\mathbb{C}$. Moreover, the elements of Z_n are $e^{\frac{2\pi i k}{n}}$ for

$k = 0, 1, \ldots, n-1$. So $Z_n = \langle \zeta_n \rangle$ for $\zeta_n = e^{\frac{2\pi i}{n}}$. Hence, $Z_n = \langle e^{\frac{2\pi ik}{n}} \rangle$ so Z_n is a cyclic group.
(Note here that we are temporarily reverting to using e for the real number, which is the base of the natural logarithm rather than the identity element in a group; the identity element of Z_n is 1. As you'll already have seen there are not enough letters to go around, so we often end up doing this type of thing.)

(d) We'll consider some subgroups of S_n where $n \in \mathbb{N}$. For these examples, recall that we write e for the identity in S_n.
(i) Let $H = \{\sigma \in S_n : n\sigma = n\}$ (remember that we're writing permutations on the right). We'll check that H is a subgroup of S_n using the first subgroup test.
First we note that $e \in H$, because $ne = n$.
Next let $\sigma, \tau \in H$. Then $n\sigma = n = n\tau$. Thus we have $n\sigma^{-1} = n$ by the definition of σ^{-1} so that $\sigma^{-1} \in H$. Also $n(\sigma\tau) = (n\sigma)\tau = n\tau = n$, so that $\sigma\tau \in H$.
Thus H is a subgroup of S_n by the first subgroup test.
In fact we can see that H looks just like S_{n-1}, because a permutation of $\{1, 2, \ldots, n\}$ that fixes n can be thought of as a permutation of $\{1, 2, \ldots, n-1\}$.
(ii) Now we determine all subgroups of S_3. The arguments below may seem a bit brief at first, so may take a bit of time to sink in.
First we note that $\{e\}$ is a subgroup of S_3. This is trivial to check; see (f) below for a more general statement.
Next we consider $\langle (12) \rangle$, which is a subgroup of S_3 by (c) above. (Remember that we omit the 1 cycles when we write elements of S_3.) By direct calculation we check that $(12)^m$ is equal to e if m is even and is equal to (12) if m is odd. Thus $\langle (12) \rangle = \{e, (12)\}$ is a subgroup of S_3 and $\langle (12) \rangle$ has order 2.
Similarly, we have that $\langle (13) \rangle = \{e, (13)\}$ and $\langle (23) \rangle = \{e, (23)\}$ are subgroups of S_3.
Now consider $\langle (123) \rangle$, which is a subgroup of S_3 by (c) above. By direct calculation we check that $(123)^m$ is equal to e if $m \equiv 0 \bmod 3$, is equal to (123) if $m \equiv 1 \bmod 3$, and is equal to (132) if $m \equiv 2 \bmod 3$. Therefore, $\langle (123) \rangle = \{e, (123), (132)\}$ is a subgroup of S_3 and $\langle (123) \rangle$ has order 3. Also using similar arguments we see that $\langle (132) \rangle = \langle (123) \rangle$.
Lastly we have that S_3 is a subgroup of itself. This is trivial to check; see (g) below for a more general statement.
Now we ask if there are any other subgroups. Lets first consider a subgroup H containing (12), which is not one of the subgroups that we have already written down. We have that $e \in H$ and also H contains another element different to e or (12), we'll just consider the case $(13) \in H$. But then we also have $(12)(13) = (123) \in H$, $(13)(12) = (132) \in H$ and $(12)(13)(12) = (23) \in H$. It follows that $H = S_3$, so that H is not a different subgroup to the ones that we have already written down.
There are a number of other cases to consider, but using similar arguments to those above, we can check that there are no more subgroups of S_3.

Thus we have determined all the subgroups of S_3. They are

$$\{e\}, \quad \langle (12) \rangle, \quad \langle (13) \rangle, \quad \langle (23) \rangle, \quad \langle (123) \rangle, \quad \text{and} \quad S_3.$$

They have orders 1, 2, 2, 2, 3 and 6 respectively. We note that all these orders divide $|S_3| = 6$. Later on we'll prove Lagrange's theorem, which says that the order of a subgroup H of a finite group G always divides the order of G.

(e) Let $n \in \mathbb{N}$ and let $\mathbb{F}$ be a field (you can think of the case $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ if you like, but remember that we now know lots of other fields). We look at a couple of subgroups of $\mathrm{GL}_n(\mathbb{F})$. In these examples we write I for the identity matrix in $\mathrm{GL}_n(\mathbb{F})$, which is the identity element of $\mathrm{GL}_n(\mathbb{F})$.
(i) For this example, you'll need to remember about the determinant of a matrix, and properties of the determinant of a matrix. We'll use that for $A, B \in M_n(\mathbb{F})$ we have $\det(AB) = \det(A)\det(B)$, and if A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.
We let $\mathrm{SL}_n(\mathbb{F}) = \{A \in \mathrm{GL}_n(\mathbb{F}) : \det A = 1\}$ and call it the *special linear group of degree n over* $\mathbb{F}$. We'll show that $\mathrm{SL}_n(\mathbb{F})$ is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ using the second subgroup test.
First we note that $\det(I) = 1$, so $I \in \mathrm{SL}_n(\mathbb{F})$ and $\mathrm{SL}_n(\mathbb{F}) \neq \varnothing$.
Now let $A, B \in \mathrm{SL}_n(\mathbb{F})$. Then $\det(A) = 1 = \det(B)$, so $\det(AB^{-1}) = \det(A)\det(B^{-1}) = 1\det(B)^{-1} = 1$.
Hence, $\mathrm{SL}_n(\mathbb{F})$ is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ by the second subgroup test.
(ii) For this example, you'll need to remember about the transpose of a matrix, and properties of the transpose of a matrix. We denote the transpose of $A \in M_n(\mathbb{F})$ by A^t. We'll use that for matrices $A, B \in M_n(\mathbb{F})$ we have $(AB)^t = B^tA^t$ and that if A is an invertible square matrix, then $(A^t)^{-1} = (A^{-1})^t$.
Let $\mathrm{O}_n(\mathbb{F}) = \{A \in \mathrm{GL}_n(\mathbb{F}) : A^t = A^{-1}\}$ and call it the *orthogonal group of degree n over* $\mathbb{F}$. We'll show that $\mathrm{O}_n(\mathbb{F})$ is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ using the second subgroup test.
First we note that $I^t = I$, so $I \in \mathrm{O}_n(\mathbb{F})$ and $\mathrm{O}_n(\mathbb{F}) \neq \varnothing$.
Now let $A, B \in \mathrm{O}_n(\mathbb{F}) = I^{-1}$. Then $A^t = A^{-1}$ and $B^t = B^{-1}$, so $(AB^{-1})^t = (B^{-1})^tA^t = (B^t)^{-1}A^{-1} = (B^{-1})^{-1}A^{-1} = (AB^{-1})^{-1}$ (you may need to think for a while about how we get all of the equalities here).
Hence, $\mathrm{O}(\mathbb{F})$ is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ by the second subgroup test.

(f) Let G be a group. The trivial group $\{e\}$ is a subgroup of G.
You should think about why this is a subgroup. Once you've thought about it for long enough you should hopefully see that it is trivial.

(g) Let G be a group. Then G is a subgroup of itself.
As in the previous example, you should think about why this is true, and should see that it is trivial.

Also as mentioned in §4.3.5, we'll check that the symmetries of a square give a subgroup of S_4. To do this we look at this group in more detail in the next example.

Example 4.18. Recall that the group of symmetries of the square is denoted by D_8, and that we view the elements of D_8 as elements of S_4 viewing them as permutations of the vertices of the square. The eight elements of D_8 are given in cycle notation by $e = \mathrm{id}$, $\rho_1 = (1234)$, $\rho_2 = (13)(24)$, $\rho_3 = (1432)$, $\sigma_1 = (12)(34)$, $\sigma_2 = (14)(23)$, $\sigma_3 = (24)$, $\sigma_4 = (13)$.

We begin by modifying our notation and let $\rho = \rho_1$ and $\sigma = \sigma_1$. Then we can calculate that

- $e = \rho^0$
- $\rho_1 = \rho$
- $\rho_2 = \rho^2$

- $\rho_3 = \rho^3$
- $\sigma_1 = \sigma$
- $\sigma_2 = \rho^2\sigma$
- $\sigma_3 = \rho\sigma$
- $\sigma_4 = \rho^3\sigma$

Therefore, the elements of D_8 can be written uniquely in the form $\rho^i\sigma^j$, where $i \in \{0,1,2,3\}$ and $j \in \{0,1\}$. Also we can check that $\sigma\rho = \rho^{-1}\sigma = \rho^3\sigma$, by calculating $(12)(34)(1234) = (13) = (1432)(12)(34)$. Further, we can calculate that $\rho^4 = e$ and $\sigma^2 = e$. From this we can deduce that $\sigma\rho^i = \rho^{-i}\sigma = \rho^{4-i}$ for $i = 2,3$.
You may want to think about what all this means geometrically, and this may help you to understand what is going on.

Next we'll see that we have enough information to calculate products in D_8. For example, for $g = \rho^3$ and $h = \rho^2\sigma$, we have

$$\begin{aligned} gh &= (\rho^3)(\rho^2\sigma) \\ &= \rho^5\sigma \\ &= \rho\sigma, \end{aligned}$$

and for $g = \rho\sigma$ and $h = \rho^2\sigma$, we calculate

$$\begin{aligned} gh &= \rho\sigma\rho^2\sigma \\ &= \rho\rho^{-2}\sigma\sigma \\ &= \rho\rho^2\sigma^2 \\ &= \rho^3. \end{aligned}$$

You may have to think for a bit to see where all of these steps come from. Similarly, we can calculate that all other products of elements in D_8 are in D_8, so we deduce that D_8 is closed under multiplication.

We can also work out inverses of elements $\rho^i\sigma^j$ of D_8. For example,

$$\begin{aligned} (\rho^3\sigma)^{-1} &= \sigma^{-1}(\rho^3)^{-1} \\ &= \sigma\rho^{-3} \\ &= \rho^3\sigma. \end{aligned}$$

Similarly, we can check that all other inverses of elements in D_8 are in D_8, so we deduce that D_8 is closed under taking inverses.

Thus we have seen that D_8 is closed under multiplication and taking inverses, and it contains the identity. Hence, D_8 is a subgroup of S_4 by the first subgroup test.

In the example above we have been looking at the group of symmetries of a square, which is a group of order 8. More generally, for $n \in \mathbb{N}$, we can consider the groups of symmetries of an n-gon, which is a group of order $2n$. This is an important group and so gets a name, the *dihedral group of order* $2n$, and is denoted by D_{2n}. Similarly to the example above we let $\rho \in D_{2n}$ to be a rotation of $\frac{2\pi}{n}$ and σ to be a reflection. Then show that each element of D_{2n} can be written uniquely in the form $\rho^i\sigma^j$, where

$i \in \{0, 1, 2, \ldots, n-1\}$ and $j \in \{0, 1\}$. Also we can show that $\sigma\rho = \rho^{-1}\sigma$, and that $\rho^n = e$ and $\sigma^2 = e$. So $D_{2n} = \{\rho^i\sigma^j : 0 \leq i < n, 0 \leq j \leq 1\}$, and we have $\sigma\rho = \rho^{-1}\sigma$, $\rho^n = e$ and $\sigma^2 = e$. This gives us enough information to be able to calculate in the group.

The rotations $\{\rho^i : 0 \leq i < n\} = \langle\rho\rangle$ give a cyclic subgroup of D_{2n}. This group is sometimes denoted C_n.

4.6 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 an idea of the proof here, as this involves the language of cosets, which is covered in the next section. Once we have covered cosets, we'll give a full proof of the theorem.

Theorem 4.19 (Lagrange's theorem)**.** *Let G be a finite group and let H be a subgroup of G. Then $|H|$ is a factor of $|G|$.*

Idea 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 show that $\{gH : g \in G\}$ is a partition of G. Thus 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 4.13, and we have that $|\langle g\rangle| = o(g)$ by Lemma 4.14. We thus obtain the following corollary as a consequence of Lagrange's theorem.

Corollary 4.20. *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. 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 4.21. *Let $p \in \mathbb{N}$ be a prime and let G be a finite group of order p. Then G is cyclic.*

Now move on to show that Fermat's little theorem and Euler's theorem can be deduced as consequences of Lagrange' theorem for the group of units $U(\mathbb{Z}_n)$ for $n \in \mathbb{N}$.

Let's start with the case where $n = p$ is prime. Then we know that $\mathbb{Z}_p$ is a field so that $U(\mathbb{Z}_p) = \mathbb{Z}_p \setminus \{[0]_p\}$ and $|U(\mathbb{Z}_p)| = p - 1$. We can now deduce Fermat's little theorem. As mentioned in 1AC Algebra 1 this is a beautiful theorem, and we see how Lagrange's theorem leads to a really neat proof.

Theorem 4.22 (Fermat's little theorem)**.** *Let $p \in \mathbb{N}$ be a prime, and let $a \in \mathbb{Z}$ such that a is coprime to p. Then $a^{p-1} \equiv 1 \bmod p$.*

Proof. We have $[a]_p \in U(\mathbb{Z}_p)$ and $|U(\mathbb{Z}_p)| = p - 1$. Therefore, by Corollary 4.20, we have $([a]_p)^{p-1} = [1]_p$, remember that $[1]_p$ is the identity element in $U(\mathbb{Z}_p)$. Therefore, $[a^{p-1}]_p = [1]_p$, so that $a^{p-1} \equiv 1 \bmod p$. □

Now we move on to consider $\mathbb{Z}_n$ for general $n \in \mathbb{N}$. First we consider the order of $U(\mathbb{Z}_n)$, which we denote by $\phi(n)$. The function $\phi : \mathbb{N} \to \mathbb{N}$ is called *Euler's ϕ-function.* As stated in §1.2.2, we know that $[a]_n \in \mathbb{Z}_n$ is a unit if and only if a is coprime to n. Thus $\phi(n) = |U(\mathbb{Z}_n)|$ is the number of $m \in \{0, 1, \dots, n-1\}$ such that $\mathrm{hcf}(m, n) = 1$.

Next we state and prove Euler's theorem. This theorem generalizes Fermat's little theorem, as the case $n = p$ is a prime of Euler's theorem is precisely Fermat's little theorem. The proof is basically the same as that of Fermat's little theorem, and is an application of Lagrange's theorem.

Theorem 4.23 (Euler's theorem)**.** *Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$. Suppose that a is coprime to n. Then $a^{\phi(n)} \equiv 1 \bmod n$.*

Proof. We have $[a]_n \in U(\mathbb{Z}_n)$ and $|U(\mathbb{Z}_n)| = \phi(n)$. Therefore, by Corollary 4.20, we have $([a]_n)^{\phi(n)} = [1]_n$. Hence, $[a^{\phi(n)}]_n = [1]_n$, so that $a^{\phi(n)} \equiv 1 \bmod n$. □

Now we move on to determine the value of $\phi(n)$ for $n \in \mathbb{N}$. We can calculate the first few values of Euler's ϕ-function and we present them in the table below.

n	1	2	3	4	5	6	7	8	9	10	11	12
$\phi(n)$	1	1	2	2	4	2	6	4	6	4	10	4

From this table we see that it is not the case that $\phi(mn) = \phi(m)\phi(n)$ for all $m, n \in \mathbb{N}$. However, the following proposition does hold, which tells us that such a multiplicative property holds under the assumption that m is coprime to n. We note that our proof below is a special case of the Chinese remainder theorem, but we include the details to make this section more self-contained. We use ring homomorphisms in the proof as we have this tool at our disposal; it is possible to avoid this, but the arguments become a bit more complicated.

Proposition 4.24. *Let $m, n \in \mathbb{N}$. Suppose that m is coprime to n. Then $\phi(mn) = \phi(m)\phi(n)$.*

Proof. We define the function

$$\theta : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$$

by

$$\theta([x]_{mn}) = ([x]_m, [x]_n).$$

Recall that as $\mathbb{Z}_m$ and $\mathbb{Z}_n$ are rings, then their direct product $\mathbb{Z}_m \times \mathbb{Z}_n$ is also a ring as set out in Definition 1.21. Further, we can check that θ is a (well defined) homomorphism.

Suppose that $[x]_{mn} \in \ker\theta$. Then $\theta([x]_{mn}) = ([0]_m, [0]_n)$, so $m \mid x$ and $n \mid x$. Thus, $mn \mid x$, because m is coprime to n (this was proved in 1AC Algebra 1 and you can look it up there). Therefore, $[x]_{mn} = [0]_{mn}$, so we have $\ker\theta = \{[0]_{mn}\}$ and θ is injective by Proposition 2.10. Since, $|\mathbb{Z}_{mn}| = mn = |\mathbb{Z}_m \times \mathbb{Z}_n|$, we deduce that θ is a bijection and therefore an isomorphism.

We have $U(\mathbb{Z}_m \times \mathbb{Z}_n) = U(\mathbb{Z}_m) \times U(\mathbb{Z}_n)$ (you should think about this and convince yourself that it is true). Since, θ is an isomorphism we deduce that θ restricts to a bijection between $U(\mathbb{Z}_{mn})$ and $U(\mathbb{Z}_m \times \mathbb{Z}_n)$. From this we obtain

$$\phi(mn) = |U(\mathbb{Z}_m \times \mathbb{Z}_n)| = |U(\mathbb{Z}_m) \times U(\mathbb{Z}_n)| = |U(\mathbb{Z}_m)||U(\mathbb{Z}_n)| = \phi(m)\phi(n). \qquad \square$$

In the theorem below we use the proposition above to deduce the value of Euler's ϕ-function from a prime factorization. First the next lemma gives the value of ϕ on prime powers.

Lemma 4.25. *Let $p \in \mathbb{N}$ be prime and let $s \in \mathbb{N}$. Then $\phi(p^s) = p^{s-1}(p-1)$.*

Proof. Let $m \in \{0, 1, \dots, p^s - 1\}$. Then $\operatorname{hcf}(m, p^s) = p^i$ for some $i \in \{0, 1, \dots, s-1\}$. Therefore, m is not coprime to p^s precisely when p is a factor of m. Thus, the natural numbers less than p^s that are not coprime to p^s are simply the multiples of p, and there are p^{s-1} of them. Hence,

$$\phi(p^s) = p^s - p^{s-1} = p^{s-1}(p-1). \qquad \square$$

Theorem 4.26. *Let $n \in \mathbb{N}$ with prime factorization*

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

where $p_1 < p_2 < \dots < p_k$ are primes and $s_1, s_2, \dots, s_k \in \mathbb{N}$. Then

$$\phi(n) = p_1^{s_1-1}(p_1 - 1)p_2^{s_2-1}(p_2 - 1) \cdots p_k^{s_k-1}(p_k - 1).$$

Proof. We have that $p_1^{s_1}$ is coprime to $p_2^{s_2} p_3^{s_3} \dots p_k^{s_k}$. Therefore, we have $\phi(n) = \phi(p_1^{s_1})\phi(p_2^{s_2} p_3^{s_3} \dots p_k^{s_k})$. Repeating this argument we obtain that

$$\phi(n) = \phi(p_1^{s_1})\phi(p_2^{s_2}) \cdots \phi(p_k^{s_k}).$$

We have $\phi(p_i^{s_i}) = p_i^{s_i-1}(p_i - 1)$ for each i. Therefore,

$$\phi(n) = p_1^{s_1-1}(p_1 - 1)p_2^{s_2-1}(p_2 - 1) \cdots p_k^{s_k-1}(p_k - 1). \qquad \square$$

We end this section by briefly discussing another generalization of Fermat's last theorem. To explain this we first recall that $\mathbb{Z}_p$ is a field, and we often denote it by $\mathbb{F}_p$. Now by Theorem 3.21, we know that there is a field of order p^d for any $d \in \mathbb{N}$.

Let $d \in \mathbb{N}$ and let $\mathbb{F}$ be a field of order p^d. Then $\mathbb{F}^\times$ is a group (under multiplication) of order $p^d - 1$. Therefore, we deduce from Lagrange's theorem that every $a \in \mathbb{F}^\times$ satisfies $a^{p^d-1} = 1$.

In fact we can prove that there is an element of $\mathbb{F}^\times$ of order $p^d - 1$, but this requires a bit more work. This is one of the key ingredients required to prove Theorem 3.21.

A good exercise is to prove that if $\mathbb{F}$ is any field and H is a finite subgroup of the multiplicative group $\mathbb{F}^\times$, then H is a cyclic. This implies the assertion made in the previous paragraph. We first an outline of how to do this, but there is some work involved in filling in the gaps. The first step in proving this is to show that if a and b are elements of an abelian group, such that $o(a)$ is coprime to $o(b)$, then $o(ab) = o(a)o(b)$. Now if we let m be the lowest common multiple of the orders of all of the elements of H, then there exists $h \in H$ with $o(h) = m$. Moreover, all elements of H satisfy $h^m = 1$. But the polynomial $X^m - 1 = 0$ has at most m roots in $\mathbb{F}$, and from this we can deduce that we must have $m = |H|$. In fact, we have shown that H is precisely the set of all mth roots of unity in $\mathbb{F}$ (and there are m of them).

4.7 Cosets and the proof of Lagrange's theorem

This section is included in the notes so that the details of the proof of Lagrange's theorem can be given. We'll most likely not have time to cover it in the lectures. It is not part of the syllabus and is not examinable. There is a lot more here than is really needed to give the proof, so you could skip through some parts if you just want to see the details of the proof of Lagrange's theorem.

We move on to looking at cosets of a subgroup of a group G. Later we'll see that these cosets give a partition of G. What we do here is similar to what we did when we looked at cosets of an ideal in a ring. There is a little bit of a complication here, because G may not be abelian, so we have to look at both left and right cosets.

Let's get on with it and have the definition.

Definition 4.27. Let G be a group, let H be a subgroup of G and let $g \in G$.
We define the *left coset of H with respect to g* to be $gH = \{gh : h \in H\}$, and say that g is a *coset representative* of the coset gH.
We define the *right coset of H with respect to g* to be $Hg = \{hg : h \in H\}$, and say that g is a *coset representative* of the coset Hg.

We note that a coset can have many different representatives, as we'll see.

Next we'll look at cosets in S_3 in some detail.

Example 4.28. In Examples 4.17 we determined all subgroups of $G = S_3$. We are going to consider the subgroups $H = \langle(12)\rangle = \{e, (12)\}$ and $K = \langle(123)\rangle = \{e, (123), (132)\}$.

First lets look at all the left cosets of H. The notation below may look a bit funny to start with, as we're considering cosets in $G = S_3$ and using cycle notation for elements of G, but for example $(12)H$ just means gH for $g = (12) \in G$.
We have

- $eH = \{e, (12)\} = H$;
- $(12)H = \{(12), e\} = \{e, (12\}$;
- $(13)H = \{(13), (132)\}$;
- $(23)H = \{(23), (123)\}$;
- $(123)H = \{(123), (23)\} = \{(23), (123)\}$; and
- $(132)H = \{(132), (13)\} = \{(13), (132)\}$.

So we see that

- $eH = (12)H$;
- $(13)H = (132)H$; and
- $(23)H = (123)H$.

Also we see that $H \cup (13)H \cup (23)H = G$, and that $H \cap (13)H = \varnothing$, $H \cap (23)H = \varnothing$ and $(13)H \cap (23)H = \varnothing$. Hence, $\{H, (13)H, (23)H\}$ is a partition of G.

Now lets look at the right cosets. We find that $He = H(12) = \{e, (12)\}$ and $H(13) = H(123) = \{(13), (123)\}$ and $H(23) = H(132) = \{(23), (132)\}$. Then we can see that $\{H, H(13), H(23)\}$ is a partition of G.

We observe that the left cosets are not equal to the right cosets in general. For example $(13)H \neq H(13)$ and $(23)H \neq H(23)$.

We move on to look at the cosets of K in G. For the left cosets we get

- $eK = (123)K = (132)K = \{e, (123), (132)\} = K$; and
- $(12)K = (13)K = (23)K = \{(12), (13), (23)\}$.

Thus we see that $\{K, (12)K\}$ is a partition of G.
The right cosets are

- $Ke = K(123) = K(132) = \{e, (123), (132)\} = H$; and
- $K(12) = K(13) = K(23) = \{(12), (13), (23)\}$.

Thus we observe that $gK = Kg$ for all $g \in G$ in this case, in contrast to the situation for H. So we also have that $\{K, K(12)\}$ is a partition of G.

Another thing to observe here is that all the cosets of H in G have the same size, namely $2 = |H|$; and all the cosets of K in G have the same size, namely $3 = |K|$.

We now proceed to prove some things about cosets, and we go along very similar lines to what we did with rings. The next proposition allows us to apply the theory of equivalence relations to understand cosets. You can look in Section 0.1 for a recap on equivalence relations. The proof of this proposition is similar to the proof of Proposition 2.23, we include the details as it is a key result, so it's good to see it all.

Proposition 4.29. *Let G be a group and H a subgroup of G.*

(a) *Define the relation $\sim_L$ on G by $x \sim_L y$ means $x^{-1}y \in H$. Then $\sim_L$ is an equivalence relation on G. Moreover, for $g \in G$ the equivalence class $[g]_{\sim_L}$ is equal to the left coset gH.*

(b) *Define the relation $\sim_R$ on G by $x \sim_R y$ means $xy^{-1} \in H$. Then $\sim_R$ is an equivalence relation on G. Moreover, for $g \in G$ the equivalence class $[g]_{\sim_R}$ is equal to the right coset Hg.*

Proof. We only prove (a), as the proof of (b) is entirely similar, though it is a good exercise for you to go through it.

We have to show that $\sim_L$ is reflexive, symmetric and transitive. Let $x, y, z \in G$.
We have $x^{-1}x = e \in H$, because H is a subgroup, so $x \sim_L x$. Thus $\sim_L$ is reflexive.
Now suppose $x \sim_L y$, so that $x^{-1}y \in H$. Then $y^{-1}x = (x^{-1}y)^{-1} \in H$, because H is a subgroup, so $y \sim_L x$. Thus $\sim_L$ is symmetric.
Suppose that $x \sim_L y$ and $y \sim_L z$, so that $x^{-1}y \in H$ and $y^{-1}z \in H$. Then $x^{-1}z = x^{-1}(yy^{-1})z = (x^{-1}y)(y^{-1}z) \in H$, because H is a subgroup, so $x \sim_L z$. Thus $\sim_L$ is transitive.
Hence, $\sim_L$ is an equivalence relation.

We have

$$\begin{aligned}
[g]_{\sim_L} &= \{x \in G : x \sim_L g\} \\
&= \{x \in G : g \sim_L x\} \\
&= \{x \in G : g^{-1}x \in H\} \\
&= \{x \in G : g^{-1}x = h \text{ for some } h \in H\} \\
&= \{x \in G : x = gh \text{ for some } h \in H\} \\
&= \{gh : h \in H\} \\
&= gH.
\end{aligned}$$

Hence, $[g]_{\sim_L} = gH$. □

The equivalence relation $\sim_L$ in the proposition above is sometimes called *left congruence modulo H*; so we would say x is left congruent to y modulo H if $x^{-1}y \in H$. Similarly, $\sim_R$ is sometimes called *right congruence modulo H*.

We have the following corollary of Proposition 4.29.

Corollary 4.30. *Let G be a group, H a subgroup of G and $x, y \in G$. Then*

(a)(i) *$x \in xH$;*
(ii) *$xH = yH$ if and only if $x^{-1}y \in H$ if and only if $xh = y$ for some $h \in H$;*
(iii) *$xH = yH$ or $xH \cap yH = \varnothing$.*
(iv) *$\{gH : g \in G\}$ is a partition of G.*

(b)(i) *$x \in Hx$;*
(ii) *$Hx = Hy$ if and only if $xy^{-1} \in H$ if and only if $x = hy$ for some $h \in H$;*
(iii) *$Hx = Hy$ or $Hx \cap Hy = \varnothing$.*
(iv) *$\{Hg : g \in G\}$ is a partition of G.*

Proof. Part (a) follows directly from Theorem 0.5 for the equivalence relation $\sim_L$ from Proposition 4.29(a), and that $[x]_{\sim_L} = xH$ for $x \in G$. Part (b) follows in the same way, where we use Proposition 4.29(b). □

Let G be a finite group and let H be a subgroup of G. Then $\{gH : g \in G\}$ is finite and we define $|G : H| = |\{gH : g \in G\}|$ and call it the *index* of H in G. We have defined $|G : H|$ using left cosets, but show now that we could have also defined $|G : H|$ using right cosets.

Lemma 4.31. *Let G be a group and H a subgroup of G. Then the function $f : \{xH : x \in G\} \to \{Hy : y \in G\}$ given by $f(xH) = Hx^{-1}$ is well defined and a bijection.*

Proof. First we check that f is well defined. So let $x, y \in G$ and suppose that $xH = yH$. Then we have $x^{-1}y \in H$, and thus also $y^{-1}(x^{-1})^{-1} = y^{-1}x = (x^{-1}y)^{-1} \in H$. Therefore, we obtain that $Hx^{-1} = Hy^{-1}$. Hence, f is well defined.

A very similar argument shows that f is an injection, and we leave this as an exercise. Finally, we note that f is a surjection because $Hy = f(y^{-1})$ for any $y \in G$. □

We record the following elementary but useful consequence about the sizes of cosets.

Lemma 4.32. *Let G be a group, let H be a finite subgroup of G and let $g \in G$. Then $|gH| = |H| = |Hg|$.*

Proof. Consider the function $f : H \to gH$ defined by $f(h) = gh$. By the definition of gH, we have that f is surjective. Also if $f(h) = f(k)$ for $h, k \in H$, then $gh = gk$ and multiplying by g^{-1} gives $h = k$, thus f is injective. Hence, f is a bijection and we deduce that $|gH| = |H|$.

The proof that $|H| = |Hg|$ is very similar. □

We are now ready to fill in the details of the proof of Lagrange's theorem. We can also now give a slightly better statement of it.

Theorem 4.33 (Lagrange's theorem)**.** *Let G be a finite group and let H be a subgroup of G. Then $|G| = |G : H||H|$. In particular, $|H|$ is a factor of $|G|$.*

Proof. By Corollary 4.30(d), we know that $\{gH \mid g \in G\}$ is a partition of G. Therefore, we can choose $g_1, g_2, \ldots, g_r$, where $r = |G : H|$, such that $G = g_1H \cup g_2H \cup \ldots \cup g_rH$ and $g_iH \cap g_jH = \varnothing$ for $i \neq j$. By Lemma 4.32, we have $|g_iH| = |H|$ for each i. Therefore,

$$\begin{aligned} |G| &= \sum_{i=1}^{|G:H|} |g_iH| \\ &= \sum_{i=1}^{|G:H|} |H| \\ &= |G : H||H|. \end{aligned}$$

□

4.8 More about group theory

That's it for group theory in this course, but I'll just give a very brief overview of what the next steps are, and then mention further some topics, which are fascinating. There is a very interesting course which delves much further in to group theory that you can take next year.

A *homomorphism* of groups is defined to be a function $\theta : G \to H$ between groups G and H such that $\theta(gh) = \theta(g)\theta(h)$ for all $g, h \in G$. An *isomorphism* is defined to be a bijective homomorphism. Two groups are said to be *isomorphic* if there is an isomorphism between, and we can think of isomorphic groups as essentially being the same, just that their elements have different names. For example, we could prove that any two cyclic groups of the same order are isomorphic.

We can also go on to define quotient groups. We can't take the quotient by any subgroup, and there is special type of subgroup called a normal subgroup. A subgroup N of a group G is said to be *normal* if $gN = Ng$ for all $g \in G$, equivalently $gng^{-1} \in N$ for all $g \in G$ and $n \in N$. Then we can proceed to define a quotient group G/N, when N is a normal subgroup, similarly to what we did with quotients of rings by ideals.

The *kernel* of a homomorphism $\theta : G \to H$ is defined to be $\ker\theta = \{g \in G : \theta(g) = e\}$, and can be shown to be a normal subgroup of G. The *image* of θ is $\operatorname{im}\theta = \{\theta(g) : g \in G\}$, and can be shown to be a subgroup of H. Then we get an isomorphism theorem for groups, which says that $G/\ker\theta \cong \operatorname{im}\theta$.

In the group theory course that you can take next year, you'll learn a lot more about the structure of finite groups and how they can be thought of as symmetries of mathematical and geometric objects. Then you'll move on to study some Galois theory, which I mention briefly below.

I'll briefly explain a really interesting part of the structure theory of finite groups. It can be shown that, in a sense, any finite group can be built up from irreducible pieces called *simple groups*; this is similar to how any integer can be factorized as a product of prime numbers. One of the biggest achievements in mathematics research in the second half of the 20th century was the classification of finite simple groups (up to isomorphism). A lot of the research for this project was done here at the University of Birmingham, and there is lots of research in group theory nowadays both here and all around the world.

The books

- Marcus Du Sautoy, *Finding Moonshine: A Mathematician's Journey Through Symmetry*
- Mark Ronan, *Symmetry and The Monster*

give an accessible account of the work that went in to the classification of finite simple groups.

A particularly interesting application of group theory and its interaction with the theory of fields, is the fascinating subject of Galois theory. Whenever we have a polynomial $f(X) \in \mathbb{Q}[X]$, we can associate a group $G = \operatorname{Gal}(f(X))$ known as the Galois group of $f(X)$. To do this we let $\mathbb{F}$ be the smallest subfield of $\mathbb{C}$ containing all the roots of $f(X)$, and then we let G be the automorphism group of $\mathbb{F}$. We can think of G as being the symmetries of the roots of the polynomial $f(X)$, because we can show that any element of G permutes the roots of $f(X)$.

A really striking consequence of Galois theory is that there is not a formula that gives the roots of a quintic polynomial. We give a very brief explanation of this below. There is an alternative brief account of this in the 1AC Algebra 1 notes.

There are formulas giving the roots of polynomials of degree 4 or less. It is possible to show that there is a formula giving the roots of $f(X)$ if and only if $G = \operatorname{Gal}(f(X))$ is what is known as a *soluble group*, which very roughly means that it has an "uncomplicated structure". When $\deg f(X) \leq 4$, then G must be soluble. However, there are quintic polynomials that have a non-soluble Galois group, so there is no formula giving their solutions.

4.9 Summary of Chapter 4

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

- calculate in S_n using cycle notation;
- state the definition of a group, including all axioms;
- check whether the axioms of a group hold in examples;
- define and understand some examples of groups and be able to calculate in them;
- define the order of a group and of a group element, and calculate it in examples;
- state the definition of a subgroup, and state, prove and apply the subgroup tests; and
- define the subgroup generated by an element and define cyclic groups.

4.10 Exercises for Chapter 4

Q4.1. Determine the cycle notation for the following permutations

(a)
$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 6 & 4 & 9 & 8 & 2 & 1 & 3 & 7 & 5 \end{pmatrix}$$

(b)
$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 4 & 5 & 8 & 9 & 6 & 1 & 3 & 7 \end{pmatrix}$$

(c)
$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 6 & 9 & 3 & 7 & 1 & 5 & 8 & 4 \end{pmatrix}$$

Q4.2. In this question you have to compose elements and calculate inverses in S_n using cycle notation. As in the lectures and the notes we omit writing $\circ$ to denote composition.
Remember that we write permutations in S_n on the right.
So we compose them from left to right, so that $\rho\sigma$ means do ρ and then σ.
Let
$$\rho = (142)(35) \quad \text{and} \quad \sigma = (15)(34)(2) \quad \text{and} \quad \tau = (152)(34)$$
be permutations in S_5 in cycle notation.

Calculate the following permutations giving your solution in cycle notation.

(a) $\rho\sigma$ (e) σ^3
(b) $\rho\tau$ (f) ρ^{-1}
(c) $\sigma\tau$ (g) σ^{-1}
(d) τ^2 (h) τ^{-3}

Powers of elements of S_n are defined as in Section 3.3 of the notes.

Q4.3. Which of the following are groups with the given binary operation $*$.

(a) $\mathbb{Z} \setminus \{0\}$ with $x * y = xy$.
(b) $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ with $x * y = xy$.
(c) $\mathbb{R}_+$ with $x * y = x^y$.
(d) $\mathbb{R}_+$ with $x * y = x^{\log_2 y}$.
(e) $\{z \in \mathbb{C} : |z| = 1\}$ with $x * y = xy$.
(f) $\mathbb{Z}$ with $x * y = x - y$.
(g) $\mathbb{R}$ with $x * y = x^2y^2$.
(h) $\mathbb{R}^3$ with $x * y = x \times y$, where $\times$ denotes the cross product of vectors as in §1.2.9.

You should justify your answers.

Q4.4. Let G be a group.

(a) Suppose that $g^2 = e$ for all $g \in G$. Prove that G is abelian.
(b) Prove that G is abelian if and only if $(gh)^2 = g^2h^2$ for all $g, h \in G$.

Q4.5. Prove the second subgroup test.

Lemma 4.34 (Second subgroup test)**.** *Let G be a group and let H be a subset of G. Then H is a subgroup of G provided*

- *H is nonempty; and*
- *for all $h, k \in H$, we have $hk^{-1} \in H$.*

Q4.6. Let G be a group. The *centre of* G is $Z(G) = \{g \in G : xg = gx \text{ for all } x \in G\}$.

(a) Prove that $Z(G)$ is a subgroup of G.
(b) Show that $Z(S_3) = \{\text{id}\}$.

Q4.7. Which of the following subsets H of $G = \mathrm{GL}_2(\mathbb{C})$ are subgroups of G.

(a) $H = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{C},\ ac \neq 0 \right\}$.

(b) $H = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a, b \in \mathbb{C},\ ab \neq 0 \right\} \cup \left\{ \begin{pmatrix} 0 & c \\ d & 0 \end{pmatrix} : c, d \in \mathbb{C},\ cd \neq 0 \right\}$.

(c) $H = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{C},\ ac \neq 0 \right\} \cup \left\{ \begin{pmatrix} r & 0 \\ s & t \end{pmatrix} : r, s, t \in \mathbb{C},\ rt \neq 0 \right\}$.

(d) $H = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} : a, b, c \in \mathbb{C},\ ac - b^2 \neq 0 \right\}$.

(e) $H = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b, c \in \mathbb{C},\ a^2 - b^2 \neq 0 \right\}$.

(f) Let $\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, $\mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\mathbf{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$, and let $H = \{\pm\mathbf{1}, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\}$.

You should justify your answers, which means that you should give either a proof or a counterexample.

Q4.8. Investigate the orders of elements in the additive group of $\mathbb{Z}_n$.

First work out the order of all the elements of $\mathbb{Z}_n$ for small n: doing $n \leq 10$ should be more than enough.
Then you should be able to make a conjecture on the order of the elements.
Then you can try to prove your conjecture. Bézout's lemma may be useful for your proof.
You'll need to be a bit careful to untangle notation as $\mathbb{Z}_n$ is written additively, and the definition of the order of an element is given multiplicatively.

Q4.9. Let G be a group and let $g \in G$. Suppose that g has finite order. Prove that $|\langle g \rangle| = o(g)$.

Q4.10. Let G be a group and let H and K be subgroups of G. Prove that $H \cap K$ is a subgroup of G.

Q4.11. Prove the following subgroup test for finite groups.

Lemma 4.35 (Subgroup test for finite groups)**.** *Let G be a finite group and let H be a subset of G. Then H is a subgroup of G provided*

- *H is nonempty; and*
- *for all $h, k \in H$, we have $hk \in H$.*

Hint: Start by showing that if $h \in H$, then $h^m \in H$ for all $m \in \mathbb{N}$. Then use the fact that any $h \in H$ has finite order to deduce that $e \in H$ and $h^{-1} \in H$, and apply the first subgroup test.

Q4.12. Prove the following lemma.

Lemma 4.36. *Let R be a ring. Then* $\mathrm{Aut}(R)$ *is a subgroup of* $\mathrm{Sym}(R)$. *In particular,* $\mathrm{Aut}(R)$ *is a group.*

Q4.13. Determine the orders of the following elements of S_7.

(a) (12345)
(b) (1234567)
(c) $(12)(345)$
(d) $(12)(3456)$
(e) $(123)(456)$
(f) $(12)(34567)$
(g) $(123)(4567)$
(h) $(12)(34)(567)$

Q4.14. Let H be a subgroup of S_3 and suppose that (12) and (123) are elements of H. Prove that $H = S_3$.
You should do this without using Lagrange's theorem.

Q4.15.

(a) Prove that S_n has order $n!$.
(b) Let $H = \{\sigma \in S_n : n\sigma = n\}$, which we have seen is a subgroup of S_n.
  (i) Determine the order of H and the index of H in S_n.
  (ii) For $n = 4$ find $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in S_n$ such that $S_n = \sigma_1 H \cup \sigma_2 H \cup \sigma_3 H \cup \sigma_4 H$.

(c) Prove that $V_4 = \{e, (12)(34), (13)(24), (14)(23)\}$ is a subgroup of S_4 and that V_4 is not cyclic.

Q4.16.

(a) Determine the order of $\begin{pmatrix} 2 & 2 \\ 2 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_5)$.

(e) Determine the order of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_7)$.

Q4.17. Let p be a prime.

(a) Determine the order of $\mathrm{GL}_2(\mathbb{F}_p)$.
To do this let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(|F_p)$. Consider the number of possibilities for the first column $\begin{pmatrix} a \\ c \end{pmatrix}$, which has to be a nonzero vector. Then the second column $\begin{pmatrix} a \\ c \end{pmatrix}$ has to be linearly independent of the first column, so how many possibilities are there?

(b) Determine the order of $\mathrm{SL}_2(\mathbb{F}_p)$ and the index $|\mathrm{GL}_2(\mathbb{F}_p) : \mathrm{SL}_2(\mathbb{F}_p)|$ of $\mathrm{SL}_2(\mathbb{F}_p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$.
You can proceed as in (a), but you'll need to think a bit more about the number of possibilities for the second column.

(c) Let $\mathrm{B}_2(\mathbb{F}_p) = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_p) \ : \ a, b, c \in \mathbb{F}_p,\ ac \neq 0 \right\}$. You may assume that $\mathrm{B}_2(\mathbb{F}_p)$ is a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Determine the order of $\mathrm{B}_2(\mathbb{F}_p)$ and the index $|\mathrm{GL}_2(\mathbb{F}_p) : \mathrm{B}_2(\mathbb{F}_p)|$ of $\mathrm{B}_2(\mathbb{F}_p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$.

Q4.18. Recall that D_8 is the group of symmetries of the square and that it can be written as $\{\rho^i \sigma^j : 0 \leq i < n, 0 \leq j \leq 1\}$, where we have $\sigma\rho = \rho^{-1}\sigma$, $\rho^n = e$ and $\sigma^2 = e$.
(In terms of permutation we can take $\rho = (1432)$ and $\sigma = (12)(34)$).

(a) Let $H = \langle \sigma \rangle$ and $K = \langle \rho \rangle$.

   (i) Determine the left cosets of H in D_8.

   (ii) Determine the left cosets of K in D_8.

(b) Determine all the subgroups of D_8.

In this question you can either work with elements of D_8 written in the form $\rho^i\sigma^j$ or writing then as permutations in cycle notation.
Part (b) is a fair amount of work, but we have seen some similar examples. You should use Lagrange's theorem for this. It may help you to start by writing out all the elements of D_8 and their orders.

Q4.19. Prove the Corollary 4.21 from the lecture notes.

Corollary 4.37. *Let $p \in \mathbb{N}$ be a prime and let G be a finite group of order p. Then G is cyclic.*

Q4.20. Let G be a finite group, let H be a subgroup of G and let K be a subgroup of H, so that K is also a subgroup of G. Prove that $|G : K| = |G : H||H : K|$.

Q4.21. Let $n \in \mathbb{N}$. In this question we'll introduce the *alternating groups of degree n* which is denoted A_n and is a subgroup of S_n.

(a) Let $m \in \mathbb{N}$. Show that the m-cycle $(12 \ldots m)$ is equal to the product of transpositions $(1m)(2m) \ldots (m-1, m)$.
You may want to consider this for specific values of m at first to see what is going on.

(b) Deduce that any $\sigma \in S_n$ can be written as a product of transpositions.
You should first write σ in cycle notation and then use (a) to deduce that each of the cycles can be written as a product of transpositions.
Let $\sigma \in S_n$. Then by (b) we know that σ can be written as a product of transpositions, but it can be written as a product of transpositions in many different ways. The number of transpositions in two different ways of writing σ as a product of transpositions may differ. However, it can be proved that the parity (even or odd) of the number of transposition does not depend on the choice of product. You are not asked to prove this, but you may be interested in looking up a proof.

(c) Let A_n be the set of all σ in S_n that can be written as a product of an even number of transpositions. Show that A_n is a subgroup of S_n.

(d) Show that $S_n = A_n \cup (12)A_n$ and deduce that the index of A_n in S_n is 2.

Q4.22. In this question we'll determine the orders of elements of S_n, where $n \in \mathbb{N}$.

(a) Let $m \in \mathbb{N}$ and let $\gamma \in S_n$ be an m-cycle. Show that γ has order m.

(b) Let $l, m \in \mathbb{N}$ and let $\gamma, \delta \in S_n$ be disjoint cycles where γ is an l-cycle and δ is an m-cycle. Prove that the order of $\gamma\delta$ is $\operatorname{lcm}(l, m)$, where $\operatorname{lcm}(l, m)$ denotes the lowest common multiple of l and m.
Remember that $\gamma\delta = \delta\gamma$, because γ and δ are disjoint; and therefore $(\gamma\delta)^k = \gamma^k\delta^k$. Also observe that as γ and δ are disjoint, $\gamma^k\delta^k = e$ if and only if $\gamma^k = e$ and $\delta^k = e$. You may want to look back at the examples you did in Q4 to get an idea of what is going on.

(c) Let $\sigma \in S_n$ write σ in cycle notation $\sigma = \gamma_1\gamma_2 \ldots \gamma_r$, where $\gamma_1, \gamma_2, \ldots, \gamma_r$ are disjoint cycles, and let m_i be the length of m_i. Prove that the order of σ is $\operatorname{lcm}(m_1, m_2, \ldots, m_r)$.

Q4.23. This question covers a couple of things about units in rings. It is interesting, but quite challenging, so you may want to ask for a hint, or look at the hints on the 2AC Canvas page.

(a) Give an example of a ring R and an element $a \in R$ such that there exists $b \in R$ with $ab = 1$, but there is no $c \in R$ with $ca = 1$.

(b) Give an example of a noncommutative ring R such that its group of units $U(R)$ is abelian.

Chapter 5

Factorization

This chapter is not part of 2AC, but it covers some interesting ring theory, so you may be interested to read it.
It had been typed, so I thought it may as well be made available, though it is a bit brief in some places, and there may be quite a few errors and typos, and it is a bit disorganized. Please let me know if you sot any mistakes, and feel free to ask.
We don't require any of the material on Group theory from Chapter 4, so this chapter actually would be better placed straight after Chapter 2.

In 1AC Algebra 1, we encountered the fundamental theorem of arithmetic, which tells us that a natural number can be factorized as a product of primes in an essentially unique way. We also saw a similar result is true for polynomials over a field, namely that any nonzero polynomial can be factorized as a product of irreducible polynomials in an essentially unique way. We know that both the integers and the polynomials over a field are examples of rings, and it is natural to consider the question of which other rings have such a "unique factorization" property. This is the emphasis of this chapter, in which we will build the general framework of factorization in rings.

We'll begin with the required definitions, and then say what it means for an integral domain to be a unique factorization domain. After that we'll consider principal ideal domains, which we already saw in Section 5.3, and show that a principal ideal domain is a unique factorization domain. Next we'll consider Euclidean domains, which roughly speaking are integral domains in which we can divide and get a quotient and remainder. We'll see that Euclidean domains are principal ideal domains, and therefore unique factorization domains. Also we'll see a number of examples of Euclidean domains, including the integers, the Gaussian integers and the ring of polynomials over a field. Then the general theory tells us that these are all unique factorization domains. This is a really good example of the power of abstraction, as we work abstractly to prove some theorems and then can apply them in many different situations.

As a highlight we will apply this material to prove a really nice theorem called Fermat's two square theorem. It tells us that any prime that is congruent to 1 modulo 4 can be written as the sum of two squares, which is a really striking statement. We mention here that it is a theorem that any natural number can be written as the sum of four squares, and this is known as Lagrange's four square theorem. We won't cover this in the notes, but you can look it up online.

Throughout this chapter, we'll restrict to considering integral domains. Factorization isn't such a useful concept in rings which have zero divisors (you could think about why this is the case) and the situation for noncommutative rings is more complicated. So all rings in this chapter are integral domains, even if I forget to say so.

Before we start properly, we state and prove a useful lemma about integral domains.

5.1 Unique factorization

In this section we define what it means for an integral domain to be a unique factorization domain. Before we can do this, we define a few things that we'll need throughout this chapter. It is convenient for us to recall the definition of a unit here even though we have already seen this in Definition 1.7.

Definition 5.1. Let R be an integral domain, and let $a, b \in R$.

(a) We say a *is a factor of* b and write $a \mid b$ if there exists $z \in R$ such that $b = az$.

(b) We say that a is a *unit* if there exists $z \in R$ such that $az = 1$.

(c) We say that a *is associate to* b (or a and b are *associates*) if there exists a unit $u \in R$ such that $a = bu$.

(d) We say that a is *irreducible* if a is not zero and not a unit and if whenever $a = cd$ for $c, d \in R$, then either c or d is a unit.

(e) We say that a is *prime* if whenever $a \mid cd$ for $c, d \in R$, then $a \mid c$ or $a \mid d$.

Let's consider irreducible and prime elements in $\mathbb{Z}$. The definition of a prime number is essentially the same as the definition of an irreducible element of $\mathbb{Z}$, except that we don't insist that irreducible elements are positive. So the irreducible elements of $\mathbb{Z}$ are just the prime numbers and their negatives. Also we note that the units of $\mathbb{Z}$ are just 1 and -1, so that irreducible elements are the associates of the prime numbers.

We have the following theorem about prime numbers, which we saw in 1AC Algebra 1.

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

So in the terminology of Definition 5.1, we have that an irreducible element of $\mathbb{Z}$ is a prime element of $\mathbb{Z}$. In other words a prime number is a prime element of $\mathbb{Z}$, and we can also see that the negative of a prime number is also a prime element.

We now prove a lemma showing that prime elements are always irreducible.

Lemma 5.2. *Let R be an integral domain and let $p \in R$. Suppose that p is prime. Then p is irreducible.*

Proof. Let $a, b \in R$, and suppose that $p = ab$. Then since p is prime and $p \mid p = ab$, we have that $p \mid a$ or $p \mid b$. Without loss of generality suppose that $p \mid a$. Then $a = pz$ for some $z \in R$, and thus $p = pzb$, so $zb = 1$ by Lemma 1.26(b). Therefore, b is a unit. Hence, p is irreducible. □

In particular, this lemma along with the theorem above, shows that irreducible elements and prime elements in $\mathbb{Z}$ coincide. So that irreducible elements and prime elements in $\mathbb{Z}$ are just the associates of prime numbers, i.e. prime numbers and their negatives.

We recall from Lemma 1.25(d) that if u is a unit in a ring R, then it's inverse is unique; and this justifies the notation u^{-1}.

The following lemma about units and associates is useful for us.

Lemma 5.3. *Let R be an integral domain, and let $u, v, a, b \in R$.*

(a) *1 is a unit and $1^{-1} = 1$.*
(b) *If u is a unit, then u^{-1} is a unit and $(u^{-1})^{-1} = u$.*
(c) *If u and v are units, then uv is a unit and $(uv)^{-1} = v^{-1}u^{-1}$.*
(d) *Define $\sim$ on R by $a \sim b$ means a is a associate to b. Then $\sim$ is an equivalence relation.*
(e) *If a is associate to b and a is irreducible, then b is irreducible.*

Proof. (a) We have $1 \cdot 1 = 1$.
(b) We have $u^{-1}u = 1$, so that u^{-1} is a unit and $(u^{-1})^{-1} = u$.
(c) We have $(uv)(v^{-1}u^{-1}) = u(vv^{-1})u^{-1} = u1u^{-1} = uu^{-1} = 1$. Therefore, uv is a unit and $(uv)^{-1} = v^{-1}u^{-1}$.
(d) Let $a, b, c \in R$. The relation $\sim$ is

- reflexive, because $a = a1$;
- symmetric, because if $a = bu$, where u is a unit, then $b = au^{-1}$ and u^{-1} is a unit; and
- transitive, because if $a = bu$ and $b = cv$, where u and v are units, then $a = c(uv)$ and uv is a unit.

Hence, $\sim$ is an equivalence relation.
(e) Suppose that $b = cd$ where $c, d \in R$. Let $u \in R$ be a unit such that $a = bu$. Then $a = (cd)u = c(du)$ and $c, du \in R$. Thus c or du is a unit, because a is irreducible. Since u is a unit, this implies that c or d is a unit, and therefore b is irreducible. □

The equivalence classes of the equivalence relation $\sim$ from Lemma 5.3(iv) are often called the *associate classes* in R.

Now let us define what it means for an integral domain to be a unique factorization domain.

Definition 5.4. Let R be an integral domain. We say that R is a *unique factorization domain*, or UFD for short, provided:

(a) for all $a \in R \setminus \{0\}$, there exist irreducible elements $p_1, \dots, p_k \in R$ and a unit $u \in R$ such that $a = up_1p_2 \dots p_k$; and
(b) if $q_1, q_2, \dots, q_l \in R$ are irreducible and $v \in R$ is a unit and $a = vq_1q_2 \dots q_l$, then $l = k$ and there is a bijection $\phi : \{1, 2, \dots, k\} \to \{1, 2, \dots, k\}$ such that p_j is associate to $q_{\phi(j)}$ for each j.

So this definition roughly says that a unique factorization domain is an integral domain in which any nonzero element can be factorized into a product of a unit and some irreducibles, and that this factorization is essentially unique. We note that the inclusion of the unit u in (a) is not strictly necessary for the factorization of a nonunit, as we can just replace p_1 by up_1, which is also irreducible. However, it is useful to keep the unit here, as this means that units have a factorization as in (a) (this is the case $k = 1$), and in some examples of unique factorization domains, it is natural to have the unit there (we'll see this is the case for integers and polynomials over fields below).

Let us think about the condition in (UFD2), which is saying that the factorization is essentially unique. Of course, we can reorder a factorization of a into irreducibles, which means that we need to have the bijection ϕ as part of the definition. Also whenever we have a factorization of a as in (UFD1) then we can multiply, any of the irreducible factors p_j by a unit w and then multiply the unit u at the front by w^{-1}. So it is necessary for us to think of two factorization where the irreducible factors pair up as associates, as being essentially the same.

Now we recall the fundamental theorem of arithmetic and in a bit we recall the theorem on unique factorization of polynomials. From these we will see that both $\mathbb{Z}$ and $\mathbb{F}[X]$ (where $\mathbb{F}$ is a field) are unique factorization domains.

Theorem (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

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

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

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

It follows from this theorem that $\mathbb{Z}$ is a unique factorization domain, though the statement of the theorem looks a bit different to Definition 5.4. First we have to note that in the statement we talk about prime numbers, as discussed earlier, these are irreducible elements in $\mathbb{Z}$. Next note that, the statement of the fundamental theorem of arithmetic only deals with $n > 1$. For $n = 1$, we can take $k = 0$, and $u = 1$ to write 1 as in (UFD1). For $n < 0$, we can consider $-n > 0$ and factorize this as a $-n = p_1 p_2 \dots p_k$, and then say $n = (-1)p_1 p_2 \dots p_k$ and that -1 is a unit. The statement in (b) also looks a bit simpler than in (UFD2), the reason for not requiring the bijection ϕ is that we order $p_1 \leq p_2 \leq \cdots \leq p_k$. Also since the units in $\mathbb{Z}$ are 1 and -1 we just choose each of our irreducible elements to be positive, so that we don't need to worry about associates.

Now let's recall the theorem on unique factorization of polynomials.

Theorem (Unique factorization of polynomials)**.** *Let $f(X) \in \mathbb{F}[X]$ with $\deg f(X) > 0$. Then:*

(a) *there exists $a \in \mathbb{F}$ and monic irreducible polynomials $p_1(X), p_2(X), \dots, p_k(X) \in \mathbb{F}[X]$ such that*

$$f(X) = a p_1(X) p_2(X) \dots p_k(X).$$

(b) *if $b \in \mathbb{F}$ and $q_1(X), q_2(X), \dots, q_l(X)$ are monic irreducible polynomials such that $f(X) = b q_1(X) q_2(X) \dots q_l(X)$, then $k = l$, $a = b$ and there is a bijection $\phi : \{1, 2, \dots, k\} \to \{1, 2, \dots, k\}$ such that $p_i(X) = q_{\phi(i)}(X)$ for each i.*

It follows from this theorem that $\mathbb{F}[X]$ is a unique factorization domain. Note again that the statement of the theorem looks a bit different Definition 5.4. The theorem actually only deals with polynomials of positive degree, but we recall that the units in $\mathbb{F}[X]$ are precisely the nonzero constant polynomials, so that they can be dealt with easily. Also we don't have to mention associates in (b), which is due to the fact that we can choose a representative of an associate class of irreducible polynomials to be monic. We can do this because if $a \neq 0$ is the leading coefficient of a polynomial $f(X)$, then $a^{-1}f(X)$ is a monic polynomial associate to $f(X)$.

Now thinking back to the proof of the fundamental theorem of arithmetic the key ingredient was the theorem recalled above saying that if p is an irreducible element of $\mathbb{Z}$, then p is a prime element of $\mathbb{Z}$. We also had a similar theorem for polynomials, which we recall below.

Theorem. *Let $f(X), g(X) \in \mathbb{F}[X]$ and let $p(X) \in \mathbb{F}[X]$ be an irreducible polynomial. Suppose that $p(X) \mid f(X)g(X)$. Then $p(X) \mid f(X)$ or $p(X) \mid g(X)$.*

These observations guide us towards the next theorem.

Theorem 5.5. *Let R be an integral domain. Suppose that:*

(a) for all $a \in R \setminus \{0\}$, there exist irreducible elements $p_1, \dots, p_k \in R$ and a unit $u \in R$ such that $a = up_1p_2 \dots p_k$; and
(b) any irreducible element of R is prime.

Then R is a unique factorization domain.

We don't include the proof of this theorem here at present, though maybe one will be added at some point. The idea of the proof is simply to emulate the proof of the fundamental theorem of arithmetic. The point is that we have abstracted the key properties required in that proof, namely the existence of factorizations into irreducible elements, and that irreducible elements are prime. The details of the proof get a little more fiddly, as we have to deal with associate classes, but the main idea is exactly the same.

We point out the converse of Theorem 5.5 is also true. In particular, we point out that in a unique factorization domain any irreducible element is prime. Combining this with Lemma 5.2, we see that irreducible elements and prime elements are the same in unique factorization domains.

Let's have a example of an integral domain, which is not a unique factorization domain. In this example we consider $\mathbb{Z}[\sqrt{-3}$ and we have to work out quite a bit about it in order to deduce that it is not a unique factorization domain.

Example 5.6. Consider $R = \mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\}$. We have $R \subseteq \mathbb{C}$, so R is an integral domain. Now we have $2 \cdot 2 = 4 = (1 + \sqrt{-3})(1 - \sqrt{-3})$.

Note that R is a subring of $\mathbb{C}$, so we can take the absolute value of elements of R. Given $a + b\sqrt{-3} \in R$, we have $|a + b\sqrt{-3}|^2 = a^2 + 3b^2 \in \mathbb{Z}$.

Now suppose that $x = a + b\sqrt{-3} \in R$ is a unit. Then we have $xy = 1$ for some $y \in R$. Thus $|x|^2|y|^2 = |xy|^2 = 1$, so that $|x|^2 = 1$. Therefore, $a^2 + 3b^2 = 1$, which is only possible if $a = \pm 1$ and $b = 0$. Therefore, $U(R) = \{\pm 1\}$.

Now consider $2 \in R$ and suppose for a contradiction that 2 is not irreducible in R. Let $2 = xy$ where $x, y \in R$ are not units. Then we have $4 = |2|^2 = |x|^2|y|^2$. Thus we must have $|x|^2$ is equal to 1, 2 or 4. We have seen that if $|x|^2 = 1$, then $x = \pm 1$ is a unit. Also if $|x|^2 = 4$, then we have $|y|^2 = 1$, so that y is a unit. Thus we must have $|x|^2 = 2$. Writing $x = a + b\sqrt{-3}$ with $a, b \in \mathbb{Z}$, we have $a + 3b^2 = 2$, and this forces $b = 0$, and $a^2 = 2$, which is not possible because $\sqrt{2} \notin \mathbb{Z}$. Therefore, we have a contradiction and deduce that 2 is irreducible.

Using the same arguments, we can also show that $1+\sqrt{-3}$ and $1-\sqrt{-3}$ are irreducible in R.

Thus we have the two factorizations $4 = 2 \cdot 2$ and $4 = (1 + \sqrt{-3})(1 - \sqrt{-3})$ into irreducibles. They are certainly not the same up to reordering and associates. Hence, R is not a UFD.

Let us finish this section, by stating an important result about unique factorization domains. The proof is beyond the scope of this course.

Theorem (Gauss's lemma)**.** *Let R be an integral domain and let X be an indeterminate. Suppose that R is a unique factorization domain. Then $R[X]$ is a unique factorization domain.*

As a consequence of Gauss's lemma, we see that $\mathbb{Z}[X]$ is a unique factorization domain. Also if $\mathbb{F}$ is a field and X and Y are indeterminates, then the ring of polynomials in two variables $\mathbb{F}[X, Y]$ is a unique factorization domain; more generally, for indeterminates $X_1, X_2, \dots, X_n$ we have that $\mathbb{F}[X_1, X_2, \dots, X_n]$ is a unique factorization domain.

5.2 Highest common factors

In this short section, we consider highest common factors in integral domains. We note that highest common factors are often called greatest common divisors in books and other places.

Definition 5.7. Let R be an integral domain and let $a, b \in R$. We say that $h \in R$ is a *highest common factor* of a and b provided

(a) $h \mid a$ and $h \mid b$; and
(b) if $k \mid a$ and $k \mid b$, then $k \mid h$.

There are two possible deficiencies in this definition. Firstly, that a highest common factor need not exist in general, and indeed this is the case for some integral domains. Secondly, if a highest common does exist, then it need not be unique. However, it is essentially unique as we show in the next lemma.

Before, it is helpful to consider the case where a or b is zero. We have that a is a highest common factor of a and 0. Also we have 0 is a highest common factor of 0 and 0 and is in fact unique.

Lemma 5.8. *Let R be an integral domain and let $a, b \in R$. Suppose that $h, k \in R$ are both highest common factors a and b. Then h is a associate to k.*

Proof. If $h = 0$, then we see must have $a = b = 0$, and then also $k = 0$. It follows that we may assume that both h and k are nonzero.

We have $h = uk$ and $k = vh$ for some $u, v \in R$. Then $h = uvh$. Thus $uv = 1$, so u and v are units and h is a associate to k. □

We note that the converse of the above lemma is also true. That is if h is a highest common factor of a and b and k is associate to h, then k is also a highest common factor of a and b. This is left as an exercise.

We also state here that highest common factors do exist in unique factorization domains. We do not include a proof here at present, and it is not examinable. The idea of the proof is to take factorizations of a and b into irreducible elements, and then take the product of as many irreducible factors that appear in both, up to associates, and this is the highest common factor. A proof of this may be added here at some point in the future.

Theorem 5.9. *Let R be a unique factorization domain and let $a, b \in R$. Then a and b have a highest common factor.*

5.3 Principal ideal domains

We recall that we defined principal ideal domains in Definition 2.30. The main objective of this section is to show (though we won't include all the details) that a principal ideal domain is a unique factorization domain. Also we seen that $\mathbb{Z}$ is an example of a principal ideal domain, as is $\mathbb{F}[X]$ for $\mathbb{F}$ a field, in Propositions 2.31 and 2.32.

A useful remark for us is that for a and b in an integral domain R we have $a \mid b$ if and only if $b \in (a)$, or in other words the principal ideal (a) is equal to $\{x \in R : a \mid x\}$. Another property that we use is that if $a, b, c, x, y \in R$ such that $a \mid b$ and $a \mid c$, then $a \mid ax + by$; this can be proved in exactly the same way as it is proved for $\mathbb{Z}$.

Now we show that highest common factors exist in principal ideal domains.

Proposition 5.10. *Let R be a principal ideal domain, and let $a, b \in R$. Then a and b have a highest common factor h and $(h) = (a) + (b)$, so $h = ax + by$ for some $x, y \in R$.*

Proof. Since $(a) + (b)$ is an ideal of R, we have $(a) + (b) = (h)$ for some $h \in R$, because R is a principal ideal domain. By definition $(a) + (b) = \{ax + by : x, y \in R\}$, so $h = ax + by$ for some $x, y \in R$ because $h \in (h) = (a) + (b)$. We show that h is a highest common factor of a and b.

We have $a \in (a) + (b) = (h)$ and thus $h \mid a$, by the remark made above. Similarly, we have $h \mid b$. Now let $k \mid a$ and $k \mid b$. Then we have $k \mid ax + by = h$. Hence, h is a highest common factor of a and b. □

We now use this proposition to deduce that an irreducible element of a principal ideal domain is prime. The proof of this theorem can be done in more or less exactly the same way as the corresponding theorem for $\mathbb{Z}$ that you saw in 1AC Algebra 1; we write it slightly different and shorter here though.

Theorem 5.11. *Let R be a principal ideal and let $p \in R$. Suppose that p is irreducible. Then p is prime.*

Proof. Let $a, b \in R$ and suppose that $p \mid ab$. We have to show that $p \mid a$ or $p \mid b$. If $p \mid a$, then we are done. So we may assume that $p \nmid a$. Since p is irreducible, any highest common of p and a must be a unit, so 1 is a highest common factor of p and a. Thus we can write $1 = px + ay$ for some $x, y \in R$ by Proposition 5.10. Thus we have $b = (bx)p + y(ab)$. Hence, $p \mid b$, because $p \mid p$ and $p \mid ab$. □

Let's move on to stating that a principal ideal domain is a unique factorization domain. We only give a sketch of the proof, as proving that any element has a factorization in to irreducibles takes a bit of work.

Theorem 5.12. *Let R be an integral domain and let $p \in R$. Suppose that R is a principal ideal domain. Then R is a unique factorization domain.*

Sketch proof. We use Theorem 5.5, and we only sketch the proof that condition (a) in that theorem holds.

Let $a \in R$. We aim to show that a can be factorized as a product of irreducible elements of R. If a is a unit or if a is irreducible, then we are done. So we may assume that $a = a_1 b_1$ for some $a_1, b_1 \in R$ which are not units, so we have $(a) \subsetneqq (a_1)$. If a_1 and b_1 can be written as a product of irreducibles, then so can a. So we assume without loss of generality that a_1 cannot be written as a product of irreducibles. We can continue to argue in this way and we either factorize a as a product of irreducibles or we get a sequence $a = a_0, a_1, a_2, a_3, \dots$ of elements of R such that $(a_i) \subsetneqq (a_{i+1})$ for each i. We may take the union of this chain of ideals to obtain $I = \bigcup_{i=0}^{\infty} (a_i)$, and we can show that I is an ideal of R. Since R is a principal ideal domain we deduce that $I = (c)$ for some $c \in R$. Then we have $c \mid a_i$ for all $i \in \mathbb{N}_0$, and also $c \in (a_j)$ so $a_j \mid c$ for some $j \in \mathbb{N}_0$. But then we deduce that $a_j \mid a_{j+1}$, which is a contradiction. It follows that we must be able to factorize a as a product of irreducibles which proves that condition (a) in Theorem 5.5 holds.

Also condition (a) in Theorem 5.5 is given by Theorem 5.11. Hence, R is a unique factorization domain. □

We mention at the end of this section that $\mathbb{Z}[X]$ is an example of a unique factorization domain that is not a principal ideal domain. At some point a justification of this assertion may be given here, but for now this is left as an exercise. More generally, we mention that if R is a UFD, then $R[X]$ is a UFD, and $R[X]$ is a PID if and only if R is a field.

5.4 Euclidean domains

This section is quite brief, and the idea is TO just to give the flavour of the material. The definition of a Euclidean domain may look a bit complicated to start with, but we'll see in some examples afterwards that it is a natural concept.

Definition 5.13. Let R be a integral domain. We say that R is a *Euclidean domain* if there exists a function $d : R \setminus \{0\} \to \mathbb{N}_0$ satisfying:

(a) $d(ab) \geq d(a)$ for all $a, b \in R \setminus \{0\}$; and
(b) if $a, b \in R$ with $b \neq 0$, then there exist $q, r \in R$, with $a = qb + r$, and $r = 0$ or $d(r) < d(b)$.

The function d is called a *Euclidean function.*

The idea of the definition of a Euclidean domain is that d gives some measure of size of elements of R. Then the first condition roughly says that if you multiply two elements together, then you get something bigger; and the second condition roughly says that you can divide a by b and get a quotient q and remainder r, and the remainder is smaller than b.

This should all make a bit more sense when we look at some examples.

Examples 5.14. (a) The ring of integers $\mathbb{Z}$ is a Euclidean domain with Euclidean function given by $d(a) = |a|$. This follows from the division theorem for integers.

(b) Let $\mathbb{F}$ be a field. Then the ring of polynomials $\mathbb{F}[X]$ is a Euclidean domain with Euclidean function d given $d(f(X)) = \deg f(X)$. This follows from the division theorem for polynomials.

(c) Let $\mathbb{Z}[i]$ be the ring of Gaussian integers. Let $d : \mathbb{Z}[i] \to \mathbb{N}_0$ be defined by $d(a+bi) = |a+bi|^2 = a^2+b^2$. We'll show that d is a Euclidean function on $\mathbb{Z}[i]$. The first condition to be a Euclidean function obviously holds. To deal with the second condition let $x, y \in \mathbb{Z}[i]$ and consider $\frac{x}{y} \in \mathbb{C}$. By considering the complex plane we see that there is an element $q \in \mathbb{Z}[i]$ such that $|\frac{x}{y} - q| \leq \frac{1}{\sqrt{2}}$. Let $r = x - qy \in \mathbb{Z}[i]$, so that $x = qy + r$. Then we have

$$d(r) = |r|^2 = |x - qy|^2 = |\frac{x}{y} - q|^2 |y|^2 \leq \frac{1}{2}|y|^2 < d(y).$$

Hence, d is a Euclidean function and $\mathbb{Z}[i]$ is an Euclidean domain.

In a similar way we can prove that $\mathbb{Z}[\omega]$ is a Euclidean domain, where $\omega = e^{\frac{2\pi i}{3}}$ is a cube root of 1. Also $\mathbb{Z}[\sqrt{-2}]$ can be dealt with in a similar way. Note that $\mathbb{Z}[\sqrt{-3}]$ is not a Euclidean domain: in Example 5.6, we saw that it is not a unique factorization domain, and thus by Theorem 5.15 below it can't be a PID. It can also be proved similarly that $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain and therefore not a Euclidean domain, as is the case for $\mathbb{Z}[\sqrt{-d}]$ when $d \in \mathbb{N}$ is square free and greater than 2.

We can also consider rings of the form $\mathbb{Z}[\sqrt{d}]$ when $d \in \mathbb{N}$ is square free. It is not too difficult to show that $\mathbb{Z}[\sqrt{2}]$ is a Euclidean domain with the Euclidean function $d(a + b\sqrt{d}) = a^2 - 2b^2$. You may be interested to look into what happens for larger d – in fact it is unknown whether there are infinitely many d for which $\mathbb{Z}[\sqrt{d}]$ is a unique factorization domain!

We now show that a Euclidean domain is a principal ideal domain. You should notice that the proof is almost identical to the proof that $\mathbb{Z}$ or $\mathbb{F}[X]$ is a principal ideal domain in Propositions 2.31 and 2.32.

Theorem 5.15. *Let R be an integral domain. Suppose that R is a Euclidean domain. Then R is a principal ideal domain.*
In particular, any irreducible element in R is a prime and R is a unique factorization domain.

Proof. Let d be a Euclidean function on R.
Let I be an ideal of $\mathbb{Z}$. We need to show $I = (m)$ for some $m \in R$.
First we consider the case where $I = \{0\}$. Then we have $I = (0)$.
So we may assume that $I \neq \{0\}$. We let $m \in I$ with $d(m)$ minimal.
Then for all $r \in R$ we have $mr \in I$, because I is an ideal of R. Therefore, $(m) \subseteq I$.
Now let $x \in I$, then there exist $q, r \in R$ with $x = qm + r$ and $r = 0$ or $d(r) < d(m)$. Then $r = x - qm \in I$. Also m was the chosen so that $d(m)$ is minimal for nonzero elements of I, so we must have $r = 0$ and $x = qm \in (m)$. Therefore, $I \subseteq (m)$.
Hence, $I = (m)$. □

Let R be a Euclidean domain with Euclidean function d. Then we can use d to show that any element of $a \in R$ can be factorized as a product of irreducibles. This goes as follows, if a is a unit or irreducible, then we are done, otherwise $a = bc$ where $b, c \in R$ with $d(b), d(c) < d(a)$, then we can apply induction to deduce that b and c can be factorized as irreducibles. Note that this is essentially the same argument as used to prove that any natural number can be factorized as a product of primes.

We move on to mention the Euclidean algorithm in a Euclidean domain. Let R be a Euclidean domain with Euclidean function d, and let $a, b \in R$. We can define the Euclidean algorithm on R in the more or less the same way that it is defined on $\mathbb{Z}$ (or $\mathbb{F}[X]$ for $\mathbb{F}$ a field). This leads to a way to calculate a highest common factor h of a and b. Moreover, there is an extended Euclidean algorithm is which we reverse the Euclidean algorithm to find $x, y \in R$ such that $h = ax + by$. We omit the details here, but just reemphasize that it is essentially the same as for $\mathbb{Z}$, and we use property (b) of the Euclidean function rather than the division theorem for integers.

We end this section we some remarks about the relationship between Euclidean domain, principal ideal domains and unique factorization domains. We have seen that a Euclidean domain is a principal ideal domain and that a principal ideal domain is a unique factorization domain. Also we saw that $\mathbb{Z}[X]$ is an example of a unique factorization domain, which is not a principal ideal domain. You may wonder if there are examples of Euclidean domains that are principal ideal domains, the answer is that there are, but it is a little tricky to show this. In fact it is true that $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ is a principal ideal domain, which is not a Euclidean domain, if you want to find out more about this, then you can look at

- Oscar A. Cámpoli, *A principal ideal domain that is not a Euclidean domain*, The American Mathematical Monthly, Vol. 95, No. 9, Nov., 1988.

5.5 Fermat's two square theorem

The goal of this section is to prove the following theorem. The proof uses quite a lot of the material that we have covered so far in this course. It is a very nice example, of how all this theory can be used to prove a really beautiful theorem.

Theorem 5.16 (Fermat's two square theorem)**.** *Let $p \in \mathbb{N}$ be a prime, and suppose that $p \equiv 1 \bmod 4$. Then there exist $x, y \in \mathbb{N}$ such that $p = x^2 + y^2$.*

Proof. We have seen in the paragraphs about "Square roots of -1 in $\mathbb{F}_p$" in Section 2.6, that there exists $z \in \mathbb{F}_p$ such that $z^2 = -1$ in $\mathbb{F}_p$. Considering z as an element of $\mathbb{Z}$, we have $z^2 \equiv -1 \bmod p$, and therefore $p \mid z^2 + 1$.

We know from Examples 5.14 that $\mathbb{Z}[i]$ is a Euclidean domain. In particular, this means that an irreducible element of $\mathbb{Z}[i]$ is prime. Also we have $z^2 + 1 = (z + i)(z - i)$, so that $p \mid (z + i)(z - i)$. Further, we have $p \nmid z + i$ and $p \nmid z - i$, therefore p is not prime in $\mathbb{Z}[i]$ and thus p is not irreducible. So let $x + iy$ be a factor of p that is not a unit. Then $p = (x + iy)a$ for some $a \in \mathbb{Z}[i]$. Taking absolute values of complex numbers, we obtain that $|p| = |x + iy||a|$, therefore that $p^2 = (x^2 + y^2)|a|^2$. Thus we have $x^2 + y^2 \in \mathbb{N}$ is a factor of p in $\mathbb{Z}$, so that $x^2 + y^2$ is equal to either 1 or p. If $x^2 + y^2 = 1$, then the only possibilities are $x = \pm 1, y = 0$ or $x = 0, y = \pm 1$, so case $x + iy \in \{\pm 1, \pm i\}$ is a unit in $\mathbb{Z}[i]$, but $x+iy$ is not a unit in $\mathbb{Z}[i]$, so this is impossible. Hence, we have $p = x^2+y^2$. □

We note that it is easy to prove that for a prime p with $p \equiv 3 \bmod 4$, we cannot write p as the sum of two square. This follows from the easily checked fact that $x^2 \equiv 0 \bmod 4$ of $x^2 \equiv 1 \bmod 4$ for any $x \in \mathbb{Z}$.

In fact it is possible to prove that a natural number n can be written as the sum of two squares if and only each prime factor p of n with $p \equiv 3 \bmod 4$, occurs with an even power in the prime factorization of n. The identity

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

is key for proving that any such n can be written as the sum of two squares: can you see why? The fact that n not of the above form cannot be written as the sum of 2 squares takes a bit more work, and goes roughly as follows, as is rather brief so don't worry if you don't understand it at first.
Suppose p is a prime factor that occurs with an odd power in n, and that $n = x^2 + y^2$. If $p \mid x$, then we must also have $p \mid y$. Then dividing everything by p^2 as many times as possible we may assume that $p \nmid x$ and $p \nmid y$, and also that p has odd power in the prime factorization of n. Therefore, viewing x and y as elements of $\mathbb{Z}_p = \mathbb{F}_p$, they are both nonzero and $x^2 + y^2 = 0$ in $\mathbb{F}_p$. As $\mathbb{F}_p$ is a field we can divide by y in $\mathbb{F}_p$ and we get $z = \frac{x}{y} \in \mathbb{F}_p$ and $z^2 = -1$ in $\mathbb{F}_p$. But this is contrary to the theorem we proved in the paragraphs about "Square roots of -1 in $\mathbb{F}_p$" in Section 2.6 that there is a square root of -1 in $\mathbb{F}_p$ if and only if $p \equiv 3 \bmod 4$.

We also mention here that it is a theorem that any natural number can be written as the sum of four squares, and this is known as Lagrange's four square theorem. We won't say any more about this here, but you can look it up online.