Introduction to Ring Theory

Mira Bernstein and Julian Gilbey

Introduction to Ring Theory

Mira Bernstein and Julian Gilbey

Based on a course taught at Mathcamp 2001

Mathematics Foundation of America

© 2002 Mathematics Foundation of America

This book may be reproduced and modified for non-commercial purposes as long as this copyright notice is attached. The book may be available electronically from the Mathcamp website `http://www.mathcamp.org/`

Preface

This book is a record of an introductory ring theory course which the authors co-taught at Mathcamp 2001, with some additional help from Alice Auld.

The Canada/USA Mathcamp is a summer camp designed for gifted and talented high school students who wish to spend a summer learning about "real" mathematics, as opposed to high-school mathematics. Many of the students who participate have either had exposure to advanced mathematics from a previous Mathcamp or from taking undergraduate-level courses while still at high school. The content of the courses taught at Mathcamp ranges from advanced high school level through to advanced undergraduate or even graduate level. The level of skill and background required for these courses also varies, but not as widely. This particular course was one of the most advanced ones offered: its content was relatively accessible, but the skills it required were very high.

This course was taught following a simple version of the Moore method. Every day, the students were given definitions and statements of results (lemmas, propositions, theorems and the occasional exercise), and they were expected to be able to present proofs the next day in class. This was intended to be very stretching for the most mathematically mature and competent students at the camp, and it proved to be the case. As the course progressed, we also invited the students to develop the skill of writing mathematics by writing out proofs of some of the results.

This book, therefore, is in two parts. The first part contains the definitions and statements of results, with a small amount of discussion. This is essentially what the students were given, although a certain amount of editing has been done based on feedback from the course. The second section contains a repeat of the first section with all of the proofs and answers provided. This was done in order to enable this part to be read smoothly without having to constantly flick back and forth. Where the students provided a written proof during the course, this has been included with acknowledgement, subject to some stylistic editing where it was felt necessary. However, no attempt has been made to impose a consistent style between the different students' proofs. Where a proof is unattributed, it has been supplied by us.

It was a great privilege to be able to teach such enthusiastic, talented young mathematicians. We hope that this book will serve as a memento of the camp if you were in those classes, and as an inspiration if you were not!

Mira Bernstein, Boston
Julian Gilbey, London
May 2002

Contents

Part I: Statements of Results

1 Basic definitions and properties • 3

2 Factorisation in rings • 9

2.1 Relations, equivalence relations and partitions • 13

2.2 Gauss' Lemma and UFDs • 16

3 Homomorphisms and ideals • 19

3.1 Orderings and Zorn's Lemma • 23

Part II: Proofs of Results

1 Basic definitions and properties • 27

2 Factorisation in rings • 41

2.1 Relations, equivalence relations and partitions • 51

2.2 Gauss' Lemma and UFDs • 57

3 Homomorphisms and ideals • 63

3.1 Orderings and Zorn's Lemma • 75

Bibliography • 79

Index • 81

Part I

Statements of Results

Chapter 1

Basic definitions and properties

Definition 1.1. A **ring (with unity)** is a set R together with two binary operations called **addition** and **multiplication**. They must satisfy the following properties:

Axioms for addition

(A0) (Closure law) For all $a, b \in R$, $a + b \in R$.

(A1) (Associative law) $a + (b + c) = (a + b) + c$ for all $a, b, c \in R$.

(A2) (Zero law) There exists $0 \in R$ such that $a + 0 = 0 + a = a$ for all $a \in R$.

(A3) (Inverse law) For all $a \in R$, there exists $b \in R$ with $a+b = b+a = 0$.

(A4) (Commutative law) $a + b = b + a$ for all $a, b \in R$.

Axioms for multiplication

(M0) (Closure law) For all $a, b \in R$, $ab \in R$.

(M1) (Associative law) $a(bc) = (ab)c$ for all $a, b, c \in R$.

(M2) (Unity/Identity law) There exists $1 \in R$ such that $a1 = 1a = a$ for all $a \in R$.

Mixed axiom

(D) (Distributive law) $(a + b)c = ac + bc$ and $c(a + b) = ca + cb$ for all $a, b, c \in R$.

Usually, a binary operation on a set X is defined to be a function $X \times X \to X$, and the closure laws are implicit in this definition. We have chosen to make them explicit, though, in this introductory course.

Also, some authors do not require rings to have unity (axiom (M2)). The primary reason for this appears to be so that ideals are subrings, but we do not see this as a good enough justification for defining a ring without requiring a unity; such objects are seldom, if ever, used.

Exercise 1.2. Construct as many different examples of rings as you are able to.

Some standard examples of rings are the integers $\mathbb{Z}$, the integers modulo n, which we will write as $\mathbb{Z}/(n)$ for reasons to be discussed later, and the polynomials in x with real coefficients, written $\mathbb{R}[x]$. The rationals $\mathbb{Q}$ and the reals $\mathbb{R}$ are also rings, as is any field.

As with most formal mathematical objects, we can define the concept of a subobject in this case.

Definition 1.3. If R and S are rings, then S is called a **subring** of R if $S \subseteq R$ as sets, the operations of addition and multiplication in S are just the restrictions of those in R (that is, if $a, b \in S$, then $a + b$ is the same whether the addition is performed in R or S, and similarly for multiplication), and the zero and identity elements of S are the same as those of R.

For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$, and both are subrings of $\mathbb{R}$. Also, a subset of a ring R will be a subring if and only if it is closed under addition and multiplication, contains 0 and 1 and has additive inverses; all of the other axioms are automatically satisfied, as they hold in R.

We now consider a few 'book-keeping' lemmas, to check that things which ought to be true really are.

Lemma 1.4. In a ring, the zero element is unique (that is, if both z_1 and z_2 satisfy (A2), then $z_1 = z_2$). [You may only use axiom (A2) in your proof.]

Lemma 1.5. (Additive) inverses are unique (that is, if both b_1 and b_2 satisfy $a + b_1 = a + b_2 = 0$, then $b_1 = b_2$). Thus we can write $-a$ for the unique additive inverse of a, and we also write $a - b$ for $a + (-b)$.

Lemma 1.6. If $a \in R$, then $-(-a) = a$.

Lemma 1.7. Let $a, b, c \in R$. If $a + c = b + c$, then $a = b$. (This is called the cancellation law for addition.)

Lemma 1.8. For any $a \in R$, $0a = a0 = 0$.

Lemma 1.9. For any $a, b \in R$, $a(-b) = (-a)b = -(ab)$ and $(-a)(-b) = ab$.

Lemma 1.10. For any $a, b, c \in R$, $a(b-c) = ab-ac$, so multiplication distributes over subtraction as well as addition.

Lemma 1.11. If $0 = 1$ in the ring R, then $R = 0$ (the zero ring).

Definition 1.12. Some rings have additional properties. Here are a few basic ones. A ring R is a **division ring** if $1 \neq 0$ in R and R satisfies

(M3) (Inverse law) For all $a \in R$ with $a \neq 0$, there exists $b \in R$ with $ab = ba = 1$.

R is a **commutative ring** if it satisfies

(M4) (Commutative law) $ab = ba$ for all $a, b \in R$.

A commutative division ring is called a **field**.

All rings from now on will be assumed to be commutative rings (with unity). We will often also assume that $1 \neq 0$.

The above lemmas relating to addition in a ring can be applied to the multiplicative structure as well, as the following exercises show.

Exercise 1.13. What is the multiplicative statement corresponding to lemma 1.4? Is it true? Give a proof or counterexample.

Exercise 1.14. What is the multiplicative statement corresponding to lemma 1.5? Is it true? Give a proof or counterexample. If an element $a \in R$ has a unique multiplicative inverse, it is denoted by a^{-1}.

Exercise 1.15. What is the multiplicative statement corresponding to lemma 1.6? Is it true? Give a proof or counterexample.

Exercise 1.16. What is the multiplicative statement corresponding to lemma 1.7? Is it true? Give a proof or counterexample.

Definition 1.17. A **zero divisor** in a ring R is a non-zero element $a \in R$ for which there exists a non-zero element $b \in R$ with $ab = 0$. A ring which has no zero divisors and with $1 \neq 0$ is called an **integral domain**. (Remember that we are now dealing only with commutative rings.)

Exercise 1.18. Would any of your answers to exercises 1.13–1.16 change if we assumed the ring to be an integral domain?

Lemma 1.19. Every field is an integral domain. (Is the converse true?)

Here is a very useful way of building new rings from old ones.

Definition 1.20. If R is a ring, then the **polynomial ring** $R[x]$ is the ring of all polynomials in x with coefficients in R. (This really requires formal definitions of addition and multiplication of polynomials; see exercise 1.21 below.) Similarly, the ring $R[x_1, \ldots, x_n]$ is the ring of all polynomials with variables $x_1, \ldots, x_n$ and coefficients in R.

Exercise 1.21. (Hard) Try to formally define a polynomial (in one variable) and define the operations of addition and multiplication of polynomials. Use your definitions to check that $R[x]$ really is a ring.

Exercise 1.22. If you managed the previous exercise, can you do the same for polynomials in more than one variable?

The next definition is critical to the whole of ring theory.

Definition 1.23. An **ideal** I of a ring R is a subset of R which satisfies axioms (A0)–(A4) and for which the following 'ideal property' holds:

(I) If $a \in R$ and $x \in I$, then $ax \in I$.

The following lemma will allow us to easily show that something is an ideal.

Lemma 1.24. Let $I \subseteq R$. Then I is an ideal of R if and only if the following conditions hold:

(I1) I is non-empty.

(I2) If $x, y \in I$, then $x + y \in I$.

(I3) If $a \in R$ and $x \in I$, then $ax \in I$.

Definition 1.25. The ideal I **generated by** a set of elements in R is the smallest ideal containing the given set. ('Smallest' means that if I' is any ideal containing the given set, then $I \subseteq I'$.) If $S \subseteq R$, then we write (S) for the ideal generated by S. In the case that $S = \{x_1, x_2, \ldots, x_n\}$, we write $(S) = (x_1, x_2, \ldots, x_n)$. A **generating set** for an ideal I is a set of elements which generates I.

Lemma 1.26. The ideal generated by the set $S = \{x_1, x_2, \ldots, x_n\}$ is given by

$$(S) = \left\{ \sum_{i=1}^{n} a_i x_i : a_i \in R \right\}.$$

Exercise 1.27. (Harder) What is the corresponding result if S might be an infinite set?

Exercise 1.28. Find all ideals of $\mathbb{Z}$ and $\mathbb{R}[x]$, and find the smallest possible generating set of each such ideal.

It turns out that the ideals in these rings have a very nice structure, and we capture this in the following definition.

Definition 1.29. An ideal which has a generating set consisting of precisely one element is called a **principal ideal**. An integral domain in which every ideal is principal is called a **principal ideal domain** (PID).

We will investigate such rings in much greater detail in the next chapter.

Exercise 1.30. What is the ideal (1)?

A useful result about ideals:

Lemma 1.31. If I_1 and I_2 are ideals of a ring R, then so are $I_1 \cap I_2$ and $I_1 + I_2$, where

$$I_1 + I_2 = \{x_1 + x_2 : x_1 \in I_1 \text{ and } x_2 \in I_2\}.$$

This result can be generalised to arbitrary intersections and sums. If A is an indexing set and I_α is an ideal for each $\alpha \in A$, then $\bigcap_{\alpha\in A} I_\alpha$ is an ideal, as is

$$\sum_{\alpha\in A} I_\alpha = \left\{ \sum_{\alpha\in A} x_\alpha : x_\alpha \in I_\alpha \text{ and almost all } x_\alpha \text{ zero} \right\}.$$

The proof is essentially the same as that of the above lemma.

Definition 1.32. An element of the ring R is called **invertible** if it has a multiplicative inverse. Invertible elements are sometimes also known as **units**, especially in number theory. The set of invertible elements is called the **group of units of** R, and is denoted by R^*.

The following proposition explains the use of the word 'group' in this definition.

Proposition 1.33. The group of units of R forms an abelian group under multiplication.

Exercise 1.34. What are the groups of units of the following rings: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}[x]$, $\mathbb{Z}[i]$? Challenge: What are the groups of units of $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[\sqrt{-2}]$?

Lemma 1.35. If u is a unit, then $(u) = (1)$. It follows that if I is an ideal of R and $u \in I$, then $I = (1)$.

We can completely describe one aspect of the elements of $\mathbb{Z}/(n)$.

Proposition 1.36. Let n be an integer with $m \geqslant 2$ and let x be an integer with $1 \leqslant x < n$. Then in the ring $\mathbb{Z}/(n)$, x is a zero divisor if and only if the greatest common divisor (gcd) of x and n is greater than 1, and is a unit if and only if their gcd is equal to 1.

Corollary 1.37. Let n be an integer with $n \geqslant 2$. Then the ring $\mathbb{Z}/(n)$ is an integral domain if and only if n is prime, and in such a case is also a field.

Chapter 2

Factorisation in rings

We showed in exercise 1.28 that $\mathbb{Z}$ and $\mathbb{R}[x]$ are both PIDs, but the proof required the fact that these two rings both have a *division algorithm*. They also have the property that factorisation in them is 'unique' (in a precise sense to be explained later). We now investigate these ideas a little more carefully, and show that whenever we have a division algorithm, we also have unique factorisation. It's a bit of a long trail, but it's worth it! We do need to build up some machinery, though.

If R is not an integral domain, then we have $ab = 0$ for some non-zero a and b, so that $ac = a(b + c)$ for all $c \in R$. So we have no hope of unique factorisation in such a case. Therefore, for this entire chapter, *we assume that R is an integral domain*.

Definition 2.1. Let $a, b \in R$. We say that a **divides** b, often written $a \mid b$, if there exists an element $c \in R$ with $ac = b$. We also say that a is a **factor** or **divisor** of b in this case.

Lemma 2.2. Let $a, b, c \in R$. If $a \mid b$ and $b \mid c$, then $a \mid c$.

Lemma 2.3. Let $a, b, c \in R$. If $a \mid b$, then $ac \mid bc$.

Lemma 2.4. Let $a, b \in R$. If $a \mid b$ and $b \mid a$, then $b = ua$ for some unit $u \in R$.

Lemma 2.5. If $u \in R$ is a unit and $a \mid u$, then a is also a unit.

Lemma 2.6. If $u \in R$ is a unit, then $u \mid a$ for all $a \in R$.

Lemma 2.7. If $a, b \in R$, then $a \mid b$ if and only if $(b) \subseteq (a)$, where (a) and (b) are the ideals generated by a and b respectively. (Note that this is somewhat counterintuitive!)

Definition 2.8. If $a = ub$ with u a unit, we say that a is **associated** to b, or that a and b are **associates**, and we denote this by $a \sim b$.

Lemma 2.9. The elements a and b are associated if and only if $a \mid b$ and $b \mid a$, that is if and only if $(a) = (b)$.

Lemma 2.10. The relation of association is an equivalence relation, that is, it satisfies the following three conditions:

- (Reflexive) If $a \in R$, then $a \sim a$.
- (Symmetric) If $a \sim b$, then $b \sim a$.
- (Transitive) If $a \sim b$ and $b \sim c$, then $a \sim c$.

(We will deal with equivalence relations in more detail below.)

We need one more definition before we can define unique factorisation, and that is what we are factoring things into. This idea is used to define 'prime numbers' in high school, but we will reserve that term for an even more fundamental property.

Definition 2.11. Let R be an integral domain. A non-zero element $a \in R$ is an **irreducible** if a is not a unit, and whenever $a = bc$ for some $b, c \in R$, one of b and c is a unit (and so the other is an associate of a).

Lemma 2.12. Every associate of an irreducible element is irreducible.

Exercise 2.13. We give the standard example of a ring which does not have unique factorisation. Consider the ring $R = \mathbb{Z}[\sqrt{-5}]$. Then we have

$$6 = 2.3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

To finish the task, show that all of 2, 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducibles in this ring and that they are not associated to each other. (It is not that easy!)

Definition 2.14. An integral domain R is called a **unique factorisation domain** (UFD) if it satisfies the following two conditions:

(UFD1) Every non-zero element $r \in R$ can be written in the form

$$r = ua_1 \cdots a_n$$

where u is a unit in R, $n \geqslant 0$ and the a_i are irreducibles of R.

(UFD2) If $ua_1 \cdots u_n = vb_1 \cdots b_m$, where u and v are units and the a_i and b_i are irreducibles, then $n = m$ and $a_i \sim b_{\pi(i)}$ for some permutation π of $\{1, 2, \ldots, n\}$.

Lemma 2.15. Every field is (trivially) a UFD.

We now give the promised definition of primes in an integral domain.

Definition 2.16. An element $p \in R$ is called a **prime** in R if p is neither zero nor a unit, and whenever $a, b \in R$ with $p \mid ab$, then either $p \mid a$ or $p \mid b$ (or both).

Example 2.17. Referring back to example 2.13, we see that in $\mathbb{Z}[\sqrt{-5}]$, irreducibles may not be prime. For $2 \mid (1+\sqrt{-5})(1-\sqrt{-5})$, yet 2 does not divide either factor.

The converse is true, though:

Proposition 2.18. Let R be an integral domain. Then every prime in R is irreducible.

Theorem 2.19. Let R be an integral domain. Then R is a UFD if and only if R satisfies (UFD1) and

(UFD2′) Every irreducible in R is prime.

Definition 2.20. A **Euclidean domain** (ED) is an integral domain together with a **Euclidean function** $\phi : R \setminus \{0\} \to \mathbb{N}$ (where $\mathbb{N} = \{0, 1, 2, \dots\}$) such that:

(ED1) If $a \mid b$ then $\phi(a) \leqslant \phi(b)$, and

(ED2) Given $b \in R$ and $a \in R \setminus \{0\}$, there exist $q, r \in R$ such that $b = aq + r$ and either $r = 0$ or $\phi(r) < \phi(a)$.

Example 2.21. The rings $\mathbb{Z}$ and $\mathbb{R}[x]$ are Euclidean domains. In fact $k[x]$ is an ED for any field k.

Exercise 2.22. In the ring $\mathbb{Z}[i]$ we can define a *norm* function by $N(a+bi) = a^2 + b^2$. Show that this norm function is a Euclidean function on $\mathbb{Z}[i]$. (This is non-trivial.) It then follows from Theorem 2.30 below that this ring is a UFD.

We can now generalise the result of Exercise 1.28 above.

Theorem 2.23. Every ED is a PID.

We now aim to show that every PID is a UFD. The proof will take several steps.

Lemma 2.24. Every PID satisfies condition (UFD2′). [Hint: Let p be irreducible, and assume $p \mid ab$ but $p \nmid a$. Consider what we can say about the (principal) ideal $(p) + (a)$.]

Definition 2.25. A ring R is said to satisfy the **ascending chain condition (a.c.c.) on ideals** if any chain of ideals

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$$

is eventually constant (that is, $I_n = I_{n+1} = \cdots$ for some n).

Lemma 2.26. If $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals of R, then $\bigcup_{n=0}^{\infty} I_n$ is an ideal of R.

Lemma 2.27. Every PID satisfies a.c.c. on ideals.

Lemma 2.28. Every PID satisfies (UFD1). [Hint: Assume it doesn't. Pick an element which doesn't satisfy (UFD1), say r_0, and try factorising it. What do you find? You will want to consider the ideal (r_0) and similar ideals.]

Theorem 2.29. Every PID is a UFD.

This result together with Theorem 2.23 also shows the following result.

Theorem 2.30. Every ED is a UFD.

This can also be proved slightly more directly, but the proof is very similar in essence to the route we took.

Corollary 2.31. If k is a field, then $k[x]$ is a UFD. Also, $\mathbb{Z}$ is a UFD.

The rings we have been considering are called Euclidean because the Euclidean algorithm can be applied to them to find gcds. So we now formally define a greatest common divisor of two elements in a ring. We have to do it in the following way, as there may not be any concept of 'greatest' in the ring.

Definition 2.32. Let $a, b \in R$. d is a **greatest common divisor** (gcd) of a and b if $d \mid a$ and $d \mid b$, and whenever $c \mid a$ and $c \mid b$, we have $c \mid d$.

There may be lots of gcds:

Lemma 2.33. If d is a gcd of a and b, then d' is also a gcd of a and b if and only if $d' \sim d$.

Exercise 2.34. Using this definition, find all the greatest common divisors of 6 and 9 in the ring $\mathbb{Z}$ and of the polynomials $(x+1)(x+2)$ and $(x+1)(x^2+2)$ in the ring $\mathbb{R}[x]$. How does this definition relate to the traditional one in the case of $\mathbb{Z}$?

Exercise 2.35. Extend the definition to define a gcd of the elements $a_1, \ldots, a_n$.

Proposition 2.36. If R is an integral domain and $a, b, c \in R$ with $(a, b) = (a) + (b) = (c)$, then c is a gcd of a and b and we have $c = xa + yb$ for some $x, y \in R$. In the case that R is a PID, then we always have $(a) + (b) = (c)$ for some $c \in R$, so any a and b in R have a gcd.

In the case that the ring is Euclidean, we can determine gcds using the **Euclidean algorithm**. Let ϕ be a Euclidean function on R, and let a and b be non-zero elements of R. We can determine a gcd of a and b by performing the

following calculation:

$$\begin{aligned}
b &= aq_1 + r_1 && 0 \leqslant \phi(r_1) < \phi(a) \\
a &= r_1q_2 + r_2 && 0 \leqslant \phi(r_2) < \phi(r_1) \\
r_1 &= r_2q_3 + r_3 && 0 \leqslant \phi(r_3) < \phi(r_2) \\
r_2 &= r_3q_4 + r_4 && 0 \leqslant \phi(r_4) < \phi(r_3) \\
&\dots \\
r_{n-3} &= r_{n-2}q_{n-1} + r_{n-1} && 0 \leqslant \phi(r_{n-1}) < \phi(r_{n-2}) \\
r_{n-2} &= r_{n-1}q_n + r_n && 0 \leqslant \phi(r_n) < \phi(r_{n-1}) \\
r_{n-1} &= r_nq_{n+1}, &&
\end{aligned}$$

terminating when the remainder is zero.

Proposition 2.37. This algorithm always terminates, and when it does, r_n is a gcd of a and b.

More interestingly, we can use this algorithm to work backwards, first expressing r_n as a linear combination of r_{n-1} and r_{n-2} using the penultimate equation, then this and the preceding equation to express r_n as a linear combination of r_{n-2} and r_{n-3}, and so on. Eventually we will end up with an expression for r_n, which is a gcd of a and b, as a linear combination of a and b. This is more constructive than the result of Proposition 2.36, in particular if we are able to explicitly determine the quotient and remainder for the division steps, as we are for $\mathbb{Z}$ and $\mathbb{Q}[x]$.

The last main result that we wish to prove in this chapter is based on Gauss' lemma, and states that if R is a UFD, then so is $R[x]$. But we will need to do some important preparatory work on equivalence relations and fields of fractions first. This material is used throughout mathematics, and we will meet it again when we consider quotient rings in Chapter 3.

2.1 Relations, equivalence relations and partitions

Definition 2.38. Let X and Y be any sets. The **Cartesian product** of X and Y is defined by

$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\},$$

the set of all ordered pairs of elements from X and Y.

Definition 2.39. A **(binary) relation** on X is a subset $\mathcal{R} \subseteq X \times X$. We often write $x \mathcal{R} y$ to mean $(x, y) \in \mathcal{R}$ and $x \not\mathcal{R} y$ to mean $(x, y) \notin \mathcal{R}$.

Example 2.40. Let $X = \mathbb{Z}$. Then the subset $\{(m, n) : m < n\} \subset \mathbb{Z} \times \mathbb{Z}$ defines the relation 'less than'. The subset $\{(m, n) : m = n\}$ is the relation of equality.

Definition 2.41. The following terms are used to describe some special features of relations. Here $\mathcal{R}$ is a relation on a set X.

- $\mathcal{R}$ is called **reflexive** if $x \mathcal{R} x$ for all $x \in X$.
- $\mathcal{R}$ is called **symmetric** if for all $x, y \in X$, $x \mathcal{R} y$ implies that $y \mathcal{R} x$.
- $\mathcal{R}$ is called **anti-symmetric** if for all $x, y \in X$, $x \mathcal{R} y$ and $y \mathcal{R} x$ together imply that $x = y$.
- $\mathcal{R}$ is called **transitive** if for all $x, y, z \in X$, $x \mathcal{R} y$ and $y \mathcal{R} z$ together imply that $x \mathcal{R} z$.

Exercise 2.42. Given the set $X = \{1, 2, 3\}$, for each of the following conditions, find a relation on X which satisfies it:

(a) The relation is reflexive and symmetric, but not transitive.

(b) The relation is reflexive and transitive, but not symmetric.

(c) The relation is symmetric and transitive, but not reflexive.

(d) The relation is reflexive, transitive and anti-symmetric.

Definition 2.43. The following special types of relations are so important that they are given special names.

- A relation which is reflexive, symmetric and transitive is called an **equivalence relation**.
- A relation which is reflexive, transitive and anti-symmetric is called a **partial order** on the set, and the set is called a **partially ordered set** or **poset**. If in addition, for every pair $x, y \in X$, either $x = y$, $x \mathcal{R} y$ or $y \mathcal{R} x$, then the relation is called a **(total) order**, and the set is called a **totally ordered set** or just an **ordered set**. (Some authors require that these relations have $(x, x) \notin \mathcal{R}$ for all $x \in X$; this is essentially the difference between the relations $\leqslant$ and $<$.)

Exercise 2.44. Give some examples of equivalence relations, partial orders which are not total orders, and total orders, either for the set $X = \{1, 2, 3\}$ or for other sets.

We will be focusing on equivalence relations. An equivalence relation essentially says that certain elements in the set X are to be considered 'the same' in a precise sense depending upon the context. We make this a little more formal, by showing that partitions and equivalence relations are essentially the same concept.

Definition 2.45. A **partition** of a set X consists of a set $\mathcal{X}$ of non-empty subsets of X satisfying the following conditions:

(a) The subsets in $\mathcal{X}$ cover X, that is

$$\bigcup_{Y \in \mathcal{X}} Y = X.$$

(b) The subsets are pairwise disjoint, that is if $Y_1, Y_2 \in \mathcal{X}$ with $Y_1 \neq Y_2$, then $Y_1 \cap Y_2 = \varnothing$.

Note that condition (b) implies that if $Y_1, Y_2 \in \mathcal{X}$ and $Y_1 \cap Y_2 \neq \varnothing$, then $Y_1 = Y_2$.

Definition 2.46. Let $\mathcal{R}$ be a relation on a set X. For $x \in X$, define $\mathcal{R}(x) = \{y \in X : x \mathcal{R} y\}$. In the case that $\mathcal{R}$ is an equivalence relation, $\mathcal{R}(x)$ is called the **equivalence class** of x. Very often one uses the notation $[x]$ or $\bar{x}$ to denote this.

Lemma 2.47. If $\mathcal{R}$ is an equivalence relation on X, then $x \in \mathcal{R}(x)$ for all $x \in X$. Furthermore, whenever $y \in \mathcal{R}(x)$, $\mathcal{R}(x) = \mathcal{R}(y)$.

Proposition 2.48. If $\mathcal{R}$ is an equivalence relation on X, then the set of equivalence classes $P(\mathcal{R}) = \{\mathcal{R}(x) : x \in X\}$ is a partition of X.

Proposition 2.49. Let $\mathcal{X}$ be a partition of a set X. Define a relation $R(\mathcal{X})$ on X by $(x, y) \in R(\mathcal{X})$ if and only if x and y lie in the same subset in $\mathcal{X}$. Then $R(\mathcal{X})$ is an equivalence relation on X.

Proposition 2.50. The operations defined in the previous two propositions are mutually inverse, that is given an equivalence relation $\mathcal{R}$ on X, $R(P(\mathcal{R})) = \mathcal{R}$, and given a partition $\mathcal{X}$ of X, $P(R(\mathcal{X})) = \mathcal{X}$.

Definition 2.51. Let $\mathcal{R}$ be an equivalence relation on X. We define the **quotient** of X by $\mathcal{R}$, written $X/\mathcal{R}$, to be the set $P(\mathcal{R})$.

Thus we can think of equivalence relations and partitions as one and the same thing: an equivalence relation partitions a set into parts called equivalence classes, where the elements of a given part are considered equal to each other, but elements in different parts are considered unequal.

Often, the notation $a \equiv b$ is used to indicate an equivalence relation. Whenever you see this notation, you should check to see that you are indeed working with an equivalence relation.

Examples 2.52. Here are some typical examples of the use of equivalence relations in pure mathematics.

- When we build the rationals from the integers and make identifications such as $\frac{1}{2} = \frac{3}{6}$, we are essentially defining an equivalence relation on pairs of integers, as we explain in more detail below.

- When we use modular arithmetic, we define an equivalence relation on the integers with m being equivalent to n whenever the modulus divides $m - n$.
- In topology, when we identify edges of a square to make a torus, we are essentially defining an equivalence relation on the points of the square.
- Later on, we will see that quotient rings are defined by equivalence relations, as are factor groups (also called quotient groups) in group theory.

We are now able to return to the point at hand, and define a generalisation of the formation of $\mathbb{Q}$ from $\mathbb{Z}$.

Let R be an integral domain. Let X be the set $R \times (R \setminus \{0\})$. We define an equivalence relation $\sim$ on X by $(a', b') \sim (a, b)$ if and only if $a'b - ab' = 0$. (Check that this really is an equivalence relation!) We define addition and multiplication on X by

$$(a, c) + (b, d) = (ad + bc, cd)$$
$$(a, c)(b, d) = (ab, cd)$$

Lemma 2.53. If $(a', c') \sim (a, c)$ and $(b', d') \sim (b, d)$, we have $(a', c') + (b', d') \sim (a, c) + (b, d)$ and $(a', c')(b', d') \sim (a, c)(b, d)$.

Hence addition and multiplication are well-defined on the set $K = X/\sim$. As with the rationals, we write a/b for the equivalence class of (a, b) in this set, so that $a/b + c/d = (ad + bc)/bd$ and $(a/b)(c/d) = ac/bd$.

Proposition 2.54. In the above setup, K is a field, called the **field of fractions** of R. The map $\phi : R \to K$ given by $\phi(a) = a/1$ is an injective (one-to-one) homomorphism. [Homomorphisms are special kinds of maps defined at the start of Chapter 3.]

Lemma 2.55. Assume R is a UFD. Then given $a, b \in R$, we can find a gcd d of a and b. Then if $a = da'$ and $b = db'$, 1 is a gcd of a' and b'. Hence in the above construction, $(a, b) \sim (a', b')$ with 1 being a gcd of a' and b'. That is, in the case of a UFD, we can always write a/b in 'lowest terms' with a and b having no common (prime) factor.

2.2 Gauss' Lemma and UFDs

We now have enough machinery to work on Gauss' lemma and UFDs. For the rest of this chapter, we let R be a UFD and K its field of fractions. Let $a \in K$ be a non-zero element of the field, and let p be a prime element in R. We can then write

$$a = p^r b$$

where $b \in K$, r is an integer and p does not divide the numerator or denominator of b when written in lowest terms. For example, if $R = \mathbb{Z}$, so $K = \mathbb{Q}$, and $p = 3$, then we can write $\frac{27}{17} = 3^3 \cdot \frac{1}{17}$ and $\frac{14}{15} = 3^{-1} \cdot \frac{14}{5}$.

Lemma 2.56. In the above equation, r is uniquely determined by a and p.

Definition 2.57. In the above equation, r is called the **order of a at p** and we write $r = \operatorname{ord}_p a$. We also define $\operatorname{ord}_p 0 = \infty$, with the conventions that $n < \infty$ and $n + \infty = \infty + n = \infty$ for all $n \in \mathbb{Z}$, and $\infty + \infty = \infty$.

Lemma 2.58. If $a, a' \in K$, then

$$\operatorname{ord}_p(aa') = \operatorname{ord}_p a + \operatorname{ord}_p a'.$$

Definition 2.59. Let $f \in K[x]$ be a polynomial over the field K, with

$$f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0.$$

As before, if $f = 0$, we define $\operatorname{ord}_p f = \infty$. Otherwise, define

$$\operatorname{ord}_p f = \min_{0 \leqslant i \leqslant n} \{\operatorname{ord}_p a_i\}.$$

Then if $r = \operatorname{ord}_p f$, we call up^r a **p-content** for f, where u is any unit of R. If $f \neq 0$, we define the **content** of f by

$$\operatorname{cont}(f) \sim \prod p^{\operatorname{ord}_p f},$$

the product being taken over all primes p such that $\operatorname{ord}_p f \neq 0$. The content of f is defined up to association, that is, up to multiplication by a unit of R. By abuse of notation, we often write $\operatorname{cont}(f) = \prod p^{\operatorname{ord}_p f}$ instead, using $=$ instead of $\sim$.

Lemma 2.60. If $b \in K$ with $b \neq 0$ and $f \neq 0$, then $\operatorname{cont}(bf) = b \operatorname{cont}(f)$. Hence we can write

$$f = c \cdot f_1$$

where $c = \operatorname{cont}(f)$ and $\operatorname{cont}(f_1) = 1$.

Definition 2.61. A polynomial with content 1 is called a **primitive polynomial**.

Lemma 2.62. If f is a primitive polynomial, then all of the coefficients of f lie in R and their gcd is 1.

Theorem 2.63 (Gauss' Lemma). Let R be a UFD and K its field of fractions. Let $f, g \in K[x]$ be non-zero polynomials in one variable over K. Then

$$\operatorname{cont}(fg) = \operatorname{cont}(f) \operatorname{cont}(g).$$

[Hint: Prove the result first for primitive polynomials, and consider $\operatorname{ord}_p(fg)$ for each prime of R.]

Corollary 2.64. Let $f(x) \in R[x]$ have a factorisation $f(x) = g(x)h(x)$ in $K[x]$. If $c_g = \text{cont}(g)$, $c_h = \text{cont}(h)$ and $g(x) = c_g g_1(x)$, $h(x) = c_h h_1(x)$, then

$$f(x) = c_g c_h g_1(x) h_1(x),$$

and $c_g c_h \in R$. In particular, if $f, g \in R[x]$ have content 1, then $h \in R[x]$ also.

Theorem 2.65. Let R be a UFD. Then the polynomial ring $R[x]$ in one variable is also a UFD. Its units are the units of R, and its prime elements are either primes of R or polynomials in $R[x]$ which are irreducible in $K[x]$ and have content 1. [Hint: Remember that we know that $K[x]$ is a UFD.]

Corollary 2.66. Let R be a UFD. Then the polynomial ring $R[x_1, \ldots, x_n]$ in n variables is also a UFD. Its units are the units of R and its prime elements are either primes of R or polynomials which are irreducible in $K[x_1, \ldots, x_n]$ and have content 1.

Chapter 3

Homomorphisms and ideals

We do not make any assumptions on our rings in this chapter except for commutativity unless otherwise stated.

Definition 3.1. Let R_1 and R_2 be rings. Then a **homomorphism** from R_1 to R_2 is a function

$$\phi : R_1 \to R_2$$

satisfying the conditions:

$$\begin{aligned} \phi(a+b) &= \phi(a) + \phi(b) \qquad \text{for all } a, b \in R \\ \phi(ab) &= \phi(a)\phi(b) \qquad \text{for all } a, b \in R \\ \phi(1) &= 1. \end{aligned}$$

Exercise 3.2. Let $n \geqslant 2$ be an integer, and for $a \in \mathbb{Z}$, write $[a]_n$ for the equivalence class of a modulo n (that is, $[a]_n = \{r \in \mathbb{Z} : a \equiv r \pmod{n}\}$). We can regard $\mathbb{Z}/(n)$ as the ring whose elements are $[0]_n$, $[1]_n$, $\ldots$, $[n-1]_n$. Then the map $f : \mathbb{Z} \to \mathbb{Z}/(n)$ defined by $f(k) = [k]_n$ is a homomorphism.

Lemma 3.3. Let $\phi : R_1 \to R_2$ be a homomorphism. Then the following hold:

(a) $\phi(0) = 0$.

(b) $\phi(-a) = -\phi(a)$.

(c) $\phi(a-b) = \phi(a) - \phi(b)$.

(d) If $a \in R_1$ is invertible, then $\phi(a)$ is also invertible and $\phi(a)^{-1} = \phi(a^{-1})$.

Lemma 3.4. There is a unique homomorphism $\mathbb{Z} \to R$ for any ring R.

Exercise 3.5. When does there exist a homomorphism from $\mathbb{Z}/(m)$ to $\mathbb{Z}/(n)$?

Definition 3.6. Consider the unique homomorphism $\phi : \mathbb{Z} \to K$ where K is a field. Either there is a unique smallest positive $p > 0$ with $\phi(p) = 0$ or no such p exists. In the former case, we say that the field has **characteristic** p, otherwise we say that the field has **characteristic** 0. The characteristic of K is denoted by $\operatorname{char} K$. We will show later that the characteristic must be zero or a prime; see Proposition 3.31.

Exercise 3.7. What are the characteristics of the fields $\mathbb{Q}$ and $\mathbb{Z}/(p)$, where p is a positive prime?

Lemma 3.8. If $\phi : R_1 \to R_2$ and $\psi : R_2 \to R_3$ are homomorphisms, then $\psi\phi : R_1 \to R_3$ is also a homomorphism.

Definition 3.9. Let $\phi : R_1 \to R_2$ be a homomorphism. If ϕ is injective, then it is called a **monomorphism**; if it is surjective, it is called an **epimorphism**, and if it is bijective, it is called an **isomorphism**. Two rings R_1 and R_2 are said to be **isomorphic**, denoted by $R_1 \cong R_2$, if there exists an isomorphism $\phi : R_1 \to R_2$.

If two rings are isomorphic, they are essentially the same, up to a possible renaming of the elements.

Proposition 3.10. Isomorphism is an equivalence relation on any set of rings.

There is a simple test to see whether a homomorphism is injective.

Proposition 3.11. Let $\phi : R_1 \to R_2$ be a homomorphism. Then ϕ is a monomorphism if and only if $\phi(a) = 0$ implies $a = 0$.

Lemma 3.12. Let $\phi : R_1 \to R_2$ and $\psi : R_2 \to R_3$ be homomorphisms. If $\psi\phi$ is a monomorphism, then ϕ is a monomorphism. If $\psi\phi$ is an epimorphism, then ψ is an epimorphism. We cannot say anything about the other homomorphism in either case.

Definition 3.13. Let $\phi : R_1 \to R_2$ be a homomorphism. Then the image of R_1, $\phi(R_1) = \{\phi(a) : a \in R_1\}$, is called the **image** of ϕ and is denoted by $\operatorname{Im}\phi$.

Proposition 3.14. If $\phi : R_1 \to R_2$ is a homomorphism, then $\operatorname{Im}\phi$ is a subring of R_2.

Definition 3.15. Let $\phi : R_1 \to R_2$ be a homomorphism. Then the inverse image of 0, $\phi^{-1}(0) = \{a \in R_1 : \phi(a) = 0\}$, is called the **kernel** of ϕ and is denoted by $\ker\phi$.

So we can restate Proposition 3.11 as: a homomorphism $\phi : R_1 \to R_2$ is injective if and only if $\ker\phi = 0$.

Proposition 3.16. The kernel of a homomorphism is an ideal.

Definition 3.17. A **coset** of an ideal I in a ring R is a set of the form

$$a + I = \{a + x : x \in I\},$$

where $a \in R$.

Lemma 3.18. Let $\phi : R_1 \to R_2$ be a homomorphism. If $b \in \phi(R_1)$, then $\phi^{-1}(b)$ is a coset of $\ker \phi$.

Exercise 3.19. If $x_1, x_2, \ldots, x_n \in \ker \phi$, then $(x_1, x_2, \ldots, x_n) \subseteq \ker \phi$.

Lemma 3.20. Let I be an ideal of R. The relation on R defined by

$$a \equiv b \pmod{I} \text{ if and only if } a - b \in I$$

is an equivalence relation. (This is called *equivalence modulo* I or *equivalence mod* I.) The equivalence classes are precisely the cosets of I in R.

Exercise 3.21. Show that in $\mathbb{Z}$, the relation $a \equiv b \pmod{(m)}$ is the same as the traditional relation $a \equiv b \pmod{m}$.

Definition 3.22. Let I be an ideal of R. Then the quotient of R by the relation defined in Lemma 3.20 is denoted by R/I. (See Definition 2.51.) We define a ring structure on R/I as follows. If $a + I, b + I \in R/I$, then

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

R/I is called the **quotient ring** of R by I, and read as 'R mod I'.

Exercise 3.23. Show that the definition makes sense, that is, if different representatives of the cosets $a + I$ and $b + I$ are chosen, the sum and product are still $(a + b) + I$ and $ab + I$ respectively. Show also that the ring axioms are satisfied for R/I.

Exercise 3.24. Explain the notation for the integers modulo n which we have been using, namely $\mathbb{Z}/(n)$.

Proposition 3.25. If $f : R \to R'$ is a homomorphism of rings and I' is an ideal of R', then $f^{-1}(I')$ is an ideal of R. However, if I is an ideal of R then $f(I)$ is *not* necessarily an ideal of R'.

Definition 3.26. Let I be an ideal of R. The homomorphism $\phi_I : R \to R/I$ defined by $\phi_I(a) = a + I$ is called the **canonical homomorphism**.

Exercise 3.27. What is the kernel of this homomorphism?

The following theorem is exceedingly important and very powerful.

Theorem 3.28 (First Isomorphism Theorem). Let $\phi : R \to R'$ be a homomorphism of rings with kernel $I = \ker \phi$. Then $\bar{\phi} : R/I \to \phi(R)$ defined by $\bar{\phi}(a + I) = \phi(a)$ is an isomorphism. Thus $\phi = \bar{\phi}\phi_I$, so that ϕ factors through the canonical homomorphism.

An alternative way of writing this isomorphism is $R/\ker\phi \cong \operatorname{Im}\phi$.

Exercise 3.29. Show that $k[x,y]/(y) \cong k[x]$, where k is a field.

Exercise 3.30. Describe the ring $\mathbb{R}[x]/(x^2+1)$.

Proposition 3.31. Let K be a field. Then $\operatorname{char} K = 0$ or $\operatorname{char} K$ is a prime. [Hint: Use Definition 3.6 and the First Isomorphism Theorem on the map. What are the possible ideals of $\mathbb{Z}$?]

Proposition 3.32. Let I, J be ideals of R with $I \subseteq J$. Let J/I be the quotient of J by the relation defined in Lemma 3.20. Then J/I is an ideal of R/I. Conversely, if J' is any ideal of R/I, then $J' = J/I$ for some ideal $J \subseteq R$ containing I. Thus there is a one-to-one correspondence between ideals of R/I and ideals of R containing I.

Definition 3.33. An ideal $I \subseteq R$ is called **prime** if $I \neq (1)$ and for any $a, b \in R$, $ab \in I$ implies that either $a \in I$ or $b \in I$.

The two uses of the term 'prime' which we have seen are closely related as the following proposition shows.

Proposition 3.34. Let R be an integral domain and let p be a non-zero element of R. Then p is a prime element of R if and only if (p) is a prime ideal of R.

Corollary 3.35. In a PID, any non-zero prime ideal is generated by a prime element.

Proposition 3.36. Let $f : R \to R'$ be a homomorphism of rings, and let I be a prime ideal of R'. If I is prime, then so is $f^{-1}(I)$.

Exercise 3.37. Is the converse of the last proposition true?

Exercise 3.38. If P is a prime ideal of a ring R, what can you say about the quotient ring R/P? Try to find a property which would make your answer an 'if and only if' statement, and prove it.

Definition 3.39. An ideal $I \subseteq R$ is called **maximal** if $I \neq (1)$ and there is no ideal $J \subset R$ such that $I \subset J \subset (1)$ (strict inclusions). In other words, the only ideal 'bigger' than I is (1).

Exercise 3.40. Can a ring have more than one maximal ideal?

The following exercises are useful for helping us to identify maximal ideals.

Exercise 3.41. What can you say about the ideals of a field? Try to phrase your answer as an 'if and only if' statement, and prove it.

Exercise 3.42. If M is a maximal ideal of a ring R, what can you say about the quotient ring R/M? Try to find a property which would make your answer an 'if and only if' statement, and prove it.

Proposition 3.43. A maximal ideal is prime.

Here is one very nice application of prime ideals. This is not quite the most general form, but it is very powerful nonetheless.

Proposition 3.44 (Eisenstein's criterion). Let R be an integral domain and let K be its field of fractions. Let $f = a_n x^n + \cdots + a_0 \in R[x]$ and let M be a maximal ideal of R. If $a_n \notin M$, $a_i \in M$ for $0 \leqslant i < n$ and $a_0 \notin M^2$, then f is irreducible in $K[x]$.

Here $M^2 = MM$ is the ideal generated by all products xy where $x, y \in M$. Also, in the case $R = \mathbb{Z}$, where every non-zero prime ideal is maximal, this proposition becomes: if p is prime with $p \nmid a_n$, $p \mid a_i$ for $0 \leqslant i < n$ and $p^2 \nmid a_0$, then f is irreducible in $\mathbb{Q}[x]$. So, for example, we see that $x^n + 5x + 5$ is irreducible for all $n \geqslant 2$. In the case $R = \mathbb{Z}$, this can be proved directly without recourse to prime ideals and the machinery we have been building up, but this approach is potentially more enlightening.

Note that f may not be irreducible in $R[x]$, for example, $2(x^n + 5x + 5)$ is reducible in $\mathbb{Z}[x]$ but irreducible in $\mathbb{Q}[x]$.

Here is one nice application of this result.

Proposition 3.45. Let $p \in \mathbb{Z}$ be a positive prime. Then $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1 = (x^p - 1)/(x - 1)$ is irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$. [Hint: Consider $\Phi_p(x + 1)$.]

3.1 Orderings and Zorn's Lemma

In this final section, we show that maximal ideals always exist, subject to assuming the Axiom of Choice.

Definition 3.46. Recall the definitions of partially ordered sets (posets) and totally ordered sets from the previous chapter (Definition 2.43). Let S be a poset, and write $a \leqslant b$ for the partial order on S. We now introduce some additional terminology regarding ordered sets.

- A **greatest element** of a partially ordered set S is an element a such that $b \leqslant a$ for all $b \in S$. **Least elements** are defined analogously.
- A **maximal element** of a partially ordered set S is an element $a \in S$ such that whenever $a \leqslant b$ with $b \in S$, we have $a = b$. (Alternatively, a maximal element is an element $a \in S$ for which there is no $b \in S$ with $a < b$.) **Minimal elements** are defined analogously.

- Let T be a subset of a partially ordered set S. An **upper bound** of T in S is an element $b \in S$ such that $x \leqslant b$ for all $x \in T$.
- A subset $T \subseteq S$ is said to be a **totally ordered subset**, or a **chain**, if every pair of elements of T are comparible in the partial order. (That is, for every $x, y \in T$, either $x \leqslant y$ or $y \leqslant x$ or both.)
- A poset S is **inductively ordered** if every non-empty totally ordered subset of S has an upper bound.

Proposition 3.47. Not every partially ordered set S has a maximal element, and even if such an element exists, it is not necessarily unique. Not every partially ordered set S has a greatest element, but if such an element exists, it is unique, and is also the unique maximal element of S.

Exercise 3.48. Let R be a commutative ring and M a maximal ideal of R. What partially ordered set is M a maximal element of?

Axiom 3.49 (Zorn's Lemma). Let S be a non-empty inductively ordered set. Then S has a maximal element.

(Some authors remove the requirement 'non-empty' from both the definition of 'inductively ordered' and the statement of Zorn's Lemma. The result is the same, as then the empty set, which is totally ordered, must have an upper bound, so S must be non-empty.)

Zorn's Lemma is usually treated as an axiom of set theory. It is, in fact, equivalent to the Axiom of Choice, which states that for any set S, there exists a *choice function*, which is a mapping $f : \mathcal{P}(S) \setminus \varnothing \to S$ with $f(X) \in X$ for every $\varnothing \neq X \subseteq S$, where $\mathcal{P}(S)$ is the power set of S, which is the set of all subsets of S. (Why is the word 'choice' used to describe this function?) Some people quibble with this axiom, as it results in a number of apparent paradoxes, such as the Banach-Tarski paradox. However, most working mathematicians do use it, since it is crucial for some fundamental results, such as the following.

Theorem 3.50. Every ideal of a ring R is contained in a maximal ideal.

Corollary 3.51. Every ring has a maximal ideal.

Another example of where Zorn's Lemma is used is in the proof that every vector space has a basis: here one takes sets of independent vectors ordered by inclusion; a maximal independent set is a basis.

We end on a lighter note ....

Bad Joke 3.52. What is yellow and equivalent to the axiom of choice? *Zorn's Lemon.*

Part II

Statements and Proofs of Results

Chapter 1

Basic definitions and properties

Definition 1.1. A **ring (with unity)** is a set R together with two binary operations called **addition** and **multiplication**. They must satisfy the following properties:

Axioms for addition

(A0) (Closure law) For all $a, b \in R$, $a + b \in R$.

(A1) (Associative law) $a + (b + c) = (a + b) + c$ for all $a, b, c \in R$.

(A2) (Zero law) There exists $0 \in R$ such that $a + 0 = 0 + a = a$ for all $a \in R$.

(A3) (Inverse law) For all $a \in R$, there exists $b \in R$ with $a+b = b+a = 0$.

(A4) (Commutative law) $a + b = b + a$ for all $a, b \in R$.

Axioms for multiplication

(M0) (Closure law) For all $a, b \in R$, $ab \in R$.

(M1) (Associative law) $a(bc) = (ab)c$ for all $a, b, c \in R$.

(M2) (Unity/Identity law) There exists $1 \in R$ such that $a1 = 1a = a$ for all $a \in R$.

Mixed axiom

(D) (Distributive law) $(a + b)c = ac + bc$ and $c(a + b) = ca + cb$ for all $a, b, c \in R$.

Usually, a binary operation on a set X is defined to be a function $X \times X \to X$, and the closure laws are implicit in this definition. We have chosen to make them explicit, though, in this introductory course.

Also, some authors do not require rings to have unity (axiom (M2)). The primary reason for this appears to be so that ideals are subrings, but we do not see this as a good enough justification for defining a ring without requiring a unity; such objects are seldom, if ever, used.

Examples 1.2. Some standard examples of rings are the integers $\mathbb{Z}$, the integers modulo n (which we will write as $\mathbb{Z}/(n)$ for reasons to be discussed later) and the polynomials in x with real coefficients, written $\mathbb{R}[x]$. The rationals, $\mathbb{Q}$, and the reals, $\mathbb{R}$, are also rings, as is any field.

There is a trivial ring, also called the *zero ring*, with only one element, 0. It is denoted by 0. It is the unique ring satisfying $1 = 0$, as we will show below.

An interesting example of a ring is

$$\mathbb{Z}_{(2)} = \{a/b : a, b \in \mathbb{Z} \text{ and } b \text{ odd}\}.$$

You can check that this is, indeed, a ring. Can you generalise this example?

Another family of examples is rings of the form $\mathbb{Z}[\alpha]$, which is made by *adjoining* the element α to the ring $\mathbb{Z}$. It is defined to be the set

$$\mathbb{Z}[\alpha] = \{a_0 + a_1\alpha + \cdots + a_n\alpha^n : n \geqslant 0,\ a_i \in \mathbb{Z}\}.$$

Check that this is a ring! If $\alpha = x$ is an indeterminate, then this ring is called the ring of polynomials in x over the integers, and is denoted by $\mathbb{Z}[x]$, as above. If $\alpha = i = \sqrt{-1}$, then this ring is called the *ring of Gaussian integers*, that is $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$. This example can be generalised in two ways: the ring $\mathbb{Z}$ can be replaced by other rings, and more than one element may be adjoined. This example will be developed a little further when we consider polynomial rings in more detail below.

Another example is constructed as follows. Let X be a set, and let the elements of our ring R be all subsets of X (so R is the power set of X). Addition is defined to be the symmetric difference and multiplication is intersection, that is

$$A + B = (A \setminus B) \cup (B \setminus A)$$
$$A \cdot B = A \cap B$$

Such a ring is known as a *Boolean ring*. It is an exercise to check that this set with the specified operations does indeed satisfy all of the ring axioms.

For our final example here, consider a topological space X. (If you don't know what a topological space is, take X to be the real numbers.) Then the set $C(X)$ of continuous real-valued functions on X forms a ring, where addition and multiplication are defined pointwise (that is, $(f + g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$). This can also be generalised in many ways to provide many interesting examples of rings, leading into areas of analysis and algebraic geometry.

As with most formal mathematical objects, we can define the concept of a subobject in this case.

Definition 1.3. If R and S are rings, then S is called a **subring** of R if $S \subseteq R$ as sets, the operations of addition and multiplication in S are just the restrictions of those in R (that is, if $a, b \in S$, then $a + b$ is the same whether the addition is performed in R or S, and similarly for multiplication), and the zero and identity elements of S are the same as those of R.

For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$, and both are subrings of $\mathbb{R}$. Also, a subset of a ring R will be a subring if and only if it is closed under addition and

multiplication, contains 0 and 1 and has additive inverses; all of the other axioms are automatically satisfied, as they hold in R.

We now consider a few 'book-keeping' lemmas, to check that things which ought to be true really are.

Lemma 1.4. *In a ring, the zero element is unique (that is, if both z_1 and z_2 satisfy* (A2)*, then $z_1 = z_2$).*

Proof (*Boris Alexeev*) Consider z_1 and z_2. By (A2), we have

$$\begin{aligned} & z_1 + z_2 = z_1 \qquad \text{because } z_2 \text{ is a zero,} \\ \text{and} \quad & z_1 + z_2 = z_2 \qquad \text{because } z_1 \text{ is a zero.} \end{aligned}$$

Hence $z_1 = z_2$. □

Lemma 1.5. (*Additive*) *inverses are unique (that is, if both b_1 and b_2 satisfy $a+b_1 = a+b_2 = 0$, then $b_1 = b_2$). Thus we can write $-a$ for the unique additive inverse of a, and we also write $a - b$ for $a + (-b)$.*

Proof. Assume that both b_1 and b_2 satisfy $a + b_1 = a + b_2 = 0$. Then we can add an additive inverse of a to both sides, say b_1, giving

$$\begin{aligned} & b_1 + (a + b_1) = b_1 + (a + b_2) & \\ \text{so} \quad & (b_1 + a) + b_1 = (b_1 + a) + b_2 & \text{by associativity} \\ \text{thus} \quad & (a + b_1) + b_1 = (a + b_1) + b_2 & \text{by commutativity} \\ \text{hence} \quad & 0 + b_1 = 0 + b_2 & \text{as } a + b_1 = 0 \end{aligned}$$

Hence $b_1 = b_2$ as required. □

Lemma 1.6. *If $a \in R$, then $-(-a) = a$.*

Proof. We have $a + (-a) = 0$ by the definition of $-a$, so we see that a is the additive inverse of $-a$, or in symbols $a = -(-a)$. □

Lemma 1.7. *Let $a, b, c \in R$. If $a + c = b + c$, then $a = b$.* (*This is called the cancellation law for addition.*)

Proof (*William Meyerson*) By axiom (A3), there exists d such that $c + d = 0$. Then

$$\begin{aligned} & (a + c) + d = (b + c) + d, & \\ \text{so} \quad & a + (c + d) = b + (c + d) & \text{by associativity (axiom (A1)),} \\ \text{so} \quad & a + 0 = b + 0, & \end{aligned}$$

yielding $a = b$ by the definition of 0 (axiom (A2)), as required. □

Lemma 1.8. *For any $a \in R$, $0a = a0 = 0$.*

Proof (*Pradeep Mouli*) We have

$$\begin{array}{lrl} & 0+0=0 & \text{by axiom (A2),} \\ \text{so} & a(0+0)=a0 & \text{by substitution.} \\ \text{Hence} & a0+a0=a0 & \text{by distributivity, axiom (D)} \\ & =0+a0 & \text{by axiom (A2).} \end{array}$$

Thus $a0 = 0$ by cancellation (Lemma 1.7). The proof that $0a = 0$ is entirely similar, so $a0 = 0a = 0$. □

Lemma 1.9. *For any $a, b \in R$, $a(-b) = (-a)b = -(ab)$ and $(-a)(-b) = ab$.*

Proof. We have

$$\begin{array}{rl} ab+a(-b)=a(b+(-b)) & \text{by distributivity, axiom (D)} \\ =a0=0 & \text{by axiom (A2) and Lemma 1.8} \end{array}$$

So by the definition and uniqueness of the additive inverse of ab, we deduce that $a(-b) = -(ab)$. The other half, $(-a)b = -(ab)$ follows similarly.

For $(-a)(-b) = ab$, note that $(-a)(-b) = -(a(-b)) = -(-(ab))$ by a double application of our result, and hence by Lemma 1.6, $(-a)(-b) = ab$. □

Lemma 1.10. *For any $a, b, c \in R$, $a(b-c) = ab-ac$, so multiplication distributes over subtraction as well as addition.*

Proof. We have $a(b - c) = a(b + (-c))$ by definition (see Lemma 1.5), so

$$\begin{array}{rll} a(b-c) & a(b+(-c)) & \\ & =ab+a(-c) & \text{by distributivity} \\ & =ab+(-(ac)) & \text{by the previous lemma} \\ & =ab-ac & \text{by definition of subtraction,} \end{array}$$

and we are done. □

Lemma 1.11. *If $0 = 1$ in the ring R, then $R = 0$ (the zero ring).*

Proof (*George Khachatryan*) Let $a \in R$. We know from Lemma 1.8 that $a0 = 0a = 0$. Then from $a0 = 0$, we have $a1 = 1$ on substituting $1 = 0$. But by the Identity Axiom (M2), this means that $a = 1$. So $a = 0$ as $1 = 0$. Hence $R = 0$ as the only element in R is 0. □

Definition 1.12. Some rings have additional properties. Here are a few basic ones. A ring R is a **division ring** if $1 \neq 0$ in R and R satisfies

(M3) (Inverse law) For all $a \in R$ with $a \neq 0$, there exists $b \in R$ with $ab = ba = 1$.

R is a **commutative ring** if it satisfies

(M4) (Commutative law) $ab = ba$ for all $a, b \in R$.

A commutative division ring is called a **field**.

All rings from now on will be assumed to be commutative rings (with unity). We will often also assume that $1 \neq 0$.

The above lemmas relating to addition in a ring can be applied to the multiplicative structure as well, as follows.

Lemma 1.13. (*Yuliya Gorlina*) *In a ring, the multiplicative identity is unique.*

Proof. Let R be a ring. Suppose there exist two elements $e, e' \in R$ such that $ea = ae = a$ and $e'a = ae' = a$ for all $a \in R$. Using these equations we have $ee' = e$ and $ee' = e'$, so $e = e'$ and the multiplicative identity is unique. We can therefore denote this unique multiplicative identity by 1. □

Lemma 1.14. (*Yuliya Gorlina*) *Multiplicative inverses are unique if they exist. We can, therefore, write a^{-1} for the unique inverse of a if it exists.*

Proof. Let R be a ring and let $a \in R$ be an invertible element. Suppose $ab = ba = 1$ and $ac = ca = 1$ for some $b, c \in R$. Then, $ab = ac$, so $b(ab) = b(ac)$ and hence $(ba)b = (ba)c$, by associativity. Thus $1b = 1c$, so that $b = c$. So multiplicative inverses are unique. □

Lemma 1.15. (*Yuliya Gorlina*) *If $a \in R$ is invertible, then a^{-1} is invertible and $(a^{-1})^{-1} = a$.*

Proof. We have $aa^{-1} = a^{-1}a = 1$ by the definition of the inverse of a. However, the equation also shows that a is the inverse of a^{-1}. So a^{-1} is invertible, and $(a^{-1})^{-1} = a$. □

Note 1.16. (Yuliya Gorlina) It is *not* true that $ac = bc$ implies that $a = b$ in general. For example, in $\mathbb{Z}/(6)$, $2.3 = 4.3$, but $2 \neq 4$.

Definition 1.17. A **zero divisor** in a ring R is a non-zero element $a \in R$ for which there exists a non-zero element $b \in R$ with $ab = 0$. A ring which has no zero divisors and with $1 \neq 0$ is called an **integral domain**. (Remember that we are now dealing only with commutative rings.)

We can recover the cancellation law for multiplication if we assume that we are working in an integral domain:

Lemma 1.18. (*Mark Lipson*) *Let R be an integral domain and $a, b, c \in R$. If $ac = bc$ and $c \neq 0$, then $a = b$. This is the cancellation law for multiplication in an integral domain.*

Proof. Since $ac = bc$, we have $bc - ac = ac - ac = 0$, so $(b - a)c = 0$ by distributivity (Lemma 1.10). Since R is an integral domain and $c \neq 0$, we must have $b - a = 0$, or $a = b$. □

Lemma 1.19. *Every field is an integral domain. The converse is false in general.*

Proof (*James Berglund*) Assume that a field F is not an integral domain. Then there exist $a, b \in F$ with a, b non-zero, such that $ab = 0$. By definition, a field has multiplicative inverses for all non-zero elements, so in particular, a has an inverse a^{-1}. Then

	$a^{-1}(ab) = 0$	
so	$(a^{-1}a)b = 0$	by associativity (M1),
hence	$1b = 0$	by the definition of inverses (M3),
and finally	$b = 0$	by the definition of unity (M2).

But this is a contradiction to our assumption that b is non-zero, hence F cannot have any zero divisors.

The converse is false, as evidenced by $\mathbb{Z}$. □

We now wish to formally define polynomial rings. We do not repeat the (somewhat vague) definition given in the questions section, but present a much more powerful setup which will allow us to define both polynomials and polynomial rings. We will not be making use of the technical aspects of this construction again, and the reader is recommended to only skim the material from here until the beginning of Definition 1.23 on page 35 on a first reading.

The most common method of defining polynomial rings is somewhat intricate, but accessible; it makes use of the idea of a *monoid*, which is essentially a group without inverses. We will work with commutative monoids, but there is no necessity to assume commutativity; the monoid rings we will be defining are commutative if and only if the monoid is commutative. (We have labelled the axioms with the same numbers we used for the definition of rings.)

Definition 1.20. A **monoid** is a set M together with one binary operation called multiplication which satisfies the axioms

- **(M0) (Closure law)** For all $a, b \in M$, $ab \in M$.
- **(M1) (Associative law)** $a(bc) = (ab)c$ for all $a, b, c \in M$.
- **(M2) (Identity law)** There exists $1 \in M$ such that $a1 = 1a = a$ for all $a \in M$.

The monoid is called **commutative** if the operation also satisfies

(M4) (Commutative law) $ab = ba$ for all $a, b \in M$.

Every group is a monoid, and the multiplicative 'part' of any ring is a monoid. The most familiar monoids which do not fall into either of these categories are $(\mathbb{N}, +)$, the set of natural numbers (including zero!) with the operation of addition, and $(\mathbb{N}, \times)$, the natural numbers with the operation of multiplication. Another very useful way of thinking about $(\mathbb{N}, +)$, which we will make good use of, is to write it as the set $\{x^0 = 1, x^1 = x, x^2, x^3, \dots\}$, with $x^m \cdot x^n = x^{m+n}$. This is called the **free monoid** on x, and we will use this to define polynomial rings.

There are two equivalent ways of defining monoid rings; we will present both for the sake of completeness.

Definition 1.21. (David Roe) Let R be a commutative ring and M a monoid. We define the **monoid ring** $R[M]$ to be the set of all functions $f : M \to R$ satisfying $f(m) = 0$ for almost all $m \in M$ (that is, $f(m)$ is non-zero for only finitely many $m \in M$). If $f, g \in R[M]$, we define $f + g$ and fg by

$$\begin{aligned}(f+g)(m) &= f(m) + g(m),\\ (fg)(m) &= \sum_{lk=m} f(l)g(k).\end{aligned}$$

Note that this is actually a finite sum, as almost all of the terms are zero.

Alternatively, we can define $R[M]$ to be the set of all 'formal sums'

$$R[M] = \left\{ \sum_{m \in M} a_m m : a_m \in R, \text{ only finitely many } a_m \text{ non-zero} \right\},$$

where the terms $a_m m$ are formal products of a ring element and a monoid element. (This means that we don't actually try to 'evaluate' these terms, as a_m is an element of R and m is in M, and we have no rules for multiplying elements from these two distinct structures.) We then define addition and multiplication by

$$\begin{aligned}\left(\sum_{m \in M} a_m m \right) + \left(\sum_{m \in M} b_m m \right) &= \sum_{m \in M} (a_m + b_m) m,\\ \left(\sum_{m \in M} a_m m \right)\left(\sum_{m \in M} b_m m \right) &= \sum_{m \in M} \left(\sum_{\substack{l,k \in M \\ lk = m}} a_l b_k \right) m.\end{aligned}$$

These two definitions are essentially equivalent. Indeed, given a function $f : M \to R$, we can produce the formal sum $\sum_{m \in M} f(m)m$, and given the formal sum $\sum_{m \in M} a_m m$, we can produce the function $f : M \to R$ given by $f(m) = a_m$.

We must prove that these definitions do, indeed, yield rings. We will content ourselves to identify the zero and unity of the ring $R[M]$ and to prove that multiplication is associative; all of the other axioms can be shown to be true

similarly (and mostly somewhat more simply). Similarly, it is easy to show that $R[M]$ *is commutative if and only if* M *is commutative*. We use the second definition of $R[M]$ in our proofs. The zero element of $R[M]$ is $\sum_{m\in M} 0m$, and unity is $\sum_{m\in M} \chi_m m$, where $\chi_m = 1$ if $m = 1$ and zero otherwise.

To show that the multiplication is associative, let $a = \sum_{m\in M} a_m m$, $b = \sum_{m\in M} b_m m$, $c = \sum_{m\in M} c_m m$. We have

$$\begin{aligned}
(ab)c &= \Big(\Big(\sum_{m\in M} a_m m\Big)\Big(\sum_{m\in M} b_m m\Big)\Big)\Big(\sum_{m\in M} c_m m\Big) \\
&= \Big(\sum_{m\in M}\Big(\sum_{\substack{l,k\in M\\ lk=m}} a_l b_k\Big)m\Big)\Big(\sum_{m\in M} c_m m\Big) \\
&= \sum_{m\in M}\Big(\sum_{\substack{i,j\in M\\ ij=m}}\Big(\sum_{\substack{l,k\in M\\ lk=i}} a_l b_k\Big)c_j\Big)m \\
&= \sum_{m\in M}\Big(\sum_{\substack{l,k,j\in M\\ lkj=m}} a_l b_k c_j\Big)m.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
a(bc) &= \Big(\sum_{m\in M} a_m m\Big)\Big(\Big(\sum_{m\in M} b_m m\Big)\Big(\sum_{m\in M} c_m m\Big)\Big) \\
&= \Big(\sum_{m\in M} a_m m\Big)\Big(\sum_{m\in M}\Big(\sum_{\substack{l,k\in M\\ lk=m}} b_l c_k\Big)m\Big) \\
&= \sum_{m\in M}\Big(\sum_{\substack{i,j\in M\\ ij=m}} a_i\Big(\sum_{\substack{l,k\in M\\ lk=j}} b_l c_k\Big)\Big)m \\
&= \sum_{m\in M}\Big(\sum_{\substack{i,l,k\in M\\ ilk=m}} a_i b_l c_k\Big)m.
\end{aligned}$$

The final expressions in both cases are identical, hence $(ab)c = a(bc)$ and the multiplication is associative.

It is common to write specific elements of $R[M]$ as explicit sums. For example, if $M = \{1, x, x^2, \dots\}$ and $a = \sum_{m\in M} a_m m$ with $a_1 = 2$ and $a_{x^2} = 3$, we could write $a = 2.1 + 3x^2$, or just $a = 2 + 3x^2 = 3x^2 + 2$. Note that this is precisely the ring $R[x]$, so we have found our desired definition:

Definition 1.22. If R is a ring, then the **polynomial ring** $R[x]$ is the monoid ring $R[M]$ where $M = \{1, x, x^2, x^3, \dots\}$. The elements of the ring $R[x]$ are called **polynomials in** x **over** R, or just polynomials. Similarly, the ring $R[x_1, \dots, x_n]$

of all polynomials over R in the variables $x_1, \ldots, x_n$ is the monoid ring $R[M]$, where $M = \{1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots\}$ is the free monoid generated by $x_1, \ldots, x_n$.

The next definition is critical to the whole of ring theory.

Definition 1.23. An **ideal** I of a ring R is a subset of R which satisfies axioms (A0)–(A4) and for which the following 'ideal property' holds:

(I) If $a \in R$ and $x \in I$, then $ax \in I$.

The following lemma will allow us to easily show that something is an ideal.

Lemma 1.24. *Let $I \subseteq R$. Then I is an ideal of R if and only if the following conditions hold:*

(I1) *I is non-empty.*

(I2) *If $x, y \in I$, then $x + y \in I$.*

(I3) *If $a \in R$ and $x \in I$, then $ax \in I$.*

Proof. These three conditions are clearly necessary: axiom (A2) requires $0 \in I$, so I is non-empty; condition (I2) is just axiom (A0), and condition (I3) is the ideal property (I).

To show that these three conditions are sufficient, we must show that the axioms (A0)–(A4) and the ideal property hold whenever these conditions hold. The ideal property (I) clearly holds, as this is condition (I3). Axiom (A0) is just condition (I2). Axioms (A1) and (A4) hold as addition in the ideal is just addition in the ring. So we need only show axioms (A2) and (A3). Let $x \in I$. (Such an x exists by (I1).) Then $0x = 0 \in I$ by the ideal property, so (A2) holds. Similarly, if $x \in I$, then using Lemma 1.9 and the ideal property, we see that $(-1)x = -(1x) = -x \in I$, so (A3) holds. □

Definition 1.25. The ideal I **generated by** a set of elements in R is the smallest ideal containing the given set. ('Smallest' means that if I' is any ideal containing the given set, then $I \subseteq I'$.) If $S \subseteq R$, then we write (S) for the ideal generated by S. In the case that $S = \{x_1, x_2, \ldots, x_n\}$, we write $(S) = (x_1, x_2, \ldots, x_n)$. A **generating set** for an ideal I is a set of elements which generates I.

Lemma 1.26. *The ideal generated by the set $S = \{x_1, x_2, \ldots, x_n\}$ is given by*

$$(S) = \left\{ \sum_{i=1}^{n} a_i x_i : a_i \in R \right\}.$$

Proof (*Ilya Grigoriev*) Let $I = \left\{\sum_{i=1}^{n} a_i x_i : a_i \in R\right\}$. We have to show that $I = (S)$. To do this we have to prove that I is an ideal containing S and that it is the smallest possible one.

We first show that I is an ideal. To do this, we shall prove that it satisfies the conditions of Lemma 1.24. Taking each $a_i = 0$ shows that I is non-empty, so that (I1) is satisfied.

For (I2), closure, let $a, b \in I$. Then, for some $a_i, b_i \in R$, $a = \sum_{i=1}^{n} a_i x_i$ and $b = \sum_{i=1}^{n} b_i x_i$. By commutativity and distributivity, $a+b = \sum_{i=1}^{n} (a_i + b_i) x_i \in I$; as R is closed under addition, $a_i + b_i \in R$, so $a + b \in I$. Hence I is closed under addition.

Finally for the ideal property (I3), let $a = \sum_{i=1}^{n} a_i x_i \in I$ and $c \in R$. Then

$$ca = c \sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} c(a_i x_i) = \sum_{i=1}^{n} (ca_i) x_i \in I$$

by distributivity and associativity. As $ca_i \in R$ for each i, we have $ca \in I$ as required. Hence I is an ideal.

It is obvious that $S \subseteq I$, and now that we have proved that I is an ideal, it follows that $(S) \subseteq I$ by the definition of (S). We are left to prove that $I \subseteq (S)$.

Let $a = \sum_{i=1}^{n} a_i x_i \in I$. But $x_i \in S \subseteq (S)$ for each i. By the ideal property of (S), $a_i x_i \in (S)$ for each i. So, by closure, $a \in (S)$.

So $I \subseteq (S)$. Hence $(S) = I$ as required. □

We can extend this result to infinite generating sets if we are careful, as follows.

Lemma 1.27. (*Theo Johnson-Freyd*) *Let R be a ring, A a (possibly infinite) indexing set and $x_\alpha \in R$ for each $\alpha \in A$. Then the ideal generated by the set $S = \{x_\alpha : \alpha \in A\}$ is given by*

$$(S) = \left\{ \sum_{\alpha \in A}^{n} a_\alpha x_\alpha : a_\alpha \in R \text{ and only finitely many } a_\alpha \text{ non-zero} \right\}.$$

Proof. The proof is essentially the same as the proof of the previous lemma. Again let

$$I = \left\{ \sum_{\alpha \in A}^{n} a_\alpha x_\alpha : a_\alpha \in R \text{ and only finitely many } a_\alpha \text{ non-zero} \right\}.$$

Properties (I1) and (I3) hold for I as in the previous proof; for closure, we note that if $a = \sum_{\alpha \in A}^{n} a_\alpha x_\alpha$ and $b = \sum_{\alpha \in A}^{n} b_\alpha x_\alpha$ with only finitely many a_α non-zero and only finitely many b_α non-zero, then only finitely many $a_\alpha + b_\alpha$ can be non-zero, so $a + b \in I$, proving that (I2) holds. Hence I is an ideal.

We have $(S) \subseteq I$ as before, for $x_\alpha = \sum_{\beta \in A} \chi_\beta x_\beta$ where $\chi_\beta = 1$ if $\beta = \alpha$ and zero otherwise, so $S \subseteq I$; we must show that $I \subseteq (S)$. The proof runs almost exactly as before: let $a = \sum_{\alpha \in A}^{n} a_\alpha x_\alpha \in I$. Note that $x_\alpha \in S \subseteq (S)$ for each α. By the ideal property of (S), $a_\alpha x_\alpha \in (S)$ for each α. So, by closure, $a \in (S)$, and we are allowed to form this sum as only finitely many terms are non-zero. It follows from this that $I \subseteq (S)$, and so $(S) = I$ as required. □

Exercise 1.28. Find all ideals of $\mathbb{Z}$ and $\mathbb{R}[x]$, and find the smallest possible generating set of each such ideal.

The following proof essentially shows that these two rings are Euclidean domains, and then goes on to argue that they are PIDs. This argument is generalised in Chapter 2.

Proof (*Cristina Domnisoru*) We show that the ideals of $\mathbb{Z}$ are all of the form (n) for some $n > 0$ and that the ideals of $\mathbb{R}[x]$ are all of the form (f) for some $f \in \mathbb{R}[x]$.

Beginning with $\mathbb{Z}$, let I be an ideal of $\mathbb{Z}$. If the only element of I is 0, then $I = (0)$ and we are done. Otherwise, there exists a non-zero $n \in I$. It follows that $-n \in I$, and as one of n and $-n$ is positive, we may assume that $n > 0$. So I contains a positive integer, hence I contains a least positive integer, l say. We will prove that $I = (l)$, and so every ideal is principal.

As $l \in I$, $(l) \subseteq I$. Assume that $I \neq (l)$. Then there exists an $n \in I$ with $n \notin (l)$ (so that n is not a multiple of l). We can write

$$n = cl + r, \qquad \text{where } c \in \mathbb{Z} \text{ and } 0 < r < l,$$

and as $n \in I$ and $cl \in (l) \subseteq I$, we must have $n - cl = r \in I$. Now l was chosen to be the smallest positive integer in I, but r is positive and smaller than l, which is a contradiction. Thus $I = (l)$ and we are done.

The argument for $\mathbb{R}[x]$ is similar. Again, if $I = 0$ we are done. If not, we consider the *degrees* of the non-zero polynomials in the ideal I, and pick a polynomial with smallest degree, say f. (The degree of the non-zero polynomial $\sum_{i=0}^{\infty} a_i x^i$ is the largest i with $a_i \neq 0$.) We claim that $I = (f)$, and argue as above. We certainly have $(f) \subseteq I$; if equality holds, we are done. Otherwise, let $g \in I$ with $g \notin (f)$. Then we can write

$$g = hf + r, \qquad \text{where } h \in \mathbb{R}[x] \text{ and } 0 \leqslant \deg(r) < \deg(f).$$

As $g \in I$ and $hf \in I$, we must have $g - hf = r \in I$. But $\deg(r) < \deg(f)$, contradicting the choice of f. Thus $I = (f)$ and we are done. □

Given this result, we are inspired to make the following definition.

Definition 1.29. An ideal which has a generating set consisting of precisely one element is called a **principal ideal**. An integral domain in which every ideal is principal is called a **principal ideal domain** (PID).

We will investigate such rings in much greater detail in the next chapter.

Example 1.30. (Cristina Domnisoru) The ideal (1) is the ideal that contains all the elements of R.

A useful result about ideals:

Lemma 1.31. *If I_1 and I_2 are ideals of a ring R, then so are $I_1 \cap I_2$ and $I_1 + I_2$, where*

$$I_1 + I_2 = \{x_1 + x_2 : x_1 \in I_1 \text{ and } x_2 \in I_2\}.$$

Proof (*Justin Blanchard*) We use the result of Lemma 1.24 to prove these two statements.

Consider first the set $I = I_1 \cap I_2$. By (A2), $0 \in I_1$ and $0 \in I_2$, therefore $0 \in I$, hence I is non-empty and so satisfies (I1). Now let $x, y \in I$. Clearly $x, y \in I_1$ and $x, y \in I_2$, and by additive closure in ideals, $x + y \in I_1$ and $x + y \in I_2$, therefore $x + y \in I$. Thus I satisfies (I2). Finally, let $x \in I$ and $a \in R$. Now $x \in I_1$ and therefore $ax \in I_1$; similarly $x \in I_2$ implies $ax \in I_2$, therefore $ax \in I$, and I satisfies the ideal property (I3). Hence $I = I_1 \cap I_2$ is an ideal.

Now consider $I = I_1 + I_2$. By (A2), $0 \in I_1$ and $0 \in I_2$, thus $0 + 0 = 0 \in I$, and I satisfies (I1). Now let $x, y \in I$, so $x = x_1 + x_2$ and $y = y_1 + y_2$ for some $x_1, y_1 \in I_1$ and $x_2, y_2 \in I_2$. By closure, $x_1 + y_1 \in I_1$ and $x_2 + y_2 \in I_2$, thus $(x_1+y_1)+(x_2+y_2) = (x_1+x_2)+(y_1+y_2) = x+y \in I$, thus (I2) holds in I. Finally, let $x \in I$ and $a \in R$. Then $x = x_1 + x_2$ for some $x_1 \in I_1$, $x_2 \in I_2$. By the ideal property, $ax_1 \in I_1$ and $ax_2 \in I_2$, so $ax = a(x_1 + x_2) = ax_1 + ax_2 \in I_1 + I_2 = I$, so I satsifies the ideal property (I3). Hence $I = I_1 + I_2$ is an ideal. □

This result can be generalised to arbitrary intersections and sums. If A is an indexing set and I_α is an ideal for each $\alpha \in A$, then $\bigcap_{\alpha \in A} I_\alpha$ is an ideal, as is

$$\sum_{\alpha \in A} I_\alpha = \left\{ \sum_{\alpha \in A} x_\alpha : x_\alpha \in I_\alpha \text{ and almost all } x_\alpha \text{ zero} \right\}.$$

The proof is essentially the same as that of the above lemma.

Definition 1.32. An element of the ring R is called **invertible** if it has a multiplicative inverse. Invertible elements are sometimes also known as **units**, especially in number theory. The set of invertible elements is called the **group of units of** R, and is denoted by R^*.

The following proposition explains the use of the word 'group' in this definition.

Proposition 1.33. *The group of units of R forms an abelian group under multiplication.*

Proof. We must check that the group axioms are satisfied. We must show that the set of units is closed under multiplication, that multiplication is associative and commutative, has an identity and that every element has an inverse. Since multiplication is associative and commutative in our ring, it will also be so in the set of units, so we may assume these from now on.

We start with closure. Let $u, v \in R^*$. Then there exist elements $u', v' \in R$ with $uu' = vv' = 1$. Then $(uv)(u'v') = (uu')(vv') = 1.1 = 1$ using associativity

and commutativity. Thus uv is invertible, so $uv \in R^*$ and R^* is closed under multiplication.

As $1.1 = 1$, 1 is invertible, so $1 \in R^*$; thus we have an identity element. Finally, if $u \in R^*$, there exists $v \in R$ with $uv = 1$. This also shows that v is invertible, so $v \in R^*$, and R^* has inverses. □

Exercise 1.34. What are the groups of units of the following rings: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}[x]$, $\mathbb{Z}[i]$? Challenge: What are the groups of units of $\mathbb{Z}[\sqrt{2}]$ and $\mathbb{Z}[\sqrt{-2}]$?

Answer. The units of $\mathbb{Z}$ are just ± 1. Every non-zero element of $\mathbb{Q}$ is a unit. In $\mathbb{R}[x]$, the units are just the non-zero reals regarded as degree 0 polynomials, and in $\mathbb{Z}[i]$, the units are ± 1 and $\pm i$, which can be shown by considering the norms of elements as in Example 2.22 below.

The units of $\mathbb{Z}[\sqrt{-2}]$ can easily be shown to be just ± 1 by a consideration of the norm function $N(a + b\sqrt{-2}) = a^2 + 2b^2$; to have $N(a + b\sqrt{-2}) = 1$ then requires $a^2 = 1$ and $b = 0$. However, for $\mathbb{Z}[\sqrt{2}]$, the situation is a lot more complex: we now require $a + b\sqrt{2}$ to satisfy $a^2 - 2b^2 = 1$, which is a case of Pell's equation. It turns out that the units of this ring are given by $\pm(1 + \sqrt{2})^n$ where $n \in \mathbb{Z}$. □

Lemma 1.35. *If u is a unit, then $(u) = (1)$. It follows that if I is an ideal of R and $u \in I$, then $I = (1)$.*

Proof. As u is a unit, there exists $v \in R$ with $vu = 1$. Hence $1 = vu \in (u)$, and so $(1) \subseteq (u)$. But $(u) \subseteq (1) = R$, so $(u) = (1)$. Finally, if $u \in I$, then $(u) \subseteq I$, hence $(1) \subseteq I \subseteq (1) = R$, so we must have equality and $I = (1)$. □

We can completely describe one aspect of the elements of $\mathbb{Z}/(n)$.

Proposition 1.36. *Let n be an integer with $m \geqslant 2$ and let x be an integer with $1 \leqslant x < n$. Then in the ring $\mathbb{Z}/(n)$, x is a zero divisor if and only if the greatest common divisor (gcd) of x and n is greater than 1, and is a unit if and only if their gcd is equal to 1.*

Proof (*Taktin Oey*) Let $d = \gcd(x, n)$. As $d \mid n$, n/d is an integer and therefore $x \mid x(\frac{n}{d})$. Similarly, x/d is an integer and $n \mid n(\frac{x}{d}) = x(\frac{n}{d})$. Now assume $d = \gcd(x, n) > 1$ and write $a = n/d$. Then as $n \mid xa$, we have $xa \equiv 0 \pmod{n}$. As $1 \leqslant x < n$ and $1 \leqslant a < n$, it follows that $x \not\equiv 0 \pmod{n}$, so x is a zero divisor in $\mathbb{Z}/(n)$.

Conversely, if x is a zero divisor in $\mathbb{Z}/(n)$, there must exist integers k and m such that $xk = mn$ and $n \nmid k$. But $\gcd(x, n) = 1$ would imply that $n \mid k$, so we must have $\gcd(x, n) > 1$.

For the second part of the proposition, if $\gcd(x, n) = 1$, then the Diophantine equation $xc + nl = 1$ can be solved for c and l in integers using the Euclidean algorithm (see the discussion around Proposition 2.37 in Chapter 2). But $1 - nl \equiv 1 \pmod{n}$, so $xc \equiv 1 \pmod{n}$ and x is a unit in $\mathbb{Z}/(n)$.

Conversely, if x is a unit, there exists c with $xc \equiv 1 \pmod{n}$, so $xc = nl + 1$ for some $l \in \mathbb{Z}$, so that $xc - nl = 1$ and $\gcd(x, n) = 1$. □

Corollary 1.37. *Let n be an integer with $n \geqslant 2$. Then the ring $\mathbb{Z}/(n)$ is an integral domain if and only if n is prime, and in such a case is also a field.*

Proof. If n is prime, we show that $\mathbb{Z}/(n)$ is a field by showing that all of $1, 2, \ldots, n-1$ are units (that is, they are invertible). But the gcd of n and any of these numbers is 1 as n is prime, so the result follows from our proposition. Since every field is an integral domain by Lemma 1.19, we are done in this direction.

Conversely, if n is not prime, let $n = ab$ be a non-trivial factorisation of n into positive integers. Then a is a zero divisor in $\mathbb{Z}/(n)$ by the proposition, and so this ring cannot be an integral domain. □

The restriction $n \geqslant 2$ is not actually necessary in this corollary, but we have not yet explained what we mean by $\mathbb{Z}/(n)$ when $n < 2$.

Chapter 2

Factorisation in rings

We showed in exercise 1.28 that $\mathbb{Z}$ and $\mathbb{R}[x]$ are both PIDs, but the proof required the fact that these two rings both have a *division algorithm*. They also have the property that factorisation in them is 'unique' (in a precise sense to be explained later). We now investigate these ideas a little more carefully, and show that whenever we have a division algorithm, we also have unique factorisation. It's a bit of a long trail, but it's worth it! We do need to build up some machinery, though.

If R is not an integral domain, then we have $ab = 0$ for some non-zero a and b, so that $ac = a(b + c)$ for all $c \in R$. So we have no hope of unique factorisation in such a case. Therefore, for this entire chapter, *we assume that R is an integral domain.*

Definition 2.1. Let $a, b \in R$. We say that a **divides** b, often written $a \mid b$, if there exists an element $c \in R$ with $ac = b$. We also say that a is a **factor** or **divisor** of b in this case.

Lemma 2.2. *Let $a, b, c \in R$. If $a \mid b$ and $b \mid c$, then $a \mid c$.*

Proof (*Ryan Hendrickson*) As $a \mid b$, there exists $m \in R$ such that $am = b$. Similarly, there exists $n \in R$ such that $bn = c$. Thus $amn = bn = c$, and as $mn \in R$ (rings are closed under multiplication), we have $a \mid c$ as required. □

Lemma 2.3. *Let $a, b, c \in R$. If $a \mid b$, then $ac \mid bc$.*

Proof (*Ryan Hendrickson*) As $a \mid b$, there exists $m \in R$ such that $am = b$. Thus $acm = amc = bc$, so we have $ac \mid bc$. □

Lemma 2.4. *Let $a, b \in R$. If $a \mid b$ and $b \mid a$, then $b = ua$ for some unit $u \in R$.*

Proof (*Ryan Hendrickson*) As $a \mid b$, there exists $m \in R$ such that $am = b$. Similarly, there exists $n \in R$ such that $bn = a$. Thus $amn = bn = a$, so either $mn = 1$ or $a = 0$ (by Lemma 1.18). If $mn = 1$, then m is a unit as required, and if $a = 0$, then $b = am = 0$, so $b = 1a$. □

Lemma 2.5. *If $u \in R$ is a unit and $a \mid u$, then a is also a unit.*

Proof (*Michael Kolodrubetz*) There exists $b \in R$ such that $u = ab$. Therefore, $1 = uu^{-1} = abu^{-1} = a(bu^{-1})$, so a is a unit. □

Lemma 2.6. *If $u \in R$ is a unit, then $u \mid a$ for all $a \in R$.*

Proof (*Ben Aisen*) Let $u \in R$ be a unit, with inverse u^{-1}. By definition, $uu^{-1} = 1$. Take any $a \in R$. Then $u(u^{-1}a) = (uu^{-1})a = 1a = a$, so $u \mid a$. □

Lemma 2.7. *If $a, b \in R$, then $a \mid b$ if and only if $(b) \subseteq (a)$, where (a) and (b) are the ideals generated by a and b respectively. Note that this is somewhat counterintuitive!*

Proof (*Boyko Kakaradov*) Assume $a \mid b$, then there exists $m \in R$ such that $am = b$. By definition, $(a) = \{ac : c \in R\}$ and $(b) = \{bd : d \in R\}$. Take any $e \in (b)$, then $e = bd$ for some $d \in R$. Then $e = (am)d = a(md)$, therefore $e \in (a)$. So $(b) \subseteq (a)$.

Conversely, if $(b) \subseteq (a)$, then as $b \in (b)$, we have $b \in (a)$. Since $(a) = \{ac : c \in R\}$, it follows that $b = ac$ for some $c \in R$. So $a \mid b$. □

Definition 2.8. If $a = ub$ with u a unit, we say that a is **associated** to b, or that a and b are **associates**, and we denote this by $a \sim b$.

Lemma 2.9. *The elements a and b are associated if and only if $a \mid b$ and $b \mid a$, that is if and only if $(a) = (b)$.*

Proof (*George Khachatryan*) By Lemma 2.4, if $b \mid a$ and $a \mid b$, then $a = ub$ for some unit $u \in R$, so a is associated to b. Conversely, if a is associated to b, then $a = ub$ with u a unit, so $b = u^{-1}a$. These two equations show that $b \mid a$ and $a \mid b$. Finally, note that if $(a) = (b)$, then $(a) \subseteq (b)$ and $(b) \subseteq (a)$, so $b \mid a$ and $a \mid b$ by the previous lemma. The converse is also true similarly. □

Lemma 2.10. *The relation of association is an equivalence relation, that is, it satisfies the following three conditions:*

- (*Reflexive*) *If $a \in R$, then $a \sim a$.*
- (*Symmetric*) *If $a \sim b$, then $b \sim a$.*
- (*Transitive*) *If $a \sim b$ and $b \sim c$, then $a \sim c$.*

(*We will deal with equivalence relations in more detail below.*)

Proof (*Jimmy Jia*) Reflexivity: For every $a \in R$, $a = 1.a$ and 1 is a unit, so $a \sim a$.

Symmetry: As $a \sim b$, $a = ub$ with u being a unit. Thus $b = u^{-1}a$ and u^{-1} is a unit, so $b \sim a$.

Transitivity: As $a \sim b$, $a = u_1 b$ with u_1 a unit. Similarly, $b \sim c$ so $b = u_2 c$ with u_2 a unit. Thus $a = u_1 b = (u_1 u_2)c$, and as $u_1 u_2$ is a unit (as $(u_1 u_2)(u_2^{-1} u_1^{-1}) = 1$), we have $a \sim c$. □

We need one more definition before we can define unique factorisation, and that is what we are factoring things into. This idea is used to define 'prime numbers' in high school, but we will reserve that term for an even more fundamental property.

Definition 2.11. Let R be an integral domain. A non-zero element $a \in R$ is an **irreducible** if a is not a unit, and whenever $a = bc$ for some $b, c \in R$, one of b and c is a unit (and so the other is an associate of a).

Lemma 2.12. *Every associate of an irreducible element is irreducible.*

Proof (*Dan Strimpel*) Let b be irreducible and let $b \sim a$, so $b = au$ with u a unit. If a is not irreducible, then $a = de$ for some non-units d and e, so $b = deu$. Now eu is not a unit, for if it were, then $e = (eu)u^{-1}$ would be a unit, as the units form a group (Proposition 1.33). So $b = d(eu)$ with d and eu non-units, contradicting the irreducibility of b. So a is irreducible. □

Example 2.13. We give the standard example of a ring which does not have unique factorisation. Consider the ring $R = \mathbb{Z}[\sqrt{-5}]$. Then we have

$$6 = 2.3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

To finish the task, we must show that all of 2, 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducibles in this ring and that they are not associated to each other.

Proof (*Theo Johnson-Freyd*) Consider the *norm* function

$$\begin{array}{rccc} N: & \mathbb{Z}[\sqrt{-5}] & \to & \mathbb{Z} \\ & a + b\sqrt{-5} & \mapsto & a^2 + 5b^2 \end{array}$$

It is not hard to check that $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}]$, and that $N(\alpha) \geqslant 0$ with equality if and only if $\alpha = 0$. Using this, we see that whenever $\alpha, \beta \neq 0$, then $N(\alpha\beta) \geqslant N(\alpha)$, with equality if and only if β is a unit. For if β is a unit, then $N(\alpha\beta)N(\beta^{-1}) = N(\alpha\beta\beta^{-1}) = N(\alpha)$ so $N(\alpha) \geqslant N(\alpha\beta) \geqslant N(\alpha)$ and thus $N(\alpha\beta) = N(\alpha)$. Conversely, if $N(\alpha\beta) = N(\alpha)$, then $N(\beta) = 1$. But if we write $\beta = a + b\sqrt{-5}$, then $N(\beta) = a^2 + 5b^2 = (a + b\sqrt{-5})(a - b\sqrt{-5}) = 1$, so that β is a unit with inverse $a - b\sqrt{-5}$.

Now consider $N(2) = 4$, $N(3) = 9$ and $N(1 \pm \sqrt{-5}) = 6$. If any of 2, 3, $1 \pm \sqrt{-5}$ can be written as the product of two non-units, then the norms of each of these non-units must properly divide 4, 6 or 9. To show that this is impossible, it suffices to show that $N(\alpha)$ cannot equal 2 or 3. But $N(\alpha) = a^2 + 5b^2$ for some integers a and b. This is clearly impossible, and as neither 2 nor 3 divides $1 \pm \sqrt{-5}$ in $\mathbb{Z}[\sqrt{-5}]$, which we can see by considering their norms, we are done. □

Definition 2.14. An integral domain R is called a **unique factorisation domain** (UFD) if it satisfies the following two conditions:

(UFD1) Every non-zero element $r \in R$ can be written in the form

$$r = ua_1 \cdots a_n$$

where u is a unit in R, $n \geqslant 0$ and the a_i are irreducibles of R.

(UFD2) If $ua_1 \cdots u_n = vb_1 \cdots b_m$, where u and v are units and the a_i and b_i are irreducibles, then $n = m$ and $a_i \sim b_{\pi(i)}$ for some permutation π of $\{1, 2, \ldots, n\}$.

Lemma 2.15. *Every field is* (*trivially*) *a UFD.*

Proof (*Brian Rice*) A field F is a division ring, so any non-zero element $r \in F$ has an inverse, and is thus a unit. So (UFD1) is trivially satisfied. Also, since all non-zero elements are units, there are no irreducibles in F. So when an element is written in the form $ua_1a_2 \cdots a_n$ with u a unit and the a_i irreducibles, we must have $n = 0$. Therefore (UFD2) is satisfied. □

We now give the promised definition of primes in an integral domain.

Definition 2.16. An element $p \in R$ is called a **prime** in R if p is neither zero nor a unit, and whenever $a, b \in R$ with $p \mid ab$, then either $p \mid a$ or $p \mid b$ (or both).

Example 2.17. Referring back to example 2.13, we see that in $\mathbb{Z}[\sqrt{-5}]$, irreducibles may not be prime. For $2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5})$, yet 2 does not divide either factor.

The converse is true, though:

Proposition 2.18. *Let R be an integral domain. Then every prime in R is irreducible.*

Proof (*William Meyerson*) Suppose there exists a prime $p \in R$ which is not irreducible. Then it can be written in the form $p = ab$ for some non-units $a, b \in R$. By the definition of a prime, since $p \mid ab$, either $p \mid a$ or $p \mid b$. Without loss of generality, we may assume $p \mid a$. Then there is some $k \in R$ such that $pk = a$, therefore, $(ab)k = pk = a$, so $a(bk) = a = a.1$. By Lemma 1.18, $bk = 1$, and therefore b is a unit by the definition of a unit. However, this contradicts the choice of b, so p must be an irreducible and we are done. □

Theorem 2.19. *Let R be an integral domain. Then R is a UFD if and only if R satisfies* (UFD1) *and*

(UFD2′) *Every irreducible in R is prime.*

Proof (*Justin Blanchard*) We first prove that in an integral domain R satisfying (UFD1), (UFD2′) implies (UFD2). Suppose this is not the case. Of all of those elements $r \in R$ which have two factorisations $r = ua_1a_2 \cdots a_n = vb_1b_2 \cdots b_m$

which do not satisfy (UFD2), choose one with the minimum $m + n$. We will then derive a contradiction.

If $n = 0$, then r is a unit, so $m = 0$ too (as only units divide units, see Lemma 2.5 above), and thus these factorisations would satisfy (UFD2). So we may assume $n > 0$.

We first show that none of the b_i are associates of a_1. Assume there is some b_i such that $a_1 \sim b_i$. Then $b_i = wa_1$ for some unit w, and by rearranging the order of factors, we can assume further that $i = 1$. Thus $a_1(ua_2 \cdots a_n) = a_1((vw)b_2 \cdots b_m)$, and by multiplicative cancellation in an integral domain, we have $ua_2 \cdots a_n = (vw)b_2 \cdots b_m$. Since R^* is an abelian group under multiplication, vw is a unit, and we have another pair of factorisations which have a smaller $m + n$. (UFD2) cannot hold here, because if it did the original factorisations would also satisfy (UFD2), contrary to our assumption. Thus a_1 has no associates among the b_i.

Now $vb_1b_2 \cdots b_m = (a_1)(ua_2a_3 \cdots a_n)$ implies that $a_1 \mid vb_1b_2 \cdots b_m$. Let i be the smallest non-negative integer such that $a_1 \mid vb_1b_2 \cdots b_i$. If $i > 0$, then $a_1 \mid (vb_1b_2 \cdots b_{i-1})b_i$ but $a_1 \nmid vb_1b_2 \cdots b_{i-1}$. As a_1 is irreducible and thus prime by (UFD2′), either $a_1 \mid vb_1b_2 \cdots b_{i-1}$ or $a_1 \mid b_i$. As the former is not true by the choice of i, we must have $a_1 \mid b_i$. Since b_i is an irreducible, either a_1 is a unit, which is not the case, or $a_1 \sim b_i$, which we have shown to be impossible. Thus we must have $i = 0$, so we have $a_1 \mid v$, hence a_1 is a unit, again by Lemma 2.5. This is impossible because a_1 is an irreducible and so by definition not a unit. This contradiction proves that no such pair of factorisations can exist, and so R must satisfy (UFD2).

Conversely, assume that (UFD1) and (UFD2) hold; we must show that (UFD2′) also holds. Let $a \in R$ be an irreducible and assume that $a \mid bc$. If either b or c is zero, then a divides it, so we may assume that both b and c are non-zero. By definition, there must exist $d \in R$ such that $ad = bc$; moreover, $d \neq 0$ because R is an integral domain and so $bc \neq 0$. Using (UFD1), we can write $b = u_1b_1b_2 \cdots b_m$, $c = u_2c_1c_2 \cdots c_n$ and $d = u_3d_1d_2 \ldots d_r$. Since $ad = bc$, we have

$$a(u_3d_1d_2 \cdots d_r) = (u_1b_1b_2 \cdots b_m)(u_2c_1c_2 \cdots c_n),$$

so that

$$u_3ad_1d_2 \cdots d_r = (u_1u_2)b_1b_2 \cdots b_mc_1c_2 \cdots c_n.$$

By (UFD2), either some $b_i \sim a$ or some $c_j \sim a$. If the former is true, clearly $a \mid b$; if the latter is true, then $a \mid c$. Thus a is prime, and all irreducibles in R are prime. □

Definition 2.20. A **Euclidean domain** (ED) is an integral domain together with a **Euclidean function** $\phi : R \setminus \{0\} \to \mathbb{N}$ (where $\mathbb{N} = \{0, 1, 2, \dots\}$) such that:

(ED1) If $a \mid b$ then $\phi(a) \leqslant \phi(b)$, and

(ED2) Given $b \in R$ and $a \in R \setminus \{0\}$, there exist $q, r \in R$ such that $b = aq + r$ and either $r = 0$ or $\phi(r) < \phi(a)$.

Example 2.21. The rings $\mathbb{Z}$ and $\mathbb{R}[x]$ are Euclidean domains. In fact $k[x]$ is an ED for any field k.

Proof (*Taktin Oey*) We show that $\mathbb{Z}$ is a Euclidean domain. Define $\phi : \mathbb{Z} \to \mathbb{N}$ by $\phi(n) = |n|$. Then if $a, b \in \mathbb{Z}$ are non-zero and $a \mid b$, there exists a non-zero c with $ac = b$, so $|c| \geqslant 1$. Thus $|ac| = |b|$ implies $|b| = |a||c| \geqslant |a|$, so condition (ED1) is satisfied. For (ED2), if $a \mid b$, then $b = aq + 0$ for some $q \in \mathbb{Z}$ as required. So assume $a \nmid b$ and $a \neq 0$. Then the set $\{aq : q \in \mathbb{Z} \text{ and } aq \leqslant b\}$ is non-empty and bounded above: $|a|(|b| + 1) > |a||b| \geqslant |b|$ as $|a| \geqslant 1$, so if $a > 0$, then aq is greater than b for all $q \geqslant |b| + 1$ and less than b for all $q \leqslant -(|b| + 1)$, and vice versa if $a < 0$. Thus this set has a greatest element, say $g = aq_0 \leqslant b$. Let $r = b - g$, and note that $r \geqslant 0$. Suppose that $r \geqslant |a|$. If $a > 0$, then $r = b - g \geqslant a$, so $q + a = a(q_0 + 1) \leqslant b$, but this contradicts the choice of g, as $g = aq_0 < a(q_0 + 1) \leqslant b$. Similarly, if $a < 0$, then $r = b - g \geqslant -a$, so $g = aq_0 < a(q_0 - 1) \leqslant b$, again contradicting the choice of g. Hence $|r| < |a|$ and $b = aq_0 + r$ as required for (ED2).

We now show similarly that $k[x]$ is a Euclidean domain. We define the *degree* of a non-zero polynomial $f = \sum_{k=0}^{\infty} a_k x^k$ by $\deg f = \max\{k : a_k \neq 0\}$. Then we let our Euclidean function be $\phi(f) = \deg f$. If $f \mid g$ for $f, g \in k[x]$, then there is some non-zero $h \in k[x]$ with $fh = g$. It is easy to show that $\deg g = \deg f + \deg h$, so $\deg g \geqslant \deg f$ as $\deg h \geqslant 0$, and (ED1) is satisfied. For (ED2), we argue similarly to the above. Assume $f, g \in k[x]$ with $f \neq 0$. If $f \mid g$, then $g = fh + 0$ and we are done, so assume that $f \nmid g$. Consider the set $S = \{g - fh : h \in k[x]\}$. Since the degree function takes non-negative integer values, there must exist some polynomial in S whose degree is minimal in S. Say $\deg(g - fh_0) \leqslant \deg(g - fh)$ for all $h \in k[x]$ and set $r = g - fh_0$. If we can show that $\deg r < \deg f$, then since we have $g = fh_0 + r$, we will be done. So assume $\deg r \geqslant \deg f$. Then let $a_m x^m$ be the leading term of f and $b_n x^n$ be the leading term of r. By assumption, $m \leqslant n$, so consider $r_1 = r - (b_n a_m^{-1} x^{n-m}) f = g - f(h_0 + b_n a_m^{-1} x^{n-m})$. The coefficient of x^n in r_1 is clearly 0, and since $\deg((b_n a_m^{-1} x^{n-m}) f) = (n - m) + \deg f = n$, we must have $\deg r_1 < n$, contradicting our assumption that $\deg r$ was minimal. Hence $\deg r < \deg f$, and we are done. □

Exercise 2.22. In the ring $\mathbb{Z}[i]$ we can define a *norm* function by $N(a + bi) = a^2 + b^2$. This norm function is a Euclidean function on $\mathbb{Z}[i]$. It then follows from Theorem 2.30 below that this ring is a UFD.

Proof (*Boris Alexeev*) By definition 2.20, the first condition on a Euclidean function is that it maps $\mathbb{Z}[i] \setminus \{0\} \to \mathbb{N}$. This is clear from our definition of $N(a + bi)$. The other two conditions are proven as follows.

(ED1) We first notice that our norm function can be written as $N(z) = z\bar{z}$ (where $\bar{z}$ denotes the complex conjugate of z). By the properties of the complex conjugate, we get $N(\alpha\beta) = (\alpha\beta)\overline{(\alpha\beta)} = (\alpha\bar{\alpha})(\beta\bar{\beta}) = N(\alpha)N(\beta)$. (This can

easily be proven directly.) Also, we note that $N(\alpha) \geqslant 1$ for all non-zero α. To show (ED1), assume that $\alpha \mid \beta$, so there exists $\gamma \in \mathbb{Z}[i]$ such that $\alpha\gamma = \beta$. Thus $N(\alpha\gamma) = N(\alpha)N(\gamma) = N(\beta)$, and since γ is trivially non-zero, $N(\gamma) \geqslant 1$, giving $N(\beta) = N(\alpha)N(\gamma) \geqslant N(\alpha)$, so $N(\alpha) \leqslant N(\beta)$ as required.

(ED2) For this part, we will give a constructive division algorithm. Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\alpha = a + bi \neq 0$ and $\beta = c + di$. If $\alpha \mid \beta$, then $\beta = \alpha\gamma$ and (ED2) is satisfied. Otherwise, we have $N(\alpha) \neq 0$ and

$$\frac{\beta}{\alpha} = \frac{c+di}{a+bi} = \frac{(c+di)(a-bi)}{a^2+b^2} = \frac{ac+bd}{N(\alpha)} + \frac{ad-bc}{N(\alpha)}i.$$

Let e and f be the nearest integers to the rational numbers $(ac+bd)/N(\alpha)$ and $(ad-bc)/N(\alpha)$ respectively, and set $\gamma = e + fi$. Then $\beta/\alpha - \gamma = x + yi$ with $|x|, |y| \leqslant \frac{1}{2}$, so $x^2 + y^2 \leqslant \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Thus if we set $\delta = \beta - \alpha\gamma$, we have $N(\delta) = N(\alpha)(x^2+y^2) \leqslant \frac{1}{2}N(\alpha) < N(\alpha)$, so we have $\beta = \alpha\gamma + \delta$ with $N(\delta) < N(\alpha)$ as required.

For example, take $\alpha = 5 + 2i$ and $\beta = 8 + 12i$. Then

$$\frac{8+12i}{5+2i} = \frac{(8+12i)(5-2i)}{29} = \frac{64+44i}{29} = 2 + 2i + \frac{6-14i}{29},$$

so the quotient is $2 + 2i$ and the remainder is $(8 + 12i) - (2 + 2i)(5 + 2i) = 2 - 2i$. As a check, $N(2 - 2i) = 8$ is less than $N(5 + 2i) = 29$. □

We can now generalise the result of Exercise 1.28 above.

Theorem 2.23. *Every ED is a PID.*

Proof (*David Roe*) Let I be an ideal in a Euclidean domain R. Either $I = 0 = (0)$, in which case I is principal and we are done, or there is some non-zero $a \in I$. Since $X = \{\phi(a) : a \in I,\ a \neq 0\}$ is therefore a nonempty subset of $\mathbb{N}$, it contains a smallest element n. Pick some $a_0 \in I$ with $\phi(a_0) = n$. We will show that $I = (a_0)$. Assume there is some $k \in I$ with $a_0 \nmid k$. Since R is a Euclidean domain, we can find $q, r \in R$ with $k = a_0 q + r$ and either $r = 0$ or $\phi(r) < \phi(a_0)$. If $r = 0$, then $a_0 \mid k$, contrary to the assumption that $a_0 \nmid k$. Thus $\phi(r) < \phi(a_0)$. But $r = k - a_0 q \in I$, so $\phi(r) \in X$ and $\phi(r) < n$, contradicting the fact that n is the smallest element of X. Therefore $a_0 \mid k$ for all $k \in I$, so $I \subseteq (a_0)$. But $a_0 \in I$ so $(a_0) \subseteq I$. Thus $I = (a_0)$ and I is principal, and therefore R is a principal ideal domain. □

We now aim to show that every PID is a UFD. The proof will take several steps.

Lemma 2.24. *Every PID satisfies condition* (UFD2′).

Proof (*Theo Johnson-Freyd*) Let $p \in R$ be irreducible. We must show that p is prime. So assume $p \mid ab$ but $p \nmid a$; we must then show that $p \mid b$. As R is a PID, we have $(p) + (a) = (c)$ for some $c \in R$. So $c \mid ax + py$ for

all $x, y \in R$ (since $(p) = \{py : y \in R\}$ and similarly for (a) and (c), and so $(p)+(a) = \{p'+a' : p' \in (p) \text{ and } a' \in (a)\} = \{ax+py : x, y \in R\}$). In particular, $c \mid p$, so c is either a unit or an associate of p. If $c \sim p$, then $p \mid c$, and as $c \mid a$, we have $p \mid a$, contrary to our assumption. So c must be a unit and hence $(p) + (a) = (1)$, so $ax + py = 1$ for some $x, y \in R$. So $b = abx + pby$, and as p divides both terms of this sum, we have $p \mid b$ as required. □

Definition 2.25. A ring R is said to satisfy the **ascending chain condition (a.c.c.) on ideals** if any chain of ideals

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$$

is eventually constant (that is, $I_n = I_{n+1} = \cdots$ for some n).

Lemma 2.26. *If $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals of R, then $\bigcup_{n=0}^{\infty} I_n$ is an ideal of R.*

Proof (*Pradeep Mouli*) Let $I = \bigcup_{n=0}^{\infty} I_n$. We show that I is an ideal using Lemma 1.24. I is non-empty, as $0 \in I_0$, so $0 \in I$, hence condition (I1) holds. To show that I is closed under addition (condition (I2)), let $x, y \in I$. Then for some m and n, we have $x \in I_m$ and $y \in I_n$. Let $N = \max\{m, n\}$, then $I_m, I_n \subseteq I_N$ as the chain of ideals is ascending. Thus $x, y \in I_N$, so $x + y \in I_N \subseteq I$ as I_N is closed under addition, and (I2) holds. Finally, for the ideal condition (I3), let $x \in I$ and $a \in R$. Then $x \in I_n$ for some n, so $ax \in I_n \subseteq I$, and hence (I3) is satisfied. □

Lemma 2.27. *Every PID satisfies a.c.c. on ideals.*

Proof (*Cristin Kenney*) Let R be a PID and $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ be an ascending chain of ideals of R. Then by the previous lemma, $I = \bigcup_{n=0}^{\infty} I_n$ is an ideal of R. Since R is a PID, $I = (c)$ for some $c \in R$. Since $c \in I$, we must have $c \in I_n$ for some n, by the definition of I. But then $(c) \subseteq I_n \subseteq I = (c)$, so $I_n = I_{n+1} = \cdots = I = (c)$. Therefore R satisfies a.c.c. on ideals. □

Lemma 2.28. *Every PID satisfies* (UFD1).

Proof. Assume that R be a PID which does not satisfy (UFD1). Then there is some element $r_0 \in R$ which does not have a factorisation into irreducibles. It is not possible for r_0 to be irreducible, for then $r_0 = 1.r_0$ would be our required factorisation. So we can write $r_0 = r_1 r_1'$ with neither r_1 nor r_1' being a unit. If both r_1 and r_1' had factorisations into irreducibles, then r_0 would also. So at least one of them, say r_1, does not have a factorisation into irreducibles. We can then repeat this argument with r_1, writing $r_1 = r_2 r_2'$ with neither r_2 nor r_2' a unit, and we may assume that r_2 does not have a factorisation into irreducibles. Continuing this, we get a sequence r_0, r_1, …, with $r_{i+1} \mid r_i$ for each i, and none of the r_i having a factorisation into irreducibles. As $r_{i+1} \mid r_i$, we have $(r_i) \subseteq (r_{i+1})$ for each i by Lemma 2.7. We cannot have equality here, as if

we did, then we would have $r_i \sim r_{i+1}$ by Lemma 2.9, implying $r_{i+1} = ur_i$ for some unit u. But as $r_i = r_{i+1}r'_{i+1}$, this would give $r_i = ur_ir'_{i+1}$, so $ur'_{i+1} = 1$ by cancellation, contrary to the assumption that r'_{i+1} is not a unit. Thus $(r_i) \subset (r_{i+1})$ for each i, and we have found an infinite ascending chain of ideals $(r_0) \subset (r_1) \subset (r_2) \subset \cdots$, which is impossible by the previous lemma. Thus every PID must satisfy (UFD1). □

We have therefore proved

Theorem 2.29. *Every PID is a UFD.*

Proof (*Cristin Kenney*) This follows from Lemma 2.24, which states that every PID satisfies (UFD2′), Lemma 2.28, which states that every PID satisfies (UFD1), and Theorem 2.19, which states that any integral domain which satisfies (UFD1) and (UFD2′) is a UFD. □

This result together with Theorem 2.23 also shows the following result.

Theorem 2.30. *Every ED is a UFD.* □

This can also be proved slightly more directly, but the proof is very similar in essence to the route we took.

Corollary 2.31. *If k is a field, then $k[x]$ is a UFD. Also, $\mathbb{Z}$ is a UFD.*

Proof. This follows from Example 2.21 and the above theorem, or from Exercise 1.28 and Theorem 2.29. □

The rings we have been considering are called Euclidean because the Euclidean algorithm can be applied to them to find gcds. So we now formally define a greatest common divisor of two elements in a ring. We have to do it in the following way, as there may not be any concept of 'greatest' in the ring.

Definition 2.32. Let $a, b \in R$. d is a **greatest common divisor** (gcd) of a and b if $d \mid a$ and $d \mid b$, and whenever $c \mid a$ and $c \mid b$, we have $c \mid d$.

There may be lots of gcds:

Lemma 2.33. *If d is a gcd of a and b, then d' is also a gcd of a and b if and only if $d' \sim d$.*

Proof (*George Khachatryan*) If d' is also a gcd of a and b, then by the definition of a gcd, we have $d' \mid a$ and $d' \mid b$. But by the other part of the definition, this means that $d' \mid d$. Similarly, as $d \mid a$ and $d \mid b$, we have $d \mid d'$ as d' is a gcd. Then by Lemma 2.9, $d' \sim d$.

Conversely, assume that $d' \sim d$. Now as $d \mid a$ and $d \mid b$, we have $a = dn$ and $b = dm$ for some $m, n \in R$. As $d = d'u$ with u a unit, we have $a = d'un$ and $b = d'um$, so $d' \mid a$ and $d' \mid b$, so d' is a common divisor of a and b. Now assume

$c \in R$ with $c \mid a$ and $c \mid b$, so $c \mid d$. But then $d = ck$ for some $k \in R$, so $d'u = ck$ and $d' = cku^{-1}$, so $c \mid d'$. Therefore the two conditions for a gcd are satisfied by d', so d' is a gcd of a and b. □

Exercise 2.34. Using this definition, find all the greatest common divisors of 6 and 9 in the ring $\mathbb{Z}$ and of the polynomials $(x+1)(x+2)$ and $(x+1)(x^2+2)$ in the ring $\mathbb{R}[x]$. How does this definition relate to the traditional one in the case of $\mathbb{Z}$?

Answer. The greatest common divisors of 6 and 9 in $\mathbb{Z}$ are ± 3, and of the polynomials $(x+1)(x+2)$ and $(x+1)(x^2+2)$ in $\mathbb{R}[x]$ are all polynomials of the form $\lambda(x+1)$ where $\lambda \in \mathbb{R}$ is non-zero. For the case of $\mathbb{Z}$, the traditional definition would give a unique gcd, here $\gcd(6,9) = 3$, whereas our definition regards both 3 and -3 as gcds.

Exercise 2.35. Extend the definition to define a gcd of the elements $a_1, \ldots, a_n$.

Answer. Let $a_1, \ldots, a_n \in R$. d is a greatest common divisor of the a_i if $d \mid a_i$ for $i = 1, \ldots, n$, and whenever $c \mid a_i$ for $i = 1, \ldots, n$, we have $c \mid d$.

Proposition 2.36. *If R is an integral domain and $a, b, c \in R$ with $(a,b) = (a) + (b) = (c)$, then c is a gcd of a and b and we have $c = xa + yb$ for some $x, y \in R$. In the case that R is a PID, then we always have $(a) + (b) = (c)$ for some $c \in R$, so any a and b in R have a gcd.*

Proof. We have $(a) \subseteq (a,b) = (c)$, so $c \mid a$; similarly $c \mid b$, so c is a common divisor of a and b. Furthermore, $c \in (c) = (a,b)$, so $c = xa + yb$ for some $x, y \in R$ by Lemma 1.26. If $c' \mid a$ and $c' \mid b$, then we have $c' \mid xa + yb = c$. Thus c is a gcd of a and b. □

In the case that the ring is Euclidean, we can determine gcds using the **Euclidean algorithm**. Let ϕ be a Euclidean function on R, and let a and b be non-zero elements of R. We can determine a gcd of a and b by performing the following calculation:

$$\begin{aligned}
b &= aq_1 + r_1 && 0 \leqslant \phi(r_1) < \phi(a) \\
a &= r_1 q_2 + r_2 && 0 \leqslant \phi(r_2) < \phi(r_1) \\
r_1 &= r_2 q_3 + r_3 && 0 \leqslant \phi(r_3) < \phi(r_2) \\
r_2 &= r_3 q_4 + r_4 && 0 \leqslant \phi(r_4) < \phi(r_3) \\
&\cdots \\
r_{n-3} &= r_{n-2} q_{n-1} + r_{n-1} && 0 \leqslant \phi(r_{n-1}) < \phi(r_{n-2}) \\
r_{n-2} &= r_{n-1} q_n + r_n && 0 \leqslant \phi(r_n) < \phi(r_{n-1}) \\
r_{n-1} &= r_n q_{n+1}, &&
\end{aligned}$$

terminating when the remainder is zero.

Proposition 2.37. *This algorithm always terminates, and when it does, r_n is a gcd of a and b.*

Proof (*Theo Johnson-Freyd*) Were it not to terminate, the $\phi(r_i)$ would form a strictly decreasing sequence of non-negative integers, which is impossible. Now we will show by induction that r_n is a common divisor of a and b. To simplify the notation, we will write $r_0 = a$ and $r_{-1} = b$, and we aim to show that $r_n \mid r_{n-i}$ for $i = 0, 1, \ldots, n+1$. Clearly $r_n \mid r_n$ and $r_n \mid r_{n-1} = r_n q_{n+1}$, which shows that the statement is true for $i = 0$ and $i = 1$. Now assume that $r_n \mid r_{n-j}$ for $0 \leqslant j \leqslant i$, where $i \geqslant 1$. Then we have $r_n \mid r_{n-i}q_{n-i+1} + r_{n-(i-1)} = r_{n-(i+1)}$. Thus, by induction, we deduce that $r_n \mid r_{n-i}$ for all $i \geqslant 0$, and in particular, $r_n \mid r_0 = a$ and $r_n \mid r_{-1} = b$, so that r_n is a common divisor of a and b.

To show that r_n is a gcd of a and b, assume that $c \mid a$ and $c \mid b$. We show by induction that $c \mid r_i$ for each $i \geqslant -1$. This is true for $i = -1$ and $i = 0$ by assumption, so assume that $c \mid r_j$ for $-1 \leqslant j \leqslant i$, where $i \geqslant 0$. Then $c \mid r_{i-1} - r_i q_i = r_{i+1}$, so by induction, $c \mid r_i$ for $-1 \leqslant i \leqslant n$. In particular, $c \mid r_n$, so that r_n is a gcd of a and b. □

More interestingly, we can use this algorithm to work backwards, first expressing r_n as a linear combination of r_{n-1} and r_{n-2} using the penultimate equation, then this and the preceding equation to express r_n as a linear combination of r_{n-2} and r_{n-3}, and so on. Eventually we will end up with an expression for r_n, which is a gcd of a and b, as a linear combination of a and b. This is more constructive than the result of Proposition 2.36, in particular if we are able to explicitly determine the quotient and remainder for the division steps, as we are for $\mathbb{Z}$ and $\mathbb{Q}[x]$.

The last main result that we wish to prove in this chapter is based on Gauss' lemma, and states that if R is a UFD, then so is $R[x]$. But we will need to do some important preparatory work on equivalence relations and fields of fractions first. This material is used throughout mathematics, and we will meet it again when we consider quotient rings in Chapter 3.

2.1 Relations, equivalence relations and partitions

Definition 2.38. Let X and Y be any sets. The **Cartesian product** of X and Y is defined by

$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\},$$

the set of all ordered pairs of elements from X and Y.

Definition 2.39. A **(binary) relation** on X is a subset $\mathcal{R} \subseteq X \times X$. We often write $x \mathcal{R} y$ to mean $(x, y) \in \mathcal{R}$ and $x \not\mathcal{R} y$ to mean $(x, y) \notin \mathcal{R}$.

Example 2.40. Let $X = \mathbb{Z}$. Then the subset $\{(m, n) : m < n\} \subset \mathbb{Z} \times \mathbb{Z}$ defines the relation 'less than'. The subset $\{(m, n) : m = n\}$ is the relation of equality.

Definition 2.41. The following terms are used to describe some special features of relations. Here $\mathcal{R}$ is a relation on a set X.

- $\mathcal{R}$ is called **reflexive** if $x \mathcal{R} x$ for all $x \in X$.
- $\mathcal{R}$ is called **symmetric** if for all $x, y \in X$, $x \mathcal{R} y$ implies that $y \mathcal{R} x$.
- $\mathcal{R}$ is called **anti-symmetric** if for all $x, y \in X$, $x \mathcal{R} y$ and $y \mathcal{R} x$ together imply that $x = y$.
- $\mathcal{R}$ is called **transitive** if for all $x, y, z \in X$, $x \mathcal{R} y$ and $y \mathcal{R} z$ together imply that $x \mathcal{R} z$.

Exercise 2.42. Given the set $X = \{1, 2, 3\}$, for each of the following conditions, find a relation on X which satisfies it:

(a) The relation is reflexive and symmetric, but not transitive.

(b) The relation is reflexive and transitive, but not symmetric.

(c) The relation is symmetric and transitive, but not reflexive.

(d) The relation is reflexive, transitive and anti-symmetric.

Answer (*Michael Kolodrubetz*) We given relation tables for the four relations. In each table, a '$*$' in row i and column j indicates that $(i, j) \in \mathcal{R}_n$. The four tables respresenting $\mathcal{R}_1$ through $\mathcal{R}_4$ correspond to the four questions.

$\mathcal{R}_1$	1	2	3
1	$*$	$*$	
2	$*$	$*$	$*$
3		$*$	$*$

$\mathcal{R}_2$	1	2	3
1	$*$	$*$	$*$
2		$*$	$*$
3			$*$

$\mathcal{R}_3$	1	2	3
1			
2			
3			

$\mathcal{R}_4$	1	2	3
1	$*$		
2	$*$	$*$	
3	$*$	$*$	$*$

Note that $\mathcal{R}_2$ and $\mathcal{R}_4$ are the relations given by $x \leqslant y$ and $x \geqslant y$ respectively. □

Definition 2.43. The following special types of relations are so important that they are given special names.

- A relation which is reflexive, symmetric and transitive is called an **equivalence relation**.
- A relation which is reflexive, transitive and anti-symmetric is called a **partial order** on the set, and the set is called a **partially ordered set** or **poset**. If in addition, for every pair $x, y \in X$, either $x = y$, $x \mathcal{R} y$ or $y \mathcal{R} x$, then the relation is called a **(total) order**, and the set is called a **totally ordered set** or just an **ordered set**. (Some authors require that these relations have $(x, x) \notin \mathcal{R}$ for all $x \in X$; this is essentially the difference between the relations $\leqslant$ and $<$.)

Exercise 2.44. Give some examples of equivalence relations, partial orders which are not total orders, and total orders, either for the set $X = \{1, 2, 3\}$ or for other sets.

Answer (*William Meyerson*) We give examples which work equally on the set X and on any subset of the reals containing X such as $\mathbb{R}$, $\mathbb{Q}$ or $\mathbb{Z}$.

(a) The relation given by $a \mathcal{R} b$ if and only if $a = b$ is an equivalence relation: it is reflexive since $a \mathcal{R} a$ for every $a \in X$ (as $a = a$), symmetric since if $a \mathcal{R} b$, then $b \mathcal{R} a$ (as $b = a$ whenever $a = b$), and transitive since if $a \mathcal{R} b$ and $b \mathcal{R} c$, then $a \mathcal{R} c$ (as $a = c$ whenever $a = b$ and $b = c$).

(b) The relation given by $a \mathcal{R} b$ if and only if $b = a + 2n$ for some non-negative integer n is a partial order. It is easily seen to be reflexive (as $a = a + 2.0$ for every $a \in X$), anti-symmetric (the only time both $a \mathcal{R} b$ and $b \mathcal{R} a$ hold is when $a = b$), and transitive (if $a \mathcal{R} b$ and $b \mathcal{R} c$ then $b = a + 2n$ and $c = b + 2m$, so $c = a + 2(m + n)$ and $a \mathcal{R} c$). However, it is not a total order because 1 and 2 are incomparable (that is, $1 \neq 2$, $1 \not\mathcal{R} 2$ and $2 \not\mathcal{R} 1$).

(c) The relation given by $a \leqslant b$ is a total order. It is a partial order because it can easily be checked to be reflexive (as $a \leqslant a$ for every a), anti-symmetric (the only time both $a \leqslant b$ and $b \leqslant a$ hold is when $a = b$), and transitive (if $a \leqslant b$ and $b \leqslant c$, then $a \leqslant c$). It is clearly a total order as well because for any two numbers a and b, either $a \leqslant b$ or $b \leqslant a$ (or both if $a = b$).

We will be focusing on equivalence relations. An equivalence relation essentially says that certain elements in the set X are to be considered 'the same' in a precise sense depending upon the context. We make this a little more formal, by showing that partitions and equivalence relations are essentially the same concept.

Definition 2.45. A **partition** of a set X consists of a set $\mathcal{X}$ of non-empty subsets of X satisfying the following conditions:

(a) The subsets in $\mathcal{X}$ cover X, that is

$$\bigcup_{Y \in \mathcal{X}} Y = X.$$

(b) The subsets are pairwise disjoint, that is if $Y_1, Y_2 \in \mathcal{X}$ with $Y_1 \neq Y_2$, then $Y_1 \cap Y_2 = \varnothing$.

Note that condition (b) implies that if $Y_1, Y_2 \in \mathcal{X}$ and $Y_1 \cap Y_2 \neq \varnothing$, then $Y_1 = Y_2$.

Definition 2.46. Let $\mathcal{R}$ be a relation on a set X. For $x \in X$, define $\mathcal{R}(x) = \{y \in X : x \mathcal{R} y\}$. In the case that $\mathcal{R}$ is an equivalence relation, $\mathcal{R}(x)$ is called the **equivalence class** of x. Very often one uses the notation $[x]$ or $\bar{x}$ to denote this.

Lemma 2.47. *If $\mathcal{R}$ is an equivalence relation on X, then $x \in \mathcal{R}(x)$ for all $x \in X$. Furthermore, whenever $y \in \mathcal{R}(x)$, $\mathcal{R}(x) = \mathcal{R}(y)$.*

Proof (*Pradeep Mouli*) As $\mathcal{R}$ is an equivalence relation and so $\mathcal{R}$ is reflexive, we have $x \mathcal{R} x$ for all $x \in X$, so $x \in \mathcal{R}(x)$ for all $x \in X$. Now assume $y \in \mathcal{R}(x)$, so $x \mathcal{R} y$ by the definition of $\mathcal{R}(x)$, and thus $y \mathcal{R} x$ by symmetry. Now assume $z \in \mathcal{R}(x)$, so $x \mathcal{R} z$. But as $y \mathcal{R} x$, we have $y \mathcal{R} z$ by transitivity, hence $z \in \mathcal{R}(y)$. Thus $\mathcal{R}(x) \subseteq \mathcal{R}(y)$. Conversely, if $z \in \mathcal{R}(y)$, then $y \mathcal{R} z$, and since $x \mathcal{R} y$, it follows that $x \mathcal{R} z$, so $z \in \mathcal{R}(x)$. Thus $\mathcal{R}(y) \subseteq \mathcal{R}(x)$. It follows that $\mathcal{R}(x) = \mathcal{R}(y)$ as required. □

Proposition 2.48. *If $\mathcal{R}$ is an equivalence relation on X, then the set of equivalence classes $P(\mathcal{R}) = \{\mathcal{R}(x) : x \in X\}$ is a partition of X.*

Proof (*Justin Blanchard*) By definition, $\mathcal{R}(x) \subseteq X$, therefore

$$\bigcup_{Y \in P(\mathcal{R})} Y = \bigcup_{x \in X} \mathcal{R}(x) \subseteq X.$$

We also know that $\varnothing \notin P(\mathcal{R})$ because $x \in \mathcal{R}(x)$ for each $x \in X$, so all equivalence classes are non-empty. This fact also leads to the conclusion that

$$X = \bigcup_{x \in X} \{x\} \subseteq \bigcup_{x \in X} \mathcal{R}(x) = \bigcup_{Y \in P(\mathcal{R})} Y.$$

Since we have inclusions both ways, it follows that $P(\mathcal{R})$ covers X, so condition (a) of Definition 2.45 is satisfied. Finally, let $Y_1, Y_2 \in P(\mathcal{R})$. By definition, $Y_1 = \mathcal{R}(x_1)$ and $Y_2 = \mathcal{R}(x_2)$ for some $x_1, x_2 \in X$. Assume that Y_1 and Y_2 are not disjoint. Choose some element $x \in Y_1 \cap Y_2$. Then as $x \in \mathcal{R}(x_1)$, we have $\mathcal{R}(x) = \mathcal{R}(x_1)$ by the preceding lemma, and similarly $x \in \mathcal{R}(x_2)$ implies that $\mathcal{R}(x) = \mathcal{R}(x_2)$. Therefore $Y_1 = Y_2$, and it follows that distinct subsets in $P(\mathcal{R})$ are disjoint, so condition (b) is satisfied. Thus $P(\mathcal{R})$ is a partition of X. □

Proposition 2.49. *Let $\mathcal{X}$ be a partition of a set X. Define a relation $R(\mathcal{X})$ on X by $(x, y) \in R(\mathcal{X})$ if and only if x and y lie in the same subset in $\mathcal{X}$. Then $R(\mathcal{X})$ is an equivalence relation on X.*

Proof. We must show that $R(\mathcal{X})$ is reflexive, symmetric and transitive. Let $x \in X$. Then as $\mathcal{X}$ is a partition, x lies in some subset of $\mathcal{X}$ by condition (a) of Definition 2.45, so x and x lie in the same subset in $\mathcal{X}$ giving $(x, x) \in R(\mathcal{X})$. Now if $(x, y) \in R(\mathcal{X})$, then x and y lie in the same subset of $\mathcal{X}$, and so y and x also lie in the same subset of $\mathcal{X}$, that is $(y, x) \in R(\mathcal{X})$. Finally, if $(x, y), (y, z) \in R(\mathcal{X})$, then x and y lie in the same subset of $\mathcal{X}$, say Y_1, and y and z also lie in the same subset of $\mathcal{X}$, say Y_2. Now $y \in Y_1 \cap Y_2$, so that $Y_1 = Y_2$ as $\mathcal{X}$ is a partition. Hence x and z both lie in the same subset of $\mathcal{X}$, so that $(x, z) \in R(\mathcal{X})$. Thus $R(\mathcal{X})$ is an equivalence relation as required. □

Proposition 2.50. *The operations defined in the previous two propositions are mutually inverse, that is given an equivalence relation $\mathcal{R}$ on X, $R(P(\mathcal{R})) = \mathcal{R}$, and given a partition $\mathcal{X}$ of X, $P(R(\mathcal{X})) = \mathcal{X}$.*

Proof. Noting that both equivalence relations and partitions are defined as sets, we recall that to show that sets A and B are equal, we must show that $A \subseteq B$ and $B \subseteq A$, or that $x \in A$ implies $x \in B$ and conversely.

We first show that, given an equivalence relation $\mathcal{R}$ on X, $R(P(\mathcal{R})) = \mathcal{R}$. Let $x, y \in X$. Assume that $(x, y) \in R(P(\mathcal{R}))$. This means that x and y lie in the same part of the partition $P(\mathcal{R})$ by definition of $R(P(\mathcal{R}))$. But as $x \in \mathcal{R}(x)$ and $y \in \mathcal{R}(y)$, this means that $\mathcal{R}(x) = \mathcal{R}(y)$ by the definition of $P(\mathcal{R})$. So $y \in \mathcal{R}(x)$, or $(x, y) \in \mathcal{R}$. Thus $R(P(\mathcal{R})) \subseteq \mathcal{R}$.

Conversely, assume $(x, y) \in \mathcal{R}$. Then $y \in \mathcal{R}(x)$ by definition of $\mathcal{R}(x)$, so by Lemma 2.47, we have $\mathcal{R}(x) = \mathcal{R}(y)$. Thus as $x \in \mathcal{R}(x)$ and $y \in \mathcal{R}(y)$, x and y lie in the same part of $P(\mathcal{R})$, and so $(x, y) \in R(P(\mathcal{R}))$. Thus $\mathcal{R} \subseteq R(P(\mathcal{R}))$, and it follows that $R(P(\mathcal{R})) = \mathcal{R}$ as required.

Moving on now to the second case: we are given a partition $\mathcal{X}$ of X and must show that $P(R(\mathcal{X})) = \mathcal{X}$. Assume $Y \in P(R(\mathcal{X}))$. Then Y is of the form $R(\mathcal{X})(x)$ for some $x \in X$. Thus $Y = \{y \in X : (x, y) \in R(\mathcal{X})\}$. But by the definition of $R(\mathcal{X})$, we have $(x, y) \in R(\mathcal{X})$ precisely when x and y lie in the same subset of $\mathcal{X}$, so Y is the set of all $y \in X$ lying in the same subset of $\mathcal{X}$ as x. Since x is in a unique subset of $\mathcal{X}$, it follows that Y is this unique subset, so $Y \in \mathcal{X}$. Thus $P(R(\mathcal{X})) = \mathcal{X}$.

Conversely, assume that $Y \in \mathcal{X}$. Let $x \in Y$. Then consider the subset $Y' \in P(R(\mathcal{X}))$ containing x. We show that $Y' = Y$, from which it will follow that $Y \in P(R(\mathcal{X}))$. If $y \in Y$, then x and y are in the same part of $\mathcal{X}$, from which we have $(x, y) \in R(\mathcal{X})$. Then $y \in R(\mathcal{X})(x)$ by definition, so x and y lie in the part $R(\mathcal{X})(x)$ of $P(R(\mathcal{X}))$, showing that $y \in Y'$. If we now assume $y' \in Y'$, then $y' \in R(\mathcal{X})(x)$ (as this is the part of $P(R(\mathcal{X}))$ containing x), so $(x, y') \in R(\mathcal{X})$ and hence x and y' lie in the same part of $\mathcal{X}$, that is, $y' \in Y$. Hence $Y' = Y$, and it follows that $\mathcal{X} \subseteq P(R(\mathcal{X}))$, giving $P(R(\mathcal{X})) = \mathcal{X}$ as required. □

Definition 2.51. Let $\mathcal{R}$ be an equivalence relation on X. We define the **quotient** of X by $\mathcal{R}$, written $X/\mathcal{R}$, to be the set $P(\mathcal{R})$.

Thus we can think of equivalence relations and partitions as one and the same thing: an equivalence relation partitions a set into parts called equivalence classes, where the elements of a given part are considered equal to each other, but elements in different parts are considered unequal.

Often, the notation $a \equiv b$ is used to indicate an equivalence relation. Whenever you see this notation, you should check to see that you are indeed working with an equivalence relation.

Examples 2.52. Here are some typical examples of the use of equivalence relations in pure mathematics.

- When we build the rationals from the integers and make identifications such as $\frac{1}{2} = \frac{3}{6}$, we are essentially defining an equivalence relation on pairs of integers, as we explain in more detail below.
- When we use modular arithmetic, we define an equivalence relation on the integers with m being equivalent to n whenever the modulus divides $m - n$.
- In topology, when we identify edges of a square to make a torus, we are essentially defining an equivalence relation on the points of the square.
- Later on, we will see that quotient rings are defined by equivalence relations, as are factor groups (also called quotient groups) in group theory.

We are now able to return to the point at hand, and define a generalisation of the formation of $\mathbb{Q}$ from $\mathbb{Z}$.

Let R be an integral domain. Let X be the set $R \times (R \setminus \{0\})$. We define an equivalence relation $\sim$ on X by $(a', b') \sim (a, b)$ if and only if $a'b - ab' = 0$. (Check that this really is an equivalence relation!) We define addition and multiplication on X by

$$\begin{aligned}(a, c) + (b, d) &= (ad + bc, cd)\\ (a, c)(b, d) &= (ab, cd)\end{aligned}$$

Lemma 2.53. *If $(a', c') \sim (a, c)$ and $(b', d') \sim (b, d)$, we have $(a', c') + (b', d') \sim (a, c) + (b, d)$ and $(a', c')(b', d') \sim (a, c)(b, d)$.*

Proof (*Hubert Hwang*) By definition, we have $a'c - ac' = 0$ and $b'd - bd' = 0$.

To show $(a', c')+(b', d') \sim (a, c)+(b, d)$, we must show that $(a'd'+b'c', c'd') \sim (ad + bc, cd)$. We have

$$\begin{aligned}(a'd' + b'c')(cd) - (ad + bc)(c'd') &= a'cdd' + b'cc'd - ac'dd' - bcc'd'\\ &= (a'c - ac')dd' + (b'd - bd')cc'\\ &= 0.dd' + 0.cc' = 0,\end{aligned}$$

so that $(a'd' + b'c', c'd') \sim (ad + bc, cd)$ as required. Similarly, to show that $(a'b', c'd') \sim (ab, cd)$, we check

$$\begin{aligned}a'b'cd - abc'd' &= (a'c)(b'd) - (ac')(bd')\\ &= (a'c - ac')(b'd) + (ac')(b'd - bd')\\ &= 0.b'd + (ac').0 = 0\end{aligned}$$

as required. □

Hence addition and multiplication are well-defined on the set $K = X/\sim$. As with the rationals, we write a/b for the equivalence class of (a, b) in this set, so that $a/b + c/d = (ad + bc)/bd$ and $(a/b)(c/d) = ac/bd$.

Proposition 2.54. *In the above setup, K is a field, called the* ***field of fractions*** *of R. The map $\phi : R \to K$ given by $\phi(a) = a/1$ is an injective (one-to-one) homomorphism. [Homomorphisms are special kinds of maps defined at the start of Chapter 3.]*

Proof. This is only a sketch proof; many of the book-keeping details are omitted.

We must show that the axioms of a ring are all satisfied in this case. Let $a/b, c/d \in K$. Axiom (A0) follows by observing that as $(a/b) + (c/d) = (ad + bc)/bd$ and R itself is closed under addition and multiplication, both $ad + bc$ and bd are elements of R. Furthermore, as R is an integral domain, $bd \neq 0$ as $b, d \neq 0$. Hence $(ad + bc)/bd \in K$. Axiom (A1) is routine algebra. For (A2), it suffices to show that $0/1$ is the zero element of K, and for (A3), we show that $(-a)/b$ is the additive inverse of a/b. Axiom (A4) is also elementary algebra.

The multiplicative axioms are similar; here $1/1$ is the multiplicative identity, and b/a is the multiplicative inverse of a/b whenever $a/b \neq 0/1$. Finally, the distributive law is again elementary algebra. Thus K is a field.

To show that ϕ is injective, we need only note that if $a/1 = b/1$, then this means that $(a, 1) \sim (b, 1)$, so that $a.1 = b.1$, or $a = b$. To show that it is a homomorphism, we simply observe that $\phi(a + b) = (a + b)/1 = a/1 + b/1 = \phi(a) + \phi(b)$ and $\phi(ab) = ab/1 = (a/1)(b/1) = \phi(a)\phi(b)$. Also, we have $\phi(1) = 1/1$, which is the multiplicative identity. Thus ϕ is a homomorphism and R embeds naturally into its field of fractions. □

Lemma 2.55. *Assume R is a UFD. Then given $a, b \in R$, we can find a gcd d of a and b. Then if $a = da'$ and $b = db'$, 1 is a gcd of a' and b'. Hence in the above construction, $(a, b) \sim (a', b')$ with 1 being a gcd of a' and b'. That is, in the case of a UFD, we can always write a/b in 'lowest terms' with a and b having no common (prime) factor.*

Proof (*Theo Johnson-Freyd*) We factorise a and b into primes: $a = u \prod_{i=1}^{n} p_i^{\alpha_i}$ and $b = v \prod_{i=1}^{n} p_i^{\beta_i}$, where $p_1, \ldots, p_n$ are the distinct primes dividing a or b and u and v are units. It is easy to see that $d = \prod_{i=1}^{n} p_i^{\min\{\alpha_i, \beta_i\}}$ is a gcd of a and b: that it is a common divisor is obvious, and that it is a greatest common divisor follows from uniqueness of factorisation, as any common factor of a and b must be of the form $c = w \prod_{i=1}^{n} p_i^{\gamma_i}$ with $\gamma_i \leqslant \min\{\alpha_i, \beta_i\}$ for each i and w being a unit.

For the second part, let d' be a gcd of a' and b'. Then dd' divides both a and b, so must divide d. So $dd'k = d$ for some k, giving $d'k = 1$, and hence d' is a unit. Thus 1 is a gcd of a' and b' by Lemma 2.33 as $d' \sim 1$. □

2.2 Gauss' Lemma and UFDs

We now have enough machinery to work on Gauss' lemma and UFDs. For the rest of this chapter, we let R be a UFD and K its field of fractions. Let $a \in K$

be a non-zero element of the field, and let p be a prime element in R. We can then write

$$a = p^r b$$

where $b \in K$, r is an integer and p does not divide the numerator or denominator of b when written in lowest terms. For example, if $R = \mathbb{Z}$, so $K = \mathbb{Q}$, and $p = 3$, then we can write $\frac{27}{17} = 3^3 \cdot \frac{1}{17}$ and $\frac{14}{15} = 3^{-1} \cdot \frac{14}{5}$.

The rest of the proofs of this chapter are due to Theo Johnson-Freyd.

Lemma 2.56. *In the above equation, r is uniquely determined by a and p.*

Proof. That r exists follows from considering $a = s/t$, where $s, t \in R$. By unique factorisation, $s = p^{r_1} s'$ and $t = p^{r_2} t'$, where $p \nmid s'$ and $p \nmid t'$. Then $a = p^{r_1 - r_2}(s'/t')$. That r is unique is shown by assuming $a = p^r b = p^{r'} b'$ where $b = m/n$ and $b' = m'/n'$ and p divides none of m, n, m' or n'. Without loss of generality, assume $r \leqslant r'$. Multiplying by nn' and dividing by p^r gives $mn' = p^{r'-r} nm'$. Since p is prime, we must have $r' = r$, as required. □

Definition 2.57. In the above equation, r is called the **order of** a **at** p and we write $r = \operatorname{ord}_p a$. We also define $\operatorname{ord}_p 0 = \infty$, with the conventions that $n < \infty$ and $n + \infty = \infty + n = \infty$ for all $n \in \mathbb{Z}$, and $\infty + \infty = \infty$.

Lemma 2.58. *If $a, a' \in K$, then*

$$\operatorname{ord}_p(aa') = \operatorname{ord}_p a + \operatorname{ord}_p a'.$$

Proof. If one of a, a' is zero, then the equation is true by our conventions about infinity. So assume $a, a' \neq 0$ and let $r = \operatorname{ord}_p a$ and $r' = \operatorname{ord}_p a'$. Then $aa' = p^r(m/n).p^{r'}(m'/n') = p^{r+r'}(mm'/nn')$. Since p is prime and $p \nmid m, n, m', n'$, it follows that $p \nmid mm', nn'$. Hence, by Lemma 2.56, we have $\operatorname{ord}_p(aa') = r + r'$. □

Definition 2.59. Let $f \in K[x]$ be a polynomial over the field K, with

$$f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0.$$

As before, if $f = 0$, we define $\operatorname{ord}_p f = \infty$. Otherwise, define

$$\operatorname{ord}_p f = \min_{0 \leqslant i \leqslant n} \{\operatorname{ord}_p a_i\}.$$

Then if $r = \operatorname{ord}_p f$, we call up^r a **p-content** for f, where u is any unit of R. If $f \neq 0$, we define the **content** of f by

$$\operatorname{cont}(f) \sim \prod p^{\operatorname{ord}_p f},$$

the product being taken over all primes p such that $\operatorname{ord}_p f \neq 0$. The content of f is defined up to association, that is, up to multiplication by a unit of R. By abuse of notation, we often write $\operatorname{cont}(f) = \prod p^{\operatorname{ord}_p f}$ instead, using $=$ instead of $\sim$.

Lemma 2.60. *If $b \in K$ with $b \neq 0$ and $f \neq 0$, then* $\operatorname{cont}(bf) = b \operatorname{cont}(f)$. *Hence we can write*

$$f = c \cdot f_1$$

where $c = \operatorname{cont}(f)$ *and* $\operatorname{cont}(f_1) = 1$.

Proof. By Lemma 2.58, $\operatorname{ord}_p(ba_i) = \operatorname{ord}_p b + \operatorname{ord}_p a_i$, where $f = a_n x^n + \cdots + a_0$. Then

$$\begin{aligned}
\operatorname{ord}_p(bf) &= \min_i\{\operatorname{ord}_p(ba_i)\} \\
&= \min_i\{\operatorname{ord}_p b + \operatorname{ord}_p a_i\} \\
&= \operatorname{ord}_p b + \min_i\{\operatorname{ord}_p a_i\} \\
&= \operatorname{ord}_p b + \operatorname{ord}_p f.
\end{aligned}$$

Thus

$$\begin{aligned}
\operatorname{cont}(bf) &= \prod p^{\operatorname{ord}_p(bf)} \\
&= \prod p^{\operatorname{ord}_p b + \operatorname{ord}_p f} \\
&= \prod p^{\operatorname{ord}_p b} \cdot \prod p^{\operatorname{ord}_p f} \\
&\sim b \cdot \operatorname{cont}(f),
\end{aligned}$$

since unique factorisation implies that $b \sim \prod p^{\operatorname{ord}_p b}$. Hence $\operatorname{cont}(bf) = b \operatorname{cont}(f)$ up to association. For the second part, setting $f_1 = (1/\operatorname{cont}(f))f$ gives $\operatorname{cont}(f_1) = (1/\operatorname{cont}(f)) \operatorname{cont}(f) = 1$ as required. □

Definition 2.61. A polynomial with content 1 is called a **primitive polynomial**.

Lemma 2.62. *If f is a primitive polynomial, then all of the coefficients of f lie in R and their gcd is* 1.

Proof. Assume f has a coefficient $a_j \notin R$. Then we can write a_j in lowest terms as $a_j = m/n$, with $n \neq 1$. So there is some prime p with $p \mid n$ but $p \nmid m$. Then $\operatorname{ord}_p a_j < 0$, so $\operatorname{ord}_p f = \min_i\{\operatorname{ord}_p a_i\} < 0$. But as $\operatorname{cont}(f) = 1$, we see by unique factorisation that all of the $\operatorname{ord}_p f$ are zero. Hence all of the coefficients of f lie in R.

If the coefficients of f had a gcd which was not a unit, then we must have some prime p with $p \mid a_i$ for all i, so $\operatorname{ord}_p a_i \geqslant 1$ for all i, giving $\operatorname{ord}_p f = \min_i\{\operatorname{ord}_p a_i\} \geqslant 1$, so f cannot have content 1, contrary to assumption. Hence the gcd of the coefficients is 1.

Essentially, since $f \in R[x]$ with $f = \sum_{i=0}^n a_i x^i$, unique factorisation shows that $\operatorname{cont}(f) = \gcd(a_0, \ldots, a_n)$, since both equal $\prod p^{\min_i\{\operatorname{ord}_p a_i\}}$. □

Theorem 2.63 (Gauss' Lemma). *Let R be a UFD and K its field of fractions. Let $f, g \in K[x]$ be non-zero polynomials in one variable over K. Then*

$$\operatorname{cont}(fg) = \operatorname{cont}(f)\operatorname{cont}(g).$$

Proof. We will show first that the product of two primitive polynomials is primitive. Let f, g be primitive, so their coefficients lie in R and have gcd 1 (by the preceding lemma). Write $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{j=0}^{m} b_j x^j$, with $a_i, b_j \in R$ for all i, j, and with $\gcd(\{a_i\}) = \gcd(\{b_j\}) = 1$. Let p be a prime and let a_j be such that $p \nmid a_j$ but $p \mid a_i$ for all $i < j$, and similarly for b_k. Then the coefficient of x^{j+k} in fg is:

$$\sum_{i=0}^{j+k} a_i b_{j+k-i} = a_j b_k + \sum_{i=0}^{j-1} a_i b_{j+k-i} + \sum_{i=j+1}^{j+k} a_i b_{j+k-i}.$$

Now p divides each of the sums on the right, as $p \mid a_i$ for $i < j$ and $p \mid b_{j+k-i}$ for $i > j$. But $p \nmid a_j b_k$ as p is prime, so $p \nmid \sum_{i=0}^{j+k} a_i b_{j+k-i}$. Therefore, p does not divide all of the coefficients of fg, so $\operatorname{ord}_p(fg) = 0$. As p was arbitrary, it follows that $\operatorname{cont}(fg) = 1$.

Now let f, g be not necessarily primitive, so we can write $f = c_f f_1$ and $g = c_g g_1$ with f_1 and g_1 primitive, by Lemma 2.60. Then

$$\begin{aligned}
\operatorname{cont}(fg) &= \operatorname{cont}(c_f f_1 c_g g_1) \\
&= c_f c_g \operatorname{cont}(f_1 g_1) && \text{again by Lemma 2.60} \\
&= c_f c_g && \text{by the first part} \\
&= \operatorname{cont}(f)\operatorname{cont}(g),
\end{aligned}$$

and we are done. □

Corollary 2.64. *Let $f(x) \in R[x]$ have a factorisation $f(x) = g(x)h(x)$ in $K[x]$. If $c_g = \operatorname{cont}(g)$, $c_h = \operatorname{cont}(h)$ and $g(x) = c_g g_1(x)$, $h(x) = c_h h_1(x)$, then*

$$f(x) = c_g c_h g_1(x) h_1(x),$$

and $c_g c_h \in R$. In particular, if $f, g \in R[x]$ have content 1, then $h \in R[x]$ also.

Proof. We have only to show that $c_g c_h \in R$ and that if both f and g have content 1, then so does h. By the previous lemmas, $\operatorname{cont}(f) = \operatorname{cont}(g)\operatorname{cont}(h) = c_g c_h$. But $f \in R[x]$, so $\operatorname{cont}(f) \in R$. Now, if f and g both have content 1, then $1 = 1.c_h$, so $c_h = 1$ and $h \in R[x]$. □

Theorem 2.65. *Let R be a UFD. Then the polynomial ring $R[x]$ in one variable is also a UFD. Its units are the units of R, and its prime elements are either primes of R or polynomials in $R[x]$ which are irreducible in $K[x]$ and have content 1.*

Proof. We show first that a primitive non-constant polynomial in $R[x]$ is irreducible in $R[x]$ if and only if it is irreducible in $K[x]$. ('Non-constant' means its degree is greater than zero.) Let $f \in R[x]$ be such a polynomial, and assume that it is irreducible in $K[x]$ but factorises as $f = gh$ in $R[x]$. But then this is a non-trivial factorisation of f in $K[x]$ unless one of g and h is a unit in $K[x]$, say g is a unit. Then $g \in K$ (easy exercise), so $g \in R$. But then $\operatorname{cont}(f) = \operatorname{cont}(g)\operatorname{cont}(h)$, so $1 = g\operatorname{cont}(h)$ and g must be a unit in R. Thus f is irreducible in $R[x]$. Conversely, assume that it is irreducible in $R[x]$ but that it factorises as $f = gh$ in $K[x]$. By the previous corollary, we can then write $f = (c_g c_h)g_1 h_1$, which gives a factorisation of f in $R[x]$. Thus we must have either g_1 or h_1 being a unit in $R[x]$, so g or h must be a unit in $K[x]$ as well, and so the factorisation is trivial. Thus f is irreducible in $K[x]$.

Now given a polynomial $f \in R[x]$, we factorise it into irreducibles in $K[x]$. Recall also that $K[x]$ is a UFD by Example 2.21 and Theorem 2.30, so that this factorisation into irreducibles is unique (in the sense of Definition 2.14). After replacing each irreducible polynomial p in the factorisation with $c_p p_1$, where p_1 is primitive, we see from the previous corollary and induction that $\prod c_p \in R$, so that we now have a factorisation of f in $R[x]$ of the form

$$f = \operatorname{cont}(f).\prod(\text{irreducibles in } K[x] \text{ of content } 1).$$

Since R is a UFD, we can write $\operatorname{cont}(f) \in R$ as a product of irreducibles, and the primitive irreducibles in $K[x]$ are irreducible in $R[x]$ by the above. So every element of $R[x]$ has a factorisation into irreducibles.

Also, this factorisation is unique. If a non-constant polynomial $g \in R[x]$ has content other than 1, it can be written as a product $g = c_g g_1$, and neither c_g nor g_1 is a unit. So the irreducibles of $R[x]$ must either be constants or primitive polynomials. Then consider another factorisation of f into irreducibles in $R[x]$; again we can write it as

$$f = \prod(\text{elements of } R).\prod(\text{primitive non-constant polynomials in } R[x]).$$

But this is then a factorisation in $K[x]$ into a unit (the product of elements of R) and irreducibles by the above. So by uniqueness of factorisation in $K[x]$, the second product must be the same as in the original factorisation, and by the uniqueness of factorisation in R, so must the first product.

Thus factorisation into irreducibles in unique in $R[x]$, so $R[x]$ is a UFD as required.

Finally, the units of $R[x]$ are clearly constant polynomials (elements of R), for if a polynomial f has degree $\deg f > 0$, then $\deg(fg) = \deg f + \deg g > 0 = \deg 1$ for all non-zero $g \in R[x]$, so that f cannot be a unit. Then the units must be precisely the units of R by a similar argument. By the above, the primes of $R[x]$ are seen to be the primes of R and those primitive polynomials of $K[x]$ which are irreducible in $K[x]$ (recalling that every prime is irreducible and vice versa in a UFD). □

Corollary 2.66. *Let R be a UFD. Then the polynomial ring $R[x_1, \ldots, x_n]$ in n variables is also a UFD. Its units are the units of R and its prime elements are either primes of R or polynomials which are irreducible in $K[x_1, \ldots, x_n]$ and have content* 1.

Proof. This result follows from the previous theorem and induction, since $R[x_1, \ldots, x_n]$ and $R[x_1, \ldots, x_{n-1}][x_n]$ are easily seen to be isomorphic. □

Chapter 3

Homomorphisms and ideals

We do not make any assumptions on our rings in this chapter except for commutativity unless otherwise stated.

Definition 3.1. Let R_1 and R_2 be rings. Then a **homomorphism** from R_1 to R_2 is a function

$$\phi : R_1 \to R_2$$

satisfying the conditions:

$$\begin{aligned} \phi(a+b) &= \phi(a) + \phi(b) \qquad \text{for all } a, b \in R \\ \phi(ab) &= \phi(a)\phi(b) \qquad \text{for all } a, b \in R \\ \phi(1) &= 1. \end{aligned}$$

Example 3.2. Let $n \geqslant 2$ be an integer, and for $a \in \mathbb{Z}$, write $[a]_n$ for the equivalence class of a modulo n (that is, $[a]_n = \{r \in \mathbb{Z} : a \equiv r \pmod{n}\}$). We can regard $\mathbb{Z}/(n)$ as the ring whose elements are $[0]_n$, $[1]_n$, ..., $[n-1]_n$. Then the map $f : \mathbb{Z} \to \mathbb{Z}/(n)$ defined by $f(k) = [k]_n$ is a homomorphism.

Proof. Let $a, b \in \mathbb{Z}$. Then $f(a+b) = [a+b]_n = [a]_n + [b]_n = f(a) + f(b)$, and similarly $f(ab) = [ab]_n = [a]_n[b]_n = f(a)f(b)$. Finally, $f(1) = [1]_n$, which is the multiplicative identity in $\mathbb{Z}/(n)$. □

Lemma 3.3. *Let $\phi : R_1 \to R_2$ be a homomorphism. Then the following hold:*

(a) $\phi(0) = 0$.

(b) $\phi(-a) = -\phi(a)$.

(c) $\phi(a-b) = \phi(a) - \phi(b)$.

(d) *If $a \in R_1$ is invertible, then $\phi(a)$ is also invertible and $\phi(a)^{-1} = \phi(a^{-1})$.*

Proof (*Mark Lipson*) (a) We have $\phi(0) = \phi(0+0) = \phi(0) + \phi(0)$, so subtracting $\phi(0)$ from both sides gives $\phi(0) = 0$.

(b) We have $0 = \phi(0) = \phi(a + (-a)) = \phi(a) + \phi(-a)$. By the definition of additive inverse, this gives $\phi(-a) = -\phi(a)$.

(c) We have $\phi(a-b) = \phi(a+(-b)) = \phi(a) + \phi(-b) = \phi(a) - \phi(b)$ by the above and the definition of subtraction.

(d) Since a is invertible, we have $1 = \phi(1) = \phi(aa^{-1}) = \phi(a)\phi(a^{-1})$. Thus $\phi(a)$ is invertible with inverse $\phi(a)^{-1} = \phi(a^{-1})$. □

Lemma 3.4. *There is a unique homomorphism $\mathbb{Z} \to R$ for any ring R.*

Proof. We show uniqueness first. Let $\phi : \mathbb{Z} \to R$ be a homomorphism. Then $\phi(1) = 1$ by the third homomorphism condition. Let $n \in \mathbb{Z}$ with $n > 0$. Then $n = 1+\cdots+1$ with n ones, so $\phi(n) = \phi(1+\cdots+1) = \phi(1)+\cdots+\phi(1) = 1+\cdots+1$, so $\phi(n)$ is determined uniquely. Furthermore, if $n < 0$, we have $-n > 0$ and $\phi(n) = \phi(-(-n)) = -\phi(-n)$ by the previous lemma, which is uniquely determined as $-n > 0$. Finally, $\phi(0) = 0$ by the previous lemma, so ϕ is uniquely determined.

For the existence of such a homomorphism, we see that the map defined in this way ($\phi(0) = 0$, $\phi(n) = 1 + \cdots + 1$ for $n > 0$ and $\phi(n) = -\phi(-n)$ for $n < 0$) is actually a homomorphism, for with some slightly tedious work, it is not hard to show using standard properties of the integers that $\phi(m + n) = \phi(m) + \phi(n)$ and $\phi(mn) = \phi(m)\phi(n)$. □

Exercise 3.5. When does there exist a homomorphism from $\mathbb{Z}/(m)$ to $\mathbb{Z}/(n)$?

Answer. If $\phi : \mathbb{Z}/(m) \to \mathbb{Z}/(n)$ is a homomorphism, then $\phi([1]_m) = [1]_n$, and as $[1]_m + \cdots + [1]_m = [0]_m$, where we have summed m ones, we must have $\phi([1]_m + \cdots + [1]_m) = \phi([0]_m)$, or $[1]_n + \cdots + [1]_n = [0]_n$. For this, we require $n \mid m$. And if this is the case, then there exists a unique homomorphism defined in this way, which is again not hard to show, but a little tedius. □

Definition 3.6. Consider the unique homomorphism $\phi : \mathbb{Z} \to K$ where K is a field. Either there is a unique smallest positive $p > 0$ with $\phi(p) = 0$ or no such p exists. In the former case, we say that the field has **characteristic** p, otherwise we say that the field has **characteristic** 0. The characteristic of K is denoted by $\operatorname{char} K$. We will show later that the characteristic must be zero or a prime; see Proposition 3.31.

Exercise 3.7. What are the characteristics of the fields $\mathbb{Q}$ and $\mathbb{Z}/(p)$, where p is a positive prime?

Answer. 0 and p respectively.

Lemma 3.8. *If $\phi : R_1 \to R_2$ and $\psi : R_2 \to R_3$ are homomorphisms, then $\psi\phi : R_1 \to R_3$ is also a homomorphism.*

Proof. Let $a, b \in R_1$. Then

$$\begin{aligned}
\psi\phi(a+b) &= \psi(\phi(a+b)) && \text{by definition of } \psi\phi \\
&= \psi(\phi(a)+\phi(b)) && \text{as } \phi \text{ is a homomorphism} \\
&= \psi(\phi(a))+\psi(\phi(b)) && \text{as } \psi \text{ is a homomorphism} \\
&= \psi\phi(a)+\psi\phi(b)
\end{aligned}$$

as we required. The proof that $\psi\phi(ab) = \psi\phi(a)\psi\phi(b)$ is entirely similar. Finally, $\psi\phi(1) = \psi(1) = 1$. Thus $\psi\phi$ is a homomorphism. □

Definition 3.9. Let $\phi : R_1 \to R_2$ be a homomorphism. If ϕ is injective, then it is called a **monomorphism**; if it is surjective, it is called an **epimorphism**, and if it is bijective, it is called an **isomorphism**. Two rings R_1 and R_2 are said to be **isomorphic**, denoted by $R_1 \cong R_2$, if there exists an isomorphism $\phi : R_1 \to R_2$.

If two rings are isomorphic, they are essentially the same, up to a possible renaming of the elements.

Proposition 3.10. *Isomorphism is an equivalence relation on any set of rings.*

Proof. If R is a ring, then the identity map $\mathrm{id}_R : R \to R$ is clearly an isomorphism, so R is isomorphic to itself. Transitivity is also easy: if $R_1 \cong R_2$ and $R_2 \cong R_3$, then there exist isomorphisms $\phi : R_1 \to R_2$ and $\psi : R_2 \to R_3$. Then $\psi\phi : R_1 \to R_3$ is a homomorphism by the above lemma, it is injective as both ϕ and ψ are, and is likewise surjective. Thus $\psi\phi$ is an isomorphism.

Symmetry is a little more difficult to show. Let $\phi : R_1 \to R_2$ be an isomorphism. As ϕ is bijective, we can consider its inverse $\psi = \phi^{-1}$. We certainly know that ψ is bijective, but we must show that it is a homomorphism. Let $a', b' \in R_2$. As ϕ is surjective, we have $a' = \phi(a)$ and $b' = \phi(b)$ for some $a, b \in R_1$. Then we have $\psi(a'+b') = \psi(\phi(a)+\phi(b)) = \psi(\phi(a+b)) = a+b = \psi(a')+\psi(b')$ as required, using the facts that ϕ is a homomorphism and that $\psi\phi$ is the identity map. Similarly, we have $\psi(a'b') = \psi(a')\psi(b')$, and $\psi(1') = \psi(\phi(1)) = 1$, where here we write $1'$ for the multiplicative identity in R_2. Thus ψ is a homomorphism and so also an isomorphism, and so isomorphism is a symmetric relation.

Thus isomorphism is an equivalence relation. □

There is a simple test to see whether a homomorphism is injective.

Proposition 3.11. *Let $\phi : R_1 \to R_2$ be a homomorphism. Then ϕ is a monomorphism if and only if $\phi(a) = 0$ implies $a = 0$.*

Proof. If ϕ is a monomorphism, then $\phi(a) = \phi(0) = 0$ implies that $a = 0$. Conversely, assume that $\phi(a) = 0$ implies $a = 0$. Then if $\phi(b) = \phi(c)$, we have $\phi(b-c) = \phi(b) - \phi(c) = 0$, so that $b - c = 0$ and hence $b = c$. Thus ϕ is a monomorphism. □

Lemma 3.12. *Let $\phi : R_1 \to R_2$ and $\psi : R_2 \to R_3$ be homomorphisms. If $\psi\phi$ is a monomorphism, then ϕ is a monomorphism. If $\psi\phi$ is an epimorphism, then ψ is an epimorphism. We cannot say anything about the other homomorphism in either case.*

Proof (*Ilya Grigoriev*) Assume $\psi\phi$ to be injective and let $x_0, x_1 \in R_1$ with $\phi(x_0) = \phi(x_1)$. Then $\psi\phi(x_0) = \psi\phi(x_1)$. Hence $x_0 = x_1$ because $\psi\phi$ is injective. So we have proved that ϕ is a monomorphism.

Now assume $\psi\phi$ is surjective and let $z \in R_3$. Then there is some $x \in R_1$ with $\psi\phi(x) = z$ by surjectivity, so $\psi(\phi(x)) = z$. Thus we have proved that ψ is an epimorphism.

We now give examples to show that ψ is not necessarily injective if $\psi\phi$ is, and ϕ is not necessarily surjective if $\psi\phi$ is. The corresponding result for sets is easier, as there is no extra structure to deal with, so we will do that first, and then give an example with rings.

Consider the sets and maps $\{1\} \xrightarrow{\phi} \{1, 2\} \xrightarrow{\psi} \{1\}$, where $\phi(1) = 1$ and $\psi(1) = \psi(2) = 1$. Then clearly $\psi\phi$ is injective, yet ϕ is injective and ψ is not. Also, $\psi\phi$ is surjective, yet ψ is surjective and ϕ is not.

The case with rings is essentially the same, but we have to be careful that the maps are indeed homomorphisms. Here is an example which parallels the above. Consider $R \xrightarrow{\phi} R \times R \xrightarrow{\psi} R$, where R is any non-zero ring and $R \times R = \{(x, y) : x, y \in R\}$ is the *direct product* (or Cartesian product) of two copies of R, with the ring operations defined componentwise (you should check this is a ring!). The function ϕ is just the standard diagonal embedding $\phi(x) = (x, x)$, and ψ is projection onto the first component, so $\psi((x, y)) = x$. It is clear, therefore, that $\psi\phi : R \to R$ is the identity map, and yet ϕ is not surjective and ψ is not injective. □

Definition 3.13. Let $\phi : R_1 \to R_2$ be a homomorphism. Then the image of R_1, $\phi(R_1) = \{\phi(a) : a \in R_1\}$, is called the **image** of ϕ and is denoted by $\operatorname{Im}\phi$.

Proposition 3.14. *If $\phi : R_1 \to R_2$ is a homomorphism, then $\operatorname{Im}\phi$ is a subring of R_2.*

Proof. We only need to check that $0, 1 \in \operatorname{Im}\phi$, that $\operatorname{Im}\phi$ is closed under addition and multiplication, and that $\operatorname{Im}\phi$ has additive inverses. We have $\phi(0) = 0$ and $\phi(1) = 1$, so that $0, 1 \in \operatorname{Im}\phi$. Now let $a', b' \in \operatorname{Im}\phi$, and let $a, b \in R_1$ be such that $\phi(a) = a'$ and $\phi(b) = b'$. For closure, we note that $a' + b' = \phi(a) + \phi(b) = \phi(a + b) \in \operatorname{Im}\phi$, and similarly $a'b' = \phi(a)\phi(b) = \phi(ab) \in \operatorname{Im}\phi$. Finally, we have additive inverses in $\operatorname{Im}\phi$, as $-a' = -\phi(a) = \phi(-a) \in \operatorname{Im}\phi$, so we are done. □

Definition 3.15. Let $\phi : R_1 \to R_2$ be a homomorphism. Then the inverse image of 0, $\phi^{-1}(0) = \{a \in R_1 : \phi(a) = 0\}$, is called the **kernel** of ϕ and is denoted by $\ker\phi$.

So we can restate Proposition 3.11 as: a homomorphism $\phi : R_1 \to R_2$ is injective if and only if $\ker \phi = 0$.

Proposition 3.16. *The kernel of a homomorphism is an ideal.*

Proof (*Taktin Oey*) Let ϕ be the homomorphism. Using Lemma 1.24, we must show that $\ker \phi$ is non-empty, is closed under addition and satisfies the ideal property. Firstly, $\phi(0) = 0$, so $0 \in \ker \phi$ and $\ker \phi$ is non-empty. Now let $x, y \in \ker \phi$. Then $\phi(x + y) = \phi(x) + \phi(y) = 0 + 0 = 0$, so $x + y \in \ker \phi$ and $\ker \phi$ is closed under addition. Finally, if $x \in \ker \phi$ and $a \in R$, then $\phi(ax) = \phi(a)\phi(x) = 0$, so $ax \in \ker \phi$ and the ideal condition is satisfied. □

Definition 3.17. A **coset** of an ideal I in a ring R is a set of the form

$$a + I = \{a + x : x \in I\},$$

where $a \in R$.

Lemma 3.18. *Let $\phi : R_1 \to R_2$ be a homomorphism. If $b \in \phi(R_1)$, then $\phi^{-1}(b)$ is a coset of* $\ker \phi$.

Proof. As $b \in \phi(R_1)$, there is some $a \in R_1$ with $b = \phi(a)$. If $\phi(a') = b$, then $\phi(a' - a) = \phi(a') - \phi(a) = b - b = 0$, so $a' - a \in \ker \phi$. Conversely, if $a' - a \in \ker \phi$, then $\phi(a' - a) = 0$, so $\phi(a') - \phi(a) = 0$ and $\phi(a') = \phi(a) = b$. Thus $a' - a \in \ker \phi$ if and only if $\phi(a') = b$. So $\phi^{-1}(b) = \{a' \in R_1 : a' - a \in \ker \phi\} = a + \ker \phi$, a coset of $\ker \phi$. □

Exercise 3.19. If $x_1, x_2, \ldots, x_n \in \ker \phi$, then $(x_1, x_2, \ldots, x_n) \subseteq \ker \phi$.

Proof. By Proposition 3.16 above, we know that $\ker \phi$ is an ideal. The definition of the ideal generated by a set (Definition 1.25) states that $(x_1, x_2, \ldots, x_n)$ is the smallest ideal containing the elements x_1, x_2, …, x_n. But we observe that the ideal $\ker \phi$ contains these elements, so we must have $(x_1, x_2, \ldots, x_n) \subseteq \ker \phi$ as claimed. □

Lemma 3.20. *Let I be an ideal of R. The relation on R defined by*

$$a \equiv b \pmod{I} \text{ if and only if } a - b \in I$$

is an equivalence relation. (*This is called* equivalence modulo I *or* equivalence mod I.) *The equivalence classes are precisely the cosets of I in R.*

Proof. We must show first that this relation is indeed an equivalence relation. Let $a, b, c \in R$. We have $a - a = 0 \in I$ by the definition of an ideal (Definition 1.23), so $a \equiv a \pmod{I}$ and the relation is reflexive. For symmetry, if $a \equiv b \pmod{I}$, then $a - b \in I$, so $b - a = -(a - b) \in I$, again by the properties of an ideal, so $b \equiv a \pmod{I}$. Finally, to show that the relation is transitive, assume that

$a \equiv b \pmod{I}$ and $b \equiv c \pmod{I}$. Then $a - b \in I$ and $b - c \in I$. As I is closed under addition, we have $(a - b) + (b - c) = a - c \in I$, so $a \equiv c \pmod{I}$.

The equivalence class of $a \in R$ is given by

$$\begin{aligned}[a] &= \{b \in R : a \equiv b \pmod{I}\} \\ &= \{b \in R : b \equiv a \pmod{I}\} \\ &= \{b \in R : b - a \in I\} \\ &= a + I,\end{aligned}$$

which is a coset of I in R, as required. This also shows that $a + I = [a]$ for any $a \in R$, so that every coset of I in R is an equivalence class. □

Exercise 3.21. Show that in $\mathbb{Z}$, the relation $a \equiv b \pmod{(m)}$ is the same as the traditional relation $a \equiv b \pmod{m}$.

Proof. We have $a \equiv b \pmod{(m)}$ if and only if $a - b \in (m)$. But by Lemma 1.26, $(m) = \{rm : r \in \mathbb{Z}\}$, so $c \in (m)$ if and only if $c = rm$ for some $r \in \mathbb{Z}$, that is $c \in (m)$ if and only if $m \mid c$. Thus $a \equiv b \pmod{(m)}$ if and only if $m \mid a - b$, which is the traditional definition of $a \equiv b \pmod{m}$. □

Definition 3.22. Let I be an ideal of R. Then the quotient of R by the relation defined in Lemma 3.20 is denoted by R/I. (See Definition 2.51.) We define a ring structure on R/I as follows. If $a + I, b + I \in R/I$, then

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

R/I is called the **quotient ring** of R by I, and read as 'R mod I'.

Exercise 3.23. Show that the definition makes sense, that is, if different representatives of the cosets $a + I$ and $b + I$ are chosen, the sum and product are still $(a + b) + I$ and $ab + I$ respectively. Show also that the ring axioms are satisfied for R/I.

Answer. Let $a' \in a + I$ and $b' \in b + I$, so that $a' + I = a + I$ and $b' + I = b + I$. (It is easy to show that $a' + I = a + I$ if and only if $a' \in a + I$, for the cosets of I form a partition of R, so we can use the results of the previous chapter.) We must show that $(a' + b') + I = (a + b) + I$ and that $a'b' + I = ab + I$. We have $a' = a + x$ and $b' = b + y$ for some $x, y \in I$. Thus $a' + b' = (a + b) + (x + y)$, so $a' + b' \in (a + b) + I$ as $x + y \in I$. Similarly, $a'b' = (a + x)(b + y) = ab + (ay + xb + xy)$, and the parenthesised term is in I by the ideal property and the closure of I under addition. Thus $a'b' \in ab + I$. It follows that $(a' + b') + I = (a + b) + I$ and $a'b' + I = ab + I$ as required, so the operations are well-defined.

For the ring axioms, we only sketch the proof. For (A0), we have $(a + I) + (b + I) = (a + b) + I$ which is in R/I, so R/I is closed under addition. Multiplicative closure, (M0), follows similarly. Associativity and commutativity of addition

and multiplication hold in R/I because they hold in R. For (A2), $0 + I$ is the additive identity, as $(a + I) + (0 + I) = (a + 0) + I = a + I$, and likewise, $1 + I$ is the multiplicative identity. Similarly, $(-a) + I$ is the additive inverse of $a + I$. The distributive law is also easy to verify. Thus R/I is a ring.

Exercise 3.24. Explain the notation for the integers modulo m which we have been using, namely $\mathbb{Z}/(n)$.

Answer. By the answer to exercise 3.21 above, we see that the equivalence classes of integers modulo m are the cosets of (m) in $\mathbb{Z}$. The operations of addition and multiplication modulo m are defined by the corresponding operations on coset representatives, and so we are actually working in the ring $\mathbb{Z}/(m)$, the quotient of $\mathbb{Z}$ by the ideal (m).

Proposition 3.25. *If $f : R \to R'$ is a homomorphism of rings and I' is an ideal of R', then $f^{-1}(I')$ is an ideal of R. However, if I is an ideal of R then $f(I)$ is* not *necessarily an ideal of R'.*

Proof. Let $I = f^{-1}(I')$. We must show that I is an ideal of R. By Lemma 1.24, we need only show that I is non-empty, closed under addition and satisfies the ideal property. We have $0 \in I'$, and as $f(0) = 0$, we have $0 \in I$. If $x, y \in I$, then $f(x + y) = f(x) + f(y) \in I'$ as $f(x), f(y) \in I'$, so $x + y \in I$. Finally, if $a \in R$ and $x \in I$, then $f(ax) = f(a)f(x) \in I'$ as $f(x) \in I'$ and I' satisfies the ideal property. Thus $ax \in I$ and I is an ideal as required.

For the converse, consider the example $f : \mathbb{Z} \to \mathbb{Q}$. The ideal (1) in $\mathbb{Z}$ maps under f to the image of $\mathbb{Z}$ in $\mathbb{Q}$. But this is not an ideal in the field $\mathbb{Q}$, as it does not satisfy the ideal property; for example, $\frac{1}{2} \in \mathbb{Q}$ and $1 \in \mathbb{Z}$, but $\frac{1}{2} \cdot 1 \notin \mathbb{Z}$. □

Definition 3.26. Let I be an ideal of R. The homomorphism $\phi_I : R \to R/I$ defined by $\phi_I(a) = a + I$ is called the *canonical homomorphism.*

Exercise 3.27. What is the kernel of this homomorphism?

Answer. The kernel is I, for if $a \in I$, then $\phi_I(a) = a + I = I$, and if $\phi_I(a) = a + I = I$, then we must have $a \in I$.

Theorem 3.28 (First Isomorphism Theorem). *Let $\phi : R \to R'$ be a homomorphism of rings with kernel $I = \ker \phi$. Then $\bar{\phi} : R/I \to \phi(R)$ defined by $\bar{\phi}(a + I) = \phi(a)$ is an isomorphism. Thus $\phi = \bar{\phi}\phi_I$, so that ϕ factors through the canonical homomorphism.*

An alternative way of writing this isomorphism is $R/\ker \phi \cong \operatorname{Im} \phi$.

Proof (*Michael Kolodrubetz*) We must first show that $\bar{\phi}$ is a well-defined map. Assume $a+I=b+I$; we need to show that $\phi(a)=\phi(b)$. Now if $a+I=b+I$, then $a=a+0\in b+I$, so $a=b+x$ for some $x\in I$. Then $\phi(a)=\phi(b+x)=\phi(b)+\phi(x)=\phi(b)$ as $\phi(x)=0$. Thus $\bar{\phi}$ is well-defined.

Next, we show that $\bar{\phi}$ is, indeed, a homomorphism, which we do by showing that the three defining conditions are satisfied. Firstly, let $a,b\in R$. Then $\bar{\phi}(a+I)+\bar{\phi}(b+I)=\phi(a)+\phi(b)=\phi(a+b)=\bar{\phi}((a+b)+I)=\bar{\phi}((a+I)+(b+I))$ as required. Similarly, we have $\bar{\phi}(a+I)\bar{\phi}(b+I)=\phi(a)\phi(b)=\phi(ab)=\bar{\phi}(ab+I)=\bar{\phi}((a+I)(b+I))$, and so the second condition is satisfied. Finally, noting that $1+I$ is the unity element of R/I, we have $\bar{\phi}(1+I)=\phi(1)=1$ as required.

Now we show that $\bar{\phi}$ is a bijection. First we prove that $\bar{\phi}$ is injective. To do this, we use Proposition 3.11 and show that if $\bar{\phi}(a+I)=0$, then $a+I=0+I$. But if $\bar{\phi}(a+I)=\phi(a)=0$, then $a\in\ker\phi=I$, so $a+I=0+I$. Thus $\bar{\phi}$ is injective.

To prove surjectivity, we must show that for any element $b\in\phi(R)$, there exists an element of R/I that will be mapped to it by $\bar{\phi}$. But $b=\phi(a)$ for some $a\in R$ by the definition of $\phi(R)$. Then $\bar{\phi}(a+I)=b$ as required, so $\bar{\phi}$ is surjective.

Thus $\bar{\phi}:R/I\to\phi(R)$ is an isomorphism, and we are done. □

An alternative way of doing the above can be achieved by noting that there are actually two things to be proven, firstly that the map is a well-defined bijection and secondly that it is a homomorphism. If we split them, we discover that the first part is actually a purely set-theoretic result, as follows. If $f:X\to Y$ is a map of sets, we consider the equivalence relation defined by $x_1\sim x_2$ if $f(x_1)=f(x_2)$. Then we can show that the map $\bar{f}:X/\!\sim\to f(X)$ defined by $\bar{f}(\bar{x})=f(x)$ is a well-defined bijection, where $\bar{x}$ is the equivalence class of x. If we now return to a homomorphism of rings, we have $\phi:R\to R'$, and the equivalence classes under the relation just described are precisely the cosets $a+I$ where $I=\ker\phi$, and hence we will have shown that $\bar{\phi}:R/I\to\phi(R)$ is a bijection. This argument will also work when considering the First Isomorphism Theorem for groups, modules and other more exotic examples.

Exercise 3.29. Show that $k[x,y]/(y)\cong k[x]$, where k is a field.

Answer. The most enlightening proof uses the First Isomorphism Theorem, although it can be proven directly. Consider the evaluation map $\phi_0:k[x,y]\to k[x]$ given by mapping $f(x,y)\in k[x,y]$ to $f(x,0)\in k[x]$, that is, we 'set y to be 0'. It is not hard to check that ϕ_0 is a homomorphism (all such evaluation maps are). The image of ϕ_0 is clearly the whole of $k[x]$, as any polynomial in $k[x]$ can be thought of as a polynomial in $k[x,y]$. The kernel of ϕ_0 is the set of polynomials in $k[x,y]$ which map to 0 when y is set to zero. Thinking of $k[x,y]$ as $k[x][y]$, we can write a polynomial $f(x,y)$ as $f(x,y)=f_0(x)+f_1(x)y+f_2(x)y^2+\cdots+f_n(x)y^n$, we have $f(x,0)=f_0(x)$, so $f(x,0)=0$ if and only if $f_0(x)=0$. Thus $\ker\phi_0$ is the set of all polynomials of the form $f_1(x)y+f_2(x)y^2+\cdots+f_n(x)y^n=(f_1(x)+f_2(x)y+\cdots+f_n(x)y^{n-1})y$. But this is just the ideal (y). So the First

Isomorphism Theorem gives $k[x,y]/\ker\phi_0 \cong \operatorname{Im}\phi_0$, that is $k[x,y]/(y) \cong k[x]$ as required.

Exercise 3.30. Describe the ring $\mathbb{R}[x]/(x^2+1)$.

Answer. Note first that x^2+1 is irreducible in $\mathbb{R}[x]$. We claim that $\mathbb{R}[x]/(x^2+1)$ is the field obtained by *adjoining* a root of x^2+1 to $\mathbb{R}$. Of course, the roots of x^2+1 are $\pm i$, so we should get a field isomorphic to $\mathbb{R}[i] = \mathbb{C}$. Again, the simplest way to show this is to use the First Isomorphism Theorem. Consider the evaluation map homomorphism $\phi : \mathbb{R}[x] \to \mathbb{C}$ defined by $\phi(f) = f(i)$. It is surjective as $a+bi$ is the image of $a+bx$ for any $a, b \in \mathbb{R}$. The kernel of the map is all polynomials satisfying $f(i) = 0$. Certainly $x^2+1 \in \ker\phi$, as $i^2+1 = 0$, so $(x^2+1) \subseteq \ker\phi$. Assume that there is some other polynomial $f \in \ker\phi$ which is not in (x^2+1). Then by subtracting an appropriate multiple of x^2+1 from f, we will get a polynomial of degree either 1 or 0 in $\ker\phi$. But $ai+b \neq 0$ if $a \neq 0$, and certainly $b \neq 0$ maps to a non-zero element of $\mathbb{C}$. So this is impossible, and $\ker\phi = (x^2+1)$. Thus by the First Isomorphism Theorem, $\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}$.

Proposition 3.31. *Let K be a field. Then* $\operatorname{char} K = 0$ *or* $\operatorname{char} K$ *is a prime.*

Proof. Consider the unique homomorphism $\phi : \mathbb{Z} \to K$ of Definition 3.6. Then by the First Isomorphism Theorem, $\operatorname{Im}\phi \cong \mathbb{Z}/\ker\phi$. Now the ideals of $\mathbb{Z}$ are 0 and (n) for some $n > 0$, by the result of Exercise 1.28. If $\ker\phi = 0$, then $\phi(m) \neq 0$ for all $m > 0$, so $\operatorname{char} K = 0$. Otherwise, if $\ker\phi = (n)$, then $\phi(m) = 0$ for $m = kn$, $k \in \mathbb{Z}$, so $\operatorname{char} K = n$. Note that $\phi(1) = 1 \neq 0$, so $n > 1$. From the above, we see that $\operatorname{Im}\phi \cong \mathbb{Z}/(n)$. But $\operatorname{Im}\phi$ is a subring of K by Proposition 3.14, so it must be an integral domain as K is an integral domain (Lemma 1.19). Now we know from Corollary 1.37 that for $n > 1$, the ring $\mathbb{Z}/(n)$ is an integral domain precisely when n is prime, so we must have n prime in our case, and $\operatorname{char} K$ is prime. □

We can observe more from this proof. In the case that $\operatorname{char} K = p$ is prime, the image of the homomorphism, $\operatorname{Im}\phi \cong \mathbb{Z}/(p)$, is actually a *field* by the same corollary, and is called the **prime subfield** of K. In the case of a field of characteristic 0, $\operatorname{Im}\phi \cong \mathbb{Z}$, and this is called the **prime subring** of K. In this case, we can extend the mapping $\phi : \mathbb{Z} \to K$ to a mapping $\bar{\phi} : \mathbb{Q} \to K$ by defining $\bar{\phi}(m/n) = \phi(m)/\phi(n)$. (Check that this is well-defined!) Then $\operatorname{Im}\bar{\phi}$ is called the prime subfield of K. Thus in either case, the prime subfield is the smallest subfield of K.

Proposition 3.32. *Let I, J be ideals of R with $I \subseteq J$. Let J/I be the quotient of J by the relation defined in Lemma 3.20. Then J/I is an ideal of R/I. Conversely, if J' is any ideal of R/I, then $J' = J/I$ for some ideal $J \subseteq R$ containing I. Thus there is a one-to-one correspondence between ideals of R/I and ideals of R containing I.*

Proof. We first show that J/I is an ideal of R/I. Before we do this, we note that if $a \in J$, then $a + I \subseteq J$. (This is because $I \subseteq J$, and so $a + x \in J$ for every $x \in I$ as J is an ideal.) Similarly, if $a + I \in J/I$, then clearly $a \in J$. Next, as $I \subseteq J$, we see that $I \in J/I$, so J/I is non-empty. Now, if $a + I$ and $b + I$ both lie in J/I, then as $a, b \in J$, we have $a + b \in J$, and so $(a + b) + I \in J/I$, showing that J/I is closed under addition. Finally, for the ideal property, if $a + I \in J/I$ and $r + I \in R/I$, then $a \in J$ and so $ra \in J$ by the ideal property for J. Thus $(r + I)(a + I) = ra + I \in J/I$ and J/I satisfies the ideal property. Hence J/I is an ideal of R/I.

For the converse, let J' be an ideal of R/I. Then J' is a set of cosets of I in R; let J be the union of these cosets. Clearly $J' = J/I$; we need only show that J is an ideal of R. As $I \in J'$, we have $I \subseteq J$, so J is non-empty. Also, J is closed under addition: if $a, b \in J$, then $a + I, b + I \in J'$, so $(a + b) + I \in J'$ and thus $a + b \in J$. Finally, J satisfies the ideal property, for if $a \in J$ and $r \in R$, then $a + I \in J'$, so $(r + I)(a + I) = ra + I \in J'$ as J' is an ideal, so $ra \in J$. Thus J is an ideal as required.

Finally, it is clear that given an ideal $J \supseteq I$ of R, the ideal we constructed in the second part from the ideal J/I of R/I is precisely J, and similarly the other way around, so that we do indeed have a bijective correspondence between the ideals of R containing I and the ideals of R/I. □

Definition 3.33. An ideal $I \subseteq R$ is called **prime** if $I \neq (1)$ and for any $a, b \in R$, $ab \in I$ implies that either $a \in I$ or $b \in I$.

The two uses of the term 'prime' which we have seen are closely related as the following proposition shows.

Proposition 3.34. *Let R be an integral domain and let p be a non-zero element of R. Then p is a prime element of R if and only if (p) is a prime ideal of R.*

Proof. Recall that $a \in (p)$ if and only if $p \mid a$. Let $a, b \in R$. Then if p is prime, we see that $ab \in (p)$ implies $p \mid ab$, so either $p \mid a$ or $p \mid b$, that is, either $a \in (p)$ or $b \in (p)$, so (p) is a prime ideal of R. Conversely, if (p) is a prime ideal of R, then if $p \mid ab$, we have $ab \in (p)$, so either $a \in (p)$ or $b \in (p)$, and therefore either $p \mid a$ or $p \mid b$. □

Corollary 3.35. *In a PID, any non-zero prime ideal is generated by a prime element.*

Proof. As every non-zero ideal is principal, let this prime ideal be (a). But (a) is prime if and only if a is prime by the previous result, so a must be prime. □

Here is a direct proof of the same result.

Proof (*Christopher Lee*) Let I be a non-zero prime ideal in the PID R, so $I = (p)$ for some element $p \in R$.

Assume that p is not prime. Then by the definition of a prime element (Definition 2.16), we know that there are $a, b \in R$ with $p \mid ab$ but $p \nmid a$ and $p \nmid b$. Because $p \mid ab$, we have $ab = pc$ for some $c \in R$, so $ab \in (p) = I$. Similarly, we can deduce from $p \nmid a$ and $p \nmid b$ that $a, b \notin (p) = I$. But by the definition of a prime ideal, as $ab \in I$, we must have $a \in I$ or $b \in I$. We have arrived at a contradiction, so our original assumption was incorrect, and the generating element p must be prime. □

Proposition 3.36. *Let $f : R \to R'$ be a homomorphism of rings, and let I be a prime ideal of R'. If I is prime, then so is $f^{-1}(I)$.*

Proof. Write $J = f^{-1}(I)$ and let $a, b \in R$. We must show that $ab \in J$ implies that either $a \in J$ or $b \in J$. Now $ab \in J$ implies that $f(ab) = f(a)f(b) \in I$, so either $f(a) \in I$ or $f(b) \in I$ by primality of I. It follows that either $a \in J$ or $b \in J$ respectively, so J is prime. □

Exercise 3.37. Is the converse of the last proposition true?

Answer. No. For example, consider the homomorphism $\delta : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$, where the latter ring is the set of all ordered pairs of integers, with addition and multiplication defined componentwise, and $\delta(n) = (n, n)$. Now consider the ideal $I = 0$ in $\mathbb{Z} \times \mathbb{Z}$. This is not prime, as $(1, 0)(0, 1) = (0, 0) \in I$, but $\delta^{-1}(I) = 0$ is a prime ideal in $\mathbb{Z}$, as $\mathbb{Z}$ is an integral domain.

Proposition 3.38. *If P is an ideal of a ring R, then P is prime if and only if R/P is an integral domain.*

Proof. Assume P is prime and that $a+P, b+P \in R/P$ with $(a+P)(b+P) = P$. (Recall that P is the zero element of R/P.) Then $ab + P = P$, so $ab \in P$. As P is prime, it follows that $a \in P$ or $b \in P$, so $a + P = P$ or $b + P = P$, that is R/P is an integral domain.

Conversely, assume that R/P is an integral domain. Let $a, b \in R$ and assume that $ab \in P$. Then $ab + P = (a + P)(b + P) = P$, so either $a + P = P$ or $b + P = P$, that is, either $a \in P$ or $b \in P$. Thus P is a prime ideal. □

Definition 3.39. An ideal $I \subseteq R$ is called **maximal** if $I \neq (1)$ and there is no ideal $J \subset R$ such that $I \subset J \subset (1)$ (strict inclusions). In other words, the only ideal 'bigger' than I is (1).

Exercise 3.40. Can a ring have more than one maximal ideal?

Answer. Yes, for example, every non-zero prime ideal in $\mathbb{Z}$ is maximal.

Here is a useful result which will help us to identify maximal ideals.

Proposition 3.41. (*Pradeep Mouli*) *A ring R is a field if and only if the only ideals of R are 0 and (1).*

Proof. Let k be a field and let I be an ideal of k. If $I = 0$, we are done, otherwise there is some non-zero $x \in I$. Since every non-zero element of a field is a unit, it follows from Lemma 1.35 that $I = (1)$.

Conversely, let R be a ring whose only ideals are 0 and (1). Then for any non-zero $x \in R$, we have $(x) = (1)$, so there exists $y \in R$ with $xy = 1$. Thus every non-zero element is invertible, so R is a field. □

Proposition 3.42. *If M is an ideal of a ring R, then M is maximal if and only if R/M is a field.*

Proof. Assume M is maximal. The ideals of R/M are in one-to-one correspondence with the ideals of R containing M by Proposition 3.32 above. But the only ideals containing M are M and (1), so the only ideals of R/M are $\{M\}$ (which is the zero ideal) and the image of (1), which is R/M, so that R/M is a field.

Conversely, if R/M is a field, then the only ideals of R/M are the zero ideal $\{M\}$ and R/M, so the only ideals of R containing M are M and R, so M is maximal. □

Proposition 3.43. *A maximal ideal is prime.*

Proof. This follows from the previous proposition and Proposition 3.38, for if M is a maximal ideal in R, then R/M is a field, and hence is an integral domain, so M is prime. □

Here is one very nice application of prime ideals. This is not quite the most general form, but it is very powerful nonetheless.

Proposition 3.44 (Eisenstein's criterion). *Let R be an integral domain and let K be its field of fractions. Let $f = a_n x^n + \cdots + a_0 \in R[x]$ and let M be a maximal ideal of R. If $a_n \notin M$, $a_i \in M$ for $0 \leqslant i < n$ and $a_0 \notin M^2$, then f is irreducible in $K[x]$.*

Here $M^2 = MM$ is the ideal generated by all products xy where $x, y \in M$. Also, in the case $R = \mathbb{Z}$, where every non-zero prime ideal is maximal, this proposition becomes: if p is prime with $p \nmid a_n$, $p \mid a_i$ for $0 \leqslant i < n$ and $p^2 \nmid a_0$, then f is irreducible in $\mathbb{Q}[x]$. So, for example, we see that $x^n + 5x + 5$ is irreducible for all $n \geqslant 2$. In the case $R = \mathbb{Z}$, this can be proved directly without recourse to prime ideals and the machinery we have been building up, but this approach is potentially more enlightening.

Note that f may not be irreducible in $R[x]$, for example, $2(x^n + 5x + 5)$ is reducible in $\mathbb{Z}[x]$ but irreducible in $\mathbb{Q}[x]$.

Proof. Assume that $f = gh$ with $g, h \in K[x]$, where $g = b_m x^m + \cdots + b_0$ and $h = c_{n-m} x^{n-m} + \cdots + c_0$. We may assume, using Corollary 2.64, that $g, h \in R[x]$. Note that as $a_n = b_m c_{n-m}$, we have $b_m, c_{n-m} \notin M$ as M is prime. We reduce the polynomial modulo M, that is, we take the image of the polynomials in

$(R/M)[x]$ under the reduction map $R[x] \to (R/M)[x]$ given by $rx^s \mapsto (r+M)x^s$. It is not hard to check that this map is a homomorphism. Writing a bar to indicate the image under this map, we thus have $\bar{f} = \bar{g}\bar{h}$. But $\bar{f} = \bar{a}_n x^n$ as $a_i \in M$ for $0 \leqslant i < n$. Furthermore, R/M is a field by Proposition 3.42, so $(R/M)[x]$ is a UFD (Corollary 2.31). As x is irreducible in $(R/M)[x]$ and $\bar{a}_n$ is a unit, we have $\bar{f} = \bar{a}_n x \cdots x$ as the unique factorisation of $\bar{f}$ into irreducibles. Thus we must have $\bar{g} = \bar{b}_m x^m$ and $\bar{h} = \bar{c}_{n-m} x^{n-m}$. Now if $0 < m < n$, we must have $b_0 \in M$ and $c_0 \in M$, so $a_0 = b_0 c_0 \in M^2$, contrary to hypothesis. Thus either $m = 0$ or $m = n$, so either g or h is a unit in $K[x]$, proving that f is irreducible in $K[x]$. □

Here is one nice application of this result.

Proposition 3.45. *Let $p \in \mathbb{Z}$ be a positive prime. Then $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1 = (x^p - 1)/(x - 1)$ is irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.*

Proof. If this polynomial is irreducible in $\mathbb{Q}[x]$, then it is irreducible in $\mathbb{Z}[x]$ as it has content 1, by Theorem 2.65. We would like to apply Eisenstein's criterion, but it is not immediately clear how to do so. There is a trick, however: consider $\Phi_p(x+1)$. Clearly, this is irreducible if and only if $\Phi_p(x)$ is. But using the binomial theorem, we have

$$\begin{aligned}\Phi_p(x+1) &= \frac{(x+1)^p - 1}{(x+1) - 1} \\ &= \tfrac{1}{x}\left(x^p + \tbinom{p}{p-1}x^{p-1} + \cdots + \tbinom{p}{1}x + 1 - 1\right) \\ &= x^{p-1} + \tbinom{p}{p-1}x^{p-2} + \cdots + \tbinom{p}{1}\end{aligned}$$

This now satisfies Eisenstein's criterion: the leading coefficient is 1, every other coefficient is a multiple of p (as when p is prime, $p \mid \binom{p}{r}$ for $0 < r < p$), and the constant coefficient is $\binom{p}{1} = p$, which is not a multiple of p^2. Thus $\Phi_p(x+1)$ is irreducible in $\mathbb{Q}[x]$ as required. □

3.1 Orderings and Zorn's Lemma

Definition 3.46. Recall the definitions of partially ordered sets (posets) and totally ordered sets from the previous chapter (Definition 2.43). Let S be a poset, and write $a \leqslant b$ for the partial order on S. We now introduce some additional terminology regarding ordered sets.

- A **greatest element** of a partially ordered set S is an element a such that $b \leqslant a$ for all $b \in S$. **Least elements** are defined analogously.
- A **maximal element** of a partially ordered set S is an element $a \in S$ such that whenever $a \leqslant b$ with $b \in S$, we have $a = b$. (Alternatively, a maximal element is an element $a \in S$ for which there is no $b \in S$ with $a < b$.) **Minimal elements** are defined analogously.

- Let T be a subset of a partially ordered set S. An **upper bound** of T in S is an element $b \in S$ such that $x \leqslant b$ for all $x \in T$.
- A subset $T \subseteq S$ is said to be a **totally ordered subset**, or a **chain**, if every pair of elements of T are comparible in the partial order. (That is, for every $x, y \in T$, either $x \leqslant y$ or $y \leqslant x$ or both.)
- A poset S is **inductively ordered** if every non-empty totally ordered subset of S has an upper bound.

Proposition 3.47. *Not every partially ordered set S has a maximal element, and even if such an element exists, it is not necessarily unique. Not every partially ordered set S has a greatest element, but if such an element exists, it is unique, and is also the unique maximal element of S.*

Proof. An example of a poset with no maximal element is given by $\mathbb{N}$, although this does have a minimal element. $\mathbb{Z}$ has neither. In a poset in which no two distinct elements are comparable, every element is maximal, so maximal elements are not necessarily unique.

Clearly a greatest element is maximal, so not every poset has a greatest element. But assume that S has a greatest element x and a maximal element y, then by definition, we have $y \leqslant x$ as x is a greatest element, and we cannot have $y < x$ as y is maximal. So we must have $x = y$. Thus x is the unique maximal element, and as every greatest element is maximal, x is also the unique greatest element. □

Exercise 3.48. Let R be a commutative ring and M a maximal ideal of R. What partially ordered set is M a maximal element of?

Answer. The set of ideals of R excluding (1), ordered by inclusion.

Axiom 3.49 (Zorn's Lemma). Let S be a non-empty inductively ordered set. Then S has a maximal element.

(Some authors remove the requirement 'non-empty' from both the definition of 'inductively ordered' and the statement of Zorn's Lemma. The result is the same, as then the empty set, which is totally ordered, must have an upper bound, so S must be non-empty.)

Zorn's Lemma is usually treated as an axiom of set theory. It is, in fact, equivalent to the Axiom of Choice, which states that for any set S, there exists a *choice function*, which is a mapping $f : \mathcal{P}(S) \setminus \varnothing \to S$ with $f(X) \in X$ for every $\varnothing \neq X \subseteq S$, where $\mathcal{P}(S)$ is the power set of S, which is the set of all subsets of S. (Why is the word 'choice' used to describe this function?) Some people quibble with this axiom, as it results in a number of apparent paradoxes, such as the Banach-Tarski paradox. However, most working mathematicians do use it, since it is crucial for some fundamental results, such as the following.

Theorem 3.50. *Every ideal of a ring R is contained in a maximal ideal.*

Proof. Let I be an ideal of R, and consider the set S of ideals of R containing I, ordered by inclusion and excluding (1). This is inductively ordered, for given a non-empty totally ordered subset $T \subseteq S$, we claim that $J_T = \bigcup_{J \in T} J$ is an upper bound for T. Firstly, it is clear that $J \subseteq J_T$ for every $J \in T$, and it also follows that $I \subseteq J_T$ as T is non-empty. So we need only show that it is an ideal. The proof follows almost identically to that of Lemma 2.26. We have $I \subseteq J_T$, so J_T is non-empty. If $x, y \in J_T$, then $x \in J$ and $y \in J'$ for some $J, J' \in T$. But as T is totally ordered, either $J \subseteq J'$ or $J' \subseteq J$. Thus either $x, y \in J'$ or $x, y \in J$. Either way, it follows that $x + y \in J_T$. Finally, if $a \in R$ and $x \in J_T$, then $x \in J$ for some $J \in T$, so $ax \in J$ as J is an ideal, hence $ax \in J_T$. It follows that J_T is an ideal containing I which is an upper bound for T. Thus every non-empty totally ordered subset of S has an upper bound, so S is inductively ordered. Then by Zorn's Lemma, S has a maximal element, which is a maximal ideal containing I. (Any greater ideal would also contain I, and hence either be (1) or have to be in S, which is impossible.) □

Corollary 3.51. *Every ring has a maximal ideal.*

Proof. Apply the previous theorem to the ideal 0. □

Another example of where Zorn's Lemma is used is in the proof that every vector space has a basis: here one takes sets of independent vectors ordered by inclusion; a maximal independent set is a basis.

We end on a lighter note ....

Bad Joke 3.52. What is yellow and equivalent to the axiom of choice? *Zorn's Lemon.*

Bibliography

Here are a few books for further study of this and related subjects. There are many, many books available; these are just some of the authors' favourites. They are listed in roughly increasing order of difficulty.

- *?., Fraleigh, A first course in abstract algebra.* This covers elementary group theory, up to things like the Jordan–Hölder theorem and the structure of finitely generated abelian groups, and elementary ring theory up to the results we have presented, and on to Galois theory. There are many elementary exercises, providing a wealth of opportunity to gain familiarity with the material.
- *B. Hartley and T. O. Hawkes, Rings, Modules and Linear Algebra.* An introduction to ring theory, focusing on the results of our chapter 2, and going on to discuss canonical forms for matrices using the language of module theory.
- *P. J. Cameron, Introduction to algebra.* This covers elementary ring theory, group theory, module theory, Galois theory and more, at a reasonable pace.
- *I. N. Stewart and D. O. Tall, Algebraic Number Theory and Fermat's Last Theorem.* A nice introduction to algebraic number theory, which is one of the major (and beautiful) applications of the ring theory we have been studying.
- *M. Atiyah and I. G. Macdonald, Introduction to commutative algebra.* The advanced undergraduate or beginning graduate level introduction to ring theory and module theory. A classic in the field, with superb (and highly challenging) exercises.

Index

In this index, pages where the terms are defined are indicated in bold face.

addition, **3**
adjoining elements, **28**
almost all ..., **33**
anti-symmetric, **14**
ascending chain condition, **11**
associated, **9**
associates, **9**
associative law, **3**
Axiom of Choice, 24

binary operation, 3
binary relation, **13**
Boolean ring, **28**

$\mathbb{C}$, 71
cancellation law
    addition, 4
    multiplication, 32
canonical homomorphism, **21**
Cartesian product, **13**, 66
chain, *see* totally ordered subset
characteristic 0, **19**, 22
characteristic p, **19**, 22
closure law, **3**
commutative law, **3**, **5**
commutative ring, **5**
content, **17**
coset, **20**

direct product, 66
distributive law, **3**
    subtraction, 4
divides, **9**
division algorithm, 9
division ring, **4**
divisor, **9**

ED, *see* Euclidean domain
Eisenstein's criterion, 23
epimorphism, **20**
equivalence class, **15**
equivalence modulo I, **21**
equivalence relation, 10, **14**
    and partitions, 15
Euclidean algorithm, **12**
Euclidean domain, **11**
    is a PID, 11
Euclidean function, **11**
evaluation map, 70

factor, **9**
factor groups, 16
factorisation
    unique, 9
field, **5**, 10
    characteristic, **19**, 22
    ideals, 73
    is a Euclidean domain, 11
field of fractions, **16**
First Isomorphism Theorem, **21**
fraction
    lowest terms, 16

Gauss' Lemma, **17**
Gaussian integers, 28
    norm, 11
    units of, 39
gcd, *see* greatest common divisor
generated ideal, **6**
generating set, **6**
greatest common divisor, **12**, 13, 50
greatest element, **23**
group of units, **7**
    examples, 39

homomorphism, **19**
    image, *see* image
    kernel, *see* kernel

ideal, **6**

ideal (1), 37
ideal property, **6**
ideal test, 6
identity element
    uniqueness, 31
image, **20**, **66**
inductively ordered, **24**
integers modulo n, *see* $\mathbb{Z}/(n)$
integral domain, **5**, 9
inverse
    uniqueness
        additive, 4
        multiplicative, 31
inverse law, **3, 4**
invertible, **7**
irreducible, **10**
    and prime elements, 11
    Eisenstein's criterion, 23
isomorphic, **20**
isomorphism, **20**

kernel, **20**

least element, **23**
lowest terms, *see* fraction

maximal, **22**
maximal element, **23**
maximal ideal, 24
minimal element, **23**
modular arithmetic, 16
modulo I, *see* equivalence modulo I
mod I, *see* equivalence modulo I
monoid, **32**
    free, **33**
monoid ring, **33**
monomorphism, **20**
multiplication, **3**

norm, 43

order, *see* total order
order at a prime, **17**
ordered set, *see* totally ordered set

p-content, **17**
partial order, **14**
partially ordered set, *see* poset
partition, **14**
    and equivalence relations, 15
PID, *see* principal ideal domain
polynomial, **34**
    degree, **37**, **46**
    primitive, **17**
polynomial ring, **5**, **34**
    and unique factorisation, 18
    primes, 18
    units, 18
poset, **14**, 23
prime
    element, **10**
        and irreducibles, 11
    ideal, **22**
prime subfield, **71**
prime subring, **71**
primitive polynomial, **17**
principal ideal, **6**
Principal ideal domain
    is a UFD, 11–12
principal ideal domain, **6**, 22

quotient
    by a relation, **15**
quotient ring, **21**
quotient rings, 16

$\mathbb{R}[x]$
    is a Euclidean domain, 11
rationals, 15
reflexive, **14**
relation, *see* binary relation
ring
    axioms, **3**
    definition, **3**
    examples, 27

subring, **4**, **28**
subtraction, **4**
symmetric, **14**

topology, 16
total order, **14**
totally ordered set, **14**

totally ordered subset, **24**
transitive, **14**

UFD, *see* unique factorisation domain
unique factorisation
    failure, 10
unique factorisation domain, **10**
unit, **7**, 9
unity, 3
unity/identity law, **3**
upper bound, **24**

vector space, 24, 77

$\mathbb{Z}$
    is a Euclidean domain, 11
$\mathbb{Z}/(n)$, **4**
    elements, 19
    notation, 69
    structure, 7
    when a field, 7
$\mathbb{Z}[\alpha]$, 28
$\mathbb{Z}[i]$, *see* Gaussian integers
$\mathbb{Z}[x]$, 28
zero divisor, **5**
zero element
    uniqueness, 4
zero law, **3**
zero ring, **28**
Zorn's Lemma, 24