MT3505 Algebra: Rings and Fields

April 29, 2016

Contents

Chapter 1

Some historical background

A *ring* is a set, together with two binary operations + and ·, satisfying various natural axioms.

The prototype example is the set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ with the usual arithmetic. The fact that this example on its own yields the whole of number theory shows what a rich structure rings can have.

In fact, many of the familiar examples of systems where one can 'add' and 'multiply' give us rings. For example: the integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$, real valued functions and so on...

However, starting with the axioms and looking for examples of things that satisfy them is not the way rings first came into mathematics.

In about 1630, Fermat was reading a recently published translation of *Arithmetica* by Diophantus of Alexandria. He was making notes in the margin and at one point he entered:

> *To divide a cube into two other cubes, a fourth power or in general any power whatever into two powers of the same denomination above the second is impossible, and I have assuredly found an admirable proof of this, but the margin is too narrow to contain it.*

That is, if $n > 2$ then there are no integer solutions x, y and z of the equation $x^n + y^n = z^n$. This is the famous Fermat's Last Theorem which resisted all attempts to prove it until recently. Investigations of this result led to much interesting mathematics, including some of the first systematic investigations of ring theory.

Fermat was able to prove the case $n = 4$ (by something called the *method of descent*) but all other cases proved much harder.

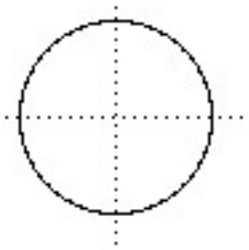
In his 1770 book *Algebra*, Euler published a proof of the $n = 3$ case. The following is a sketch of this proof. (We will return to this later in the semester!)

Assume that $x^3 + y^3 = z^3$, and that x and y are both odd and coprime. Then put $x = p + q$, $y = p - q$ and $z = 2r$ to get $(p+q)^3 + (p-q)^3 = 8r^3$, which simplifies to $p(p^2 + 3q^2) = 4r^3$. Since p and q are coprime, with a little work one can deduce that p is divisible by 4 and $p^2 + 3q^2$ is a perfect cube. Then Euler factorised $p^2 + 3q^2$ into $(p + q\sqrt{-3})(p - q\sqrt{-3})$ and observed that in the ring R (he didn't use that term!) of complex numbers of the form $\{a + b\sqrt{-3} \ : \ a, b \in \mathbb{Z}\}$ these two factors were coprime and so each factor is a perfect cube. That leads to a contradiction.

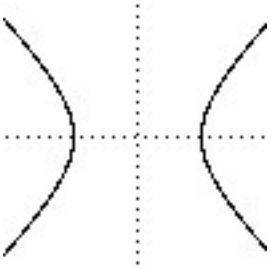
Unfortunately, Euler's proof relies on the fact that in the ring R we can factor elements into a unique product of primes, just as one can in $\mathbb{Z}$. But in R we can write $4 = 2 \times 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$, and these two factorisations are distinct. Hence this proof is incorrect, though nobody noticed the problem at the time.

It only became apparent much later, in 1847, when Lamé claimed he had proved Fermat's Last Theorem, by factoring $x^n + y^n$ as $(x - \eta)(x - \eta^2)...(x - \eta^{n-1})$, where η is an n^{th} complex root of 1. It was swiftly realised that the rings in which this factorisation was done do not have the property that each number factorises uniquely into primes, and it was left to Kummer to introduce the idea of an 'ideal number' to restore unique factorisation and allow Fermat's Theorem to be proved for some values of n. This invention of Kummer led to the development of the idea of an *ideal* of a ring, which we will meet later.

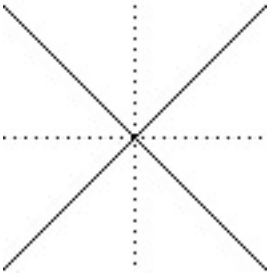
Another area of mathematics which was important in the development of ring theory is geometry. Many interesting curves and surfaces have equations which involve polynomials.



A circle$x^2 + y^2 - 1 = 0$



A hyperbola $x^2 - y^2 - 1 = 0$



Two lines $x + y = 0$

A single point $x^2 + y^2 = 0$

A paraboloid $x^2 + y^2 - z = 0$

A hyperboloid $x^2 + y^2 - z^2 - 1 = 0$

It turns out that geometric objects like this are associated with particular rings of polynomials and the algebra of these rings gives insight into the geometric properties. This area of mathematics is called *algebraic geometry*.

In the 19th century it was realised that the complex numbers parametrise the plane $\mathbb{R}^2$ in a very useful way, so people wanted to find a similar way to parametrise the space $\mathbb{R}^3$. The mathematician William Rowan Hamilton worked on this for a long time with no success. Each day at breakfast his daughter would ask:

Well, Papa can you multiply triplets?

but he had to admit that he could still only add and subtract them.

His breakthrough came in two stages. First he realised that one had to move from 3 to 4 dimensions and so (by analogy with the complex numbers) one had numbers of the form $a + ib + jc + kd$ for a, b, c, d and $i^2 = j^2 = k^2 = -1$. Then one can put $ij = k$, but to avoid a contradiction one has to make the multiplication non-commutative, with $ji = -k$.

> *And here there dawned on me the notion that we must admit, in some sense, a fourth dimension of space for the purpose of calculating with triples... An electric circuit seemed to close, and a spark flashed forth.*

This revelation came to Hamilton when he was walking with his wife by a canal in Dublin in 1843 and he was so taken with it that he stopped and carved two rules for multiplication on the Brougham bridge.

This system is called the *quaternions* or *hamiltonians*. These proved to be very important in the development of mechanics and other areas of applied mathematics in the 19th century. In fact, the quaternions contain what we now call the scalar and vector product of 3-dimensional vectors and it is these products which are now used.

This was the first example of a non-commutative ring. When Cayley and Sylvester developed the ideas of matrices later in the century they gave further examples of such structures.

Chapter 2

The ring of integers

Before beginning our study of abstract rings, we recall some well-known properties of the integers $\mathbb{Z}$.

If $a, b \in \mathbb{Z}$ satisfy $a = qb$ for some $q \in \mathbb{Z}$, then b *divides* a, and we write $b \mid a$. Note that $1 \mid a$, $-1 \mid a$, $a \mid 0$ for all $a \in \mathbb{Z}$, and $0 \mid a$ if and only if $a = 0$.

Example 0.1. The numbers

$$1, 2, 3, 5, 6, 10, 15, 30$$

divide 30 and

$$1, 2, 19, 38$$

divide 38.

1. The division and euclidean algorithms

Theorem 1.1 (The division algorithm) *If $a, b \in \mathbb{Z}$ with $b > 0$, then there exist a unique quotient $q \in \mathbb{Z}$ and a unique remainder $r \in \mathbb{Z}$ such that*

$$a = qb + r$$

where $0 \leq r < b$.

Proof. To prove that r exists, let $M = \{a - qb \ : \ q \in \mathbb{Z}\}$ and let r be the smallest number in M such that $r \geq 0$ (exercise: why does r exist?). Then $r = a - qb$ for some $q \in \mathbb{Z}$. If $r \geq b$, then $r > r - b \geq 0$ and $r - b = a - qb - b = a - (q + 1)b \in M$. This contradicts our assumption that r is the smallest number in M with $r \geq 0$.

To prove that r is unique, assume that

$$a = q_1 b + r_1 \ \text{ and } \ a = q_2 b + r_2.$$

We must prove that $r_1 = r_2$. Seeking a contradiction, assume that $r_1 < r_2$. It follows that $(q_1 - q_2)b = r_2 - r_1 > 0$. But then b divides $r_2 - r_1 \leq r_2 < b$, a contradiction. Since r is unique, $q = (a - r)/b$ is also unique. ■

Example Let $a = 13281$ and $b = 17$. Then $a = 781b+4$. So, the quotient is 781 and the remainder is 4.

In this course, the natural numbers $\mathbb{N}$ are taken to *include* zero.

Definition 1.2. Let $a, b \in \mathbb{Z}$. Then $d \in \mathbb{N}$ is the *greatest common divisor* of a and b if $d \mid a$ and $d \mid b$ and for all d' such that $d' \mid a$ and $d' \mid b$ we have $d' \leq d$. The greatest common divisor of a and b is denoted $\gcd(a, b)$.

The next lemma will help us find a method for determining the gcd of any two integers.

Lemma 1.3. *Let $a, b \in \mathbb{Z}$ with $a \neq 0$. Then*

(i) $\gcd(a, 0) = |a|$;

(ii) $\gcd(a, b) = \gcd(a - qb, b)$ *for all* $q \in \mathbb{Z}$.

Proof. **(i).** All integers divide 0 and the largest number dividing a is $|a|$. Hence $\gcd(a, 0) = |a|$.

(ii). Let $d_1 = \gcd(a, b)$ and $d_2 = \gcd(a - qb, b)$. Then $d_1 \mid a$ and $d_1 \mid b$. Hence d_1 divides $a - qb$ and so $d_1 \leq d_2$. Conversely, if $n \mid a - qb$ and $n \mid b$ for some $n \in \mathbb{N}$, then $n \mid a$. In particular, $d_2 \mid a$ and $d_2 \mid b$ and so $d_2 \leq d_1$. ■

Example 1.4. [The extended euclidean algorithm] Let $a = 76$ and $b = 32$. Then the divisors of a are

$$1, 2, 4, 19, 38, 76$$

and the divisors of b are

$$1, 2, 4, 8, 16, 32.$$

So, $\gcd(a, b) = 4$. The *extended euclidean algorithm* allows us to find x and y such that $xa + yb = \gcd(a, b) = 4$.

$$\begin{array}{rrl|l} & a = 76 & 32 = b & \\ & 2b = 64 & 24 = 2a - 4b & \\ \hline r_1 = & a - 2b = 12 & 8 = -2a + 5b & = r_2 \\ & -2a + 5b = 8 & 8 = 6a - 14b & \\ \hline r_3 = & 3a - 7b = 4 & 0 = -8a + 19b & \end{array}$$

The greatest common divisor of a and b is the last non-zero remainder, $r_3 = 4$. So $x = 3$ and $y = -7$.

Theorem 1.5. *Let $a, b \in \mathbb{Z}$ and $d = \gcd(a, b)$. Then there exist $x, y \in \mathbb{Z}$ such that $d = ax + by$.*

Proof. If $a, b \in \mathbb{N}$, then the extended euclidean algorithm can be used, as in Example 1.4, to find x and y. See Tutorial Sheet 1 for a proof that the extended euclidean algorithm is correct.

If $a \in \mathbb{Z} \setminus \mathbb{N}$, then $-a \in \mathbb{N}$. So there exist $x, y \in \mathbb{Z}$ such that $(-a)x + yb = d$. Hence $(-x)a + by = d$, as required. The case $b \in \mathbb{Z} \setminus \mathbb{N}$ is similar. ■

Definition 1.6. Let $a, b \in \mathbb{Z}$. Then a and b are called *coprime* if $\gcd(a, b) = 1$.

Lemma 1.7. *Let $a, b, c \in \mathbb{Z}$ such that a and b are coprime. Then*

(i) *if $a \mid bc$, then $a \mid c$;*

(ii) *if $a \mid c$ and $b \mid c$, then $ab \mid c$;*

(iii) *if a and c are coprime, then a and bc are coprime.*

Proof. See Tutorial Sheet 1. ■

2. Modular arithmetic

In this section we will recall the basic ideas of modular arithmetic. Let $m \in \mathbb{N}$, and let $\mathbb{Z}/(m)$ denote the set

$$\{n \in \mathbb{N} \ : \ 0 \leq n < m\} = \{0, 1, 2, \ldots, m - 1\}.$$

(The reason for choosing the notation $\mathbb{Z}/(m)$ will become apparent later in the course!)

For $x, y \in \mathbb{Z}$, we will write

$$x \equiv y \bmod m$$

if x and y leave the same remainder on division by m. Note that this is equivalent to the fact that $x - y$ is divisible by m.

If $n \in \mathbb{N}$ and $n \equiv r \bmod m$ with $0 \leq r < m$, then we may also say that r is n *reduced modulo* m. Sometimes we denote the modulo m reduction of a number by putting a bar over the number (if it is clear from the context modulo which number we reduce).

If $a, b \in \mathbb{Z}/(m)$, then $a + b, a \cdot b \in \mathbb{N}$ and so by the division algorithm (Theorem 1.1) we can write

$$a + b = q_1 m + r_1 \text{ and } a \cdot b = q_2 m + r_2$$

where r_1, r_2 are the unique remainders such that $0 \leq r_1, r_2 < m$, that is, $r_1, r_2 \in \mathbb{Z}/(m)$. This defines two operations *addition modulo* m and *multiplication modulo* m on $\mathbb{Z}/(m)$, namely

$$a + b \equiv r_1 \bmod m \text{ and } a \cdot b \equiv r_2 \bmod m.$$

Example 2.1.

$$\begin{aligned} 1+1 &\equiv 2 \bmod 7 \\ 4+6 &\equiv 3 \bmod 7 \\ 3+4 &\equiv 0 \bmod 7 \\ 2 \cdot 5 &\equiv 3 \bmod 7 \\ \overline{17} &= 3 \qquad \text{(if it is clear that we are reducing modulo 7).} \end{aligned}$$

Lemma 2.2. *Working modulo* m*,*

$$\overline{x \cdot y} = \overline{\overline{x} \cdot \overline{y}}$$

and

$$\overline{x + y} = \overline{\overline{x} + \overline{y}}$$

for all $x, y \in \mathbb{Z}$*.*

Proof. See Tutorial Sheet 1. ■

From this follows that addition and multiplication modulo m satisfy the following properties:

$$\overline{\overline{(x+y)} + z} = \overline{x + \overline{(y+z)}}$$

$$\overline{\overline{(xy)}z} = \overline{x\overline{(yz)}}$$

$$\overline{x + 0} = \overline{0 + x} = x$$

$$\overline{1 \cdot x} = \overline{x \cdot 1} = x$$

$$\overline{x + (m - x)} = \overline{(m - x) + x} = 0$$

$$\overline{x + y} = \overline{y + x}$$

$$\overline{xy} = \overline{yx}$$

$$\overline{x \cdot 0} = \overline{0 \cdot x} = 0$$

for all $x, y, z \in \mathbb{Z}/(m)$.

Example 2.3. The following is the table of multiplication for $\mathbb{Z}/(7)$ modulo 7.

$\cdot$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

3. The Chinese Remainder Theorem

Example 3.1. Consider the following game. Alice thinks of any natural number x less than 30 and calculates the remainder of x modulo 2, modulo 3, and modulo 5. Alice then tells Bob the three remainders. For example, Alice might tell Bob that $x \equiv 0 \bmod 2$, $x \equiv 0 \bmod 3$, and $x \equiv 3 \bmod 5$. Bob wins if he can tell Alice what the secret number x is, and Alice wins if Bob cannot.

Is there a way for Bob to always win this game? We'll see in this section that the answer is yes!

Lemma 3.2. *Let $q_1, q_2, \ldots, q_t \in \mathbb{N} \setminus \{0\}$. Assume that $q_1, q_2, \ldots, q_t$ are pairwise coprime, let $N = q_1 q_2 \cdots q_t$, and define*

$$\rho : \mathbb{Z}/(N) \longrightarrow \mathbb{Z}/(q_1) \times \mathbb{Z}/(q_2) \times \cdots \times \mathbb{Z}/(q_t)$$

by

$$\rho(x) = (x \bmod q_1, x \bmod q_2, \ldots, x \bmod q_t).$$

Then ρ is a bijection.

Proof. Let $x, y \in \mathbb{Z}/(N)$ such that $\rho(x) = \rho(y)$. Then $x \equiv y \bmod q_i$ for all i. It follows that $x - y \equiv 0 \mod q_i$ for all i and so $q_i \mid x - y$ for all i. Now, $q_1, q_2, \ldots, q_t$ are pairwise coprime and so, by repeatedly applying Lemma 1.7(ii) and (iii), we obtain $N = q_1 \cdots q_t \mid x - y$. But $x, y \in \mathbb{Z}/(N)$ and so $-N < x - y < N$. Hence $x - y = 0$ and so $x = y$, and ρ is injective.

Moreover, $|\mathbb{Z}/(N)| = N = |\mathbb{Z}/(q_1) \times \mathbb{Z}/(q_2) \times \cdots \times \mathbb{Z}/(q_t)|$, so ρ is surjective. ■

So, in Example 3.1, we know that the secret number x is in 1–1 correspondence with the triple $(0, 0, 3)$. To recover x from $(0, 0, 3)$ we have to apply ρ^{-1}.

Theorem 3.3 (Chinese Remainder Theorem) *Let $q_1, q_2, \ldots, q_t \in \mathbb{N}\setminus\{0\}$, such that $q_1, q_2, \ldots, q_t$ are pairwise coprime. Let $N = q_1 q_2 \cdots q_t$, and let $a_1, a_2, \ldots, a_t \in \mathbb{Z}$ be arbitrary. Then the system of equations*

$$\begin{aligned} x &\equiv a_1 \bmod q_1 \\ x &\equiv a_2 \bmod q_2 \\ &\vdots \\ x &\equiv a_t \bmod q_t \end{aligned}$$

has a solution $x \in \mathbb{Z}$. Moreover, $y \in \mathbb{Z}$ is also a solution if and only if $x \equiv y \bmod N$.

Proof. We start by showing how to construct such an x.

By repeatedly applying Lemma 1.7(iii) we deduce that q_i and N/q_i are coprime for all i. Hence by the extended euclidean algorithm (Theorem 1.5), for all i there exist integers x_i and y_i such that

$$x_i q_i + y_i (N/q_i) = 1$$

Set $b_i = y_i(N/q_i)$ for all i, and notice that $q_j \mid (N/q_i)$ so

$$b_i \equiv 0 \bmod q_j$$

for all $i \neq j$. Also,

$$b_i = y_i(N/q_i) = 1 - x_i q_i \equiv 1 \bmod q_i,$$

So, $a_i b_i \equiv a_i \bmod q_i$, and hence

$$x = a_1 b_1 + a_2 b_2 + \cdots + a_t b_t \in \mathbb{Z}$$

is a solution to the equations in the theorem.

An integer y is a solution if and only if x and y leave the same remainder on division by each of the q_i. Let ρ be as in Lemma 3.2. Then, reducing modulo N, we see that $\rho(\overline{x}) = \rho(\overline{y})$. The map ρ is a bijection on $\mathbb{Z}/(N)$, so $x \equiv y \bmod N$. ■

The proof of Theorem 3.3 can be used to find the solutions to specific examples of systems of equations like those given in Theorem 3.3.

Example 3.1 ctd. In the notation of Theorem 3.3, Alice's secret number x satisfies the equations

$$\begin{aligned} x &\equiv 0 \bmod 2 \\ x &\equiv 0 \bmod 3 \\ x &\equiv 3 \bmod 5. \end{aligned}$$

So, $N = 30$, $a_1 = a_2 = 0$, $a_3 = 5$, $q_1 = 2$, $q_2 = 3$ and $q_3 = 5$. Our solution is of the form $a_1b_1 + a_2b_2 + a_3b_3$, so (since $a_1 = a_2 = 0$) it suffices to apply the extended euclidean algorithm to the pair $(5, 6)$, to obtain:

$N/q_3 = 6$	$5 = q_3$
$q_3 = 5$	
$(N/q_3) - q_3 = 1$	

So, $y_3 = 1$, and $b_3 = 1 \times N/q_3 = 6$. Thus the solution we are looking for is

$$x = 0b_1 + 0b_2 + a_3b_3 = 3 \cdot 6 = 18.$$

4. Prime Numbers

A *prime number* is a natural number $p > 1$ that is only divisible by itself and 1. The first few primes are

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

They can be found as follows using a method called the *sieve of Eratosthenes*. Write out all the natural numbers bigger than 1

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25$$

and then cross out the numbers that are multiples of 2 but not 2 itself

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, 9, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, 15, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, 21, \cancel{22}, 23, \cancel{24}, 25.$$

The numbers not crossed out are not multiples of 2. Continue by crossing out the multiples of 3

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, \cancel{15}, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, \cancel{21}, \cancel{22}, 23, \cancel{24}, 25.$$

then 5, then 7, then 11 and so on

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, \cancel{15}, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, \cancel{21}, \cancel{22}, 23, \cancel{24}, \cancel{25}.$$

Repeating this procedure over and over again you can eventually tell if any number is prime or not. The largest known prime (as of January 2016!) is

$$2^{74207281} - 1$$

and it has roughly 22 million digits!

Theorem 4.1. *There are infinitely many prime numbers.*

Proof. Seeking a contradiction, assume the contrary. Then the primes can be listed as

$$p_1, p_2, \ldots, p_m$$

for some natural number $m \geq 1$. Now, the natural number $n = p_1p_2 \cdots p_m + 1$ is a product of primes and hence divisible by some p_i for $1 \leq i \leq m$. But then $p_i \mid n$ and $p_i \mid p_1p_2 \cdots p_m$ and so $p_i \mid (n - p_1p_2 \cdots p_m) = 1$, a contradiction as 1 is not a prime. ■

A crucial property of the primes is given in the following lemma. We will revisit this lemma again in later sections.

Lemma 4.2. *If $p \in \mathbb{N}$ is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof. If $p \mid a$, then we are finished, so assume that $p \nmid a$. Then $\gcd(p, a) = 1$, and hence $p \mid b$ by Lemma 1.7. ■

Theorem 4.3 (The fundamental theorem of arithmetic) *Every non-zero natural number n is a product of prime numbers, and this product is unique up to the order of the factors.*

Proof. We will first use induction to prove that n is a product of primes. The number 1 is the empty product of primes. Assume that for all $m < n$ we know that m is a product of primes. If n is a prime number, then it is a product of primes (of length 1). If n is not a prime, then there exist $a, b \in \mathbb{N}$ such that $a, b > 1$ and $n = ab$. Then by induction a and b are products of primes and hence so is n.

Now for the uniqueness claim. By way of contradiction, assume that there exist natural numbers that have two distinct factorisations into products of primes. Let n be the smallest such number, with distinct factorisations

$$n = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_r.$$

It is immediate that $n > 1$, and hence $m \geq 1$ and $r \geq 1$. Now, $p_1 \mid n$, so by repeatedly applying Lemma 4.2 we find that $p_1 \mid q_j$ for some j. But q_j is a prime, so $p_1 = q_j$. Hence $n/p_1 = n/q_j$ is a smaller natural number with two distinct prime factorisations, a contradiction. ■

Chapter 3

Introduction to rings

1. Definitions and first examples

The example of the integers in the previous chapter serves as a prototype for the more abstract study we start in this chapter.

Definition 1.1. A *ring* is a set R together with a pair of binary operations $+$ and $*$ satisfying the axioms:

A1. $(R, +)$ is a group, with identity element 0;

A2. $r + s = s + r$ for all $r, s \in R$;

M1. $r * (s * t) = (r * s) * t$ for all $r, s, t \in R$;

D. $(r + s) * t = r * t + s * t$ and $r * (s + t) = r * s + r * t$ for all $r, s, t \in R$.

The element 0 from Axiom **A1** is called the *zero* of the ring R.

A1, A2, M1, D are called the *ring axioms*. The final Axiom **D** is called the *distributive law*. Note that we will write $r - s$ to mean $r + (-s)$. The use of 0 in **A1** is symbolic: it does not always mean the integer 0.

[Aside: $(R, +, *)$ is a ring if $(R, +)$ is an abelian group, $(R, *)$ is a semigroup, and R satisfies the distributive law.]

Example 1.2. The sets $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ with the usual addition and multiplication are all rings.

Definition 1.3. The *order* of a ring R is the number of elements it contains, that is, $|R|$.

In the previous example, all the rings had infinite order.

Example 1.4. For all $n \in \mathbb{N} \setminus \{0\}$, the integers $\mathbb{Z}/(n) = \{0, 1, \ldots, n-1\}$ with addition and multiplication modulo n satisfy the ring axioms. Hence $\mathbb{Z}/(n)$ is a ring. The order of $\mathbb{Z}/(n)$ is n.

Example 1.5. Let $M_2(\mathbb{R})$ denote the set of all 2×2 real matrices. Then $M_2(\mathbb{R})$ forms a ring under the usual matrix addition $+$ and multiplication $*$. For instance,

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 2 & 4 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 5 & 8 \end{pmatrix}.$$

In fact, the set of $n \times n$ matrices with entries in any ring R forms a ring, denoted $M_n(R)$ (see Tutorial Sheet 2).

Example 1.6. The *quaternions* are the set $\mathbb{H} = \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \ : \ a, b, c, d \in \mathbb{R}\}$ with addition

$$(a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}) + (a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) = (a_1 + a_2) + (b_1 + b_2)\mathbf{i} + (c_1 + c_2)\mathbf{j} + (d_1 + d_2)\mathbf{k}$$

and multiplication

$$\begin{aligned}(a_1+b_1\mathbf{i}+c_1\mathbf{j}+d_1\mathbf{k})*(a_2+b_2\mathbf{i}+c_2\mathbf{j}+d_2\mathbf{k}) = (a_1a_2-b_1b_2-c_1c_2-d_1d_2)+(a_1b_2+b_1a_2+c_1d_2-d_1c_2)\mathbf{i}\\ + (a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2)\mathbf{j} + (a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2)\mathbf{k}.\end{aligned}$$

Sometimes the quaternions are denoted Q_8 or $\mathbf{H}$.

The above multiplication is rather cumbersome. It is often more useful to multiply elements of $\mathbb{H}$ together as if they are polynomials (with real coefficients and indeterminates **i, j,** and **k**) and then apply the rules given in the following table:

*	1	**i**	**j**	**k**
1	1	**i**	**j**	**k**
i	**i**	-1	**k**	**-j**
j	**j**	**-k**	-1	**i**
k	**k**	**j**	**-i**	-1

By rather laborious calculations it is possible to show that $\mathbb{H}$, together with the addition and multiplication defined above, satisfy the ring axioms. Thus $\mathbb{H}$ is a ring.

Example 1.7. Let R be a finite set. Then it is possible to specify operations $+$ and $*$ on R using addition and multiplication tables.

For example, if $R = \{0, a, b, c\}$, then such tables might look like

$+$	0	a	b	c
0	0	a	b	c
a	a	0	c	b
b	b	c	0	a
c	c	b	a	0

$*$	0	a	b	c
0	0	0	0	0
a	0	0	a	a
b	0	0	b	b
c	0	0	c	c

.

It is almost impossible to check by hand that these tables satisfy the ring axioms. But they do, and R is a ring.

2. Further axioms for rings

Definition 2.1. Let R be a ring satisfying the axiom

M2. there exists $1 \in R$ such that $1 * r = r * 1 = r$ for all $r \in R$.

Then R is called a *ring with identity* or a *ring with* 1. The element $1 \in R$ is referred to as the *multiplicative identity* or *one* of R.

The 1 in Axiom **M2** is symbolic and not necessarily the integer 1. Some mathematicians include the existence of a multiplicative identity as an axiom in the definition of a ring. In this course we will not assume, unless otherwise stated, that our rings have a multiplicative identity.

Example 2.2. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/(n)$ all have multiplicative identity $1 \in \mathbb{Z}$. The one of the ring $M_2(\mathbb{R})$ is the *identity matrix*

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The multiplicative identity of the quaternions $\mathbb{H}$ is $1 + 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k} = 1$. The ring R in Example 1.7 has no multiplicative identity.

Definition 2.3. If R is a ring with identity 1, then the *multiplicative inverse* of an element $x \in R$ is an element $x^{-1} \in R$ such that

$$x * x^{-1} = x^{-1} * x = 1.$$

If $x \in R$ has a multiplicative inverse, then x is called a *unit*.

Even if a ring R has a multiplicative identity, it may not be possible to find a multiplicative inverse for every element in R. In particular, if $|R| > 1$, then the element 0 will never have an inverse.

Example 2.4.

1. Let $x \in \mathbb{Z}$ be arbitrary. Then x is a unit if there exists $y \in \mathbb{Z}$ such that $x * y = y * x = 1$. It follows that the only units in $\mathbb{Z}$ are 1 and -1.
2. A matrix $A \in M_2(\mathbb{R})$ is invertible (i.e. is a unit) if and only if $\det(A) \neq 0$.
3. Let $x \in \mathbb{Z}/(n)$. Then x is a unit if and only if x is coprime to n: see Tutorial Sheet 2.
4. It can be shown (see Tutorial Sheet 2) that the multiplicative inverse of a non-zero element $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ in the quaternions $\mathbb{H}$ is

$$(a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}) \cdot \frac{1}{a^2 + b^2 + c^2 + d^2}.$$

Definition 2.5. Let R be a ring with identity, such that $0 \neq 1$. If R satisfies the axiom

M3. every $r \in R \setminus \{0\}$ is a unit

then R is a *division ring* (or sometimes a *skew field*).

Example 2.6. Every non-zero element of $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ is a unit. Hence the rings $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ are division rings. The ring $\mathbb{Z}$ is not a division ring. The ring $\mathbb{Z}/(n)$ is a division ring if and only if n is prime. The ring $M_2(\mathbb{R})$ is not a division ring. The quaternions form a division ring.

Definition 2.7. A ring R is *commutative* if R satisfies the axiom

M4. $r * s = s * r$ for all $r, s \in R$.

Example 2.8. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ and $\mathbb{Z}/(n)$ are commutative. In $M_2(\mathbb{R})$

$$\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 11 \\ 3 & 5 \end{pmatrix} \neq \begin{pmatrix} 4 & 6 \\ 5 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

Hence $M_2(\mathbb{R})$ is a non-commutative ring.

Since $\mathbf{i} * \mathbf{j} = \mathbf{k} \neq -\mathbf{k} = \mathbf{j} * \mathbf{i}$, it follows that $\mathbb{H}$ is non-commutative.

Definition 2.9. A commutative division ring is called a *field*. That is, a field is a set F together with a pair of binary operations $+$ and $*$ satisfying the axioms: **A1, A2, M1, M2, M3, M4**, and **D**.

Example 2.10. So, $\mathbb{Z}$ is not a field but $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are fields. The ring $\mathbb{Z}/(n)$ is a field if and only if n is a prime number. The ring $M_2(\mathbb{R})$ fails to satisfy **M4**, so $M_2(\mathbb{R})$ is not a field. Likewise, the quaternions $\mathbb{H}$ fail **M4**, so $\mathbb{H}$ is not a field.

Definition 2.11. If R is a ring and $a, b \in R \setminus \{0\}$ satisfy $a * b = 0$, then a and b are *zero divisors*.

Example 2.12. In $\mathbb{Z}/(6)$ we have $2 * 3 = 0$ and so 2 and 3 are zero divisors. More generally, $x \in \mathbb{Z}/(n)$ is a zero divisor if and only if $\gcd(x, n) \neq 1$: see Tutorial Sheet 2.

Definition 2.13. An *integral domain* is a ring R such that

- R is commutative;
- R has a multiplicative identity $1 \neq 0$;
- no element of R is a zero divisor.

So, if R is an integral domain and $a * b = 0$, then $a = 0$ or $b = 0$.

Example 2.14. The commutative ring $\mathbb{Z}$ is an integral domain, as are the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$. In $M_2(\mathbb{R})$

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and so $M_2(\mathbb{R})$ is not an integral domain. The ring of quaternions $\mathbb{H}$ is not an integral domain as $*$ is not commutative. However, $\mathbb{H}$ has no zero divisors (see Tutorial Sheet 2).

Theorem 2.15. *Let a be a unit in a ring R. Then a is not a zero divisor. Hence, every field is an integral domain.*

Proof. Let $a, b \in R$ with $a * b = 0$. Then $0 = a^{-1} * 0 = a^{-1} * a * b = 1 * b = b$. Thus a is not a zero divisor. A field is a commutative ring with 1 in which every nonzero element is a unit, so the result follows. ■

The following table shows the different axioms satisfied by the different types of rings defined in this section. **Z** denotes that the ring has no zero divisors.

	A1	**A2**	**M1**	**D**	**M2**	**M3**	**M4**	**Z**	
1	✓	✓	✓	✓	✗	✗	✗	✗	ring
2	✓	✓	✓	✓	✗	✗	✓	✗	commutative ring
3	✓	✓	✓	✓	✓	✗	✗	✗	ring with identity
4	✓	✓	✓	✓	✗	✗	✗	✓	ring with no zero divisors
5	✓	✓	✓	✓	✓	✗	✓	✗	comm. ring with one
6	✓	✓	✓	✓	✗	✗	✓	✓	comm. ring with no zero divisors
7	✓	✓	✓	✓	✓	✗	✗	✓	ring with one and no zero divisors
8	✓	✓	✓	✓	✓	✗	✓	✓	integral domain
9	✓	✓	✓	✓	✓	✓	✗	✓	division ring
10	✓	✓	✓	✓	✓	✓	✓	✓	field

The following table shows which of the properties given in the table above are satisfied by the examples we have followed throughout this section.

	1	2	3	4	5	6	7	8	9	10	
$\mathbb{Z}$	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	
$\mathbb{Q}, \mathbb{R}, \mathbb{C}$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
$\mathbb{Z}/(n)$	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	n composite
$\mathbb{Z}/(p)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	p prime
$M_2(\mathbb{R})$	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	
$\mathbb{H}$	✓	✗	✓	✓	✗	✗	✓	✗	✓	✗	

We have seen that the following are equivalent

(i) n is a prime;

(ii) $\mathbb{Z}/(n)$ is a division ring;

(iii) $\mathbb{Z}/(n)$ is a field;

(iv) $\mathbb{Z}/(n)$ is an integral domain.

In fact, if $\mathbb{Z}/(n)$ is replaced with any finite ring R in (ii), (iii), and (iv), then they are still equivalent! Clearly (iii) implies (ii) and (iv). The following shows that (iv) implies (iii), but proving that (ii) implies (iii) is very hard.

Theorem 2.16. *Every finite integral domain I is a field.*

Proof. Since I is an integral domain, we know that I is commutative and has a multiplicative identity. Hence the only thing we need to show is that an arbitrary non-zero element $a \in I$ has a multiplicative inverse. The sequence $a, a^2, a^3, \ldots$ can only contain finitely many elements of I, since there are only finitely many elements in I. Therefore $a^m = a^n$ for some $m < n$ (say). Then $0 = a^m - a^n = a^m(1 - a^{n-m})$. Since there are no zero divisors and $a \neq 0$ it follows that $a^m \neq 0$. Hence $1 - a^{n-m} = 0$ and so $1 = a * a^{n-m-1}$. It follows that $a^{n-m-1} = a^{-1}$ is a multiplicative inverse for a. ■

Example 2.17. Let $\mathbb{Z}[i]$ denote the subset $\{a + bi \ : \ a, b \in \mathbb{Z}\}$ of $\mathbb{C}$ where $i = \sqrt{-1}$, with the usual addition and multiplication of complex numbers. Then $(a + bi) - (c + di) \in \mathbb{Z}[i]$ for all $a + bi, c + di \in \mathbb{Z}[i]$, so $(\mathbb{Z}[i], +)$ is an additive subgroup of $\mathbb{C}$, and is commutative (since $\mathbb{C}$ is commutative). Furthermore, $(a+bi)*(c+di) \in \mathbb{Z}[i]$, and the multiplication is associative, because multiplication is associative in $\mathbb{C}$. Distributivity is also inherited from $\mathbb{C}$. Hence $\mathbb{Z}[i]$ is a ring, called the *Gaussian integers*.

The integer $1 = 1 + 0i$ is the identity of $\mathbb{C}$, and is an element of $\mathbb{Z}[i]$. Hence $\mathbb{Z}[i]$ has a multiplicative identity. We will determine the units of $\mathbb{Z}[i]$ in a later section. Also, multiplication in $\mathbb{C}$ is commutative, so $\mathbb{Z}[i]$ is a commutative ring. Since $\mathbb{C}$ has no zero divisors, it follows that $\mathbb{Z}[i]$ has no zero divisors. Hence $\mathbb{Z}[i]$ is an integral domain.

Example 2.18. Let M denote the set of all real 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}.$$

If

$$A = \begin{pmatrix} a_1 & b_1 \\ 0 & d_1 \end{pmatrix}, B = \begin{pmatrix} a_2 & b_2 \\ 0 & d_2 \end{pmatrix} \in M,$$

then

$$A + B = \begin{pmatrix} a_1 + a_2 & b_1 + b_1 \\ 0 & d_1 + d_2 \end{pmatrix} \in M \text{ and } A * B = \begin{pmatrix} a_1 a_2 & a_1 b_2 + b_1 d_2 \\ 0 & d_1 d_2 \end{pmatrix} \in M.$$

So, it makes sense to talk about the usual matrix operations of $+$ and $*$ on M. Notice that the zero matrix is an element of M, and that M is closed under negation. Hence, it follows that M satisfies Axioms **A1, A2, M1** and **D**, as $M_2(\mathbb{R})$ does. The identity of M is the identity matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and $A \in M$ has an inverse if and only if $a \neq 0$ and $d \neq 0$. To prove that M is not commutative it suffices to see that

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

The matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

are zero divisors in M. Hence M is a non-commutative ring with one and zero divisors that is not a division ring.

3. Polynomial rings

Polynomials are the source of some of the most important examples of rings.

Definition 3.1. Let R be a ring. Then a *polynomial over* R is an expression of the form

$$f = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where n is a non-negative integer, the *coefficients* $a_0, a_1, \ldots, a_n$ are elements of R and x is a symbol not in R called the *indeterminate*. Two polynomials are equal if and only if all of their coefficients are equal.

To reiterate, the indeterminate x is not a member of R, and neither are x^2, $x^3, \ldots$. They are simply markers that indicate how to add and multiply.

Definition 3.2. The *degree* of a polynomial f, denoted $\deg(f)$, is the largest n such that $a_n \neq 0$. The *zero polynomial* is written 0, and has all coefficients equal to 0. By convention, $\deg(0) = -\infty$. Polynomials of degree 0 are called *constant polynomials*.

Addition and multiplication of polynomials is done in the usual way. That is, if $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$ (taking coefficients to be 0 if necessary to ensure that the degrees are equal), then

$$f + g = \sum_{i=0}^{n} (a_i + b_i) x^i$$

and

$$f * g = \sum_{k=0}^{2n} c_k x^k \text{ where } c_k = \sum_{0 \leq i,j \leq n \; i+j=k} a_i b_j.$$

Example 3.3. If $f = 1 + 13x + 3x^2 + x^3$ and $g = x + 3x^3$ are polynomials over the ring $\mathbb{Z}/(14)$, then

$$f + g = 1 + 3x^2 + 4x^3 \text{ and } f * g = x + 13x^2 + 6x^3 + 12x^4 + 9x^5 + 3x^6.$$

Definition 3.4. We denote by $R[x]$ the set of polynomials over R with the operations $+$ and $*$ given above.

Theorem 3.5. *Let R be a ring. Then $R[x]$ is a ring called the* ring of polynomials over R, *and its zero element is the zero polynomial. The ring $R[x]$ is commutative if and only if R is commutative.*

Proof. See Tutorial Sheet 2. ■

Example 3.6. Let us consider the polynomial ring $\mathbb{Z}[x]$. The one of $\mathbb{Z}[x]$ is just $1 \in \mathbb{Z}[x]$ (the polynomial of degree 0) and $\mathbb{Z}[x]$ is commutative as $\mathbb{Z}$ is commutative.

What are the units of $\mathbb{Z}[x]$? If $f = \sum_{i=0}^{n} a_i x^i, g = \sum_{i=0}^{m} b_i x^i \in \mathbb{Z}[x]$ and $f * g = 1$, then the only non-zero coefficients must be a_0 and b_0. Hence either $f = 1$ and $g = 1$ or $f = -1$ and $g = -1$. So, $\mathbb{Z}[x]$ is not a field.

Let $f, g \in \mathbb{Z}[x]$ be non-zero polynomials with $\deg(f) = m$ and $\deg(g) = n$. Then $\deg(f * g) = m + n \in \mathbb{N}$. In particular, $f * g \neq 0$ and so $\mathbb{Z}[x]$ is an integral domain.

You can prove using an analogous argument that $\mathbb{R}[x]$ is a commutative ring with identity, that the units of $\mathbb{R}[x]$ are the constant polynomials, and that $\mathbb{R}[x]$ is an integral domain.

The following lemma is straightforward.

Lemma 3.7. *Let $f, g \in R[x]$. Then*

$$\deg(f + g) \leq \max(\deg(f), \deg(g)) \text{ and } \deg(f * g) \leq \deg(f) + \deg(g).$$

If R is an integral domain, then

$$\deg(f * g) = \deg(f) + \deg(g).$$

Chapter 4

Subrings and quotient rings

Subrings and ideals play the same role in ring theory as subgroups and normal subgroups do in group theory.

1. Subrings

Definition 1.1. A *subring* S of a ring R is a subset of R which is a ring under the same operations as R. We will write $S \leq R$ to denote that S is a subring of R.

Rather than rechecking all the ring axioms for S we can simply apply the following lemma.

Lemma 1.2. *Let S be a non-empty subset of a ring R. The set S is a subring of R if and only if $a - b, a * b \in S$ for all $a, b \in S$.*

Proof. The forward direction is clear, so we prove the other direction.

The fact that S closed under $a - b$ implies that $(S, +)$ is a subgroup of $(R, +)$, so in particular $(S, +)$ is a group, and Axiom **A1** holds. Addition is commutative in S (Axiom **A2**) because it is commutative in R.

We have assumed that S is closed under multiplication. Associativity of multiplication (Axiom **M1**), and the distributive laws (Axiom **D**) in S follow from the respective properties of R. ■

Example 1.3. Let R denote the subset $\{a + b\sqrt{5} \ : \ a, b \in \mathbb{Z}\}$ of $\mathbb{R}$. If $+$ and $*$ denote the usual operations on $\mathbb{R}$, then

$$(a + b\sqrt{5}) - (c + d\sqrt{5}) = (a - c) + (b - d)\sqrt{5} \in R$$

and

$$(a + b\sqrt{5})(c + d\sqrt{5}) = (ac + 5bd) + (ad + bc)\sqrt{5} \in R.$$

Hence R is a subring of $\mathbb{R}$. An analogous argument shows that $\{x + y\sqrt{5} \ : \ x, y \in \mathbb{Q}\}$ with $+$ and $*$ is also a subring of $\mathbb{R}$.

Example 1.4. Let (2) denote the even integers. Then

$$2i - 2j = 2(i - j) \in (2)$$

and

$$2i * 2j = 2 * 2ij \in (2).$$

Hence, by Lemma 1.2, (2) is a subring of $\mathbb{Z}$.

More generally, if n is any integer, then the same reasoning shows that the set (n) of all multiples of n is a subring (n) of $\mathbb{Z}$. On the other hand, the odd integers do not form a subring of $\mathbb{Z}$ since $5 - 3 = 2$ is not odd.

Example 1.5. The subsets $\{0, 2, 4\}$ and $\{0, 3\}$ are subrings of $\mathbb{Z}/(6)$.

Lemma 1.6. *Let R be a ring and let S and T be subrings of R. Then $S \cap T$ is a subring of R.*

Proof. Since $0 \in S \cap T$, it follows that $S \cap T$ is non-empty. If $s, t \in S \cap T$, then in particular $s, t \in S$ so $s - t \in S$ and $s * t \in S$. Similarly, $s - t \in T$ and $s * t \in T$, so $s - t \in S \cap T$ and $s * t \in S \cap T$. The result now follows from Lemma 1.2. ■

2. Ideals

Definition 2.1. A subring I of a ring R is an *ideal* of R if $r * a, a * r \in I$, for all $a \in I$ and for all $r \in R$.

If R is any ring, then $\{0\}$ and R are ideals in R. An ideal of a ring R is called *proper* if it is not equal to R, and *nontrivial* if it is not equal to $\{0\}$. [Some books use "proper" to mean nontrivial as well as not equal to R.]

Example 2.2. In Example 1.4, we saw that the even integers (2) are a subring of $\mathbb{Z}$. If $i \in (2)$ and $r \in \mathbb{Z}$, then $i = 2 * j$, for some j, and so $i * r = r * i = 2 * jr \in (2)$. Hence (2) is an ideal of $\mathbb{Z}$.

In Example 1.5, we showed that $I = \{0, 2, 4\}$ and $J = \{0, 3\}$ are subrings of $\mathbb{Z}/(6)$. If $i \in I$ and $r \in \mathbb{Z}/(6)$, then from the table below, $i * r = r * i \in I$.

*	0	1	2	3	4	5
0	0	0	0	0	0	0
2	0	2	4	0	2	4
4	0	4	2	0	4	2

Similarly, J is an ideal of $\mathbb{Z}/(6)$.

Example 2.3. Let $\mathbb{Z}[i]$ be the Gaussian integers, from Chapter 3, Example 2.17. Let $a+bi \in \mathbb{Z}[i]$, and let $r = c/d \in \mathbb{Q}$. If $d \nmid ac$ then $(c/d) * (a + bi) \notin \mathbb{Z}[i]$. Hence $\mathbb{Z}[i]$ is not an ideal of $\mathbb{C}$ even though it is a subring.

Likewise $\mathbb{Z}[\sqrt{5}] = \{a + b\sqrt{5} \ : \ a, b \in \mathbb{Z}\}$ is not an ideal of $\mathbb{R}$ or $\mathbb{C}$, although it is a subring of both.

Example 2.4. Let R be a commutative ring with 1, and let $I \subseteq R[x]$ be all the polynomials that have constant coefficient 0. That is,

$$I = \{\sum_{i=0}^{n} a_i x^i \in R[x] \ : \ a_0 = 0\}.$$

Let $f_1, f_2 \in I$. Then

$$f_1 = x * g_1, \ f_2 = x * g_2$$

for some $g_1, g_2 \in R[x]$. Hence

$$f_1 - f_2 = x * g_1 - x * g_2 = x * (g_1 - g_2) \in I,$$

and, for all $h \in R[x]$,

$$f_1 * h = x * g_1 * h = x * (g_1 * h) \in I \text{ and } h * f_1 = h * x * g_1 = x * (h * g_1) \in I$$

It follows that I is an ideal in $R[x]$.

Lemma 2.5. *Let I be an ideal of a ring R with identity. Then*

(i) *if $1 \in I$, then $I = R$;*

(ii) *if R is a division ring, then $I = R$ or $I = \{0\}$.*

Proof. **(i).** If $r \in R$, then $r = r * 1 \in I$. Hence $I = R$.
(ii). Assume that $I \neq \{0\}$. Then there exists $a \in I$ such that $a \neq 0$. Hence $1 = a^{-1} * a \in I$. It follows that from part (i) that $I = R$. ■

Corollary 2.6. *Let R be a commutative ring with $1 \neq 0$. Then $R[x]$ is not a division ring.*

Proof. We saw in Example 2.4 that the polynomials with 0 constant coefficient form a proper nontrivial ideal in $R[x]$. Hence, by Lemma 2.5(ii), $R[x]$ is not a division ring. ■

Example 2.7. Let $R = M_2(\mathbb{R})$, and let I denote the set of matrices of the form

$$\begin{pmatrix} 0 & a \\ 0 & b \end{pmatrix}.$$

Now,

$$\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} - \begin{pmatrix} 0 & y \\ 0 & t \end{pmatrix} = \begin{pmatrix} 0 & b-y \\ 0 & d-t \end{pmatrix} \in I, \text{ and } \begin{pmatrix} x & y \\ z & t \end{pmatrix} * \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} 0 & xb+yd \\ 0 & zb+td \end{pmatrix} \in I.$$

Thus I is a subring and $r * i \in I$ for all $i \in I$ and for all $r \in R$. But

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \notin I.$$

Hence it is not true that $i * r \notin I$ for all $i \in I$ and for all $r \in R$. It follows that I is not an ideal of R. In fact, R has no proper ideals except $\{0\}$.

Lemma 2.8. *Let R be a commutative ring with identity, and $r_1, r_2, \ldots, r_n \in R$. Then*

$$(r_1, r_2, \ldots, r_n) = \{\lambda_1 r_1 + \lambda_2 r_2 + \cdots + \lambda_n r_n \ : \ \lambda_1, \lambda_2, \ldots, \lambda_n \in R\}.$$

is an ideal.

Proof. If $i = \lambda_1 r_1 + \lambda_2 r_2 + \cdots + \lambda_n r_n, j = \mu_1 r_1 + \mu_2 r_2 + \cdots + \mu_n r_n \in (r_1, r_2, \ldots, r_n)$, then

$$i - j = (\lambda_1 - \mu_1) r_1 + (\lambda_2 - \mu_2) r_2 + \cdots + (\lambda_n - \mu_n) r_n \in (r_1, r_2, \ldots, r_n).$$

If $x \in R$ is arbitrary, then

$$xi = ix = (x\lambda_1) r_1 + (x\lambda_2) r_2 + \cdots + (x\lambda_n) r_n \in (r_1, r_2, \ldots, r_n).$$

The result follows. ■

The assumption that R has an identity implies that $r_1, r_2, \ldots, r_n \in (r_1, r_2, \ldots, r_n)$.

Definition 2.9. We call $(r_1, \ldots, r_n)$ the ideal *generated by* $r_1, \ldots, r_n$. The ideal (r) generated by a single element $r \in R$ is called the *principal ideal generated by* r.

Example 2.10. We have already seen that the even integers (2) form an ideal in $\mathbb{Z}$. Moreover,

$$(2) = \{2n \ : \ n \in \mathbb{Z}\}$$

and so (2) is the principal ideal generated by 2.

Example 2.11. Let R be a commutative ring with $1 \neq 0$. In Example 2.4 we showed that the set I of all polynomials in $R[x]$ with 0 constant coefficient is an ideal. We also saw that $f \in I$ if and only if $f = x * g$ for some $g \in R[x]$. It follows that

$$I = (x) = \{x * g \ : \ g \in R[x]\}.$$

We finish with a lemma that will be useful in the next few sections.

Lemma 2.12. *(i) Let I and J be ideals in a ring R. Then $I \cap J$ and*

$$I + J = \{i + j \ : \ i \in I, j \in J\}$$

are ideals of R.

(ii) Let R be a commutative ring with identity, let I be an ideal of R, and let $r_1, r_2, \ldots, r_n \in I$. Then $(r_1, r_2, \ldots, r_n) \subseteq I$.

Proof. For (i), it follows from Lemma 1.6 that $I \cap J$ is a subring of R. To see that $I \cap J$ is an ideal, let $r \in R$ and $x \in I \cap J$. Then $r * x, x * r \in I$ and $r * x, x * r \in J$, so $r * x, x * r$ in $I \cap J$.

To see that $I + J$ is an ideal, first note that $I \subset I + J$ so $I + J$ is nonempty. Next, we calculate that for all $i_1, i_2 \in I$, $j_1, j_2 \in J$ and $r \in R$

$$\begin{gathered}(i_1 + j_1) - (i_2 + j_2) = (i_1 - i_2) + (j_1 - j_2) \in I + J \\ r * (i_1 + j_1) = r * i_1 + r * j_1 \in I + J \\ (i_1 + j_1) * r = i_1 * r + j_1 * r \in I + J.\end{gathered}$$

Hence $I + J$ is closed under subtraction, and is closed under multiplication on the right and left by all elements of R, so is an ideal.

For (ii), let $a \in (r_1, r_2, \ldots, r_n)$. We must show that $a \in I$. By definition of $(r_1, \ldots, r_n)$, there exist $\lambda_i \in R$ such that

$$a = \lambda_1 r_1 + \cdots \lambda_n r_n.$$

We see that $s_i = \lambda_i r_i \in I$ for all i, since $r_i \in I$ and I is an ideal. The fact that I is closed under addition now implies that $a = s_1 + s_2 + \cdots + s_n \in I$. ■

The second part of Lemma 2.12 states that $(r_1, r_2, \ldots, r_n)$ is the least ideal containing $r_1, r_2, \ldots, r_n$.

3. Quotients of rings

Definition 3.1. Let R be a ring, let I be an ideal of R, and let $a \in R$. Then the *coset* (or *residue class*) of I with *representative* a is the subset

$$a + I = \{a + s \ : \ s \in I\}$$

of R.

Theorem 3.2. *Let R be a ring and let I be an ideal of R. Then the following hold, for all $a, b \in R$:*

(i) *$a + I = b + I$ if and only if $a - b \in I$;*

(ii) *any two cosets of I are either equal or disjoint: either $a + I = b + I$ or $(a + I) \cap (b + I) = \emptyset$;*

(iii) *R is a disjoint union of the cosets of I;*

(iv) *the map $f_a : r \mapsto a + r$ is a bijection from I to the coset $a + I$.*

Proof. We are only considering the additive structure of R. The ideal I is a subgroup of the group $(R, +)$, so the result follows (see MT2002 or MT2505 or MT4003 or MT4516 for the proof of the result for groups). ■

Theorem 3.3. *Let R be a ring, I be an ideal of R, and let R/I denote the set*

$$\{a + I \ : \ a \in R\}$$

of cosets of the ideal I. Then R/I is a ring under the operations defined by

$$(a + I) + (b + I) = (a + b) + I \text{ and } (a + I) * (b + I) = (a * b) + I.$$

Proof. The group $(R, +)$ is abelian, so $(I, +)$ is a normal subgroup of $(R, +)$. Hence $(R/I, +)$ is a group, and is abelian because $(R, +)$ is abelian. Thus **A1** and **A2** hold.

We must prove that the operation $*$ is well-defined and that $(R/I, +, *)$ satisfies the remaining ring axioms.

To show that $*$ is well-defined we must prove that if $a_1 + I = a_2 + I$ and $b_1 + I = b_2 + I$, then

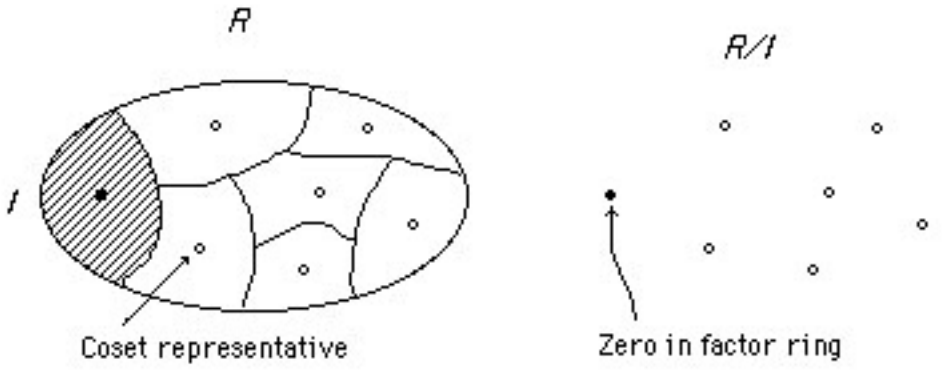
$$(a_1 b_1) + I = (a_2 b_2) + I.$$

From the direct implication of Theorem 3.2(i), $a_1 - a_2 \in I$ and $b_1 - b_2 \in I$. We deduce that $(a_1 - a_2)b_1 + a_2(b_1 - b_2) = a_1 b_1 - a_2 b_2 \in I$, and so $(a_1 b_1) + I = (a_2 b_2) + I$, and $*$ is well-defined.

Associativity of $*$ (Axiom **M1**) and the distributive law (Axiom **D**) follow easily from the corresponding properties for R. ■

Definition 3.4. The ring R/I is called a *factor ring* (or sometimes a *quotient ring* or *residue class ring*); we may also say that R/I is *the quotient of R by I*. We may also refer to the factor ring R/I as *R modulo I*.

The way you should think of the factor ring R/I is that it is the ring R where all the elements in the ideal I have been 'made' into zero. Here are two pictures:



Example 3.5. Let $n \in \mathbb{Z}$ and let (n) be the principal ideal generated by n. Then we can form the quotient $\mathbb{Z}/(n)$ of $\mathbb{Z}$ by (n). Two elements $x, y \in \mathbb{Z}$ are representatives of the same coset in $\mathbb{Z}/(n)$ if and only if $x - y \in (n)$. This happens if and only if $x \equiv y \bmod n$. It follows that the cosets in $\mathbb{Z}/(n)$ are

$$0 + (n), 1 + (n), \ldots, n - 1 + (n)$$

and we add and multiply the representatives of cosets modulo n. For the sake of brevity, we omit the (n) when referring to elements of $\mathbb{Z}/(n)$ and we get back to the elements $\{0, 1, 2, \ldots, n - 1\}$ modulo n!

The moral of the story: you've been working with factor rings since the start of the course!

We will now consider another very important class of factor rings. We showed on Tutorial Sheet 2 that if p is a prime, then $\mathbb{Z}/(p)$ is a field.

Definition 3.6. If p is a prime number, we will denote $\mathbb{Z}/(p)$ by $\mathbb{F}_p$ and we will refer to it as the *Galois field of order p*.

Example 3.7. Let $f = x \in \mathbb{F}_2[x]$. Then the elements of $\mathbb{F}_2[x]/(f)$ are the cosets

$$g + (f)$$

where $g \in \mathbb{F}_2[x]$. The zero of $\mathbb{F}_2[x]/(f)$ is the coset

$$0 + (f) = x + (f).$$

So, if $g \in \mathbb{F}_2[x]$, then

$$g + (f) \in \{0 + (f), 1 + (f)\},$$

as whenever an x appears in g it can be replaced by 0. Thus the order of $\mathbb{F}_2[x]/(f)$ is 2. For example,

$$x^4 + x^2 + 1 + (f) = 0 + 0 + 1 + (f) = 1 + (f).$$

Example 3.8. Let $f = x^2 + x + 1 \in \mathbb{F}_2[x]$. Then the elements of $\mathbb{F}_2[x]/(f)$ are

$$0 + (f), 1 + (f), x + (f), x + 1 + (f)$$

as again whenever x^2 appears it can be replaced by $x + 1$. For example,

$$[1 + x + (f)] * [1 + x + (f)] = 1 + 2x + x^2 + (f) = x + (f).$$

The multiplicative inverses of $1+(f), x+(f), 1+x+(f)$ are $1+(f), 1+x+(f), x+(f)$, respectively. So, $\mathbb{F}_2[x]/(f)$ is a field with 4 elements.

We will return to the study of polynomial factor rings later in the course.

4. Principal ideals, prime ideals, and maximal ideals

In this section we will briefly introduce three special types of ideals. They will enable us to determine when a factor ring R/I is an integral domain, and when it a field.

Definition 4.1. A *principal ideal domain (PID)* is an integral domain where every ideal is principal.

Lemma 4.2. *The ring of integers $\mathbb{Z}$ is a principal ideal domain.*

Proof. We saw in Chapter 3, Example 2.14 that $\mathbb{Z}$ is an integral domain. Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$, then $I = (0)$. Otherwise, assume that $I \neq \{0\}$ and pick a to be the minimal positive element of I. We know from Lemma 2.12(ii) that $(a) \subseteq I$. Now, assume there is an element $b \in I \backslash (a)$. By the division algorithm we have $b = q * a + r$ with $0 < r < a$, but $r = b - q * a \in I$, a contradiction to a being minimal. ■

Definition 4.3. An ideal I of a ring R is said to be *prime* if $I \neq R$, and $ab \in I$ implies that $a \in I$ or $b \in I$, for all $a, b \in R$.

Theorem 4.4. *Let I be an ideal of a commutative ring R with one. Then I is a prime ideal if and only if R/I is an integral domain.*

Proof. ($\Leftarrow$) Assume that R/I is an integral domain. Then R/I contains a (non-zero) multiplicative identity, and so R/I has order greater than 1. Hence $R \neq I$. If $a, b \in R$ with $ab \in I$, then $ab + I = (a + I)(b + I) = 0 + I$. Hence, since R/I is an integral domain, either $a + I = 0 + I$ or $b + I = 0 + I$. Thus either $a \in I$ or $b \in I$ by Theorem 3.2(i) and so I is a prime ideal.

($\Rightarrow$) See Tutorial Sheet 3. ■

The zero ideal $\{0\}$ is prime if and only if R is an integral domain. Hence if $I = \{0\}$, then Theorem 4.4 says that I is prime if and only if $R/I = R$ is an integral domain (i.e. it says nothing!).

Definition 4.5. An ideal I of a ring R is said to be *maximal* if $I \neq R$ and $I \subsetneq J$ for some ideal J of R implies $J = R$ (i.e. I is not contained in any other proper ideals of R).

Theorem 4.6. *Let I be an ideal of a commutative ring R with one. Then I is a maximal ideal if and only if R/I is a field.*

Proof. ($\Leftarrow$) Assume that R/I is a field. Then $R/I \neq \{0\}$ and for all non-zero $a + I \in R/I$ (i.e. $a \notin I$) there exists $b + I \in R/I$ such that $(a + I)(b + I) = ab + I = 1 + I$. Hence for all $a \in R \setminus I$ there exists $b \in R \setminus I$ such that $ab - 1 \in I$.

Let J be an ideal of R where $I \subseteq J \subseteq R$. If $a \in J \setminus I$, then there exists $b \notin I$ such that $ab - 1 \in I \subseteq J$. But $ab \in J$ since $a \in J$ and J is an ideal. Hence $1 = ab - (ab - 1) \in J$ and so $J = R$. We have shown that I is a maximal ideal of R.

($\Rightarrow$) Assume that I is a maximal ideal of R. Let $a + I \in R/I$, such that $a + I \neq 0 + I$. Then in particular $a \notin I$. Set

$$J = I + (a).$$

By Lemma 2.12, J is an ideal of R and it is clear that $I \subsetneq J$. Since I is maximal, it follows that $J = R$. Hence $1 \in J$ and so $1 = i + ra$ for some $i \in I$ and $r \in R$. Now,

$$1 + I = [i + ra] + I = ra + I = (r + I)(a + I)$$

It follows that $a + I$ is a unit in R/I and, since a was arbitrary, R/I is a field. ■

In a commutative ring with one, the zero ideal $\{0\}$ is maximal if and only if $(r) = R$ for any $r \in R \setminus \{0\}$. This happens if and only if $1 \in (r)$ for any $r \in R \setminus \{0\}$, which happens if and only if every nonzero element of R is a unit. But this is the case if and only if R is a field. Hence if $I = \{0\}$, then Theorem 4.6 says that I is maximal if and only if $R/I = R$ is a field (i.e. it says nothing!).

Corollary 4.7. *Let I be a maximal ideal in a commutative ring with identity. Then I is a prime ideal.*

Proof. If I is maximal then by Theorem 4.6 the quotient ring R/I is a field and hence an integral domain. Hence by Theorem 4.4 the ideal I is prime. ■

Example 4.8. The ring of integers $\mathbb{Z}$ is a principal ideal domain. Hence every ideal of $\mathbb{Z}$ is of the form (n) for some $n \in \mathbb{Z}$. If $p \in \mathbb{Z}$ is a prime number and $n \in \mathbb{Z}$ such that $(p) \subseteq (n)$, then $p \in (n)$ and so $n \mid p$. Thus $n = \pm 1$ or $n = \pm p$. It follows that $(n) = \mathbb{Z}$ or $(n) = (p)$, and so (p) is a maximal ideal.

On the other hand, if $a, b \in \mathbb{Z}$ with $a, b \neq \pm 1$, then $ab \in (ab)$ but $a \notin (ab)$ and $b \notin (ab)$. Hence (ab) is not a prime ideal and hence not a maximal ideal.

We have proved the following result.

Lemma 4.9. *Let $n \in \mathbb{Z}$. Then the following are equivalent:*

(i) *n or $-n$ is a prime number;*

(ii) *(n) is a prime ideal;*

(iii) *(n) is a maximal ideal.*

Chapter 5

Homomorphisms and Isomorphisms

1. Homomorphisms

Just as in many branches of mathematics, maps that preserve the structure of a ring are very important.

Definition 1.1. A map $f : R \to S$ between rings is called a *ring homomorphism* if

$$f(x+y) = f(x) + f(y) \text{ and } f(x*y) = f(x) * f(y) \text{ for all } x, y \in R.$$

If f is a bijection, then f is called a *ring isomorphism*. In this case we say that the rings R and S are *isomorphic* and write $R \cong S$.

Note that $+$ in $f(x+y)$ is the operation in R and $+$ in $f(x)+f(y)$ is the operation in S. Likewise, $*$ is the operation in R in $f(x*y)$ and in S in $f(x)*f(y)$.

Isomorphic rings have identical ring-theoretic properties, that is, commutativity, the existence of an identity, being a field etc. Any pair of isomorphic rings can be regarded as the same.

Definition 1.2. Let R and S be rings and let $f : R \to S$ be a homomorphism. Then the *kernel* of f is

$$\ker(f) = \{r \in R \ : \ f(r) = 0_S\},$$

and the *image* of f is

$$\text{im}\,(f) = \{s \in S \ : \ f(r) = s \text{ for some } r \in R\}.$$

Example 1.3. Let $f_n : \mathbb{Z} \to \mathbb{Z}/(n)$ be defined by $f_n : x \mapsto x \bmod n$. Then

$$f_n(i+j) = (i+j) \bmod n \equiv (i \bmod n) + (j \bmod n) = f_n(i) + f_n(j)$$

and

$$f_n(i*j) = i*j \bmod n = (i \bmod n) * (j \bmod n) = f_n(i) * f_n(j)$$

for all $i, j \in \mathbb{Z}$. Hence f_n is a ring homomorphism. The map f_n is not a ring isomorphism since it is not injective: $f_n(0) = 0 = f_n(n)$. The kernel of f_n is the set (n) of all multiples of n, and the image of f_n is the whole of $\mathbb{Z}/(n)$.

Example 1.4. Let $f : \mathbb{Z} \to (2)$ be defined by $f(x) = 2x$. Then

$$f(i+j) = 2(i+j) = 2i + 2j = f(i) + f(j)$$

for all $i, j \in \mathbb{Z}$. However,

$$f(2*2) = 8 \neq 16 = f(2) * f(2)$$

so f is not a homomorphism.

Lemma 1.5. *Let $f : R \to S$ be a ring homomorphism. Then*

(i) *if 0_R and 0_S are the zeros in R and S, respectively, then $f(0_R) = 0_S$;*

(ii) $\ker(f) = \{0\}$ *if and only if f is injective;*

(iii) $\ker(f)$ *is an ideal of R, and* $\operatorname{im}(f)$ *is a subring of S;*

(iv) *if $g : S \to T$ is also a ring homomorphism, then so is $g \circ f : R \to T$.*

Proof. (i) For all $a \in R$, $f(a) = f(0_R + a) = f(0_R) + f(a)$, since f is a homomorphism. So $f(0_R) = 0_S$.

(ii) ($\Rightarrow$) We have to prove that f is injective. Let $f(r) = f(s)$ for some $r, s \in R$. Then $f(r - s) = f(r) - f(s) = f(r) - f(r) = 0$ and so $r - s \in \ker(f)$ and so $r - s = 0$. It follows that $r = s$ and so f is injective.

($\Leftarrow$) If $a \in \ker(f)$ and $a \neq 0$, then $f(a) = f(0) = 0$ and so f is not injective.

(iii), (iv): See Tutorial Sheet 4. ■

Example 1.6. Let $f : \mathbb{Z} \to M_2(\mathbb{R})$ be defined by

$$f(x) = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$f(x + y) = \begin{pmatrix} x + y & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} = f(x) + f(y)$$

and

$$f(x * y) = \begin{pmatrix} x * y & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} * \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} = f(x) * f(y),$$

so f is a homomorphism.

The kernel of f is $\{0\}$, so f is injective by Lemma 1.5 (ii). However, f is not surjective and so is not an isomorphism.

Notice that the identity of $\mathbb{Z}$ is 1 and the identity of $M_2(\mathbb{R})$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = f(1).$$

Example 1.7. Let $\phi : \mathbb{R}[x] \to \mathbb{R}$ be defined by

$$\phi(a_0 + a_1 x + \cdots + a_n x^n) = a_0.$$

Let $f = a_0 + a_1 x + \cdots + a_n x^n$ and $g = b_0 + b_1 x + \cdots + b_n x^n$. Then

$$\phi(f) + \phi(g) = a_0 + b_0 = \phi(f + g)$$

and

$$\phi(f) * \phi(g) = a_0 b_0 = \phi(f * g),$$

so ϕ is a ring homomorphism.

The kernel of ϕ is the set of polynomials with zero constant term: we proved that $\ker(\phi)$ is the principal ideal (x) in Chapter 4, Example 2.11.

Example 1.8. Let $\phi_2 : \mathbb{Z}[x] \to \mathbb{Z}/(2)$ be defined by

$$\phi_2(a_0 + a_1 x + \cdots + a_n x^n) = a_0 \bmod 2.$$

Then $\phi_2 = f_2 \circ \phi$, where f_2 is as in Example 1.3, and ϕ is as in Example 1.7, so ϕ_2 is a homomorphism by Lemma 1.5 (iv). The kernel of ϕ_2 is the set of polynomials with even constant terms.

Example 1.9. Let R denote the set of all real-valued matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Consider the map f from $\mathbb{C}$ to R defined by

$$a + bi \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Let $a + bi, c + di \in \mathbb{C}$. Then

$$f(a+bi) + f(c+di) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ -b-d & a+c \end{pmatrix} = f((a+bi)+(c+di))$$

and

$$f(a+bi)f(c+di) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} ac-bd & ad+bc \\ -(ad+bc) & ac-bd \end{pmatrix} = f((a+bi)(c+di))$$

so f is a homomorphism.

If

$$f(a+bi) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

then $a = b = 0$. Hence $\ker(f) = \{0\}$, and f is injective by Lemma 1.5 (ii). It is clear that f is surjective, so f is an isomorphism, and so R is isomorphic to $\mathbb{C}$!

Example 1.10. The set $C[0,1]$ of real-valued continuous functions on the interval $[0,1]$ is a ring under the operations $+$ and $*$ defined by

$$(f+g)(x) = f(x) + g(x)$$

and

$$(f*g)(x) = f(x) * g(x).$$

The 'evaluation at $1/2$' map $\phi : C[0,1] \to \mathbb{R}$ defined by $\phi(f) = f(1/2)$ is a ring homomorphism since

$$\phi(f+g) = (f+g)(1/2) = f(1/2) + g(1/2) = \phi(f) + \phi(g)$$

and

$$\phi(f*g) = (f*g)(1/2) = f(1/2) * g(1/2) = \phi(f) * \phi(g).$$

The kernel of ϕ is

$$\{f \in C[0,1] \ : \ f(1/2) = 0\}.$$

Example 1.11. Let $\phi : \mathbb{Q}[x] \to \mathbb{Q}$ be defined by

$$\phi(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0 + a_1 + \cdots + a_n.$$

Then ϕ is a ring homomorphism.

Example 1.12. Let $\phi : \mathbb{Z}[x] \to \mathbb{R}$ given by

$$\phi(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0 + a_1\sqrt{2} + \cdots + a_n\sqrt{2}^n.$$

Then ϕ is a ring homomorphism.

2. Three Isomorphism Theorems

In MT2505 you will have met the First Isomorphism Theorem for groups, and if you have already taken honours group theory you will also have seen the Second and Third Isomorphism Theorems. The theorems for rings are essentially identical, and the proofs are also very similar.

Theorem 2.1 (First Isomorphism Theorem) *Let R and S be rings and let $f : R \to S$ be a ring homomorphism. Then*

$$R/\ker(f) \cong \mathrm{im}\,(f).$$

Proof. We prove this by defining a map

$$\phi : R/\ker(f) \to \mathrm{im}\,(f), \quad a + \ker(f) \mapsto f(a),$$

and showing that ϕ is an isomorphism of rings.

Well-defined: If $a+\ker(f) = b+\ker(f)$, then $a-b \in \ker(f)$ and thus $0 = f(a-b) = f(a)-f(b)$, because f is a homomorphism. Hence $f(a) = f(b)$, and so $\phi(a + \ker(f)) = f(a) = f(b) = \phi(b + \ker(f))$.

Homomorphism: Let $a + \ker(f), b + \ker(f) \in R/\ker(f)$. Then

$$\begin{aligned}\phi(a + \ker(f)) + \phi(b + \ker(f)) = f(a) + f(b) = f(a + b) = \phi((a + b) + \ker(f)) \\ = \phi((a + \ker(f)) + (b + \ker(f)))\end{aligned}$$

and

$$\phi(a+\ker(f))*\phi(b+\ker(f)) = f(a)*f(b) = f(a*b) = \phi((a*b)+\ker(f)) = \phi((a+\ker(f))*(b+\ker(f))).$$

Surjective: If $y \in \mathrm{im}\,(f)$, then there exists $x \in R$ such that $f(x) = y$. It follows that $y = f(x) = \phi(x + \ker(f))$ and so ϕ is surjective.

Injective: It is straightforward to prove this directly. Alternatively, $\phi(x + \ker(f)) = f(x) = 0$ if and only if $x \in \ker(f)$. Hence $\ker(\phi) = \{0 + \ker(f)\}$. But $0 + \ker(f)$ is the zero of $R/\ker(f)$ and so, by Lemma 1.5 (ii), ϕ is injective. ■

There is a close relationship between ideals and kernels of homomorphisms.

Lemma 2.2. *Let R be a ring.*

(i) *Let I be an ideal of R. Then $f_I : R \to R/I, \quad f_I(r) = r + I$ is a homomorphism, with kernel I. Hence every ideal I of a ring is the kernel of a homomorphism.*

(ii) *Every subring of R is the image of a homomorphism.*

Proof. (i). To verify that f is a homomorphism, we check:

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

for all $a, b \in R$. Moreover,

$$\ker(f_I) = \{r \in R \ : \ f_I(r) = I\} = \{r \in R \ : \ r \in I\} = I.$$

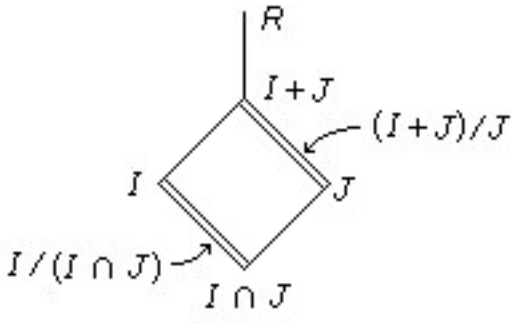
(ii). Let S be a subring of R. Then the map $\iota_S : S \to R, \ s \mapsto s$ is a ring homomorphism. ■

Definition 2.3. The homomorphism f_I in Lemma 2.2 is the *natural homomorphism* from R to R/I. The homomorphism ι_S is called the *inclusion map* from S to R.

Theorem 2.4 (Second Isomorphism Theorem) *Let I and J be ideals of a ring R. Then*

$$I/(I \cap J) \cong (I + J)/J.$$

Proof. We first note that by Chapter 4, Lemma 2.12, the set $I + J := \{i + j \mid i \in I, j \in J\}$ is a subring of R and $I \cap J$ is an ideal in I. Here is a picture showing the inclusions. The double lines represent the two factor rings.



To prove the result we will use the First Isomorphism Theorem (Theorem 2.1). Define

$$f : I \to (I + J)/J, \quad i \mapsto i + J.$$

Homomorphism: The map f is the composition of the inclusion homomorphism $\iota_I : I \to I + J$ with the natural map $f_J : I + J \to (I + J)/J$ and so is a ring homomorphism by Lemma 1.5 (iv).
Surjective: Consider an arbitrary element $i + j \in I + J$, that is, choose any $i \in I$ and $j \in J$. Since $(i + j) + J = i + J$, it follows that $f(i) = i + J = (i + j) + J$.
Kernel: An element $i \in I$ is mapped to $0 + J$ by f if and only if it lies in J and therefore in $I \cap J$, so $\ker(f) = I \cap J$.

It follows by the First Isomorphism Theorem (Theorem 2.1) that

$$\begin{aligned} I/\ker(f) &\cong \operatorname{im}(f) \\ I/(I \cap J) &\cong (I + J)/J. \end{aligned}$$

■

Example 2.5. Let $m, n \in \mathbb{Z}$ and consider the principal ideals (m) and (n). Then

$$(m) + (n) = (\gcd(m, n))$$

and

$$(m) \cap (n) = (\operatorname{lcm}(m, n)).$$

From the Second Isomorphism Theorem (Theorem 2.4) we deduce that

$$(\gcd(m, n))/(n) \cong (m)/(\operatorname{lcm}(m, n)).$$

These two rings are finite, so they have equal orders, and hence

$$n/\gcd(m, n) = \operatorname{lcm}(m, n)/m.$$

We conclude that

$$\gcd(m, n)\ \operatorname{lcm}(m, n) = mn.$$

Theorem 2.6 (Third Isomorphism Theorem) *Let I and J be ideals of a ring R with $I \subseteq J$. Then J/I is an ideal of R/I and*

$$(R/I)/(J/I) \cong R/J.$$

Proof. First we show that J/I is an ideal of R/I. Let $j_1 + I$, $j_2 + I \in J/I$. Then

$$(j_1 + I) - (j_2 + I) = (j_1 - j_2) + I$$

so J/I is closed under subtraction. Let $j + I \in J/I$ and $r + I \in R/I$. Then $j * r, r * j \in J$ since J is an ideal. It follows that

$$(j + I) * (r + I) = (j * r) + I \in J/I \text{ and } (r + I) * (j + I) = (r * j) + I.$$

Thus J/I is an ideal of R/I.

We will now use the First Isomorphism Theorem (Theorem 2.1) to prove that $(R/I)/(J/I) \cong R/J$. Define $\phi : R/I \to R/J$ by $a + I \to a + J$ for any coset $a + I$ of I in R.
Well-defined: Let $a, b \in R$ such that $a + I = b + I$. Then $a - b \in I$ and we assumed that $I \subseteq J$, so $a - b \in J$, and so $a + J = b + J$. Hence

$$\phi(a + I) = a + J = b + J = \phi(b + I).$$

Homomorphism: We check

$$\phi(a + I) + \phi(b + I) = (a + J) + (b + J) = (a + b) + J = \phi((a + I) + (b + I)),$$

and

$$\phi(a + I) * \phi(b + I) = (a + J) * (b + J) = ab + J = \phi(ab + I) = \phi((a + I) * (b + I)).$$

Kernel: An element $a + I$ is in the kernel of ϕ if and only if $\phi(a + I) = J$, which happens if and only if $a \in J$. Thus $\ker(\phi) = J/I$.
Surjective: Let $a + J \in R/J$. Then $a + J = \phi(a + I)$, so ϕ is surjective,

Thus by the First Isomorphism Theorem (Theorem 2.1)

$$\begin{aligned} (R/I)/\ker(\phi) &\cong \operatorname{im}(\phi) \\ (R/I)/(J/I) &\cong R/J. \end{aligned}$$

■

Chapter 6

Factorisation

In this part of the course we will study the notion of factorisation in rings, as a generalisation of the unique factorisation of integers into prime powers.

The rings we will consider in this chapter will all be integral domains: so all our rings are commutative, have a (non-zero) multiplicative identity, and have no zero divisors. From now on, we will write R^* to denote the set of all units in an integral domain R.

1. Divisibility and associates

Definition 1.1. Let R be an integral domain and let $a, b \in R$. If there exists $c \in R$ such that $a = bc$ then b *divides* a. We denote this by $b \mid a$.

Example 1.2. In the ring $\mathbb{Z}$, this definition of division exactly matches the standard notion of divisibility of integers. In the ring $\mathbb{Z}[x]$, we can write $x^2 - 1 = (x-1)(x+1)$, so both $(x+1)$ and $(x-1)$ are divisors of $x^2 - 1$.

Example 1.3. Let R be an integral domain, and let $a \in R$. Then $a = a1$, so a divides a and 1 divides a. If $a \mid b$ and $b \mid c$, then $b = ax$ and $c = by$ for some $x, y \in R$. Hence $c = by = axy$, so a divides c. Also note that if $0 \mid a$, then $a = 0x = 0$, so the only element that is divisible by 0 is 0 itself.

Definition 1.4. Let R be an integral domain. Elements $a, b \in R$ are *associates* if $a \mid b$ and $b \mid a$. We denote this $a \sim b$.

Example 1.5. Every element is an associate of itself. The only associate of 0 is 0 itself, by Example 1.3.

Lemma 1.6. *Let R be an integral domain, and let $a, b, c \in R$. If $ac = bc$ and $c \neq 0$, then $a = b$.*

Proof. Since $ac = bc$, we deduce that $ac - bc = (a-b)c = 0$. But R is an integral domain and $c \neq 0$, so $a - b = 0$. ■

Hence, although not all elements of an integral domain are units, we can normally cancel elements that occur on both sides of an equation.

Lemma 1.7. *Let $a, b \in R$.*

(i) If $a \mid b$ and b is a unit, then a is a unit. In other words, the only divisors of units are units.

(ii) $a \sim b$ if and only if $a = bu$ for some unit $u \in R^*$.

(iii) Let $u_1, u_2 \in R^*$. Then $u_1 \sim u_2$.

Proof. (i) b is a unit if and only if there exists $c \in R$ such that $bc = 1$, which happens if and only if $b \mid 1$. Hence if $a \mid b$ and b is a unit, then $a \mid 1$ and so a is a unit.

(ii) ($\Rightarrow$) First notice that $a = 0$ if and only if $b = 0$, in which case $a = b * 1$.

If $a, b \neq 0$, then as $a \sim b$, $a \mid b$ and $b \mid a$. Hence there exist $c, d \in R$ such that $b = ac$ and $a = bd$. Thus $a = acd$ and so by Lemma 1.6, $cd = 1$. Therefore both c and d are units.

($\Leftarrow$) From $a = bu$, we see that $b \mid a$. Also, $b = au^{-1}$ and so $a \mid b$.

(iii) Since u_1 and u_2 are units, there exist $u_1^{-1}, u_2^{-1} \in R$. So $u_2 = u_1(u_1^{-1}u_2)$, and $u_1 = u_2(u_2^{-1}u_1)$, and hence $u_1 \mid u_2$ and $u_2 \mid u_1$, as required. ■

Example 1.8. The units in $\mathbb{Z}$ are ± 1, so the associates of $a \in \mathbb{Z}$ are a and $-a$. Similarly, the units in $\mathbb{Z}[x]$ are the constant polynomials ± 1, so the associates of $\sum a_i x^i$ are itself and $\sum(-a_i)x^i$.

Lemma 1.9. *Let R be an integral domain and let $a, b \in R$ be arbitrary. Then the following hold:*

(i) *$b \mid a$ if and only if $(a) \subseteq (b)$;*

(ii) *$(a) = (b)$ if and only if $a \sim b$.*

Proof. (i) $b \mid a$ if and only if there exists $x \in R$ such that $a = bx$, which happens if and only if $a \in (b)$. By Chapter 4, Lemma 2.12(ii), this happens if and only if $(a) \subseteq (b)$.

(ii) $a \sim b$ if and only if $a \mid b$ and $b \mid a$. By Part (i) this happens if and only if $(a) \subseteq (b)$ and $(b) \subseteq (a)$, that is, if and only if $(a) = (b)$. ■

2. Primes and irreducibles

Definition 2.1. Let R be an integral domain, and let $a \in R$. We say that a is *irreducible* if all of the following hold:

(i) $a \neq 0$;

(ii) $a \notin R^*$;

(iii) $a = bc$ for $b, c \in R$ implies that b or c is a unit.

If $x \in R$ is irreducible and $u \in R^*$ is a unit, then xu is also irreducible: see Tutorial Sheet 5.

Example 2.2. If $p \in \mathbb{Z}$ is a prime number and $p = ab$ for some $a, b \in \mathbb{Z}$, then $a = \pm 1$ and $b = \pm p$ or vice versa. Since 1 and -1 are units, it follows that p is irreducible.

Conversely, suppose that $x \in \mathbb{Z}$ is irreducible. If $x = ab$ for some $a, b \in \mathbb{Z}$, then a or b is a unit and so $a = \pm 1$ or $b = \pm 1$. Hence $|x|$ is a prime number.

We have shown that the irreducible elements of $\mathbb{Z}$ are precisely the prime numbers and their negatives.

Definition 2.3. Let R be an integral domain, and let $p \in R$. We say that p is *prime* if all of the following hold:

(i) $p \neq 0$;

(ii) $p \notin R^*$;

(iii) if $p \mid ab$ for some $a, b \in R$, then $p \mid a$ or $p \mid b$.

Example 2.4. Let $p \in \mathbb{Z}$ be such that $|p|$ is a prime number. Then p is a prime element of the integral domain $\mathbb{Z}$.

Conversely, let $n \in \mathbb{Z}$ be such that $|n|$ is not 0 or 1 and $|n|$ is not a prime number. Then $n = ab$ for some $a, b \in \mathbb{Z}$ with $1 < |a| < n$ and $1 < |b| < n$. So n does not divide a or b, and hence n is not a prime element of the integral domain $\mathbb{Z}$.

Lemma 2.5. *Let R be an integral domain. Then every prime element of R is irreducible.*

Proof. Let $p \in R$ be a prime element and assume that $p = ab$ for some $a, b \in R$. Since $p \neq 0$, it follows that $a \neq 0$ and $b \neq 0$. Certainly, $p \mid ab$ and so $p \mid a$ or $p \mid b$. If $p \mid a$, then $a = px$ for some $x \in R$. But then $a = abx$ and so $bx = 1$ by Lemma 1.6, and b is a unit. Similarly, if $p \mid b$, then a is a unit. Hence p is irreducible. ■

Example 2.6. The ring of integers $\mathbb{Z}$ is an integral domain. In Example 2.2 we saw that the irreducible elements of $\mathbb{Z}$ are precisely the prime numbers and their negatives, and in Example 2.4 we saw that the same is true for the prime elements of $\mathbb{Z}$.

In order to determine the irreducible elements in some rings other than $\mathbb{Z}$, we take a short detour.

Definition 2.7. Let R be a ring. A function $N : R \to \mathbb{Z}$ is called a *norm* if

$$N(ab) = N(a)N(b)$$

for all $a, b \in R$.

Lemma 2.8. *Let $n \in \mathbb{Z}$ be arbitrary and define*

$$\mathbb{Z}[\sqrt{n}] = \{a + b\sqrt{n} \ : \ a, b \in \mathbb{Z}\}.$$

Then $\mathbb{Z}[\sqrt{n}]$ is an integral domain.

If n is a non-square in $\mathbb{Z}$ (i.e. $n \neq m^2$ for all $m \in \mathbb{Z}$), then $N : \mathbb{Z}[\sqrt{n}] \longrightarrow \mathbb{Z}$ given by

$$N(a + b\sqrt{n}) = a^2 - nb^2$$

is a norm on $\mathbb{Z}[\sqrt{n}]$.

Proof. We first check that $\mathbb{Z}[\sqrt{n}]$ is a subring of the integral domain $\mathbb{C}$:

$$\begin{aligned} (a + b\sqrt{n}) - (c + d\sqrt{n}) &= (a - c) + (b - d)\sqrt{n} \in \mathbb{Z}[\sqrt{n}] \\ (a + b\sqrt{n}) * (c + d\sqrt{n}) &= (ac + nbd) + (bc + ad)\sqrt{n} \in \mathbb{Z}[\sqrt{n}], \end{aligned}$$

so $\mathbb{Z}[\sqrt{n}]$ is a subring of $\mathbb{C}$. The ring $\mathbb{Z}[\sqrt{n}]$ is commutative, because $\mathbb{C}$ is commutative. The multiplicative identity of $\mathbb{C}$ lies in $\mathbb{Z}[\sqrt{n}]$, and $\mathbb{Z}[\sqrt{n}]$ has no zero divisors, because $\mathbb{C}$ has no zero divisors. So $\mathbb{Z}[\sqrt{n}]$ is an integral domain.

See Tutorial Sheet 5 for the proof that N is a norm when n is a non-square. ■

Example 2.9. If the n in Lemma 2.8 is a square, then there is no way of writing $a+b\sqrt{n}$ uniquely. For example, in $\mathbb{Z}[\sqrt{4}]$, we can write $4 + 0\sqrt{4} = 0 + 2\sqrt{4}$. Hence $N(4 + 0\sqrt{4}) = 4^2 - 4 \cdot 0^2 = 16$ and $N(0 + 2\sqrt{4}) = 0^2 - 4 \cdot 2^2 = -16$, and so $N : \mathbb{Z}[\sqrt{n}] \longrightarrow \mathbb{Z}$ is not a function.

Example 2.10. Consider the integral domain

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \ : \ a, b \in \mathbb{Z}\}.$$

The norm on $\mathbb{Z}[\sqrt{-5}]$ is $N(a + b\sqrt{-5}) = a^2 + 5b^2$.

Let us start by seeing how the norm can be used to determine the units of $\mathbb{Z}[\sqrt{-5}]$. If $\alpha = a + b\sqrt{5} \in \mathbb{Z}[\sqrt{-5}]$ is a unit, then there exists $\beta \in \mathbb{Z}[\sqrt{-5}]$ such that $\alpha\beta = 1$, Hence

$$N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1.$$

Therefore $N(\alpha) = a^2 + 5b^2 = \pm 1$, and from $a, b \in \mathbb{Z}$ we deduce that $a = \pm 1$ and $b = 0$. Hence the only units in $\mathbb{Z}[\sqrt{-5}]$ are the "obvious" ones, namely ± 1, and these are precisely the elements of norm 1.

We will now see how the norm can be used to show that $2+\sqrt{-5}$ is irreducible in $\mathbb{Z}[\sqrt{-5}]$. If $a+b\sqrt{-5} \mid 2+\sqrt{-5}$, then there exists $c+d\sqrt{-5}$ such that $(a+b\sqrt{-5})(c+d\sqrt{-5}) = 2+\sqrt{-5}$. Hence $N(a+b\sqrt{-5}) \mid N(2+\sqrt{-5}) = 9$ and $N(c+d\sqrt{-5}) \mid 9$. Hence $N(a+b\sqrt{-5}) = 1, 3$, or 9.

If $N(a+b\sqrt{-5}) = 1$, then $a+b\sqrt{-5}$ is a unit. Similarly, if $N(a+b\sqrt{-5}) = 9$, then $N(c+d\sqrt{-5}) = 1$ and so $c+d\sqrt{-5}$ is a unit. If $N(a+b\sqrt{-5}) = 3$, then $a, b \in \mathbb{Z}$ satisfy $a^2+5b^2 = 3$. However, there are no integer solutions to this equation. Hence $2+\sqrt{-5}$ is irreducible in $\mathbb{Z}[\sqrt{-5}]$.

However, $2+\sqrt{-5}$ is not prime in $\mathbb{Z}[\sqrt{-5}]$. To see this, we calculate

$$(2+\sqrt{-5}) \mid 9 = 3 \cdot 3.$$

Notice that $N(2+\sqrt{-5}) = 9 = N(3)$, so if there exists $(c+d\sqrt{-5})$ such that $(2+\sqrt{-5})(c+d\sqrt{-5}) = 3$, then $c+d\sqrt{-5}$ is a unit. However, $3 \neq \pm 1 \cdot (2+\sqrt{-5})$, so $2+\sqrt{-5}$ does not divide 3. Hence $2+\sqrt{-5}$ is not prime.

Example 2.11. The norm on $\mathbb{Z}[\sqrt{10}]$ is defined by

$$N(a+b\sqrt{10}) = a^2 - 10b^2.$$

As in Example 2.10, we start by showing that the elements of norm 1 are units. We do so using a very general argument. If $N(a+b\sqrt{10}) = 1$, then $a^2-10b^2 = 1$. Hence $(a+b\sqrt{10})(a-b\sqrt{10}) = a^2-10b^2 = 1$, and so $a+b\sqrt{10}$ is a unit.

We will now show that 2, 3, $(4+\sqrt{10})$ and $(4-\sqrt{10})$ are all irreducible in $\mathbb{Z}[\sqrt{10}]$. Their norms are

$$N(2) = 4, \quad N(3) = 9, \quad N(4+\sqrt{10}) = N(4-\sqrt{10}) = 6,$$

so if they are divisible by any elements other than units, there must exist elements of $\mathbb{Z}[\sqrt{10}]$ of norm ± 2 or ± 3.

If $a+b\sqrt{10} \in \mathbb{Z}[\sqrt{10}]$, then

$$N(a+b\sqrt{10}) \bmod 10 = a^2 \bmod 10.$$

We calculate that, working modulo 10:

$$0^2 \equiv 0,\ (\pm 1)^2 \equiv 1,\ (\pm 2)^2 \equiv 4,\ (\pm 3)^2 \equiv 9,\ (\pm 4)^2 \equiv 6,\ 5^2 \equiv 5.$$

Hence $(x^2 \bmod 10) \in \{0,1,4,5,6,9\}$ for all $x \in \mathbb{Z}$ and so $(N(a+b\sqrt{10}) \bmod 10) \in \{0,1,4,5,6,9\}$. Thus $N(a+b\sqrt{10}) \neq \pm 2$ or ± 3. It follows that 2, 3, $(4+\sqrt{10})$, and $(4-\sqrt{10})$ are irreducible in $\mathbb{Z}[\sqrt{10}]$.

Example 2.12. Recall the Gaussian integers $\mathbb{Z}[i] = \mathbb{Z}[\sqrt{-1}]$ from Chapter 3, Example 2.17. The norm on $\mathbb{Z}[i]$ is just $N(a+bi) = a^2+b^2$.

Let p be a prime number in $\mathbb{N}$ such that there exist $a, b, c \in \mathbb{Z}$ with $\gcd(p,c) = 1$ and

$$a^2 + b^2 = cp.$$

We will show by contradiction that p is not prime in $\mathbb{Z}[i]$.

First notice that

$$cp = a^2 + b^2 = (a+bi)(a-bi),$$

and $p \mid cp$, so $p \mid (a+bi)(a-bi)$. We assumed that p is prime, so $p \mid (a+bi)$ or $p \mid (a-bi)$. If $p \mid (a+bi)$ then $a+bi = (u+vi)p$ for some $u, v \in \mathbb{Z}$, and so $a = pu$ and $b = pv$. Hence $p \mid (a-bi)$, and so $p^2 \mid (a+bi)(a-bi) = cp$, and so $p \mid c$. But we assumed that $\gcd(p,c) = 1$, giving a contradiction.

3. Factorisation and unique factorisation

Definition 3.1. Let R be an integral domain, and let $x \in R \setminus R^*$ be non-zero. Then x has a *factorisation into irreducibles* if there exist irreducible $p_1, p_2, \ldots, p_m \in R$ such that $x = p_1 p_2 \cdots p_m$. The integral domain R is a *factorisation domain* if every non-zero element in $R \setminus R^*$ has a factorisation into irreducibles.

Example 3.2. [**Every field is a factorisation domain.**] Let R be a field. Then every nonzero element of R is a unit and so $R \setminus R^* = \{0\}$. Hence it is vacuously true that every non-zero element in $R \setminus R^*$ (there are none!) has a factorisation into irreducibles. So, every field is a factorisation domain.

Before giving some more examples of factorisation domains, we state the central definition in this part of the course. In $\mathbb{Z}$, $10 = 2 \cdot 5 = (-2) \cdot (-5)$, and so strictly speaking factorisation is not unique. However, we might consider the two factorisations $2 \cdot 5 = (-2) \cdot (-5)$ as being the same since 2 and -2 are associates, and so are 5 and -5.

Definition 3.3. Let R be an integral domain and let $x \in R \setminus R^*$ be non-zero. Then x has a *unique factorisation* into irreducibles if x has a factorisation into irreducibles, and whenever

$$\begin{aligned} x &= p_1 p_2 \cdots p_m \\ &= q_1 q_2 \cdots q_n \end{aligned}$$

are two factorisations into irreducibles, there is a bijection $\phi : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that $p_i \mid q_{\phi(i)}$ for all i. That is, $n = m$ and we can reorder $q_1, \ldots, q_n$ so that $p_i \mid q_i$ for all i.

Note that if $p_i \mid q_j$, then $q_j = p_i u_i$ for some $u_i \in R$. But q_j is irreducible and so u_i is a unit. Hence by Lemma 1.7 (ii) the irreducibles p_i and q_j are associates.

Example 3.4. In the integers $\mathbb{Z}$, we can write

$$6 = 2 \cdot 3 = 3 \cdot 2 = (-2) \cdot (-3) = (-3) \cdot (-2)$$

but there is a bijection from the first factorisation $(2 \cdot 3)$ to each of the others, sending 2 to ± 2 and 3 to ± 3.

Definition 3.5. An integral domain R is a *unique factorisation domain* or *UFD* if every non-zero element of $R \setminus R^*$ has a unique factorisation into irreducibles.

Example 3.6. Every field is a unique factorisation domain, as there are no non-zero elements that are not units.

Theorem 3.7. *Let R be a factorisation domain. Then R is a unique factorisation domain if and only if every irreducible element in R is a prime.*

Proof. ($\Rightarrow$) Let R be a unique factorisation domain, and let $a \in R$ be an irreducible. We must show that a is prime, so assume that $a \mid bc$ for some $b, c \in R$. Then $ad = bc$ for some $d \in R$. We must prove that $a \mid b$ or $a \mid c$.

If $b = 0$ or $c = 0$, then $ad = bc = 0$, so a divides whichever of b or c is 0, as required. If b or c is a unit, then $c = b^{-1}ad$ or $b = adc^{-1}$, and so $a \mid c$ or $a \mid b$, respectively. If neither b nor c is a unit, then there exist unique factorisations into irreducibles for both b and c. But a is an irreducible and so, by unique factorisation, a is an associate of one of the irreducibles dividing b or c. Hence a divides b or c, as required.

($\Leftarrow$) Let $x \in R \setminus R^*$ be non-zero, and suppose that

$$x = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n$$

where $p_1, p_2, \ldots, p_m, q_1, q_2, \ldots, q_n \in R$ are irreducible. We may assume without loss of generality that $m \leq n$, Since every irreducible element of R is prime, $p_1, p_2, \ldots, p_m$ are prime.

Since $p_1 \mid q_1 \cdots q_n$ and p_1 is prime, it follows that there is $j \in \{1, \ldots, n\}$ such that $p_1 \mid q_j$. By re-ordering the q_i, we can assume that $p_1 \mid q_1$. It follows that $q_1 = p_1 u_1$ for some unit u_1 and so

$$p_1 \cdots p_m = p_1 u_1 q_2 \cdots q_n.$$

By Lemma 1.6, we can cancel the p_1 on both sides, to get

$$p_2 \cdots p_m = u_1 q_2 \cdots q_n.$$

Repeating this process for $p_2, \ldots, p_m$, we obtain

$$1 = u_1 \cdots u_m \cdot q_{m+1} \cdots q_n.$$

If $m < n$ then $q_{m+1}, \ldots, q_n \in R$ are units, contradicting the fact that they are irreducible. Therefore $m = n$, and we were able to reorder the q_i so that $p_i \mid q_i$ for all i, as required. ■

Example 3.8. The ring of integers $\mathbb{Z}$ is a unique factorisation domain, since every irreducible element of $\mathbb{Z}$ is prime, by Example 2.6.

Example 3.9. We saw in Example 2.10 that the element $2 + \sqrt{-5}$ is irreducible in

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \ : \ a, b \in \mathbb{Z}\},$$

and that it is not prime. Hence Theorem 3.7 shows that $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorisation domain.

Another way to see this is as follows: similar arguments to Example 2.10 show that $2 - \sqrt{-5}$ and 3 are irreducible in $\mathbb{Z}[\sqrt{-5}]$ since $N(2 - \sqrt{-5}) = N(3) = 9$. Hence

$$9 = (2 + \sqrt{-5})(2 - \sqrt{-5}) = 3 \cdot 3$$

has two essentially different factorisations into irreducibles.

Example 3.10. We saw in Example 2.11 that the elements 2, 3, $4 + \sqrt{10}$ and $4 - \sqrt{10}$ are irreducible in $\mathbb{Z}[\sqrt{10}]$. Now,

$$6 = 2 \cdot 3 = (4 + \sqrt{10})(4 - \sqrt{10})$$

but $2 \nmid (4 + \sqrt{10})$ and $2 \nmid (4 - \sqrt{10})$ as $N(2) = 4 \nmid 6 = N(4 + \sqrt{10}) = N(4 - \sqrt{10})$. Hence $\mathbb{Z}[\sqrt{10}]$ is not a unique factorisation domain.

Chapter 7

Properties of unique factorisation domains

1. Principal ideal domains and unique factorisation domains

In Chapter 4, Definition 4.1, we defined a *principal ideal domain* to be an integral domain where every ideal is principal. In Chapter 4, Lemma 4.2, we showed that the ring of integers $\mathbb{Z}$ is a principal ideal domain. We also showed in Chapter 6, Example 3.8 that $\mathbb{Z}$ is a unique factorisation domain. Motivated by the archetypal example of the integers, in this section, we will show that every principal ideal domain is a unique factorisation domain.

Certainly every unique factorisation domain is a factorisation domain and so, as a first step to proving that every principal ideal domain is a unique factorisation domain, we show that every principal ideal domain is a factorisation domain.

We require the following two slightly technical results to prove that every principal ideal domain is a factorisation domain.

Lemma 1.1. *Let $I_1, I_2, \ldots$ be ideals of a ring R, such that $I_1 \subseteq I_2 \subseteq \cdots$. Then*

$$I = \bigcup_{n=1}^{\infty} I_n$$

is an ideal.

Proof. See Tutorial Sheet 6. ■

Proposition 1.2. *Let R be a principal ideal domain. Then for any ideals $I_1, I_2, \ldots$ in R such that $I_1 \subseteq I_2 \subseteq \cdots$ there exists $N \in \mathbb{N}$ such that $I_N = I_{N+n}$ for all $n \in \mathbb{N}$. That is, the sequence $I_1, I_2, \ldots$ is eventually constant.*

Proof. Let R be a principal ideal domain and let $I_1, I_2 \ldots$ be ideals in R such that

$$I_1 \subseteq I_2 \subseteq \cdots .$$

By Lemma 1.1, the union of an ascending chain of ideals is an ideal and so

$$I = \bigcup_{i=1}^{\infty} I_i$$

is an ideal of R. Since R is a principal ideal domain, it follows that $I = (d)$ for some $d \in R$. From the definition of I, there exists $N \in \mathbb{N}$ such that $d \in I_N$. Thus $I = (d) \subseteq I_m \subseteq I$ for all $m \geq N$ and so $I_N = I_{N+1} = I_{N+2} = \cdots = I$. ■

Example 1.3. The ring of integers, $\mathbb{Z}$ is a principal ideal domain. Let $(n_1) \subseteq (n_2) \subseteq (n_3) \subseteq \cdots$ be ideals in $\mathbb{Z}$, and assume without loss of generality that $n_1, n_2, \ldots \geq 0$. Since $(n_1) \subseteq (n_2)$, we see that $n_1 \in (n_2)$, so $n_2 \mid n_1$. In particular, $n_2 \leq n_1$, and in general $n_{i+1} \leq n_i$ for all i. Hence the sequence of non-negative integers $n_1 \geq n_2 \geq \ldots$ must contain a minimal element n_N such that $n_m = n_N$ for all $m \geq N$. Hence $(n_m) = (n_N)$ for all $m \geq N$.

These two technical results enable us to prove that every principal ideal domain is a factorisation domain.

Theorem 1.4. *Let R be a principal ideal domain. Then R is a factorisation domain.*

Proof. Let $r_0 \in R$ be such that $r_0 \neq 0$ and r_0 is not a unit. We must prove that r_0 has a factorisation into irreducibles.

Assume the contrary. Then r_0 is not itself an irreducible, and so there exist nonzero $r_1, s_1 \in R \setminus R^*$ such that $r_0 = r_1 s_1$. Thus $r_1 \mid r_0$ and so $r_0 \in (r_1)$. Hence by Chapter 4, Lemma 2.12(ii), $(r_0) \subseteq (r_1)$. If $(r_0) = (r_1)$, then $r_1 \in (r_0)$, and so $r_0 \mid r_1$. Hence $r_0 \sim r_1$, and so $r_1 = r_0 u$ for some unit u, by Chapter 6, Lemma 1.7(ii). But then $r_0 = r_1 s_1 = r_0 u s_1$ and by Chapter 6, Lemma 1.6, since R is an integral domain, we can cancel the r_0 on each side to get $us_1 = 1$. In other words, $s_1 \in R^*$, a contradiction. Hence $(r_0) \subsetneq (r_1)$. By an identical argument, $(r_0) \subsetneq (s_1)$.

If both r_1 and s_1 have factorisations into irreducibles, then so does r_0, which contradicts our assumption that r_0 has no such factorisation. So without loss of generality we may assume that r_1 has no factorisation into irreducibles. Then repeating the above argument, but replacing r_0 with r_1, we find $r_2, s_2 \in R \setminus R^*$ such that $r_1 = r_2 s_2$, $(r_1) \subsetneq (r_2)$, and $(r_1) \subsetneq (s_2)$. Again not both r_2 and s_2 can have factorisations into irreducibles, and so without loss of generality r_2 has no such factorisation.

Continuing in this way we obtain

$$(r_0) \subsetneq (r_1) \subsetneq \cdots$$

with $(r_i) \neq (r_{i+1})$ for all $i \in \mathbb{N}$. This contradicts Proposition 1.2, and so r_0 must have a factorisation into irreducibles. ■

The next theorem relates the notions of maximal ideals and irreducible elements of a principal ideal domain.

Theorem 1.5. *Let R be a principal ideal domain that is not a field and let $r \in R$. Then the ideal (r) is maximal if and only if r is an irreducible element of R.*

Proof. ($\Rightarrow$) Assume that (r) is maximal, and that $r = ab$ for some $a, b \in R$. We must show that a or b is a unit. From $r = ab$ we deduce that $r \in (a)$ and so $(r) \subseteq (a)$. But (r) is maximal and so either $(a) = R$ or $(a) = (r)$. If $(a) = R$ then $1 \in (a)$, and so a is a unit and we are done. If $(r) = (a)$ then $r \mid a$ and $a \mid r$, so there is a unit c such that $a = rc$. Hence $r = ab = rcb$. Now by Chapter 6, Lemma 1.6, since R is an integral domain we can cancel r to get $1 = cb$. In other words, b is a unit, and we are done.

($\Leftarrow$) Assume that r is irreducible, and let $s \in R$ be such that $(r) \subseteq (s)$. We must show that $(s) = (r)$ or $(s) = R$. From $(r) \subseteq (s)$ we deduce that $r = sx$ for some $x \in R$. But r is irreducible and so either s or x is a unit. If s is a unit, then $(s) = R$. If x is a unit, then $s = rx^{-1}$, so $s \in (r)$ and hence $(s) = (r)$. Hence (r) is a maximal ideal. ■

Example 1.6. If R is a field, then $(0) = \{0\}$ is a maximal ideal but 0 is not irreducible, so the conclusion of Theorem 1.5 does not hold.

Example 1.7. The irreducible elements in $\mathbb{Z}$ are the prime numbers and their negatives. We saw in Chapter 4, Lemma 4.9 that the maximal ideals of $\mathbb{Z}$ are exactly the ideals (p) where p or $-p$ is a prime number.

Corollary 1.8. *Let R be a principal ideal domain that is not a field. Then an ideal I in R is maximal if and only if I is a prime ideal and $I \neq \{0\}$.*

Proof. ($\Rightarrow$) Let R be a principal ideal domain that is not a field, and assume that I is a maximal ideal of R. By Chapter 4, Corollary 4.7, every maximal ideal of R is a prime ideal, so we only need to show that $I \neq \{0\}$.

Seeking a contradiction, suppose that $I = \{0\}$. Then for all $r \in R$ such that $r \neq 0$, it follows that $I = \{0\} \subsetneq (r) \subseteq R$. The fact that I is maximal then implies that $(r) = R$. But this in turn implies that r is a unit. This holds for all nonzero $r \in R$, and so R is a field, a contradiction. Thus $I \neq \{0\}$.

($\Leftarrow$) Let (r) be a prime ideal of R with $(r) \neq \{0\}$. We must show that (r) is maximal. If r is a unit, then $(r) = R$, so r is not a unit. By Theorem 1.4 the ring R is a factorisation domain, so it follows that there exist irreducibles $x_1, x_2, \ldots, x_n \in R$ such that

$$r = x_1 x_2 \cdots x_n.$$

But (r) is prime and so $x_i \in (r)$ for some i. Now, by Theorem 1.5, (x_i) is maximal in R. But $(x_i) \subseteq (r)$, and (r) is prime, so $(r) \neq R$. Hence $(x_i) = (r)$ is a maximal ideal, as required. ■

We are now ready to prove the main result of this section.

Theorem 1.9. *Let R be a principal ideal domain. Then R is a unique factorisation domain.*

Proof. We proved in Theorem 1.4 that R is a factorisation domain. Hence it suffices by Chapter 6, Theorem 3.7 to show that every irreducible element in R is prime.

Let $r \in R$ be irreducible, and suppose that $r \mid ab$ for some $a, b \in R$. We must show that $r \mid a$ or $r \mid b$. From $r \mid ab$, we deduce that $rs = ab$ for some $s \in R$, and so $ab \in (r)$. But (r) is a maximal ideal by Theorem 1.5 and hence a prime ideal by Corollary 1.8. It follows that either $a \in (r)$ or $b \in (r)$. Therefore $r \mid a$ or $r \mid b$ and so r is prime. ■

The converse of Theorem 1.9 is not true, as we will show in Theorem **??**: there exist unique factorisation domains that are *not* principal ideal domains.

Example 1.10. [**A non-principal ideal in $\mathbb{Z}[\sqrt{-5}]$.**] In Chapter 6, Examples 2.10 and 2.11, we showed that $\mathbb{Z}[\sqrt{-5}]$ and $\mathbb{Z}[\sqrt{10}]$ are not unique factorisation domains. Hence they are not principal ideal domains by Theorem 1.9.

We can also show that $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain by explicitly showing that $I := (2, 1 + \sqrt{-5})$ is not a principal ideal.

An arbitrary element of I is of the form $2\alpha + (1 + \sqrt{-5})\beta$, where $\alpha, \beta \in \mathbb{Z}[\sqrt{-5}]$. Let $\alpha = w + x\sqrt{-5}$ and $\beta = y + z\sqrt{-5}$, where $w, x, y, z \in \mathbb{Z}$. Then rearranging, we see that an arbitrary element of I is of the form:

$$\begin{aligned} 2(w + x\sqrt{-5}) + (1 + \sqrt{-5})(y + z\sqrt{-5}) &= 2w + y - 5z + (2x + y + z)\sqrt{-5} \\ &= 2(w - x - 3z) + (2x + y + z) + (2x + y + z)\sqrt{-5}. \end{aligned}$$

Substituting $a = w - x - 3z$ and $b = 2x + y + z$, we obtain

$$2\alpha + (1 + \sqrt{-5})\beta = (2a + b) + b\sqrt{-5}.$$

Hence

$$I = \{(2a + b) + b\sqrt{-5} \ : \ a, b \in \mathbb{Z}\}.$$

In particular, if $(2a+b)+b\sqrt{-5} \in \mathbb{Z}$, for integers a and b, then $b = 0$. Hence the only integers in I are even, so $1 \notin I$. Hence $I \neq \mathbb{Z}[\sqrt{-5}]$. Suppose that $I = (r)$ for some $r = s + t\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$. Then since $I \neq \mathbb{Z}[\sqrt{-5}]$, it follows that r is not a unit and so

$$N(r) = s^2 + 5t^2 > 1.$$

Since r divides every element of I, and by definition $2 \in I$, we see that r divides 2. Hence $N(r) \mid N(2) = 4$.

The positive integer divisors of 4 are $1, 2$, and 4. However, $N(r) \neq 1$ since r is not a unit, and $s^2 + 5t^2 = 2$ has no integer solutions. Hence $N(r) = 4$ and so $r = \pm 2$, and $I = (2)$. But $N(2) = 4 \nmid 6 = N(1 + \sqrt{-5})$ and so $2 \nmid (1 + \sqrt{-5})$. Therefore $1 + \sqrt{-5} \notin (2) = I = (2, 1 + \sqrt{-5})$, a contradiction. Hence I is not a principal ideal.

2. Greatest common divisors in integral domains

One consequence of unique factorisation in the integers is that every pair of integers has a greatest common divisor. In this section, we explore the notion of greatest common divisors in integral domains.

Definition 2.1. An element d of an integral domain R is said to be a *greatest common divisor* of $a, b \in R$ if

(i) $d \mid a$ and $d \mid b$; and

(ii) if $c \mid a$ and $c \mid b$, then $c \mid d$.

In general, a and b may have none, one or more than one greatest common divisor; we write $\gcd(a, b)$ to denote *any* of the greatest common divisors of a and b, if one exists.

Example 2.2. Let $R = \mathbb{Z}$, and let $a, b \in \mathbb{Z}$. Then the greatest common divisors of a and b are $\pm d$, where d is the greatest common divisor of a and b as defined in Chapter 2.

Example 2.3. Let $R = \mathbb{Z}[i]$, the Gaussian integers. We shall show that $1+i$ is a greatest common divisor of $2+2i$ and $1+3i$. Notice that $2+2i = 2(1+i)$ and $1+3i = (1+i)(2+i)$, so $1+i$ is a common divisor of both $2+2i$ and $1+2i$.

We calculate $N(2+2i) = 2^2+2^2 = 8$, so if $c \in \mathbb{Z}[i]$ divides $2+2i$ then c has norm $1, 2, 4$ or 8. Similarly, $N(1+2i) = 1^2+3^2 = 10$, so if c divides $1+3i$ then c has norm 1, 2, 5 or 10. Hence if $c \in \mathbb{Z}[i]$ divides both $2+2i$ and $1+2i$, then $N(c) = 1$ or 2. If $N(c) = 1$ then c is a unit, so $c \mid 1+i$. If $N(c) = 2$ then $c = \pm 1 \pm i$, and so c is an associate of $1+i$. Hence in particular $c \mid 1+i$, and so $1+i$ is a greatest common divisor of $2+2i$ and $1+3i$.

Lemma 2.4. *Let R be an integral domain and let $a, b \in R$.*

(i) *Let $d = \gcd(a, b)$. An element $d' \in R$ is also a greatest common divisor of a and b if and only if $d' \sim d$.*

(ii) *Let $a \in R^*$ and $b \in R$. An element $d \in R$ is a greatest common divisor of a and b if and only if d is a unit.*

(iii) *An element $d \in R$ is a greatest common divisor of a and 0 if and only if $d \sim a$.*

Proof. (i) ($\Rightarrow$) Let d' be a greatest common divisor of a and b. Then $d \mid a$ and $d \mid b$ implies that $d \mid d'$. Similarly, $d' \mid a$ and $d' \mid b$ so $d' \mid d$. Hence $d \sim d'$.

($\Leftarrow$) If $d' \sim d$ then in particular $d' \mid d$. Hence from $d \mid a$ and $d \mid b$ we deduce that $d' \mid a$ and $d' \mid b$. Let x be a divisor of a and b. Then $x \mid d$, so $xy = d$ for some $y \in R$. From $d' \sim d$ we deduce that $d = d'u$ for some unit u, so $xy = d = d'u$, and $x(yu^{-1}) = d'$. Hence $x \mid d'$, and so d' is a greatest common divisor of a and b.

(ii) ($\Rightarrow$) Let $d = \gcd(a, b)$. Then $d \mid a$ and a is a unit, so d is a unit.

($\Leftarrow$). Let d be a unit. Then $d \mid a$ and $d \mid b$. If $d' \mid a$, then d' is a unit, and so $d' \mid d$. Hence $d = \gcd(a, b)$.

(iii) $a \mid a$ and $a \mid 0$, so a is a common divisor of a and 0. If $c \mid a$ and $c \mid 0$, then in particular $c \mid a$, so a is a greatest common divisor of a and 0. By Part (i), all other greatest common divisors are associates of a. ■

An integral domain R that is not a unique factorisation domain always contains $a, b \in R \setminus R^*$ such that a and b have no greatest common divisor; more details follow in Theorem 2.8.

Example 2.5. [**An integral domain where greatest common divisors do not always exist.**] We will show that in $\mathbb{Z}[\sqrt{-3}]$ the greatest common divisor of $2+2\sqrt{-3}$ and 4 does not exist.

First, let's find the units. The norm is $N(a+b\sqrt{-3}) = a^2+3b^2$, so all norms are non-negative, and the units are the elements of norm 1. That is, the units are 1 and -1. It follows that if $\alpha, \beta \in \mathbb{Z}[\sqrt{-3}]$ satisfy both $N(\alpha) = N(\beta)$ and $\beta \mid \alpha$, then $\alpha = \pm\beta$.

We will now show that there is no greatest common divisor of $2+2\sqrt{-3}$ and 4. Note that

$$N(4) = 16 = 2^2 + 3 \cdot 2^2 = N(2+2\sqrt{-3})$$

and so any factor $a+b\sqrt{-3}$ of 4 or $2+2\sqrt{-3}$ has norm that divides 16. The divisors of 16 are $1, 2, 4, 8$, and 16. Since $a^2+3b^2 = 2$ or 8 has no solutions $a, b \in \mathbb{Z}$, we only have to consider the cases when $a^2+3b^2 = 1, 4$, and 16.

$\mathbf{a^2+3b^2=1}$. In this case, $a = \pm 1$ and $b = 0$.

$\mathbf{a^2+3b^2=4}$. In this case, either $a = \pm 1$ and $b = \pm 1$, or $a = \pm 2$ and $b = 0$.

$\mathbf{a^2+3b^2=16}$. If $b > 2$, then $a^2+3b^2 \geq a^2+27 > 16$, so $-2 \leq b \leq 2$. If $b = 0$, then $a^2 = 16$ and so $a = \pm 4$. If $b = \pm 1$, then $a^2+3b^2 = a^2+3 = 16$ and so $a^2 = 13$, which is not possible. If $b = \pm 2$, then $a^2+12 = 16$ and so $a = \pm 2$.

It follows that the only potential common divisors of 4 and $2+2\sqrt{-3}$ are:

$$\pm 1, \quad \pm 2, \quad \pm 1 \pm \sqrt{-3}, \quad \pm 4, \quad \pm 2 \pm 2\sqrt{-3}.$$

Hence, if $d = \gcd(4, 2+2\sqrt{-3})$ exists, then d must be one of these.

$\mathbf{d \neq \pm 1, \pm 2}$. Since $(1+\sqrt{-3}) \mid 4$ and $(1+\sqrt{-3}) \mid (2+2\sqrt{-3})$ but $(1+\sqrt{-3}) \nmid \pm 1$ and $(1+\sqrt{-3}) \nmid \pm 2$, it follows that $d \neq \pm 1, \pm 2$.

$\mathbf{d \neq \pm 1 \pm \sqrt{-3}}$. Since $2 \mid 4$ and $2 \mid (2+2\sqrt{-3})$ but $2 \nmid (\pm 1 \pm \sqrt{-3})$, it follows that $d \neq \pm 1 \pm \sqrt{-3}$.

$\mathbf{d \neq \pm 4, \pm 2 \pm 2\sqrt{-3}}$. Since $N(4) = N(2+2\sqrt{-3}) = 16$, it follows that $4 \nmid (2+2\sqrt{-3})$ and $(2+2\sqrt{-3}) \nmid 4$ and so $d \neq \pm 4, \pm 2 \pm 2\sqrt{-3}$.

Hence $2+2\sqrt{-3}$ and 4 have no greatest common divisor.

In the next proof, we see one way to calculate greatest common divisors.

Theorem 2.6. *Let R be a unique factorisation domain and let $a, b \in R$. Then a and b have a greatest common divisor.*

Proof. If either of a and b are units, then by Lemma 2.4(ii) their greateast common divisors are the units; so assume that neither of them are units. Similarly, if either is 0, then by Lemma 2.4(iii) their greatest common divisors are the associates of the other one; so assume that neither of them are zero.

Since a and b are non-zero and not units, they both have factorisations into irreducibles. By taking the original factorisations, and multiplying the irreducibles in the factorisation by units, if required, a and b can be written as follows

$$a = u_1 x_1^{m_1} x_2^{m_2} \cdots x_k^{m_k}$$

$$b = u_2 x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}.$$

Here the u_i are units, the m_i and n_i are non-negative integers (possibly 0), and if $i \neq j$ then x_i and x_j are not associates of one another.

Consider the element

$$d = x_1^{p_1} x_2^{p_2} \cdots x_k^{p_k}$$

where $p_i = \min(m_i, n_i)$ for all i. It is clear that $d \mid a$ and $d \mid b$, so d is a common divisor of both a and b.

Let $c \in R$ be such that $c \mid a$. Then $cs = a$, for some $s \in R$, and so

$$cs = u_1 x_1^{m_1} x_2^{m_2} \cdots x_k^{m_k}$$

By Chapter 6, Theorem 3.7, the fact that R is a unique factorisation domain implies that every irreducible element is prime. In particularly the x_i are prime, so each x_i divides c or s. Hence

$$c = u_3 x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}$$

where $r_i \leq m_i$ for all i. Similarly, if c divides b then $r_i \leq n_i$ for all i.

Thus if c divides both a and b then $r_i \leq \min(m_i, n_i)$ for all i, and so c divides d. Thus d is a greatest common divisor of a and b. ■

However, except for very small examples, even in the integers it exceeds human patience, and soon after that the ability of a computer, to use this method to find greatest common divisors. In the integers we use the division algorithm and the euclidean algorithm instead. The euclidean algorithm does not necessarily exist in an arbitrary integral domain: we shall explore this in the next chapter.

Lemma 2.7. *Let R be an integral domain and let $a, b, c \in R \setminus \{0\}$, such that d is a greatest common divisor of a and b.*

(i) *If $\gcd(ca, cb)$ exists then $c \cdot \gcd(a, b) \sim \gcd(ca, cb)$.*

(ii) *If the following greatest common divisors all exist, then*

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

(iii) *Let $x \in R$ be irreducible and assume $\gcd(a, x)$ exists. Then $x \nmid a$ if and only if $\gcd(a, x)$ is a unit.*

(iv) *If $a \sim c$, then $d = \gcd(c, b)$.*

Proof. **(i).** Assume that $\gcd(ca, cb)$ exists. From $d \mid a$ and $d \mid b$, we deduce that $cd \mid ca$ and $cd \mid cb$. Hence $cd \mid \gcd(ca, cb)$. In particular, $\gcd(ca, cb) = cdx$ for some $x \in R$. But then $cdx \mid ca$, and so $cdxy = ca$, for some $y \in R$. Since R is an integral domain, by Chapter 6, Lemma 1.6 we can cancel the c on each side to see that $dxy = a$. Hence $dx \mid a$. By an identical argument, we can deduce that $dx \mid b$. Hence $dx \mid \gcd(a, b) = d$, and so x is a unit. Hence $\gcd(ca, cb) = cdx \sim cd = c \cdot \gcd(a, b)$.

(ii). Let $x = \gcd(\gcd(a, b), c)$ and $y = \gcd(a, \gcd(b, c))$. We show that $x \mid y$: the argument that $y \mid x$ is very similar.

Since $x = \gcd(\gcd(a, b), c)$, we see that $x \mid \gcd(a, b)$, and hence $x \mid a$ and $x \mid b$. In addition, from $x = \gcd((a, b), c)$ we see that $x \mid c$. Hence x divides both b and c, and so $x \mid \gcd(b, c)$. From this we see that $x \mid \gcd(a, \gcd(b, c)) = y$.

(iii). We shall prove the opposite statement: that $x \mid a$ if and only if $\gcd(a, x)$ is not a unit.

($\Rightarrow$) By assumption, $x \mid a$ and certainly $x \mid x$, so $x \mid \gcd(a, x)$. Since divisors of units are units, and x is not a unit, it follows that $\gcd(a, x)$ is not a unit.

($\Leftarrow$) Assume that x is irreducible, and that $d = \gcd(a, x)$ is not a unit. Then in particular $d \mid x$, so there exists $y \in R$ such that $x = dy$. But x is irreducible, and so y is a unit, and hence $x \mid d$. By definition, $d \mid a$, and so the transitivity of division implies that $x \mid a$.

(iv). From $d = \gcd(a, b)$ we deduce that $d \mid a$. From $c \sim a$ we deduce that $a \mid c$. So the transitivity of division implies that $d \mid c$. So d divides both b and c.

Let $r \in R$ be such that $r \mid b$ and $r \mid c$. From $c \sim a$ we deduce that $c \mid a$. Hence $r \mid a$, and so $r \mid \gcd(a, b) = d$.

Thus d is a greatest common divisor of b and c. ■

We can now prove that the factorisation domains where a greatest common divisor exists for every pair of elements are precisely the unique factorisation domains.

Theorem 2.8. *Let R be a factorisation domain. Then R is a unique factorisation domain if and only if every $a, b \in R$ have a greatest common divisor.*

Proof. ($\Rightarrow$) This is Theorem 2.6.

($\Leftarrow$) Let R be a factorisation domain in which every pair of elements have a greatest common divisor. By Theorem 3.7, since R is a factorisation domain, it is sufficient to prove that every irreducible in R is prime. That is, we need to show that if x is irreducible, and $x \mid rs$ for some $r, s \in R$, then $x \mid r$ or $x \mid s$. We shall prove the contrapositive, so let x be irreducible, and assume that $x \nmid a$ and $x \nmid b$, for some $a, b \in R$. We must show that $x \nmid ab$.

Clearly $x \mid xb$ and $x \mid x$ and so $x \mid \gcd(xb, x)$. By definition, $\gcd(xb, x) \mid x$, so $x \sim \gcd(xb, x)$. Hence, by Lemma 2.7(iv),

$$\gcd(ab, x) = \gcd(ab, \gcd(xb, x)).$$

Next, we use Lemma 2.7(ii) to see that

$$\gcd(ab, \gcd(xb, x)) \sim \gcd(\gcd(ab, xb), x).$$

By Lemma 2.7(i), $\gcd(ab, xb) \sim \gcd(a, x)b$, so by Lemma 2.7(iv),

$$\gcd(\gcd(ab, xb), x) = \gcd(\gcd(a, x)b, x).$$

Since x is irreducible and does not divide a, it follows from Lemma 2.7(iii) that $\gcd(a, x)$ is a unit. Hence $b \sim \gcd(a, x)b$, and hence again by Lemma 2.7(iv)

$$\gcd(\gcd(a, x)b, x) = \gcd(b, x).$$

Since x is irreducible and does not divide b, it follows from Lemma 2.7(iii) that $\gcd(b, x)$ is a unit.

We can put these four equations together to conclude that $\gcd(ab, x)$ is a unit. We assumed that x is irreducible, so by Lemma 2.7(iii) $x \nmid ab$, as required.

We have shown that every irreducible element in the factorisation domain R is prime, and so R is a unique factorisation domain. ■

Chapter 8

Euclidean domains

1. Euclidean domains

To describe the class of integral domains where the euclidean algorithm exists we require the following definition.

Definition 1.1. Let R be an integral domain. Let $\phi : R \setminus \{0\} \longrightarrow \mathbb{N}$ be any function such that for all $a \in R$ and $b \in R \setminus \{0\}$ there exist $q, r \in R$ such that

$$a = qb + r$$

and either $r = 0$ or $\phi(r) < \phi(b)$. Then ϕ is called a *euclidean function*.

Definition 1.2. If R is an integral domain such that there exists a euclidean function from $R \setminus \{0\}$ to $\mathbb{N}$, then R is called a *euclidean domain*.

Example 1.3. By the division algorithm (Chapter 2, Theorem 1.1), the integers $\mathbb{Z}$ have a euclidean function $\phi : \mathbb{Z} \setminus \{0\} \to \mathbb{N}$ defined by $\phi(x) = |x|$. Hence the integers are a euclidean domain.

Theorem 1.4. *Let R be a euclidean domain. Then R is a principal ideal domain and hence a unique factorisation domain.*

Proof. Let $\phi : R \setminus \{0\} \to \mathbb{N}$ be the euclidean function for R. Let $I \neq \{0\}$ be an ideal in R and let $x \in I$ be a non-zero element such that $\phi(x)$ is the minimum possible amongst the elements of I. We will prove that $I = (x)$.

Certainly, $(x) \subseteq I$ since $x \in I$. For the converse, if $y \in I$, then there exist $q, r \in R$ such that $y = qx + r$, where $r = 0$ or $\phi(r) < \phi(x)$. But $\phi(x)$ is minimal for elements in I and $r = y - qx \in I$, and so $r = 0$. In other words, $y = qx \in (x)$ and so $I \subseteq (x)$.

Thus R is a principal ideal domain. By Chapter 7, Theorem 1.9, every principal ideal domain is a unique factorisation domain, so R is a unique factorisation domain. ■

The converse of Theorem 1.4 does not hold. The ring $\mathbb{Z}[(1 + \sqrt{-19})/2]$ is an example of a principal ideal domain that is not a euclidean domain, although it takes a little work to prove this and we will not do so in this course.

Recall the following basic definitions. The *modulus* of $a + bi \in \mathbb{C}$ is

$$|a + bi| = \sqrt{a^2 + b^2}.$$

The distance from $a + bi$ to $c + di$ in the complex plane is just

$$|(a + bi) - (c + di)| = \sqrt{(a - c)^2 + (b - d)^2}.$$

One property of the modulus that we will use later is

$$\left|\frac{a+bi}{c+di}\right| = \frac{|a+bi|}{|c+di|}.$$

For $a \in \mathbb{Q}$, we write $\lfloor a \rfloor$ to denote the integer that we get by rounding a down to the nearest integer.

Theorem 1.5. *The Gaussian integers $\mathbb{Z}[i]$ is a euclidean domain.*

Proof. Let $N : \mathbb{Z}[i] \to \mathbb{N}$ be the usual norm on $\mathbb{Z}[i]$, i.e. $N(a+bi) = a^2 + b^2 = |a+bi|^2$. We shall show that N is also a euclidean function for $\mathbb{Z}[i]$. To do so we must prove that for all $\alpha \in \mathbb{Z}[i]$ and $\beta \in \mathbb{Z}[i] \setminus \{0\}$ there exist $q, r \in \mathbb{Z}[i]$ such that

$$\alpha = q\beta + r$$

and either $r = 0$ or $N(r) < N(\beta)$.

Since $\alpha, \beta \in \mathbb{Z}[i] \subset \mathbb{C}$ and $\beta \neq 0$, the number $\alpha/\beta = a+bi$ is an element of $\mathbb{C}$. Let $a' = \lfloor a \rfloor \in \mathbb{Z}$ and $b' = \lfloor b \rfloor$. Then in the complex plane, $\alpha/\beta = a + bi$ sits in the square of side 1 whose corners are

$$a' + b'i, (a'+1) + b'i, a' + (b'+1)i \text{ and } (a'+1) + (b'+1)i \in \mathbb{Z}[i].$$

Since the square has side length 1, the length of the diagonal is $\sqrt{2}$. It follows that the distance in the complex plane from α/β to at least one of the corners of the square is at most $\sqrt{2}/2$. Denote one such corner by $q = c + di \in \mathbb{Z}[i]$. Then

$$\left|\frac{\alpha}{\beta} - q\right|^2 \leq \left(\frac{\sqrt{2}}{2}\right)^2 = \frac{1}{2} < 1$$

and so

$$\left|\frac{\alpha}{\beta} - q\right|^2 = \left|\frac{\alpha - q\beta}{\beta}\right|^2 = \frac{|\alpha - q\beta|^2}{|\beta|^2} = \frac{N(\alpha - q\beta)}{N(\beta)} < 1. \tag{8.1}$$

Hence multiplying both sides of the inequality by $N(\beta)$

$$N(\alpha - q\beta) < N(\beta).$$

Let $r = \alpha - q\beta$. Then

$$\alpha = q\beta + r$$

and $N(r) < N(\beta)$. Hence N is a euclidean function on the integral domain $\mathbb{Z}[i]$, and so $\mathbb{Z}[i]$ is a euclidean domain. ■

Corollary 1.6. *$\mathbb{Z}[i]$ is a principal ideal domain, and $\mathbb{Z}[i]$ is a unique factorisation domain.*

Proof. Every euclidean domain is a principal ideal domain and a unique factorisation domina, by Theorem 1.4. ■

Example 1.7. We will find $q, r \in \mathbb{Z}[i]$ such that $1 + 8i = (1+2i)q + r$ where $r = 0$ or $N(r) < N(1+2i) = 5$. Dividing $1+8i$ by $1+2i$ in $\mathbb{C}$ we obtain

$$\frac{1+8i}{1+2i} = \frac{1+8i}{1+2i}\,\frac{1-2i}{1-2i} = \frac{17}{5} + \frac{6i}{5} \in \mathbb{Q}[i].$$

If we round down $\frac{17}{5}$ we get 3, and if we round down $\frac{6}{5}$ we get 1, so the four corners of the square are $3+i$, $4+i$, $3+2i$, and $4+2i$. We can choose as q any of these that are distance less than 1 from $\frac{1+8i}{1+2i}$, so we try the first one:

$$\left|\frac{17}{5} + \frac{6i}{5} - (3+i)\right|^2 = \left|\frac{2}{5} + \frac{i}{5}\right|^2 = \left(\frac{2}{5}\right)^2 + \left(\frac{1}{5}\right)^2 = \frac{5}{25} < 1$$

For practice, we will calculate all of the other distances as well, although this is not necessary:

$$\begin{aligned}\left|\frac{17}{5}+\frac{6i}{5}-(4+i)\right|^2 &= \left|\frac{-3}{5}+\frac{i}{5}\right|^2=\frac{10}{25}<1\\ \left|\frac{17}{5}+\frac{6i}{5}-(3+2i)\right|^2 &= \left|\frac{2}{5}-\frac{4i}{5}\right|^2=\frac{20}{25}<1\\ \left|\frac{17}{5}+\frac{6i}{5}-(4+2i)\right|^2 &= \left|\frac{-3}{5}-\frac{4i}{5}\right|^2=\frac{25}{25}\not<1.\end{aligned}$$

So any of $3+i$, $4+i$, or $3+2i$ can be used as the quotient q, but $4+2i$ cannot. If $q=3+i$, then

$$r=1+8i-(1+2i)q=1+8i-(1+2i)(3+i)=1+8i-(1+7i)=i$$

and $N(r)=1<5=N(1+2i)$, as required.

Note that if $\alpha,\beta\in\mathbb{Z}[i]$, then this process will find at *at least* one and *at most* four quotients q and remainders r such that $\alpha=q\beta+r$ and either $r=0$ or $N(r)<N(\beta)$: you only need one of them.

We now take a slight diversion before proving that the euclidean algorithm returns the correct answer in arbitrary euclidean domains.

Proposition 1.8. *Let R be a principal ideal domain and let $a,b\in R$. Then the greatest common divisor of a and b exists, and $d=\gcd(a,b)$ if and only if $(a,b)=(d)$.*

Proof. First note that since R is a principal idea domain, by Chapter 7, Theorem 1.9 the ring R is also a unique factorisation domain. Hence by Chapter 7, Theorem 2.8 every pair of elements of R has a greatest common divisor.

($\Rightarrow$) If $d=\gcd(a,b)$, then $d\mid a$ and $d\mid b$ and so $(a)\subseteq(d)$ and $(b)\subseteq(d)$. It follows that $(a,b)\subseteq(d)$. But R is a principal ideal domain, so $(a,b)=(c)$ for some $c\in R$, and hence $a,b\in(c)$. In particular, $c\mid a$ and $c\mid b$ so, by definition of a greatest common divisor, $c\mid d$. Thus $(d)\subseteq(c)=(a,b)$ and so $(a,b)=(d)$, as required.

($\Leftarrow$) By assumption $(a,b)=(d)$. We must show that d is a greatest common divisor of a and b. From $(a,b)=(d)$ we see that $(a)\subseteq(d)$, and so in particular $a\in(d)$ and hence $d\mid a$. Similarly, $d\mid b$, and so d is a common divisor of a and b. If $c\in R$ divides a and b, then $(a)\subseteq(c)$ and $(b)\subseteq(c)$. Thus $(d)=(a,b)\subseteq(c)$ and so $c\mid d$. It follows that d is a greatest common divisor of a and b. ■

Corollary 1.9. *Let R be a principal ideal domain. The ideals (a,b) and (a',b') are equal if and only if $\gcd(a,b)\sim\gcd(a',b')$.*

Proof. ($\Rightarrow$) Assume that $(a,b)=(a',b')$, and let $d=\gcd(a,b)$ and $d'=\gcd(a',b')$. Then by Proposition 1.8, $(d)=(a,b)=(a',b')=(d')$. In particular $(d)\subseteq(d')$, so $d'\mid d$. Similarly, $d\mid d'$, so $d\sim d'$.

($\Leftarrow$) Since every associate of d is also a greatest common divisor of a and b, we may assume that $\gcd(a,b)=\gcd(a',b')$. The result is then immediate from Proposition 1.8. ■

We are now ready to prove that the euclidean algorithm works in an arbitrary euclidean domain, so let R be a euclidean domain with euclidean function ϕ, and let $a_0,a_1\in R$.

If $a_0=0$, then $\gcd(a_0,a_1)=\gcd(0,a_1)\sim a_1$, by Chapter 7, Lemma 2.4(i). Similarly, if $a_1=0$, then $\gcd(a_0,a_1)\sim a_0$. In either case, we have found a greatest common divisor and we can stop.

Hence we may assume that $a_0,a_1\neq0$, and without loss of generality $\phi(a_0)\geq\phi(a_1)$. By the definition of a euclidean domain, there exist $q_1,a_2\in R$ such that

$$a_0=q_1a_1+a_2 \quad\text{and}\quad a_2=0 \text{ or } \phi(a_2)<\phi(a_1).$$

Hence $a_0 \in (a_1, a_2)$ and $a_2 = a_0 - q_1 a_1 \in (a_0, a_1)$ and so $(a_0, a_1) = (a_1, a_2)$. Thus $\gcd(a_0, a_1) = \gcd(a_1, a_2)$ by Corollary 1.9.

If $a_2 = 0$, then $\gcd(a_0, a_1) = \gcd(a_1, a_2) = a_1$. Otherwise we repeatedly apply this procedure to produce $a_3, a_4, \ldots$ until $a_j = 0$ for some j. This is guaranteed to happen as $\phi(a_0) > \phi(a_1) > \phi(a_2) > \cdots \geq 0$. It follows that

$$\gcd(a_0, a_1) = \gcd(a_1, a_2) = \cdots = \gcd(a_{j-1}, a_j) = \gcd(a_{j-1}, 0) = a_{j-1}.$$

Hence we have proved the following:

Theorem 1.10. *Let R be a euclidean domain Then the output of the euclidean algorithm, on input $a, b \in R$, is a greatest common divisor of a and b.*

Example 1.11. [**Finding the greatest common divisor of two Gaussian integers.**] Let $a_0 = -1 + 15i$ and $a_1 = -1 + 9i$. Then in $\mathbb{C}$

$$\frac{a_0}{a_1} = \frac{-1+15i}{-1+9i} \cdot \frac{-1-9i}{-1-9i} = \frac{136-6i}{82} = \frac{68}{41} - \frac{3i}{41}.$$

Rounding down $68/41$ gives 1 and rounding down $-3/41$ gives -1, so we find the four possible quotients $1-i, 2-i, 1, 2$ of $\mathbb{Z}[i]$. Since

$$\left|\frac{68}{41} - \frac{3i}{41} - 1\right|^2 = \left(\frac{27}{41}\right)^2 + \left(\frac{3}{41}\right)^2 < 1,$$

we can choose 1 as our quotient, to get

$$a_2 = -1 + 15i - ((-1+9i) \cdot 1) = 6i.$$

We check that $N(a_2) = N(6i) = 36 < 82 = N(a_1)$. (If we had chosen a different quotient, we would have a different value of a_2.)

Now repeat the previous steps on a_1 and a_2. That is

$$\frac{a_1}{a_2} = \frac{-1+9i}{6i} \cdot \frac{-6i}{-6i} = \frac{54+6i}{36} = \frac{9}{6} + \frac{i}{6}.$$

Since $\lfloor 9/6 \rfloor = 1$ and $\lfloor 1/6 \rfloor = 0$, we can choose 1 again as our quotient, to get

$$a_3 = (-1+9i) - (6i \cdot 1) = (-1+3i).$$

We check that $N(a_3) = N(-1+3i) = 10 < 36 = N(a_2)$.

Repeat the previous steps on a_2 and a_3:

$$\frac{a_2}{a_3} = \frac{6i}{-1+3i} \cdot \frac{-1-3i}{-1-3i} = \frac{18-6i}{10} = \frac{9}{5} - \frac{3i}{5}.$$

Since $\lfloor 9/5 \rfloor = 1$ and $\lfloor -3/5 \rfloor = -1$, we can choose $2-i$ as our quotient, to get

$$a_4 = 6i - (-1+3i)(2-i) = 6i - (1+7i) = -1-i.$$

We check that $N(a_4) = N(-1-i) = 2 < 10 = N(a_3)$.

Repeat the previous steps on a_3 and a_4. Then

$$\frac{a_3}{a_4} = \frac{-1+3i}{-1-i} \cdot \frac{-1+i}{-1+i} = \frac{-2-4i}{2} = -1-2i$$

At this point our only choice for a quotient is $-1-2i$, and

$$a_5 = -1+3i - (-1-i)(-1-2i) = -1+3i - (-1+3i) = 0.$$

Hence $\gcd(-1+15i, -1+9i)$ is the last nonzero a_i, namely $a_4 = -1-i$.

If d' is any other gcd of a_0 and a_1, then $d' \sim d$ and so there exists a unit u such that $d = d'u$. In $\mathbb{Z}[i]$ the units are ± 1 and $\pm i$ and so $\pm 1 \pm i$ are the only gcds of a_0 and a_1.

Chapter 9

Polynomial rings over unique factorisation domains

The aim of this chapter is to prove a famous theorem, due to Gauss. The theorem states that if a ring R is a unique factorisation domain, then so is the polynomial ring $R[x]$.

1. Polynomial division

We make a few additional definitions relating to polynomials.

Definition 1.1. If $f = a_0 + a_1x + \cdots + a_nx^n \in R[x]$ and $a_n \neq 0$, then the coefficient a_n is called the *leading coefficient* of f and a_nx^n is called the *leading term* of f. A polynomial is *monic* if the leading coefficient is the multiplicative identity of R.

Proposition 1.2. *Let R be a ring and let $f, g \in R[x]$ be any non-zero polynomials. If the leading coefficient of f or g is not a zero divisor, then* $\deg(fg) = \deg(f) + \deg(g)$.

Proof. See Tutorial Sheet 8. ■

Corollary 1.3. *Let R be an integral domain. Then $R[x]$ is an integral domain and the units in $R[x]$ are just the units in R.*

Proof. The ring $R[x]$ is commutative, because R is, and has the same identity as R. So to show that $R[x]$ is an integral domain, it suffices to show that $R[x]$ has no zero divisors. Let $f, g \in R[x]$. Then the leading coefficient of f (and g) is not a zero divisor and so by Proposition 1.2, $\deg(fg) = \deg(f) + \deg(g)$. It follows that $\deg(fg) = \deg(0) = -\infty$ if and only if $\deg(f) = -\infty$ or $\deg(g) = -\infty$, that is, if and only if $f = 0$ or $g = 0$. Thus $R[x]$ has no zero divisors and so is an integral domain.

If $f \in R \subsetneq R[x]$ is a unit, then there exists $g \in R \subsetneq R[x]$ such that $fg = 1$. Hence f is a unit in $R[x]$. Conversely, if $f \in R[x]$ is a unit, then there exists $g \in R[x]$ such that $fg = 1$. Hence $0 = \deg(1) = \deg(fg) = \deg(f) + \deg(g)$ and so $\deg(f) = \deg(g) = 0$. Thus $f \in R[x]$ is a unit in R. ■

Example 1.4. If R is not an integral domain, then the units in $R[x]$ may not be just the units in R. For example, let $R = \mathbb{Z}/(9)$ and let $f = 1 - 3x$, $g = 1 + 3x \in R[x]$. Then

$$fg = (1 - 3x)(1 + 3x) = 1 - 9x^2 = 1.$$

So f and g are units.

Theorem 1.5. *Let R be a commutative ring with one and let $f, g \in R[x]$ be arbitrary non-zero polynomials. If the leading coefficient of g is not a zero divisor, then there exist $q, r \in R[x]$ such that*

$$f = qg + r$$

and either $r = 0$ or none of the terms in r is divisible by the leading term of g.

Proof. See Tutorial Sheet 8. ■

Corollary 1.6. *Let R be a commutative ring with one and let $f, g \in R[x] \setminus \{0\}$. If the leading coefficient a_n of g is a unit, then there exist unique $q, r \in R[x]$ such that*

$$f = qg + r$$

and $\deg(r) < \deg(g)$.

Proof. By Theorem 1.5 there exist $q, r \in R[x]$ such that $f = qg + r$ and either $r = 0$ or none of the terms in r is divisible by the leading term of g. The leading coefficient a_n of g is a unit (by assumption), and so a_n divides every element of R. Hence if $\deg(r) \geq \deg(g)$, then $a_n x^n$ divides the leading term of r. This contradicts Theorem 1.5, so $\deg(r) < \deg(g)$.

It remains to show that q and r are unique. Assume that

$$f = q_1 g + r_1 = q_2 g + r_2$$

for some $q_1, q_2, r_1, r_2 \in R[x]$, and that $\deg(r_1) \leq \deg(r_2) < \deg(g)$. Then $(q_1 - q_2)g = r_2 - r_1$. If $q_2 - q_1 \neq 0$, then by Proposition 1.2, since a_n is a unit

$$\deg(g) \leq \deg(q_1 - q_2) + \deg(g) = \deg((q_1 - q_2)g) = \deg(r_2 - r_1) \leq \deg(r_2).$$

Hence $\deg(g) \leq \deg(r_2)$, contradicting Theorem 1.5. Thus $q_1 = q_2$ and $r_1 = r_2$, as required. ■

Theorem 1.7. *Let R be a commutative ring with one.*

(i) $R[x]$ is a principal ideal domain if and only if R is a field.

(ii) If R is a field, then $R[x]$ is a euclidean domain and a unique factorisation domain.

Proof. **(i).** ($\Rightarrow$) Since $R[x]$ is a principal ideal domain, it is certainly an integral domain.

We claim that (x) is a prime ideal in $R[x]$. To see this, let $f, g \in R[x]$ such that $fg \in (x)$. Then $x \mid fg$, and so the constant coefficient of fg is 0. But the constant coefficient of fg is the product of the constant coefficients of f and of g. The ring R is an integral domain, so at least one of the constant coefficients of f or g is 0, and hence $f \in (x)$ or $g \in (x)$, and so x is a prime ideal.

By Chapter 7, Corollary 1.8, every nonzero prime ideal of a principle ideal domain that is not a field is maximal, so (x) is a maximal ideal. By Chapter 4, Theorem 4.6 it follows that $R[x]/(x)$ is a field. The map $\phi : R[x] \to R$, $\phi : a_0 + a_1 x + \cdots + a_n x^n \mapsto a_0$ is a ring homomorphism with kernel (x), and so by the First Isomorphism Theorem

$$R[x]/(x) \cong R$$

and so R is a field.

($\Leftarrow$) Every field is an integral domain, so by Corollary 1.3 the ring $R[x]$ is an integral domain. By Corollary 1.6, the function $\deg : R[x] \setminus \{0\} \to \mathbb{N}$ is a euclidean function. Hence $R[x]$ is a euclidean domain. By Chapter 8, Theorem 1.4 every euclidean domain is a principal ideal domain.

(ii). We have just shown that if R is a field then $R[x]$ is a euclidean domain. By Chapter 8, Theorem 1.4, every euclidean domain is a unique factorisation domain. ■

As a consequence of the previous theorem, we can perform the euclidean algorithm in $R[x]$ when R is a field.

Example 1.8. [**Calculating a** gcd **in a polynomial ring.**] Let $a_0 = x^5 + x + 1, a_1 = x^4 + x^3 + x + 1 \in \mathbb{F}_2[x]$. Then using the division algorithm for polynomials we get:

$$a_0 = (x+1)a_1 + (x^3 + x^2 + x).$$

Set $a_2 = x^3 + x^2 + x$ and divide again:

$$a_1 = x \cdot a_2 + (x^2 + x + 1).$$

Set $a_3 = x^2 + x + 1$ and divide again:

$$a_2 = x \cdot a_3 + 0.$$

Hence $\gcd(x^5 + x + 1, x^4 + x^3 + x + 1) = x^2 + x + 1$.

Example 1.9. [$\mathbb{Z}[x]$ **is not a principal ideal domain.**] Since $\mathbb{Z}$ is not a field, it follows from Theorem 1.7 that $\mathbb{Z}[x]$ is not a principal ideal domain. We can also prove this directly as follows. Let $I = (2, x)$. Then

$$I = \{2f + xg \ : \ f, g \in \mathbb{Z}[x]\} = \{f \in \mathbb{Z}[x] \ : \ \text{constant coefficient of } f \text{ is even}\}.$$

Suppose that I is principal. Then there exists $f \in \mathbb{Z}[x]$ such that $I = (f)$. Hence every element of I has degree at least $\deg(f)$. Since $2 \in I$, it follows that $\deg(f) = 0$. Since $I \neq \{0\}$ and $I \neq \mathbb{Z}[x]$, it follows that $f \neq 0$, and $f \neq \pm 1$ (f is not a unit). Therefore if $g \in I = (f)$, then the coefficients of g are multiples of $f \in \mathbb{Z}$. But $x + 2 \in I$, a contradiction. Thus $I = (2, x)$ is not a principal ideal and so $\mathbb{Z}[x]$ is not a principal ideal domain.

2. Fields of fractions

Before we begin the proof of Gauss' theorem, we study a method for constructing a field that contains a given integral domain.

Lemma 2.1. *Let R be an integral domain, and let*

$$P = \{(a, b) \ : \ a, b \in R, \ b \neq 0_R\}.$$

Define a relation $\sim$ on P by $(a, b) \sim (c, d)$ if $ad = cb$ in R.

(i) For all $a, b \in R \setminus \{0\}$, we have $(0, a) \sim (c, b)$ if and only if $c = 0$.

(ii) The relation $\sim$ is an equivalence relation on P.

(iii) For all $a \in R$ and $b, c \in R \setminus \{0\}$, $(a, b) \sim (ac, bc)$.

Proof. **(i)** If $(0, a) \sim (c, b)$ then $0 \cdot b = ac$, so $c = 0$. The converse is clear.
(ii). It is easy to check that $\sim$ is reflexive and symmetric. For transitivity, if $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, then $ad = cb$ and $cf = de$. Hence $a(cf) = a(de) = (ad)e = cbe$. If $c \neq 0$ then since we are in an integral domain we can cancel the cs to deduce that $af = be$ and so $(a, b) \sim (e, f)$. If $c = 0$ then $a = 0$ and $e = 0$, by Part (i), so $(a, b) = (0, b) \sim (0, f) = (e, f)$.
(iii). See Tutorial Sheet 8. ■

Notation We write a/b for the equivalence class containing (a, b). Notice that by Lemma 2.1(iii), if $c \neq 0$ then $\frac{a}{b} = \frac{ac}{bc}$.

Theorem 2.2. *Let R be an integral domain, and let*

$$F_R = \{\frac{a}{b} \ : \ a, b \in R \text{ and } b \neq 0\}.$$

Define

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \qquad \frac{a}{b} * \frac{c}{d} = \frac{ac}{bd}.$$

Then with these operations, F_R is a field. Furthermore, F_R contains a subring isomorphic to R.

Proof. We will break the proof down into a series of steps.
Claim: $+$ and $*$ are well defined Assume that $a/b = a_1/b_1$ and $c/d = c_1/d_1$. So $ab_1 = ba_1$ and $cd_1 = dc_1$. Then

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}.$$

Hence

$$\frac{ad + bc}{bd} = \frac{(ad + bc)b_1d_1}{bdb_1d_1} = \frac{ab_1dd_1 + cd_1bb_1}{bdb_1d_1} = \frac{ba_1dd_1 + dc_1bb_1}{bdb_1d_1} = \frac{a_1d_1 + b_1c_1}{b_1d_1} = \frac{a_1}{b_1} + \frac{c_1}{d_1},$$

as required.

Similarly,

$$\frac{a}{b} * \frac{c}{d} = \frac{ac}{bd} = \frac{(ac)(b_1d_1)}{(bd)(b_1d_1)} = \frac{(ab_1)(cd_1)}{(bd)(b_1d_1)} = \frac{(ba_1)(dc_1)}{(bd)(b_1d_1)} = \frac{(bd)(a_1c_1)}{(bd)(b_1d_1)} = \frac{a_1c_1}{b_1d_1} = \frac{a_1}{b_1} * \frac{c_1}{d_1}.$$

So both operations are well defined.
Claim: $(F_R, +)$ is an abelian group. To see that $+$ is associative, we check

$$\begin{aligned}
\left(\tfrac{a}{b} + \tfrac{c}{d}\right) + \tfrac{e}{f} &= \tfrac{ad+bc}{bd} + \tfrac{e}{f} \\
&= \tfrac{(ad+bc)f+(bd)e}{bdf} \\
&= \tfrac{adf+bcf+bde}{bdf} \\
&= \tfrac{adf+b(cf+de)}{bdf} \\
&= \tfrac{a}{b} + \tfrac{cf+de}{df} \\
&= \tfrac{a}{b} + \left(\tfrac{c}{d} + \tfrac{e}{f}\right).
\end{aligned}$$

The zero of F_R is $\frac{0}{1}$. Since R is an integral domain and $b \neq 0$, $b^2 \neq 0$, so

$$\frac{a}{b} + \frac{-a}{b} = \frac{ab - ba}{b^2} = \frac{0}{b^2} = \frac{0}{1}.$$

Now

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} = \frac{cb + da}{db} = \frac{c}{d} + \frac{a}{b},$$

so $(F_R, +)$ is an abelian group.
Claim: $(F_R \setminus \{\frac{0}{1}\}, *)$ is an abelian group. To see that $*$ is associative, we check

$$\left(\frac{a}{b} * \frac{c}{d}\right) * \frac{e}{f} = \frac{ac}{bd} * \frac{e}{f} = \frac{ace}{bdf} = \frac{a}{b} * \frac{ce}{df} = \frac{a}{b} * \left(\frac{c}{d} * \frac{e}{f}\right).$$

The multiplicative identity is $1/1$. Let a/b be a nonzero element of F_R. Then $a \neq 0$ and $b \neq 0$, so b/a is also an element of F_R, and

$$\frac{a}{b} * \frac{b}{a} = \frac{ab}{ba} = \frac{ab}{ab} = \frac{1}{1}$$

so $(a/b)^{-1} = (b/a)$. Finally,

$$\frac{a}{b} * \frac{c}{d} = \frac{ac}{bd} = \frac{ca}{db} = \frac{c}{d} * \frac{a}{b}.$$

Hence $(F_R \setminus \{\frac{0}{1}\}, *)$ is an abelian group.
Claim: $(F_R, +, *)$ is a field. It only remains to check that the operations are distributive, and since we have already shown that F_R is commutative we need only check one form of the distributive law.

$$\begin{aligned}
\left(\tfrac{a}{b} + \tfrac{c}{d}\right) * \tfrac{e}{f} &= \tfrac{ad+bc}{bd} * \tfrac{e}{f} \\
&= \tfrac{(ad+bc)e}{bdf} \\
&= \tfrac{(ae)(df)+(bf)(ce)}{bdf^2} \\
&= \tfrac{ae}{bf} + \tfrac{ce}{df} \\
&= \left(\tfrac{a}{b} * \tfrac{e}{f}\right) + \left(\tfrac{c}{d} * \tfrac{e}{f}\right),
\end{aligned}$$

as required.

Claim: F_R contains a subring isomorphic to R

We can define a map ϕ from R to F_R by setting $\phi(a) = a/1$, for all $a \in R$. We check that ϕ is a ring homomorphism:

$$\begin{aligned} \phi(a) + \phi(b) &= \tfrac{a}{1} + \tfrac{b}{1} = \tfrac{a\cdot 1 + 1\cdot b}{1^2} = \tfrac{a+b}{1} \\ &= \phi(a+b) \\ \phi(a) * \phi(b) &= \tfrac{a}{1} * \tfrac{b}{1} = \tfrac{ab}{1} = \phi(ab) \end{aligned}$$

The kernel of ϕ is just $\{0_R\}$, and the image of ϕ is the set

$$R' = \{\frac{a}{1} \ : \ a \in R\}.$$

Hence $R \cong R'$, and R' is the required subring of F_R. ■

Definition 2.3. The field F_R constructed in Theorem 2.2 is called the *field of fractions* of R. We normally identify R with the subring $R' = \{\frac{r}{1} \ : \ r \in R\}$ of F_R, and think of R as a subring of F_R.

Example 2.4. Consider the integral domain $\mathbb{Z}$. The field $F_\mathbb{Z}$ has elements of the form a/b, where $a, b \in \mathbb{Z}$ and $b \neq 0$. That is, $F_\mathbb{Z} = \mathbb{Q}$. We normally think of the subring $\{\frac{n}{1} \ : \ n \in \mathbb{Z}\}$ as just being the ring $\mathbb{Z}$.

Example 2.5. The field of fractions of $\mathbb{Z}[\sqrt{2}]$ is

$$F_{\mathbb{Z}[\sqrt{2}]} = \{\frac{u}{v} \ : \ u, v \in \mathbb{Z}[\sqrt{2}], \ v \neq 0\}.$$

We can write

$$\frac{a+b\sqrt{2}}{c+d\sqrt{2}} = \frac{(a+b\sqrt{2})(c-d\sqrt{2})}{c^2-2d^2} = \frac{ac-2bd}{c^2-2d^2} + \frac{bc-ad}{c^2-2d^2}\sqrt{2} = r + s\sqrt{2}$$

where $r, s \in \mathbb{Q}$. So the field of fractions of $\mathbb{Z}[\sqrt{2}]$ is (isomorphic to) $\mathbb{Q}[\sqrt{2}]$.

3. Gauss' theorem

The aim of this section is to prove that if a ring R is a unique factorisation domain, so is $R[x]$. We saw in the previous section how to make a field F_R that contains a subring isomorphic to R, and we saw in Theorem 1.7 that a polynomial ring over a field is a unique factorisation domain. The basic idea of the proof of Gauss' Theorem is to use the facts that R and $F_R[x]$ are unique factorisation domains and that $R[x]$ is somewhere in between

Definition 3.1. Let R be a unique factorisation domain and let $f = a_n x^n + \cdots + a_0 \in R[x]$. Then the *content* of f is $\gcd(a_n, \ldots, a_0)$; we denote the content of f by $\mathrm{cont}(f)$. If $\mathrm{cont}(f) \sim 1$, then we say that f is *primitive*.

The content of a polynomial with coefficients in a unique factorisation domain is well-defined since the greatest common divisor of any pair of elements in a unique factorisation domain exists (Chapter 7, Theorem 2.8).

Example 3.2. Let $R = \mathbb{Z}$ and let $f = 5x^2 + 3x + 2 \in R[x]$. Then

$$\mathrm{cont}(f) = \gcd(5, 3, 2) = 1$$

and so f is primitive in $\mathbb{Z}[x]$.

If $R = \mathbb{Z}[i]$ and $h = (1+i)x^2 + (2+2i) \in R[x]$, then

$$\mathrm{cont}(h) = \gcd(1+i, 2+2i) = 1+i.$$

Hence h is not primitive. Notice that we can factorise h as $h = (1+i)(x^2+2) = \mathrm{cont}(h)(x^2+2)$, and x^2+2 is primitive.

Lemma 3.3. *Let R be an integral domain, let $r \in R$, and let $f = a_0 + \cdots + a_n x^n \in R[x]$. Then $r \mid f$ if and only if $r \mid \text{cont}(f)$.*

Proof. ($\Rightarrow$) Since $r \mid f$, there exists $g = b_0 + \cdots + b_m x^m \in R[x]$ such that $f = r \cdot g$. By Proposition 1.2, $\deg(f) = \deg(rg) = \deg(r) + \deg(g) = \deg(g)$, so $m = n$. Hence $rg = rb_0 + \cdots + rb_m x^m = a_0 + \cdots + a_n x^n$, and so r divides $a_0, a_1, \ldots, a_n$. Hence r divides $\text{cont}(f)$.

($\Leftarrow$) If $r \mid \text{cont}(f)$ then $r \mid a_i$ for all i, so $r \mid f$. ■

Lemma 3.4 (Gauss' Lemma) *Let R be an integral domain and let $p \in R$ be prime. Then p is prime in $R[x]$.*

Proof. Let $f = a_0 + \cdots + a_m x^m, g = b_0 + \cdots + b_n x^n \in R[x]$ be such that p divides fg, and p does not divide f. We must show that $p \mid g$.

Since $p \nmid f$, it follows from Lemma 3.3 that $p \nmid \text{cont}(f)$, so there is a coefficient of f that p does not divide. Let r be maximal, subject to $p \nmid a_r$. Starting at b_n, we will show that p divides every coefficient of g, so that $p \mid \text{cont}(g)$. It will then follow from Lemma 3.3 that $p \mid g$.

The coefficient of x^{n+r} in fg is

$$c_{n+r} = \sum_{i+j=n+r} b_i a_j = b_n a_r + b_{n-1} a_{r+1} + b_{n-2} a_{r+2} + \cdots + b_{n-k} a_{r+k},$$

where $k = \min\{n, m-r\}$. Since r is maximal subject to $p \nmid a_r$, it follows that p divides all of $a_{r+1}, a_{r+2}, \ldots, a_{r+k}$. But $p \mid fg$ and so by Lemma 3.3, $p \mid c_{n+r}$. Hence $p \mid b_n a_r$ and since we assumed that $p \nmid a_r$ and p is prime, it follows that $p \mid b_n$.

We now show that $p \mid b_i$, as i goes from $n-1$ down to 1. Let $k < n$ and assume that $p \mid b_l$ for all $l > k$. The coefficient of x^{k+r} in fg is:

$$c_{k+r} = \sum_{i+j=k+r} b_i a_j = \cdots + b_{k+1} a_{r-1} + b_k a_r + b_{k-1} a_{r+1} + \cdots .$$

By assumption, $p \mid b_l$ for all $l > k$, so it follows that p divides all terms to the left of $b_k a_r$ in the above sum. If follows from our choice of r that $p \mid a_s$ for all $s > r$, and so p divides all terms to the right of $b_k a_r$ in the above sum. Hence p divides every term of c_{k+r} except $b_k a_r$. But $p \mid c_{k+r}$, and so $p \mid b_k a_r$. We assumed that $p \nmid a_r$ and p is prime, so $p \mid b_k$.

Thus p divides $b_n, b_{n-1}, \ldots, b_0$, and so $p \mid g$, as required. ■

Proposition 3.5. *Let R be a unique factorisation domain and let $f, g \in R[x]$ be primitive. Then fg is primitive.*

Proof. If fg is not primitive, then there is an irreducible $p \in R$ such that $p \mid \text{cont}(fg)$, and so by Lemma 3.3, $p \mid fg$. Since R is a unique factorisation domain, the irreducible element p is prime in R. Hence, by Gauss' Lemma 3.4, p is also prime in $R[x]$, and so $p \mid f$ or $p \mid g$. Hence $p \mid \text{cont}(f)$ or $p \mid \text{cont}(g)$. But $\text{cont}(f) \sim 1$ and $\text{cont}(g) \sim 1$ and so p is a unit, a contradiction. ■

Proposition 3.6. *Let R be a unique factorisation domain with field of fractions F_R, and let $f \in F_R[x]$.*

(i) There exists a $c \in F_R$ such that $f = c \cdot f_1$ where $f_1 \in R[x]$ is primitive.

(ii) If $f = c_1 f_1 = c_2 f_2$, where both f_1 and f_2 are primitive in $R[x]$, then $c_1 = c_2 u$ for some unit u in R.

(iii) If $f \in R[x]$ then $c \in R$.

Proof. **(i)** Since $f \in F_R[x]$ we can write

$$f = \frac{p_n}{q_n} x^n + \cdots + \frac{p_0}{q_0}.$$

Let $\alpha = q_1 q_2 \cdots q_n \in R \setminus \{0\}$, and let $f' = \alpha f$. Then each coefficient in f' is an element of R, so $f' \in R[x]$. Let $f' = \text{cont}(f') f_1$, where $f_1 \in R[x]$ is primitive, and let $c = \frac{\text{cont}(f')}{\alpha} \in F_R$. Then

$$f = \frac{\alpha}{\alpha} f = \frac{1}{\alpha} f' = \frac{\text{cont}(f')}{\alpha} f_1 = c f_1$$

as required.

(ii) Let $c_1 = \frac{a_1}{b_1}$ and $c_2 = \frac{a_2}{b_2}$. Then

$$b_1 b_2 f = a_1 b_2 f_1 = a_2 b_1 f_2 \in R[x].$$

Hence

$$a_1 b_2 \sim \text{cont}(a_1 b_2 f_1) \sim \text{cont}(a_2 b_1 f_2) \sim a_2 b_1,$$

and so $a_1 b_2 = u a_2 b_1$ for some unit $u \in R$. Hence $\frac{a_1}{b_1} = u \frac{a_2}{b_2}$, as required.

(iii) If $f \in R[x]$, then $\alpha = 1$ in the proof of Part (i) and so $c = \beta \in R$. By Part (ii), this holds for all such decompositions of f. ■

Theorem 3.7. *Let R be a unique factorisation domain with field of fractions F_R, and let $f \in R[x]$ such that $\deg(f) > 0$. Then f is irreducible in $R[x]$ if and only if f is primitive and f is irreducible in $F_R[x]$.*

Proof. ($\Rightarrow$) We begin by showing that f is primitive. If $m \in R \setminus \{0\}$ divides $\text{cont}(f)$, then $f = m \cdot g$ for some g with $\deg(g) = \deg(f) > 0$. It follows that g is not a unit and so, since f is irreducible, m is a unit. Hence $\text{cont}(f)$ is a unit, and so f is primitive.

Suppose that there exist $g, h \in F_R[x]$ such that $f = gh$. By Proposition 3.6(i), we may write $g = cg'$ and $h = dh'$ for some $c, d \in F_R$ and some primitive $g', h' \in R[x]$. Hence $f = gh = (cg')(dh') = (cd)(g'h')$. By Proposition 3.5, $g'h'$ is primitive in $R[x]$, and so by Proposition 3.6(iii), $f = (cd)(g'h')$ implies that $cd \in R$. Hence we can write $f = (cdg')h'$ with $cdg', h' \in R[x]$. Since f is irreducible in $R[x]$, either cdg' or h' is a unit in $R[x]$, and hence by Corollary 1.3 has degree 0. Hence g or h has degree 0, and so is a unit in $F_R[x]$. Thus f is irreducible in $F_R[x]$.

($\Leftarrow$) Suppose there exist $g, h \in R[x]$ such that $f = gh$. We must show that g or h is a unit in $R[x]$. By assumption, f is irreducible in $F_R[x]$, so without loss of generality g is a unit in $F_R[x]$. Then by Corollary 1.3, g has degree 0. But $g, h \in R[x]$, so $g \in R$. Since $f = gh$, if follows from Lemma 3.3 that $g \mid \text{cont}(f)$. By assumption, f is primitive, so $\text{cont}(f) \sim 1$ and g is a unit in R. Thus g is a unit in $R[x]$, as required. ■

Theorem 3.8 (Gauss' theorem) *Let R be a unique factorisation domain. Then $R[x]$ is a unique factorisation domain.*

Proof. We begin by showing that $R[x]$ is a factorisation domain. By Corollary 1.3, since R is an integral domain, $R[x]$ is an integral domain. So we only need to prove that every non-zero non-unit polynomial $f \in R[x]$ has a factorisation into irreducibles in $R[x]$. Since $f \in F_R[x]$, it follows from Theorem 1.7(ii) that there exists a factorisation of f into irreducibles in $F_R[x]$, say

$$f = q_1 q_2 \cdots q_n$$

where every $q_i \in F_R[x]$ is irreducible. By Proposition 3.6(i), for $1 \leq i \leq n$ we may write $q_i = c_i q_i'$, where $c_i \in F_R$ and $q_i' \in R[x]$ is primitive. Since $q_i' = c_i^{-1} q_i$, it follows that q_i' irreducible in $F_R[x]$ for $1 \leq i \leq n$. By assumption q_i' is primitive, so by Theorem 3.7, every q_i' is irreducible in $R[x]$.

Now write

$$f = c_1 \cdots c_n q_1' \cdots q_n' = c q_1' \cdots q_n'.$$

By Proposition 3.5, $q_1' \cdots q_n'$ is primitive, so $c \in R$ by Proposition 3.6(iii). Write $c = p_1 \cdots p_m \in R$, where the p_i are primes in R. Then by Gauss' Lemma 3.4, each p_i is also prime in $R[x]$. All prime elements of an integral domain are irreducible (Chapter 6, Lemma 2.5), so

$$f = p_1 \cdots p_m q_1' \cdots q_n'$$

is a factorisation of f into irreducibles in $R[x]$.

It remains to show that the factorisation is unique. By Theorem 3.7, any factorisation of f into irreducibles in $R[x]$ consists of a factorisation of $\mathrm{cont}(f)$ and a factorisation of the primitive polynomial $f' = f/\mathrm{cont}(f)$ into primitive irreducibles of degree at least 1. The factorisation of $\mathrm{cont}(f)$ is into elements of R, and so is unique because R is a unique factorisation domain. Let $f' = f_1 \cdots f_n = g_1 \dots g_n$ in $R[x]$, where the f_i, g_i are primitive irreducibles in $R[x]$. By Theorem 1.7 $F_R[x]$ is a unique factorisation domain, so without loss of generality $f_i = \frac{a_i}{b_i} g_i$ for some $\frac{a_i}{b_i} \in F_R$. Since f_i and g_i are primitive, by Proposition 3.6(ii), $\frac{a_i}{b_i}$ is a unit in R, so we are done. ■

Chapter 10

Introduction to fields

Finite fields are one of the few examples of algebraic structures that are completely classified. That is, the finite fields of any given order are classified. No such classification is known for finite rings in general. No such classification is known for finite groups or finite semigroups.

In this part of the course we will only briefly introduce the very rich topic of fields. If you'd like to learn more, take MT5836 Galois Theory!

1. Characteristic

Definition 1.1. Let F be a field with multiplicative identity 1. If there exists $n > 0$ with

$$n \cdot 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = 0$$

then the minimum such n is called the *characteristic* of F. If no such n exists, then F has *characteristic* 0.

Example 1.2. The fields $\mathbb{Q}$ and $\mathbb{C}$ have characteristic 0.

The fields $\mathbb{F}_2 = \mathbb{Z}/(2)$ or $\mathbb{F}_5 = \mathbb{Z}/(5)$ have characteristic 2 and 5, respectively.

Theorem 1.3. *Let F be a field of positive characteristic. Then F has prime characteristic.*

Proof. Since F contains non-zero elements, F has characteristic $n \geq 2$. If n is not prime, then $n = km$ with $k, m \in \mathbb{Z}$, $1 < k, m < n$. Then $0 = n \cdot 1 = (km) \cdot 1 = (k \cdot 1)(m \cdot 1)$. Since F has no zero divisors, either $k \cdot 1 = 0$ or $m \cdot 1 = 0$, a contradiction. ■

Corollary 1.4. *Every finite field has prime characteristic.*

Proof. By Theorem 1.3, we only have to show that a finite field F has positive characteristic. Consider the sums $1 \cdot 1, 2 \cdot 1, 3 \cdot 1, \ldots$ of the identity. Since F has finitely many elements, there must exist integers k and m with $1 \leq k < m$ such that $k \cdot 1 = m \cdot 1$. Then $(k - m) \cdot 1 = 0$, and thus F has positive characteristic. ■

2. Polynomials again

Let F be a field. Recall that a non-zero, non-unit element $f \in F[x]$ is *irreducible* if whenever $f = gh$ for some $g, h \in F[x]$, either f or g is a unit (and so has degree 0). A polynomial is *reducible* if it is not irreducible.

Example 2.1. We show that in $\mathbb{R}[x]$, $x^2 + 1$ is irreducible. Assume not, then $x^2 + 1 = (ax + b)(cx + d)$ for some $a, b, c, d \in \mathbb{R}$ with $a, c \neq 0$. So $x^2 + 1 = acx^2 + (ad + bc)x + bd$, and hence $c = 1/a$ and $d = 1/b$. Then $ad + bc = 0$ implies that $ad = -1/ad$, and so $(ad)^2 = -1$, a contradiction.

Notice that the field of coefficients is very important: in $\mathbb{C}[x]$

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

is reducible.

Example 2.2. The polynomial $x^2 + 1$ is irreducible in $\mathbb{F}_3[x]$, for arguing as in Example 2.1 we would need $(ad)^2 = -1 = 2$. But in $\mathbb{F}_3$ we have $0^2 = 0$, $1^2 = 1$, $2^2 = 1$, so this equation has no solutions.

Again, notice that

$$x^2 + 1 = (x - 2)(x - 3) \in \mathbb{F}_5[x].$$

Theorem 2.3. *Let F be a field and let $f \in F[x]$ be arbitrary. Then $F[x]/(f)$ is a field if and only if f is an irreducible element of $F[x]$.*

Proof. By Chapter 9, Theorem 1.7, $F[x]$ is a principal ideal domain. The ring $F[x]$ is not a field since by Chapter 9, Corollary 1.3 the units are the units in F, namely $F \setminus \{0\}$. Hence by Chapter 7, Theorem 1.5, (f) is maximal if and only if f is an irreducible element of $F[x]$. Moreover, by Chapter 4, Theorem 4.6, (f) is maximal if and only if $F[x]/(f)$ is a field. ■

Definition 2.4. A *root* of a polynomial $f(x) \in F[x]$ is an element $a \in F$ such that $f(a) = 0$.

Example 2.5. (i) The elements $2, 3 \in \mathbb{Q}$ are roots of $x^2 - 5x + 6 \in \mathbb{Q}[x]$.

(ii) The polynomial $x^2 + 1 \in \mathbb{Q}[x]$ has no roots, but $x^2 + 1 \in \mathbb{C}[x]$ has two roots $\pm i$.

Theorem 2.6. *Let F be a field, let $f \in F[x]$ and let $a \in F$. Then a is a root of $f \in F[x]$ if and only if $(x - a) \mid f$.*

Proof. ($\Rightarrow$) By Chapter 9, Theorem 1.7, $F[x]$ is a euclidean domain. So we can divide f by $x - a$ to get

$$f = q \cdot (x - a) + r$$

with $q \in F[x]$ and $r \in F$. Substituting $x = a$, we get $f(a) = r$. By assumption, $f(a) = 0$, so $r = 0$ and $f = q \cdot (x - a)$. So $(x - a) \mid f$, as required.
($\Leftarrow$) $(x-a) \mid f$ implies that $f = g \cdot (x-a)$ for some $g \in F[x]$. Hence $f(a) = g(a) \cdot (a-a) = g(a) \cdot 0 = 0$, and so a is a root of f. ■

Corollary 2.7. *Let F be a field, and let $f \in F[x]$. Then*

(i) *if $\deg(f) = 1$, then f is irreducible;*

(ii) *if f is irreducible and $\deg(f) > 1$, then f has no roots;*

(iii) *if $\deg(f) = 2$ or 3, then f is irreducible if and only if it has no roots.*

Proof. See Tutorial Sheet 9. ■

The following is a summary of the steps required to describe the field $F[x]/(f)$, where f is irreducible:

- the elements of $F[x]/(f)$ are cosets $g + (f)$ with $g \in F[x]$;
- two cosets $g + (f)$ and $h + (f)$ are equal if and only if $g - h \in (f)$; i.e. if and only if $f \mid g - h$;
- given $g \in F[x]$, by Chapter 9, Corollary 1.6 there exists a unique $r \in F[x]$ such that $\deg(r) < \deg(f)$ and $g = qf + r$; that is, there exists a unique $r \in F[x]$ with $\deg(r) < \deg(f)$ and $g + (f) = r + (f)$;

- so the coset representatives in $F[x]/(f)$ are precisely $r + (f)$ where r runs through all the polynomials in $F[x]$ with $\deg(r) < \deg(f)$;
- hence if $F = \mathbb{F}_p$ and $\deg(f) = n$, then $\mathbb{F}_p[x]/(f)$ has p^n elements, as there are p^n choices for the coefficients of the coset representatives.

Example 2.8. Let $f = x + 1 \in \mathbb{F}_7[x]$. Then $\deg(f) = 1$ and so by by Corollary 2.7(i), f is irreducible. Hence $\mathbb{F}_7[x]/(f)$ is a field with $7^{\deg(f)} = 7$ elements:

$$0 + (f), 1 + (f), 2 + (f), 3 + (f), 4 + (f), 5 + (f), 6 + (f).$$

Example 2.9. Let $f = x^2 + x + 1 \in \mathbb{F}_5[x]$. Then

$$f(0) = 1,\ f(1) = 3, f(2) = 2, f(3) = 3, f(4) = 1$$

and so f is irreducible by Corollary 2.7(iii). Hence $\mathbb{F}_5[x]/(f)$ is a field of order $5^{\deg(f)} = 25$. Its elements correspond to the elements of $\mathbb{F}_5[x]$ with degree at most 1. That is, its elements are all cosets of the form $ax + b + (f)$, where $a, b \in \mathbb{F}_5$.

3. Field extensions

Definition 3.1. Let F be a field, and let $K \subseteq F$ be a subset of F that is a field with respect to the same operations. Then K is a *subfield* of F, and F is an *extension field* of K. If $K \neq F$, then K is a *proper* subfield.

Example 3.2. $\mathbb{Q}$ is a proper subfield of $\mathbb{C}$, $\mathbb{R}$ is an extension field of $\mathbb{Q}$, and $\mathbb{F}_2$ is a proper subfield of $\mathbb{F}_2[x]/(x^2 + x + 1)$.

Definition 3.3. Let F be a field. The *prime subfield* of F is the intersection of all subfields of F.

Lemma 3.4. *Let F be a field, and let K be the prime subfield of F. Then K is a field, and K contains no proper subfields.*

Proof. As a set, K is the intersection of all subfields of F. The intersection of subgroups is a group, so $(K, +)$ is an abelian group, and $(K \setminus \{0\}, *)$ is an abelian group. The distributive law is inherited from F, so K is a field.

If F' is a subfield of K, then F' is a subfield of F, so $K = K \cap F' = F'$, as required. ■

The following theorem proves that very few fields can be prime subfields.

Theorem 3.5. *Let F be a field, and let K be the prime subfield of F. Then*

(i) *if F has characteristic 0, then $K \cong \mathbb{Q}$;*

(ii) *if F has characteristic $p > 0$, then $K \cong \mathbb{F}_p$.*

Proof. **(i)** If $n \in \mathbb{Z}$ and $n < 0$, we write $n \cdot 1$ for $|n| \cdot (-1)$. If $m, n \in \mathbb{Z}$ such that $m \neq n$ and $m \cdot 1 = n \cdot 1$, then $(m - n) \cdot 1 = 0$. This implies that the characteristic of F is non-zero, a contradiction. Hence the elements $n \cdot 1$ ($n \in \mathbb{Z}$) are all distinct. Now $(n \cdot 1) - (m \cdot 1) = (n - m) \cdot 1$, and $(n \cdot 1) \cdot (m \cdot 1) = (nm) \cdot 1$, so $\{n \cdot 1 \ :\ n \in \mathbb{Z}\}$ is a subring of F isomorphic to $\mathbb{Z}$. Consider the set

$$Q = \{m \cdot 1 \cdot (n \cdot 1)^{-1} : m, n \in \mathbb{Z}, n \neq 0\}.$$

If $n \leq t$ then

$$(m \cdot 1)(n \cdot 1)^{-1} - (s \cdot 1)(t \cdot 1)^{-1} = (m(t - n) - s)(t \cdot 1)^{-1}$$

and similarly if $n > t$, so Q is closed under subtraction. It is easy to see that Q is closed under multiplication, and under taking inverses of nonzero elements. Hence Q is a subfield of F, and $Q \cong \mathbb{Q}$. Any subfield of F must contain 1 and so $Q \subseteq K$. Since Q is itself a subfield of F, we also have $K \subseteq Q$ and so in fact $Q = K$.

(ii) See Tutorial Sheet 9. ■

Definition 3.6. Let F be a field, let V be an abelian group, and let there be an external multiplication of elements from V by elements from F. Then V is said to be a *vector space* over F if the following axioms are satisfied, for all $\alpha, \beta \in F$ and all $x, y \in V$.

(V1) $(\alpha + \beta)x = \alpha x + \beta x$.

(V2) $\alpha(x + y) = \alpha x + \alpha y$.

(V3) $(\alpha\beta)x = \alpha(\beta x)$.

(V4) $1x = x$.

Theorem 3.7. *Let K be a field and let L be an extension field of K. Then L is a vector space over K.*

Proof. We set $L = V$ and $K = F$ in the above definition. The multiplication of vectors by scalars is just multiplication in L, so **(V1)** and **(V2)** follow from the distributive law in L, **(V3)** follows from associativity, and **(V4)** from the fact that the identity of K is also the identity of L. ■

A *basis* of a vector space V over F is a subset $\{v_1, \ldots, v_n\}$ of vectors in V such that every $v \in V$ can be *uniquely* written as

$$v = a_1 v_1 + \cdots + a_n v_n$$

where $a_1, \ldots, a_n \in F$. Vector spaces can have many different bases, but there are always the same number of basis vectors; this number is the *dimension* of V over F.

Definition 3.8. Let K be a field, and let L be an extension field of K. The dimension of the vector space L over K is written $[L : K]$. If $[L : K]$ is finite, then L is a *finite extension* of K.

Example 3.9. Let $L = \mathbb{C}$ and K be the subfield $\mathbb{R}$. Then $\mathbb{C}$ is a vector space over $\mathbb{R}$. Since $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$, it is clear that $\{1, i\}$ is a basis and so $[\mathbb{C} : \mathbb{R}] = 2$.

Let $L = \mathbb{F}_5[x]/(x^2 + x + 1)$, from Example 2.9, and let $K = \mathbb{F}_5$. Then a basis for L over K is $\{1, x\}$, so $[L : K] = 2$.

Theorem 3.10 (The Fundamental Theorem for finite fields) *(i) Let F be a finite field. Then F has p^n elements, where the prime p is the characteristic of F and n is the dimension of F over its prime subfield.*

(ii) If p is any prime and n is any positive integer, then up to isomorphism there exists a unique field with p^n elements.

Proof. (i) Since F is finite, it has characteristic p for some prime p by Corollary 1.4. Hence by Theorem 3.5 the prime subfield of F is isomorphic to $\mathbb{F}_p$. Thus F is an extension field of $\mathbb{F}_p$. In particular, by Theorem 3.7, F is a vector space over $\mathbb{F}_p$ of dimension n, for some $n \in \mathbb{N}$. Hence $|F| = p^n$, as required.

(ii) Proof omitted. See MT5836 Galois Theory. ■