

Notes on Rings and Arithmetic

by PETER M NEUMANN (Queen's College, Oxford)

Preface

These notes are intended as a rough guide to the course *Rings and Arithmetic* which is a part of the Oxford 2nd year undergraduate course in mathematics. Please do not expect a polished account. They are lecture notes, not a carefully checked text-book. Nevertheless, I hope they may be of some help.

By ‘arithmetic’ we mean, in the first instance, operations with numbers—addition, multiplication, subtraction, division (where possible). Abstracting from familiar systems, such as the integers \mathbb{Z} , the rational numbers \mathbb{Q} , the real numbers \mathbb{R} , the complex numbers \mathbb{C} , in which these natural human activities take place, we generalise to *commutative rings, integral domains, fields*. It is these that form the subject of these notes.

The word ‘arithmetic’ has come to have a second meaning, however. It refers also to elementary number theory—properties of the natural numbers with respect to divisibility and related matters. Sometimes this is called ‘The Higher Arithmetic’ as in the title of a famous and charming little book by H Davenport. The beginnings of that area are part of the syllabus of this course because they have a considerable number of applications outside of number theory. And not just in other areas of pure mathematics: in recent decades they have become also of considerable importance in several areas of modern applied mathematics—design of experiments in statistics, design and analysis of codes in information theory, design and implementation of cryptographic systems used throughout the IT industry for security systems, for example.

It is a pleasure to acknowledge with warm thanks that these notes have benefitted from comments and suggestions by Dr Jan Grabowski. Any remaining errors, infelicities and obscurities are of course my own responsibility—I would welcome feedback.

HMN: Queen's: 30.ix.2007

CONTENTS

Part I: Rings

Axioms for Commutative Rings	1
Subrings	5
Units	6
Fields	7
Integral domains	7
The characteristic of a ring	10
Ideals	10
Quotient Rings	11
Ring Homomorphisms	13
Image and kernel of a homomorphism: Isomorphism Theorems	13
The Chinese Remainder Theorem	19
Some further exercises	20

Part II: Arithmetic	22
Divisibility	22
Euclidean rings	24
The Euclidean Algorithm	28
Some applications to polynomial rings	30
The Gaussian integers	33
Further exercises	36

Part I: Rings

As has been indicated, rings are the mathematical objects in which arithmetic takes place. Recall that addition and multiplication are functions of two variables that return an entity of the same kind as what is substituted for those variables. Thus if the set of ‘numbers’ under consideration is R then, in the set-theoretic language taught and learned in first-year mathematics, they are functions $R \times R \rightarrow R$. Such functions are known as *binary operations* on R . Similarly, the function which assigns to a ‘number’ its negative is a *unary operation*. The word ‘number’ is here in inverted commas because we already know examples where the elements to be added, multiplied or subtracted, are not numbers. For example, we know how to add, multiply or subtract polynomials with real coefficients, or real-valued functions of a real variable. But in all these contexts the arithmetic operations have much in common. For example, they are always associative and commutative. Bear this, and the familiar examples, in mind when you come to the general notion of a ring.

Axioms for Commutative Rings

A *commutative ring* (with 1) is

- a set R equipped with
- two binary operations $+$: $R \times R \rightarrow R$ and \times : $R \times R \rightarrow R$
(notation: $+$: $(a, b) \mapsto a + b$ and \times : $(a, b) \mapsto ab$)
- a unary operation $-$: $R \rightarrow R$
(notation: $a \mapsto -a$)
- distinguished elements 0 and 1
- satisfying the following conditions:
 - (1) $a + (b + c) = (a + b) + c$ [+ is associative]
 - (2) $a + b = b + a$ [+ is commutative]
 - (3) $a + 0 = a$
 - (4) $a + (-a) = 0$ [- is additive inverse]
 - (5) $a(bc) = (ab)c$ [\times is associative]
 - (6) $ab = ba$ [\times is commutative]
 - (7) $a1 = 1a = a$
 - (9) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ [\times distributes over +]

NOTE 1: All these axioms are to be read as holding **for all** $a, b, c \in R$.

NOTE 2: In general a *ring* is a system satisfying all these conditions except (6); it is this axiom that distinguishes *commutative* rings. And it is for this reason that Axioms (7) and (9) have been expressed in ‘two-sided’ form, even although Axiom (6) would permit them to be abbreviated.

NOTE 3: A few authors choose not to postulate existence of 1.

NOTE 4: The alert and inquisitive reader will wonder where Axiom (8) has gone. That (and it) will emerge shortly.

The first four axioms capture the essential properties of addition. They tell us that under the single binary operation of addition R forms a commutative *group* (as defined in the first-year course). The next three axioms capture the essential properties of multiplication, and Axiom (9) connects addition and multiplication.

EXAMPLES. The systems \mathbb{Z} , \mathbb{Q} , \mathbb{R} , \mathbb{C} , $\mathbb{R}[x]$ (polynomials in x with real coefficients), with their usual operations, are commutative rings.

For any set X the set \mathbb{R}^X of real-valued functions on X (that is, functions $X \rightarrow \mathbb{R}$) with 0 , 1 being the appropriate constant functions, and with pointwise negative, pointwise addition and pointwise multiplication, is a commutative ring. [If you have never seen this proved you should check it—once, but only once, in your life.]

The set $\{0\}$ with the only possible operations is a ring—as a simple exercise you may like to prove that it is the only ring in which $0 = 1$. We call this the *trivial* ring.

NON-EXAMPLES. The system \mathbb{N} with its usual operations is not a commutative ring (negatives do not exist, subtraction is not always possible).

The set $2\mathbb{Z}$ of even integers with its usual operations is not a commutative ring (no multiplicative identity).

The set $M_{d \times d}(\mathbb{R})$ of $d \times d$ matrices with real coefficients, equipped with the usual matrix addition and multiplication is a ring, but is not commutative if $d \geq 2$.

NOTATIONAL AND TERMINOLOGICAL CONVENTIONS. The notations $+$, \times (usually elided in formulae, so that $a \times b$ is written ab) are absolutely standard and the same the world over for addition and multiplication respectively. The same is true of $-$ (extended so that $a + (-b)$ is written as $a - b$) for subtraction, and of 0 , 1 as symbols for zero and unity.

To describe a ring we should give the set of its elements and specify also the particular operations $+$, \times , $-$, and the particular elements 0 , 1 that we have in mind. Conventionally, however, we use a name for the set of elements as a name for the ring, leaving the operations to be understood or inferred from the context. This works well provided we remember to specify the operations in any situation where ambiguity might arise.

EXERCISE 1: Which of the following are rings (commutative, with unity), which are not?

- (i) the set $27\mathbb{Z}$ of all integers divisible by 27 with the usual addition and multiplication;
- (ii) \mathbb{Z} with addition given by $a \oplus b := a + b + 1$, multiplication $a \otimes b := ab + a + b$, and with appropriate definitions of 0 , 1 and $-$;
- (iii) $M_{2 \times 2}(\mathbb{Z})$, the set of all 2×2 matrices with integer coefficients with its usual addition and multiplication.

EXERCISE 2: Let X be a set. Define ‘addition’ and ‘multiplication’ on its power set $\mathcal{P}X$ by $u + v := (u \cup v) \setminus (u \cap v)$ and $uv := u \cap v$. Prove that, with suitable definitions of 0 , 1 and $-$, this turns $\mathcal{P}X$ into a commutative ring with 1 in which $x^2 = x$ for all elements x .

The axioms express familiar properties of arithmetic as we know it for numbers, polynomials, functions, and the like. What is astonishing is that these few very basic properties are enough to pin down pretty well all the basic facts of arithmetic. For example:

OBSERVATION. Let R be any commutative ring and let $a, b, c \in R$. Then

- (i) if $a + c = b + c$ then $a = b$;
- (ii) if $a + a = a$ then $a = 0$;
- (iii) $-(-a) = a$;
- (iv) $0a = 0$;
- (v) $(-a)b = -(ab) = a(-b)$;
- (vi) $-a = (-1)a$.

Proof. Suppose that $a + c = b + c$. Then $(a + c) - c = (b + c) - c$ (recall that $x - y$ is an abbreviation for $x + (-y)$), so $a + (c - c) = b + (c - c)$ by Axiom (1), that is $a + 0 = b + 0$ by Axiom (4), whence $a = b$ by Axiom (3).

Suppose that $a + a = a$. Then $(a + a) - a = a - a$. Therefore $a + (a - a) = 0$ by Axioms (1) and (4). Now Axioms (4) and (3) yield that $a = 0$.

We know that $a + (-a) = 0$ and $(-a) + (-(-a)) = 0$ by Axiom (4). Applying Axiom (2) we see that $a + (-a) = (-(-a)) + (-a)$, and now Part (i) yields that $a = -(-a)$.

We leave Parts (iv) and (v) as exercises for the reader (see below).

To prove (vi) we write $(1 + (-1))a$ in two ways. On the one hand it is $0a$ by Axiom (4), and this is 0 by (iv). On the other hand it is $1a + (-1)a$ by Axiom (9), that is, $a + (-1)a$ by Axiom (7). Thus $a + (-1)a = 0 = a + (-a)$ by Axiom (4), so $(-1)a + a = (-a) + a$ by Axiom (2), and it follows that $(-1)a = -a$ by (i).

EXERCISE 3: Let R be a ring.

- (i) Prove that $0a = 0$ for all $a \in R$.
- (ii) Prove that $(-a)b = -(ab) = a(-b)$ for all $a, b \in R$.

MORE EXAMPLES: POLYNOMIAL RINGS. Let R be a commutative ring. A *polynomial* over R in the indeterminate (or ‘variable’) x is a formal expression of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0, a_1, a_2, \dots, a_n \in R$. The elements $a_0, a_1, a_2, \dots, a_n$ are known as the *coefficients*, R as the *coefficient ring*. If $a_n = 0$ then

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1},$$

so we can delete terms with coefficient 0 from the end of a polynomial. Likewise we can of course add such terms. The zero polynomial is the one all of whose coefficients are 0. If $f(x)$ is a non-zero polynomial $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ in which $a_n \neq 0$ then we define the *degree* $\deg f$ to be n . Given

$$f(x) = a_0 + a_1x + \cdots + a_nx^n \quad \text{and} \quad g(x) = b_0 + b_1x + \cdots + b_mx^m,$$

we define

$$(-f)(x) := \sum (-a_r)x^r,$$

we define

$$(f + g)(x) := \sum (a_r + b_r)x^r$$

with the convention that $a_r = 0$ if $r > n$ and $b_r = 0$ if $r > m$, and we define

$$(fg)(x) := \sum_{r=0}^{m+n} \left(\sum_{s+t=r} a_s b_t \right) x^r.$$

These are exactly the same rules for addition and multiplication of polynomials that should be familiar from the first-year treatment of $\mathbb{R}[x]$. And now,

EXERCISE 4 (worth doing carefully once in one's life, but not more than once—unless an examiner offers marks for it): with 1 defined to be the polynomial of degree 0 such that $a_0 = 1$, prove that if R is a commutative ring then $R[x]$ is a commutative ring.

Note that the ring $R[x, y]$ of formal polynomials in two variables can be constructed as $(R[x])[y]$, the ring of polynomials in y with coefficients that are polynomials in x . We can describe polynomial rings $R[x_1, \dots, x_n]$ in many variables in a similar way.

MORE EXAMPLES: DIRECT PRODUCTS. Let R, S be commutative rings. The *direct product* $R \times S$ is defined as follows:

$$\text{set } := R \times S = \{(a, b) \mid a \in R, b \in S\};$$

$$0 := (0, 0) \text{ and } 1 := (1, 1);$$

$$-(a, b) := (-a, -b) \text{ and } (a_1, b_1) + (a_2, b_2) := (a_1 + a_2, b_1 + b_2);$$

$$(a_1, b_1)(a_2, b_2) := (a_1 a_2, b_1 b_2).$$

Thus arithmetic in $R \times S$ is defined componentwise. Note that in the definition $0 := (0, 0)$ the symbol 0 has three different meanings. But it should be clear from the context that the first instance is the zero element of $R \times S$, the second is the zero of R and the third the zero of S . Similar comments apply to all the other definitions.

EXERCISE 5 (worth doing carefully once in one's life, but not more than once—unless an examiner offers marks for it): if R and S are commutative rings then so is the direct product $R \times S$.

Subrings

A *subring* of a commutative ring R is a subset S of R such that

- (1) $0, 1 \in S$;
- (2) $a \in S \Rightarrow -a \in S$;
- (3) $a, b \in S \Rightarrow a + b \in S$;
- (4) $a, b \in S \Rightarrow ab \in S$.

We say that S is *closed* with respect to the operations of R and write $S \leq R$ (rather than $S \subseteq R$).

EXAMPLES. The ring \mathbb{Z} of integers is a subring of \mathbb{Q} , which is a subring of \mathbb{R} , which itself is a subring of \mathbb{C} . That is, $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$.

If R is any commutative ring and R is identified with the set of constant polynomials then $R \leq R[x]$.

If R and S are commutative rings then, identifying R with $R \times \{0\} \subseteq R \times S$ and S with $\{0\} \times S \subseteq R \times S$, we have $R \leq R \times S$ and $S \leq R \times S$.

EXERCISE 6: Let $d \in \mathbb{Z}$. Define $\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} \in \mathbb{C} \mid a, b \in \mathbb{Z}\}$. Check that $\mathbb{Z}[\sqrt{d}] \leq \mathbb{C}$ (of course if $d \geq 0$ then $\mathbb{Z}[\sqrt{d}] \leq \mathbb{R}$).

In particular, $\mathbb{Z}[\sqrt{-1}]$ is a very important ring. Its members are known as *Gaussian integers*.

The following proposition reduces (if only by a small amount) the labour required to check for subrings.

PROPOSITION. *Let R be a commutative ring and let S be a subset of R . Then S is a subring if and only if*

- (i) $1 \in S$,
- (ii) $a, b \in S \Rightarrow a - b \in S$, and
- (iii) $a, b \in S \Rightarrow ab \in S$.

Proof. That a subring satisfies (i), (ii) and (iii) should be clear. The force of the assertion is the converse. So suppose that S satisfies (i), (ii) and (iii). From (i) and (ii) we see that $0 = 1 - 1 \in S$. Then from (ii) we see that if $a \in S$ then $0 - a \in S$, that is, $-a \in S$. Finally, if $a, b \in S$ then $a - (-b) \in S$, that is, $a + b \in S$. Thus S is a subring, as we claimed.

EXERCISE 7 (Identification of all the subrings of \mathbb{Q}): Let Π be a set of prime numbers. A positive integer n is said to be a Π -number if all its prime factors lie in Π (note that 1 will always be a Π number). Define

$$\mathbb{Z}_{\Pi} := \{m/n \in \mathbb{Q} \mid m, n \in \mathbb{Z}, n > 0, \text{ and } n \text{ is a } \Pi\text{-number}\}.$$

- (i) Show that \mathbb{Z}_{Π} is a subring of \mathbb{Q} .

(ii) What is \mathbb{Z}_Π when $\Pi = \emptyset$? And when Π consists of all the prime numbers?

Now let R be any subring (with 1) of \mathbb{Q} .

(iii) Let $m/n \in R$, with m, n co-prime, and let p be a prime divisor of n . Show that $p^{-1} \in R$. [Hint: You may find it helpful to recall from Mods that, since m and p are co-prime, there exist $a, b \in \mathbb{Z}$ such that $am + bp = 1$. Then divide through by p , etc.]

(iv) Deduce that $\mathbb{Z}_{\{p\}} \leq R$.

(v) Prove that there exists a unique set Π of prime numbers such that $R = \mathbb{Z}_\Pi$.

Units

Let R be a commutative ring. An element $u \in R$ is said to be a *unit* if there exists $v \in R$ such that $uv = vu = 1$. Often such elements are said to be *invertible* and we write u^{-1} for the element v . This is justified by the observation that, although multiplicative inverses need not exist, when they do exist they are unique:

LEMMA. *Suppose that R is a commutative ring and $u, v, w \in R$. If $uv = uw = 1$ then $v = w$.*

Proof. Suppose that $uv = uw = 1$. Then $v1 = v(uw) = (vu)w = (uv)w = 1w$, that is $v = w$, as required.

As was said a few lines ago, multiplicative inverses need not exist. For example, 0 has an inverse if and only if $R = \{0\}$; if R is non-trivial then 0 does not have an inverse. For any commutative ring R we define

$$U(R) := \{u \in R \mid u \text{ is invertible}\}.$$

the group of units or *unit-group* of R .

EXAMPLES. $U(\mathbb{Z}) = \{\pm 1\}$, $U(\mathbb{Q}) = \mathbb{Q} \setminus \{0\}$.

EXERCISE 8: Let R, S be rings (commutative and with 1). Identify $U(R \times S)$ in terms of $U(R)$ and $U(S)$.

LEMMA. *If R is a commutative ring then $U(R)$ is a commutative group under multiplication.*

The proof is left as an exercise—you need to observe that the product of two units is a unit, and that the inverse of a unit is a unit; then associativity and the other axioms for groups come from the properties of ring multiplication.

EXERCISE 9: Let R be a commutative ring. Define an equivalence relation \sim on R by $a \sim b : \iff \exists u \in U(R) : a = bu$. Show that \sim is an equivalence relation on R .

NOTE: if $a \sim b$ then a, b are known as *associates*.

Fields

Fields are a particularly important kind of commutative ring. A *field* is a commutative ring in which every non-zero element is invertible. Thus fields F are characterised among commutative rings by the equation $U(F) = F \setminus \{0\}$. When F is a field we often write F^\times for its multiplicative group $F \setminus \{0\}$. Fields are sufficiently important that it is worth writing down a self-contained description of the concept.

Axioms for Fields

A *field* is a set F equipped with two binary operations $+$ and \times , a unary operation $-$, and distinguished elements 0 and 1 , satisfying the following conditions for all $a, b, c \in F$:

- (1) $a + (b + c) = (a + b) + c$ [$+$ is associative]
- (2) $a + b = b + a$ [$+$ is commutative]
- (3) $a + 0 = a$
- (4) $a + (-a) = 0$ [$-$ is additive inverse]
- (5) $a(bc) = (ab)c$ [\times is associative]
- (6) $ab = ba$ [\times is commutative]
- (7) $a1 = 1a = a$
- (8) $a \neq 0 \Rightarrow \exists x \in F : ax = 1$
- (9) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ [\times distributes over $+$]
- (10) $0 \neq 1$.

The number systems \mathbb{Q} , \mathbb{R} , \mathbb{C} are familiar and very important examples of fields. Note the very important *caveat* in Axiom 8: it is only if $a \neq 0$ that it has an inverse a^{-1} . It corresponds to

THE FIRST LAW OF COMMONSENSE: *Thou shalt not divide by zero.*

After all, we know that $0x = 0y$ for every $x, y \in F$ so if division by 0 were permitted then we'd have $x = y$ for all $x, y \in F$ and F would be of no interest at all.

Integral domains

Let R be a commutative ring. An element $a \in R \setminus \{0\}$ is said to be a *zero-divisor* if there exists $b \in R \setminus \{0\}$ such that $ab = 0$.

EXAMPLE. Let X, Y be non-trivial commutative rings and let $R := X \times Y$. If $x \in X \setminus \{0\}$ and $y \in Y \setminus \{0\}$, and if $a := (x, 0) \in R$, $b := (0, y) \in R$ then $ab = 0$. Thus there are many zero-divisors in a direct product.

An *integral domain* is a commutative ring in which $0 \neq 1$ and there are no divisors of zero.

The number systems \mathbb{Z} , \mathbb{Q} , \mathbb{R} , \mathbb{C} are familiar and very important examples of integral domains. Indeed, \mathbb{Z} is the prototypical integral domain. Notice that any field F is an integral domain because if $a \in F \setminus \{0\}$ and $ab = 0$ then, multiplying by a^{-1} we see that $b = 0$. As in the case of fields, integral domains are sufficiently important that it is worth writing down a self-contained description of the concept:

Axioms for Integral Domains

An *integral domain* is a set R equipped with two binary operations $+$ and \times , a unary operation $-$, and distinguished elements 0 and 1 , satisfying the following conditions for all $a, b, c \in R$:

- (1) $a + (b + c) = (a + b) + c$ [+ is associative]
- (2) $a + b = b + a$ [+ is commutative]
- (3) $a + 0 = a$
- (4) $a + (-a) = 0$ [- is additive inverse]
- (5) $a(bc) = (ab)c$ [\times is associative]
- (6) $ab = ba$ [\times is commutative]
- (7) $a1 = 1a = a$
- (8') $(a \neq 0 \ \& \ b \neq 0) \Rightarrow ab \neq 0$
- (9) $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ [\times distributes over +]
- (10) $0 \neq 1$.

We emphasize that

$$\{\text{fields}\} \subset \{\text{integral domains}\} \subset \{\text{commutative rings}\}.$$

It is worth having an alternative characterisation of integral domains. They are precisely the commutative rings in which cancellation is possible:

LEMMA. *Let R be a commutative ring. Then R is an integral domain if and only if non-zero elements can be cancelled, that is, it satisfies*

$$(8'') \quad (ac = bc \ \& \ c \neq 0) \Rightarrow a = b$$

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

Proof. Suppose that (8'') holds. Suppose also that $a \in R \setminus \{0\}$ and $ab = 0$ for some $b \in R$. Then $0a = ba$ and so by (8'') (with 0 substituted for a and a for c), $b = 0$. Thus a is not a zero-divisor. Hence R is an integral domain.

Now suppose conversely that R is an integral domain. Suppose also that $a, b, c \in R$, that $ac = bc$ and that $c \neq 0$. Then $(a - b)c = 0$ and since c is not a zero-divisor we must have $a - b = 0$, that is, $a = b$. Thus (8'') holds.

EXERCISE 10: Prove that a finite integral domain is a field. [*Hint*: let R be a finite integral domain and let $a \in R \setminus \{0\}$. The problem is to show that a has an inverse. Show that the map $x \mapsto ax$ is one-one on R to R and exploit this.]

OBSERVATION. *Any subring of an integral domain is an integral domain.*

This is almost obvious: if R is an integral domain and $S \leq R$ then, since there are no zero-divisors in R there cannot be any in S . Note, in particular, that any subring of \mathbb{R} or of \mathbb{C} is an integral domain—thus we see for example that the rings $\mathbb{Z}[\sqrt{d}]$ introduced on p. 5 are integral domains.

In fact, integral domains can be characterised as being those commutative rings that arise as subrings of fields. We have seen that any subring of a field is an integral domain. The converse is as follows:

PROPOSITION [not syllabus material]. *Let R be an integral domain. There exists a field F such that $R \leq F$.*

Sketch proof. Define $A := R \times (R \setminus \{0\})$ —note that this is simply the product set; it is not a ring. Define a binary relation \sim on A by the rule

$$(a, b) \sim (a', b') :\Leftrightarrow ab' = ba'.$$

It is easy to check that \sim is reflexive, that it is symmetric, and that it is transitive. (Note that it is for proving transitivity that we need R to be an integral domain.) Thus \sim is an equivalence relation.

Write a/b for the equivalence class of the pair (a, b) . We would like to think of a/b as being a fraction in the same way as rational numbers are fractions of integers (with non-zero denominator). Thus we'd like to define $-(a/b)$ to be $-a/b$, we'd like to define $(a/b) + (c/d)$ to be $(ad + bc)/bd$, and we'd like to define $(a/b)(c/d)$ to be $(ac)/(bd)$. For these specifications to make sense however, that is, to be unambiguous, we need to know that

$$\begin{aligned} &\mathbf{if} (a, b) \sim (a', b') \mathbf{and} (c, d) \sim (c', d') \mathbf{then} \\ &(-a, b) \sim (-a', b') \mathbf{and} (ad + bc, bd) \sim (a'd' + b'c', b'd') \mathbf{and} (ac, bd) \sim (a'c', b'd'). \end{aligned}$$

This is true and its proof is a routine exercise. Once that is done we define

$$F := \{a/b \mid (a, b) \in A\}, \quad 0 := 0/1, \quad 1 := 1/1,$$

and

$$-(a/b) := -a/b, \quad (a/b) + (c/d) := (ad + bc)/bd, \quad (a/b)(c/d) := (ac)/(bd).$$

What is then left to do is to show that Axioms (1)–(10) hold. Again, this is routine and easy, if rather a long calculation. The upshot, however, is that F is a field. It is known as the *field of fractions* of R . If we then identify R with the set of elements $a/1$ of F then we find that we have a copy (anticipating a little, an isomorphic copy) of R embedded as a subring of F . It is in this sense that $R \leq F$.

The characteristic of a ring

Let R be a commutative ring and let $a \in R$. We define na inductively for $n \in \mathbb{N}$:

$$0a := 0 \text{ and } (n+1)a = na + a \text{ for } n \geq 0.$$

Thus $na = a + a + \cdots + a$ with n summands on the right of the equation. As one would expect, one can easily prove that $(m+n)a = ma + na$ and $(ma)(nb) = (mn)(ab)$. Consider now what happens to such multiples of 1: it is possible that $1, 1+1, 1+1+1, \dots$ are all different. In this case we say that R has *characteristic 0*. If, on the other hand two of them coincide, say $r1 = s1$ with $r < s$ then $(s-r)1 = 0$. In this case, then, there are positive integers n such that $n1 = 0$ and we define the *characteristic* of R to be the smallest such n . Thus the characteristic is defined by

$$\text{char } R := \begin{cases} 0 & \text{if } n1 \neq 0 \text{ for all } n > 0, \\ \min\{n \in \mathbb{N} \mid n > 0, n1 = 0\} & \text{otherwise.} \end{cases}$$

Thus the characteristic of R is the additive order of its multiplicative identity 1, or 0 if the additive order of 1 is infinite.

In general we can say little more. For integral domains, however, there is a very strong restriction on what the characteristic may be:

PROPOSITION. *Let R be an integral domain and let $p := \text{char } R$. If $p \neq 0$ then p is a prime number.*

Moreover, if $a \in R \setminus \{0\}$ then the additive order of a is p (infinite if $p = 0$).

Proof. Suppose that $p \neq 0$ and that $p = rs$ where (without loss of generality) $1 \leq r \leq s \leq p$ and $r, s \in \mathbb{N}$. Then $(r1)(s1) = (rs)1 = p1 = 0$ and so since R is an integral domain, either $r1 = 0$ or $s1 = 0$. Since p is minimal such that $p1 = 0$ it follows that $r = 1$ and $s = p$. Thus p is prime.

Now let $a \in R \setminus \{0\}$. Then $pa = (p1)a = 0a = 0$, so if $p > 0$ then the additive order of a divides p and, since it is not 1, it must be p . If $p = 0$ then $n1 \neq 0$ for $n > 0$ and so, since $na = (n1)a$, also $na \neq 0$ (since R is an integral domain); thus in this case also a has infinite additive order.

Ideals

A subset A of a commutative ring R is said to be an *ideal* if

- (1) $0 \in A$ and $a, b \in A \Rightarrow a + b, -a \in A$ (so A is an additive subgroup);
- (2) $(a \in A, x \in R) \Rightarrow xa \in A$.

EXAMPLES: $\{0\}$, R are always ideals. An ideal different from $\{0\}$ is called *non-trivial* or *non-zero*; an ideal different from R is called *proper*.

For any integer n the set $n\mathbb{Z}$ of all multiples of n is an ideal in \mathbb{Z} .

Generally, if R is any commutative ring and $a \in R$ then aR , the set of all multiples of a , is an ideal.

NOTE: Such an ideal aR (or Ra) is known as the *principal ideal generated by the element a* . It is sometimes denoted (a) or $\langle a \rangle_R$.

NOTE: We thus have two natural notations for the trivial ideal, $\{0\}$ and (0) . Generally, I prefer the former as being more basic. Many authors simply write 0 , which is an abuse of notation—though perhaps a harmless one.

EXERCISE 11: Let R be an integral domain and let $a, b \in R$. Show that $aR = bR$ (that is, a, b generate the same principal ideal) if and only if a, b are associates (see p. 6).

EXERCISE 12: Let R, S be rings (commutative and with 1). Show that if X is an ideal in R and Y an ideal in S then $X \times Y$ is an ideal in $R \times S$. Show conversely that if A is an ideal in $R \times S$ then there exist ideals X in R and Y in S such that $A = X \times Y$.

NOTE: If A is an ideal and $1 \in A$ then, since $x = x1 \in A$ for all $x \in R$, we must have $A = R$. Thus a proper ideal is never a subring according to our conventions. Be aware, however, that if the concept of ring does not require the existence of 1 then the theory is a little different at this point—in his book *Introduction to Algebra* (OUP 1998) Peter J Cameron goes to some pains to emphasize that for him an ideal is a subring (of a particular kind). But his rings do not necessarily have identity elements 1, and so even in a ring that does happen to have an identity 1, subrings need not contain 1.

Quotient Rings

Let A be an ideal in the commutative ring R . The *quotient ring* R/A is defined as follows:

$$\begin{aligned} \text{Set} &:= \{x + A \mid x \in R\} && \text{[additive cosets]} \\ 0 &:= A \\ 1 &:= 1 + A \\ -(x + A) &:= (-x) + A \\ (x + A) + (y + A) &:= (x + y) + A \\ (x + A)(y + A) &:= (xy) + A. \end{aligned}$$

Of course it should be checked that this specification does describe a ring. The issues are:

- are $-$, $+$ and \times well-defined?
- do the axioms for commutative rings (see p. 1) hold?

As an example we check here that \times is well-defined, leaving the rest to the conscientious reader. Suppose that $x + A = x' + A$ and $y + A = y' + A$. It is conceivable that $xy + A$ is a different coset from $x'y' + A$, and if this were so then the definition would not make sense. However, $x' = x + a$ and $y' = y + b$ for some $a, b \in A$. Therefore $x'y' = xy + ay + xb + ab$. Because A is an ideal, each of ay, xb, ab lies in A and therefore so does $ay + xb + ab$. Thus $x'y' = xy + c$ where $c \in A$, and so $xy + A = x'y' + A$. This ensures that \times is well-defined.

IMPORTANT EXAMPLE: For a natural number $n > 0$ we define $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$, the ring of integers modulo n . It is a very important ring and worth the little time it takes to become familiar with it.

First, what are the additive cosets of $n\mathbb{Z}$ in \mathbb{Z} ? Given any integer x the set of all $y \in \mathbb{Z}$ for which $x - yn \geq 0$ is bounded above. Therefore we may define

$$q := \max\{y \in \mathbb{Z} \mid x - yn \geq 0\} \quad \text{and} \quad r := x - qn.$$

From the definition we see that $x - (q+1)n < 0$, that is, $r - n < 0$. Thus $x = qn + r$ where $0 \leq r \leq n-1$. It follows that x lies in one of the cosets $r + n\mathbb{Z}$ for this range of r , and so this is a list of all the cosets of $n\mathbb{Z}$ in \mathbb{Z} . Also, however, if $r_1 + n\mathbb{Z} = r_2 + n\mathbb{Z}$, where (without loss of generality) $0 \leq r_1 \leq r_2 \leq n-1$, then, since $r_2 - r_1$ is a multiple of n , we must have $r_1 = r_2$. Thus $n\mathbb{Z}$ has exactly n cosets in \mathbb{Z} , namely

$$n\mathbb{Z}, 1 + n\mathbb{Z}, 2 + n\mathbb{Z}, \dots, (n-1) + n\mathbb{Z},$$

and \mathbb{Z}_n has size n . These cosets are often called the *residue classes* modulo n .

To calculate in \mathbb{Z}_n it is convenient to use the notation of congruence introduced by C F Gauss in his historic book *Disquisitiones arithmeticae* (Arithmetical investigations) of 1801. For $a, b \in \mathbb{Z}$ define $a \equiv b \pmod{n}$ (in words, a is congruent to b modulo n) to mean $\exists c \in \mathbb{Z} : a - b = cn$. Thus $a \equiv b \pmod{n}$ means that a, b lie in the same coset of $n\mathbb{Z}$ in \mathbb{Z} . Congruence modulo n is an equivalence relation, and it mirrors arithmetic in \mathbb{Z}_n . In particular, if $a_1 \equiv a_2 \pmod{n}$ and $b_1 \equiv b_2 \pmod{n}$ then $a_1 + b_1 \equiv a_2 + b_2 \pmod{n}$ and $a_1 b_1 \equiv a_2 b_2 \pmod{n}$. For most of us it is easier to work with ordinary integers (as representatives of the elements of \mathbb{Z}_n) and congruence in \mathbb{Z} than with residue classes modulo n and equality in \mathbb{Z}_n .

As an example of calculation in \mathbb{Z}_n (though, as it happens, not of the use of Gauss's congruence notation), we prove the following useful fact.

PROPOSITION. $U(\mathbb{Z}_n) = \{\bar{r} \in \mathbb{Z}_n \mid \text{hcf}(r, n) = 1\}$, where \bar{r} denotes the coset $r + n\mathbb{Z}$.

Proof. We want to find those cosets $u + n\mathbb{Z}$ for which there exists a coset $v + n\mathbb{Z}$ such that $uv + n\mathbb{Z} = 1 + n\mathbb{Z}$. This is equivalent to finding those integers u such that there exist $v, w \in \mathbb{Z}$ such that $uv - nw = 1$. Clearly, then, u and n must be co-prime. Conversely, however, we know from the Euclidean Algorithm in \mathbb{Z} (part of the First-Year syllabus) that if u and n are co-prime then there will exist integers v, w such that $uv - nw = 1$, which means that $v + n\mathbb{Z}$ is inverse to $u + n\mathbb{Z}$ in \mathbb{Z}_n . Thus $U(\mathbb{Z}_n)$ consists of the cosets $r + n\mathbb{Z}$ where $0 \leq r \leq n-1$ and r is co-prime with n .

EXAMPLE. In particular, for example, $U(\mathbb{Z}_{18})$ consists of the cosets $r + 18\mathbb{Z}$ where r is divisible neither by 2 nor by 3, that is, $r \equiv 1, 5, 7, 11, 13$ or $17 \pmod{18}$.

COROLLARY. *The ring \mathbb{Z}_n is a field if and only if n is prime.*

Proof. The ring \mathbb{Z}_n is a field if and only if for $1 \leq r \leq n-1$ all the residue classes $r + n\mathbb{Z}$ are invertible. By the proposition, this is true if and only if all the integers from 1 to $n-1$ are co-prime with n , that is, n is prime.

Ring Homomorphisms

Let R, S be commutative rings. A function $\varphi : R \rightarrow S$ is said to be a *homomorphism* if

- (0) $\varphi(0) = 0, \quad \varphi(1) = 1,$
- (1) $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(-a) = -\varphi(a)$ for all $a, b \in R$, and
- (2) $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.

EXAMPLES. The identity map $\text{id}_R : R \rightarrow R$ is a homomorphism.

If R is a ring and A an ideal in R then the map $x \mapsto x + A$ is a homomorphism $R \rightarrow R/A$. It is known as the *natural projection* or *natural epimorphism*. In particular, the map $\mathbb{Z} \rightarrow \mathbb{Z}_n$ where $x \mapsto \bar{x}$ (and \bar{x} is the residue class of x modulo n) is a surjective homomorphism.

NOTE. If $\varphi : R \rightarrow S$ and $\psi : S \rightarrow T$ are ring homomorphisms then also $\psi \circ \varphi : R \rightarrow T$ is a homomorphism. You should check this.

NOTE. If $\varphi : R \rightarrow S$ is a ring homomorphism then $\varphi(U(R)) \subseteq U(S)$. For if $u \in U(R)$ then there exists $v \in R$ (in fact, of course, $v \in U(R)$) such that $uv = 1$, and then $\varphi(u)\varphi(v) = \varphi(uv) = \varphi(1) = 1$, so $\varphi(u)$ is invertible with (inverse $\varphi(v)$).

We define an *isomorphism* to be an invertible homomorphism, in other words, a homomorphism $\varphi : R \rightarrow S$ which has a two-sided inverse, that is a homomorphism $\psi : S \rightarrow R$ such that $\psi \circ \varphi = \text{id}_R$ and $\varphi \circ \psi = \text{id}_S$. We write $R \cong S$ to mean that there exists an isomorphism $R \rightarrow S$, and then R, S are said to be *isomorphic* (to each other).

NOTE. A ring homomorphism $\varphi : R \rightarrow S$ is an isomorphism if and only if it is one-one and onto (injective and surjective). The proof of this is essentially the same as the proofs of the corresponding facts for vector spaces and for groups, and we leave it as an important but easy exercise (Exercise 13—see below). Our reason for preferring the definition given is that it works for all categories of mathematical objects, whereas there are some such in which the definition “bijective structure-preserving map” does not work. For example, a bijective map from one combinatorial graph (network) to another that preserves adjacency need not be an isomorphism; a bijective map from one partially ordered set to another that preserves $<$ need not be an isomorphism; or again, the inverse of a bijective continuous map from one topological space to another need not be continuous.

EXERCISE 13: Prove that a ring homomorphism $\varphi : R \rightarrow S$ is an isomorphism if and only if it is one-one and onto.

Image and kernel of a homomorphism: Isomorphism Theorems

Let $\varphi : R \rightarrow S$ be a ring homomorphism. We define the *image* and the *kernel* of φ by

$$\text{Im } \varphi := \{y \in S \mid \exists x \in R: \varphi(x) = y\}, \quad \text{Ker } \varphi := \{a \in R \mid \varphi(a) = 0\}.$$

IMPORTANT OBSERVATION. If $\varphi : R \rightarrow S$ is a ring homomorphism then $\text{Im}\varphi$ is a subring of S and $\text{Ker}\varphi$ is an ideal in R .

Proof. Let $\varphi : R \rightarrow S$ be a ring homomorphism. Certainly $0, 1 \in \text{Im}\varphi$ since $\varphi(0) = 0$ and $\varphi(1) = 1$. If $y \in \text{Im}\varphi$ then there exists $x \in R$ such that $\varphi(x) = y$ and then $\varphi(-x) = -\varphi(x) = -y$: thus $-y \in \text{Im}\varphi$. Suppose that $y_1, y_2 \in \text{Im}\varphi$, as witnessed by elements $x_1, x_2 \in R$ for which $\varphi(x_1) = y_1$ and $\varphi(x_2) = y_2$. Then $\varphi(x_1 + x_2) = \varphi(x_1) + \varphi(x_2) = y_1 + y_2$, so $x_1 + x_2$ witnesses that $y_1 + y_2 \in \text{Im}\varphi$. Similarly, $x_1 x_2$ witnesses that $y_1 y_2 \in \text{Im}\varphi$. This shows that $\text{Im}\varphi$ is a subring of S .

Now consider $\text{Ker}\varphi$. Certainly $0 \in \text{Ker}\varphi$ since $\varphi(0) = 0$. If $a, b \in \text{Ker}\varphi$ then $\varphi(a - b) = \varphi(a) - \varphi(b) = 0 - 0 = 0$, so $a - b \in \text{Ker}\varphi$. Therefore $\text{Ker}\varphi$ is an additive subgroup of R . Also, if $a \in \text{Ker}\varphi$ and $x \in R$ then $\varphi(ax) = \varphi(a)\varphi(x) = 0\varphi(x) = 0$, so $ax \in \text{Ker}\varphi$. Thus $\text{Ker}\varphi$ is an ideal in R .

THE FIRST ISOMORPHISM THEOREM FOR RINGS. If $\varphi : R \rightarrow S$ is a ring homomorphism then $\text{Im}\varphi \cong R/\text{Ker}\varphi$.

Proof. Let $\varphi : R \rightarrow S$ be a ring homomorphism, and, to simplify notation, let $S' := \text{Im}\varphi$, $A := \text{Ker}\varphi$. We have just observed that S' is a subring of S and A is an ideal in R . We seek to define an isomorphism $\psi : R/A \rightarrow S'$. From the data available there is really only one sensible way to proceed: for each $x \in R$ define

$$\psi(x + A) := \varphi(x).$$

Since the coset representative x is not uniquely defined by the coset $x + A$ we must check that ψ is well-defined. But if $x + A = x' + A$ then $x' = x + a$ for some $a \in A$, and then $\varphi(x') = \varphi(x) + \varphi(a) = \varphi(x) + 0$, so $\varphi(x') = \varphi(x)$. This confirms that ψ is well-defined.

Next we must check that ψ is a ring homomorphism. Again, this is completely straightforward: $\psi(0) = \psi(A) = \varphi(0) = 0$; $\psi(1) = \psi(1 + A) = \varphi(1) = 1$;

$$\begin{aligned} \psi((x + A) + (x' + A)) &= \psi((x + x') + A) \\ &= \varphi(x + x') \\ &= \varphi(x) + \varphi(x') \\ &= \psi(x + A) + \psi(x' + A); \end{aligned}$$

and similarly

$$\begin{aligned} \psi((x + A)(x' + A)) &= \psi((xx') + A) \\ &= \varphi(xx') \\ &= \varphi(x)\varphi(x') \\ &= \psi(x + A)\psi(x' + A). \end{aligned}$$

Finally, we need that ψ is bijective. Suppose that $\psi(x + A) = \psi(x' + A)$. Then $\varphi(x) = \varphi(x')$, so $\varphi(x - x') = 0$. Thus $x - x' \in A$, and so $x + A = x' + A$. This shows that ψ is one-one (injective). That ψ is surjective is obvious since if $y \in S'$ then, by definition of $\text{Im}\varphi$, there exists $x \in R$ such that $\varphi(x) = y$, whence $\psi(x + A) = y$. Putting these pieces together we see that ψ is an isomorphism, so $R/\text{Ker}\varphi \cong \text{Im}\varphi$.

EXAMPLE. If R is a commutative ring of characteristic m then

$$\begin{cases} \mathbb{Z} \leq R & \text{if } m = 0, \\ \mathbb{Z}_m \leq R & \text{if } m > 0. \end{cases}$$

In particular, if R is an integral domain of characteristic p , where $p \neq 0$, then $\mathbb{Z}_p \leq R$.

Proof. Let R be a commutative ring. Recall from p.10 the definition of $n1$ for $n \in \mathbb{N}$. The map $\mathbb{Z} \rightarrow R$ given by $n \mapsto n1$ if $n \geq 0$ and $n \mapsto -(-n)1$ if $n < 0$ is easily seen to be a homomorphism [you should check this with care] and its kernel is $m\mathbb{Z}$, where $m = \text{char } R$ (note that this is one reason why we define $\text{char } R$ to be 0 if 1 has infinite additive order). Thus by the First Isomorphism Theorem $\mathbb{Z} \leq R$ if $m = \text{char } R = 0$, and otherwise $\mathbb{Z}_m \leq R$.

A WORKED EXAMPLE. (Part of Oxford FHS 1987, I, 5.) Let D be the ring of all differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ with the operations of pointwise addition and multiplication. Show that if $I := \{f \in D : f(0) = f'(0) = 0\}$ then I is an ideal in D .

Let $\mathbb{R}[x]$ denote the ring of polynomials in the indeterminate x with real coefficients, and (x^2) the ideal generated by the polynomial x^2 . Show that there is a homomorphism from $\mathbb{R}[x]$ onto D/I and deduce that $D/I \cong \mathbb{R}[x]/(x^2)$.

Suggested response. Certainly the constant function 0 lies in I . Also, if $f, g \in I$ and $h = f - g$ then $h(0) = f(0) - g(0) = 0 - 0 = 0$ and $h'(0) = f'(0) - g'(0) = 0 - 0 = 0$, so $f - g \in I$: thus I is an additive subgroup of D . Now let $f \in I$ and $g \in D$. If $h = fg$, then $h(0) = f(0)g(0) = 0g(0) = 0$ and $h'(0) = f'(0)g(0) + f(0)g'(0) = 0g(0) + 0g'(0) = 0$, and therefore $fg \in I$. This shows that I is an ideal in D .

Now let $\varphi : D \rightarrow D/I$ be the natural homomorphism, $\varphi : f \mapsto f + I$. Consider $\mathbb{R}[x]$ as a subring of D and let $\psi : \mathbb{R}[x] \rightarrow D/I$ be the restriction of φ to this subring. Our aim is to show that ψ is surjective. So let $f + I$ be any element of D/I , that is, let f be any element of D . Define $a := f(0)$, $b := f'(0)$ and let $g(x) := a + bx$. Clearly, g is a (linear) polynomial, so $g \in \mathbb{R}[x]$. Now if $h := f - g$ then $h(0) = f(0) - g(0) = a - a = 0$ and $h'(0) = f'(0) - g'(0) = b - b = 0$. Therefore $h \in I$, whence $g + I = f + I$, and so $\psi(g) = f + I$. Thus $\psi : \mathbb{R}[x] \rightarrow D/I$ is surjective.

It follows from the First Isomorphism Theorem that $D/I \cong \mathbb{R}[x]/\text{Ker } \psi$. To finish the question we must identify $\text{Ker } \psi$. But if $g(x) \in \text{Ker } \psi$ then $g(x)$ is a polynomial $c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ with the property that $g(0) = g'(0) = 0$ (because $\text{Ker } \psi \subseteq \text{Ker } \varphi = I$). Thus $c_0 = c_1 = 0$ and so $g(x) = x^2(c_2 + \cdots + c_nx^{n-2}) \in (x^2)$. Conversely of course, if $g(x) \in (x^2)$ then $g(0) = g'(0) = 0$, so $g \in \text{Ker } \psi$. Thus $\text{Ker } \psi = (x^2)$ and so $D/I \cong \mathbb{R}[x]/(x^2)$, as required.

The First Isomorphism Theorem has very many applications. In particular, just as in the case of groups, it can be used to prove further useful isomorphism theorems, and those are what we turn to now.

THE SECOND ISOMORPHISM THEOREM FOR RINGS. Let R be a commutative ring, S a subring, and A an ideal in R . Define $S + A := \{s + a \in R \mid s \in S, a \in A\}$. Then:

- (1) $S + A$ is a subring of R ;

- (2) $S \cap A$ is an ideal in S ;
- (3) $(S + A)/A \cong S/(S \cap A)$.

Proof. We leave it to the reader to check that $S + A$ is an additive subgroup of R . Now if $s_1, s_2 \in S$ and $a_1, a_2 \in A$ then $(s_1 + a_1)(s_2 + a_2) = s_1 s_2 + (s_1 a_2 + a_1 s_2 + a_1 a_2)$. Since S is a subring $s_1 s_2 \in S$ and since A is an ideal $s_1 a_2 + a_1 s_2 + a_1 a_2 \in A$. Therefore $(s_1 + a_1)(s_2 + a_2) \in S + A$, that is, $S + A$ is closed under multiplication, and hence $S + A$ is a subring of R as claimed in (1).

That $S \cap A$ is an additive subgroup of S should be clear (if not, check it). We check the second defining property of ideals: let $a \in S \cap A$, $x \in S$; then $ax \in S$ since S is a subring, and $ax \in A$ since A is an ideal of R ; thus $ax \in S \cap A$, as required for (2).

For (3) we let $\varphi : R \rightarrow R/A$ be the natural homomorphism and consider what it does to S and to $S + A$. Clearly, $\varphi(S + A) = \varphi(S) + \varphi(A) = \varphi(S) + \{0\} = \varphi(S)$, so the images of S and of $S + A$ are the same. Now consider the kernel of the restriction of φ to S . Clearly this is $\{a \in S \mid \varphi(a) = 0\}$, which is $\{a \in S \mid a \in A\}$, that is, $S \cap A$. From the First Isomorphism Theorem we deduce that $\varphi(S) \cong S/(S \cap A)$. Next consider the kernel of the restriction of φ to $S + A$. Since $A \subseteq S + A$ this clearly is A . Using the First Isomorphism Theorem again we have that $\varphi(S + A) \cong (S + A)/A$. We have seen already that $\varphi(S) = \varphi(S + A)$ and therefore $(S + A)/A \cong S/(S \cap A)$, as stated.

AN IMPORTANT EXAMPLE. Let R be a commutative ring and let $f \in R[x]$ with $\deg f \geq 1$. Then $R \leq R[x]/(f)$.

Recall that (f) denotes the principal ideal of $R[x]$ generated by f , that is, the set of all multiples of f . Note that by $R \leq R[x]/(f)$ we mean that there is a “natural” injective homomorphism $R \rightarrow R[x]/(f)$ (of course this is an isomorphism with its image). Identifying R with the set of constant polynomials we have $R \leq R[x]$. As such $R \cap (f) = \{0\}$, and so the fact that $R \leq R[x]/(f)$ comes directly from the Second Isomorphism Theorem.

THE THIRD ISOMORPHISM THEOREM FOR RINGS *Let R be a commutative ring and let A be an ideal in R .*

- (1) *If B is an ideal in R and $A \subseteq B$ then B/A is an ideal of R/A .*
- (2) *The map $B \mapsto B/A$ is a one-one correspondence (that is, a bijection) between the set of all ideals of R that contain A and the set of all ideals of R/A .*
- (3) *If B is an ideal and $A \subseteq B$ then $(R/A)/(B/A) \cong R/B$.*

Proof. Let B be an ideal of R such that $A \subseteq B$. As usual, it should be clear that B/A is an additive subgroup of R/A , so all we check is that it enjoys the second defining property of idealism. Take $b \in B$ so that $b + A$ is a typical element of B/A , and take $x \in R$ so that $x + A$ is a typical element of R/A . Then $(b + A)(x + A) = bx + A$, and, since B is an ideal, $bx \in B$, so that $bx + A \in B/A$. This shows that B/A is indeed an ideal of R/A .

Now let $\varphi : R \rightarrow R/A$ be the natural homomorphism. Let \mathcal{B} be the set of all those ideals B of R that contain A and let \mathcal{C} be the set of all ideals of R/A . We have seen that

if $B \in \mathcal{B}$ then $\varphi(B) \in \mathcal{C}$. Thus φ induces a function $\beta : \mathcal{B} \rightarrow \mathcal{C}$ and, since $\beta(B) = B/A$, our task is to prove that this is one-one and onto. We begin by proving that if $C \in \mathcal{C}$ then $\varphi^{-1}(C) \in \mathcal{B}$ (recall that $\varphi^{-1}(C)$ is defined to be the inverse image of C in R , that is $\varphi^{-1}(C) := \{x \in R \mid \varphi(x) \in C\}$). For suppose that C is an ideal of R/A . That $\varphi^{-1}(C)$ is an additive subgroup of R is routine, so, as usual, we focus on the second condition for idealism. Let $c \in \varphi^{-1}(C)$ and $x \in R$. Then $\varphi(cx) = \varphi(c)\varphi(x) \in C$ since $\varphi(c) \in C$ and C is an ideal of R/A ; therefore $cx \in \varphi^{-1}(C)$, so $\varphi^{-1}(C)$ is an ideal of R . Clearly, $\varphi^{-1}(0) \subseteq \varphi^{-1}(C)$, that is, $A \subseteq \varphi^{-1}(C)$. Therefore $\varphi^{-1}(C) \in \mathcal{B}$. Now define $\gamma : \mathcal{C} \rightarrow \mathcal{B}$ by $\gamma(C) := \varphi^{-1}(C)$. It is easy to see that if $B \in \mathcal{B}$ then $\gamma(\beta(B)) = B$ and if $C \in \mathcal{C}$ then $\beta(\gamma(C)) = C$. Thus β and γ are inverse maps and it follows that each is a bijection. In particular, β is a bijection and this is what (2) states.

We already have the natural homomorphism $\varphi : R \rightarrow R/A$; let ψ be the natural homomorphism $R/A \rightarrow (R/A)/(B/A)$, and define $\theta := \psi \circ \varphi : R \rightarrow (R/A)/(B/A)$. Since φ and ψ are surjective (onto), so is θ . Now

$$\begin{aligned} \text{Ker } \theta &= \{a \in R \mid \psi(\varphi(a)) = 0\} \\ &= \{a \in R \mid \varphi(a) \in \text{Ker } \psi\} \\ &= \{a \in R \mid \varphi(a) \in B/A\} = B, \end{aligned}$$

and it follows from the First Isomorphism Theorem that $R/B \cong \text{Im } \theta = (R/A)/(B/A)$. This proves (3) and completes the proof of the Third Isomorphism Theorem.

NOTE. As in the theorem, let R be a commutative ring and let A be an ideal in R . It is also true (1) that if S is a subring of R and $A \subseteq S$ then S/A is a subring of R/A , and (2) that the map $S \mapsto S/A$ is a bijection from the set of all subrings of R that contain A to the set of all subrings of R/A . The proof is very similar to the proof of the corresponding statements about ideals, and is left as an exercise for the reader.

Let R be a commutative ring. An ideal A is said to be *maximal* if $A \neq R$ and the only ideals B such that $A \subseteq B$ are A and R .

THEOREM. *Let R be a commutative ring and let A be an ideal of R . Then A is maximal if and only if R/A is a field.*

To prove this we'll use the following subsidiary result:

LEMMA. *Let R be a commutative ring. The ideal $\{0\}$ is maximal if and only if R is a field.*

Proof. Suppose first that R is a field, and let A be any non-trivial ideal. Choose $a \in A \setminus \{0\}$. Since R is a field a has a multiplicative inverse a^{-1} . And since A is an ideal, $1 = aa^{-1} \in A$, so for any $x \in R$, $x = 1x \in A$. Thus $A = R$, and this shows that $\{0\}$ is a maximal ideal in R .

Now suppose conversely that $\{0\}$ is a maximal ideal in R . Then $R \neq \{0\}$ since this is one of the conditions for a maximal ideal, and it follows that $0 \neq 1$. Let a be any non-zero element of R . Our task is to show that a is a unit in R . Consider the principal ideal aR . This is non-trivial since $a = a1 \in aR$. Therefore $aR = R$ since $\{0\}$ is maximal. In particular, there exists $b \in R$ such that $ab = 1$. Thus every non-zero element of R is invertible, and $0 \neq 1$, and so R is a field.

Proof of the theorem. It follows from the Third Isomorphism Theorem that A is a maximal ideal if and only if A/A , that is $\{0\}$, is maximal in R/A . Thus, by the lemma, A is a maximal ideal in R if and only if R/A is a field.

EXAMPLE. As an important example we get another proof of the corollary on p. 12 that *the ring \mathbb{Z}_n is a field if and only if n is prime*, as follows.

Suppose that n is not prime, say $n = rs$ where $r \geq 2$ and $s \geq 2$. Then $n\mathbb{Z} \subset r\mathbb{Z} \subset \mathbb{Z}$ and (as the notation is intended to convey) each of these containments is proper, so $n\mathbb{Z}$ is not maximal in \mathbb{Z} , and therefore \mathbb{Z}_n is not a field.

Suppose now that n is prime. Then, since the additive group of \mathbb{Z}_n has no non-trivial proper subgroups (a simple consequence of Lagrange's Theorem in group theory), $\{0\}$ is a maximal ideal, and so \mathbb{Z}_n is a field.

It is worth seeing, as a further example of the power of calculation with congruences and in the rings \mathbb{Z}_n , how the algebra we have been discussing may be used to prove a famous and important number-theoretic fact discovered by Pierre de Fermat around 1640 and first satisfactorily proved by Euler about a hundred years later. We begin with a preparatory lemma.

LEMMA. *Let p be an odd prime number. If $a \in \mathbb{Z}$ and $a^2 \equiv 1 \pmod{p}$ then $a \equiv \pm 1 \pmod{p}$.*

Proof. Let a be an integer such that $a^2 \equiv 1 \pmod{p}$. This means that p divides $a^2 - 1$. Now $a^2 - 1 = (a - 1)(a + 1)$ and so since p is prime, it must divide either $a - 1$ or $a + 1$. In other words, $a \equiv \pm 1 \pmod{p}$, as required.

THEOREM. *If p is an odd prime number then there exists $a \in \mathbb{N}$ such that $a^2 \equiv -1 \pmod{p}$ if and only if $p \equiv 1 \pmod{4}$.*

Proof. Let p be an odd prime number. Suppose first that $a \in \mathbb{N}$ is such that $a^2 \equiv -1 \pmod{p}$. Then $a^4 \equiv 1 \pmod{p}$ and $a^2 \not\equiv 1 \pmod{p}$, so when we think of the residue of a in \mathbb{Z}_p it is an element of the multiplicative group \mathbb{Z}_p^\times that has order 4. By Lagrange's Theorem, 4 must divide the order of this group. But $|\mathbb{Z}_p^\times| = p - 1$ (since $\mathbb{Z}_p^\times = \mathbb{Z}_p \setminus \{0\}$), so 4 divides $p - 1$, that is, $p \equiv 1 \pmod{4}$.

Now suppose conversely that $p \equiv 1 \pmod{4}$. Define $\sigma : \mathbb{Z}_p^\times \rightarrow \mathbb{Z}_p^\times$ to be the squaring map, $\sigma(x) := x^2$. Since $(xy)^2 = x^2y^2$ this is a group homomorphism. We ask first what is $\text{Ker } \sigma$? Well, $\text{Ker } \sigma = \{x \in \mathbb{Z}_p^\times \mid x^2 = 1\}$, and so, by the lemma, $\text{Ker } \sigma = \{\pm 1\}$ (note that the congruence $x \equiv \pm 1 \pmod{p}$ becomes the equality $x = \pm 1$ when we pass from congruences in \mathbb{Z} to equality in \mathbb{Z}_p). Since $|\text{Ker } \sigma| = 2$, by the First Isomorphism Theorem for groups, $|\text{Im } \sigma| = |\mathbb{Z}_p^\times|/2 = \frac{1}{2}(p - 1)$. Now recall that if G is a group of even order then there exists an element of order 2 in G , that is an element $t \in G \setminus \{1\}$ such that $t^2 = 1$. [This was Qn 1 on Exercise Sheet 8 of the course on Sets and Groups for Mods, TT 2007: recall that it may be proved by pairing elements not satisfying the equation $x^2 = 1$ with their inverses, so there are an even number of these; therefore since $|G|$ is even the number of solutions of $x^2 = 1$ must also be even, and therefore there is a non-trivial solution.] Now $\frac{1}{2}(p - 1)$ is even since $p \equiv 1 \pmod{4}$ (this is where this assumption is crucial), and so $\text{Im } \sigma$ contains an element of order 2. By the lemma,

however, -1 is the only element of order 2 in \mathbb{Z}_p^\times . It follows that $-1 \in \text{Im}\sigma$, that is, there exists $a \in \mathbb{Z}$ such that $a^2 \equiv -1 \pmod{p}$.

The Chinese Remainder Theorem

Ideals A, B of a commutative ring R are said to be *co-prime* if $A + B = R$.

EXAMPLE. The ideals $12\mathbb{Z}, 7\mathbb{Z}$ are co-prime in \mathbb{Z} . For, $1 = 12 \times 3 + 7 \times (-5)$ and so $x = 12 \times (3x) + 7 \times (-5x)$ for any $x \in \mathbb{Z}$, whence $\mathbb{Z} = 12\mathbb{Z} + 7\mathbb{Z}$.

Generally, if m, n are co-prime integers then $m\mathbb{Z}, n\mathbb{Z}$ are co-prime ideals in \mathbb{Z} . For we know as a consequence of the Euclidean Algorithm that when m, n are co-prime there exist $u, v \in \mathbb{Z}$ such that $mu + nv = 1$. Then $x = m(ux) + n(vx)$ for any $x \in \mathbb{Z}$, and this shows that $m\mathbb{Z} + n\mathbb{Z} = \mathbb{Z}$.

CHINESE REMAINDER THEOREM—GENERAL FORM. *If A, B are co-prime ideals in a commutative ring R then $R/(A \cap B) \cong R/A \times R/B$.*

Proof. Let R be a commutative ring and let A, B be co-prime ideals in R . Let $\alpha : R \rightarrow R/A$ and $\beta : R \rightarrow R/B$ be the natural homomorphisms. Define $\varphi : R \rightarrow R/A \times R/B$ by $\varphi : x \mapsto (\alpha(x), \beta(x))$ for all $x \in R$. It is easy to check that φ is a homomorphism, and this is left to the reader to do. We propose to identify $\text{Ker}\varphi$ and to show that φ is surjective. Now

$$\begin{aligned} \text{Ker}\varphi &= \{x \in R \mid \varphi(x) = (0, 0)\} \\ &= \{x \in R \mid x \in \text{Ker}\alpha \text{ and } x \in \text{Ker}\beta\} = \text{Ker}\alpha \cap \text{Ker}\beta. \end{aligned}$$

But $\text{Ker}\alpha = A$ and $\text{Ker}\beta = B$, so $\text{Ker}\varphi = A \cap B$.

To show that φ is surjective we need to prove that any pair $(\alpha(y), \beta(z)) \in R/A \times R/B$ lies in $\text{Im}\varphi$. So let $y, z \in R$. Since $A + B = R$ there exist $u \in A$ and $v \in B$ such that $u + v = 1$. Note that $\alpha(u) = 0$, $\alpha(v) = \alpha(1) - \alpha(u) = 1$, $\beta(u) = 1$ and $\beta(v) = 0$, and so $\varphi(u) = (0, 1)$ and $\varphi(v) = (1, 0)$. This suggests that we might take $x := uz + vy$. And indeed, with this choice of x we find that

$$\alpha(x) = \alpha(u)\alpha(z) + \alpha(v)\alpha(y) = 0\alpha(z) + 1\alpha(y) = \alpha(y)$$

and

$$\beta(x) = \beta(u)\beta(z) + \beta(v)\beta(y) = 1\beta(z) + 0\beta(y) = \beta(z),$$

so that $\varphi(x) = (\alpha(x), \beta(x)) = (\alpha(y), \beta(z))$, as required.

Finally, since the First Isomorphism Theorem tells us that $R/\text{Ker}\varphi \cong \text{Im}\varphi$, we have the advertised result, that $R/(A \cap B) \cong R/A \times R/B$.

APPLICATION. *If m, n are co-prime integers then $m\mathbb{Z} \cap n\mathbb{Z} = mn\mathbb{Z}$ and therefore*

$$\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n.$$

Using induction on k we see that if $n = m_1 m_2 \cdots m_k$ where m_1, m_2, \dots, m_k are pair-wise co-prime integers, then $\mathbb{Z}_n \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$. In particular, if $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, where p_1, \dots, p_k are distinct prime numbers then

$$\mathbb{Z}_n \cong \mathbb{Z}_{q_1} \times \cdots \times \mathbb{Z}_{q_k}, \quad \text{where } q_i := p_i^{\alpha_i}.$$

COROLLARY. [Chinese Remainder Theorem—classical form]: If m_1, m_2, \dots, m_k are pairwise co-prime integers and $c_1, c_2, \dots, c_k \in \mathbb{Z}$ then there exists $x \in \mathbb{Z}$ such that $x \equiv c_i \pmod{m_i}$ for $1 \leq i \leq k$.

Moreover, x is unique up to addition of multiples of $m_1 m_2 \cdots m_k$.

EXERCISE 14 [G H Hardy and E M Wright, *An Introduction to the Theory of Numbers* (5th edition, OUP, Oxford 1979) p.95]: Six professors begin courses of lectures on Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday, and announce their intentions of lecturing at intervals of two, three, four, one, six and five days respectively. The regulations of the university forbid Sunday lectures (so that a Sunday lecture must be omitted). When first will all six professors find themselves compelled to omit a lecture? [In case of emergency consult the cited reference for a solution.]

Some further exercises

EXERCISE 15: A ring (possibly non-commutative in the first instance, but with 1) in which $x^2 = x$ for all elements x is known as a *boolean algebra*. [For examples of such rings see Exercise 2 on p. 3.] Show that in a boolean algebra $2x = 0$ and $xy = yx$ for all x, y .

EXERCISE 16: Let B be a boolean algebra (for terminology see Exercise 15).

- (i) For $b, c \in B$ we define $b \leq c$ if $bc = b$. Show that this binary relation is a partial order on B . [Recall that a binary relation \leq is called a partial order if it is antisymmetric ($a \leq b$ & $b \leq a \Leftrightarrow a = b$) and transitive ($a \leq b$ & $b \leq c \Rightarrow a \leq c$).]
- (ii) Now suppose that B is finite. By an *atom* of B we mean an element of B which, in the partial order on B , is minimal subject to being non-zero: thus a is an atom if $a \neq 0$ and $0 \leq x \leq a \Rightarrow (x = 0 \text{ or } x = a)$. Prove that if $x \in B$ and a_1, \dots, a_k are the atoms a for which $a \leq x$ then $x = a_1 + \cdots + a_k$.
- (iii) Hence or otherwise show that if B is finite then B has 2^n elements for some n , and is isomorphic to the ring of all subsets of a suitable set X of size n as described in Exercise 2 (see p. 3). [Hint: take X to be the set of all atoms in B .]

EXERCISE 17: Let R be an integral domain. Prove that if $f(x), g(x) \in R[x]$ then $\deg(fg) = \deg f + \deg g$. Deduce

- (i) that $R[x]$ is an integral domain, and then deduce further that there is an infinite integral domain of characteristic 2;
- (ii) that $U(R[x]) = U(R)$ (where, recall, these are the groups of units of the rings $R[x]$ and R respectively).

EXERCISE 18: An element x of a ring R is said to be *nilpotent* if $x^k = 0$ for some positive integer k ; it is said to be *idempotent* if $x^2 = x$. Let $n \in \mathbb{N}$, $n > 1$. We abuse language a little and use m to denote both an integer m and the member of \mathbb{Z}_n that it represents.

- (i) Show that m is a unit in \mathbb{Z}_n if and only if m, n are coprime.
- (ii) Show that m is a zero-divisor in \mathbb{Z}_n if and only if m, n are not coprime.
- (iii) Identify the nilpotent elements of \mathbb{Z}_{12} .
- (iv) Identify the idempotent elements of \mathbb{Z}_{12} .

EXERCISE 19: For $n \in \mathbb{N}$ identify

- (i) the nilpotent elements in \mathbb{Z}_n , and
- (ii) the idempotent elements in \mathbb{Z}_n .

EXERCISE 20: For a ring R (commutative as always) we define

$$N(R) := \{a \in R \mid a \text{ is nilpotent}\},$$

the so-called *nilradical* of R .

- (i) Show that $N(R)$ is an ideal of R .
- (ii) Show also that $N(R/N(R)) = \{0\}$.
- (iii) Show that if $a \in N(R)$ then $1 + a \in U(R)$ and deduce that if $u \in U(R)$ and $a \in N(R)$ then $u + a \in U(R)$.

EXERCISE 21: Let R be a ring (commutative and with 1, as always) and let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. Show that $f \in U(R[x])$ if and only if $a_0 \in U(R)$ and $a_1, \dots, a_n \in N(R)$.

EXERCISE 22: Show that if p is prime and $1 \leq r \leq p-1$ then the binomial coefficient $\binom{p}{r}$ is a multiple of p .

Now let R be an integral domain of characteristic p where $p > 0$ (so, recall, p must be prime). Define $\Phi : R \rightarrow R$ by $\Phi : a \mapsto a^p$ for all $a \in R$. Show that Φ is a ring homomorphism and that $\text{Ker } \Phi = \{0\}$. Give an example where Φ is not surjective. [*Hint*: consider the polynomial ring $\mathbb{Z}_p[x]$.]

EXERCISE 23: Let R be a commutative ring with 1. An ideal A in R is said to be *prime* if $A \neq R$ and $ab \in A \Rightarrow a \in A$ **or** $b \in A$. Thus, for example, $\{0\}$ is a prime ideal if and only if R is an integral domain.

- (i) Show that an ideal A is prime if and only if R/A is an integral domain.
- (ii) Show that a maximal ideal is prime.
- (iii) Show that if A, B are prime ideals then $A \cap B$ is prime if and only if $A \subseteq B$ or $B \subseteq A$.
- (iv) Which ideals are prime in \mathbb{Z} ?

Part II: Arithmetic

We turn now to higher arithmetic—the study of divisibility and factorisation. Unless it is explicitly specified otherwise, throughout this part we assume that

R is an integral domain.

Divisibility

The element b is said to *divide* the element a in R if $\exists c \in R : a = bc$. Then we write $b|a$.

Elements $a, b \in R$ are said to be *associates* if $\exists u \in U(R) : a = ub$. Then we write $a \sim b$ (recall the note on p.6).

NOTES: (1) For $a, b \in R$, $a \sim b$ if and only if $a|b$ & $b|a$.

For, if $a \sim b$ then there exists $u \in U(R)$ such that $a = ub$, and then $b = u^{-1}a$, which shows that $a|b$ and $b|a$. Conversely, suppose that $a|b$ and $b|a$. If either a or b is 0 then the other must also be 0, so certainly $a \sim b$. So suppose that $b \neq 0$. There exist $c, d \in R$ such that $b = ad$ and $a = bc$. Therefore $b = b(cd)$, and since R is an integral domain and $b \neq 0$ this implies that $cd = 1$. Thus $c, d \in U(R)$ and so $a \sim b$.

(2) The relation \sim is an equivalence relation on R . See Exercise 9 on p.6.

(3) For $a, b \in R$, $a \sim b$ if and only if $aR = bR$. See Exercise 11 on p.11.

We are taught quite early in our school lives that in the arithmetic of \mathbb{N} a natural number > 1 is prime if it is divisible only by 1 and itself. Not much later we learn that a prime number p which divides a product rs of two natural numbers must divide either r or s . One of the subtle discoveries made by Eduard Kummer in 1845 and developed by Richard Dedekind in about 1875, both very great mathematicians, is that these two properties of a number are worth distinguishing—for elements of other domains they may be different.*

The element $a \in R$ is said to be *irreducible* if $a \neq 0$, $a \notin U(R)$, and

$$a = bc \Rightarrow b \in U(R) \text{ or } c \in U(R).$$

The element $a \in R$ is said to be *prime* if $a \neq 0$, $a \notin U(R)$, and

$$a|bc \Rightarrow a|b \text{ or } a|c.$$

LEMMA. *In an integral domain primes are always irreducible.*

Proof. Let a be a prime in the integral domain R . Suppose that $a = bc$. By the definition of prime, either $a|b$ or $a|c$. If $a|b$ then, as we have seen above (in the proof of Note (1) to the definition of divisibility) $c \in U(R)$; similarly, if $a|c$ then $b \in U(R)$. Thus a is irreducible.

*E. Kummer, 'Zur Theorie der complexen Zahlen', *Journal für die reine und angew. Math.* 35 (1847), 319–320; R. Dedekind, Eleventh Supplement to Dirichlet's *Vorlesungen über Zahlentheorie*, and *Sur la théorie des nombres entiers algébriques*, Gauthier-Villars, Paris, 1877, and other works.

IMPORTANT NOTE. In many integral domains, but not in all, irreducible elements are prime. An example where there are irreducible elements that are not prime will be discussed shortly.

A *highest common factor* (also known as *greatest common divisor*) of elements a, b of R is an element d with the properties:

$$(1) d|a \ \& \ d|b \quad \text{and} \quad (2) c|a \ \& \ c|b \Rightarrow c|d.$$

NOTE. The highest common factor of $a, b \in R$ need not exist (see the example below). Even when it does exist it is generally not unique. If d is a highest common factor of a, b then so is any associate of d ; conversely, if d_1, d_2 are highest common factors of a, b then $d_1 \sim d_2$. Nevertheless, it is conventional to abuse notation a little and write $d = \text{hcf}(a, b)$ (instead of, say, $d \sim \text{hcf}(a, b)$) to mean that d is a highest common factor of a and b .

NOTE. The general definition differs from the classical concept in \mathbb{N} , where the highest common factor is defined to be the numerically largest amongst the common factors.

EXERCISE 24. Explain why in \mathbb{N} the general version of $\text{hcf}(a, b)$ is essentially the same (the same up to sign) as the numerically largest amongst the common factors of a and b .

EXAMPLE. Let $R := \mathbb{Z}[\sqrt{-5}] := \{x + y\sqrt{-5} \mid x, y \in \mathbb{Z}\}$. Note that $R \leq \mathbb{C}$, so R is an integral domain (see Exercise 6 on p. 5). In R we have

- $6 = 2 \times 3 = (1 + \sqrt{-5}) \times (1 - \sqrt{-5})$,
- each of $2, 3, 1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ is irreducible, but they are not prime,
- 6 and $2 + 2\sqrt{-5}$ have no highest common factor.

Proof. For $a = x + y\sqrt{-5} \in R$ define $N(a) := x^2 + 5y^2$. Note that $N(a) = a\bar{a} = |a|^2$ (where bar denotes complex conjugation as usual), and therefore $N(ab) = N(a)N(b)$ for all $a, b \in R$.

Suppose that $2 = bc$ with $b, c \in R$. Then $N(b)N(c) = N(2) = 4$, and it follows that $N(b)$ is 1, 2 or 4. However, $N(b) \neq 2$ since it is obviously impossible to find $x, y \in \mathbb{Z}$ such that $x^2 + 5y^2 = 2$. If $N(b) = 1$ then $b = \pm 1$, so $b \in U(R)$. And if $N(b) = 4$ then $N(c) = 1$ and so $c \in U(R)$. This shows that 2 is irreducible in R . Very similar arguments show that $3, 1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible.

Now $2|(1 + \sqrt{-5})(1 - \sqrt{-5})$ but clearly 2 divides neither $1 + \sqrt{-5}$ nor $1 - \sqrt{-5}$ in R . Therefore 2 is not prime. Similarly, $3, 1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible elements that are not prime in R .

Finally, consider the common factors of 6 and $2 + 2\sqrt{-5}$. Amongst them are 2 and $1 + \sqrt{-5}$. If $d \in R$ is divisible by both of these then $N(d)$ must be divisible both by $N(2)$, which is 4, and by $N(1 + \sqrt{-5})$, which is 6, and therefore it is divisible by 12. On the other hand, if d is a common factor of 6 and $2 + 2\sqrt{-5}$ then $N(d)$ divides both 36 and 24. Thus if $d = \text{hcf}(6, 2 + 2\sqrt{-5})$ then $N(d) = 12$. But the equation $x^2 + 5y^2 = 12$

has no solution in integers x, y : if $y = 0$ there is no solution since 12 is not a square; if $y = \pm 1$ there is no solution since 7 is not a square; and if $|y| \geq 2$ then $x^2 + 5y^2 \geq 20$. Therefore 6 and $2 + 2\sqrt{-5}$ have no highest common factor in R .

Euclidean rings

Part of the syllabus for the Mods course studied in Hilary Term is the euclidean algorithm. This was presented in two forms, one to find the highest common factor of two integers, the other to find the highest common factor of two polynomials with real coefficients. Although polynomials and integers are vastly different, the two forms of the algorithm had much in common. In particular, they were based on division with remainder in the two domains. It is this that we propose to study and generalise to a wider context.

We begin by establishing the following terminology: A *euclidean ring* (or *euclidean domain*) is an integral domain R equipped with a function $v : R \setminus \{0\} \rightarrow \mathbb{N} \cup \{0\}$ such that

- (1) $v(ab) \geq v(a)$ for all $a, b \in R \setminus \{0\}$, and
- (2) for all $a, b \in R$, if $b \neq 0$ then there exist $q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $v(r) < v(b)$.

Clause (2) of this definition is what is known as *division with remainder*. Note that the whole idea is to be able to divide a by b and have a remainder that is ‘smaller than’ b : but what should ‘smaller than’ mean in an abstract context? It implies some notion of ‘size’, and that is what the function v , often called a *euclidean valuation*, is designed to supply.

EXAMPLE 1: $R = \mathbb{Z}$ with $v(a) = |a|$. That (1) holds is immediate from the fact that $v(ab) = |b|v(a)$ and $|b| \geq 1$. To prove (2) we let

$$S := \{m \in \mathbb{Z} \mid m \geq 0 \text{ and } m = a - nb \text{ for some } n \in \mathbb{Z}\},$$

we show that $S \neq \emptyset$, we define $r := \min S$, and we define q by the equation $r = a - qb$. Then we use subtraction to show that $0 \leq r < |b|$. Note that to *find* the quotient q and remainder r we use long division.

EXAMPLE 2: $R = F[x]$, where F is a field, with $v(f) = \deg f$. To prove division with remainder and to *find* the quotient q and remainder r we use the *Division Algorithm* for polynomials (which is a version of long division).

EXERCISE 25: Let F be a field. By an ‘arithop’ [NOT standard terminology!] we’ll mean an operation of addition, subtraction, multiplication or division of two non-zero members of F . Suppose that $f, g \in F[x] \setminus \{0\}$. Let $n := \deg f$, $m := \deg g$, and suppose that $m \leq n$.

- (i) We want to calculate $c \in F$ and a polynomial $h \in F[x]$ having the properties that $f(x) = cx^{n-m}g(x) + h(x)$ and either $h = 0$ or $\deg h \leq n - 1$. Show that this can be done using at most $2m + 1$ arithops.

- (ii) deduce that polynomials $q, r \in F[x]$ such that $f(x) = q(x)g(x) + r(x)$ and either $r = 0$ or $\deg r < m$ can be calculated at a cost of at most $(2m + 1)(n - m + 1)$ arithops.

EXAMPLE 3: Let $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}[i] := \{x + yi \mid x, y \in \mathbb{Z}\} \leq \mathbb{C}$, the ring of Gaussian integers (see p.5). Then $\mathbb{Z}[i]$ is euclidean.

Proof. For $a \in \mathbb{Z}[i] \setminus \{0\}$ define $N(a) := |a|^2$. Thus if $a = x + yi$ then $N(a) = x^2 + y^2 \in \mathbb{N}$. We propose to show that N will serve as euclidean valuation for $\mathbb{Z}[i]$. Since $N(ab) = N(a)N(b)$ and $N(b) \geq 1$, certainly $N(ab) \geq N(a)$. To see that we have division with remainder, let $a, b \in \mathbb{Z}[i]$ and suppose that $b \neq 0$. Write $a/b = \xi + \eta i$, where $\xi, \eta \in \mathbb{Q}$. There exist $u, v \in \mathbb{Z}$ such that $|u - \xi| \leq \frac{1}{2}$ and $|v - \eta| \leq \frac{1}{2}$. Define $q := u + vi \in \mathbb{Z}[i]$. Clearly then $|a/b - q|^2 \leq 1/4 + 1/4 = 1/2$, so if we define $r := a - qb \in \mathbb{Z}[i]$ then we have $a = qb + r$ and $0 \leq |r|^2 \leq \frac{1}{2}|b|^2$, that is, either $r = 0$ or $N(r) < N(b)$, as required.

EXERCISE 26: Let $R := \mathbb{Z}[\sqrt{2}] \leq \mathbb{R}$. For $a \in R \setminus \{0\}$ define $N(a) := |x^2 - 2y^2|$ where $a = x + y\sqrt{2}$. Show that N is a euclidean valuation on R .

The point of the concept of euclidean ring is to generalise basic facts about \mathbb{Z} and $F[x]$, such as that highest common factors exist and that factorisations are unique, to other useful and important integral domains: we shall derive many theorems for the price of one or two. Before we do this, however, we consider some useful little facts that follow from the definition.

FACT 1: Define $v_0 := v(1)$. Then

- (1) $v_0 \leq v(a)$ for all $a \in R \setminus \{0\}$, that is, v_0 is minimal amongst values of v ; and
- (2) $v(a) = v_0$ if and only if $a \in U(R)$.

Proof. From condition (1) for a euclidean valuation we know that $v(1a) \geq v(1)$ for all $a \in R \setminus \{0\}$, and this is the first assertion.

For the second, suppose first that $a \in U(R)$. Then there exists b such that $ab = 1$, so $v_0 = v(1) = v(ab) \geq v(a)$, and by minimality of v_0 we must have $v(a) = v_0$. Now suppose conversely that $v(a) = v_0$. Divide 1 by a : that is, write $1 = qa + r$ where either $r = 0$ or $v(r) < v(a) = v_0$. Again by minimality of v_0 , we must in fact have $r = 0$, so there exists $q \in R$ such that $1 = qa$, that is $a \in U(R)$.

FACT 2: Let $b, c \in R \setminus \{0\}$. If $c \notin U(R)$ then $v(bc) > v(b)$.

Proof. Suppose that $v(bc) \leq v(b)$. By condition (1) for a euclidean valuation it must then be the case that $v(bc) = v(b)$. Use division with remainder to write $b = q(bc) + r$, where $q, r \in R$ and either $r = 0$ or $v(r) < v(bc)$. Then $r = b(1 - qc)$ and so by condition (1) again, if $1 - qc \neq 0$ then $v(r) \geq v(b)$, which is not true. Therefore $qc = 1$ and so $c \in U(R)$. It follows that if $c \notin U(R)$ then $v(bc) > v(b)$, as we claimed.

We use this to prove the simple but very important fact that in a euclidean ring every non-zero element is (associated with) a product of irreducible elements. Note that the product of no factors is naturally taken to be 1, and so units, which are the associates

of 1, can be (and will be) thought of as products of an empty set of irreducibles. Notice also that in the following statement, if $k > 0$ then there is no need for the unit u since it could be absorbed into a_1 . Nevertheless, for purposes of induction, and when we come to compare factorisations later, we will find it useful.

OBSERVATION: *Let $a \in R \setminus \{0\}$. Then there exists $k \geq 0$, there exists $u \in U(R)$, and there exist irreducible elements a_1, a_2, \dots, a_k in R such that $a = u a_1 a_2 \cdots a_k$.*

Proof. We prove this by induction on $v(a)$. It is vacuously true if $v(a) = v_0$ (the smallest value taken by v) since then, as we have seen, $a \in U(R)$. So suppose that the assertion is known to be true for elements $x \in R \setminus \{0\}$ with $v(x) < v(a)$. If a is irreducible then it satisfies the conclusion with $k = 1$ and so there is nothing more to be done. Suppose, then, that it is reducible: that is, there exist $b, c \in R \setminus U(R)$ such that $a = bc$. Since $c \notin U(R)$, by Fact 2 we have $v(b) < v(bc) = v(a)$, and similarly, since $b \notin U(R)$ we have $v(c) < v(a)$. By our induction hypothesis there exist units $u_1, u_2 \in U(R)$ and irreducible elements b_1, \dots, b_m and c_1, \dots, c_n in R such that $b = u_1 b_1 \cdots b_m$ and $c = u_2 c_1 \cdots c_n$. But then with $u := u_1 u_2$ we have $a = u b_1 \cdots b_m c_1 \cdots c_n$, that is, a is a product of irreducible elements of R , as required.

We turn now to ideals in euclidean rings and prove the following very important fact:

PRINCIPAL IDEAL THEOREM. *Every ideal in a euclidean ring is principal.*

Proof. Let R be a euclidean ring and let $v : R \setminus \{0\} \rightarrow \mathbb{N} \cup \{0\}$ be a euclidean valuation. Let A be an ideal in R . If $A = \{0\}$ then $A = 0R$ and certainly A is principal. Suppose therefore that $A \neq \{0\}$. Then A contains non-zero elements and, since $\mathbb{N} \cup \{0\}$ is well-ordered there exists $d \in A \setminus \{0\}$ such that $v(d)$ is minimal, that is $v(d) \leq v(x)$ for all $x \in A \setminus \{0\}$. Then certainly $dR \subseteq A$ (by the second defining property of an ideal). On the other hand, if $x \in A$ then there exist $q, r \in R$ such that $x = qd + r$ and either $r = 0$ or $v(r) < v(d)$. But $r = x - qd$ and so $r \in A$. Therefore the latter possibility is excluded by the minimality of $v(d)$, and so $r = 0$. Thus $x = qd$ and this shows that $x \in dR$. Hence $A \subseteq dR$ and so in fact A is the principal ideal dR generated by d .

NOTE. An integral domain in which every ideal is principal is known as a *principal ideal domain* (PID). Thus we see that every euclidean ring is a PID.

EXAMPLE: The ring $\mathbb{Z}[x]$ is not a PID and therefore cannot be euclidean. For, let

$$A := \{a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[x] \mid a_0 \text{ is even}\}.$$

One checks very easily that A is an ideal in $\mathbb{Z}[x]$. Suppose it were principal, generated by a polynomial g . Since A contains the constant polynomial 2 we would have to have that $\deg g = 0$ and $g \mid 2$, so $g(x) = \pm 1$ or $g(x) = \pm 2$. If $g(x) = \pm 1$ then $(g) = \mathbb{Z}[x]$, so $(g) \neq A$. But if $g(x) = \pm 2$ then $g(x)$ does not divide the polynomial x which is in A . Thus A is not principal. Thus $\mathbb{Z}[x]$ is not a PID and therefore cannot be euclidean.

EXERCISE 27: Let R be an integral domain and let A_0 be an ideal in R . Define

$$A := \{a_0 + a_1 x + \cdots + a_n x^n \in R[x] \mid a_0 \in A_0\} \subseteq R[x].$$

Show that A is an ideal in $R[x]$ and that A is principal if and only if either $A_0 = \{0\}$ or $A_0 = R$. Deduce that $R[x]$ is a euclidean ring if and only if R is a field.

HIGHEST COMMON FACTOR THEOREM. *In a principal ideal domain R , and in particular in a euclidean ring R , highest common factors exist. Moreover, if $a, b \in R$ and $d = \text{hcf}(a, b)$ then there exist $u, v \in R$ such that $d = ua + vb$.*

Proof. Let $a, b \in R$. Define $A := \{xa + yb \mid x, y \in R\}$. It is easy and routine to check that A is an ideal in R . By assumption (for a general PID), or by the previous theorem (for a euclidean ring), there exists $d \in A$ such that $A = dR$. Thus every element of A is a multiple of d : but certainly $a, b \in A$ and so d is a common divisor of a and b . By definition of A there exist $u, v \in R$ such that $d = ua + vb$. Now if c divides both a and b then certainly it divides $ua + vb$, and so $c \mid d$. Thus $d = \text{hcf}(a, b)$ and the proof is complete.

THEOREM. *In a principal ideal domain R , and in particular in a euclidean ring R , irreducible elements are prime.*

Proof. Let R be a euclidean ring (or more generally a PID) and let a be an irreducible element of R . Suppose that $b, c \in R$ and $a \mid bc$. If a does not divide b then, since a is irreducible, $\text{hcf}(a, b) = 1$ and so by the previous theorem there exist $u, v \in R$ such that $ua + vb = 1$. But then $c = (uc)a + v(bc)$, and since a divides each summand, $a \mid c$. Thus we see that $a \mid b$ or $a \mid c$. Hence a is prime.

Next we show that factorisation in euclidean rings is unique. It is useful to have some more terminology. Let $a \in R$ where R is an integral domain. We'll say that factorisations

$$a = up_1 \cdots p_k = vq_1 \cdots q_m,$$

where $u, v \in U(R)$ and $p_1, \dots, p_k, q_1, \dots, q_m$ are irreducible elements of R , are *essentially the same* if $k = m$ and the elements q_i can be re-labelled so that $q_i \sim p_i$ for $1 \leq i \leq k$. We'll say that a non-zero element a of a commutative ring R is *uniquely factorisable into irreducibles* if there exists such a factorisation and all factorisations into irreducibles are essentially the same. A *unique factorisation domain* (UFD) is an integral domain in which every non-zero element is uniquely factorisable into irreducibles.

UNIQUE FACTORISATION THEOREM. *A euclidean ring is a unique factorisation domain.*

Proof. Let R be a euclidean ring and let $a \in R \setminus \{0\}$. We have already seen in the observation on p.26 that a can be factorised as a product of irreducible elements of R . The issue therefore is uniqueness. Suppose then that

$$a = up_1 \cdots p_k = vq_1 \cdots q_m,$$

where $u, v \in U(R)$ and $p_1, \dots, p_k, q_1, \dots, q_m$ are irreducible elements of R . If $k = 0$ then $a \in U(R)$, $m = 0$, and there is nothing to prove. We use induction on k therefore, and suppose that $k \geq 1$ and that products of fewer than k irreducible elements have unique factorisations. Now p_k , being irreducible, is, by the previous theorem, prime. Since it divides the product $vq_1 \cdots q_m$ it must divide one of the factors. Permuting

the q_j if necessary, we may assume that $p_k | q_m$. Then since q_m is irreducible it follows that $q_m = w p_k$ for some $w \in U(R)$. Cancelling p_k (which is permissible since R is an integral domain) we see that

$$u p_1 \cdots p_{k-1} = v' q_1 \cdots q_{m-1},$$

where $v' = v w \in U(R)$. By inductive hypothesis we then have $k - 1 = m - 1$; moreover q_1, \dots, q_{k-1} can be re-labelled so that $p_i \sim q_i$ for $1 \leq i \leq k - 1$. Since $k = m$ and $p_k \sim q_m$, the result we want is immediate.

NOTE: As a special case we get the “Fundamental Theorem of Arithmetic”, the uniqueness of factorisation in \mathbb{Z} or in \mathbb{N} . As another special case we get uniqueness of factorisation in the polynomial ring $F[x]$ where F is any field.

NOTE: In fact, although this is just a little beyond the scope of this course, every PID is a UFD. Proof of existence of a factorisation is non-trivial—and you should find it to be an interesting challenge. Proof of uniqueness, however, is exactly the same as in the case of euclidean rings.

EXERCISE 28: Let R be an integral domain with the property that every non-zero element a has a factorisation $u p_1 \cdots p_k$ where $u \in U(R)$ and p_1, \dots, p_k are irreducible elements of R . Show that R is a UFD if and only if every irreducible element of R is prime.

EXERCISE 29: We’ll call a real number a *dyadic root of 2* (not standard terminology!) if it is of the form $2^{m/2^n}$, where m, n are positive integers and m is odd. Then we’ll say that the real number a is of *dyadic root character* if $a = k_1 a_1 + \cdots + k_r a_r$ where $r \in \mathbb{N}$, $k_i \in \mathbb{Z}$ and a_i is a dyadic root of 2 for $1 \leq i \leq r$. Define

$$R := \{a \in \mathbb{R} \mid a \text{ is of dyadic root character}\}.$$

Show that R is a subring of \mathbb{R} . Show also that 2 has no factorisation as a product of irreducible elements of R .

The Euclidean Algorithm

In a UFD (and in particular in a euclidean ring) highest common factors can be described in terms of factorisations. For, given such a ring R , and given $a, b \in R$ we can let p_1, \dots, p_k be a list of the distinct irreducibles (primes) that divide either a or b and then write

$$a = u p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}, \quad b = v p_1^{\beta_1} p_2^{\beta_2} \cdots p_k^{\beta_k},$$

where $u, v \in U(R)$ and the exponents $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k$ are non-negative integers. It follows from unique factorisation that then

$$\text{hcf}(a, b) = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_k^{\gamma_k}, \quad \text{where } \gamma_i = \min(\alpha_i, \beta_i) \text{ for } 1 \leq i \leq k.$$

Unfortunately—or perhaps fortunately, since this is the foundation of much of modern cryptography—factorisation is usually impracticable. But factorisation can be avoided and this is what the Euclidean Algorithm is for.

NOTATION. Let R be a euclidean ring with euclidean valuation v , and let $a, b \in R$. If $a = qb + r$ where $r = 0$ or $v(r) < v(b)$ we write $a \div b$ for the quotient q and $\text{rema} \pmod{b}$ for the remainder r .

In some euclidean rings there is a *Division Algorithm* which provides a method whereby given a, b , division of a by b (when $b \neq 0$) to **find** $a \div b$ and $\text{rema} \pmod{b}$ can actually be achieved computationally. In \mathbb{Z} , where integers are represented in decimal notation, this is done by long division. In the polynomial ring $F[x]$ over a field F it is done by the Division Algorithm for polynomials, which is an analogue of long division. And the point is that the Euclidean Algorithm works as well in a general euclidean ring R as it does in \mathbb{N} , in \mathbb{Z} , or in $\mathbb{R}[x]$, provided that there is a Division Algorithm in R . The basis of the Euclidean Algorithm is the following very simple fact:

OBSERVATION. Let R be a euclidean ring, let $a, b \in R$, and let $r := \text{rema} \pmod{b}$. Then $\text{hcf}(a, b) = \text{hcf}(b, r)$.

Proof. Let $q := a \div b$, so that $a = qb + r$, and let $d := \text{hcf}(a, b)$, $d' := \text{hcf}(b, r)$. Since $d|a$ and $d|b$, also $d|r$, and therefore $d|d'$. Similar reasoning, however, tells us that $d|d'$. Therefore $d \sim d'$, that is, $\text{hcf}(a, b) = \text{hcf}(b, r)$.

EUCLIDEAN ALGORITHM. Let R be a euclidean ring, and assume a Division Algorithm in R .

```

INPUT: elements  $a, b$  of  $R$ ;
OUTPUT:  $\text{hcf}(a, b)$ .
BEGIN:
    Initialise  $a_1 := a, b_1 := b$ ;
    while  $b_1 \neq 0$  do
         $r := \text{rema}_1 \pmod{b_1}$     [use Division Algorithm];
         $a_1 := b_1$ ;
         $b_1 := r$ ;
    endwhile;
    return  $a_1$ ;
END.
```

Proof of termination and correctness. An algorithm is no good if it fails to finish. In this case, however, one of two things can happen in a given run of the while-loop. Either r is found to be 0, in which case b_1 becomes 0 and the loop (hence the computation) terminates, or r is non-zero, but $v(r) < v(b_1)$. Thus each run of the while-loop reduces $v(b_1)$ until b_1 becomes 0. Since $v(b_1) \in \mathbb{N}$ either the calculation finishes immediately (if $b = 0$) or it finishes after at most $n + 1$ passes of the while-loop, where $n := v(b)$.

If $b = 0$ then $\text{hcf}(a, b) = a$. At each pass of the while-loop the pair (a_1, b_1) is replaced by the pair (b_1, r) , where $r := \text{rema}_1 \pmod{b_1}$. By the preceding observation, $\text{hcf}(a_1, b_1)$ is unchanged after each pass of the while-loop (it is said to be an *invariant* of the computation). Therefore in the end, when the values of the variables a_1, b_1 are, say, $d, 0$, we have $d = \text{hcf}(d, 0) = \text{hcf}(a, b)$. Thus the algorithm not only terminates but also gives the correct answer.

EXERCISE 30: Let $f, g \in F[x]$ where F is a field, and suppose that $\deg f \leq n$ and $\deg g \leq m$. Recall from Exercise 25 (p. 24) the notion of an ‘arithop’. Using the result

of that exercise prove that the Euclidean Algorithm discovers $\text{hcf}(f, g)$ using at most $2mn + m + n + 1$ arithops.

In practice (for example, in cryptography) one wants not merely the highest common factor d of two natural numbers or polynomials a, b , but also the coefficients in an expression $d = ua + vb$ of d as a linear combination of a and b . Those can be calculated using an extension of the euclidean algorithm, sometimes called Bezout's Algorithm:

THE EXTENDED EUCLIDEAN ALGORITHM Let R be a euclidean ring, and assume a Division Algorithm in R .

```

INPUT: elements  $a, b$  of  $R$ ;
OUTPUT:  $\text{hcf}(a, b)$  and elements  $u, v \in R$  such that  $\text{hcf}(a, b) = ua + vb$ .
BEGIN:
    set  $a_1 := a, b_1 := b$ ;
    set  $u := 1, v := 0$ ;
    set  $u' := 0, v' := 1$ ;
    [COMMENT: so at this point  $a_1 = ua + vb$  and  $b_1 = u'a + v'b$ ];

    while  $b \neq 0$  do
         $q := a_1 \div b_1; r := \text{rema}_1 \pmod{b_1}$     [use Division Algorithm];
         $a_1 := b_1$ ;
         $b_1 := r$ ;
         $u_{\text{old}} := u, v_{\text{old}} := v$ ;
         $u := u'; v := v'$ ;
         $u' := qu' - u_{\text{old}}; v' := qv' - v_{\text{old}}$ ;
    endwhile;
    return  $a, u, v$ ;
END.
```

The proof of termination is the same as for the euclidean algorithm. The proof of correctness is left to the reader: note that again we have invariants of the computation— not only is $\text{hcf}(a_1, b_1)$ an invariant, but also the equations $a_1 = ua + vb, b_1 = u'a + v'b$ remain true throughout.

Some applications to polynomial rings

We begin with a very useful fact.

OBSERVATION. Let R be a euclidean ring and $a \in R$. If a is irreducible in R then $R/(a)$ is a field.

Proof. Suppose that a is irreducible in R . Let B be an ideal of R such that $(a) \subseteq B \subseteq R$. By the Principal Ideal Theorem (p.26), $B = (b)$ for some $b \in R$. Since $a \in B$ we have that $b|a$. Therefore $b \sim a$ or $b \in U(R)$. In the former case $B = (a)$, while in the latter $B = R$. Thus (a) is a maximal ideal in R and so by the theorem on p.17, $R/(a)$ is a field.

COROLLARY. *In particular, if f is irreducible in $F[x]$, where F is a field, then $F[x]/(f)$ is a field.*

THE REMAINDER THEOREM *Let F be a field, let $f \in F[x]$, and let $\alpha \in F$. Then there exists $g \in F[x]$ such that*

$$f(x) = (x - \alpha)g(x) + f(\alpha).$$

Proof. Use division with remainder in the polynomial ring $F[x]$ to write $f(x) = (x - \alpha)g(x) + r$, where either $r = 0$ or $\deg r < \deg(x - \alpha)$. Then either $r = 0$ or $\deg r = 0$; in either case r is a constant polynomial, that is, $r \in F$. Now set $x = \alpha$ to see that $r = f(\alpha)$. Thus $f(x) = (x - \alpha)g(x) + f(\alpha)$.

COROLLARY. *If $f \in F[x]$ where F is a field, and $f(\alpha) = 0$ where $\alpha \in F$, then $(x - \alpha) \mid f(x)$ in $F[x]$.*

THEOREM. *Let F be a field and let $f \in F[x] \setminus \{0\}$. Suppose that f has k distinct zeros in F . Then $k \leq \deg f$.*

Proof. The assertion is trivially true if $k = 0$, and this is the base of an inductive argument. Suppose therefore that it is true for polynomials that have $k - 1$ distinct zeros, where $k \geq 1$. Let f be a polynomial with k distinct zeros in F . Choose $a \in F$ to be one of the zeros of f . By the preceding corollary $f(x) = (x - a)g(x)$ for some $g \in F[x]$. If b is a zero of f other than a then $(b - a)g(b) = 0$, so $g(b) = 0$. Thus g has at least $k - 1$ distinct zeros in F , so by inductive hypothesis $\deg g \geq k - 1$. But $\deg g = \deg f - 1$. Therefore $\deg f \geq k$, as required.

NOTE 1. The same holds if F is replaced by an integral domain R .

NOTE 2. An illuminating alternative proof uses the so-called Vandermonde determinant:

EXERCISE 31: The Vandermonde matrix V_{n+1} of size $n + 1$ is the $(n + 1) \times (n + 1)$ matrix whose (i, j) entry is x_{i-1}^{j-1} where x_0, \dots, x_n are variables (or numbers). Prove that

$$\det V_{n+1} = \prod_{0 \leq i < j \leq n} (x_j - x_i),$$

and deduce that a non-zero polynomial of degree n over a field F can have at most n distinct roots in F .

In theory a non-zero polynomial in $F[x]$ can be expressed as a product of irreducible polynomials. In practice, given a polynomial it can be hard to find such a factorisation even when F is such a simple field as \mathbb{Q} . Gauss's Lemma is a powerful tool which is of considerable help with this problem—and others. A polynomial $c_0 + c_1x + c_2x^2 + \dots + c_nx^n$ in $\mathbb{Z}[x] \setminus \{0\}$ is said to be *primitive* if $\text{hcf}(c_0, c_1, c_2, \dots, c_n) = 1$.

GAUSS'S LEMMA, VERSION 1: *If $f, g \in \mathbb{Z}[x] \setminus \{0\}$ are primitive and $h(x) = f(x)g(x)$ then h is primitive.*

Proof. Write

$$f(x) = a_0 + a_1x + \cdots + a_mx^m \quad \text{and} \quad g(x) = b_0 + b_1x + \cdots + b_nx^n.$$

Then $h(x) = c_0 + c_1x + \cdots + c_{m+n}x^{m+n}$ where $c_k = \sum_{i+j=k} a_i b_j$. Suppose that f and g are primitive and let p be any prime number. Since p cannot divide all coefficients a_i we can choose r to be the smallest index such that $p \nmid a_r$. Similarly, let s be the smallest index such that $p \nmid b_s$. Examine c_{r+s} : we find that $p \mid a_i b_j$ if $i < r$ or $j < s$, but $p \nmid a_r b_s$, and therefore $p \nmid c_{r+s}$. Thus there is no prime number that divides every coefficient of h , and so h is primitive as Gauss's Lemma states.

We take this a step further as follows. For $f \in \mathbb{Z}[x] \setminus \{0\}$ define the *content* by

$$c(f) := \text{hcf}(a_0, a_1, \dots, a_m) \quad \text{where} \quad f(x) = a_0 + a_1x + \cdots + a_mx^m.$$

GAUSS'S LEMMA, VERSION 2: *If $f, g \in \mathbb{Z}[x] \setminus \{0\}$ then $c(fg) = c(f)c(g)$.*

Proof. We can write $f(x) = c(f)f_0(x)$ and $g(x) = c(g)g_0(x)$, where $f_0, g_0 \in \mathbb{Z}[x]$ and f_0, g_0 are primitive. Then

$$f(x)g(x) = c(f)c(g)h_0(x) \quad \text{where} \quad h_0(x) = f_0(x)g_0(x),$$

and h_0 is primitive by Version 1 of Gauss's Lemma. Therefore $c(fg) = c(f)c(g)$, as required.

GAUSS'S LEMMA, VERSION 3: *Let $f \in \mathbb{Z}[x] \setminus \{0\}$. Then f can be factorised as a product of polynomials of degrees r, s in $\mathbb{Q}[x]$ if and only if f can be factorised as a product of polynomials of degrees r, s in $\mathbb{Z}[x]$.*

Proof. This is trivial one way round, so we focus on the non-trivial—and somewhat surprising—fact that even although $\mathbb{Q}[x]$ is very much bigger than $\mathbb{Z}[x]$ it offers no real scope for more refined factorisation. Suppose that

$$f(x) = g(x)h(x) \quad \text{where} \quad g, h \in \mathbb{Q}[x] \quad \text{and} \quad \deg g = r, \deg h = s.$$

Let a be the least common multiple of the denominators of the non-zero coefficients of g and let b be the least common multiple of the denominators of the non-zero coefficients of h . Thus if $g_1(x) := ag(x)$ and $h_1(x) := bh(x)$ then $g_1, h_1 \in \mathbb{Z}[x]$. We can write $f(x) = cf_0(x)$, where $c := c(f)$ and f_0 is a primitive polynomial in $\mathbb{Z}[x]$. Then $(abc)f_0(x) = g_1(x)h_1(x)$ and, since the content of the polynomial on the left of this equation is abc while, by Gauss's Lemma, the content of the polynomial on the right is $c(g_1)c(h_1)$, these must be equal. Thus if $g_0(x) := g_1(x)/c(g_1)$ and $h_0(x) := h_1(x)/c(h_1)$, then $f_0(x) = g_0(x)h_0(x)$, and so $f(x) = (cg_0(x))h_0(x)$. This is a factorisation of f as a product of polynomials of degrees r and s respectively in $\mathbb{Z}[x]$, as required.

NOTE: All these versions of Gauss's Lemma work for an arbitrary field F and integral domain R of which F is the field of fractions, so $F = \{a/b \mid a \in R, b \in R \setminus \{0\}\}$ (see p.9), provided that arithmetic in R is well-behaved—what is required is that R is a unique factorisation domain so that highest common factors exist in R .

EXERCISE 32: Factorise $x^6 + 1$ in each of $\mathbb{Q}[x]$, $\mathbb{R}[x]$, $\mathbb{C}[x]$, $\mathbb{Z}_2[x]$, $\mathbb{Z}_3[x]$, and $\mathbb{Z}_5[x]$.

EXERCISE 33: Let f be an irreducible element of $\mathbb{Z}[x]$. Show that $f(x) = \pm p$, where p is an ordinary integer prime, or $\deg f \geq 1$, $c(f) = 1$, and f is irreducible in $\mathbb{Q}[x]$.

We return to the question with which this section began: given $f \in F[x]$, how can we factorise f ? The answer is, with great difficulty. But here are some pointers.

OBSERVATION. *If $\deg f$ is 2 or 3 then f is reducible in $F[x]$ if and only if it has a root in F .*

For, in this case f is reducible if and only if it has a factor of degree 1, and we know that this happens if and only if f has a root in F .

If F is finite and small then it is not unreasonable to use trial division. *E.g.*

$$x^7 - 1 = (x - 1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1)$$

in $\mathbb{Z}_2[x]$, and by simply trying factors of small degree we find that

$$\begin{aligned} x^7 - 1 &= (x - 1)(x^6 + x^5 + x^4 + x^3 + x^2 + x + 1) \\ &= (x - 1)(x^3 + x^2 + 1)(x^3 + x + 1) \end{aligned}$$

in $\mathbb{Z}_2[x]$.

In $\mathbb{Q}[x]$ we find that Gauss's Lemma helps greatly. *E.g.* Since $x^2 - 2$ is irreducible in $\mathbb{Z}[x]$ (easy!), it is irreducible in $\mathbb{Q}[x]$. (And this gives a proof that $\sqrt{2}$ is irrational!) Here is a challenge which illustrates that factorisation is not always straightforward, however, even with Gauss's Lemma available (and which also demonstrates the danger of misprints in mathematical formulae): show that $x^9 + x^3 + 1$ is irreducible in $\mathbb{Z}[x]$, therefore in $\mathbb{Q}[x]$. [See Herstein, *Topics in Algebra*, 1st Ed. (1964), p. 186].

The Gaussian integers

Recall from p. 5 that a *Gaussian integer* is a complex number of the form $x + yi$ where $x, y \in \mathbb{Z}$. Then $\mathbb{Z}[i]$, the set $\{x + yi \in \mathbb{C} \mid x, y \in \mathbb{Z}\}$, is a subring of \mathbb{C} and therefore an integral domain. On p. 25 we defined $N : \mathbb{Z}[i] \rightarrow \mathbb{N} \cup \{0\}$ by $N(a) := |a|^2$ for all $a \in \mathbb{Z}[i]$, so that if $a = x + yi$ then $N(a) = x^2 + y^2$; $N(a)$ is often called the (Gaussian) *norm* of a ; and N the (Gaussian) *norm* function. What we were able to show was that N is a euclidean valuation, so $\mathbb{Z}[i]$ is a euclidean ring. As it happens, N is a special kind of euclidean valuation in that it is *multiplicative*: $N(ab) = N(a)N(b)$ for all $a, b \in \mathbb{Z}[i]$. This considerably simplifies arithmetic in $\mathbb{Z}[i]$. What we want to know is:

- What are the units in $\mathbb{Z}[i]$?
- What are the primes in $\mathbb{Z}[i]$?
- How can we factorise in $\mathbb{Z}[i]$?

We begin with the Gaussian units:

THEOREM: $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$.

Proof. It is clear that these four numbers are units in $\mathbb{Z}[i]$, so our task is to prove that there are no more. Let $u \in U(\mathbb{Z}[i])$, and let $v := u^{-1} \in \mathbb{Z}[i]$, so that $uv = 1$. Then $N(uv) = N(1) = 1$, so $N(u)N(v) = 1$. It follows that $N(u) = 1$ and so if $u = x + yi$ with $x, y \in \mathbb{Z}$, then $x^2 + y^2 = 1$. This equation obviously only has the solutions $x = \pm 1, y = 0$ and $x = 0, y = \pm 1$. Therefore $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$ as the theorem states.

Next we seek the Gaussian primes.

LEMMA: *Let a be a prime in $\mathbb{Z}[i]$. Then there is a prime p in \mathbb{N} such that $a|p$ in $\mathbb{Z}[i]$. Moreover, either $N(a) = p$ or $a = up$ for some $u \in U(\mathbb{Z}[i])$.*

Proof. Factorise $N(a)$ as a product of ordinary primes in \mathbb{N} . Since $N(a) = a\bar{a}$ in $\mathbb{Z}[i]$ and a is prime in this ring, a must divide one of the ordinary prime factors p of $N(a)$. This is the first assertion of the lemma.

Now let p be a prime in \mathbb{N} such that $a|p$ in $\mathbb{Z}[i]$, say $p = ab$ with $b \in \mathbb{Z}[i]$. Taking norms we find that $p^2 = N(a)N(b)$. Since $N(a) > 1$ it follows that either $N(a) = p$ or $N(a) = p^2$. In the latter case $N(b) = 1$, and so b is a unit. This completes the proof.

This lemma tells us that to find the Gaussian primes we should study ordinary primes in \mathbb{N} and seek their factorisations in $\mathbb{Z}[i]$. Here is what emerges:

THEOREM: *Let p be an ordinary prime number in \mathbb{N} .*

- *If $p = 2$ then $p = (-i)(1 + i)^2$.*
- *If $p \equiv 3 \pmod{4}$ then p remains prime in $\mathbb{Z}[i]$.*
- *If $p \equiv 1 \pmod{4}$ then p becomes reducible in $\mathbb{Z}[i]$ —in fact p factorises as a product of two distinct primes in $\mathbb{Z}[i]$.*

Proof. The fact that $2 = (-i)(1 + i)^2$ is a simple calculation with complex numbers, and then we see that $1 + i$ is a Gaussian prime since its norm is 2, which is prime in \mathbb{N} .

Suppose that p is reducible in $\mathbb{Z}[i]$. This means that $p = ab$ where a and b are non-units in $\mathbb{Z}[i]$. Taking norms we see that $p^2 = N(a)N(b)$, and since $N(a) > 1$ and $N(b) > 1$ it follows that $N(a) = N(b) = p$. Writing $a = x + yi$, we have $p = x^2 + y^2$. Every square is congruent to 0 or 1 modulo 4, and so $p \not\equiv 3 \pmod{4}$. Turning this around: if $p \equiv 3 \pmod{4}$ then p must be irreducible in $\mathbb{Z}[i]$, hence prime.

Now suppose that $p \equiv 1 \pmod{4}$. We have seen in the theorem on p.18 that there exists $x \in \mathbb{Z}$ such that $x^2 + 1 \equiv 0 \pmod{p}$. Then p divides $(x + i)(x - i)$, in $\mathbb{Z}[i]$ but it clearly divides neither factor. Therefore p is not prime in $\mathbb{Z}[i]$, and so, since this is a euclidean ring, by the theorem on p.27, p must be reducible. As we have just seen, it then follows that $p = x^2 + y^2 = (x + yi)(x - yi)$ for suitable natural numbers x, y and it is easy to see that these cannot be associates of each other. Thus in this case p is a product of two distinct primes in $\mathbb{Z}[i]$.

As an immediate corollary we have the following information:

THEOREM: *The primes in $\mathbb{Z}[i]$ are (associates of):*

- $1 + i$;
- primes p of \mathbb{N} of the form $4m + 3$; and
- numbers $x + yi$ where $x, y \in \mathbb{N}$ and $x^2 + y^2$ is prime.

EXAMPLES: $1 + i, 3, 2 + i, 2 - i, 7, 11, 3 + 2i, 3 - 2i, 4 + i, 4 - i, 19, 23, \dots$ are primes in $\mathbb{Z}[i]$.

EXERCISE 34: Factorise $2, 5, 8 - i, 3 + 15i$ in $\mathbb{Z}[i]$.

A classical application is to identify those positive integers which may be written as a sum of two squares:

FERMAT'S TWO SQUARES THEOREM: *Every ordinary prime of the form $4m + 1$ is a sum of two squares.*

THEOREM: *Let $n \in \mathbb{N}$. Factorise n as $p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$ where p_1, \dots, p_k are distinct prime numbers. There exist $x, y \in \mathbb{N} \cup \{0\}$ such that $n = x^2 + y^2$ if and only if m_i is even whenever $p_i \equiv 3 \pmod{4}$.*

Proof. Suppose first that n is a sum $x^2 + y^2$ of two squares. Factorise $x + yi$ as $q_1 q_2 \cdots q_r$ where q_1, q_2, \dots, q_r are primes in $\mathbb{Z}[i]$. Then $x - yi = \bar{q}_1 \bar{q}_2 \cdots \bar{q}_r$. Without loss of generality we may suppose that q_1, \dots, q_s are ordinary primes $\equiv 3 \pmod{4}$ and q_{s+1}, \dots, q_r are such that $N(q_i)$ is prime in \mathbb{N} and not $\equiv 3 \pmod{4}$. Then $n = x_1^2 \cdots x_s^2 N(x_{s+1}) \cdots N(x_r)$. We read off from this that if p_i is a prime divisor of n such that $p_i \equiv 3 \pmod{4}$ then its exponent m_i must be even.

Suppose conversely that p_1, \dots, p_t are 3 modulo 4 and m_1, \dots, m_t are even, and that p_{t+1}, \dots, p_k are not 3 modulo 4. We know then that there are Gaussian primes q_{t+1}, \dots, q_k such that $p_i = N(q_i)$ for $t + 1 \leq i \leq k$. Define

$$n_0 := p_1^{m_1/2} \cdots p_t^{m_t/2} q_{t+1}^{m_{t+1}} \cdots q_k^{m_k}.$$

Then $n = n_0 \bar{n}_0$ and so if $n_0 = x + yi$ then $n = x^2 + y^2$.

EXERCISE 35: Which of the numbers between 100 and 120 can be expressed as sums of two squares?

Further exercises

EXERCISE 36: Let $\omega \in \mathbb{C}$ be a primitive cube root of 1. Thus $\omega^2 + \omega + 1 = 0$. Define $\mathbb{Z}[\omega] := \{a + b\omega \mid a, b \in \mathbb{Z}\}$. Show that $\mathbb{Z}[\omega] \leq \mathbb{C}$.

EXERCISE 37: Which of the following are euclidean rings? As always, justify your answers.

- (i) $\mathbb{Q}[x, y]$, the ring of polynomials in two variables with rational coefficients;
- (ii) $\mathbb{Z}[\sqrt{-5}]$;
- (iii) $\mathbb{Z}[\omega]$. [*Hint*: show that if $N(a)$ is defined for $a \in \mathbb{Z}[\omega]$ by $N(a) := |a|^2$ then N is a euclidean valuation on $\mathbb{Z}[\omega] \setminus \{0\}$.]

EXERCISE 38: Consider the quadratic polynomial $q(x) = ax^2 + bx + c \in \mathbb{R}[x]$, where $a \neq 0$.

- (i) Show that q is irreducible in $\mathbb{R}[x]$ if and only if $b^2 < 4ac$.
- (ii) Show that if $b^2 < 4ac$ then $\mathbb{R}[x]/(q) \cong \mathbb{C}$. [*Hint*: let α be a root of q in \mathbb{C} and consider the evaluation map $\varphi : \mathbb{R}[x] \rightarrow \mathbb{C}$ given by $\varphi(f) := f(\alpha)$. Show that φ is a ring homomorphism, that it is surjective, and that its kernel is (q) ; then use the First Isomorphism Theorem.]

EXERCISE 39: The so-called Fundamental Theorem of Algebra (which is actually a theorem of Analysis much more than Algebra) states that every non-constant polynomial in $\mathbb{C}[x]$ has a root in \mathbb{C} . Show that this assertion is equivalent to the assertion that if f is a polynomial with real coefficients and f is irreducible in $\mathbb{R}[x]$ then $\deg f$ is 1 or 2.

EXERCISE 40: Let $R := \mathbb{Z}[i]$, let p be an ordinary prime number, and let A be the principal ideal (p) generated by p in R .

- (i) Show that if $p \equiv 3 \pmod{4}$ then R/A is a field with p^2 elements.
- (ii) Show that if $p \equiv 1 \pmod{4}$ then $R/A \cong \mathbb{Z}_p \times \mathbb{Z}_p$.
[*Hint*: use the fact that p is reducible in R , together with the Chinese Remainder Theorem.]
- (iii) What can you say about R/A when $p = 2$?

E N D