Chapter 1

Primes and factorization

This chapter gives a brief introduction to some ideas in algebra and number theory. The central objects of study in number theory are the integers $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$, and the rational numbers $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}, b \neq 0\}$. These objects are among the simplest examples of algebraic structures that we shall study throughout the book. The integers are one of the simplest examples of a group (under the operation of addition), and they also form a ring under addition and multiplication. The rationals are one of the simplest examples of a field. We shall begin by defining these notions carefully. Our immediate goal after that will be to discuss prime numbers and factorization. In particular, we shall show that integers admit a unique factorization into prime numbers, but we will develop the notions and proofs so that they generalize to other interesting rings as well—for example, to the ring of Gaussian integers $\mathbb{Z}[i]$, and to the ring of polynomials over a field (both defined below).

1.1. Groups

Definition 1.1. A *group* is a set G with a binary operation, denoted $\cdot$ (or $*$, or $+$, or $\times$, or just omitted), satisfying the following properties:

- If a and b are in G then $a \cdot b$ is also in G.
- Associativity: For any a, b, c in G we have

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

• There is an identity element (denoted e) with the property that for any $a \in G$ one has

$$a \cdot e = e \cdot a = a.$$

• For every $a \in G$ there is an inverse element a^{-1} such that

$$a \cdot a^{-1} = a^{-1} \cdot a = e.$$

Note that in our definition we do not insist that $a \cdot b = b \cdot a$ for all a and b. Groups in which $a \cdot b = b \cdot a$ are called *commutative* (or *abelian*) groups.

In our definition of a group, we only required the existence of an identity element e, but in fact one can see that such an identity element must be unique. For, if e_1 and e_2 were two identity elements for a group G, then we must have $e_1 \cdot e_2 = e_1$ (since e_2 is an identity), and also that $e_1 \cdot e_2 = e_2$ (since e_1 is an identity), and therefore $e_1 = e_2$. Similarly you should check that there is a unique inverse for any element $a \in G$ (see Exercise 1(i) below).

Another useful property that follows from the definition is the *cancellation law*. If a, b, c are any elements of a group G with $ab = ac$, then we can "cancel a on both sides" and conclude that $b = c$. Precisely, multiply both sides of the relation $ab = ac$ (on the left) with a^{-1}, obtaining $a^{-1}(ab) = a^{-1}(ac)$. Using the associative property we find $a^{-1}(ab) = (a^{-1}a)b = eb = b$ and similarly $a^{-1}(ac) = (a^{-1}a)c = ec = c$, and thus the cancellation law is justified.

Example 1.2. The set of integers $\mathbb{Z}$ with the usual addition operation forms an abelian group. The identity is 0 and the inverse of a number n is $-n$.

The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are all examples of abelian groups under the usual addition operation. The non-zero rational numbers (denoted $\mathbb{Q}^\times$), non-zero real numbers $\mathbb{R}^\times$, and non-zero complex numbers $\mathbb{C}^\times$ are groups under the usual multiplication operation (with the identity being 1 now).

Example 1.3. Let G be a group with the operation denoted by $\cdot$. If g is an element of G, then we can "multiply" g with itself (precisely, we are using the operation $\cdot$ on g repeatedly), arriving at elements $g \cdot g$ which we denote naturally by g^2, $g \cdot g \cdot g = g^3$, and so on. Considering also the inverse of g, namely g^{-1}, we are led to elements g^{-2}, g^{-3}, and so on.

Note that the inverse of g^n is simply g^{-n}, and that the "law of exponents" holds: $g^i \cdot g^j = g^{i+j}$. Consider the set $H = \{g^n : n \in \mathbb{Z}\}$, which is a subset of G, and in fact is also a group in its own right under the same operation $\cdot$ (check that the properties in the definition hold). The set H is an example of a *subgroup* of G, and is known as the *cyclic group* generated by the element g.

For example, if we consider the group $\mathbb{Z}$ under addition, then the subgroup generated by 2 consists of all the even numbers $\{2n : n \in \mathbb{Z}\}$. For other examples, consider the group $\mathbb{C}^*$ of non-zero complex numbers under multiplication. The group generated by π consists of the infinite set $\{\pi^n : n \in \mathbb{Z}\}$. We can also obtain finite subgroups: the group generated by 1 is simply $\{1\}$, while the group generated by -1 has two elements $\{1, -1\}$. More generally, for any natural number n we can start with the complex number $e^{2\pi i/n}$ (an n-th root of unity), and this generates the n-element group $\{e^{2\pi i/n}, e^{4\pi i/n}, e^{6\pi i/n}, \ldots e^{2\pi in/n} = 1\}$. This gives one way of thinking about the cyclic group of size n.

Example 1.4. While we will only be concerned with the simplest groups (like $\mathbb{Z}$) and most of our discussions will involve abelian groups, we give a few important examples of non-abelian groups. As one example of a group that is not abelian, (and which you might have encountered before in linear algebra) look at 2×2 matrices with real entries and determinant not equal to zero. The group operation here is matrix multiplication, and the identity element of the group is the identity matrix. The condition that the determinant is not zero allows one to invert matrices. This group is denoted as $GL_2(\mathbb{R})$ (here GL stands for General Linear), and you can similarly think of $n \times n$ matrices with real entries and non-zero determinant obtaining the group $GL_n(\mathbb{R})$. Another related example is to look at $n \times n$ matrices with real entries and determinant equal to 1, and again with matrix multiplication as the group operation—this group is denoted by $SL_n(\mathbb{R})$ (with SL standing for Special Linear, and "special" indicating here the specification that the determinant is 1). A third example is the symmetric group S_n of all permutations of an n-element set (usually thought of as $\{1, 2, \ldots, n\}$). By a "permutation" we mean a bijective function on the n-element set, and the group operation here is composition of functions. You may have encountered permutations while discussing the determinant in linear algebra.

1.2. Rings

Definition 1.5. A *ring* R is a set together with two binary operations, usually denoted by $+$ and $\times$, and satisfying the following properties:

- Under the operation $+$, the set R forms an abelian group. The (additive) identity of this group is denoted by 0.
- The operation $\times$ is associative $a \times (b \times c) = (a \times b) \times c$.
- Multiplication is distributive over addition:

$$a \times (b + c) = a \times b + a \times c, \qquad \text{and} \qquad (a + b) \times c = a \times c + b \times c.$$

Two other desirable properties, which need not be satisfied by general rings, are:

- Commutativity of multiplication: $a \times b = b \times a$.
- Existence of a multiplicative identity: There exists an element 1 with $a \times 1 = 1 \times a = a$ for all $a \in R$.

A ring which satisfies the last two properties above is called a commutative ring with identity. We will only be interested in such commutative rings with identity, but it may be useful to have one example of a non-commutative ring. A natural example, related to Example 1.4 for groups, is the ring $M_n(\mathbb{R})$ of $n \times n$ matrices with real entries with the usual operations of matrix addition and multiplication.

From now on, ring will always mean, for us, a commutative ring with identity. We will remind you of this assumption from time to time, but it is assumed throughout the text.

In any ring R, $0 \times a = 0$ for all $a \in R$. To see this, note that $0 \times a = (0 + 0) \times a = 0 \times a + 0 \times a$ by the distributive law. Canceling one $0 \times a$ from both sides of the relation $0 \times a = 0 \times a + 0 \times a$ (recall that we are allowed to cancel in a group), we obtain $0 \times a = 0$.

Example 1.6. If the multiplicative identity 1 is the same as the additive identity 0, then the ring can have only one element 0: indeed, we must have $1 \times a = a = 0 \times a = 0$. This is a trivial example of a ring, called the *zero ring*; it consists of one element 0, and is described by the boring properties $0 + 0 = 0 \times 0 = 0$. We shall henceforth assume that $0 \neq 1$, to avoid this example.

Example 1.7. Note that $\mathbb{Z}$ is a commutative ring with identity for the usual addition and multiplication operations. The additive identity is 0 and the multiplicative identity is 1.

Example 1.8. The Gaussian integers are defined by $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$, and this forms a commutative ring with identity under the usual operations of addition and multiplication. Precisely, here i is a symbol denoting $\sqrt{-1}$, so that any occurrence of $i \times i$ may be replaced with -1. Adding $a+bi$ to $c+di$ results in $(a+b)+(c+d)i$, and multiplying $(a + bi)$ and $(c + di)$ results in $ac + adi + bci + bdi \times i$ (as demanded by the distributive law), which simplifies to $(ac - bd) + (ad + bc)i$.

Similarly you may check that $\mathbb{Z}[\sqrt{-5}] = \{a+b\sqrt{-5} : a, b \in \mathbb{Z}\}$ (for example) is also a ring, under usual addition and multiplication. Again, think of $\sqrt{-5}$ as standing for some symbol which when multiplied with itself yields -5. Later on, while discussing quotient rings, we shall give a more precise description of what exactly we are doing in these two examples.

Example 1.9. You may have seen something about congruences in the integers, which we will discuss in more detail and generality later. Let $n \geq 2$ be a natural number. We say that two integers a and b are congruent mod n if n divides their difference $a - b$. By a congruence class a mod n we mean the set of all integers that are congruent to a mod n. Any integer lies in precisely one of the congruence classes 0 mod n, 1 mod n, …, $n - 1$ mod n (the remainder obtained upon dividing by n). These n congruence classes inherit operations + and × from addition and subtraction in the integers. By this we mean that if we add any two integers in the congruence classes a mod n and b mod n, then we will obtain an integer in the congruence class $(a + b)$ mod n; and similarly if we multiply two such integers, we would obtain an integer in the congruence class ab mod n. The ring obtained in this way is denoted by $\mathbb{Z}/n\mathbb{Z}$ and is a finite ring of size n.

As mentioned already, we will discuss this notion more precisely later (see Section 3.2), and work towards understanding the structure of this ring. For the present, you may wish to consider special cases such as $n = 2$, 3, or 6 and check how the ring operations work in these cases.

Example 1.10. Given a ring R, we can form an important example of a ring by considering polynomials in a variable x with coefficients in the ring R. This is known as the *polynomial ring over R* and is denoted by

$R[x]$. The elements of $R[x]$ are polynomials of the form $f(x) = a_0 + a_1x + \ldots + a_nx^n$, where n is a non-negative integer, and a_0, …, a_n are elements of the ring R. Usually one has in mind that $a_n \neq 0$, so that x^n is the leading power of x in the polynomial $f(x)$, but be careful to allow for the zero polynomial $f(x) = 0$ where all the coefficients are 0. If $g(x) = b_0+b_1x+\ldots+b_mx^m$ is another polynomial with coefficients in R, then their sum $(f+g)$ is defined as the polynomial $(f+g)(x) = \sum_j c_jx^j$ with $c_j = a_j + b_j$ (with the understanding that $a_j = 0$ for $j > n$, and $b_j = 0$ for $j > m$); although we haven't specified the range of values for j, clearly $c_j = 0$ if $j > \max(m, n)$. Similarly the product of the two polynomials f and g is given by

$$(fg)(x) = a_0b_0 + (a_1b_0 + a_0b_1)x + \ldots + a_nb_mx^{m+n}.$$

You would already be familiar with polynomials whose coefficients are real numbers (the polynomial ring $\mathbb{R}[x]$) or complex numbers (the ring $\mathbb{C}[x]$), and we can now consider further examples such as $\mathbb{Z}[x]$, or the more exotic $(\mathbb{Z}/6\mathbb{Z})[x]$.

1.3. Integral domains and fields

Let R be a ring (as always, commutative with identity and with $0 \neq 1$). Since R forms a group under addition, we have the cancellation law $a + b = a + c$ implies $b = c$. Is there a cancellation law for multiplication? Since $0\times a = 0$ for all elements $a \in R$, we may have $0\times b = 0\times c$ without necessarily having $b = c$. Less trivially, even if $a \neq 0$ it may happen that $ab = ac$ without b being equal to c. For example, in the ring $\mathbb{Z}/6\mathbb{Z}$ we have $2 \bmod 6 \times 3 \bmod 6 = 4 \bmod 6 \times 3 \bmod 6$ (both are $0 \bmod 6$) but $2 \bmod 6 \neq 4 \bmod 6$. The problem is that it is possible for rings R to have non-zero elements a and b such that the product ab equals 0. Indeed in $\mathbb{Z}/6\mathbb{Z}$ we have $2 \bmod 6 \times 3 \bmod 6 = 0 \bmod 6$. We isolate this undesired behavior, and define a class of rings that are better behaved and permit cancellation with respect to multiplication.

Definition 1.11. Let R be a commutative ring with identity, and with $0 \neq 1$. A non-zero element a of R is called a *zero divisor* if there is a non-zero element b with $ab = 0$. A ring R that has no zero divisors is called an *integral domain*.

Example 1.12. The ring $\mathbb{Z}$, and the ring of Gaussian integers $\mathbb{Z}[i]$ are both integral domains. To see why $\mathbb{Z}[i]$ is an integral domain, note that

$(a+bi)\times(c+di) = 0$ implies that $(a-bi)(a+bi)(c+di)(c-di) = (a^2+b^2)(c^2+d^2) = 0$. The last relation gives that a product of non-negative integers is 0, so that either $a^2+b^2 = 0$ (so that $a = b = 0$) or $c^2+d^2 = 0$ (so that $c = d = 0$).

Lemma 1.13. *Let R be an integral domain, and let a be a non-zero element of R. If $ab = ac$ then $b = c$.*

Proof. Rewrite the relation $ab = ac$ as $ab-ac = 0$, or $a(b-c) = 0$ (here by $b-c$ we naturally mean $b+(-c)$). Since R is an integral domain, the relation $a(b-c) = 0$ implies that either $a = 0$ or $b-c = 0$. By assumption $a \neq 0$, and so we must have $b-c = 0$, or $b = c$. □

Note that this proof is different from that of the cancellation law in a group, because the element $a \in R$ may not have a multiplicative inverse; nevertheless, ruling out zero divisors is sufficient to make the cancellation law work.

Definition 1.14. Let R be a ring, and f be a non-zero polynomial in $R[x]$. Write $f = a_0 + a_1x + \ldots + a_nx^n$, with $a_n \neq 0$. Then we call n the *degree* of the polynomial f, and denote it by $\deg(f)$. Note that the degree of the zero polynomial is left undefined; one convention is to define it to be $-\infty$.

We may expect that if two non-zero polynomials f and g are multiplied, then the degree of fg should be the sum of the degree of f and the degree of g. But this may fail owing to zero divisors in R: for example the polynomials $2x$ and $3x$ in $(\mathbb{Z}/6\mathbb{Z})[x]$ both have degree 1, but their product is the zero polynomial of undefined degree. For integral domains, our expectation about degrees is true.

Proposition 1.15. *Let R be an integral domain. Then the polynomial ring $R[x]$ is also an integral domain. Moreover, if f and g are non-zero polynomials in $R[x]$ then the degree of fg equals the sum of the degrees of f and g.*

Proof. Let f be a non-zero polynomial. Then we may write $f(x) = a_nx^n + a_{n-1}x^{n-1} + \ldots + a_0$, where $a_j \in R$, and $a_n \neq 0$. The degree of f is then n, which is a non-negative integer. Let g be another non-zero polynomial $Q(x) = b_mx^m + \ldots + b_0$ with $b_m \neq 0$, so that g has degree m. Then $f(x)g(x) = a_nb_mx^{m+n}$ + lower powers of x, and since R is an

integral domain $a_n b_m \neq 0$. Thus $f(x)g(x)$ is a non-zero polynomial of degree $m+n$, proving our proposition. □

Example 1.16. Thus $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, $\mathbb{C}[x]$, $\mathbb{Z}[x,y] = (\mathbb{Z}[x])[y]$ are all integral domains.

Definition 1.17. In a ring R, elements that have multiplicative inverses are called *units*. Thus, $u \in R$ is a unit if there exists $v \in R$ with $uv = 1$.

Since $0 \times a = 0$ for all $a \in R$, we cannot expect 0 to be a unit (recall that we are ignoring the zero ring where $0 = 1$). Further, if a is a zero divisor, then a cannot be a unit. To see this, suppose there is a multiplicative inverse a^{-1} of a (thus $aa^{-1} = 1$), and also a non-zero element $b \in R$ with $ab = 0$. Then we must have $0 = a^{-1} \times 0 = a^{-1} \times ab = b$, which contradicts b being non-zero.

Check that the units of a ring R (always commutative with identity) form a group under multiplication: this group is denoted by $R^\times$.

Example 1.18. The units of $\mathbb{Z}$ are just ± 1. For instance, we may see this by considering the size, or absolute value, of integers. If a is a non-zero integer, with an inverse a^{-1}, then $1 = aa^{-1}$ and so $1 = |a| \times |a^{-1}|$. Thus either a or a^{-1} must have absolute value at most 1, but the only non-zero integers with absolute value at most 1 are ± 1.

The units in the Gaussian integers $\mathbb{Z}[i]$ are ± 1, and $\pm i$. Again we may see this by using a notion of size, or absolute value; this time using the absolute value of complex numbers, or more precisely the square of the absolute value in the complex numbers. Consider $N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$ defined by $N(a+bi) = a^2 + b^2$, and N is often called a norm function. You should check that the norm is multiplicative, by which we mean that $N(\alpha\beta) = N(\alpha)N(\beta)$. Therefore if u is a unit then $N(u) = 1$ (for $uv = 1$ and so $N(u)N(v) = 1$, and both $N(u)$ and $N(v)$ are non-negative integers). Since $a^2 + b^2 = 1$ only if $a = \pm 1$ and $b = 0$, or $a = 0$ and $b = \pm 1$, it follows that the units in $\mathbb{Z}[i]$ are ± 1 and $\pm i$.

Definition 1.19. A *field* is an integral domain R where all non-zero elements are units.

Example 1.20. You would already be familiar with the field of rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$.

Example 1.21. A less familiar example may be

$$\mathbb{Q}(i) = \{a + bi : \quad a, b \in \mathbb{Q}\}.$$

You should check that this is a field (see Exercise 7 below), and note that this field bears the same relation to the ring of Gaussian integers $\mathbb{Z}[i]$ that the field of rational numbers $\mathbb{Q}$ bears to the ring $\mathbb{Z}$. Recall $\mathbb{Q}$ is obtained from $\mathbb{Z}$ by considering *fractions* a/b (with $a, b \in \mathbb{Z}$, and $b \neq 0$) with the understanding that two fractions a_1/b_1 and a_2/b_2 are equal if $a_1b_2 = a_2b_1$. Similarly $\mathbb{Q}(i)$ may be obtained from $\mathbb{Z}[i]$ by considering fractions $(a + bi)/(c + di)$, with $c + di \neq 0$.

More generally, starting with an integral domain R we may construct a *field of fractions* by considering expressions a/b with $a, b \in R$ and $b \neq 0$, with the understanding that a_1/b_1 and a_2/b_2 are the same if $a_1b_2 = a_2b_1$ (as in the familiar example $\mathbb{Q}$). One adds and multiplies such fractions in the usual way $a_1/b_1 + a_2/b_2 = (a_1b_2 + a_2b_1)/(b_1b_2)$, and $a_1/b_1 \times a_2/b_2 = (a_1a_2)/(b_1b_2)$.

You may be familiar with another example of this construction: Starting with the polynomial ring $\mathbb{R}[x]$, which is an integral domain, we obtain the field of *rational functions* $\mathbb{R}(x)$ which consists of expressions $f(x)/g(x)$ where f, g are elements of $\mathbb{R}[x]$ with $g \neq 0$.

Example 1.22. Check that $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are fields, but $\mathbb{Z}/6\mathbb{Z}$ is not a field (indeed, it is not an integral domain). These give our first examples of finite fields, and one of our goals in this book is to determine and describe all such finite fields.

Example 1.23. If $\mathbb{F}$ is a field, then the units of the polynomial ring $\mathbb{F}[x]$ are the non-zero constants in $\mathbb{F}$.

1.4. Divisibility: primes and irreducibles

With these preliminaries in place, we turn to the main goal of this chapter, which is to develop ideas of divisibility and factorization in rings, generalizing the familiar notion of prime numbers in the integers and the factorization of integers into prime numbers. Let us begin with the definition (and notation) for divisibility.

Definition 1.24. Let R be a ring, and let a and b be elements of R. We say that *a divides b*, and write $a|b$, if there is an element $c \in R$ such that $b = ac$.

Example 1.25. Since all our rings have a multiplicative identity 1, note that $a|a$ for any $a \in R$. If $a|b$ and $b|c$ then check that $a|c$. Further note that $a|0$ for any $a \in R$.

Example 1.26. If a in R is a unit, then $a|b$ for any $b \in R$ (since we can write $b = a(a^{-1}b)$). This remark implies that the notion of divisibility is not interesting in a field. Indeed, in a field every non-zero element is a unit, and therefore all non-zero elements divide all elements of a field.

Example 1.27. A natural question that arises from our definition is whether c is unique when we write $b = ac$. Note that if $a = 0$, then b must also be 0, but c may be an arbitrary element of the ring. Let us avoid this pathological case, and ask what happens when $a \neq 0$. Consider the ring $R = \mathbb{Z}/15\mathbb{Z}$, and take $a = 3 \bmod 15$ and $b = 0 \bmod 15$. Note that $a|b$ here, but we may write $b = ac$ with $c = 0, 5$, or $10 \bmod 15$. Another weird feature of this ring is that $3 \bmod 15$ divides $6 \bmod 15$, but also $6 \bmod 15$ divides $3 \bmod 15 = 3 \times 6 \bmod 15$. This allows us to factor $3 \bmod 15$ indefinitely: $3 \bmod 15 = 3 \times 6 \bmod 15 = 3 \times 6 \times 6 \bmod 15$, and so on.

The weirdness in this example arises from zero divisors, and to avoid such pitfalls, we shall develop ideas of divisibility and factorizations in the context of integral domains. If R is an integral domain, and $a|b$ with $a \neq 0$, then there is a unique way to write $b = ac$. Indeed, if $b = ac_1 = ac_2$, then we may use Lemma 1.13 to cancel a and conclude that $c_1 = c_2$.

Lemma 1.28. *Let R be an integral domain. If a and b are non-zero elements of R and $a|b$ and $b|a$ then $a = bu$ for a unit u.*

Proof. Since $a|b$ we may write $b = ac$. Since $b|a$ we may write $a = bd$. Therefore $a = bd = acd$. Since R is an integral domain, and $a \neq 0$ we may use Lemma 1.13 to cancel a from both sides of the relation $a = acd$. Thus we obtain $1 = cd$, so that c and d are units. This proves the lemma. □

If a and b are elements of a ring R with $a = bu$ for a unit u, then a and b are called *associates*.

Our observations so far suggest that to develop ideas of factorization and irreducibility in rings, we should focus on integral domains: the theory for fields is uninteresting, while the presence of zero divisors leads to pathologies as in Example 1.27. In the next few sections we will develop

a satisfactory theory of factorization into primes or irreducibles, which will cover important examples such as the integers $\mathbb{Z}$, the Gaussian integers $\mathbb{Z}[i]$, and the polynomial ring $\mathbb{F}[x]$ over any field $\mathbb{F}$. We begin by defining the notions of *prime* and *irreducible*, which will turn out to be the same in some important examples (such as the integers), but which in general are different notions.

Definition 1.29. Let R be an integral domain. An element a, not zero and not a unit, is called *irreducible* if $a = bc$ implies that either b or c is a unit. An element a (not zero or a unit) is called *reducible* if it is not irreducible.

In other words, an irreducible element cannot be factored as a product of two elements in R, except in trivial ways writing it as a unit times an associate. In the integers, this definition of an irreducible gives numbers n that are only divisible by ± 1 and $\pm n$.

Definition 1.30. Let R be an integral domain. An element p, not zero and not a unit, is called *prime* if $p|ab$ implies $p|a$ or $p|b$.

Lemma 1.31. *In any integral domain, all primes are irreducibles.*

Proof. Suppose p is prime, and write $p = ab$. We will show that a or b must necessarily be a unit, so that p would be irreducible. Since p is prime and $p|ab$, either $p|a$ or $p|b$. Say $p|a$, so that $a = pc$. Then $p = ab = pbc$, and cancelling p from both sides of $p = pbc$ we obtain $bc = 1$. Therefore b is a unit, completing our proof. □

Example 1.32. The converse to Lemma 1.31 is not true in general, and there are integral domains in which not all irreducibles are primes. For instance, in the integral domain $\mathbb{Z}[\sqrt{-5}]$ one can show that 2, 3, $(1 + \sqrt{-5})$ and $(1-\sqrt{-5})$ are all irreducible (see Exercise 11 below). However, 2 divides $(1+\sqrt{-5})\times(1-\sqrt{-5}) = 6$ but 2 does not divide either $(1+\sqrt{-5})$ or $(1-\sqrt{-5})$. In other words, in $\mathbb{Z}[\sqrt{-5}]$ the element 2 is irreducible but not prime.

In the next section we shall describe a particularly nice class of integral domains in which the notions of primes and irreducibles match. The point of the two definitions (as we shall soon see) is that it is often easy to prove the existence of a factorization of elements into irreducibles, and it is often easy to prove that a factorization into primes is unique. So it would indeed be nice if the two notions were the same!

1.5. Ideals and Principal Ideal Domains (PIDs)

We begin with the definition of an *ideal* which will be a key concept in our later discussions.

Definition 1.33. Let R be a ring (as always commutative with identity). A non-empty subset I of R is called an *ideal* if

(i) $a + b$ belongs to I for all a and b in I, and

(ii) ar belongs to I for all $a \in I$ and all $r \in R$.

Example 1.34. Since ideals are non-empty, every ideal contains some element a, and therefore contains $0 \times a = 0$. Thus every ideal contains 0, and the set $\{0\}$ itself forms an ideal, called the zero ideal. Further, the whole ring R is also an ideal.

If an ideal I contains a unit u, then it must contain $uu^{-1} = 1$, and hence must contain all elements in R (upon using property (ii)). Thus if R is a field, then there are only two ideals in R, namely $\{0\}$ and R.

Example 1.35. If a is any element in R, then the set of multiples of a, namely $\{ar \ : \ r \in R\}$, forms an ideal. We denote this ideal by (a), and call this the ideal generated by a. More generally, if $a_1, \ldots, a_n$ are elements of R, then the ideal generated by them is

$$(a_1, \ldots, a_n) = \{a_1r_1 + a_2r_2 + \ldots + a_nr_n \ : \ r_1, \ldots, r_n \in R\}.$$

You should check that this is indeed an ideal.

Definition 1.36. In any ring R an ideal (a) generated by one element is called a *principal ideal*. An integral domain where every ideal is principal is called a *Principal Ideal Domain* (abbreviated PID).

Example 1.37. The integers form a basic example of a PID. To see this, suppose I is an ideal in $\mathbb{Z}$. If $I = \{0\}$ then it is clearly principal. Suppose then that I contains non-zero elements, and let n be the smallest positive integer in I. We claim that $I = (n)$ is the set of multiples of n. If this is not true then there must be some integer $m \in I$ which is not a multiple of n. Divide m by n to extract a quotient and remainder: thus $m = nq + r$ with $1 \leq r < n$. Since m and nq are in the ideal I, it follows that r must also be in I. But this contradicts the assumption that n was the smallest positive integer in I. In Section 1.8 we shall generalize this idea and give further examples of PIDs.

Example 1.38. The polynomial ring over the integers $\mathbb{Z}[x]$ gives an example of a familiar integral domain that is not a PID. Consider the ideal I generated by 2 and x. Thus I consists of all polynomials of the form $2f + xg$ with f and $g \in \mathbb{Z}[x]$. Or, in other words, the elements of I are all polynomials $a_0 + a_1x + \ldots + a_nx^n$ with $a_i \in \mathbb{Z}$ and satisfying the extra condition that the constant coefficient a_0 is even. Suppose I is principal, and generated by $h \in \mathbb{Z}[x]$. Since $2 \in I$, we must have $h|2$, forcing h to be ± 1, or ± 2. But $h = \pm 1$ is not possible since I is not all of $\mathbb{Z}[x]$ (for instance $1 \notin I$), and $h = \pm 2$ is not possible since $2 + x \in I$.

1.6. Greatest common divisors

Definition 1.39. Let a and b be two elements in an integral domain R, with at least one of a or b being non-zero. An element $d \in R$ that divides both a and b is called a *common divisor* of a and b. A common divisor g of a and b is called a *greatest common divisor* if every common divisor of a and b also divides g.

Note, we have not said anything about the existence or uniqueness of the greatest common divisor. Indeed in Exercise 11 below, you will find an example of an integral domain where there are are elements that do not have a greatest common divisor. Further, if a greatest common divisor g exists, then you should check that gu is also a greatest common divisor for any unit u. But apart from this, the greatest common divisor (if it exists) is unique—for if g_1 and g_2 are two greatest common divisors then $g_1|g_2$ (since g_1 is a common divisor and g_2 is a greatest common divisor) and similarly $g_2|g_1$, and now use Lemma 1.28 to conclude that g_1 and g_2 are associates. We may sometimes refer to "the greatest common divisor" (when a greatest common divisor exists), but this refers to an arbitrary choice among the associates.

We now show that in a PID, the greatest common divisor of two elements can always be found, and moreover it is a linear combination of the two elements.

Proposition 1.40. *If R is a PID then there exists a greatest common divisor g for any two elements a and b (not both zero). Further we may write*

$$g = ax + by$$

for some elements x, y in R.

Proof. Given a and b consider the ideal $I = (a, b)$ generated by a and b. That is, $I = \{ax + by \ : \ x, y \in R\}$. Since R is a PID, the ideal I must be principal. Say $I = (d)$. We claim that d is a gcd of a and b (and all other gcd's are associates of d).

Note that I consists of the multiples of d, and since I contains a and b, it follows that a and b are both multiples of d. Thus d is a common divisor of a and b.

If f is a common divisor of a and b, then f divides all elements of the form $ax + by$; that is, f divides all elements of I. Therefore f must divide d. This proves that d is a gcd, and the proposition follows. □

Example 1.41. In the integral domain $\mathbb{Z}[x]$ the only common divisors of 2 and x are the units ± 1. Therefore their gcd may be taken as 1. However note that 1 cannot be written as a linear combination $2f + xg$ with $f, g \in \mathbb{Z}[x]$. This is in keeping with what we already saw in Example 1.38: $\mathbb{Z}[x]$ is not a PID.

Recall that in Lemma 1.31 we established that in any integral domain all primes are irreducible. We now establish a partial converse, showing that in a PID all irreducibles are prime.

Proposition 1.42. *Let R be a principal ideal domain. An element of R is irreducible if and only if it is prime.*

Proof. We already know that primes are irreducible, so what remains is to show that irrreducibles are prime. Let p be an irreducible in R, and we wish to show that p is prime. Suppose p divides ab and p does not divide a; we now show that p must divide b, which will complete the proof.

Consider the gcd of p and a. Since p is irreducible, it has no factors besides units and associates of p. Since p does not divide a, it follows that the gcd of p and a can only be a unit, and so we may take the gcd to be 1 (which is associate to all units). Therefore Proposition 1.40 tells us that

$$1 = ax + py$$

for some elements x and y in R. Multiplying both sides by b we find that $b = abx + pby$. Since $p|ab$, we have $p|abx$, and obviously p divides pby. Therefore p must divide $b = abx + pby$, which is what we wanted. □

1.7. Unique factorization

We are now ready to address the questions of the existence and uniqueness of factorization into irreducibles in integral domains. Let us begin by defining the problem precisely.

Let R be an integral domain. By factoring an element $a \in R$ (non-zero) into irreducibles, we mean writing

$$a = up_1p_2 \cdots p_k,$$

where u is a unit, and the p_i are irreducibles (possibly with repetitions). The first question is whether such a factorization exists. If it does, the next question is whether it is unique. To clarify what uniqueness means, suppose

$$a = up_1p_2 \cdots p_k = vq_1q_2 \cdots q_\ell$$

are two factorizations. Then we would like to assert that $k = \ell$, and that each p_i can be paired with an associate q_j—that is, apart from units/associates the p_i's and q_j's are just permutations of the same list of elements.

Definition 1.43. An integral domain where every non-zero element has a unique factorization into irreducibles as above is called a *Unique Factorization Domain* (UFD).

Proposition 1.44. *In a UFD, primes and irreducibles are the same. Further, any two elements a and b (not both 0) have a gcd.*

Proof. Suppose R is a UFD, and let $p \in R$ be irreducible. We wish to show that p is prime. Suppose p divides ab. Factor a into irreducibles $a = up_1 \cdots p_k$, and b into irreducibles $b = vq_1 \cdots q_\ell$. Thus $ab = uvp_1 \cdots p_kq_1 \cdots q_\ell$ is the unique factorization of ab into irreducibles. Since p is an irreducible dividing ab, it must be the case that p is an associate of one of $p_1, \ldots, p_k, q_1, \ldots, q_\ell$. If it is an associate of one of the p_i's then $p|a$, and if it is an associate of one of the q_j's then $p|b$. Thus we have shown that $p|ab$ implies $p|a$ or $p|b$; in other words, p is prime.

To show that the gcd of any two elements a and b exists, factor a and b into irreducibles (or, what we now know to be the same, primes). Let us express these factorizations as $a = up_1^{e_1}p_2^{e_2} \cdots p_k^{e_k}$ and $b = vp_1^{f_1}p_2^{f_2} \cdots p_k^{f_k}$ where the $p_1, \ldots, p_k$ are distinct primes (all the

primes appearing in the factorization of either a or b) and the exponents e_i and f_i are non-negative integers. Then you should check that $p_1^{\min(e_1,f_1)} p_2^{\min(e_2,f_2)} \dots p_k^{\min(e_k,f_k)}$ is the gcd of a and b. □

Theorem 1.45. *Every PID is a UFD.*

Proof of the existence of a factorization. Let R be a PID, and take a non-zero element a in R. If a is a unit or is irreducible, then we may stop. Else we can find a factor a_1 of a with a_1 not being a unit, and a_1 not an associate of a (that is, $a = a_1 b_1$ with both a_1 and b_1 not being units). If a_1 is irreducible, then look at whether b_1 is irreducible. Else extract a factor a_2 of a_1, which again is neither a unit nor an associate of a_1. Keep proceeding in this manner. If the process terminates then we would have found a factorization into irreducibles. If the process does not terminate, then we must have a chain $a, a_1, a_2, \dots$ with $a_{i+1} | a_i$, and a_{i+1} not a unit, and not an associate of a_i. We need to show that this last situation cannot happen.

Since $a_1 | a$, it follows that the ideal (a) (being the set of multiples of a) is contained in the ideal (a_1) (because a multiple of a is automatically a multiple of a_1). Thus the discussion above gives a chain of ideals

$$(a) \subset (a_1) \subset (a_2) \dots .$$

We will now show that this chain stabilizes and gives the same ideal from some point onwards. Let I denote the union $\cup_n (a_n)$. We claim that I is an ideal. Indeed if $c \in I$ then for some n we must have $c \in (a_n)$, and therefore $rc \in (a_n)$ for any element $r \in R$, which implies $rc \in I$. Similarly if c and d are in I then $c \in (a_n)$ and $d \in (a_m)$ for some n and m, and if $n \leq m$ (say) then both are contained in (a_m), and therefore so is their sum, which must now also be in I. This verifies that I is an ideal. Since R is a PID it follows that $I = (r)$ from some $r \in I$. But then r must be contained in some (a_n). So for any $m \geq n$, we have $(r) \subset (a_n) \subset (a_m) \subset I = (r)$, and so all ideals from (a_n) onwards are equal to $I = (r)$, and the chain has stabilized.

Once the chain stabilizes we have $(a_n) = (a_{n+1})$, which means that a_n and a_{n+1} are multiplies of each other, and therefore must be associates. But this contradicts our assumption, and thus completes the proof of the existence of a factorization. □

The same proof of the existence of a factorization into irreducibles would work in integral domains where every ideal is generated by finitely many elements — such rings are called *Noetherian*, after the mathematician Emmy Noether.

Proof of the uniqueness of factorization. Suppose that a can be factored into irreducibles as $up_1 \cdots p_k$ and also as $vq_1 \cdots q_\ell$ where the p_i and q_j are irreducibles. By Proposition 1.42 we know that the irreducibles p_i and q_j are also primes. Now p_1 divides $q_1 \cdots q_\ell$, and since p_1 is prime we must have p_1 divides q_j for some j. Since q_j is irreducible, this forces p_1 to be an associate of q_j. Since we're in an integral domain, we can "cancel" (but recall how we did this in Lemma 1.13) p_1 and q_j from both sides of the equation $up_1 \cdots p_k = vq_1 \cdots q_\ell$, and repeat the argument. This proves the uniqueness part. □

At present, we know from Example 1.37 that the integers $\mathbb{Z}$ form a PID and are therefore a UFD. In the next section, we shall see more examples of PIDs by generalizing the ideas in Example 1.37.

There is no converse to Theorem 1.45: there are UFDs that are not PIDs. Without going into details, let us point out that the polynomial ring $\mathbb{Z}[x]$ is a UFD (this may not be surprising to you, but does require proof), but we saw already in Example 1.38 that $\mathbb{Z}[x]$ is not a PID.

1.8. Euclidean domains

A particularly nice family of rings (which will all be PIDs) are *Euclidean domains*, which generalize the idea in Example 1.37.

Definition 1.46. An integral domain R is said to have a *division algorithm* if there is a "norm function" $N : R - \{0\} \to \mathbb{Z}_{\geq 0}$ with the following property:

If a and b are elements of R with $b \neq 0$, then there exists a "quotient" q and a "remainder" r such that $a = qb + r$, and either $r = 0$, or $N(r) < N(b)$.

An integral domain R is called *Euclidean* if it possesses a *division algorithm*.

The key fact in the division algorithm is that the remainder r can be made smaller in size (using the norm function N as a notion of size) than b.

Example 1.47. The integers $\mathbb{Z}$ satisfy a division algorithm with the norm N being the absolute value of an integer a. One way to form the remainder when dividing a by b is to subtract an appropriate multiple of b so that one lands inside the interval $[0, |b|)$. This is what we discussed in Example 1.37. Another possibility is to use signed remainders, and ensure that $-|b|/2 \leq r < |b|/2$, so that here $|r| \leq |b|/2$. One way to think of the division algorithm is that we are looking at the rational number a/b and the quotient q is the largest integer below a/b (also known as the floor $\lfloor a/b \rfloor$). For the signed remainder case take instead the quotient to be the integer nearest to a/b.

Example 1.48. We now show that the Gaussian integers $\mathbb{Z}[i]$ are also a Euclidean domain. The norm map is $N(a + bi) = a^2 + b^2$, which is the square of the absolute value of the complex number $a + bi$, and we claim that with this map the Gaussian integers satisfy a division algorithm. To see this, suppose α and $\beta \neq 0$ are in $\mathbb{Z}[i]$. Note that we can divide α by β in the field $\mathbb{Q}(i) = \{x + iy : x, y \in \mathbb{Q}\}$ (see Example 1.21): one does this by "rationalizing the denominator", that is, multiplying numerator and denominator by the complex conjugate $\overline{\beta}$. So we can find rational numbers x and y such that

$$\frac{\alpha}{\beta} = x + iy.$$

Now take the nearest integer r to x, and s to y and set $\rho = r + si \in \mathbb{Z}[i]$. Note that

$$\alpha = \frac{\alpha}{\beta}\beta = \rho\beta + \Big(\frac{\alpha}{\beta} - \rho\Big)\beta,$$

and we are thinking of $\rho \in \mathbb{Z}[i]$ as the quotient and

$$\alpha - \rho\beta = \Big(\frac{\alpha}{\beta} - \rho\Big)\beta = ((x - r) + i(y - s))\beta \in \mathbb{Z}[i]$$

as the remainder. Since $|r - x| \leq 1/2$ and $|s - y| \leq 1/2$, we obtain

$$N(\alpha - \rho\beta) = N(\beta)((r - x)^2 + (s - y)^2) \leq \Big(\frac{1}{4} + \frac{1}{4}\Big)N(\beta) = \frac{N(\beta)}{2}.$$

Thus we have found a remainder with smaller norm than β, and so the division algorithm holds. Note that above we made use of the fact that the norm $N(x + iy) = x^2 + y^2$ may be thought of also as a function on

$\mathbb{Q}(i)$ and satisfies the multiplicative property that $N(\alpha\beta) = N(\alpha)N(\beta)$ (see Example 1.18).

We should add a warning here that even though the same word "norm" is used in Example 1.18 and in the definition of the division algorithm, the two notions are distinct. In particular, the norm in the definition of the division algorithm need not be multiplicative (see the next example).

Exercises 17 and 18 will give further examples of Euclidean domains where variants of this technique work. It can be quite difficult to determine whether a given integral domain is Euclidean or not. For instance, for a long time it was unknown whether the ring $\mathbb{Z}[\sqrt{14}]$ is Euclidean, and only recently has this been determined to be Euclidean (due to M. Harper [**13**]).

Example 1.49. The polynomial ring over a field $\mathbb{F}$, namely $\mathbb{F}[x]$, is our third (and important) example of a Euclidean domain. The Euclidean norm function here is the degree of a polynomial. The division algorithm is given by long division of polynomials. Suppose we want to divide $f(x) = a_n x^n + \ldots + a_0$ by $g(x) = b_m x^m + \ldots + b_0$ (with $g(x) \neq 0$) and extract a remainder of degree $< m$. If $n < m$, then simply write $f(x) = 0 \cdot g(x) + f(x)$. If $n \geq m$, then note that $f(x) - (a_n/b_m)x^{n-m}g(x)$ is a polynomial of degree $\leq (n-1)$, and we can now try to divide this polynomial by $g(x)$ and extract a remainder. So by induction the proof goes through.

Notice that the norm used here, the degree of a polynomial, is not multiplicative. Indeed the degree of the product of two polynomials is the sum of the degrees of the factors.

Note that the key property used here is that (a_n/b_m) makes sense because we are working over a field $\mathbb{F}$. It would not be enough to work just over an integral domain. For example in the polynomial ring $\mathbb{Z}[x]$ we cannot divide x^2 by $2x$ and get a remainder of degree < 1. In fact, we shall see shortly that $\mathbb{Z}[x]$ is not a Euclidean domain.

Proposition 1.50. *Every Euclidean domain is a principal ideal domain.*

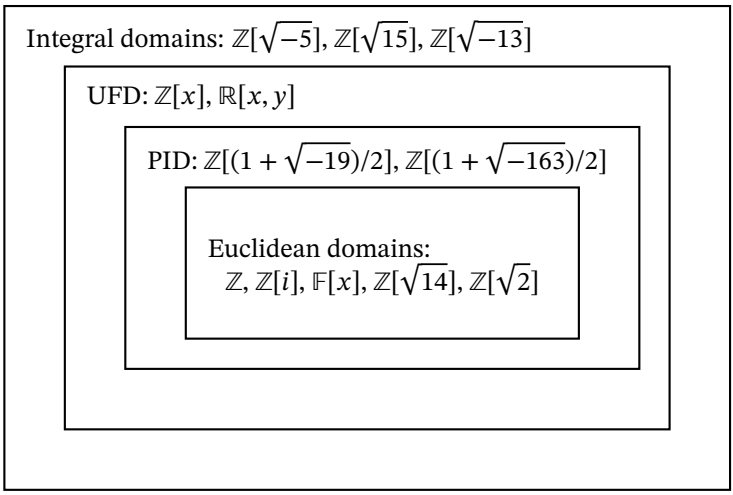
Proof. The proof follows closely the argument in Example 1.37. Suppose R is a Euclidean domain, and let I be an ideal in R. If $I = \{0\}$ there is nothing to prove. Suppose then that I is larger, and look at the norms

of all the non-zero elements of I. All these norms lie in the set of non-negative integers, and so we may find an element $b \in I$ (with $b \neq 0$) of smallest norm.

We claim that the ideal I is the set of multiples of b. Suppose instead that a is an element of I with b not dividing a. Then we may write (by the division algorithm) $a = bq + r$ with $r \neq 0$ and $N(r) < N(b)$. Since $r = a - bq$, we must also have $r \in I$, but this contradicts the minimality of $N(b)$. Therefore I is the principal ideal (b). □

Example 1.51. The ring $\mathbb{Z}[x]$ is not a principal ideal domain, and therefore not a Euclidean domain. Exercise 19 shows that the ring $R = \mathbb{Z}[(1+\sqrt{-19})/2]$ is not a Euclidean domain. However one can show that this ring is a PID; thus the converse to Proposition 1.50 does not hold.

Since every Euclidean domain is a PID, and every PID is a UFD, we conclude that the Gaussian integers and the polynomial ring over a field are both UFDs. In the next chapter, we shall discuss primes in the usual integers. Later in §3.4 we shall discuss what primes look like in the Gaussian integers, and in §4.1 we shall discuss primes in the polynomial ring over a field, which will be of importance in our construction of finite fields.



The figure above depicts the inclusions among our notions of integral domains, UFD's, PID's, and Euclidean domains, and also gives examples (just for information, and not with complete proofs) to show that

these inclusions are strict. Some, but by no means all, of these examples will be discussed further in the exercises.

We end this chapter with one last remark on gcd's. In a UFD we saw that the gcd of any two elements a and b (not both zero) exists, and in a PID we saw that the gcd may be expressed as a linear combination $ax + by$ with $x, y \in R$. In a Euclidean domain, we can go one step better and give an *algorithm* to compute the gcd, and to find x and y as well. This is known as the Euclidean algorithm.

The Euclidean algorithm. Let a and b be two elements (not both zero) in a Euclidean domain R.

If $b = 0$ then the gcd is a (or an associate of a), and clearly the gcd is $a \times 1 + b \times 0$.

If $b \neq 0$, then use the division algorithm to write $a = qb + r$; if $r = 0$ then b is the gcd. If $r \neq 0$, then $N(r) < N(b)$ (by the division algorithm), and now note that the gcd of a and b is the same as the gcd of b and r (check this carefully!). Said differently, the ideal (a, b) is the same as the ideal (b, r).

Now use the same procedure with the pair a, b replaced by the pair b, r. Note that if you have an expression for the gcd of b and r as $bv + rw$ then substituting $r = a - qb$ we obtain a linear combination of a and b, namely $bv + (a - bq)w = aw + b(v - qw)$.

Note that the Euclidean algorithm works by progressively finding elements of smaller norm in the ideal (a, b) until we find a non-zero element with smallest norm. Compare this with the proof of Proposition 1.50.

Finally note that in $\mathbb{Z}$ (use signed remainders), or $\mathbb{Z}[i]$, the Euclidean algorithm is very rapid since at each step the norm decreases (at least) by a factor of 2. We haven't discussed precisely what it means to be a rapid algorithm, but we will turn to this in Chapter 8.

1.9. Exercises

1. Let G be a group.

(i) Show that every element $g \in G$ has a unique inverse.

(ii) Suppose that for any two elements x and y in G we have $(xy)^{-1} = x^{-1}y^{-1}$. Show that G is abelian.

2. Consider $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ with operations of $+$ and $\times$ defined by component-wise addition and multiplication. Give a brief explanation of why $\mathbb{R}^2$ is a ring with these operations. Is this ring an integral domain? Describe the units and zero divisors (if any) in this ring.

3. Let $\epsilon \neq 0$ denote a symbol with $\epsilon^2 = 0$. Define a ring $\mathbb{Z}[\epsilon] = \{a + b\epsilon : a, b \in \mathbb{Z}\}$ with the natural way of adding and multiplying (subject to the $\epsilon \times \epsilon = 0$ requirement). This is vague, but what I really want is for you to work out what is intended, and it should remind you of calculus and "infinitesimals". Is this ring an integral domain? Describe the units in this ring.

4. In any ring R, show that if u is a unit then so are the powers u^n for any $n \in \mathbb{Z}$. (Interpret u^n as u multiplied by itself n times, for positive integers n; interpret u^0 as 1; and u^{-n} as $(u^{-1})^n$.)

5. Show that $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$, $\mathbb{Z}[\sqrt{3}] = \{a + b\sqrt{3} : a, b \in \mathbb{Z}\}$ and $\mathbb{Z}[\sqrt{7}] = \{a + b\sqrt{7} : a, b \in \mathbb{Z}\}$ are all rings, and indeed integral domains (with usual addition and multiplication). In explaining why these are integral domains, you may assume that $\sqrt{2}$, $\sqrt{3}$ and $\sqrt{7}$ are irrational, but you must explain why that is relevant. In this problem, I don't want you to think of $\sqrt{2}$, $\sqrt{3}$, $\sqrt{7}$ as real numbers (and therefore of these rings as *subrings* of the real numbers), but instead as just symbols whose squares equal 2, 3 and 7, rather like ϵ in Problem 3 which we could have thought of as $\sqrt{0}$.

6. Show that the rings $\mathbb{Z}[\sqrt{2}]$, $\mathbb{Z}[\sqrt{3}]$ and $\mathbb{Z}[\sqrt{7}]$ all have infinitely many units.

7. Define $\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$ and $\mathbb{Q}(\sqrt{7}) = \{a + b\sqrt{7} : a, b \in \mathbb{Q}\}$. Show that these are examples of fields.

8. Let R be a finite ring. Let a be an element of R, and assume that $a \neq 0$ and that a is not a zero divisor. Show that the map $m_a : R \to R$ defined by $m_a(r) = ar$ (thus m_a is the map "multiplication by a") is a bijection. Conclude that a is a unit.

9. Let I and J be two ideals in a ring R. Prove that $I \cap J$ is also an ideal in R.

10. Given two ideals (m) and (n) in the integers $\mathbb{Z}$, describe the ideal $(m) \cap (n)$. Is $(m) \cup (n)$ necessarily an ideal? Describe the smallest ideal that contains both (m) and (n).

11. Consider the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : \ a, b \in \mathbb{Z}\}$ and define the norm $N(a+b\sqrt{-5}) = a^2+5b^2$. (Note: we are only calling this function a norm, but it is not required to satisfy the properties of a (Euclidean) norm as in Definition 1.46. Indeed the point of this exercise is to show that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD, and hence not a PID, and hence not a Euclidean domain.)

(i) Prove that the norm is multiplicative: that is, $N(\alpha\beta) = N(\alpha)N(\beta)$ for any α, β in the ring. Determine the units in the ring. Show that if the norm of an element is prime (as an integer) then that element is irreducible.

(ii) Prove that 2, 3, $1+\sqrt{-5}$ and $1-\sqrt{-5}$ are all irreducibles, and so $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ is a genuine failure of uniqueness of factorization into irreducibles. Thus $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

(iii) Give two elements a, b in this ring for which no greatest common divisor exists.

(iv) Give an example of an ideal in this ring that is not principal.

12. Using the norm in Exercise 11 as a notion of size, show that every non-zero element in $\mathbb{Z}[\sqrt{-5}]$ can be factored into irreducibles.

13. Let R be a ring, and A and B be two ideals in R. Define AB to be the set of all elements in R of the form $\sum_{j=1}^{n} a_j b_j$ for any natural number n, and with $a_j \in A$ and $b_j \in B$. Show that AB is an ideal of R.

14. Let R be the ring $\mathbb{Z}[\sqrt{-5}]$ and define the four ideals

$$A = (2, 1 + \sqrt{-5}), \ B = (3, 1 + \sqrt{-5}),$$
$$C = (2, 1 - \sqrt{-5}), \ D = (3, 1 - \sqrt{-5}).$$

(i) Show that $A = C$, and compute the products (as defined in Exercise 13)

$$AB, AC, BD, \text{ and } CD.$$

(ii) As ideals in R, note the factorizations

$$(6) = (2) \times (3) = (1 + \sqrt{-5}) \times (1 - \sqrt{-5}).$$

How does your work in part (i) suggest a way to restore unique factorization (at the level of ideals)? Explain briefly.

Historically, ideals originated in attempts to rectify the failure of unique factorization that was observed in rings such as $\mathbb{Z}[\sqrt{-5}]$. Nineteenth century mathematicians were motivated by problems such as *Fermat's last theorem* to study factorization in general integral domains, and recognized that the failure of unique factorization foiled many attempts at proving Fermat's last theorem. This story is part of *algebraic number theory*, and see [**17**] for an introduction.

15. A ring (commutative with identity, as usual) is called Noetherian if every ideal can be generated by finitely many elements in the ring. Let R be an integral domain, and suppose R is Noetherian. Show that all non-zero elements in R admit a factorization into irreducibles.

16. Let R be a Euclidean domain with associated "norm function" N. If $N(a) = 0$ for some non-zero element a of R, show that a must be a unit.

17. (i) Let k be a positive integer congruent to 3 (mod 4). Show that $\mathbb{Z}[(1+\sqrt{-k})/2] = \{a + b(1+\sqrt{-k})/2 : a, b \in \mathbb{Z}\}$ is an integral domain.

(ii) Define the norm

$$N(a + b(1+\sqrt{-k})/2) = \left(a + \frac{b}{2}\right)^2 + \frac{kb^2}{4}.$$

Prove that this function takes values in the non-negative integers, and is multiplicative $N(\alpha\beta) = N(\alpha)N(\beta)$ for any two elements in the ring.

(iii) Prove that when $k = 3$, 7, and 11 these rings are Euclidean.

18. Show that the rings $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$, $\mathbb{Z}[\sqrt{3}] = \{a + b\sqrt{3} : a, b \in \mathbb{Z}\}$, and $\mathbb{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} : a, b \in \mathbb{Z}\}$ are all Euclidean domains. Hint: Try to generalize the notion of norm from Exercise 17 (multiply by an appropriate "conjugate") and see whether it satisfies the division algorithm.

19. This exercise shows that the ring $R = \mathbb{Z}[(1+\sqrt{-19})/2]$ is not Euclidean.

(i) Prove that the only units of R are ± 1.

(ii) Suppose there is a norm function on R that makes R Euclidean (this need not be the same function as in Exercise 17). Let s be an element in R with $s \neq 0, \pm 1$ and having smallest norm. Prove that for any $a \in R$, we must have $s|a$ or $s|(a+1)$ or $s|(a-1)$.

(iii) Taking $a = 2$, and now using the norm in Exercise 17 (or otherwise), show that s must be ± 2 or ± 3. Show that neither ± 2 nor ± 3 has the property given in (ii) above, completing the proof.