7
Rings

This chapter introduces the notion of a ring, more specifically, a commutative ring with unity. While there is a lot of terminology associated with rings, the basic ideas are fairly simple. Intuitively speaking, a ring is an algebraic structure with addition and multiplication operations that behave as one would expect.

7.1 Definitions, basic properties, and examples

Definition 7.1. *A **commutative ring with unity** is a set R together with addition and multiplication operations on R, such that:*

(i) the set R under addition forms an abelian group, and we denote the additive identity by 0_R;

(ii) multiplication is associative; that is, for all $a, b, c \in R$, we have $a(bc) = (ab)c$;

(iii) multiplication distributes over addition; that is, for all $a, b, c \in R$, we have $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$;

(iv) there exists a multiplicative identity; that is, there exists an element $1_R \in R$, such that $1_R \cdot a = a = a \cdot 1_R$ for all $a \in R$;

(v) multiplication is commutative; that is, for all $a, b \in R$, we have $ab = ba$.

There are other, more general (and less convenient) types of rings — one can drop properties (iv) and (v), and still have what is called a **ring**. We shall not, however, be working with such general rings in this text. Therefore, to simplify terminology, **from now on, by a "ring," we shall always mean a commutative ring with unity**.

Let R be a ring. Notice that because of the distributive law, for any fixed $a \in R$, the map from R to R that sends $b \in R$ to $ab \in R$ is a group homomorphism with respect to the underlying additive group of R. We call this the ***a*-multiplication map**.

We first state some simple facts:

Theorem 7.2. *Let R be a ring. Then:*

(i) *the multiplicative identity 1_R is unique;*

(ii) *$0_R \cdot a = 0_R$ for all $a \in R$;*

(iii) *$(-a)b = -(ab) = a(-b)$ for all $a, b \in R$;*

(iv) *$(-a)(-b) = ab$ for all $a, b \in R$;*

(v) *$(ka)b = k(ab) = a(kb)$ for all $k \in \mathbb{Z}$ and $a, b \in R$.*

Proof. Part (i) may be proved using the same argument as was used to prove part (i) of Theorem 6.2. Parts (ii), (iii), and (v) follow directly from parts (i), (ii), and (iii) of Theorem 6.19, using appropriate multiplication maps, discussed above. Part (iv) follows from part (iii), along with part (iv) of Theorem 6.3: $(-a)(-b) = -(a(-b)) = -(-(ab)) = ab$. □

Example 7.1. The set $\mathbb{Z}$ under the usual rules of multiplication and addition forms a ring. □

Example 7.2. For $n \geq 1$, the set $\mathbb{Z}_n$ under the rules of multiplication and addition defined in §2.5 forms a ring. □

Example 7.3. The set $\mathbb{Q}$ of rational numbers under the usual rules of multiplication and addition forms a ring. □

Example 7.4. The set $\mathbb{R}$ of real numbers under the usual rules of multiplication and addition forms a ring. □

Example 7.5. The set $\mathbb{C}$ of complex numbers under the usual rules of multiplication and addition forms a ring. Every $\alpha \in \mathbb{C}$ can be written (uniquely) as $\alpha = a+bi$, where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$. If $\alpha' = a' + b'i$ is another complex number, with $a', b' \in \mathbb{R}$, then

$$\alpha + \alpha' = (a + a') + (b + b')i \text{ and } \alpha\alpha' = (aa' - bb') + (ab' + a'b)i.$$

The fact that $\mathbb{C}$ is a ring can be verified by direct calculation; however, we shall see later that this follows easily from more general considerations.

Recall the **complex conjugation** operation, which sends α to $\overline{\alpha} := a - bi$. One can verify by direct calculation that complex conjugation is both additive and multiplicative; that is, $\overline{\alpha + \alpha'} = \overline{\alpha} + \overline{\alpha}'$ and $\overline{\alpha \cdot \alpha'} = \overline{\alpha} \cdot \overline{\alpha}'$.

The **norm** of α is $N(\alpha) := \alpha\overline{\alpha} = a^2 + b^2$. So we see that $N(\alpha)$ is a non-negative real number, and is zero if and only if $\alpha = 0$. Moreover, from the multiplicativity of complex conjugation, it is easy to see that the norm is multiplicative as well: $N(\alpha\alpha') = \alpha\alpha'\overline{\alpha\alpha'} = \alpha\alpha'\overline{\alpha}\overline{\alpha}' = \alpha\overline{\alpha}\alpha'\overline{\alpha}' = N(\alpha)N(\alpha')$. □

Example 7.6. Consider the set $\mathcal{F}$ of all arithmetic functions, that is, functions mapping positive integers to reals. Let us define addition of arithmetic functions point-wise (i.e., $(f+g)(n) = f(n) + g(n)$ for all positive integers n) and multiplication using the Dirichlet product, introduced in §2.9. The reader should verify that with addition and multiplication so defined, $\mathcal{F}$ forms a ring, where the all-zero function is the additive identity, and the special function I defined in §2.9 is the multiplicative identity. □

Example 7.7. Generalizing Example 6.18, if $R_1, \ldots, R_k$ are rings, then we can form the **direct product** $S := R_1 \times \cdots \times R_k$, which consists of all k-tuples $(a_1, \ldots, a_k)$ with $a_1 \in R_1, \ldots, a_k \in R_k$. We can view S in a natural way as a ring, with addition and multiplication defined component-wise. The additive identity is $(0_{R_1}, \ldots, 0_{R_k})$ and the multiplicative identity is $(1_{R_1}, \ldots, 1_{R_k})$. When $R = R_1 = \cdots = R_k$, the k-wise direct product of R is denoted $R^{\times k}$. □

Example 7.8. Generalizing Example 6.19, if I is an arbitrary set and R is a ring, then $\mathrm{Map}(I, R)$, which is the set of all functions $f : I \to R$, may be naturally viewed as a ring, with addition and multiplication defined point-wise: for $f, g \in \mathrm{Map}(I, R)$, we define

$$(f+g)(i) := f(i) + g(i) \text{ and } (f \cdot g)(i) := f(i) \cdot g(i) \text{ for all } i \in I.$$

We leave it to the reader to verify that $\mathrm{Map}(I, R)$ is indeed a ring, where the additive identity is the all-zero function, and the multiplicative identity is the all-one function. □

A ring R may be **trivial**, meaning that it consists of the single element 0_R, with $0_R + 0_R = 0_R$ and $0_R \cdot 0_R = 0_R$. Certainly, if R is trivial, then $1_R = 0_R$. Conversely, if $1_R = 0_R$, then for all $a \in R$, we have $a = 1_R \cdot a = 0_R \cdot a = 0_R$, and hence R is trivial. Trivial rings are not very interesting, but they naturally arise in certain constructions.

For $a_1, \ldots, a_k \in R$, the product $a_1 \cdots a_k$ needs no parentheses, because multiplication is associative; moreover, we can reorder the a_i's without changing the value of the product, since multiplication is commutative. We can also write this product as $\prod_{i=1}^{k} a_i$. By convention, such a product is defined to be 1_R when $k = 0$. When $a = a_1 = \cdots = a_k$, we can write this product as a^k. The reader may verify the usual power laws: for all $a, b \in R$, and all non-negative integers k and ℓ, we have

$$(a^{\ell})^k = a^{k\ell} = (a^k)^{\ell}, \; a^{k+\ell} = a^k a^{\ell}, \; (ab)^k = a^k b^k. \tag{7.1}$$

For all $a_1, \ldots, a_k, b_1, \ldots, b_\ell \in R$, the distributive law implies

$$(a_1 + \cdots + a_k)(b_1 + \cdots + b_\ell) = \sum_{\substack{1 \le i \le k \\ 1 \le j \le \ell}} a_i b_j.$$

A ring R is in particular an abelian group with respect to addition. We shall call a subgroup of the additive group of R an **additive subgroup** of R. The **characteristic** of R is defined as the exponent of this group (see §6.5). Note that for all $m \in \mathbb{Z}$ and $a \in R$, we have

$$ma = m(1_R \cdot a) = (m \cdot 1_R)a,$$

so that if $m \cdot 1_R = 0_R$, then $ma = 0_R$ for all $a \in R$. Thus, if the additive order of 1_R is infinite, the characteristic of R is zero, and otherwise, the characteristic of R is equal to the additive order of 1_R.

Example 7.9. The ring $\mathbb{Z}$ has characteristic zero, $\mathbb{Z}_n$ has characteristic n, and $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$ has characteristic $\mathrm{lcm}(n_1, n_2)$. □

When there is no possibility for confusion, one may write "0" instead of "0_R" and "1" instead of "1_R." Also, one may also write, for example, 2_R to denote $2 \cdot 1_R$, 3_R to denote $3 \cdot 1_R$, and so on; moreover, where the context is clear, one may use an implicit "type cast," so that $m \in \mathbb{Z}$ really means $m \cdot 1_R$.

EXERCISE 7.1. Show that the familiar **binomial theorem** (see §A2) holds in an arbitrary ring R; that is, for all $a, b \in R$ and every positive integer n, we have

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

EXERCISE 7.2. Let R be a ring. For additive subgroups A and B of R, we define their **ring-theoretic product** AB as the set of all elements of R that can be expressed as

$$a_1 b_1 + \cdots + a_k b_k$$

for some $a_1, \ldots, a_k \in A$ and $b_1, \ldots, b_k \in B$; by definition, this set includes the "empty sum" 0_R. Show that for all additive subgroups A, B, and C of R:

(a) AB is also an additive subgroup of R;

(b) $AB = BA$;

(c) $A(BC) = (AB)C$;

(d) $A(B + C) = AB + AC$.

7.1.1 Divisibility, units, and fields

For elements a, b in a ring R, we say that a **divides** b if $ar = b$ for some $r \in R$. If a divides b, we write $a \mid b$, and we may say that a is a **divisor** of b, or that b is a **multiple** of a, or that b is **divisible by** a. If a does not divide b, then we write $a \nmid b$. Note that Theorem 1.1 holds for an arbitrary ring.

We call $a \in R$ a **unit** if $a \mid 1_R$, that is, if $ar = 1_R$ for some $r \in R$. Using the same argument as was used to prove part (ii) of Theorem 6.2, it is easy to see that r is uniquely determined; it is called the **multiplicative inverse** of a, and we denote it by a^{-1}. Also, for $b \in R$, we may write b/a to denote ba^{-1}. Evidently, if a is a unit, then $a \mid b$ for every $b \in R$.

We denote the set of units by R^*. It is easy to see that $1_R \in R^*$. Moreover, R^* is closed under multiplication; indeed, if a and b are elements of R^*, then $(ab)^{-1} = a^{-1}b^{-1}$. It follows that with respect to the multiplication operation of the ring, R^* is an abelian group, called the **multiplicative group of units** of R. If $a \in R^*$ and k is a positive integer, then $a^k \in R^*$; indeed, the multiplicative inverse of a^k is $(a^{-1})^k$, which we may also write as a^{-k} (which is consistent with our notation for abelian groups). For all $a, b \in R^*$, the identities (7.1) hold for *all* integers k and ℓ.

If R is non-trivial and every non-zero element of R has a multiplicative inverse, then R is called a **field**.

Example 7.10. The only units in the ring $\mathbb{Z}$ are ± 1. Hence, $\mathbb{Z}$ is not a field. □

Example 7.11. Let n be a positive integer. The units in $\mathbb{Z}_n$ are the residue classes $[a]_n$ with $\gcd(a, n) = 1$. In particular, if n is prime, all non-zero residue classes are units, and if n is composite, some non-zero residue classes are not units. Hence, $\mathbb{Z}_n$ is a field if and only if n is prime. The notation $\mathbb{Z}_n^*$ introduced in this section for the group of units of the ring $\mathbb{Z}_n$ is consistent with the notation introduced in §2.5. □

Example 7.12. Every non-zero element of $\mathbb{Q}$ is a unit. Hence, $\mathbb{Q}$ is a field. □

Example 7.13. Every non-zero element of $\mathbb{R}$ is a unit. Hence, $\mathbb{R}$ is a field. □

Example 7.14. For non-zero $\alpha = a + bi \in \mathbb{C}$, with $a, b \in \mathbb{R}$, we have $c := N(\alpha) = a^2 + b^2 > 0$. It follows that the complex number $\bar{\alpha}c^{-1} = (ac^{-1}) + (-bc^{-1})i$ is the multiplicative inverse of α, since $\alpha \cdot \bar{\alpha}c^{-1} = (\alpha\bar{\alpha})c^{-1} = 1$. Hence, every non-zero element of $\mathbb{C}$ is a unit, and so $\mathbb{C}$ is a field. □

Example 7.15. For rings $R_1, \ldots, R_k$, it is easy to see that the multiplicative group of units of the direct product $R_1 \times \cdots \times R_k$ is equal to $R_1^* \times \cdots \times R_k^*$. Indeed, by definition, $(a_1, \ldots, a_k)$ has a multiplicative inverse if and only if each individual a_i does. □

Example 7.16. If I is a set and R is a ring, then the units in $\mathrm{Map}(I, R)$ are those functions $f : I \to R$ such that $f(i) \in R^*$ for all $i \in I$. □

Example 7.17. Consider the ring $\mathcal{F}$ of arithmetic functions defined in Example 7.6. By the result of Exercise 2.54, $\mathcal{F}^* = \{f \in \mathcal{F} : f(1) \neq 0\}$. □

7.1.2 Zero divisors and integral domains

Let R be a ring. If a and b are non-zero elements of R such that $ab = 0$, then a and b are both called **zero divisors**. If R is non-trivial and has no zero divisors, then it is called an **integral domain**. Note that if a is a unit in R, it cannot be a zero divisor (if $ab = 0$, then multiplying both sides of this equation by a^{-1} yields $b = 0$). In particular, it follows that every field is an integral domain.

Example 7.18. $\mathbb{Z}$ is an integral domain. □

Example 7.19. For $n > 1$, $\mathbb{Z}_n$ is an integral domain if and only if n is prime. In particular, if n is composite, so $n = ab$ with $1 < a < n$ and $1 < b < n$, then $[a]_n$ and $[b]_n$ are zero divisors: $[a]_n[b]_n = [0]_n$, but $[a]_n \neq [0]_n$ and $[b]_n \neq [0]_n$. □

Example 7.20. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, and hence are also integral domains. □

Example 7.21. For two non-trivial rings R_1, R_2, an element $(a_1, a_2) \in R_1 \times R_2$ is a zero divisor if and only if a_1 is a zero divisor, a_2 is a zero divisor, or exactly one of a_1 or a_2 is zero. In particular, $R_1 \times R_2$ is not an integral domain. □

The next two theorems establish certain results that are analogous to familiar facts about integer divisibility. These results hold in a general ring, provided one avoids zero divisors. The first is a **cancellation law**:

Theorem 7.3. *If R is a ring, and $a, b, c \in R$ such that $a \neq 0$ and a is not a zero divisor, then $ab = ac$ implies $b = c$.*

Proof. $ab = bc$ implies $a(b - c) = 0$. The fact that $a \neq 0$ and a is not a zero divisor implies that we must have $b - c = 0$, and so $b = c$. □

Theorem 7.4. *Let R be a ring.*

(i) Suppose $a, b \in R$, and that either a or b is not a zero divisor. Then $a \mid b$ and $b \mid a$ if and only if $ar = b$ for some $r \in R^*$.

(ii) Suppose $a, b \in R$, $a \mid b$, $a \neq 0$, and a is not a zero divisor. Then there exists a unique $r \in R$ such that $ar = b$, which we denote by b/a.

Proof. For the first statement, if $ar = b$ for some $r \in R^*$, then we also have $br^{-1} = a$; thus, $a \mid b$ and $b \mid a$. For the converse, suppose that $a \mid b$ and $b \mid a$. We may assume that b is not a zero divisor (otherwise, exchange the roles of a and b). We may also assume that b is non-zero (otherwise, $b \mid a$ implies $a = 0$, and so the conclusion holds with any r). Now, $a \mid b$ implies $ar = b$ for some $r \in R$, and $b \mid a$ implies $br' = a$ for some $r' \in R$, and hence $b = ar = br'r$. Canceling b from both sides of the equation $b = br'r$, we obtain $1 = r'r$, and so r is a unit.

For the second statement, $a \mid b$ means $ar = b$ for some $r \in R$. Moreover, this value of r is unique: if $ar = b = ar'$, then we may cancel a, obtaining $r = r'$. □

Of course, in the previous two theorems, if the ring is an integral domain, then there are no zero divisors, and so the hypotheses may be simplified in this case, dropping the explicit requirement that certain elements are not zero divisors. In particular, if a, b, and c are elements of an integral domain, such that $ab = ac$ and $a \neq 0$, then we can cancel a, obtaining $b = c$.

The next two theorems state some facts which pertain specifically to integral domains.

Theorem 7.5. *The characteristic of an integral domain is either zero or a prime.*

Proof. By way of contradiction, suppose that D is an integral domain with characteristic m that is neither zero nor prime. Since, by definition, D is not a trivial ring, we cannot have $m = 1$, and so m must be composite. Say $m = st$, where $1 < s < m$ and $1 < t < m$. Since m is the additive order of 1_D, it follows that $(s \cdot 1_D) \neq 0_D$ and $(t \cdot 1_D) \neq 0_D$; moreover, since D is an integral domain, it follows that $(s \cdot 1_D)(t \cdot 1_D) \neq 0_D$. So we have

$$0_D = m \cdot 1_D = (st) \cdot 1_D = (s \cdot 1_D)(t \cdot 1_D) \neq 0_D,$$

a contradiction. □

Theorem 7.6. *Every finite integral domain is a field.*

Proof. Let D be a finite integral domain, and let a be any non-zero element of D. Consider the a-multiplication map that sends $b \in D$ to ab, which is a group homomorphism on the additive group of D. Since a is not a zero-divisor, it follows that the kernel of the a-multiplication map is $\{0_D\}$, hence the map is injective, and by finiteness, it must be surjective as well. In particular, there must be an element $b \in D$ such that $ab = 1_D$. □

Theorem 7.7. *Every finite field F must be of cardinality p^w, where p is prime, w is a positive integer, and p is the characteristic of F.*

Proof. By Theorem 7.5, the characteristic of F is either zero or a prime, and since F is finite, it must be prime. Let p denote the characteristic. By definition, p is the exponent of the additive group of F, and by Theorem 6.43, the primes dividing the exponent are the same as the primes dividing the order, and hence F must have cardinality p^w for some positive integer w. □

Of course, for every prime p, $\mathbb{Z}_p$ is a finite field of cardinality p. As we shall see later (in Chapter 19), for every prime p and positive integer w, there exists a field of cardinality p^w. Later in this chapter, we shall see some specific examples of finite fields of cardinality p^2 (Examples 7.40, 7.59, and 7.60).

EXERCISE 7.3. Let R be a ring, and let $a, b \in R$ such that $ab \neq 0$. Show that ab is a zero divisor if and only if a is a zero divisor or b is a zero divisor.

EXERCISE 7.4. Suppose that R is a non-trivial ring in which the cancellation law holds in general: for all $a, b, c \in R$, if $a \neq 0$ and $ab = ac$, then $b = c$. Show that R is an integral domain.

EXERCISE 7.5. Let R be a ring of characteristic $m > 0$, and let n be an integer. Show that:

(a) if $\gcd(n, m) = 1$, then $n \cdot 1_R$ is a unit;

(b) if $1 < \gcd(n, m) < m$, then $n \cdot 1_R$ is a zero divisor;

(c) otherwise, $n \cdot 1_R = 0$.

EXERCISE 7.6. Let D be an integral domain, $m \in \mathbb{Z}$, and $a \in D$. Show that $ma = 0$ if and only if m is a multiple of the characteristic of D or $a = 0$.

EXERCISE 7.7. Show that for all $n \geq 1$, and for all $a, b \in \mathbb{Z}_n$, if $a \mid b$ and $b \mid a$, then $ar = b$ for some $r \in \mathbb{Z}_n^*$. Hint: this result *does not* follow from part (i) of Theorem 7.4, as we allow a and b to be zero divisors here; first consider the case where n is a prime power.

EXERCISE 7.8. Show that the ring $\mathcal{F}$ of arithmetic functions defined in Example 7.6 is an integral domain.

EXERCISE 7.9. This exercise depends on results in §6.6. Using the fundamental theorem of finite abelian groups, show that the additive group of a finite field of characteristic p and cardinality p^w is isomorphic to $\mathbb{Z}_p^{\times w}$.

7.1.3 Subrings

Definition 7.8. *A subset S of a ring R is called a* ***subring*** *if*

(i) S is an additive subgroup of R,

(ii) S is closed under multiplication, and

(iii) $1_R \in S$.

It is clear that the operations of addition and multiplication on a ring R make a subring S of R into a ring, where 0_R is the additive identity of S and 1_R is the multiplicative identity of S. One may also call R an **extension ring** of S.

Some texts do not require that 1_R belongs to a subring S, and instead require only that S contains a multiplicative identity, which may be different than that of R. This is perfectly reasonable, but for simplicity, we restrict ourselves to the case where $1_R \in S$.

Expanding the above definition, we see that a subset S of R is a subring if and only if $1_R \in S$ and for all $a, b \in S$, we have

$$a + b \in S, \quad -a \in S, \quad \text{and} \quad ab \in S.$$

In fact, to verify that S is a subring, it suffices to show that $-1_R \in S$ and that S is closed under addition and multiplication; indeed, if $-1_R \in S$ and S is closed under multiplication, then S is closed under negation, and further, $1_R = -(-1_R) \in S$.

Example 7.22. $\mathbb{Z}$ is a subring of $\mathbb{Q}$. □

Example 7.23. $\mathbb{Q}$ is a subring of $\mathbb{R}$. □

Example 7.24. $\mathbb{R}$ is a subring of $\mathbb{C}$. Note that for all $\alpha := a+bi \in \mathbb{C}$, with $a, b \in \mathbb{R}$, we have $\bar{\alpha} = \alpha \iff a + bi = a - bi \iff b = 0$. That is, $\bar{\alpha} = \alpha \iff \alpha \in \mathbb{R}$. □

Example 7.25. The set $\mathbb{Z}[i]$ of complex numbers of the form $a + bi$, with $a, b \in \mathbb{Z}$, is a subring of $\mathbb{C}$. It is called the ring of **Gaussian integers**. Since $\mathbb{C}$ is a field, it contains no zero divisors, and hence $\mathbb{Z}[i]$ contains no zero divisors either. Hence, $\mathbb{Z}[i]$ is an integral domain.

Let us determine the units of $\mathbb{Z}[i]$. Suppose $\alpha \in \mathbb{Z}[i]$ is a unit, so that there exists $\alpha' \in \mathbb{Z}[i]$ such that $\alpha\alpha' = 1$. Taking norms, we obtain

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

Since the norm of any Gaussian integer is itself a non-negative integer, and since $N(\alpha)N(\alpha') = 1$, we must have $N(\alpha) = 1$. Now, if $\alpha = a + bi$, with $a, b \in \mathbb{Z}$, then $1 = N(\alpha) = a^2 + b^2$, which implies that $\alpha = \pm 1$ or $\alpha = \pm i$. Conversely, it is easy to see that ± 1 and $\pm i$ are indeed units, and so these are the only units in $\mathbb{Z}[i]$. □

Example 7.26. Let m be a positive integer, and let $\mathbb{Q}^{(m)}$ be the set of rational numbers which can be written as a/b, where a and b are integers, and b is relatively prime to m. Then $\mathbb{Q}^{(m)}$ is a subring of $\mathbb{Q}$, since for all $a, b, c, d \in \mathbb{Z}$ with $\gcd(b, m) = 1$ and $\gcd(d, m) = 1$, we have

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},$$

and since $\gcd(bd, m) = 1$, it follows that the sum and product of any two elements of $\mathbb{Q}^{(m)}$ are again in $\mathbb{Q}^{(m)}$. Clearly, $\mathbb{Q}^{(m)}$ contains -1, and so it follows that $\mathbb{Q}^{(m)}$ is a subring of $\mathbb{Q}$. The units of $\mathbb{Q}^{(m)}$ are precisely those rational numbers of the form a/b, where $\gcd(a, m) = \gcd(b, m) = 1$. □

Example 7.27. Suppose R is a non-trivial ring. Then the set $\{0_R\}$ is not a subring of R: although it satisfies the first two requirements of the definition of a subring, it does not satisfy the third. □

Generalizing the argument in Example 7.25, it is clear that every subring of an integral domain is itself an integral domain. However, it is not the case that a subring of a field is always a field: the subring $\mathbb{Z}$ of $\mathbb{Q}$ is a counter-example. If F' is a subring of a field F, and F' is itself a field, then we say that F' is a **subfield** of F, and that F is an **extension field** of F'. For example, $\mathbb{Q}$ is a subfield of $\mathbb{R}$, which in turn is a subfield of $\mathbb{C}$.

EXERCISE 7.10. Show that if S is a subring of a ring R, then a set $T \subseteq S$ is a subring of R if and only if T is a subring of S.

EXERCISE 7.11. Show that if S and T are subrings of R, then so is $S \cap T$.

EXERCISE 7.12. Let S_1 be a subring of R_1, and S_2 a subring of R_2. Show that $S_1 \times S_2$ is a subring of $R_1 \times R_2$.

EXERCISE 7.13. Suppose that S and T are subrings of a ring R. Show that their ring-theoretic product ST (see Exercise 7.2) is a subring of R that contains $S \cup T$, and is the smallest such subring.

EXERCISE 7.14. Show that the set $\mathbb{Q}[i]$ of complex numbers of the form $a + bi$, with $a, b \in \mathbb{Q}$, is a subfield of $\mathbb{C}$.

EXERCISE 7.15. Consider the ring $\text{Map}(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$, with addition and multiplication defined point-wise.

(a) Show that $\text{Map}(\mathbb{R}, \mathbb{R})$ is not an integral domain, and that $\text{Map}(\mathbb{R}, \mathbb{R})^*$ consists of those functions that never vanish.

(b) Let $a, b \in \text{Map}(\mathbb{R}, \mathbb{R})$. Show that if $a \mid b$ and $b \mid a$, then $ar = b$ for some $r \in \text{Map}(\mathbb{R}, \mathbb{R})^*$.

(c) Let C be the subset of $\text{Map}(\mathbb{R}, \mathbb{R})$ of continuous functions. Show that C is a subring of $\text{Map}(\mathbb{R}, \mathbb{R})$, and that all functions in C^* are either everywhere positive or everywhere negative.

(d) Find elements $a, b \in C$, such that in the ring C, we have $a \mid b$ and $b \mid a$, yet there is no $r \in C^*$ such that $ar = b$.

7.2 Polynomial rings

If R is a ring, then we can form the **ring of polynomials** $R[X]$, consisting of all polynomials $g = a_0 + a_1X + \cdots + a_kX^k$ in the **indeterminate**, or "formal" variable, X, with coefficients a_i in R, and with addition and multiplication defined in the usual way.

Example 7.28. Let us define a few polynomials over the ring $\mathbb{Z}$:

$$a := 3 + X^2,\ b := 1 + 2X - X^3,\ c := 5,\ d := 1 + X,\ e := X,\ f := 4X^3.$$

We have

$$a+b = 4+2X+X^2-X^3,\ a\cdot b = 3+6X+X^2-X^3-X^5,\ cd+ef = 5+5X+4X^4.\ \Box$$

As illustrated in the previous example, elements of R are also considered to be polynomials. Such polynomials are called **constant polynomials**. The set R of constant polynomials forms a subring of $R[X]$. In particular, 0_R is the additive identity in $R[X]$ and 1_R is the multiplicative identity in $R[X]$. Note that if R is the trivial ring, then so is $R[X]$; also, if R is a subring of E, then $R[X]$ is a subring of $E[X]$.

So as to keep the distinction between ring elements and indeterminates clear, we shall use the symbol "X" only to denote the latter. Also, for a polynomial $g \in R[X]$, we shall in general write this simply as "g," and not as "$g(X)$." Of course, the choice of the symbol "X" is arbitrary; occasionally, we may use another symbol, such as "Y," as an alternative.

7.2.1 Formalities

For completeness, we present a more formal definition of the ring $R[X]$. The reader should bear in mind that this formalism is rather tedious, and may be more distracting than it is enlightening. Formally, a polynomial $g \in R[X]$ is an infinite sequence $\{a_i\}_{i=0}^{\infty}$, where each $a_i \in R$, but only finitely many of the a_i's are non-zero (intuitively, a_i represents the coefficient of X^i). For each non-negative integer j, it will be convenient to define the function $\varepsilon_j : R \to R[X]$ that maps $c \in R$ to the sequence $\{c_i\}_{i=0}^{\infty} \in R[X]$, where $c_j := c$ and $c_i := 0_R$ for $i \neq j$ (intuitively, $\varepsilon_j(c)$ represents the polynomial cX^j).

For

$$g = \{a_i\}_{i=0}^{\infty} \in R[X] \text{ and } h = \{b_i\}_{i=0}^{\infty} \in R[X],$$

we define

$$g + h := \{s_i\}_{i=0}^{\infty} \text{ and } gh := \{p_i\}_{i=0}^{\infty},$$

where for $i = 0, 1, 2, \ldots,$

$$s_i := a_i + b_i \tag{7.2}$$

and

$$p_i := \sum_{i=j+k} a_j b_k, \tag{7.3}$$

the sum being over all pairs (j, k) of non-negative integers such that $i = j + k$ (which is a finite sum). We leave it to the reader to verify that $g + h$ and gh are polynomials (i.e., only finitely many of the s_i's and p_i's are non-zero). The reader may also verify that all the requirements of Definition 7.1 are satisfied: the additive identity is the all-zero sequence $\varepsilon_0(0_R)$, and the multiplicative identity is $\varepsilon_0(1_R)$.

One can easily verify that for all $c, d \in R$, we have

$$\varepsilon_0(c + d) = \varepsilon_0(c) + \varepsilon_0(d) \text{ and } \varepsilon_0(cd) = \varepsilon_0(c)\varepsilon_0(d).$$

We shall identify $c \in R$ with $\varepsilon_0(c) \in R[X]$, viewing the ring element c as simply "shorthand" for the polynomial $\varepsilon_0(c)$ in contexts where a polynomial is expected. Note that while c and $\varepsilon_0(c)$ are not the same mathematical object, there will be no confusion in treating them as such. Thus, from a narrow, legalistic point of view, R is not a subring of $R[X]$, but we shall not let such annoying details prevent us from continuing to speak of it as such. Indeed, by appropriately renaming elements, we can make R a subring of $R[X]$ in the literal sense of the term.

We also define $X := \varepsilon_1(1_R)$. One can verify that $X^i = \varepsilon_i(1_R)$ for all $i \geq 0$. More generally, for any polynomial $g = \{a_i\}_{i=0}^{\infty}$, if $a_i = 0_R$ for all i exceeding some value k, then we have $g = \sum_{i=0}^{k} \varepsilon_0(a_i)X^i$. Writing a_i in place of $\varepsilon_0(a_i)$, we have $g = \sum_{i=0}^{k} a_i X^i$, and so we can return to the standard practice of writing polynomials as we did in Example 7.28, without any loss of precision.

7.2.2 Basic properties of polynomial rings

Let R be a ring. For non-zero $g \in R[X]$, if $g = \sum_{i=0}^{k} a_i X^i$ with $a_k \neq 0$, then we call k the **degree** of g, denoted $\deg(g)$, we call a_k the **leading coefficient** of g, denoted $\mathrm{lc}(g)$, and we call a_0 the **constant term** of g. If $\mathrm{lc}(g) = 1$, then g is called **monic**.

Suppose $g = \sum_{i=0}^{k} a_i X^i$ and $h = \sum_{i=0}^{\ell} b_i X^i$ are polynomials such that $a_k \neq 0$ and $b_\ell \neq 0$, so that $\deg(g) = k$ and $\mathrm{lc}(g) = a_k$, and $\deg(h) = \ell$ and $\mathrm{lc}(h) = b_\ell$. When we multiply these two polynomials, we get

$$gh = a_0 b_0 + (a_0 b_1 + a_1 b_0)X + \cdots + a_k b_\ell X^{k+\ell}.$$

In particular, $\deg(gh) \leq \deg(g) + \deg(h)$. If either of a_k or b_ℓ are not zero divisors, then $a_k b_\ell$ is not zero, and hence $\deg(gh) = \deg(g) + \deg(h)$. However, if both a_k

and b_ℓ are zero divisors, then we may have $a_k b_\ell = 0$, in which case, the product gh may be zero, or perhaps $gh \neq 0$ but $\deg(gh) < \deg(g) + \deg(h)$.

For the zero polynomial, we establish the following conventions: its leading coefficient and constant term are defined to be 0_R, and its degree is defined to be $-\infty$. With these conventions, we may succinctly state that

> *for all* $g, h \in R[X]$, *we have* $\deg(gh) \leq \deg(g) + \deg(h)$, *with equality guaranteed to hold unless the leading coefficients of both g and h are zero divisors.*

In particular, if the leading coefficient of a polynomial is not a zero divisor, then the polynomial is not a zero divisor. In the case where the ring of coefficients is an integral domain, we can be more precise:

Theorem 7.9. *Let D be an integral domain. Then:*

(i) *for all* $g, h \in D[X]$, *we have* $\deg(gh) = \deg(g) + \deg(h)$*;*

(ii) $D[X]$ *is an integral domain;*

(iii) $(D[X])^* = D^*$*.*

Proof. Exercise. □

An extremely important property of polynomials is a division with remainder property, analogous to that for the integers:

Theorem 7.10 (Division with remainder property). *Let R be a ring. For all* $g, h \in R[X]$ *with* $h \neq 0$ *and* $\mathrm{lc}(h) \in R^*$, *there exist unique* $q, r \in R[X]$ *such that* $g = hq + r$ *and* $\deg(r) < \deg(h)$*.*

Proof. Consider the set $S := \{g - ht : t \in R[X]\}$. Let $r = g - hq$ be an element of S of minimum degree. We must have $\deg(r) < \deg(h)$, since otherwise, we could subtract an appropriate multiple of h from r so as to eliminate the leading coefficient of r, obtaining

$$r' := r - h \cdot (\mathrm{lc}(r)\,\mathrm{lc}(h)^{-1} X^{\deg(r) - \deg(h)}) \in S,$$

where $\deg(r') < \deg(r)$, contradicting the minimality of $\deg(r)$.

That proves the existence of r and q. For uniqueness, suppose that $g = hq + r$ and $g = hq' + r'$, where $\deg(r) < \deg(h)$ and $\deg(r') < \deg(h)$. This implies $r' - r = h \cdot (q - q')$. However, if $q \neq q'$, then

$$\deg(h) > \deg(r' - r) = \deg(h \cdot (q - q')) = \deg(h) + \deg(q - q') \geq \deg(h),$$

which is impossible. Therefore, we must have $q = q'$, and hence $r = r'$. □

If $g = hq + r$ as in the above theorem, we define $g \bmod h := r$. Clearly, $h \mid g$ if and only if $g \bmod h = 0$. Moreover, note that if $\deg(g) < \deg(h)$, then $q = 0$ and $r = g$; otherwise, if $\deg(g) \geq \deg(h)$, then $q \neq 0$ and $\deg(g) = \deg(h) + \deg(q)$.

7.2.3 Polynomial evaluation

A polynomial $g = \sum_{i=0}^{k} a_i X^i \in R[X]$ naturally defines a polynomial function on R that sends $x \in R$ to $\sum_{i=0}^{k} a_i x^i \in R$, and we denote the value of this function as $g(x)$ (note that "X" denotes an indeterminate, while "x" denotes an element of R). It is important to regard polynomials over R as formal expressions, and *not* to identify them with their corresponding functions. In particular, two polynomials are equal if and only if their coefficients are equal, while two functions are equal if and only if their values agree at all points in R. This distinction is important, since there are rings R over which two different polynomials define the same function. One can of course define the ring of polynomial functions on R, but in general, that ring has a different structure from the ring of polynomials over R.

Example 7.29. In the ring $\mathbb{Z}_p$, for prime p, by Fermat's little theorem (Theorem 2.14), we have $x^p = x$ for all $x \in \mathbb{Z}_p$. However, the polynomials X^p and X are not the same polynomials (in particular, the former has degree p, while the latter has degree 1). □

More generally, suppose R is a subring of a ring E. Then every polynomial $g = \sum_{i=0}^{k} a_i X^i \in R[X]$ defines a polynomial function from E to E that sends $\alpha \in E$ to $\sum_{i=0}^{k} a_i \alpha^i \in E$, and, again, the value of this function is denoted $g(\alpha)$. We say that α is a **root** of g if $g(\alpha) = 0$.

An obvious, yet important, fact is the following:

Theorem 7.11. *Let R be a subring of a ring E. For all $g, h \in R[X]$ and $\alpha \in E$, if $s := g + h \in R[X]$ and $p := gh \in R[X]$, then we have*

$$s(\alpha) = g(\alpha) + h(\alpha) \text{ and } p(\alpha) = g(\alpha)h(\alpha).$$

Also, if $c \in R$ is a constant polynomial, then $c(\alpha) = c$ for all $\alpha \in E$.

Proof. The statement about evaluating a constant polynomial is clear from the definitions. The proof of the statements about evaluating the sum or product of polynomials is really just symbol pushing. Indeed, suppose $g = \sum_i a_i X^i$ and $h = \sum_i b_i X^i$. Then $s = \sum_i (a_i + b_i) X^i$, and so

$$s(\alpha) = \sum_i (a_i + b_i)\alpha^i = \sum_i a_i \alpha^i + \sum_i b_i \alpha^i = g(\alpha) + h(\alpha).$$

Also, we have

$$p = \Big(\sum_i a_i X^i\Big)\Big(\sum_j b_j X^j\Big) = \sum_{i,j} a_i b_j X^{i+j},$$

and employing the result for evaluating sums of polynomials, we have

$$p(\alpha) = \sum_{i,j} a_i b_j \alpha^{i+j} = \Big(\sum_i a_i \alpha^i\Big)\Big(\sum_j b_j \alpha^j\Big) = g(\alpha)h(\alpha). \quad \square$$

Example 7.30. Consider the polynomial $g := 2X^3 - 2X^2 + X - 1 \in \mathbb{Z}[X]$. We can write $g = (2X^2 + 1)(X - 1)$. For any element α of $\mathbb{Z}$, or an extension ring of $\mathbb{Z}$, we have $g(\alpha) = (2\alpha^2 + 1)(\alpha - 1)$. From this, it is clear that in $\mathbb{Z}$, g has a root only at 1; moreover, it has no other roots in $\mathbb{R}$, but in $\mathbb{C}$, it also has roots $\pm i/\sqrt{2}$. □

Example 7.31. If $E = R[X]$, then evaluating a polynomial $g \in R[X]$ at a point $\alpha \in E$ amounts to polynomial composition. For example, if $g := X^2 + X$ and $\alpha := X + 1$, then

$$g(\alpha) = g\big(X + 1\big) = (X + 1)^2 + (X + 1) = X^2 + 3X + 2. \quad \square$$

The reader is perhaps familiar with the fact that over the real or the complex numbers, every polynomial of degree k has at most k distinct roots, and the fact that every set of k points can be interpolated by a unique polynomial of degree less than k. As we will now see, these results extend to much more general, though not completely arbitrary, coefficient rings.

Theorem 7.12. *Let R be a ring, $g \in R[X]$, and $x \in R$. Then there exists a unique polynomial $q \in R[X]$ such that $g = (X - x)q + g(x)$. In particular, x is a root of g if and only if $(X - x)$ divides g.*

Proof. If R is the trivial ring, there is nothing to prove, so assume that R is non-trivial. Using the division with remainder property for polynomials, there exist unique $q, r \in R[X]$ such that $g = (X - x)q + r$, with $q, r \in R[X]$ and $\deg(r) < 1$, which means that $r \in R$. Evaluating at x, we see that $g(x) = (x - x)q(x) + r = r$. That proves the first statement. The second follows immediately from the first. □

Note that the above theorem says that $X - x$ divides $g - g(x)$, and the polynomial q in the theorem may be expressed (using the notation introduced in part (ii) of Theorem 7.4) as

$$q = \frac{g - g(x)}{X - x}.$$

Theorem 7.13. *Let D be an integral domain, and let $x_1, \ldots, x_k$ be distinct elements of D. Then for every polynomial $g \in D[X]$, the elements $x_1, \ldots, x_k$ are roots of g if and only if the polynomial $\prod_{i=1}^{k}(X - x_i)$ divides g.*

Proof. One direction is trivial: if $\prod_{i=1}^{k}(X - x_i)$ divides g, then it is clear that each x_i is a root of g. We prove the converse by induction on k. The base case $k = 1$ is just Theorem 7.12. So assume $k > 1$, and that the statement holds for $k - 1$. Let $g \in D[X]$ and let $x_1, \ldots, x_k$ be distinct roots of g. Since x_k is a root of g, then by Theorem 7.12, there exists $q \in D[X]$ such that $g = (X - x_k)q$. Moreover, for each $i = 1, \ldots, k - 1$, we have

$$0 = g(x_i) = (x_i - x_k)q(x_i),$$

and since $x_i - x_k \neq 0$ and D is an integral domain, we must have $q(x_i) = 0$. Thus, q has roots $x_1, \ldots, x_{k-1}$, and by induction $\prod_{i=1}^{k-1}(X - x_i)$ divides q, from which it then follows that $\prod_{i=1}^{k}(X - x_i)$ divides g. □

Note that in this theorem, we can slightly weaken the hypothesis: we do not need to assume that the coefficient ring is an integral domain; rather, all we really need is that for all $i \neq j$, the difference $x_i - x_j$ is not a zero divisor.

As an immediate consequence of this theorem, we obtain:

Theorem 7.14. *Let D be an integral domain, and suppose that $g \in D[X]$, with $\deg(g) = k \geq 0$. Then g has at most k distinct roots.*

Proof. If g had $k + 1$ distinct roots $x_1, \ldots, x_{k+1}$, then by the previous theorem, the polynomial $\prod_{i=1}^{k+1}(X - x_i)$, which has degree $k + 1$, would divide g, which has degree k—an impossibility. □

Theorem 7.15 (Lagrange interpolation). *Let F be a field, let $x_1, \ldots, x_k$ be distinct elements of F, and let $y_1, \ldots, y_k$ be arbitrary elements of F. Then there exists a unique polynomial $g \in F[X]$ with $\deg(g) < k$ such that $g(x_i) = y_i$ for $i = 1, \ldots, k$, namely*

$$g := \sum_{i=1}^{k} y_i \frac{\prod_{j \neq i}(X - x_j)}{\prod_{j \neq i}(x_i - x_j)}.$$

Proof. For the existence part of the theorem, one just has to verify that $g(x_i) = y_i$ for the given g, which clearly has degree less than k. This is easy to see: for $i = 1, \ldots, k$, evaluating the ith term in the sum defining g at x_i yields y_i, while evaluating any other term at x_i yields 0. The uniqueness part of the theorem follows almost immediately from Theorem 7.14: if g and h are polynomials of degree less than k such that $g(x_i) = y_i = h(x_i)$ for $i = 1, \ldots, k$, then $g - h$ is a polynomial of degree less than k with k distinct roots, which, by the previous theorem, is impossible. □

Again, we can slightly weaken the hypothesis of this theorem: we do not need to assume that the coefficient ring is a field; rather, all we really need is that for all $i \neq j$, the difference $x_i - x_j$ is a unit.

EXERCISE 7.16. Let D be an infinite integral domain, and let $g, h \in D[X]$. Show that if $g(x) = h(x)$ for all $x \in D$, then $g = h$. Thus, for an infinite integral domain D, there is a one-to-one correspondence between polynomials over D and polynomial functions on D.

EXERCISE 7.17. Let F be a field.

(a) Show that for all $b \in F$, we have $b^2 = 1$ if and only if $b = \pm 1$.

(b) Show that for all $a, b \in F$, we have $a^2 = b^2$ if and only if $a = \pm b$.

(c) Show that the familiar **quadratic formula** holds for F, assuming F has characteristic other than 2, so that $2_F \neq 0_F$. That is, for all $a, b, c \in F$ with $a \neq 0$, the polynomial $g := aX^2 + bX + c \in F[X]$ has a root in F if and only if there exists $e \in F$ such that $e^2 = d$, where d is the **discriminant** of g, defined as $d := b^2 - 4ac$, and in this case the roots of g are $(-b \pm e)/2a$.

EXERCISE 7.18. Let R be a ring, let $g \in R[X]$, with $\deg(g) = k \geq 0$, and let x be an element of R. Show that:

(a) there exist an integer m, with $0 \leq m \leq k$, and a polynomial $q \in R[X]$, such that

$$g = (X - x)^m q \text{ and } q(x) \neq 0,$$

and moreover, the values of m and q are uniquely determined;

(b) if we evaluate g at $X + x$, we have

$$g\big(X + x\big) = \sum_{i=0}^{k} b_i X^i,$$

where $b_0 = \cdots = b_{m-1} = 0$ and $b_m = q(x) \neq 0$.

Let $m_x(g)$ denote the value m in the previous exercise; for completeness, one can define $m_x(g) := \infty$ if g is the zero polynomial. If $m_x(g) > 0$, then x is called a root of g of **multiplicity** $m_x(g)$; if $m_x(g) = 1$, then x is called a **simple root** of g, and if $m_x(g) > 1$, then x is called a **multiple root** of g.

The following exercise refines Theorem 7.14, taking into account multiplicities.

EXERCISE 7.19. Let D be an integral domain, and suppose that $g \in D[X]$, with $\deg(g) = k \geq 0$. Show that

$$\sum_{x \in D} m_x(g) \leq k.$$

EXERCISE 7.20. Let D be an integral domain, let $g, h \in D[X]$, and let $x \in D$. Show that $m_x(gh) = m_x(g) + m_x(h)$.

7.2.4 Multi-variate polynomials

One can naturally generalize the notion of a polynomial in a single variable to that of a polynomial in several variables.

Consider the ring $R[X]$ of polynomials over a ring R. If Y is another indeterminate, we can form the ring $R[X][Y]$ of polynomials in Y whose coefficients are themselves polynomials in X over the ring R. One may write $R[X, Y]$ instead of $R[X][Y]$. An element of $R[X, Y]$ is called a **bivariate polynomial**.

Consider a typical element $g \in R[X, Y]$, which may be written

$$g = \sum_{j=0}^{\ell}\Big(\sum_{i=0}^{k} a_{ij}X^i\Big)Y^j. \tag{7.4}$$

Rearranging terms, this may also be written as

$$g = \sum_{\substack{0\le i\le k\\ 0\le j\le \ell}} a_{ij}X^iY^j, \tag{7.5}$$

or as

$$g = \sum_{i=0}^{k}\Big(\sum_{j=0}^{\ell} a_{ij}Y^j\Big)X^j. \tag{7.6}$$

If g is written as in (7.5), the terms X^iY^j are called **monomials**. The **total degree** of such a monomial X^iY^j is defined to be $i + j$, and if g is non-zero, then the **total degree** of g, denoted $\mathrm{Deg}(g)$, is defined to be the maximum total degree among all monomials X^iY^j appearing in (7.5) with a non-zero coefficient a_{ij}. We define the total degree of the zero polynomial to be $-\infty$.

When g is written as in (7.6), one sees that we can naturally view g as an element of $R[Y][X]$, that is, as a polynomial in X whose coefficients are polynomials in Y. From a strict, syntactic point of view, the rings $R[Y][X]$ and $R[X][Y]$ are not the same, but there is no harm done in blurring this distinction when convenient. We denote by $\deg_X(g)$ the degree of g, viewed as a polynomial in X, and by $\deg_Y(g)$ the degree of g, viewed as a polynomial in Y.

Example 7.32. Let us illustrate, with a particular example, the three different forms — as in (7.4), (7.5), and (7.6) — of expressing a bivariate polynomial. In the ring $\mathbb{Z}[X, Y]$ we have

$$\begin{aligned} g &= (5X^2 - 3X + 4)Y + (2X^2 + 1) \\ &= 5X^2Y + 2X^2 - 3XY + 4Y + 1 \\ &= (5Y + 2)X^2 + (-3Y)X + (4Y + 1). \end{aligned}$$

We have $\mathrm{Deg}(g) = 3$, $\deg_X(g) = 2$, and $\deg_Y(g) = 1$. □

More generally, we can form the ring $R[X_1, \dots, X_n]$ of **multi-variate polynomials** over R in the variables $X_1, \dots, X_n$. Formally, we can define this ring recursively as $R[X_1, \dots, X_{n-1}][X_n]$, that is, the ring of polynomials in the variable X_n, with coefficients in $R[X_1, \dots, X_{n-1}]$. A **monomial** is a term of the form $X_1^{e_1} \cdots X_n^{e_n}$, and the **total degree** of such a monomial is $e_1 + \cdots + e_n$. Every non-zero multi-variate polynomial g can be expressed uniquely (up to a re-ordering of terms) as $a_1\mu_1 + \cdots + a_k\mu_k$, where each a_i is a non-zero element of R, and each μ_i is a monomial; we define the **total degree** of g, denoted $\mathrm{Deg}(g)$, to be the maximum of the total degrees of the μ_i's. As usual, the zero polynomial is defined to have total degree $-\infty$.

Just as for bivariate polynomials, the order of the indeterminates is not important, and for every $i = 1, \dots, n$, one can naturally view any $g \in R[X_1, \dots, X_n]$ as a polynomial in X_i over the ring $R[X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$, and define $\deg_{X_i}(g)$ to be the degree of g when viewed in this way.

Just as polynomials in a single variable define polynomial functions, so do polynomials in several variables. If R is a subring of E, $g \in R[X_1, \dots, X_n]$, and $\alpha_1, \dots, \alpha_n \in E$, we define $g(\alpha_1, \dots, \alpha_n)$ to be the element of E obtained by evaluating the expression obtained by substituting α_i for X_i in g. Theorem 7.11 carries over directly to the multi-variate case.

EXERCISE 7.21. Let R be a ring, and consider the ring of multi-variate polynomials $R[X_1, \dots, X_n]$. For $m \geq 0$, define H_m to be the subset of polynomials that can be expressed as $a_1\mu_1 + \cdots + a_k\mu_k$, where each a_i belongs to R and each μ_i is a monomial of total degree m (by definition, H_m includes the zero polynomial, and $H_0 = R$). Polynomials that belong to H_m for some m are called **homogeneous polynomials**. Show that:

(a) if $g, h \in H_m$, then $g + h \in H_m$;

(b) if $g \in H_\ell$ and $h \in H_m$, then $gh \in H_{\ell+m}$;

(c) every non-zero polynomial g can be expressed uniquely as $g_0 + \cdots + g_d$, where $g_i \in H_i$ for $i = 0, \dots, d$, $g_d \neq 0$, and $d = \mathrm{Deg}(g)$;

(d) for all polynomials g, h, we have $\mathrm{Deg}(gh) \leq \mathrm{Deg}(g) + \mathrm{Deg}(h)$, and if R is an integral domain, then $\mathrm{Deg}(gh) = \mathrm{Deg}(g) + \mathrm{Deg}(h)$.

EXERCISE 7.22. Suppose that D is an integral domain, and g, h are non-zero, multi-variate polynomials over D such that gh is homogeneous. Show that g and h are also homogeneous.

EXERCISE 7.23. Let R be a ring, and let $x_1, \dots, x_n$ be elements of R. Show that every polynomial $g \in R[X_1, \dots, X_n]$ can be expressed as

$$g = (X_1 - x_1)q_1 + \cdots + (X_n - x_n)q_n + g(x_1, \dots, x_n),$$

where $q_1, \dots, q_n \in R[X_1, \dots, X_n]$.

EXERCISE 7.24. This exercise generalizes Theorem 7.14. Let D be an integral domain, and let $g \in D[X_1, \dots, X_n]$, with $\mathrm{Deg}(g) = k \geq 0$. Let S be a finite, non-empty subset of D. Show that the number of elements $(x_1, \dots, x_n) \in S^{\times n}$ such that $g(x_1, \dots, x_n) = 0$ is at most $k|S|^{n-1}$.

7.3 Ideals and quotient rings

Definition 7.16. *Let R be a ring. An* ***ideal of*** *R is an additive subgroup I of R such that $ar \in I$ for all $a \in I$ and $r \in R$ (i.e., I is closed under multiplication by elements of R).*

Expanding the above definition, we see that a non-empty subset I of R is an ideal of R if and only if for all $a, b \in I$ and $r \in R$, we have

$$a + b \in I, \quad -a \in I, \quad \text{and} \quad ar \in I.$$

Since R is commutative, the condition $ar \in I$ is equivalent to $ra \in I$. The condition $-a \in I$ is redundant, as it is implied by the condition $ar \in I$ with $r := -1_R$. In the case when R is the ring $\mathbb{Z}$, this definition of an ideal is consistent with that given in §1.2.

Clearly, $\{0_R\}$ and R are ideals of R. From the fact that an ideal I is closed under multiplication by elements of R, it is easy to see that $I = R$ if and only if $1_R \in I$.

Example 7.33. For each $m \in \mathbb{Z}$, the set $m\mathbb{Z}$ is not only an additive subgroup of the ring $\mathbb{Z}$, it is also an ideal of this ring. □

Example 7.34. For each $m \in \mathbb{Z}$, the set $m\mathbb{Z}_n$ is not only an additive subgroup of the ring $\mathbb{Z}_n$, it is also an ideal of this ring. □

Example 7.35. In the previous two examples, we saw that for some rings, the notion of an additive subgroup coincides with that of an ideal. Of course, that is the exception, not the rule. Consider the ring of polynomials $R[X]$. Suppose g is a non-zero polynomial in $R[X]$. The additive subgroup generated by g contains only polynomials whose degrees are at most that of g. However, this subgroup is not an ideal, since every ideal containing g must also contain $g \cdot X^i$ for all $i \geq 0$, and must therefore contain polynomials of arbitrarily high degree. □

Example 7.36. Let R be a ring and $x \in R$. Consider the set

$$I := \{g \in R[X] : g(x) = 0\}.$$

It is not hard to see that I is an ideal of $R[X]$. Indeed, for all $g, h \in I$ and $q \in R[X]$, we have

$$(g + h)(x) = g(x) + h(x) = 0 + 0 = 0 \text{ and } (gq)(x) = g(x)q(x) = 0 \cdot q(x) = 0.$$

Moreover, by Theorem 7.12, we have $I = \{(X - x)q : q \in R[X]\}$. □

We next develop some general constructions of ideals.

Theorem 7.17. *Let R be a ring and let $a \in R$. Then $aR := \{ar : r \in R\}$ is an ideal of R.*

Proof. This is an easy calculation. For all $ar, ar' \in aR$ and $r'' \in R$, we have $ar + ar' = a(r + r') \in aR$ and $(ar)r'' = a(rr'') \in aR$. □

The ideal aR in the previous theorem is called the **ideal of R generated by** a. An ideal of this form is called a **principal ideal**. Since R is commutative, one could also write this ideal as $Ra := \{ra : r \in R\}$. This ideal is the smallest ideal of R containing a; that is, aR contains a, and every ideal of R that contains a must contain everything in aR.

Corresponding to Theorems 6.11 and 6.12, we have:

Theorem 7.18. *If I_1 and I_2 are ideals of a ring R, then so are $I_1 + I_2$ and $I_1 \cap I_2$.*

Proof. We already know that $I_1 + I_2$ and $I_1 \cap I_2$ are additive subgroups of R, so it suffices to show that they are closed under multiplication by elements of R. The reader may easily verify that this is the case. □

Let $a_1, \ldots, a_k$ be elements of a ring R. The ideal $a_1R + \cdots + a_kR$ is called the **ideal of R generated by** $a_1, \ldots, a_k$. When the ring R is clear from context, one often writes $(a_1, \ldots, a_k)$ to denote this ideal. This ideal is that smallest ideal of R containing $a_1, \ldots, a_k$.

Example 7.37. Let n be a positive integer, and let x be any integer. Define $I := \{g \in \mathbb{Z}[X] : g(x) \equiv 0 \pmod{n}\}$. We claim that I is the ideal $(X - x, n)$ of $\mathbb{Z}[X]$. To see this, consider any fixed $g \in \mathbb{Z}[X]$. Using Theorem 7.12, we have $g = (X - x)q + g(x)$ for some $q \in \mathbb{Z}[X]$. Using the division with remainder property for integers, we have $g(x) = nq' + r$ for some $r \in \{0, \ldots, n - 1\}$ and $q' \in \mathbb{Z}$. Thus, $g(x) \equiv r \pmod{n}$, and if $g(x) \equiv 0 \pmod{n}$, then we must have

$r = 0$, and hence $g = (X - x)q + nq' \in (X - x, n)$. Conversely, if $g \in (X - x, n)$, we can write $g = (X - x)q + nq'$ for some $q, q' \in \mathbb{Z}[X]$, and from this, it is clear that $g(x) = nq'(x) \equiv 0 \pmod{n}$. □

Let I be an ideal of a ring R. Since I is an additive subgroup of R, we may adopt the congruence notation in §6.3, writing $a \equiv b \pmod{I}$ to mean $a - b \in I$, and we can form the additive quotient group R/I of cosets. Recall that for $a \in R$, the coset of I containing a is denoted $[a]_I$, and that $[a]_I = a + I = \{a + x : x \in I\}$. Also recall that addition in R/I was defined in terms of addition of coset representatives; that is, for $a, b \in I$, we defined

$$[a]_I + [b]_I := [a + b]_I.$$

Theorem 6.16 ensured that this definition was unambiguous.

Our goal now is to make R/I into a ring by similarly defining multiplication in R/I in terms of multiplication of coset representatives. To do this, we need the following multiplicative analog of Theorem 6.16, which exploits in an essential way the fact that an ideal is closed under multiplication by elements of R; in fact, this is one of the main motivations for defining the notion of an ideal as we did.

Theorem 7.19. *Suppose I is an ideal of a ring R. For all $a, a', b, b' \in R$, if $a \equiv a' \pmod{I}$ and $b \equiv b' \pmod{I}$, then $ab \equiv a'b' \pmod{I}$.*

Proof. If $a = a' + x$ for some $x \in I$ and $b = b' + y$ for some $y \in I$, then $ab = a'b' + a'y + b'x + xy$. Since I is closed under multiplication by elements of R, we see that $a'y, b'x, xy \in I$, and since I is closed under addition, $a'y + b'x + xy \in I$. Hence, $ab - a'b' \in I$. □

Using this theorem we can now unambiguously define multiplication on R/I as follows: for $a, b \in R$,

$$[a]_I \cdot [b]_I := [ab]_I.$$

Once that is done, it is straightforward to verify that all the properties that make R a ring are inherited by R/I — we leave the details of this to the reader. The multiplicative identity of R/I is the coset $[1_R]_I$.

The ring R/I is called the **quotient ring** or **residue class ring of R modulo I**. Elements of R/I may be called **residue classes**.

Note that if $I = dR$, then $a \equiv b \pmod{I}$ if and only if $d \mid (a - b)$, and as a matter of notation, one may simply write this congruence as $a \equiv b \pmod{d}$. We may also write $[a]_d$ instead of $[a]_I$.

Finally, note that if $I = R$, then R/I is the trivial ring.

Example 7.38. For each $n \geq 1$, the ring $\mathbb{Z}_n$ is precisely the quotient ring $\mathbb{Z}/n\mathbb{Z}$. □

Example 7.39. Let f be a polynomial over a ring R with $\deg(f) = \ell \geq 0$ and $\mathrm{lc}(f) \in R^*$, and consider the quotient ring $E := R[X]/fR[X]$. By the division with remainder property for polynomials (Theorem 7.10), for every $g \in R[X]$, there exists a unique polynomial $h \in R[X]$ such that $g \equiv h \pmod{f}$ and $\deg(h) < \ell$. From this, it follows that every element of E can be written uniquely as $[h]_f$, where $h \in R[X]$ is a polynomial of degree less than ℓ. Note that in this situation, we will generally prefer the more compact notation $R[X]/(f)$, instead of $R[X]/fR[X]$. □

Example 7.40. Consider the polynomial $f := X^2 + X + 1 \in \mathbb{Z}_2[X]$ and the quotient ring $E := \mathbb{Z}_2[X]/(f)$. Let us name the elements of E as follows:

$$\texttt{00} := [0]_f,\ \texttt{01} := [1]_f,\ \texttt{10} := [X]_f,\ \texttt{11} := [X+1]_f.$$

With this naming convention, addition of two elements in E corresponds to just computing the bit-wise exclusive-or of their names. More precisely, the addition table for E is the following:

+	00	01	10	11
00	00	01	10	11
01	01	00	11	10
10	10	11	00	01
11	11	10	01	00

Note that 00 acts as the additive identity for E, and that as an additive group, E is isomorphic to the additive group $\mathbb{Z}_2 \times \mathbb{Z}_2$.

As for multiplication in E, one has to compute the product of two polynomials, and then reduce modulo f. For example, to compute $\texttt{10} \cdot \texttt{11}$, using the identity $X^2 \equiv X + 1 \pmod{f}$, one sees that

$$X \cdot (X + 1) \equiv X^2 + X \equiv (X + 1) + X \equiv 1 \pmod{f};$$

thus, $\texttt{10} \cdot \texttt{11} = \texttt{01}$. The reader may verify the following multiplication table for E:

·	00	01	10	11
00	00	00	00	00
01	00	01	10	11
10	00	10	11	01
11	00	11	01	10

Observe that 01 acts as the multiplicative identity for E. Notice that every non-zero element of E has a multiplicative inverse, and so E is in fact a field. Observe that E^* is cyclic: the reader may verify that both 10 and 11 have multiplicative order 3.

This is the first example we have seen of a finite field whose cardinality is not prime. □

EXERCISE 7.25. Show that if F is a field, then the only ideals of F are $\{0_F\}$ and F.

EXERCISE 7.26. Let a, b be elements of a ring R. Show that

$$a \mid b \iff b \in aR \iff bR \subseteq aR.$$

EXERCISE 7.27. Let R be a ring. Show that if I is a non-empty subset of $R[X]$ that is closed under addition, multiplication by elements of R, and multiplication by X, then I is an ideal of $R[X]$.

EXERCISE 7.28. Let I be an ideal of R, and S a subring of R. Show that $I \cap S$ is an ideal of S.

EXERCISE 7.29. Let I be an ideal of R, and S a subring of R. Show that $I + S$ is a subring of R, and that I is an ideal of $I + S$.

EXERCISE 7.30. Let I_1 be an ideal of R_1, and I_2 an ideal of R_2. Show that $I_1 \times I_2$ is an ideal of $R_1 \times R_2$.

EXERCISE 7.31. Write down the multiplication table for $\mathbb{Z}_2[X]/(X^2 + X)$. Is this a field?

EXERCISE 7.32. Let I be an ideal of a ring R, and let x and y be elements of R with $x \equiv y \pmod{I}$. Let $g \in R[X]$. Show that $g(x) \equiv g(y) \pmod{I}$.

EXERCISE 7.33. Let R be a ring, and fix $x_1, \ldots, x_n \in R$. Let

$$I := \{g \in R[X_1, \ldots, X_n] : g(x_1, \ldots, x_n) = 0\}.$$

Show that I is an ideal of $R[X_1, \ldots, X_n]$, and that $I = (X_1 - x_1, \ldots, X_n - x_n)$.

EXERCISE 7.34. Let p be a prime, and consider the ring $\mathbb{Q}^{(p)}$ (see Example 7.26). Show that every non-zero ideal of $\mathbb{Q}^{(p)}$ is of the form (p^i), for some uniquely determined integer $i \geq 0$.

EXERCISE 7.35. Let p be a prime. Show that in the ring $\mathbb{Z}[X]$, the ideal (X, p) is not a principal ideal.

EXERCISE 7.36. Let F be a field. Show that in the ring $F[X, Y]$, the ideal (X, Y) is not a principal ideal.

EXERCISE 7.37. Let R be a ring, and let $\{I_i\}_{i=0}^{\infty}$ be a sequence of ideals of R such that $I_i \subseteq I_{i+1}$ for all $i = 0, 1, 2, \ldots$. Show that the union $\bigcup_{i=0}^{\infty} I_i$ is also an ideal of R.

EXERCISE 7.38. Let R be a ring. An ideal I of R is called **prime** if $I \subsetneq R$ and if for all $a, b \in R$, $ab \in I$ implies $a \in I$ or $b \in I$. An ideal I of R is called **maximal** if $I \subsetneq R$ and there are no ideals J of R such that $I \subsetneq J \subsetneq R$. Show that:

(a) an ideal I of R is prime if and only if R/I is an integral domain;

(b) an ideal I of R is maximal if and only if R/I is a field;

(c) all maximal ideals of R are also prime ideals.

EXERCISE 7.39. This exercise explores some examples of prime and maximal ideals. Show that:

(a) in the ring $\mathbb{Z}$, the ideal $\{0\}$ is prime but not maximal, and that the maximal ideals are precisely those of the form $p\mathbb{Z}$, where p is prime;

(b) in an integral domain D, the ideal $\{0\}$ is prime, and this ideal is maximal if and only if D is a field;

(c) if p is a prime, then in the ring $\mathbb{Z}[X]$, the ideal (X, p) is maximal, while the ideals (X) and (p) are prime, but not maximal;

(d) if F is a field, then in the ring $F[X, Y]$, the ideal (X, Y) is maximal, while the ideals (X) and (Y) are prime, but not maximal.

EXERCISE 7.40. It is a fact that every non-trivial ring R contain at least one maximal ideal. Showing this in general requires some fancy set-theoretic notions. This exercise develops a simple proof in the case where R is countable (see §A3).

(a) Show that if R is non-trivial but finite, then it contains a maximal ideal.

(b) Assume that R is countably infinite, and let $a_1, a_2, a_3, \ldots$ be an enumeration of the elements of R. Define a sequence of ideals $I_0, I_1, I_2, \ldots,$ as follows. Set $I_0 := \{0_R\}$, and for each $i \geq 0$, define

$$I_{i+1} := \begin{cases} I_i + a_i R & \text{if } I_i + a_i R \subsetneq R; \\ I_i & \text{otherwise.} \end{cases}$$

Finally, set $I := \bigcup_{i=0}^{\infty} I_i$, which by Exercise 7.37 is an ideal of R. Show that I is a maximal ideal of R. Hint: first, show that $I \subsetneq R$ by assuming that $1_R \in I$ and deriving a contradiction; then, show that I is maximal by assuming that for some $i = 1, 2, \ldots,$ we have $I \subsetneq I + a_i R \subsetneq R$, and deriving a contradiction.

EXERCISE 7.41. Let R be a ring, and let I and J be ideals of R. With the ring-theoretic product as defined in Exercise 7.2, show that:

(a) IJ is an ideal;

(b) if I and J are principal ideals, with $I = aR$ and $J = bR$, then $IJ = abR$, and so is also a principal ideal;

(c) $IJ \subseteq I \cap J$;

(d) if $I + J = R$, then $IJ = I \cap J$.

EXERCISE 7.42. Let R be a subring of E, and I an ideal of R. Show that the ring-theoretic product IE is an ideal of E that contains I, and is the smallest such ideal.

EXERCISE 7.43. Let M be a maximal ideal of a ring R, and let $a, b \in R$. Show that if $ab \in M^2$ and $b \notin M$, then $a \in M^2$. Here, $M^2 := MM$, the ring-theoretic product.

EXERCISE 7.44. Let F be a field, let $f \in F[X, Y]$, and let $E := F[X, Y]/(f)$. Define $V(f) := \{(x, y) \in F \times F : f(x, y) = 0\}$.

(a) Every element α of E naturally defines a function from $V(f)$ to F, as follows: if $\alpha = [g]_f$, with $g \in F[X, Y]$, then for $P = (x, y) \in V(f)$, we define $\alpha(P) := g(x, y)$. Show that this definition is unambiguous, that is, $g \equiv h \pmod{f}$ implies $g(x, y) = h(x, y)$.

(b) For $P = (x, y) \in V(f)$, define $M_P := \{\alpha \in E : \alpha(P) = 0\}$. Show that M_P is a maximal ideal of E, and that $M_P = \mu E + \nu E$, where $\mu := [X - x]_f$ and $\nu := [Y - y]_f$.

EXERCISE 7.45. Continuing with the previous exercise, now assume that the characteristic of F is not 2, and that $f = Y^2 - \phi$, where $\phi \in F[X]$ is a non-zero polynomial with no multiple roots in F (see definitions after Exercise 7.18).

(a) Show that if $P = (x, y) \in V(f)$, then so is $\overline{P} := (x, -y)$, and that $P = \overline{P} \iff y = 0 \iff \phi(x) = 0$.

(b) Let $P = (x, y) \in V(f)$ and $\mu := [X - x]_f \in E$. Show that $\mu E = M_P M_{\overline{P}}$ (the ring-theoretic product). Hint: use Exercise 7.43, and treat the cases $P = \overline{P}$ and $P \neq \overline{P}$ separately.

EXERCISE 7.46. Let R be a ring, and I an ideal of R. Define $\mathrm{Rad}(I)$ to be the set of all $a \in R$ such that $a^n \in I$ for some positive integer n.

(a) Show that $\mathrm{Rad}(I)$ is an ideal of R containing I. Hint: show that if $a^n \in I$ and $b^m \in I$, then $(a + b)^{n+m} \in I$.

(b) Show that if $R = \mathbb{Z}$ and $I = (d)$, where $d = p_1^{e_1} \cdots p_r^{e_r}$ is the prime factorization of d, then $\mathrm{Rad}(I) = (p_1 \cdots p_r)$.

7.4 Ring homomorphisms and isomorphisms

Definition 7.20. *A function ρ from a ring R to a ring R' is called a* ***ring homomorphism*** *if*

(i) ρ is a group homomorphism with respect to the underlying additive groups of R and R',

(ii) $\rho(ab) = \rho(a)\rho(b)$ for all $a, b \in R$, and

(iii) $\rho(1_R) = 1_{R'}$.

Expanding the definition, the requirements that ρ must satisfy in order to be a ring homomorphism are that for all $a, b \in R$, we have $\rho(a + b) = \rho(a) + \rho(b)$ and $\rho(ab) = \rho(a)\rho(b)$, and that $\rho(1_R) = 1_{R'}$.

Note that some texts do not require that a ring homomorphism satisfies part (iii) of our definition (which is not redundant—see Examples 7.49 and 7.50 below). Since a ring homomorphism is also an additive group homomorphism, we use the same notation and terminology for image and kernel.

Example 7.41. If S is a subring of a ring R, then the inclusion map $i : S \to R$ is obviously a ring homomorphism. □

Example 7.42. Suppose I is an ideal of a ring R. Analogous to Example 6.36, we may define the **natural map** from the ring R to the quotient ring R/I as follows:

$$\begin{aligned} \rho :\quad & R \to R/I \\ & a \mapsto [a]_I. \end{aligned}$$

Not only is this a surjective homomorphism of additive groups, with kernel I, it is a *ring* homomorphism. Indeed, we have

$$\rho(ab) = [ab]_I = [a]_I \cdot [b]_I = \rho(a) \cdot \rho(b),$$

and $\rho(1_R) = [1_R]_I$, which is the multiplicative identity in R/I. □

Example 7.43. For a given positive integer n, the natural map from $\mathbb{Z}$ to $\mathbb{Z}_n$ sends $a \in \mathbb{Z}$ to the residue class $[a]_n$. This is a surjective ring homomorphism, whose kernel is $n\mathbb{Z}$. □

Example 7.44. Let R be a subring of a ring E, and fix $\alpha \in E$. The **polynomial evaluation map**

$$\begin{aligned} \rho :\quad & R[X] \to E \\ & g \mapsto g(\alpha) \end{aligned}$$

is a ring homomorphism (see Theorem 7.11). The image of ρ consists of all polynomial expressions in α with coefficients in R, and is denoted $R[\alpha]$. As the reader may verify, $R[\alpha]$ is a subring of E containing α and all of R, and is the smallest such subring of E. □

Example 7.45. We can generalize the previous example to multi-variate polynomials. If R is a subring of a ring E and $\alpha_1, \ldots, \alpha_n \in E$, then the map

$$\begin{aligned} \rho : \quad R[X_1, \ldots, X_n] &\to E \\ g &\mapsto g(\alpha_1, \ldots, \alpha_n) \end{aligned}$$

is a ring homomorphism. Its image consists of all polynomial expressions in $\alpha_1, \ldots, \alpha_n$ with coefficients in R, and is denoted $R[\alpha_1, \ldots, \alpha_n]$. Moreover, this image is a subring of E containing $\alpha_1, \ldots, \alpha_n$ and all of R, and is the smallest such subring of E. Note that $R[\alpha_1, \ldots, \alpha_n] = R[\alpha_1, \ldots, \alpha_{n-1}][\alpha_n]$. □

Example 7.46. Let $\rho : R \to R'$ be a ring homomorphism. We can extend the domain of definition of ρ from R to $R[X]$ by defining $\rho(\sum_i a_i X^i) := \sum_i \rho(a_i) X^i$. This yields a ring homomorphism from $R[X]$ into $R'[X]$. To verify this, suppose $g = \sum_i a_i X^i$ and $h = \sum_i b_i X^i$ are polynomials in $R[X]$. Let $s := g + h \in R[X]$ and $p := gh \in R[X]$, and write $s = \sum_i s_i X^i$ and $p = \sum_i p_i X^i$, so that

$$s_i = a_i + b_i \text{ and } p_i = \sum_{i=j+k} a_j b_k.$$

Then we have

$$\rho(s_i) = \rho(a_i + b_i) = \rho(a_i) + \rho(b_i),$$

which is the coefficient of X^i in $\rho(g) + \rho(h)$, and

$$\rho(p_i) = \rho\Big(\sum_{i=j+k} a_j b_k\Big) = \sum_{i=j+k} \rho(a_j b_k) = \sum_{i=j+k} \rho(a_j)\rho(b_k),$$

which is the coefficient of X^i in $\rho(g)\rho(h)$.

Sometimes a more compact notation is convenient: we may prefer to write $\bar{a}$ for the image of $a \in R$ under ρ, and if we do this, then for $g = \sum_i a_i X^i \in R[X]$, we write $\bar{g}$ for the image $\sum_i \bar{a}_i X^i$ of g under the extension of ρ to $R[X]$. □

Example 7.47. Consider the natural map that sends $a \in \mathbb{Z}$ to $\bar{a} := [a]_n \in \mathbb{Z}_n$ (see Example 7.43). As in the previous example, we may extend this to a ring homomorphism from $\mathbb{Z}[X]$ to $\mathbb{Z}_n[X]$ that sends $g = \sum_i a_i X^i \in \mathbb{Z}[X]$ to $\bar{g} = \sum_i \bar{a}_i X^i \in \mathbb{Z}_n[X]$. This homomorphism is clearly surjective. Let us determine its kernel. Observe that if $g = \sum_i a_i X^i$, then $\bar{g} = 0$ if and only if $n \mid a_i$ for each i; therefore, the kernel is the ideal $n\mathbb{Z}[X]$ of $\mathbb{Z}[X]$. □

Example 7.48. Let R be a ring of prime characteristic p. For all $a, b \in R$, we have (see Exercise 7.1)

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^{p-k} b^k.$$

However, by Exercise 1.14, all of the binomial coefficients are multiples of p, except for $k = 0$ and $k = p$, and hence in the ring R, all of these terms vanish, leaving us with

$$(a+b)^p = a^p + b^p.$$

This result is often jokingly referred to as the "freshman's dream," for somewhat obvious reasons.

Of course, as always, we have

$$(ab)^p = a^p b^p \text{ and } 1_R^p = 1_R,$$

and so it follows that the map that sends $a \in R$ to $a^p \in R$ is a ring homomorphism from R into R. □

Example 7.49. Suppose R is a non-trivial ring, and let $\rho : R \to R$ map everything in R to 0_R. Then ρ satisfies parts (i) and (ii) of Definition 7.20, but not part (iii). □

Example 7.50. In special situations, part (iii) of Definition 7.20 may be redundant. One such situation arises when $\rho : R \to R'$ is surjective. In this case, we know that $1_{R'} = \rho(a)$ for some $a \in R$, and by part (ii) of the definition, we have

$$\rho(1_R) = \rho(1_R) \cdot 1_{R'} = \rho(1_R)\rho(a) = \rho(1_R \cdot a) = \rho(a) = 1_{R'}. \ \square$$

For a ring homomorphism $\rho : R \to R'$, all of the results of Theorem 6.19 apply. In particular, $\rho(0_R) = 0_{R'}$, $\rho(a) = \rho(b)$ if and only if $a \equiv b \pmod{\operatorname{Ker} \rho}$, and ρ is injective if and only if $\operatorname{Ker} \rho = \{0_R\}$. However, we may strengthen Theorem 6.19 as follows:

Theorem 7.21. *Let $\rho : R \to R'$ be a ring homomorphism.*

(i) If S is a subring of R, then $\rho(S)$ is a subring of R'; in particular (setting $S := R$), $\operatorname{Im} \rho$ is a subring of R'.

(ii) If S' is a subring of R', then $\rho^{-1}(S')$ is a subring of R.

(ii) If I is an ideal of R, then $\rho(I)$ is an ideal of $\operatorname{Im} \rho$.

(iv) If I' is an ideal of $\operatorname{Im} \rho$, then $\rho^{-1}(I')$ is an ideal of R; in particular (setting $I' := \{0_{R'}\}$), $\operatorname{Ker} \rho$ is an ideal of R.

Proof. In each part, we already know that the relevant object is an additive subgroup, and so it suffices to show that the appropriate additional properties are satisfied.

(i) For all $a, b \in S$, we have $ab \in S$, and hence $\rho(S)$ contains $\rho(ab) = \rho(a)\rho(b)$. Also, $1_R \in S$, and hence $\rho(S)$ contains $\rho(1_R) = 1_{R'}$.

(ii) If $\rho(a) \in S'$ and $\rho(b) \in S'$, then $\rho(ab) = \rho(a)\rho(b) \in S'$. Moreover, $\rho(1_R) = 1_{R'} \in S'$.

(iii) For all $a \in I$ and $r \in R$, we have $ar \in I$, and hence $\rho(I)$ contains $\rho(ar) = \rho(a)\rho(r)$.

(iv) For all $a \in \rho^{-1}(I')$ and $r \in R$, we have $\rho(ar) = \rho(a)\rho(r)$, and since $\rho(a)$ belongs to the ideal I', so does $\rho(a)\rho(r)$, and hence $\rho^{-1}(I')$ contains ar. □

Theorems 6.20 and 6.21 have natural ring analogs—one only has to show that the corresponding group homomorphisms satisfy the additional requirements of a ring homomorphism, which we leave to the reader to verify:

Theorem 7.22. *If $\rho : R \to R'$ and $\rho' : R' \to R''$ are ring homomorphisms, then so is their composition $\rho' \circ \rho : R \to R''$.*

Theorem 7.23. *Let $\rho_i : R \to R'_i$, for $i = 1, \ldots, k$, be ring homomorphisms. Then the map*

$$\begin{aligned} \rho : \quad & R \to R'_1 \times \cdots \times R'_k \\ & a \mapsto (\rho_1(a), \ldots, \rho_k(a)) \end{aligned}$$

is a ring homomorphism.

If a ring homomorphism $\rho : R \to R'$ is a bijection, then it is called a **ring isomorphism** of R with R'. If such a ring isomorphism ρ exists, we say that R **is isomorphic to** R', and write $R \cong R'$. Moreover, if $R = R'$, then ρ is called a **ring automorphism** on R.

Analogous to Theorem 6.22, we have:

Theorem 7.24. *If ρ is a ring isomorphism of R with R', then the inverse function ρ^{-1} is a ring isomorphism of R' with R.*

Proof. Exercise. □

Because of this theorem, if R is isomorphic to R', we may simply say that "R and R' are isomorphic." We stress that a ring isomorphism is essentially just a "renaming" of elements; in particular, we have:

Theorem 7.25. *Let $\rho : R \to R'$ be a ring isomorphism.*

(i) For all $a \in R$, a is a zero divisor if and only if $\rho(a)$ is a zero divisor.

(ii) For all $a \in R$, a is a unit if and only if $\rho(a)$ is a unit.

(iii) The restriction of R to R^* is a group isomorphism of R^* with $(R')^*$.

Proof. Exercise. □

An injective ring homomorphism $\rho : R \to E$ is called an **embedding** of R in E. In this case, $\operatorname{Im}\rho$ is a subring of E and $R \cong \operatorname{Im}\rho$. If the embedding is a natural one that is clear from context, we may simply identify elements of R with their images in E under the embedding; that is, for $a \in R$, we may simply write "a," and it is understood that this really means "$\rho(a)$" if the context demands an element of E. As a slight abuse of terminology, we shall say that R is a subring of E. Indeed, by appropriately renaming elements, we can always make R a subring of E in the literal sense of the term.

This practice of identifying elements of a ring with their images in another ring under a natural embedding is very common. We have already seen an example of this, namely, when we formally defined the ring of polynomials $R[X]$ over R in §7.2.1, we defined the map $\varepsilon_0 : R \to R[X]$ that sends $c \in R$ to the polynomial whose constant term is c, with all other coefficients zero. This map ε_0 is an embedding, and it was via this embedding that we identified elements of R with elements of $R[X]$, and so viewed R as a subring of $R[X]$. We shall see more examples of this later (in particular, Example 7.55 below).

Theorems 6.23 and 6.24 also have natural ring analogs—again, one only has to show that the corresponding group homomorphisms are also ring homomorphisms:

Theorem 7.26 (First isomorphism theorem). *Let $\rho : R \to R'$ be a ring homomorphism with kernel K and image S'. Then we have a ring isomorphism*

$$R/K \cong S'.$$

Specifically, the map

$$\begin{aligned} \bar{\rho} : \quad & R/K \to R' \\ & [a]_K \mapsto \rho(a) \end{aligned}$$

is an injective ring homomorphism whose image is S'.

Theorem 7.27. *Let $\rho : R \to R'$ be a ring homomorphism. Then for every ideal I of R with $I \subseteq \operatorname{Ker}\rho$, we may define a ring homomorphism*

$$\begin{aligned} \bar{\rho} : \quad & R/I \to R' \\ & [a]_I \mapsto \rho(a). \end{aligned}$$

Moreover, $\operatorname{Im}\bar{\rho} = \operatorname{Im}\rho$, and $\bar{\rho}$ is injective if and only if $I = \operatorname{Ker}\rho$.

Example 7.51. Returning again to the Chinese remainder theorem and the discussion in Example 6.48, if $\{n_i\}_{i=1}^k$ is a pairwise relatively prime family of positive integers, and $n := \prod_{i=1}^{k} n_i$, then the map

$$\begin{aligned} \rho : \quad \mathbb{Z} &\to \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k} \\ a &\mapsto ([a]_{n_1}, \ldots, [a]_{n_k}) \end{aligned}$$

is not just a surjective group homomorphism with kernel $n\mathbb{Z}$, it is also a *ring* homomorphism. Applying Theorem 7.26, we get a *ring* isomorphism

$$\begin{aligned} \bar{\rho} : \quad \mathbb{Z}_n &\to \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k} \\ [a]_n &\mapsto ([a]_{n_1}, \ldots, [a]_{n_k}), \end{aligned}$$

which is the same function as the function θ in Theorem 2.8. By part (iii) of Theorem 7.25, the restriction of θ to $\mathbb{Z}_n^*$ is a group isomorphism of $\mathbb{Z}_n^*$ with the multiplicative group of units of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$, which (according to Example 7.15) is $\mathbb{Z}_{n_1}^* \times \cdots \times \mathbb{Z}_{n_k}^*$. Thus, part (iii) of Theorem 2.8 is an immediate consequence of the above observations. □

Example 7.52. Extending Example 6.49, if n_1 and n_2 are positive integers with $n_1 \mid n_2$, then the map

$$\begin{aligned} \bar{\rho} : \quad \mathbb{Z}_{n_2} &\to \mathbb{Z}_{n_1} \\ [a]_{n_2} &\mapsto [a]_{n_1} \end{aligned}$$

is a surjective *ring* homomorphism. □

Example 7.53. For a ring R, consider the map $\rho : \mathbb{Z} \to R$ that sends $m \in \mathbb{Z}$ to $m \cdot 1_R$ in R. It is easily verified that ρ is a ring homomorphism. Since $\operatorname{Ker} \rho$ is an ideal of $\mathbb{Z}$, it is either $\{0\}$ or of the form $n\mathbb{Z}$ for some $n > 0$. In the first case, if $\operatorname{Ker} \rho = \{0\}$, then $\operatorname{Im} \rho \cong \mathbb{Z}$, and so the ring $\mathbb{Z}$ is embedded in R, and R has characteristic zero. In the second case, if $\operatorname{Ker} \rho = n\mathbb{Z}$ for some $n > 0$, then by Theorem 7.26, $\operatorname{Im} \rho \cong \mathbb{Z}_n$, and so the ring $\mathbb{Z}_n$ is embedded in R, and R has characteristic n.

Note that $\operatorname{Im} \rho$ is the smallest subring of R: any subring of R must contain 1_R and be closed under addition and subtraction, and so must contain $\operatorname{Im} \rho$. □

Example 7.54. We can generalize Example 7.44 by evaluating polynomials at several points. This is most fruitful when the underlying coefficient ring is a field, and the evaluation points belong to the same field. So let F be a field, and let $x_1, \ldots, x_k$ be distinct elements of F. Define the map

$$\begin{aligned} \rho : \quad F[X] &\to F^{\times k} \\ g &\mapsto (g(x_1), \ldots, g(x_k)). \end{aligned}$$

This is a ring homomorphism (as seen by applying Theorem 7.23 to the polynomial evaluation maps at the points $x_1, \ldots, x_k$). By Theorem 7.13, $\operatorname{Ker} \rho = (f)$, where

$f := \prod_{i=1}^{k}(X - x_i)$. By Theorem 7.15, ρ is surjective. Therefore, by Theorem 7.26, we get a ring isomorphism

$$\begin{aligned} \bar{\rho}: \quad F[X]/(f) &\to F^{\times k} \\ [g]_f &\mapsto (g(x_1), \ldots, g(x_k)). \quad \Box \end{aligned}$$

Example 7.55. As in Example 7.39, let f be a polynomial over a ring R with $\deg(f) = \ell$ and $\mathrm{lc}(f) \in R^*$, but now assume that $\ell > 0$. Consider the natural map ρ from $R[X]$ to the quotient ring $E := R[X]/(f)$ that sends $g \in R[X]$ to $[g]_f$. Let τ be the restriction of ρ to the subring R of $R[X]$. Evidently, τ is a ring homomorphism from R into E. Moreover, since distinct polynomials of degree less than ℓ belong to distinct residue classes modulo f, we see that τ is injective. Thus, τ is an embedding of R into E. As τ is a very natural embedding, we can identify elements of R with their images in E under τ, and regard R as a subring of E. Taking this point of view, we see that if $g = \sum_i a_i X^i$, then

$$[g]_f = \Big[\sum_i a_i X^i\Big]_f = \sum_i [a_i]_f([X]_f)^i = \sum_i a_i \xi^i = g(\xi),$$

where $\xi := [X]_f \in E$. Therefore, the natural map ρ may be viewed as the polynomial evaluation map (see Example 7.44) that sends $g \in R[X]$ to $g(\xi) \in E$.

Note that we have $E = R[\xi]$; moreover, every element of E can be expressed uniquely as $g(\xi)$ for some $g \in R[X]$ of degree less than ℓ, and more generally, for arbitrary $g, h \in R[X]$, we have $g(\xi) = h(\xi)$ if and only if $g \equiv h \pmod{f}$. Finally, note that $f(\xi) = [f]_f = [0]_f$; that is, ξ is a root of f. □

Example 7.56. As a special case of Example 7.55, let $f := X^2 + 1 \in \mathbb{R}[X]$, and consider the quotient ring $\mathbb{R}[X]/(f)$. If we set $i := [X]_f \in \mathbb{R}[X]/(f)$, then every element of $\mathbb{R}[X]/(f)$ can be expressed uniquely as $a + bi$, where $a, b \in \mathbb{R}$. Moreover, we have $i^2 = -1$, and more generally, for all $a, b, a', b' \in \mathbb{R}$, we have

$$(a + bi) + (a' + b'i) = (a + a') + (b + b')i$$

and

$$(a + bi) \cdot (a' + b'i) = (aa' - bb') + (ab' + a'b)i.$$

Thus, the rules for arithmetic in $\mathbb{R}[X]/(f)$ are precisely the familiar rules of complex arithmetic, and so $\mathbb{C}$ and $\mathbb{R}[X]/(f)$ are essentially the same, as rings. Indeed, the "algebraically correct" way of defining the field of complex numbers $\mathbb{C}$ is simply to define it to be the quotient ring $\mathbb{R}[X]/(f)$ in the first place. This will be our point of view from now on. □

Example 7.57. Consider the polynomial evaluation map

$$\begin{aligned} \rho : \quad \mathbb{R}[X] &\to \mathbb{C} = R[X]/(X^2+1) \\ g &\mapsto g(-i). \end{aligned}$$

For every $g \in \mathbb{R}[X]$, we may write $g = (X^2+1)q + a + bX$, where $q \in \mathbb{R}[X]$ and $a, b \in \mathbb{R}$. Since $(-i)^2 + 1 = i^2 + 1 = 0$, we have

$$g(-i) = ((-i)^2 + 1)q(-i) + a - bi = a - bi.$$

Clearly, then, ρ is surjective and the kernel of ρ is the ideal of $\mathbb{R}[X]$ generated by the polynomial $X^2 + 1$. By Theorem 7.26, we therefore get a ring automorphism $\bar{\rho}$ on $\mathbb{C}$ that sends $a + bi \in \mathbb{C}$ to $a - bi$. In fact, $\bar{\rho}$ is none other than the complex conjugation map. Indeed, this is the “algebraically correct” way of defining complex conjugation in the first place. □

Example 7.58. We defined the ring $\mathbb{Z}[i]$ of Gaussian integers in Example 7.25 as a subring of $\mathbb{C}$. Let us verify that the notation $\mathbb{Z}[i]$ introduced in Example 7.25 is consistent with that introduced in Example 7.44. Consider the polynomial evaluation map $\rho : \mathbb{Z}[X] \to \mathbb{C}$ that sends $g \in \mathbb{Z}[X]$ to $g(i) \in \mathbb{C}$. For every $g \in \mathbb{Z}[X]$, we may write $g = (X^2+1)q + a + bX$, where $q \in \mathbb{Z}[X]$ and $a, b \in \mathbb{Z}$. Since $i^2 + 1 = 0$, we have $g(i) = (i^2+1)q(i) + a + bi = a + bi$. Clearly, then, the image of ρ is the set $\{a + bi : a, b \in \mathbb{Z}\}$, and the kernel of ρ is the ideal of $\mathbb{Z}[X]$ generated by the polynomial $X^2 + 1$. This shows that $\mathbb{Z}[i]$ in Example 7.25 is the same as $\mathbb{Z}[i]$ in Example 7.44, and moreover, Theorem 7.26 implies that $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Z}[X]/(X^2 + 1)$.

Therefore, we can directly construct the Gaussian integers as the quotient ring $\mathbb{Z}[X]/(X^2 + 1)$. Likewise the field $\mathbb{Q}[i]$ (see Exercise 7.14) can be constructed directly as $\mathbb{Q}[X]/(X^2 + 1)$. □

Example 7.59. Let p be a prime, and consider the quotient ring $E := \mathbb{Z}_p[X]/(f)$, where $f := X^2 + 1$. If we set $i := [X]_f \in E$, then $E = \mathbb{Z}_p[i] = \{a + bi : a, b \in \mathbb{Z}_p\}$. In particular, E is a ring of cardinality p^2. Moreover, we have $i^2 = -1$, and the rules for addition and multiplication in E look exactly the same as they do in $\mathbb{C}$: for all $a, b, a', b' \in \mathbb{Z}_p$, we have

$$(a + bi) + (a' + b'i) = (a + a') + (b + b')i$$

and

$$(a + bi) \cdot (a' + b'i) = (aa' - bb') + (ab' + a'b)i.$$

The ring E may or may not be a field. We now determine for which primes p we get a field.

If $p = 2$, then $0 = 1 + i^2 = (1 + i)^2$ (see Example 7.48), and so in this case, $1 + i$ is a zero divisor and E is not a field.

Now suppose p is odd. There are two subcases to consider: $p \equiv 1 \pmod 4$ and $p \equiv 3 \pmod 4$.

Suppose $p \equiv 1 \pmod 4$. By Theorem 2.31, there exists $c \in \mathbb{Z}_p$ such that $c^2 = -1$, and therefore $f = X^2 + 1 = X^2 - c^2 = (X - c)(X + c)$, and by Example 7.45, we have a ring isomorphism $E \cong \mathbb{Z}_p \times \mathbb{Z}_p$ (which maps $a + bi \in E$ to $(a + bc, a - bc) \in \mathbb{Z}_p \times \mathbb{Z}_p$); in particular, E is not a field. Indeed, $c + i$ is a zero divisor, since $(c + i)(c - i) = c^2 - i^2 = c^2 + 1 = 0$.

Suppose $p \equiv 3 \pmod 4$. By Theorem 2.31, there is no $c \in \mathbb{Z}_p$ such that $c^2 = -1$. It follows that for all $a, b \in \mathbb{Z}_p$, not both zero, we must have $a^2 + b^2 \neq 0$; indeed, suppose that $a^2 + b^2 = 0$, and that, say, $b \neq 0$; then we would have $(a/b)^2 = -1$, contradicting the assumption that -1 has no square root in $\mathbb{Z}_p$. Therefore, $a^2 + b^2$ has a multiplicative inverse in $\mathbb{Z}_p$, from which it follows that the formula for multiplicative inverses in $\mathbb{C}$ applies equally well in E; that is,

$$(a + bi)^{-1} = \frac{a - bi}{a^2 + b^2}.$$

Therefore, in this case, E is a field. □

In Example 7.40, we saw a finite field of cardinality 4. The previous example provides us with an explicit construction of a finite field of cardinality p^2, for every prime p congruent to 3 modulo 4. As the next example shows, there exist finite fields of cardinality p^2 for all primes p.

Example 7.60. Let p an odd prime, and let $d \in \mathbb{Z}_p^*$. Let $f := X^2 - d \in \mathbb{Z}_p[X]$, and consider the ring $E := \mathbb{Z}_p[X]/(f) = \mathbb{Z}_p[\xi]$, where $\xi := [X]_f \in E$. We have $E = \{a + b\xi : a, b \in \mathbb{Z}_p\}$ and $|E| = p^2$. Note that $\xi^2 = d$, and the general rules for arithmetic in E look like this: for all $a, b, a', b' \in \mathbb{Z}_p$, we have

$$(a + b\xi) + (a' + b'\xi) = (a + a') + (b + b')\xi$$

and

$$(a + b\xi) \cdot (a' + b'\xi) = (aa' + bb'd) + (ab' + a'b)\xi.$$

Suppose that $d \in (\mathbb{Z}_p^*)^2$, so that $d = c^2$ for some $c \in \mathbb{Z}_p^*$. Then $f = (X-c)(X+c)$, and like in previous example, we have a ring isomorphism $E \cong \mathbb{Z}_p \times \mathbb{Z}_p$ (which maps $a + b\xi \in E$ to $(a + bc, a - bc) \in \mathbb{Z}_p \times \mathbb{Z}_p$); in particular, E is not a field.

Suppose that $d \notin (\mathbb{Z}_p^*)^2$. This implies that for all $a, b \in \mathbb{Z}_p$, not both zero, we have $a^2 - b^2 d \neq 0$. Using this, we get the following formula for multiplicative inverses in E:

$$(a + b\xi)^{-1} = \frac{a - b\xi}{a^2 - b^2 d}.$$

Therefore, E is a field in this case.

By Theorem 2.20, we know that $|(\mathbb{Z}_p^*)^2| = (p-1)/2$, and hence there exists $d \in \mathbb{Z}_p^* \setminus (\mathbb{Z}_p^*)^2$ for all odd primes p. Thus, we have a general (though not explicit) construction for finite fields of cardinality p^2 for all odd primes p. □

EXERCISE 7.47. Show that if $\rho : F \to R$ is a ring homomorphism from a field F into a ring R, then either R is trivial or ρ is injective. Hint: use Exercise 7.25.

EXERCISE 7.48. Verify that the "is isomorphic to" relation on rings is an equivalence relation; that is, for all rings R_1, R_2, R_3, we have:

(a) $R_1 \cong R_1$;

(b) $R_1 \cong R_2$ implies $R_2 \cong R_1$;

(c) $R_1 \cong R_2$ and $R_2 \cong R_3$ implies $R_1 \cong R_3$.

EXERCISE 7.49. Let $\rho_i : R_i \to R_i'$, for $i = 1, \ldots, k$, be ring homomorphisms. Show that the map

$$\begin{aligned} \rho : \quad R_1 \times \cdots \times R_k &\to R_1' \times \cdots \times R_k' \\ (a_1, \ldots, a_k) &\mapsto (\rho_1(a_1), \ldots, \rho_k(a_k)) \end{aligned}$$

is a ring homomorphism.

EXERCISE 7.50. Let $\rho : R \to R'$ be a ring homomorphism, and let $a \in R$. Show that $\rho(aR) = \rho(a)\rho(R)$.

EXERCISE 7.51. Let $\rho : R \to R'$ be a ring homomorphism. Let S be a subring of R, and let $\tau : S \to R'$ be the restriction of ρ to S. Show that τ is a ring homomorphism and that $\operatorname{Ker} \tau = \operatorname{Ker} \rho \cap S$.

EXERCISE 7.52. Suppose $R_1, \ldots, R_k$ are rings. Show that for each $i = 1, \ldots, k$, the projection map $\pi_i : R_1 \times \cdots \times R_k \to R_i$ that sends $(a_1, \ldots, a_k)$ to a_i is a surjective ring homomorphism.

EXERCISE 7.53. Show that if $R = R_1 \times R_2$ for rings R_1 and R_2, and I_1 is an ideal of R_1 and I_2 is an ideal of R_2, then we have a ring isomorphism $R/(I_1 \times I_2) \cong R_1/I_1 \times R_2/I_2$.

EXERCISE 7.54. Let I be an ideal of R, and S a subring of R. As we saw in Exercises 7.28, and 7.29, $I \cap S$ is an ideal of S, and I is an ideal of the subring $I + S$. Show that we have a ring isomorphism $(I + S)/I \cong S/(I \cap S)$.

EXERCISE 7.55. Let $\rho : R \to R'$ be a ring homomorphism with kernel K. Let I be an ideal of R. Show that we have a ring isomorphism $R/(I + K) \cong \rho(R)/\rho(I)$.

EXERCISE 7.56. Let n be a positive integer, and consider the natural map that sends $a \in \mathbb{Z}$ to $\bar{a} := [a]_n \in \mathbb{Z}_n$, which we may extend coefficient-wise to a ring homomorphism from $\mathbb{Z}[X]$ to $\mathbb{Z}_n[X]$, as in Example 7.47. Show that for every $f \in \mathbb{Z}[X]$, we have a ring isomorphism $\mathbb{Z}[X]/(f, n) \cong \mathbb{Z}_n[X]/(\bar{f})$.

EXERCISE 7.57. Let n be a positive integer. Show that we have ring isomorphisms $\mathbb{Z}[X]/(n) \cong \mathbb{Z}_n[X]$, $\mathbb{Z}[X]/(X) \cong \mathbb{Z}$, and $\mathbb{Z}[X]/(X, n) \cong \mathbb{Z}_n$.

EXERCISE 7.58. Let $n = pq$, where p and q are distinct primes. Show that we have a ring isomorphism $\mathbb{Z}_n[X] \cong \mathbb{Z}_p[X] \times \mathbb{Z}_q[X]$.

EXERCISE 7.59. Let p be a prime with $p \equiv 1 \pmod 4$. Show that we have a ring isomorphism $\mathbb{Z}[X]/(X^2 + 1, p) \cong \mathbb{Z}_p \times \mathbb{Z}_p$.

EXERCISE 7.60. Let $\rho : R \to R'$ be a surjective ring homomorphism. Let $\mathcal{S}$ be the set of all ideals of R that contain $\operatorname{Ker} \rho$, and let $\mathcal{S}'$ be the set of all ideals of R'. Show that the sets $\mathcal{S}$ and $\mathcal{S}'$ are in one-to-one correspondence, via the map that sends $I \in \mathcal{S}$ to $\rho(I) \in \mathcal{S}'$. Moreover, show that under this correspondence, prime ideals in $\mathcal{S}$ correspond to prime ideals in $\mathcal{S}'$, and maximal ideals in $\mathcal{S}$ correspond to maximal ideals in $\mathcal{S}'$. (See Exercise 7.38.)

EXERCISE 7.61. Let n be a positive integer whose factorization into primes is $n = p_1^{e_1} \cdots p_r^{e_r}$. What are the prime ideals of $\mathbb{Z}_n$? (See Exercise 7.38.)

EXERCISE 7.62. Let $\rho : R \to S$ be a ring homomorphism. Show that $\rho(R^*) \subseteq S^*$, and that the restriction of ρ to R^* yields a group homomorphism $\rho^* : R^* \to S^*$.

EXERCISE 7.63. Let R be a ring, and let $x_1, \ldots, x_n$ be elements of R. Show that the rings R and $R[X_1, \ldots, X_n]/(X_1 - x_1, \ldots, X_n - x_n)$ are isomorphic.

EXERCISE 7.64. This exercise and the next generalize the Chinese remainder theorem to arbitrary rings. Suppose I and J are two ideals of a ring R such that $I + J = R$. Show that the map $\rho : R \to R/I \times R/J$ that sends $a \in R$ to $([a]_I, [a]_J)$ is a surjective ring homomorphism with kernel IJ (see Exercise 7.41). Conclude that $R/(IJ)$ is isomorphic to $R/I \times R/J$.

EXERCISE 7.65. Generalize the previous exercise, showing that $R/(I_1 \cdots I_k)$ is isomorphic to $R/I_1 \times \cdots \times R/I_k$, where R is a ring, and $I_1, \ldots, I_k$ are ideals of R, provided $I_i + I_j = R$ for all i, j such that $i \neq j$.

EXERCISE 7.66. Let $\mathbb{Q}^{(m)}$ be the subring of $\mathbb{Q}$ defined in Example 7.26. Let us define the map $\rho : \mathbb{Q}^{(m)} \to \mathbb{Z}_m$ as follows. For $a/b \in \mathbb{Q}$ with b relatively prime to m, $\rho(a/b) := [a]_m([b]_m)^{-1}$. Show that ρ is unambiguously defined, and is a surjective ring homomorphism. Also, describe the kernel of ρ.

EXERCISE 7.67. Let R be a ring, $a \in R^*$, and $b \in R$. Define the map $\rho : R[X] \to R[X]$ that sends $g \in R[X]$ to $g\big(aX + b\big)$. Show that ρ is a ring automorphism.

EXERCISE 7.68. Consider the subring $\mathbb{Z}[1/2]$ of $\mathbb{Q}$. Show that $\mathbb{Z}[1/2] = \{a/2^i : a, i \in \mathbb{Z},\ i \geq 0\}$, that $(\mathbb{Z}[1/2])^* = \{2^i : i \in \mathbb{Z}\}$, and that every non-zero ideal of $\mathbb{Z}[1/2]$ is of the form (m), for some uniquely determined, *odd* integer m.

7.5 The structure of $\mathbb{Z}_n^*$

We are now in a position to precisely characterize the structure of the group $\mathbb{Z}_n^*$, for an arbitrary integer $n > 1$. This characterization will prove to be very useful in a number of applications.

Suppose $n = p_1^{e_1} \cdots p_r^{e_r}$ is the factorization of n into primes. By the Chinese remainder theorem (see Theorem 2.8 and Example 7.51), we have the ring isomorphism

$$\begin{aligned} \theta : \quad \mathbb{Z}_n &\to \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_r^{e_r}} \\ [a]_n &\mapsto ([a]_{p_1^{e_1}}, \ldots, [a]_{p_r^{e_r}}), \end{aligned}$$

and restricting θ to $\mathbb{Z}_n^*$ yields a group isomorphism

$$\mathbb{Z}_n^* \cong \mathbb{Z}_{p_1^{e_1}}^* \times \cdots \times \mathbb{Z}_{p_r^{e_r}}^*.$$

Thus, to determine the structure of the group $\mathbb{Z}_n^*$ for general n, it suffices to determine the structure for $n = p^e$, where p is prime. By Theorem 2.10, we already know the order of the group $\mathbb{Z}_{p^e}^*$, namely, $\varphi(p^e) = p^{e-1}(p - 1)$, where φ is Euler's phi function.

The main result of this section is the following:

Theorem 7.28. *If p is an odd prime, then for every positive integer e, the group $\mathbb{Z}_{p^e}^*$ is cyclic. The group $\mathbb{Z}_{2^e}^*$ is cyclic for $e = 1$ or 2, but not for $e \geq 3$. For $e \geq 3$, $\mathbb{Z}_{2^e}^*$ is isomorphic to the additive group $\mathbb{Z}_2 \times \mathbb{Z}_{2^{e-2}}$.*

In the case where $e = 1$, this theorem is a special case of the following, more general, theorem:

Theorem 7.29. *Let D be an integral domain and G a subgroup of D^* of finite order. Then G is cyclic.*

Proof. Suppose G is not cyclic. If m is the exponent of G, then by Theorem 6.41, we know that $m < |G|$. Moreover, by definition, $a^m = 1$ for all $a \in G$; that is, every element of G is a root of the polynomial $X^m - 1 \in D[X]$. But by Theorem 7.14, a polynomial of degree m over an integral domain has at most m distinct roots, and this contradicts the fact that $m < |G|$. □

This theorem immediately implies that $\mathbb{Z}_p^*$ is cyclic for every prime p, since $\mathbb{Z}_p$ is a field; however, we cannot directly use this theorem to prove that $\mathbb{Z}_{p^e}^*$ is cyclic for $e > 1$ (and p odd), because $\mathbb{Z}_{p^e}$ is not a field. To deal with the case $e > 1$, we need a few simple facts.

Lemma 7.30. *Let p be a prime. For every positive integer e, if $a \equiv b \pmod{p^e}$, then $a^p \equiv b^p \pmod{p^{e+1}}$.*

Proof. Suppose $a \equiv b \pmod{p^e}$, so that $a = b + cp^e$ for some $c \in \mathbb{Z}$. Then $a^p = b^p + pb^{p-1}cp^e + dp^{2e}$ for some $d \in \mathbb{Z}$, and it follows that $a^p \equiv b^p \pmod{p^{e+1}}$. □

Lemma 7.31. *Let p be a prime, and let e be a positive integer such that $p^e > 2$. If $a \equiv 1 + p^e \pmod{p^{e+1}}$, then $a^p \equiv 1 + p^{e+1} \pmod{p^{e+2}}$.*

Proof. Suppose $a \equiv 1 + p^e \pmod{p^{e+1}}$. By Lemma 7.30, $a^p \equiv (1+p^e)^p \pmod{p^{e+2}}$. Expanding $(1 + p^e)^p$, we have

$$(1 + p^e)^p = 1 + p \cdot p^e + \sum_{k=2}^{p-1} \binom{p}{k} p^{ek} + p^{ep}.$$

By Exercise 1.14, all of the terms in the sum on k are divisible by p^{1+2e}, and $1 + 2e \geq e + 2$ for all $e \geq 1$. For the term p^{ep}, the assumption that $p^e > 2$ means that either $p \geq 3$ or $e \geq 2$, which implies $ep \geq e + 2$. □

Now consider Theorem 7.28 in the case where p is odd. As we already know that $\mathbb{Z}_p^*$ is cyclic, assume $e > 1$. Let $x \in \mathbb{Z}$ be chosen so that $[x]_p$ generates $\mathbb{Z}_p^*$. Suppose the multiplicative order of $[x]_{p^e} \in \mathbb{Z}_{p^e}^*$ is m. We have $x^m \equiv 1 \pmod{p^e}$; hence, $x^m \equiv 1 \pmod{p}$, and so it must be the case that $p - 1$ divides m; thus, $[x^{m/(p-1)}]_{p^e}$ has multiplicative order exactly $p - 1$. By Theorem 6.38, if we find an integer y such that $[y]_{p^e}$ has multiplicative order p^{e-1}, then $[x^{m/(p-1)}y]_{p^e}$ has multiplicative order $(p - 1)p^{e-1}$, and we are done. We claim that $y := 1 + p$ does the job. Any integer between 0 and $p^e - 1$ can be expressed as an e-digit number in base p; for example, $y = (0 \cdots 0\,1\,1)_p$. If we compute successive pth powers of y modulo p^e, then by Lemma 7.31 we have

$$\begin{array}{rcllr}
y \bmod p^e & = & (0 & \cdots & 0\,1\,1)_p, \\
y^p \bmod p^e & = & (* & \cdots & *\,1\,0\,1)_p, \\
y^{p^2} \bmod p^e & = & (* & \cdots & *\,1\,0\,0\,1)_p, \\
& \vdots & & & \\
y^{p^{e-2}} \bmod p^e & = & (1\,0 & \cdots & 0\,1)_p, \\
y^{p^{e-1}} \bmod p^e & = & (0 & \cdots & 0\,1)_p.
\end{array}$$

Here, "$*$" indicates an arbitrary digit. From this table of values, it is clear (see

Theorem 6.37) that $[y]_{p^e}$ has multiplicative order p^{e-1}. That proves Theorem 7.28 for odd p.

We now prove Theorem 7.28 in the case $p = 2$. For $e = 1$ and $e = 2$, the theorem is easily verified. Suppose $e \geq 3$. Consider the subgroup $G \subseteq \mathbb{Z}_{2^e}^*$ generated by $[5]_{2^e}$. Expressing integers between 0 and $2^e - 1$ as e-digit binary numbers, and applying Lemma 7.31, we have

$$\begin{array}{rcllr}
5 \bmod 2^e & = & (0 & \cdots & 0\,1\,0\,1)_2, \\
5^2 \bmod 2^e & = & (* & \cdots & *\,1\,0\,0\,1)_2, \\
 & \vdots & & & \\
5^{2^{e-3}} \bmod 2^e & = & (1\,0 & \cdots & 0\,1)_2, \\
5^{2^{e-2}} \bmod 2^e & = & (0 & \cdots & 0\,1)_2.
\end{array}$$

So it is clear (see Theorem 6.37) that $[5]_{2^e}$ has multiplicative order 2^{e-2}. We claim that $[-1]_{2^e} \notin G$. If it were, then since it has multiplicative order 2, and since every cyclic group of even order has precisely one element of order 2 (see Theorem 6.32), it must be equal to $[5^{2^{e-3}}]_{2^e}$; however, it is clear from the above calculation that $5^{2^{e-3}} \not\equiv -1 \pmod{2^e}$. Let $H \subseteq \mathbb{Z}_{2^e}^*$ be the subgroup generated by $[-1]_{2^e}$. Then from the above, $G \cap H = \{[1]_{2^e}\}$, and hence by Theorem 6.25, $G \times H$ is isomorphic to the subgroup $G \cdot H$ of $\mathbb{Z}_{2^e}^*$. But since the orders of $G \times H$ and $\mathbb{Z}_{2^e}^*$ are equal, we must have $G \cdot H = \mathbb{Z}_{2^e}^*$. That proves the theorem.

Example 7.61. Let p be an odd prime, and let d be a positive integer dividing $p-1$. Since $\mathbb{Z}_p^*$ is a cyclic group of order $p - 1$, Theorem 6.32, implies that $(\mathbb{Z}_p^*)^d$ is the unique subgroup of $\mathbb{Z}_p^*$ of order $(p-1)/d$, and moreover, $(\mathbb{Z}_p^*)^d = \mathbb{Z}_p^*\{(p-1)/d\}$; that is, for all $\alpha \in \mathbb{Z}_p^*$, we have

$$\alpha = \beta^d \text{ for some } \beta \in \mathbb{Z}_p^* \iff \alpha^{(p-1)/d} = 1.$$

Setting $d = 2$, we arrive again at Euler's criterion (Theorem 2.21), but by a very different, and perhaps more elegant, route than that taken in our original proof of that theorem. □

EXERCISE 7.69. Show that if n is a positive integer, the group $\mathbb{Z}_n^*$ is cyclic if and only if

$$n = 1, 2, 4, p^e, \text{ or } 2p^e,$$

where p is an odd prime and e is a positive integer.

EXERCISE 7.70. Let $n = pq$, where p and q are distinct primes such that $p = 2p'+1$ and $q = 2q' + 1$, where p' and q' are themselves prime. Show that the subgroup $(\mathbb{Z}_n^*)^2$ of squares is a cyclic group of order $p'q'$.

EXERCISE 7.71. Let $n = pq$, where p and q are distinct primes such that $p \nmid (q-1)$ and $q \nmid (p-1)$.

(a) Show that the map that sends $[a]_n \in \mathbb{Z}_n^*$ to $[a^n]_{n^2} \in (\mathbb{Z}_{n^2}^*)^n$ is a group isomorphism (in particular, you need to show that this map is unambiguously defined).

(b) Consider the element $\alpha := [1+n]_{n^2} \in \mathbb{Z}_{n^2}^*$; show that for every non-negative integer k, $\alpha^k = [1 + kn]_{n^2}$; deduce that α has multiplicative order n, and also that the identity $\alpha^k = [1 + kn]_{n^2}$ holds for all integers k.

(c) Show that the map that sends $([k]_n, [a]_n) \in \mathbb{Z}_n \times \mathbb{Z}_n^*$ to $[(1+kn)a^n]_{n^2} \in \mathbb{Z}_{n^2}^*$ is a group isomorphism.

EXERCISE 7.72. This exercise develops an alternative proof of Theorem 7.29 that relies on less group theory. Let n be the order of the group G. Using Theorem 7.14, show that for all $d \mid n$, there are at most d elements in the group whose multiplicative order divides d. From this, deduce that for all $d \mid n$, the number of elements of multiplicative order d is either 0 or $\varphi(d)$. Now use Theorem 2.40 to deduce that for all $d \mid n$ (and in particular, for $d = n$), the number of elements of multiplicative order d is equal to $\varphi(d)$.