9

Rings

This chapter introduces the notion of a ring, more specifically, a commutative ring with unity. The theory of rings provides a useful conceptual framework for reasoning about a wide class of interesting algebraic structures. Intuitively speaking, a ring is an algebraic structure with addition and multiplication operations that behave like we expect addition and multiplication should. While there is a lot of terminology associated with rings, the basic ideas are fairly simple.

9.1 Definitions, basic properties, and examples

Definition 9.1. *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 9.2. *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) $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$;

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

(v) $(na)b = a(nb) = n(ab)$ for all $n \in \mathbb{Z}$ and $a, b \in R$.

Proof. Part (i) may be proved using the same argument as was used to prove part (i) of Theorem 8.2. Parts (ii), (iii), and (v) follow directly from parts (i), (ii), and (iii) of Theorem 8.20, using appropriate multiplication maps, discussed above. Part (iv) follows from parts (iii) and (iv) of Theorem 8.3. □

***Example* 9.1.** The set $\mathbb{Z}$ under the usual rules of multiplication and addition forms a ring. □

***Example* 9.2.** For $n \geq 1$, the set $\mathbb{Z}_n$ under the rules of multiplication and addition defined in §2.3 forms a ring. □

***Example* 9.3.** The set $\mathbb{Q}$ of rational numbers under the usual rules of multiplication and addition forms a ring. □

***Example* 9.4.** The set $\mathbb{R}$ of real numbers under the usual rules of multiplication and addition forms a ring. □

***Example* 9.5.** The set $\mathbb{C}$ of complex numbers under the usual rules of multiplication and addition forms a ring. Any $\alpha \in \mathbb{C}$ can be written (uniquely) as $\alpha = a + bi$, with $a, b \in \mathbb{R}$, and $i = \sqrt{-1}$. If $\alpha' = a' + b'i$ is another complex number, with $a', b' \in \mathbb{R}$, then

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

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 $\bar{\alpha} := a - bi$. One can verify by direct calculation that complex conjugation is both additive and multiplicative; that is, $\overline{\alpha + \alpha'} = \bar{\alpha} + \bar{\alpha}'$ and $\overline{\alpha \cdot \alpha'} = \bar{\alpha} \cdot \bar{\alpha}'$.

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

***Example* 9.6.** Consider the set $\mathcal{F}$ of all arithmetic functions, that is, functions mapping positive integers to real numbers. We can define addition and multiplication operations on $\mathcal{F}$ in a natural, point-wise fashion: for $f, g \in \mathcal{F}$, let $f + g$ be the function that sends n to $f(n) + g(n)$, and let $f \cdot g$ be the function that sends n to $f(n)g(n)$. These operations of addition and multiplication make $\mathcal{F}$ into a ring: the additive identity is the function that is everywhere 0, and the multiplicative identity is the function that is everywhere 1.

Another way to make $\mathcal{F}$ into a ring is to use the addition operation as above, together with the Dirichlet product, which we defined in §2.6, for the multiplication operation. In this case, the multiplicative identity is the function I that we defined in §2.6, which takes the value 1 at 1 and the value 0 everywhere else. The reader should verify that the distributive law holds. □

Note that in a ring R, 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 the ring R is **trivial**, in the sense that it consists of the single element 0_R, with $0_R + 0_R = 0_R$ and $0_R \cdot 0_R = 0_R$. If $1_R \neq 0_R$, we say that R is **non-trivial**. We shall rarely be concerned with trivial rings for their own sake; however, they do sometimes arise in certain constructions.

If $R_1, \ldots, R_k$ are rings, then the set of all k-tuples $(a_1, \ldots, a_k)$ with $a_i \in R_i$ for $i = 1, \ldots, k$, with addition and multiplication defined component-wise, forms a ring. The ring is denoted by $R_1 \times \cdots \times R_k$, and is called the **direct product** of $R_1, \ldots, R_k$.

The **characteristic** of a ring R is defined as the exponent of the underlying additive group (see §8.5). Note that for $m \in \mathbb{Z}$ 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* 9.7.** 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 $\mathrm{lcm}(n_1, n_2)$. □

For elements a, b in a ring R, we say that b **divides** a, or alternatively,

that a is **divisible by** b, if there exists $c \in R$ such that $a = bc$. If b divides a, then b is called a **divisor** of a, and we write $b \mid a$. Note Theorem 1.1 holds for an arbitrary ring.

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 $m \cdot 1_R$.

For $a \in R$ and positive integer n, the expression a^n denotes the product $a \cdot a \cdot \cdots \cdot a$, where there are n terms in the product. One may extend this definition to $n = 0$, defining a^0 to be the multiplicative identity 1_R.

Exercise 9.1. Verify the usual "rules of exponent arithmetic" for a ring R. That is, show that for $a \in R$, and non-negative integers n_1, n_2, we have

$$(a^{n_1})^{n_2} = a^{n_1 n_2} \quad \text{and} \quad a^{n_1} a^{n_2} = a^{n_1 + n_2}.$$

Exercise 9.2. Show that the familiar **binomial theorem** holds in an arbitrary ring R; that is, for $a, b \in R$ and positive integer n, we have

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

Exercise 9.3. Show that

$$\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{m} b_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j,$$

where the a_i and b_j are elements of a ring R.

9.1.1 Units and fields

Let R be a ring. We call $u \in R$ a **unit** if it divides 1_R, that is, if $uu' = 1_R$ for some $u' \in R$. In this case, it is easy to see that u' is uniquely determined, and it is called the **multiplicative inverse** of u, and we denote it by u^{-1}. Also, for $a \in R$, we may write a/u to denote au^{-1}. It is clear that a unit u divides every $a \in R$.

We denote the set of units by R^*. It is easy to verify that the set R^* is closed under multiplication, from which it follows that R^* is an abelian group, called the **multiplicative group of units** of R. If $u \in R^*$, then of course $u^n \in R^*$ for all non-negative integers n, and the multiplicative inverse

of u^n is $(u^{-1})^n$, which we may also write as u^{-n} (which is consistent with our notation for abelian groups).

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

***Example* 9.8.** The only units in the ring $\mathbb{Z}$ are ± 1. Hence, $\mathbb{Z}$ is not a field. □

***Example* 9.9.** For positive integer n, 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. Of course, 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 §2.3. □

***Example* 9.10.** Every non-zero element of $\mathbb{Q}$ is a unit. Hence, $\mathbb{Q}$ is a field. □

***Example* 9.11.** Every non-zero element of $\mathbb{R}$ is a unit. Hence, $\mathbb{R}$ is a field. □

***Example* 9.12.** 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* 9.13.** 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* 9.14.** Consider the rings of arithmetic functions defined in Example 9.6. If multiplication is defined point-wise, then an arithmetic function f is a unit if and only if $f(n) \neq 0$ for all n. If multiplication is defined in terms of the Dirichlet product, then by the result of Exercise 2.27, an arithmetic function f is a unit if and only if $f(1) \neq 0$. □

9.1.2 Zero divisors and integral domains

Let R be a ring. An element $a \in R$ is called a **zero divisor** if $a \neq 0_R$ and there exists non-zero $b \in R$ such that $ab = 0_R$.

If R is non-trivial and has no zero divisors, then it is called an **integral domain**. Put another way, a non-trivial ring R is an integral domain if

and only if the following holds: for all $a, b \in R$, $ab = 0_R$ implies $a = 0_R$ or $b = 0_R$.

Note that if u is a unit in R, it cannot be a zero divisor (if $ub = 0_R$, then multiplying both sides of this equation by u^{-1} yields $b = 0_R$). In particular, it follows that any field is an integral domain.

***Example* 9.15.** $\mathbb{Z}$ is an integral domain. □

***Example* 9.16.** For $n > 1$, $\mathbb{Z}_n$ is an integral domain if and only if n is prime. In particular, if n is composite, so $n = n_1 n_2$ with $1 < n_1 < n$ and $1 < n_2 < n$, then $[n_1]_n$ and $[n_2]_n$ are zero divisors: $[n_1]_n[n_2]_n = [0]_n$, but $[n_1]_n \neq [0]_n$ and $[n_2]_n \neq [0]_n$. □

***Example* 9.17.** $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields, and hence are also integral domains. □

***Example* 9.18.** 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. □

We have the following "cancellation law":

Theorem 9.3. *If R is a ring, and $a, b, c \in R$ such that $a \neq 0_R$ and a is not a zero divisor, then $ab = ac$ implies $b = c$.*

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

Theorem 9.4. *If D is an integral domain, then:*

(i) for all $a, b, c \in D$, $a \neq 0_D$ and $ab = ac$ implies $b = c$;

(ii) for all $a, b \in D$, $a \mid b$ and $b \mid a$ if and only if $a = bc$ for some $c \in D^*$.

(iii) for all $a, b \in D$ with $b \neq 0_D$ and $b \mid a$, there is a unique $c \in D$ such that $a = bc$, which we may denote as a/b.

Proof. The first statement follows immediately from the previous theorem and the definition of an integral domain.

For the second statement, if $a = bc$ for $c \in D^*$, then we also have $b = ac^{-1}$; thus, $b \mid a$ and $a \mid b$. Conversely, $a \mid b$ implies $b = ax$ for $x \in D$, and $b \mid a$ implies $a = by$ for $y \in D$, and hence $b = bxy$. If $b = 0_R$, then the equation $a = by$ implies $a = 0_R$, and so the statement holds for any c; otherwise, cancel b, we have $1_D = xy$, and so x and y are units.

For the third statement, if $a = bc$ and $a = bc'$, then $bc = bc'$, and cancel b. □

Theorem 9.5. *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 9.6. *Any 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 9.7. *Any 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 9.5, 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 8.42, 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 20), 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 whose cardinality is not prime (Examples 9.35 and 9.47).

Exercise 9.4. Let R be a ring of characteristic $m > 0$, and let n be any 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, $n \cdot 1_R = 0_R$.

EXERCISE 9.5. Let D be an integral domain, $m \in \mathbb{Z}$, and $a \in D$. Show that $ma = 0_D$ if and only if m is a multiple of the characteristic of D or $a = 0_D$.

EXERCISE 9.6. For $n \geq 1$, and for all $a, b \in \mathbb{Z}_n$, show that if $a \mid b$ and $b \mid a$, then $a = bc$ for some $c \in \mathbb{Z}_n^*$. Thus, part (ii) of Theorem 9.4 may hold for some rings that are not integral domains.

EXERCISE 9.7. This exercise depends on results in §8.6. 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 $\mathbb{Z}_p^{\times w}$.

9.1.3 Subrings

Definition 9.8. *A subset S of a ring R is called a* ***subring*** *if*

(i) S is a subgroup of the additive group 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 when $1_R \in S$.

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

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

If 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* 9.19.** $\mathbb{Z}$ is a subring of $\mathbb{Q}$. □

***Example* 9.20.** $\mathbb{Q}$ is a subring of $\mathbb{R}$. □

***Example* 9.21.** $\mathbb{R}$ is a subring of $\mathbb{C}$.

Note that for $\alpha := a+bi \in \mathbb{C}$, 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* 9.22.** 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. Hence, $\mathbb{Z}[i]$ is an integral domain.

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

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

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

***Example* 9.23.** Let m be a positive integer, and let $\mathbb{Q}^{(m)}$ be the set of rational numbers of the form 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 any $a, b, c, d \in \mathbb{Z}$ 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 element of $\mathbb{Q}^{(m)}$ is 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* 9.24.** If R and S are non-trivial rings, then $R' := R \times \{0_S\}$ is not a subring of $R \times S$: although it satisfies the first two requirements of the definition of a subring, it does not satisfy the third. However, R' does contain an element that acts as a multiplicative identity of R', namely $(1_R, 0_S)$, and hence could be viewed as a subring of $R \times S$ under a more liberal definition. □

Theorem 9.9. *Any subring of an integral domain is also an integral domain.*

Proof. If D' is a subring of the integral domain D, then any zero divisor in D' would itself be a zero divisor in D. □

Note that 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'.

***Example* 9.25.** $\mathbb{Q}$ is a subfield of $\mathbb{R}$, which in turn is a subfield of $\mathbb{C}$. □

EXERCISE 9.8. 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 9.9. Show that if S and S' are subrings of R, then so is $S \cap S'$.

EXERCISE 9.10. Let $\mathcal{F}$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$, and let $\mathcal{C}$ be the subset of $\mathcal{F}$ of continuous functions.

(a) Show that with addition and multiplication of functions defined in the natural, point-wise fashion, $\mathcal{F}$ is a ring, but not an integral domain.

(b) Let $a, b \in \mathcal{F}$. Show that if $a \mid b$ and $b \mid a$, then there is a $c \in \mathcal{F}^*$ such that $a = bc$.

(c) Show that $\mathcal{C}$ is a subring of $\mathcal{F}$, and show that all functions in $\mathcal{C}^*$ are either everywhere positive or everywhere negative.

(d) Define $a, b \in \mathcal{C}$ by $a(t) = b(t) = t$ for $t < 0$, $a(t) = b(t) = 0$ for $0 \leq t \leq 1$, and $a(t) = -b(t) = t - 1$ for $t > 1$. Show that in the ring $\mathcal{C}$, we have $a \mid b$ and $b \mid a$, yet there is no $c \in \mathcal{C}^*$ such that $a = bc$. Thus, part (ii) of Theorem 9.4 does not hold in a general ring.

9.2 Polynomial rings

If R is a ring, then we can form the **ring of polynomials** $R[\mathtt{X}]$, consisting of all polynomials $a_0 + a_1\mathtt{X} + \cdots + a_k\mathtt{X}^k$ in the indeterminate, or "formal" variable, $\mathtt{X}$, with coefficients in R, and with addition and multiplication being defined in the usual way.

***Example* 9.26.** Let us define a few polynomials over the ring $\mathbb{Z}$:

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

We have

$$a+b = 4+2\mathtt{X}+\mathtt{X}^2-\mathtt{X}^3,\ a\cdot b = 3+6\mathtt{X}+\mathtt{X}^2-\mathtt{X}^3-\mathtt{X}^5,\ cd+ef = 5+5\mathtt{X}+4\mathtt{X}^4.$$ □

As illustrated in the previous example, elements of R are also polynomials. Such polynomials are called **constant polynomials**; all other polynomials are called **non-constant polynomials**. The set R of constant polynomials clearly forms a subring of $R[\mathtt{X}]$. In particular, 0_R is the additive identity in $R[\mathtt{X}]$ and 1_R is the multiplicative identity in $R[\mathtt{X}]$.

For completeness, we present a more formal definition of the ring $R[\mathtt{X}]$. The reader should bear in mind that this formalism is rather tedious, and may be more distracting than it is enlightening. It is technically convenient to view a polynomial as having an *infinite* sequence of coefficients $a_0, a_1, a_2, \ldots$, where each coefficient belongs to R, but where only a finite number of the coefficients are non-zero. We may write such a polynomial as an infinite sum $\sum_{i=0}^{\infty} a_i \mathtt{X}^i$; however, this notation is best thought of "syntactic sugar": there is really nothing more to the polynomial than this sequence of coefficients. With this notation, if

$$a = \sum_{i=0}^{\infty} a_i \mathtt{X}^i \quad \text{and} \quad b = \sum_{i=0}^{\infty} b_i \mathtt{X}^i,$$

then

$$a + b := \sum_{i=0}^{\infty} (a_i + b_i) \mathtt{X}^i, \tag{9.1}$$

and

$$a \cdot b := \sum_{i=0}^{\infty} \Bigg(\sum_{k=0}^{i} a_k b_{i-k} \Bigg) \mathtt{X}^i. \tag{9.2}$$

We should first verify that these addition and multiplication operations actually produce coefficient sequences with only a finite number of non-zero terms. Suppose that for non-negative integers k and ℓ, we have $a_i = 0_R$ for all $i > k$ and $b_i = 0_R$ for all $i > \ell$. Then it is clear that the coefficient of $\mathtt{X}^i$ in $a + b$ is zero for all $i > \max\{k, \ell\}$, and it is also not too hard to see that the coefficient of $\mathtt{X}^i$ in $a \cdot b$ is zero for all $i > k + \ell$.

We leave it to the reader to verify that $R[\mathtt{X}]$, with addition and multiplication defined as above, actually satisfies the definition of a ring—this is entirely straightforward, but tedious.

For $c \in R$, we may identify c with the polynomial $\sum_{i=0}^{\infty} c_i \mathtt{X}^i$, where $c_0 = c$ and $c_i = 0_R$ for $i > 0$. Strictly speaking, c and $\sum_{i=0}^{\infty} c_i \mathtt{X}^i$ are not the same mathematical object, but there will certainly be no possible confusion in treating them as such. Thus, from a narrow, legalistic point of view, R is not a subring of $R[\mathtt{X}]$, but we shall not let such let such annoying details prevent us from continuing to speak of it as such. As one last matter of notation, we may naturally write $\mathtt{X}$ to denote the polynomial $\sum_{i=0}^{\infty} a_i \mathtt{X}^i$, where $a_1 = 1_R$ and $a_i = 0_R$ for all $i \neq 1$.

With all of these conventions and definitions, we can return to the practice of writing polynomials as we did in Example 9.26, without any loss of precision. Note that by definition, if R is the trivial ring, then so is $R[\mathtt{X}]$.

9.2.1 Polynomials versus polynomial functions

Of course, a polynomial $a = \sum_{i=0}^{k} a_i \mathtt{X}^i$ defines a polynomial function on R that sends $\alpha \in R$ to $\sum_{i=0}^{k} a_i \alpha^i$, and we denote the value of this function as $a(\alpha)$. However, 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. 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* 9.27.** In the ring $\mathbb{Z}_p$, for prime p, by Fermat's little theorem (Theorem 2.16), we have $\alpha^p - \alpha = [0]_p$ for all $\alpha \in \mathbb{Z}_p$. But consider the polynomial $a := \mathtt{X}^p - \mathtt{X} \in \mathbb{Z}_p[\mathtt{X}]$. We have $a(\alpha) = [0]_p$ for all $\alpha \in \mathbb{Z}_p$, and hence the function defined by a is the zero function, yet a is definitely *not* the zero polynomial. □

More generally, if R is a subring of a ring E, a polynomial $a = \sum_{i=0}^{k} a_i \mathtt{X}^i \in R[\mathtt{X}]$ 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 the value of this function is denoted $a(\alpha)$.

If $E = R[\mathtt{X}]$, then evaluating a polynomial $a \in R[\mathtt{X}]$ at a point $\alpha \in E$ amounts to polynomial composition. For example, if $a = \mathtt{X}^2 + \mathtt{X}$ then

$$a\big[\ \mathtt{X} + 1\ \big] = (\mathtt{X} + 1)^2 + (\mathtt{X} + 1) = \mathtt{X}^2 + 3\mathtt{X} + 2.$$

A simple, but important, fact is the following:

Theorem 9.10. *Let R be a subring of a ring E. For $a, b \in R[\mathtt{X}]$ and $\alpha \in E$, if $p := ab \in R[\mathtt{X}]$ and $s := a + b \in R[\mathtt{X}]$, then we have*

$$p(\alpha) = a(\alpha)b(\alpha) \quad \text{and} \quad s(\alpha) = a(\alpha) + b(\alpha).$$

Also, if $c \in R[\mathtt{X}]$ is a constant polynomial, then $c(\alpha) = c$ for all $\alpha \in E$.

Proof. Exercise. □

> Note that the syntax for polynomial evaluation creates some potential ambiguities: if a is a polynomial, one could interpret $a(b+c)$ as either a times $b + c$, or a evaluated at $b + c$; usually, the meaning will be clear from context, but to avoid such ambiguities, if the intended meaning is the former, we shall generally write this as, say, $a \cdot (b + c)$ or $(b + c)a$, and if the intended meaning is the latter, we shall generally write this as $a[\ b + c\ ]$.
>
> So as to keep the distinction between ring elements and indeterminates clear, we shall use the symbol "$\mathtt{X}$" only to denote the latter. Also, for a polynomial $a \in R[\mathtt{X}]$, we shall in general write this simply

as "a," and not as "$a(\mathtt{X})$." Of course, the choice of the symbol "$\mathtt{X}$" is arbitrary; occasionally, we may use other symbols, such as "$\mathtt{Y}$," as alternatives.

9.2.2 Basic properties of polynomial rings

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

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

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

In particular, $\deg(ab) \leq \deg(a) + \deg(b)$. If either of a_k or b_ℓ are not zero divisors, then $a_k b_\ell$ is not zero, and hence $\deg(ab) = \deg(a) + \deg(b)$. However, if both a_k and b_ℓ are zero divisors, then we may have $a_k b_\ell = 0_R$, in which case, the product ab may be zero, or perhaps $ab \neq 0_R$ but $\deg(ab) < \deg(a) + \deg(b)$.

***Example* 9.28.** Over the ring $\mathbb{Z}_6$, consider the polynomials $a := [1] + [2]\mathtt{X}$ and $b = [1] + [3]\mathtt{X}$. We have $ab = [1] + [5]\mathtt{X} + [6]\mathtt{X}^2 = [1] + [5]\mathtt{X}$. Thus, $\deg(ab) = 1 < 2 = \deg(a) + \deg(b)$. □

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 $a, b \in R[\mathtt{X}]$, we have $\deg(ab) \leq \deg(a) + \deg(b)$, with equality guaranteed to hold unless the leading coefficients of both a and b are zero divisors.*

In the case where the ring of coefficients is as integral domain, we can say significantly more:

Theorem 9.11. *Let D be an integral domain. Then:*

(i) for all $a, b \in D[\mathtt{X}]$, we have $\deg(ab) = \deg(a) + \deg(b)$;

(ii) $D[\mathtt{X}]$ is an integral domain;

(iii) $(D[\mathtt{X}])^* = D^*$.

Proof. Exercise. □

9.2.3 Division with remainder

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

Theorem 9.12 (Division with remainder property). *Let R be a ring. For $a, b \in R[\mathtt{X}]$ with $b \neq 0_R$ and $\mathrm{lc}(b) \in R^*$, there exist unique $q, r \in R[\mathtt{X}]$ such that $a = bq + r$ and $\deg(r) < \deg(b)$.*

Proof. Consider the set S of polynomials of the form $a - zb$ with $z \in R[\mathtt{X}]$. Let $r = a - qb$ be an element of S of minimum degree. We must have $\deg(r) < \deg(b)$, since otherwise, we would have $r' := r - (\mathrm{lc}(r)\,\mathrm{lc}(b)^{-1}\mathtt{X}^{\deg(r)-\deg(b)}) \cdot b \in S$, and $\deg(r') < \deg(r)$, contradicting the minimality of $\deg(r)$.

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

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

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

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

As a consequence of the above theorem, we have:

Theorem 9.13. *For a ring R and $a \in R[\mathtt{X}]$ and $\alpha \in R$, $a(\alpha) = 0_R$ if and only if $(\mathtt{X} - \alpha)$ divides a.*

Proof. If R is the trivial ring, there is nothing to prove, so assume that R is non-trivial. Let us write $a = (\mathtt{X} - \alpha)q + r$, with $q, r \in R[\mathtt{X}]$ and $\deg(r) < 1$, which means that $r \in R$. Then we have $a(\alpha) = (\alpha - \alpha)q(\alpha) + r = r$. Thus, $a(\alpha) = 0_R$ if and only if $a \bmod (\mathtt{X} - \alpha) = 0_R$, which holds if and only if $\mathtt{X} - \alpha$ divides a. □

With R, a, α as in the above theorem, we say that α is a **root** of a if $a(\alpha) = 0_R$.

Theorem 9.14. *Let D be an integral domain, and let $a \in D[\mathtt{X}]$, with $\deg(a) = k \geq 0$. Then a has at most k roots.*

Proof. We can prove this by induction. If $k = 0$, this means that a is a non-zero element of D, and so it clearly has no roots.

Now suppose that $k > 0$. If a has no roots, we are done, so suppose that

a has a root α. Then we can write $a = (\mathtt{X} - \alpha)q$, where $\deg(q) = k - 1$. Now, for any root β of a with $\beta \neq \alpha$, we have $0_D = a(\beta) = (\beta - \alpha)q(\beta)$, and using the fact that D is an integral domain, we must have $q(\beta) = 0_D$. Thus, the only roots of a are α and the roots of q. By induction, q has at most $k - 1$ roots, and hence a has at most k roots. □

Theorem 9.14 has many applications, among which is the following beautiful theorem that establishes an important property of the multiplicative structure of an integral domain:

Theorem 9.15. *Let D be an integral domain and G a subgroup of D^* of finite order. Then G is cyclic.*

Proof. Let n be the order of G, and suppose G is not cyclic. Then by Theorem 8.40, we have that the exponent m of G is strictly less than n. It follows that $\alpha^m = 1_D$ for all $\alpha \in G$. That is, all the elements of G are roots of the polynomial $\mathtt{X}^m - 1_D \in D[\mathtt{X}]$. But since a polynomial of degree m over D has at most m roots, this contradicts the fact that $m < n$. □

As a special case of Theorem 9.15, we have:

Theorem 9.16. *For any finite field F, the group F^* is cyclic. In particular, if p is prime, then $\mathbb{Z}_p^*$ is cyclic; that is, there is a primitive root modulo p.*

Exercise 9.11. Let D be an infinite integral domain, and let $a \in D[\mathtt{X}]$. Show that if $a(\alpha) = 0_D$ for all $\alpha \in D$, then $a = 0_D$. Thus, for an infinite integral domain D, there is a one-to-one correspondence between polynomials over D and polynomial functions on D.

Exercise 9.12. This exercise develops a message authentication scheme (see §6.7.2) that allows one to hash long messages using a relatively small set of hash functions. Let F be a finite field of cardinality q and let t be a positive integer. Let $\mathcal{A} := F^{\times t}$ and $\mathcal{Z} := F$. Define a family $\mathcal{H}$ of hash functions from $\mathcal{A}$ to $\mathcal{Z}$ as follows: let $\mathcal{H} := \{h_{\alpha,\beta} : \alpha, \beta \in F\}$, where for all $h_{\alpha,\beta} \in \mathcal{H}$ and all $(a_1, \ldots, a_t) \in \mathcal{A}$, we define

$$h_{\alpha,\beta}(a_1, \ldots, a_t) := \beta + \sum_{i=1}^{t} a_i \alpha^i \in \mathcal{Z}.$$

Show that $\mathcal{H}$ is a t/q-forgeable message authentication scheme. (Compare this with the second pairwise independent family of hash functions discussed in Example 6.25, which is much larger, but which is only $1/q$-forgeable; in practice, using the smaller family of hash functions with a somewhat higher forging probability may be a good trade-off.)

EXERCISE 9.13. This exercise develops an alternative proof of Theorem 9.15. Let n be the order of the group. Using Theorem 9.14, 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 $\phi(d)$. Now use Theorem 2.11 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 $\phi(d)$.

EXERCISE 9.14. Let F be a field of characteristic other than 2, so that the $2_F \neq 0_F$. Show that the familiar **quadratic formula** holds for F. That is, for $a, b, c \in F$ with $a \neq 0_F$, the polynomial $f := a\mathtt{X}^2 + b\mathtt{X} + c \in F[\mathtt{X}]$ has a root if and only if there exists $z \in F$ such that $z^2 = d$, where d is the **discriminant** of f, defined as $d := b^2 - 4ac$, and in this case the roots of f are

$$\frac{-b \pm z}{2a}.$$

EXERCISE 9.15. Let R be a ring, let $a \in R[\mathtt{X}]$, with $\deg(a) = k \geq 0$, and let α be an element of R.

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

$$a = (\mathtt{X} - \alpha)^m q \text{ and } \ q(\alpha) \neq 0_R.$$

(b) Show that the values m and q in part (a) are uniquely determined (by a and α).

(c) Show that $m > 0$ if and only if α is a root of a.

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

The following exercise refines Theorem 9.14, taking into account multiplicities.

EXERCISE 9.16. Let D be an integral domain, and let $a \in D[\mathtt{X}]$, with $\deg(a) = k \geq 0$. Show that

$$\sum_{\alpha \in D} m_\alpha(a) \leq k.$$

EXERCISE 9.17. Let D be an integral domain, let $a, b \in D[\mathtt{X}]$, and let $\alpha \in D$. Show that $m_\alpha(ab) = m_\alpha(a) + m_\alpha(b)$.

EXERCISE 9.18. Let R be a ring, let $a \in R[\mathtt{X}]$, with $\deg(a) = k \geq 0$, let $\alpha \in R$, and let $m := m_\alpha(a)$. Show that if we evaluate a at $\mathtt{X} + \alpha$, we have

$$a\big[\ \mathtt{X} + \alpha\ \big] = \sum_{i=m}^{k} b_i \mathtt{X}^i,$$

where $b_m, \ldots, b_k \in R$ and $b_m \neq 0_R$.

9.2.4 Formal derivatives

Let R be any ring, and let $a \in R[\mathtt{X}]$ be a polynomial. If $a = \sum_{i=0}^{\ell} a_i \mathtt{X}^i$, we define the **formal derivative** of a as

$$\mathbf{D}(a) := \sum_{i=1}^{\ell} i a_i \mathtt{X}^{i-1}.$$

We stress that unlike the "analytical" notion of derivative from calculus, which is defined in terms of limits, this definition is purely "symbolic." Nevertheless, some of the usual rules for derivatives still hold:

Theorem 9.17. *Let R be a ring. For all $a, b \in R[\mathtt{X}]$ and $c \in R$, we have*

(i) $\mathbf{D}(a + b) = \mathbf{D}(a) + \mathbf{D}(b)$;

(ii) $\mathbf{D}(ca) = c\mathbf{D}(a)$;

(iii) $\mathbf{D}(ab) = \mathbf{D}(a)b + a\mathbf{D}(b)$.

Proof. Parts (i) and (ii) follow immediately by inspection, but part (iii) requires some proof. First, note that part (iii) holds trivially if either a or b are zero, so let us assume that neither are zero.

We first prove part (iii) for **monomials**, that is, polynomials of the form $c\mathtt{X}^i$ for non-zero $c \in R$ and $i \geq 0$. Suppose $a = c\mathtt{X}^i$ and $b = d\mathtt{X}^j$. If $i = 0$, so $a = c$, then the result follows from part (ii) and the fact that $\mathbf{D}(c) = 0$; when $j = 0$, the result holds by a symmetric argument. So assume that $i > 0$ and $j > 0$. Now, $\mathbf{D}(a) = ic\mathtt{X}^{i-1}$ and $\mathbf{D}(b) = jd\mathtt{X}^{j-1}$, and $\mathbf{D}(ab) = \mathbf{D}(cd\mathtt{X}^{i+j}) = (i+j)cd\mathtt{X}^{i+j-1}$. The result follows from a simple calculation.

Having proved part (iii) for monomials, we now prove it in general on induction on the total number of monomials appearing in a and b. If the total number is 2, then both a and b are monomials, and we are in the base case; otherwise, one of a and b must consist of at least two monomials, and for concreteness, say it is b that has this property. So we can write $b = b_1 + b_2$, where both b_1 and b_2 have fewer monomials than does b. Applying part (i)

and the induction hypothesis for part (iii), we have

$$\begin{aligned}\mathbf{D}(ab) &= \mathbf{D}(ab_1 + ab_2)\\ &= \mathbf{D}(ab_1) + \mathbf{D}(ab_2)\\ &= \mathbf{D}(a)b_1 + a\mathbf{D}(b_1) + \mathbf{D}(a)b_2 + a\mathbf{D}(b_2)\\ &= \mathbf{D}(a)\cdot(b_1 + b_2) + a\cdot(\mathbf{D}(b_1) + \mathbf{D}(b_2))\\ &= \mathbf{D}(a)\cdot(b_1 + b_2) + a\cdot\mathbf{D}(b_1 + b_2)\\ &= \mathbf{D}(a)b + a\mathbf{D}(b). \quad \square\end{aligned}$$

Exercise 9.19. Let R be a ring, let $a \in R[\mathtt{X}]$, and let $\alpha \in R$ be a root of a. Show that α is a multiple root of a if and only if α is a root of $\mathbf{D}(a)$ (see Exercise 9.15).

Exercise 9.20. Let R be a ring, let $a \in R[\mathtt{X}]$ with $\deg(a) = k \geq 0$, and let $\alpha \in R$. Show that if we evaluate a at $\mathtt{X} + \alpha$, writing

$$a\big[\ \mathtt{X} + \alpha\ \big] = \sum_{i=0}^{k} b_i \mathtt{X}^i,$$

with $b_0, \ldots, b_k \in R$, then we have

$$i! \cdot b_i = (\mathbf{D}^i(a))(\alpha) \quad \text{for } i = 0, \ldots, k.$$

Exercise 9.21. Let F be a field such that every non-constant polynomial $a \in F[\mathtt{X}]$ has a root $\alpha \in F$. (The field $\mathbb{C}$ is an example of such a field, an important fact which we shall not be proving in this text.) Show that for every positive integer r that is not a multiple of the characteristic of F, there exists an element $\zeta \in F^*$ of multiplicative order r, and that every element in F^* whose order divides r is a power of ζ.

9.2.5 Multi-variate polynomials

One can naturally generalize the notion of a polynomial in a single variable to that of a polynomial in several variables. We discuss these ideas briefly here—they will play only a minor role in the remainder of the text.

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

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

$$a = \sum_{j=0}^{\ell} \Big(\sum_{i=0}^{k} a_{ij} \mathtt{X}^i \Big) \mathtt{Y}_j. \tag{9.3}$$

Rearranging terms, this may also be written as

$$a = \sum_{\substack{0 \le i \le k \\ 0 \le j \le \ell}} a_{ij} \mathtt{X}^i \mathtt{Y}^j, \tag{9.4}$$

or as

$$a = \sum_{i=0}^{k} \Big(\sum_{j=0}^{\ell} a_{ij} \mathtt{Y}^j \Big) \mathtt{X}^j. \tag{9.5}$$

If a is written as in (9.4), the terms $a_{ij}\mathtt{X}^i\mathtt{Y}^j$ with $a_{ij} \neq 0_R$ are called **monomials**. The **total degree** of such a monomial $a_{ij}\mathtt{X}^i\mathtt{Y}^j$ is defined to be $i + j$, and if a is non-zero, then the **total degree** of a, denoted $\mathrm{Deg}(a)$, is defined to be the maximum total degree of any monomial appearing in (9.4). We define the total degree of the zero polynomial to be $-\infty$. The reader may verify that for any $a, b \in R[\mathtt{X}, \mathtt{Y}]$, we have $\mathrm{Deg}(ab) \le \mathrm{Deg}(a) + \mathrm{Deg}(b)$, while equality holds if R is an integral domain.

When a is written as in (9.5), one sees that we can naturally view a as an element of $R[\mathtt{Y}][\mathtt{X}]$, that is, as a polynomial in X whose coefficients are polynomials in Y. From a strict, syntactic point of view, the rings $R[\mathtt{Y}][\mathtt{X}]$ and $R[\mathtt{X}][\mathtt{Y}]$ are not the same, but there is no harm done in blurring this distinction when convenient. We denote by $\deg_{\mathtt{X}}(a)$ the degree of a, viewed as a polynomial in $\mathtt{X}$, and by $\deg_{\mathtt{Y}}(a)$ the degree of a, viewed as a polynomial in $\mathtt{Y}$. Analogously, one can formally differentiate a with respect to either $\mathtt{X}$ or $\mathtt{Y}$, obtaining the "partial" derivatives $\mathbf{D}_{\mathtt{X}}(a)$ and $\mathbf{D}_{\mathtt{Y}}(a)$.

***Example* 9.29.** Let us illustrate, with a particular example, the three different forms—as in (9.3), (9.4), and (9.5)—of expressing a bivariate polynomial. In the ring $\mathbb{Z}[\mathtt{X}, \mathtt{Y}]$ we have

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

We have $\mathrm{Deg}(a) = 3$, $\deg_{\mathtt{X}}(a) = 2$, and $\deg_{\mathtt{Y}}(a) = 1$. □

More generally, if $\mathtt{X}_1, \ldots, \mathtt{X}_n$ are indeterminates, we can form the ring

$R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ of **multi-variate polynomials** in n variables over R. Formally, we can think of this ring as $R[\mathtt{X}_1][\mathtt{X}_2] \cdots [\mathtt{X}_n]$. Any multi-variate polynomial can be expressed uniquely as the sum of monomials of the form $c\mathtt{X}_1^{e_1} \cdots \mathtt{X}_n^{e_n}$ for non-zero $c \in R$ and non-negative integers $e_1, \ldots, e_n$; the total degree of such a monomial is defined to be $\sum_i e_i$, and the total degree of a multi-variate polynomial a, denoted $\mathrm{Deg}(a)$, is defined to be the maximum degree of its monomials. As above, for $a, b \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$, we have $\mathrm{Deg}(ab) \leq \mathrm{Deg}(a) + \mathrm{Deg}(b)$, while equality always holds if R is an integral domain.

Just as for bivariate polynomials, the order of the indeterminates is not important, and for any $i = 1, \ldots, n$, one can naturally view any $a \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ as a polynomial in $\mathtt{X}_i$ over the ring $R[\mathtt{X}_1, \ldots, \mathtt{X}_{i-1}, \mathtt{X}_{i+1}, \ldots, \mathtt{X}_n]$, and define $\deg_{\mathtt{X}_i}(a)$ to be the degree of a when viewed in this way. Analogously, one can formally differentiate a with respect to any variable $\mathtt{X}_i$, obtaining the "partial" derivative $\mathbf{D}_{\mathtt{X}_i}(a)$.

Just as polynomials in a single variable define polynomial functions, so do polynomials in several variables. If R is a subring of E, $a \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$, and $\alpha = (\alpha_1, \ldots, \alpha_n) \in E^{\times n}$, we define $a(\alpha)$ to be the element of E obtained by evaluating the expression obtained by substituting α_i for $\mathtt{X}_i$ in a. Theorem 9.10 carries over directly to the multi-variate case.

Exercise 9.22. Let R be a ring, and let $\alpha_1, \ldots, \alpha_n$ be elements of R. Show that any polynomial $a \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ can be expressed as

$$a = (\mathtt{X}_1 - \alpha_1)q_1 + \cdots + (\mathtt{X}_n - \alpha_n)q_n + r,$$

where $q_1, \ldots, q_n \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ and $r \in R$. Moreover, show that the value of r appearing in such an expression is uniquely determined (by a and $\alpha_1, \ldots, \alpha_n$).

Exercise 9.23. This exercise generalizes Theorem 9.14. Let D be an integral domain, and let $a \in D[\mathtt{X}_1, \ldots, \mathtt{X}_n]$, with $\mathrm{Deg}(a) = k \geq 0$. Let T be a finite subset of D. Show that the number of elements $\alpha \in T^{\times n}$ such that $a(\alpha) = 0$ is at most $k|T|^{n-1}$.

Exercise 9.24. Let F be a finite field of cardinality q, and let t be a positive integer. Let $\mathcal{A} := F^{\times t}$ and $\mathcal{Z} := F$. Use the result of the previous exercise to construct a family $\mathcal{H}$ of hash functions from $\mathcal{A}$ to $\mathcal{Z}$ that is an $O(\mathrm{len}(t)/q)$-forgeable message authentication scheme, where $\log_q |\mathcal{H}| = \mathrm{len}(t) + O(1)$. (See §6.7.2 and also Exercise 9.12.)

9.3 Ideals and quotient rings

Definition 9.18. *Let R be a ring. An **ideal of** R is a subgroup I of the additive group of R that is closed under multiplication by elements of R, that is, for all $a \in I$ and $r \in R$, we have $ar \in I$.*

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

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

Observe that the condition $-a \in I$ is redundant, as it is implied by the condition $ar \in I$ with $r = -1_R$. Note that 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* 9.30.** For $m \in \mathbb{Z}$, the set $m\mathbb{Z}$ is not only a subgroup of the additive group $\mathbb{Z}$, it is also an ideal of the ring $\mathbb{Z}$. □

***Example* 9.31.** For $m \in \mathbb{Z}$, the set $m\mathbb{Z}_n$ is not only a subgroup of the additive group $\mathbb{Z}_n$, it is also an ideal of the ring $\mathbb{Z}_n$. □

***Example* 9.32.** 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 polynomial $R[\mathtt{X}]$. Suppose a is a non-zero polynomial in $R[\mathtt{X}]$. The additive subgroup generated by a consists of polynomials whose degrees are at most that of a. However, this subgroup is not an ideal, since any ideal containing a must also contain $a \cdot \mathtt{X}^i$ for all $i \geq 0$, and must therefore contain polynomials of arbitrarily high degree. □

Let $a_1, \ldots, a_k$ be elements of a ring R. Then it is easy to see that the set

$$a_1R + \cdots + a_kR := \{a_1r_1 + \cdots + a_kr_k : r_1, \ldots, r_k \in R\}$$

is an ideal of R, and contains $a_1, \ldots, a_k$. It is called the **ideal of R generated by** $a_1, \ldots, a_k$. Clearly, any ideal I of R that contains $a_1, \ldots, a_k$ must contain $a_1R + \cdots + a_kR$, and in this sense, $a_1R + \cdots + a_kR$ is the smallest ideal of R containing $a_1, \ldots, a_k$. An alternative notation that is often used is to write $(a_1, \ldots, a_k)$ to denote the ideal generated by $a_1, \ldots, a_k$, when the ring R is clear from context. If an ideal I is of the form $aR = \{ar : r \in R\}$ for some $a \in R$, then we say that I is a **principal ideal**.

Note that if I and J are ideals of a ring R, then so are $I + J := \{x + y : x \in I, y \in J\}$ and $I \cap J$ (verify).

Since an ideal I of a ring R is a subgroup of the additive group R, we may adopt the congruence notation in §8.3, writing $a \equiv b \pmod{I}$ if and only if $a - b \in I$.

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 $a \equiv b \pmod{d}$.

Just considering R as an additive group, then as we saw in §8.3, we can form the additive group R/I of cosets, where $(a + I) + (b + I) := (a + b) + I$. By also considering the multiplicative structure of R, we can view R/I as a ring. To do this, we need the following fact:

Theorem 9.19. *Let I be an ideal of a ring R, and let $a, a', b, b' \in R$. 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 $x \in I$ and $b' = b + y$ for $y \in I$, then $a'b' = ab + ay + bx + xy$. Since I is closed under multiplication by elements of R, we see that $ay, bx, xy \in I$, and since it is closed under addition, $ay + bx + xy \in I$. Hence, $a'b' - ab \in I$. □

This theorem is perhaps one of the main motivations for the definition of an ideal. It allows us to define multiplication on R/I as follows: for $a, b \in R$,

$$(a + I) \cdot (b + I) := ab + I.$$

The above theorem is required to show that this definition is unambiguous. Once that is done, it is straightforward to show that all the properties that make R a ring are inherited by R/I — we leave the details of this to the reader. In particular, 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**. As a matter of notation, for $a \in R$, we define $[a]_I := a + I$, and if $I = dR$, we may write this simply as $[a]_d$. If I is clear from context, we may also just write $[a]$.

***Example* 9.33.** For $n \geq 1$, the ring $\mathbb{Z}_n$ is precisely the quotient ring $\mathbb{Z}/n\mathbb{Z}$. □

***Example* 9.34.** Let f be a monic polynomial over a ring R with $\deg(f) = \ell \geq 0$, and consider the quotient ring $E := R[\mathtt{X}]/fR[\mathtt{X}]$. By the division with remainder property for polynomials (Theorem 9.12), for every $a \in R[\mathtt{X}]$, there exists a unique polynomial $b \in R[\mathtt{X}]$ such that $a \equiv b \pmod{f}$ and

$\deg(b) < \ell$. From this, it follows that every element of E can be written uniquely as $[b]_f$, where $b \in R[\mathtt{X}]$ is a polynomial of degree less than ℓ.

The assumption that f is monic may be relaxed a bit: all that really matters in this example is that the leading coefficient of f is a unit, so that the division with remainder property applies. Also, note that in this situation, we will generally prefer the more compact notation $R[\mathtt{X}]/(f)$, instead of $R[\mathtt{X}]/fR[\mathtt{X}]$. □

***Example* 9.35.** Consider the polynomial $f := \mathtt{X}^2 + \mathtt{X} + 1 \in \mathbb{Z}_2[\mathtt{X}]$ and the quotient ring $E := \mathbb{Z}_2[\mathtt{X}]/(f)$. Let us name the elements of E as follows:

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

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 $\mathtt{10} \cdot \mathtt{11}$, using the identity $\mathtt{X}^2 \equiv \mathtt{X} + 1 \pmod{f}$, one sees that

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

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

$\cdot$	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. By Theorem 9.16, we know that E^* must be cyclic (this fact also follows from Theorem 8.32, and the fact that $|E^*| = 3$.) Indeed, 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 9.25. 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 $f \in R[\mathtt{X}]$. Show that $f(x) \equiv f(y) \pmod{I}$.

EXERCISE 9.26. Let p be a prime, and consider the ring $\mathbb{Q}^{(p)}$ (see Example 9.23). Show that any non-zero ideal of $\mathbb{Q}^{(p)}$ is of the form (p^i), for some uniquely determined integer $i \geq 0$.

EXERCISE 9.27. Let R be a ring. Show that if I is a non-empty subset of $R[\mathtt{X}]$ that is closed under addition, multiplication by elements of R, and multiplication by $\mathtt{X}$, then I is an ideal of $R[\mathtt{X}]$.

For the following three exercises, we need some definitions. An ideal I of a ring 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 a ring R is called **maximal** if $I \subsetneq R$ and there are no ideals J of R such that $I \subsetneq J \subsetneq R$.

EXERCISE 9.28. Let R be a ring. 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 9.29. This exercise explores some examples of prime and maximal ideals.

(a) Show that 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) More generally, show that in an integral domain D, the ideal $\{0\}$ is prime, and this ideal is maximal if and only if D is a field.

(c) Show that in the ring $F[\mathtt{X}, \mathtt{Y}]$, where F is a field, the ideal $(\mathtt{X}, \mathtt{Y})$ is maximal, while the ideals $(\mathtt{X})$ and $(\mathtt{Y})$ are prime, but not maximal.

EXERCISE 9.30. It is a fact that all non-trivial rings R contain at least one maximal ideal. Showing this in general requires some fancy set-theoretic notions. This exercise develops a proof in the case where R is countable (i.e., finite or countably infinite).

(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 $i \geq 0$, 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

$$I := \bigcup_{i=0}^{\infty} I_i.$$

Show that I is a maximal ideal of R. Hint: first show that I is an ideal; then show that $I \subsetneq R$ by assuming that $1_R \in I$ and deriving a contradiction; finally, 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.

For the following three exercises, we need the following definition: for subsets X, Y of a ring R, let $X \cdot Y$ denote the set of all finite sums of the form

$$x_1 y_1 + \cdots + x_\ell y_\ell \quad (\text{with } x_k \in X,\ y_k \in Y \text{ for } k = 1, \ldots, \ell, \text{ for some } \ell \geq 0).$$

Note that $X \cdot Y$ contains 0_R (the "empty" sum, with $\ell = 0$).

Exercise 9.31. Let R be a ring, and S a subset of R. Show that $S \cdot R$ is an ideal of R, and is the smallest ideal of R containing S.

Exercise 9.32. Let I and J be two ideals of a ring R. Show that:

(a) $I \cdot J$ is an ideal;

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

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

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

Exercise 9.33. Let S be a subring of a ring R. Let I be an ideal of R, and J an ideal of S. Show that:

(a) $I \cap S$ is an ideal of S, and that $(I \cap S) \cdot R$ is an ideal of R contained in I;

(b) $(J \cdot R) \cap S$ is an ideal of S containing J.

9.4 Ring homomorphisms and isomorphisms

Definition 9.20. *A function ρ from a ring R to a ring R' is called a* ***ring homomorphism*** *if it is a group homomorphism with respect to the underlying additive groups of R and R', and if in addition,*

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

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

Expanding the definition, we see that 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 $\rho(1_R) = 1_{R'}$.

Since a ring homomorphism ρ from R to R' is also an additive group homomorphism, we may also adopt the notation and terminology for image and kernel, and note that all the results of Theorem 8.20 apply as well here. In particular, $\rho(0_R) = 0_{R'}$, $\rho(a) = \rho(b)$ if and only if $a \equiv b \pmod{\ker(\rho)}$, and ρ is injective if and only if $\ker(\rho) = \{0_R\}$. However, we may strengthen Theorem 8.20 as follows:

Theorem 9.21. *Let $\rho : R \to R'$ be a ring homomorphism.*

(i) For any subring S of R, $\rho(S)$ is a subring of R'.

(ii) For any ideal I of R, $\rho(I)$ is an ideal of $\mathrm{img}(\rho)$.

(iii) $\ker(\rho)$ *is an ideal of R.*

(iv) For any ideal I' of R', $\rho^{-1}(I')$ is an ideal of R.

Proof. Exercise. □

Theorems 8.21 and 8.22 have natural ring analogs—one only has to show that the corresponding group homomorphisms are also ring homomorphisms:

Theorem 9.22. *If $\rho : R \to R'$ and $\rho' : R' \to R''$ are ring homomorphisms, then so is their composition $\rho' \circ \rho : R \to R''$.*

Proof. Exercise. □

Theorem 9.23. *Let $\rho_i : R \to R_i$, for $i = 1, \ldots, n$, be ring homomorphisms. Then the map $\rho : R \to R_1 \times \cdots \times R_n$ that sends $a \in R$ to $(\rho_1(a), \ldots, \rho_n(a))$ is a ring homomorphism.*

Proof. Exercise. □

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 8.24, we have:

Theorem 9.24. *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 ρ of R with R' is essentially just a "renaming" of elements; in particular, ρ maps units to units and zero divisors to zero divisors (verify); moreover, the restriction of the map ρ to R^* yields a *group* isomorphism of R^* with $(R')^*$ (verify).

An injective ring homomorphism $\rho : R \to E$ is called an **embedding** of R in E. In this case, $\mathrm{img}(\rho)$ is a subring of E and $R \cong \mathrm{img}(\rho)$. 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, and as a slight abuse of terminology, we shall say that R as a subring of E.

We have already seen an example of this, namely, when we formally defined the ring of polynomials $R[\mathtt{X}]$ over R, we defined the map $\rho : R \to R[\mathtt{X}]$ that sends $c \in R$ to the polynomial whose constant term is c, and all other coefficients zero. This map ρ is clearly an embedding, and it was via this embedding that we identified elements of R with elements of $R[\mathtt{X}]$, and so viewed R as a subring of $R[\mathtt{X}]$.

This practice of identifying elements of a ring with their images in another ring under a natural embedding is very common. We shall see more examples of this later (in particular, Example 9.43 below).

Theorems 8.25, 8.26, and 8.27 also have natural ring analogs—again, one only has to show that the corresponding group homomorphisms are also ring homomorphisms:

Theorem 9.25. *If I is an ideal of a ring R, then the natural map $\rho : R \to R/I$ given by $\rho(a) = a + I$ is a surjective ring homomorphism whose kernel is I.*

Proof. Exercise. □

Theorem 9.26. *Let ρ be a ring homomorphism from R into R'. Then the map $\bar{\rho} : R/\ker(\rho) \to \mathrm{img}(\rho)$ that sends the coset $a + \ker(\rho)$ for $a \in R$ to $\rho(a)$ is unambiguously defined and is a ring isomorphism of $R/\ker(\rho)$ with $\mathrm{img}(\rho)$.*

Proof. Exercise. □

Theorem 9.27. *Let ρ be a ring homomorphism from R into R'. Then for any ideal I contained in* $\ker(\rho)$*, the map $\bar{\rho} : R/I \to \text{img}(\rho)$ that sends the coset $a + I$ for $a \in R$ to $\rho(a)$ is unambiguously defined and is a ring homomorphism from R/I onto* $\text{img}(\rho)$ *with kernel* $\ker(\rho)/I$.

Proof. Exercise. □

***Example* 9.36.** For $n \geq 1$, the natural map ρ from $\mathbb{Z}$ to $\mathbb{Z}_n$ sends $a \in \mathbb{Z}$ to the residue class $[a]_n$. In Example 8.41, we noted that this is a surjective group homomorphism on the underlying additive groups, with kernel $n\mathbb{Z}$; however, this map is also a ring homomorphism. □

***Example* 9.37.** As we saw in Example 8.42, if $n_1, \ldots, n_k$ are pairwise relatively prime, positive integers, then the map from $\mathbb{Z}$ to $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$ that sends $x \in \mathbb{Z}$ to $([x]_{n_1}, \ldots, [x]_{n_k})$ is a surjective group homomorphism on the underlying additive groups, with kernel $n\mathbb{Z}$, where $n = \prod_{i=1}^{k} n_i$. However, this map is also a ring homomorphism (this follows from Example 9.36 and Theorem 9.23). Therefore, by Theorem 9.26, the map that sends $[x]_n \in \mathbb{Z}_n$ to $([x]_{n_1}, \ldots, [x]_{n_k})$ is a ring isomorphism of the ring $\mathbb{Z}_n$ with the ring $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$. It follows that the restriction of this map to $\mathbb{Z}_n^*$ yields a *group* isomorphism of the *multiplicative* groups $\mathbb{Z}_n^*$ and $\mathbb{Z}_{n_1}^* \times \cdots \times \mathbb{Z}_{n_k}^*$ (see Example 9.13). □

***Example* 9.38.** As we saw in Example 8.43, if n_1, n_2 are positive integers with $n_1 > 1$ and $n_1 \mid n_2$, then the map $\bar{\rho} : \mathbb{Z}_{n_2} \to \mathbb{Z}_{n_1}$ that sends $[a]_{n_2}$ to $[a]_{n_1}$ is a surjective group homomorphism on the underlying additive groups with kernel $n_1\mathbb{Z}_{n_2}$. This map is also a ring homomorphism. The map $\bar{\rho}$ can also be viewed as the map obtained by applying Theorem 9.27 with the natural map ρ from $\mathbb{Z}$ to $\mathbb{Z}_{n_1}$ and the ideal $n_2\mathbb{Z}$ of $\mathbb{Z}$, which is contained in $\ker(\rho) = n_1\mathbb{Z}$. □

***Example* 9.39.** Let R be a subring of a ring E, and fix $\alpha \in E$. The **polynomial evaluation map** $\rho : R[\mathtt{X}] \to E$ that sends $a \in R[\mathtt{X}]$ to $a(\alpha) \in E$ is a ring homomorphism from $R[\mathtt{X}]$ into E (see Theorem 9.10). The image of ρ consists of all polynomial expressions in α with coefficients in R, and is denoted $R[\alpha]$. Note that $R[\alpha]$ is a subring of E containing $R \cup \{\alpha\}$, and is the smallest such subring of E. □

***Example* 9.40.** 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 $\rho : R[\mathtt{X}_1, \ldots, \mathtt{X}_n] \to E$ that sends $a \in R[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ to $a(\alpha_1, \ldots, \alpha_n)$ 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 $R \cup \{\alpha_1, \ldots, \alpha_n\}$, and is the smallest such subring of E. □

***Example* 9.41.** For any ring R, consider the map $\rho : \mathbb{Z} \to R$ that sends $m \in \mathbb{Z}$ to $m \cdot 1_R$ in R. This is clearly a ring homomorphism (verify). If $\ker(\rho) = \{0\}$, then $\mathrm{img}(\rho) \cong \mathbb{Z}$, and so the ring $\mathbb{Z}$ is embedded in R, and R has characteristic zero. If $\ker(\rho) = n\mathbb{Z}$ for $n > 0$, then $\mathrm{img}(\rho) \cong \mathbb{Z}_n$, and so the ring $\mathbb{Z}_n$ is embedded in R, and R has characteristic n. Note that we have $n = 1$ if and only if R is trivial.

Note that $\mathrm{img}(\rho)$ is the smallest subring of R; indeed, since any subring of R must contain 1_R and be closed under addition and subtraction, it must contain $\mathrm{img}(\rho)$. □

***Example* 9.42.** Let R be a ring of prime characteristic p. For any $a, b \in R$, we have (see Exercise 9.2)

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

However, by Exercise 1.12, 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

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

and so it follows that the map $\rho : R \to R$ that sends $a \in R$ to a^p is a ring homomorphism. It also immediately follows that for any integer $e \geq 1$, the e-fold composition $\rho^e : R \to R$ that sends $a \in R$ to a^{p^e} is also a ring homomorphism. □

***Example* 9.43.** As in Example 9.34, let f be a monic polynomial over a ring R with $\deg(f) = \ell$, but now assume that $\ell > 0$. Consider the natural map ρ from $R[\mathtt{X}]$ to the quotient ring $E := R[\mathtt{X}]/(f)$ that sends $a \in R[\mathtt{X}]$ to $[a]_f$. If we restrict ρ to the subring R of $R[\mathtt{X}]$, we obtain an embedding of R into E. Since this is a very natural embedding, one usually simply identifies

elements of R with their images in E under ρ, and regards R as a subring of E. Taking this point of view, we see that if $a = \sum_i a_i \mathtt{X}^i$, then

$$[a]_f = [\sum_i a_i \mathtt{X}^i]_f = \sum_i a_i ([\mathtt{X}]_f)^i = a(\eta),$$

where $\eta := [\mathtt{X}]_f \in E$. Therefore, the map ρ may be viewed as the polynomial evaluation map, as in Example 9.39, that sends $a \in R[\mathtt{X}]$ to $a(\eta) \in E$. Note that we have $E = R[\eta]$; moreover, every element of E can be expressed uniquely as $b(\eta)$ for some $b \in R[\mathtt{X}]$ of degree less than ℓ, and more generally, for arbitrary $a, b \in R[\mathtt{X}]$, we have $a(\eta) = b(\eta)$ if and only if $a \equiv b \pmod{f}$. □

***Example* 9.44.** As a special case of Example 9.43, let $f := \mathtt{X}^2 + 1 \in \mathbb{R}[\mathtt{X}]$, and consider the quotient ring $\mathbb{R}[\mathtt{X}]/(f)$. If we set $i := [\mathtt{X}]_f \in \mathbb{R}[\mathtt{X}]/(f)$, then every element of $\mathbb{R}[\mathtt{X}]/(f)$ can be expressed uniquely as $a+bi$, where $a, b \in \mathbb{R}$. Moreover, we have $i^2 = -1$, and more generally, for $a, b, a', b' \in \mathbb{R}$, 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 $\mathbb{R}[\mathtt{X}]/(f)$ are precisely the familiar rules of complex arithmetic, and so $\mathbb{C}$ and $\mathbb{R}[\mathtt{X}]/(f)$ are essentially the same, as rings. Indeed, the "algebraically correct" way of defining the complex numbers $\mathbb{C}$ is simply to define them to be the quotient ring $\mathbb{R}[\mathtt{X}]/(f)$ in the first place. This will be our point of view from now on. □

***Example* 9.45.** Consider the polynomial evaluation map $\rho : \mathbb{R}[\mathtt{X}] \to \mathbb{C} = R[\mathtt{X}]/(\mathtt{X}^2 + 1)$ that sends $g \in \mathbb{R}[\mathtt{X}]$ to $g(-i)$. For any $g \in \mathbb{R}[\mathtt{X}]$, we may write $g = (\mathtt{X}^2 + 1)q + a + b\mathtt{X}$, where $q \in \mathbb{R}[\mathtt{X}]$ 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 $\mathbb{R}[\mathtt{X}]$ generated by the polynomial $\mathtt{X}^2 + 1$. By Theorem 9.26, we therefore get a ring automorphism $\bar{\rho}$ on $\mathbb{C}$ that sends $a + bi \in \mathbb{C}$ to $a - bi$. In fact, $\bar{\rho}$ it is none other than the complex conjugation map. Indeed, this is the "algebraically correct" way of defining complex conjugation in the first place. □

***Example* 9.46.** We defined the ring $\mathbb{Z}[i]$ of Gaussian integers in Example 9.22 as a subring of $\mathbb{C}$. Let us verify that the notation $\mathbb{Z}[i]$ introduced in Example 9.22 is consistent with that introduced in Example 9.39. Consider the polynomial evaluation map $\rho : \mathbb{Z}[\mathtt{X}] \to \mathbb{C}$ that sends $g \in \mathbb{Z}[\mathtt{X}]$ to $g(i) \in \mathbb{C}$.

For any $g \in \mathbb{Z}[\mathtt{X}]$, we may write $g = (\mathtt{X}^2 + 1)q + a + b\mathtt{X}$, where $q \in \mathbb{Z}[\mathtt{X}]$ 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 $\mathbb{Z}[\mathtt{X}]$ generated by the polynomial $\mathtt{X}^2 + 1$. This shows that $\mathbb{Z}[i]$ in Example 9.22 is the same as $\mathbb{Z}[i]$ in Example 9.39, and moreover, Theorem 9.26 implies that $\mathbb{Z}[i]$ is isomorphic to $\mathbb{Z}[\mathtt{X}]/(\mathtt{X}^2 + 1)$.

Thus, we can directly construct the Gaussian integers as the quotient ring $\mathbb{Z}[\mathtt{X}]/(\mathtt{X}^2 + 1)$. Likewise the field $\mathbb{Q}[i]$ (see Exercise 9.8) can be constructed directly as $\mathbb{Q}[\mathtt{X}]/(\mathtt{X}^2 + 1)$. Such direct constructions are appealing in that they are purely "elementary," as they do not appeal to anything so "sophisticated" as the real numbers. □

***Example* 9.47.** Let p be a prime, and consider the quotient ring $E := \mathbb{Z}_p[\mathtt{X}]/(\mathtt{X}^2 + 1)$. If we set $i := [\mathtt{X}]_{\mathtt{X}^2+1} \in E$, 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, the rules for addition and multiplication in E look exactly the same as they do in $\mathbb{C}$: for $a, b, a', b' \in \mathbb{Z}_p$, we have

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

and

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

Note that E may or may not be a field.

On the one hand, suppose that $c^2 = -1$ for some $c \in \mathbb{Z}_p$ (for example, $p = 2$, $p = 5$, $p = 13$). Then $(c + i)(c - i) = c^2 + 1 = 0$, and so E is not an integral domain.

On the other hand, suppose there is no $c \in \mathbb{Z}_p$ such that $c^2 = -1$ (for example, $p = 3$, $p = 7$). Then for any $a, b \in \mathbb{Z}_p$, 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 $\mathbb{Z}_p$. Since $\mathbb{Z}_p$ is a field, it follows that the same formula for multiplicative inverses in $\mathbb{C}$ applies in E, namely,

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

This construction provides us with more examples of finite fields whose cardinality is not prime. □

***Example* 9.48.** If $\rho : R \to R'$ is a ring homomorphism, then we can extend ρ in a natural way to a ring homomorphism from $R[\mathtt{X}]$ to $R'[\mathtt{X}]$, by defining $\rho(\sum_i a_i \mathtt{X}^i) := \sum_i \rho(a_i)\mathtt{X}^i$. We leave it to the reader to verify that this indeed is a ring homomorphism. □

Exercise 9.34. 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 9.35. Let R_1, R_2 be rings, and let $\rho : R_1 \times R_2 \to R_1$ be the map that sends $(a_1, a_2) \in R_1 \times R_2$ to $a_1 \in R_1$. Show that ρ is a surjective ring homomorphism whose kernel is $\{0_{R_1}\} \times R_2$.

Exercise 9.36. Let ρ be a ring homomorphism from R into R'. Show that the ideals of R containing $\ker(\rho)$ are in one-to-one correspondence with the ideals of $\mathrm{img}(\rho)$, where the ideal I of R containing $\ker(\rho)$ corresponds to the ideal $\rho(I)$ of $\mathrm{img}(\rho)$.

Exercise 9.37. 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 $\rho^* : R^* \to S^*$ whose kernel is $(1_R + \ker(\rho)) \cap R^*$.

Exercise 9.38. Show that if F is a field, then the only ideals of F are $\{0_F\}$ and F. From this, conclude the following: if $\rho : F \to R$ is a ring homomorphism from F into a non-trivial ring R, then ρ must be an embedding.

Exercise 9.39. Let n be a positive integer.

(a) Show that the rings $\mathbb{Z}[\mathtt{X}]/(n)$ and $\mathbb{Z}_n[\mathtt{X}]$ are isomorphic.

(b) Assuming that $n = pq$, where p and q are distinct primes, show that the rings $\mathbb{Z}_n[\mathtt{X}]$ and $\mathbb{Z}_p[\mathtt{X}] \times \mathbb{Z}_q[\mathtt{X}]$ are isomorphic.

Exercise 9.40. Let n be a positive integer, let $f \in \mathbb{Z}[\mathtt{X}]$ be a monic polynomial, and let $\bar{f}$ be the image of f in $\mathbb{Z}_n[\mathtt{X}]$ (i.e., $\bar{f}$ is obtained by applying the natural map from $\mathbb{Z}$ to $\mathbb{Z}_n$ coefficient-wise to f). Show that the rings $\mathbb{Z}[\mathtt{X}]/(n, f)$ and $\mathbb{Z}_n[\mathtt{X}]/(\bar{f})$ are isomorphic.

Exercise 9.41. Let R be a ring, and let $\alpha_1, \ldots, \alpha_n$ be elements of R. Show that the rings R and $R[\mathtt{X}_1, \ldots, \mathtt{X}_n]/(\mathtt{X}_1 - \alpha_1, \ldots, \mathtt{X}_n - \alpha_n)$ are isomorphic.

Exercise 9.42. Let $\rho : R \to R'$ be a ring homomorphism, and suppose that we extend ρ, as in Example 9.48, to a ring homomorphism from $R[\mathtt{X}]$ to $R'[\mathtt{X}]$. Show that for any $a \in R[\mathtt{X}]$, we have $\mathbf{D}(\rho(a)) = \rho(\mathbf{D}(a))$, where $\mathbf{D}(\cdot)$ denotes the formal derivative.

Exercise 9.43. 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 $I \cdot J$. Conclude that $R/(I \cdot J)$ is isomorphic to $R/I \times R/J$.

EXERCISE 9.44. 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 9.45. Let F be a field and let d be an element of F that is not a perfect square (i.e., there does not exist $e \in F$ such that $e^2 = d$). Let $E := F[\mathtt{X}]/(\mathtt{X}^2 - d)$, and let $\eta := [\mathtt{X}]_{\mathtt{X}^2-d}$, so that $E = F[\eta] = \{a + b\eta : a, b \in F\}$.

(a) Show that the quotient ring E is a field, and write down the formula for the inverse of $a + b\eta \in E$.

(b) Show that the map that sends $a + b\eta \in E$ to $a - b\eta$ is a ring automorphism on E.

EXERCISE 9.46. Let $\mathbb{Q}^{(m)}$ be the subring of $\mathbb{Q}$ defined in Example 9.23. 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 9.47. Let $\rho : R \to R'$ be a map from a ring R to a ring R' that satisfies all the requirements of a ring homomorphism, except that we do not require that $\rho(1_R) = 1_{R'}$.

(a) Give a concrete example of such a map ρ, such that $\rho(1_R) \neq 1_{R'}$ and $\rho(1_R) \neq 0_{R'}$.

(b) Show that $\mathrm{img}(\rho)$ is a ring in which $\rho(1_R)$ plays the role of the multiplicative identity.

(c) Show that if R' is an integral domain, and $\rho(1_R) \neq 0_{R'}$, then $\rho(1_R) = 1_{R'}$, and hence ρ satisfies our definition of a ring homomorphism.

(d) Show that if ρ is surjective, then $\rho(1_R) = 1_{R'}$, and hence ρ satisfies our definition of a ring homomorphism.