17
More rings

This chapter develops a number of other concepts concerning rings. These concepts will play important roles later in the text, and we prefer to discuss them now, so as to avoid too many interruptions of the flow of subsequent discussions.

17.1 Algebras

Let R be a ring. An R-**algebra** (or **algebra over** R) is a ring E, together with a ring homomorphism $\tau : R \to E$. Usually, the map τ will be clear from context, as in the following examples.

***Example* 17.1.** If E is a ring that contains R as a subring, then E is an R-algebra, where the associated map $\tau : R \to E$ is just the inclusion map. □

***Example* 17.2.** Let $E_1, \ldots, E_n$ be R-algebras, with associated maps $\tau_i : R \to E_i$, for $i = 1, \ldots, n$. Then the direct product ring $E := E_1 \times \cdots \times E_n$ is naturally viewed as an R-algebra, via the map τ that sends $a \in R$ to $(\tau_1(a), \ldots, \tau_n(a)) \in E$. □

***Example* 17.3.** Let E be an R-algebra, with associated map $\tau : R \to E$, and let I be an ideal of E. Consider the quotient ring E/I. If ρ is the natural map from E onto E/I, then the homomorphism $\rho \circ \tau$ makes E/I into an R-algebra, called the **quotient algebra of** E **modulo** I. □

***Example* 17.4.** As a special case of the previous example, consider the ring $R[\mathtt{X}]$, viewed as an R-algebra via inclusion, and the ideal of R generated by f, where f is a monic polynomial. Then $R[\mathtt{X}]/(f)$ is naturally viewed as an R-algebra, via the map τ that sends $c \in R$ to $[c]_f \in R[\mathtt{X}]/(f)$. If $\deg(f) > 0$, then τ is an embedding of R in $R[\mathtt{X}]/(f)$; if $\deg(f) = 0$, then $R[\mathtt{X}]/(f)$ is the trivial ring, and τ maps everything to zero. □

In some sense, an R-algebra is a generalization of the notion of an extension ring. When the map $\tau : R \to E$ is a canonical embedding, the language of R-algebras can be used if one wants to avoid the sloppiness involved in "identifying" elements of R with their image under τ in E, as we have done on occasion.

In this text, we will be particularly interested in the situation where E is an algebra over a field F. In this case, E either contains a copy of F, or is itself the trivial ring. To see this, let $\tau : F \to E$ be the associated map. Then since the kernel of τ is an ideal of F, it must either be $\{0_F\}$ or F. In the former case, τ is injective, and so E contains an isomorphic copy of F. In the latter case, our requirement that $\tau(1_F) = 1_E$ implies that $1_E = 0_E$, and so E is trivial.

Subalgebras

Let E be an R-algebra with associated map $\tau : R \to E$. A subset S of E is a **subalgebra** if S is a subring containing $\operatorname{img}(\tau)$. As an important special case, if τ is just the inclusion map, then a subring S of E is a subalgebra if and only if S contains R.

R-algebra homomorphisms

There is, of course, a natural notion of a homomorphism for R-algebras. Indeed, it is this notion that is our main motivation for introducing R-algebras in this text. If E and E' are R-algebras, with associated maps $\tau : R \to E$ and $\tau' : R \to E'$, then a map $\rho : E \to E'$ is called an **R-algebra homomorphism** if ρ is a ring homomorphism, and if for all $a \in R$, we have

$$\rho(\tau(a)) = \tau'(a).$$

As usual, if ρ is bijective, then it is called an **R-algebra isomorphism**, and if $R = R'$, it is called an **R-algebra automorphism**.

As an important special case, if τ and τ' are just inclusion maps, then a ring homomorphism $\rho : E \to E'$ is an R-algebra homomorphism if and only if the restriction of ρ to R is the identity map.

The reader should also verify the following facts. First, an R-algebra homomorphism maps subalgebras to subalgebras. Second, Theorems 9.22, 9.23, 9.24, 9.25, 9.26, and 9.27 carry over *mutatis mutandis* from rings to R-algebras.

***Example* 17.5.** Since $\mathbb{C}$ contains $\mathbb{R}$ as a subring, we may naturally view $\mathbb{C}$ as an $\mathbb{R}$-algebra. The complex conjugation map on $\mathbb{C}$ that sends $a + bi$ to $a - bi$, for $a, b \in \mathbb{R}$, is an $\mathbb{R}$-algebra automorphism on $\mathbb{C}$ (see Example 9.5). □

***Example* 17.6.** Let p be a prime, and let F be the field $\mathbb{Z}_p$. If E is an F-algebra, with associated map $\tau : F \to E$, then the map $\rho : E \to E$ that sends $\alpha \in E$ to α^p is an F-algebra homomorphism. To see this, note that E is either trivial, or contains a copy of $\mathbb{Z}_p$. In the former case, there is nothing really to prove. In the latter case, E has characteristic p, and so the fact that ρ is a ring homomorphism follows from Example 9.42 (the "freshman's dream"); moreover, by Fermat's little theorem, for all $a \in F$, we have $\tau(a)^p = \tau(a^p) = \tau(a)$. □

Polynomial evaluation

Let E be an R-algebra with associated map $\tau : R \to E$. Any polynomial $g \in R[\mathtt{X}]$ naturally defines a function on E: if $g = \sum_i g_i \mathtt{X}_i$, with each $g_i \in R$, and $\alpha \in E$, then

$$g(\alpha) := \sum_i \tau(g_i)\alpha^i.$$

For fixed $\alpha \in E$, the **polynomial evaluation map** $\rho : R[\mathtt{X}] \to E$ sends $g \in R[\mathtt{X}]$ to $g(\alpha) \in E$. It is easily verified that ρ is an R-algebra homomorphism (where we naturally view $R[\mathtt{X}]$ as an R-algebra via inclusion). The image of ρ is denoted $R[\alpha]$, and is a subalgebra of E. Indeed, $R[\alpha]$ is the smallest subalgebra of E containing α.

Note that if E contains R as a subring, then the notation $R[\alpha]$ has the same meaning as that introduced in Example 9.39.

We next state a very simple, but extremely useful, fact:

Theorem 17.1. *Let $\rho : E \to E'$ be an R-algebra homomorphism. Then for any $g \in R[\mathtt{X}]$ and $\alpha \in E$, we have*

$$\rho(g(\alpha)) = g(\rho(\alpha)).$$

Proof. Let $\tau : R \to E$ and $\tau' : R \to E'$ be the associated maps. Let

$g = \sum_i g_i \mathtt{X}^i \in R[\mathtt{X}]$. Then we have

$$\begin{aligned}\rho(g(\alpha)) &= \rho(\sum_i \tau(g_i)\alpha^i) = \sum_i \rho(\tau(g_i)\alpha^i) \\ &= \sum_i \rho(\tau(g_i))\rho(\alpha^i) = \sum_i \tau'(g_i)\rho(\alpha)^i \\ &= g(\rho(\alpha)). \quad \square\end{aligned}$$

As a special case of Theorem 17.1, if $E = R[\eta]$ for some $\eta \in E$, then every element of E can be expressed as $g(\eta)$ for some $g \in R[\mathtt{X}]$, and $\rho(g(\eta)) = g(\rho(\eta))$; hence, the action of ρ is completely determined by its action on η.

***Example* 17.7.** Let $E := R[\mathtt{X}]/(f)$ for some monic polynomial $f \in R[\mathtt{X}]$, so that $E = R[\eta]$, where $\eta := [\mathtt{X}]_f$, and let E' be any R-algebra.

Suppose that $\rho : E \to E'$ is an R-algebra homomorphism, and that $\eta' := \rho(\eta)$. The map ρ sends $g(\eta)$ to $g(\eta')$, for $g \in R[\mathtt{X}]$. Also, since $f(\eta) = 0_E$, we have $0_{E'} = \rho(f(\eta)) = f(\eta')$. Thus, η' must be a root of f.

Conversely, suppose that $\eta' \in E'$ is a root of f. Then the polynomial evaluation map from $R[\mathtt{X}]$ to E' that sends $g \in R[\mathtt{X}]$ to $g(\eta') \in E'$ is an R-algebra homomorphism whose kernel contains f, and this gives rise to the R-algebra homomorphism $\rho : E \to E'$ that sends $g(\eta)$ to $g(\eta')$, for $g \in R[\mathtt{X}]$. One sees that complex conjugation is just a special case of this construction (see Example 9.44). □

R-algebras as R-modules

If E is an R-algebra, with associated map $\tau : R \to E$, we may naturally view E as an R-module, where we define a scalar multiplication operation as follows: for $a \in R$ and $\alpha \in E$, define

$$a \cdot \alpha := \tau(a)\alpha.$$

The reader may easily verify that with scalar multiplication so defined, E is an R-module.

Of course, if E is an algebra over a field F, then it is also a vector space over F.

EXERCISE 17.1. Show that any ring E may be viewed as a $\mathbb{Z}$-algebra.

EXERCISE 17.2. Show that the only $\mathbb{R}$-algebra homomorphisms from $\mathbb{C}$ into itself are the identity map and the complex conjugation map.

Exercise 17.3. Let E be an R-algebra, viewed as an R-module as discussed above.

(a) Show that for all $a \in R$ and $\alpha, \beta \in E$, we have $a \cdot (\alpha\beta) = (a \cdot \alpha)\beta$.

(b) Show that a subring S of E is a subalgebra if and only if it is also submodule.

(c) Show that if E' is another R-algebra, then a ring homomorphism $\rho : E \to E'$ is an R-algebra homomorphism if and only if it is an R-linear map.

Exercise 17.4. This exercise develops an alternative characterization of R-algebras. Let R be a ring, and let E be a ring, together with a scalar multiplication operation, that makes E into an R-module. Further suppose that for all $a \in R$ and $\alpha, \beta \in E$, we have $a(\alpha\beta) = (a\alpha)\beta$. Define the map $\tau : R \to E$ that sends $a \in R$ to $a \cdot 1_E \in E$. Show that τ is a ring homomorphism, so that E is an R-algebra, and also show that $\tau(a)\alpha = a\alpha$ for all $a \in R$ and $\alpha \in E$.

17.2 The field of fractions of an integral domain

Let D be any integral domain. Just as we can construct the field of rational numbers by forming fractions involving integers, we can construct a field consisting of fractions whose numerators and denominators are elements of D. This construction is quite straightforward, though a bit tedious.

To begin with, let S be the set of all pairs of the form (a, b), with $a, b \in D$ and $b \neq 0_D$. Intuitively, such a pair (a, b) is a "formal fraction," with numerator a and denominator b. We define a binary relation $\sim$ on S as follows: for $(a_1, b_1), (a_2, b_2) \in S$, we say $(a_1, b_1) \sim (a_2, b_2)$ if and only if $a_1b_2 = a_2b_1$. Our first task is to show that this is an equivalence relation:

Lemma 17.2. *For all $(a_1, b_1), (a_2, b_2), (a_3, b_3) \in S$, we have*

(i) $(a_1, b_1) \sim (a_1, b_1)$;

(ii) $(a_1, b_1) \sim (a_2, b_2)$ *implies* $(a_2, b_2) \sim (a_1, b_1)$;

(iii) $(a_1, b_1) \sim (a_2, b_2)$ *and* $(a_2, b_2) \sim (a_3, b_3)$ *implies* $(a_1, b_1) \sim (a_3, b_3)$.

Proof. (i) and (ii) are rather trivial, and we do not comment on these any further. As for (iii), assume that $a_1b_2 = a_2b_1$ and $a_2b_3 = a_3b_2$. Multiplying the first equation by b_3 we obtain $a_1b_3b_2 = a_2b_3b_1$ and substituting a_3b_2 for a_2b_3 on the right-hand side of this last equation, we obtain $a_1b_3b_2 = a_3b_2b_1$. Now, using the fact that b_2 is non-zero and that D is an integral domain, we may cancel b_2 from both sides, obtaining $a_1b_3 = a_3b_1$. □

Since $\sim$ is an equivalence relation, it partitions S into equivalence classes, and for $(a, b) \in S$, we denote by $[a, b]$ the equivalence class containing (a, b), and we denote by K the collection of all such equivalence classes. Our next task is to define addition and multiplication operations on equivalence classes, mimicking the usual rules of arithmetic with fractions. We want to define the sum of $[a_1, b_1]$ and $[a_2, b_2]$ to be $[a_1b_2 + a_2b_1, b_1b_2]$, and the product of $[a_1, b_1]$ and $[a_2, b_2]$ to be $[a_1a_2, b_1b_2]$. Note that since D is an integral domain, if b_1 and b_2 are non-zero, then so is the product b_1b_2, and therefore $[a_1b_2 + a_2b_1, b_1b_2]$ and $[a_1a_2, b_1b_2]$ are indeed equivalence classes. However, to ensure that this definition is unambiguous, and does not depend on the particular choice of representatives of the equivalence classes $[a_1, b_1]$ and $[a_2, b_2]$, we need the following lemma.

Lemma 17.3. *For $(a_1, b_1), (a_1', b_1'), (a_2, b_2), (a_2', b_2') \in S$ with $(a_1, b_1) \sim (a_1', b_1')$ and $(a_2, b_2) \sim (a_2', b_2')$, we have*

$$(a_1b_2 + a_2b_1, b_1b_2) \sim (a_1'b_2' + a_2'b_1', b_1'b_2')$$

and

$$(a_1a_2, b_1b_2) \sim (a_1'a_2', b_1'b_2').$$

Proof. This is a straightforward calculation. Assume that $a_1b_1' = a_1'b_1$ and $a_2b_2' = a_2'b_2$. Then we have

$$\begin{aligned}(a_1b_2 + a_2b_1)b_1'b_2' &= a_1b_2b_1'b_2' + a_2b_1b_1'b_2' = a_1'b_2b_1b_2' + a_2'b_1b_1'b_2\\ &= (a_1'b_2' + a_2'b_1')b_1b_2\end{aligned}$$

and

$$a_1a_2b_1'b_2' = a_1'a_2b_1b_2' = a_1'a_2'b_1b_2. \quad \Box$$

In light of this lemma, we may unambiguously define addition and multiplication on K as follows: for $[a_1, b_1], [a_2, b_2] \in K$, we define

$$[a_1, b_1] + [a_2, b_2] := [a_1b_2 + a_2b_1, b_1b_2]$$

and

$$[a_1, b_1] \cdot [a_2, b_2] := [a_1a_2, b_1b_2].$$

The next task is to show that K is a ring—we leave the details of this (which are quite straightforward) to the reader.

Lemma 17.4. *With addition and multiplication as defined above, K is a ring, with additive identity $[0_D, 1_D]$ and multiplicative identity $[1_D, 1_D]$.*

Proof. Exercise. □

Finally, we observe that K is in fact a field: it is clear that $[a, b]$ is a non-zero element of K if and only if $a \neq 0_D$, and hence any non-zero element $[a, b]$ of K has a multiplicative inverse, namely, $[b, a]$.

The field K is called the **field of fractions of** D. Consider the map $\tau : D \to K$ that sends $a \in D$ to $[a, 1_D] \in K$. It is easy to see that this map is a ring homomorphism, and one can also easily verify that it is injective. So, starting from D, we can synthesize "out of thin air" its field of fractions K, which essentially contains D as a subring, via the canonical embedding $\tau : D \to K$.

Now suppose that we are given a field L that contains D as a subring. Consider the set K' consisting of all elements in L of the form ab^{-1}, where $a, b \in D$ and $b \neq 0$—note that here, the arithmetic operations are performed using the rules for arithmetic in L. One may easily verify that K' is a subfield of L that contains D, and it is easy to see that this is the smallest subfield of L that contains D. The subfield K' of L may be referred to as the **field of fractions of** D **within** L. One may easily verify that the map $\rho : K \to L$ that sends $[a, b] \in K$ to $ab^{-1} \in L$ is an unambiguously defined ring homomorphism that maps K injectively onto K'; in particular, K is isomorphic as a ring to K'. It is in this sense that the field of fractions K is the smallest field containing D as a subring.

Somewhat more generally, suppose that L is a field, and that $\tau' : D \to L$ is an embedding. One may easily verify that the map $\rho : K \to L$ that sends $[a, b] \in K$ to $\tau'(a)\tau'(b)^{-1} \in L$ is an unambiguously defined, injective ring homomorphism. Moreover, we may view K and L as D-algebras, via the embeddings $\tau : D \to K$ and $\tau' : D \to L$, and the map ρ is seen to be a D-algebra homomorphism.

From now on, we shall simply write an element $[a, b]$ of K as a fraction, a/b. In this notation, the above rules for addition, multiplication, and testing equality in K now look quite familiar:

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 b_2 + a_2 b_1}{b_1 b_2}, \quad \frac{a_1}{b_1} \cdot \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2}, \quad \text{and} \quad \frac{a_1}{b_1} = \frac{a_2}{b_2} \text{ iff } a_1 b_2 = a_2 b_1.$$

Observe that for $a, b \in D$, with $b \in 0_D$ and $b \mid a$, so that $a = bc$ for some $c \in D$, then the fraction $a/b \in K$ is equal to the fraction $c/1_D \in K$, and identifying the element $c \in D$ with its canonical image $c/1_D \in K$, we may simply write $c = a/b$. Note that this notation is consistent with that introduced in part (iii) of Theorem 9.4. A special case of this arises when $b \in D^*$, in which case $c = ab^{-1}$.

Function fields

An important special case of the above construction for the field of fractions of D is when $D = F[\mathtt{X}]$, where F is a field. In this case, the field of fractions is denoted $F(\mathtt{X})$, and is called the **field of rational functions (over F)**. This terminology is a bit unfortunate, since just as with polynomials, although the elements of $F(\mathtt{X})$ define functions, they are not (in general) in one-to-one correspondence with these functions.

Since $F[\mathtt{X}]$ is a subring of $F(\mathtt{X})$, and since F is a subring of $F[\mathtt{X}]$, we see that F is a subfield of $F(\mathtt{X})$.

More generally, we may apply the above construction to the ring $D = F[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ of multi-variate polynomials over a field F, in which case the field of fractions is denoted $F(\mathtt{X}_1, \ldots, \mathtt{X}_n)$, and is also called the field of rational functions (over F, in the variables $\mathtt{X}_1, \ldots, \mathtt{X}_n$).

Exercise 17.5. Let F be a field of characteristic zero. Show that F contains an isomorphic copy of $\mathbb{Q}$.

Exercise 17.6. Show that the field of fractions of $\mathbb{Z}[i]$ within $\mathbb{C}$ is $\mathbb{Q}[i]$. (See Example 9.22 and Exercise 9.8.)

17.3 Unique factorization of polynomials

Throughout this section, F denotes a field.

Like the ring $\mathbb{Z}$, the ring $F[\mathtt{X}]$ of polynomials is an integral domain, and because of the division with remainder property for polynomials, $F[\mathtt{X}]$ has many other properties in common with $\mathbb{Z}$. Indeed, essentially all the ideas and results from Chapter 1 can be carried over almost verbatim from $\mathbb{Z}$ to $F[\mathtt{X}]$, and in this section, we shall do just that.

Recall that for $a, b \in F[\mathtt{X}]$, we write $b \mid a$ if $a = bc$ for some $c \in F[\mathtt{X}]$, and in this case, note that $\deg(a) = \deg(b) + \deg(c)$.

The units of $F[\mathtt{X}]$ are precisely the units F^* of F, that is, the non-zero constants. We call two polynomials $a, b \in F[\mathtt{X}]$ **associate** if $a = ub$ for $u \in F^*$. It is easy to see that a and b are associate if and only if $a \mid b$ and $b \mid a$—indeed, this follows as a special case of part (ii) of Theorem 9.4. Clearly, any non-zero polynomial a is associate to a unique monic polynomial (i.e., with leading coefficient 1), called the **monic associate** of a; indeed, the monic associate of a is $\mathrm{lc}(a)^{-1} \cdot a$.

We call a polynomial p **irreducible** if it is non-constant and all divisors of p are associate to 1 or p. Conversely, we call a polynomial n **reducible** if it is non-constant and is not irreducible. Equivalently, non-constant n is reducible if and only if there exist polynomials $a, b \in F[\mathtt{X}]$ of degree strictly less that n such that $n = ab$.

Clearly, if a and b are associate polynomials, then a is irreducible if and only if b is irreducible.

The irreducible polynomials play a role similar to that of the prime numbers. Just as it is convenient to work with only positive prime numbers, it is also convenient to restrict attention to monic irreducible polynomials.

Corresponding to Theorem 1.3, every non-zero polynomial can be expressed as a unit times a product of monic irreducibles in an essentially unique way:

Theorem 17.5. *Every non-zero polynomial $n \in F[\mathtt{X}]$ can be expressed as*

$$n = u \cdot p_1^{e_1} \cdots p_r^{e_r},$$

where $u \in F^*$, the p_i are distinct monic irreducible polynomials, and the e_i are positive integers. Moreover, this expression is unique, up to a reordering of the p_i.

To prove this theorem, we may assume that n is monic, since the non-monic case trivially reduces to the monic case.

The proof of the existence part of Theorem 17.5 is just as for Theorem 1.3. If n is 1 or a monic irreducible, we are done. Otherwise, there exist $a, b \in F[\mathtt{X}]$ of degree strictly less than n such that $n = ab$, and again, we may assume that a and b are monic. By induction on degree, both a and b can be expressed as a product of monic irreducible polynomials, and hence, so can n.

The proof of the uniqueness part of Theorem 17.5 is almost identical to that of Theorem 1.3. As a special case of Theorem 9.12, we have the following division with remainder property, analogous to Theorem 1.4:

Theorem 17.6. *For $a, b \in F[\mathtt{X}]$ with $b \neq 0$, there exist unique $q, r \in F[\mathtt{X}]$ such that $a = bq + r$ and $\deg(r) < \deg(b)$.*

Analogous to Theorem 1.5, we have:

Theorem 17.7. *For any ideal $I \subseteq F[\mathtt{X}]$, there exists a unique polynomial d such that $I = dF[\mathtt{X}]$, where d is either zero or monic.*

Proof. We first prove the existence part of the theorem. If $I = \{0\}$, then $d = 0$ does the job, so let us assume that $I \neq \{0\}$. Let d be a monic polynomial of minimal degree in I. We want to show that $I = dF[\mathtt{X}]$.

We first show that $I \subseteq dF[\mathtt{X}]$. To this end, let c be any element in I. It suffices to show that $d \mid c$. Using Theorem 17.6, we may write $c = qd + r$, where $\deg(r) < \deg(d)$. Then by the closure properties of ideals, one sees that $r = c - qd$ is also an element of I, and by the minimality of the degree of d, we must have $r = 0$. Thus, $d \mid c$.

We next show that $dF[\mathtt{X}] \subseteq I$. This follows immediately from the fact that $d \in I$ and the closure properties of ideals.

That proves the existence part of the theorem. As for uniqueness, note that if $dF[\mathtt{X}] = d'F[\mathtt{X}]$, we have $d \mid d'$ and $d' \mid d$, from which it follows that d and d' are associate, and so if d and d' are both either monic or zero, they must be equal. □

For $a, b \in F[\mathtt{X}]$, we call $d \in F[\mathtt{X}]$ a **common divisor** of a and b if $d \mid a$ and $d \mid b$; moreover, we call such a d a **greatest common divisor** of a and b if d is monic or zero, and all other common divisors of a and b divide d. Analogous to Theorem 1.6, we have:

Theorem 17.8. *For any $a, b \in F[\mathtt{X}]$, there exists a unique greatest common divisor d of a and b, and moreover, $aF[\mathtt{X}] + bF[\mathtt{X}] = dF[\mathtt{X}]$.*

Proof. We apply the previous theorem to the ideal $I := aF[\mathtt{X}] + bF[\mathtt{X}]$. Let $d \in F[\mathtt{X}]$ with $I = dF[\mathtt{X}]$, as in that theorem. Note that $a, b, d \in I$ and d is monic or zero.

It is clear that d is a common divisor of a and b. Moreover, there exist $s, t \in F[\mathtt{X}]$ such that $as + bt = d$. If $d' \mid a$ and $d' \mid b$, then clearly $d' \mid (as + bt)$, and hence $d' \mid d$.

Finally, for uniqueness, if d'' is a greatest common divisor of a and b, then $d \mid d''$ and $d'' \mid d$, and hence d'' is associate to d, and the requirement that d'' is monic or zero implies that $d'' = d$. □

For $a, b \in F[\mathtt{X}]$, we denote by $\gcd(a, b)$ the greatest common divisor of a and b. Note that as we have defined it, $\mathrm{lc}(a) \gcd(a, 0) = a$. Also note that when at least one of a or b are non-zero, $\gcd(a, b)$ is the unique monic polynomial of maximal degree that divides both a and b.

An immediate consequence of Theorem 17.8 is that for all $a, b \in F[\mathtt{X}]$, there exist $s, t \in F[\mathtt{X}]$ such that $as + bt = \gcd(a, b)$, and that when at least one of a or b are non-zero, $\gcd(a, b)$ is the unique monic polynomial of minimal degree that can be expressed as $as + bt$ for some $s, t \in F[\mathtt{X}]$.

We say that $a, b \in F[\mathtt{X}]$ are **relatively prime** if $\gcd(a, b) = 1$, which is the same as saying that the only common divisors of a and b are units. It is immediate from Theorem 17.8 that a and b are relatively prime if and only if $aF[\mathtt{X}] + bF[\mathtt{X}] = F[\mathtt{X}]$, which holds if and only if there exist $s, t \in F[\mathtt{X}]$ such that $as + bt = 1$.

Analogous to Theorem 1.7, we have:

Theorem 17.9. *For $a, b, c \in F[\mathtt{X}]$ such that $c \mid ab$ and $\gcd(a, c) = 1$, we have $c \mid b$.*

Proof. Suppose that $c \mid ab$ and $\gcd(a, c) = 1$. Then since $\gcd(a, c) = 1$, by Theorem 17.8 we have $as + ct = 1$ for some $s, t \in F[\mathtt{X}]$. Multiplying this equation by b, we obtain $abs + cbt = b$. Since c divides ab by hypothesis, it follows that $c \mid (abs + cbt)$, and hence $c \mid b$. □

Analogous to Theorem 1.8, we have:

Theorem 17.10. *Let $p \in F[\mathtt{X}]$ be irreducible, and let $a, b \in F[\mathtt{X}]$. Then $p \mid ab$ implies that $p \mid a$ or $p \mid b$.*

Proof. Assume that $p \mid ab$. The only divisors of p are associate to 1 or p. Thus, $\gcd(p, a)$ is either 1 or the monic associate of p. If $p \mid a$, we are done; otherwise, if $p \nmid a$, we must have $\gcd(p, a) = 1$, and by the previous theorem, we conclude that $p \mid b$. □

Now to prove the uniqueness part of Theorem 17.5. Suppose we have

$$p_1 \cdots p_r = p_1' \cdots p_s',$$

where $p_1, \ldots, p_r$ and $p_1', \ldots, p_s'$ are monic irreducible polynomials (duplicates are allowed among the p_i and among the p_j'). If $r = 0$, we must have $s = 0$ and we are done. Otherwise, as p_1 divides the right-hand side, by inductively applying Theorem 17.10, one sees that p_1 is equal to p_j' for some j. We can cancel these terms and proceed inductively (on r).

That completes the proof of Theorem 17.5.

Analogous to Theorem 1.9, we have:

Theorem 17.11. *There are infinitely many monic irreducible polynomials in $F[\mathtt{X}]$.*

If F is infinite, then this theorem is true simply because there are infinitely many monic, linear polynomials; in any case, one can also just prove this theorem by mimicking the proof of Theorem 1.9 (verify).

For a monic irreducible polynomial p, we may define the function ν_p, mapping non-zero polynomials to non-negative integers, as follows: for polynomial $n \neq 0$, if $n = p^e m$, where $p \nmid m$, then $\nu_p(n) := e$. We may then write the factorization of n into irreducibles as

$$n = u \prod_p p^{\nu_p(n)},$$

where the product is over all monic irreducible polynomials p, with all but finitely many of the terms in the product equal to 1.

Just as for integers, we may extend the domain of definition of ν_p to include 0, defining $\nu_p(0) := \infty$. For all polynomials a, b, we have

$$\nu_p(a \cdot b) = \nu_p(a) + \nu_p(b) \text{ for all } p. \tag{17.1}$$

From this, it follows that for all polynomials a, b, we have

$$b \mid a \quad \text{if and only if} \quad \nu_p(b) \le \nu_p(a) \text{ for all } p, \tag{17.2}$$

and

$$\nu_p(\gcd(a, b)) = \min(\nu_p(a), \nu_p(b)) \text{ for all } p. \tag{17.3}$$

For $a, b \in F[\mathtt{X}]$ a **common multiple** of a and b is a polynomial m such that $a \mid m$ and $b \mid m$; moreover, such an m is the **least common multiple** of a and b if m is monic or zero, and m divides all common multiples of a and b. In light of Theorem 17.5, it is clear that the least common multiple exists and is unique, and we denote the least common multiple of a and b by $\mathrm{lcm}(a, b)$. Note that as we have defined it, $\mathrm{lcm}(a, 0) = 0$, and that when both a and b are non-zero, $\mathrm{lcm}(a, b)$ is the unique monic polynomial of minimal degree that is divisible by both a and b. Also, for all $a, b \in F[\mathtt{X}]$, we have

$$\nu_p(\mathrm{lcm}(a, b)) = \max(\nu_p(a), \nu_p(b)) \text{ for all } p, \tag{17.4}$$

and

$$\mathrm{lc}(ab) \cdot \gcd(a, b) \cdot \mathrm{lcm}(a, b) = ab. \tag{17.5}$$

Just as in §1.3, the notions of greatest common divisor and least common multiple generalize naturally from two to any number of polynomials. We also say that polynomials $a_1, \ldots, a_k \in F[\mathtt{X}]$ are **pairwise relatively prime** if $\gcd(a_i, a_j) = 1$ for all i, j with $i \neq j$.

Also just as in §1.3, any rational function $a/b \in F(\mathtt{X})$ can be expressed as a fraction a'/b' in **lowest terms**, that is, $a/b = a'/b'$ and $\gcd(a', b') = 1$, and this representation is unique up to multiplication by units.

Many of the exercises in Chapter 1 carry over naturally to polynomials—the reader is encouraged to look over all of the exercises in that chapter, determining which have natural polynomial analogs, and work some of these out.

Exercise 17.7. Show that for $f \in F[\mathtt{X}]$ of degree 2 or 3, we have f irreducible if and only if f has no roots in F.

17.4 Polynomial congruences

Throughout this section, F denotes a field.

Specializing the congruence notation introduced in §9.3 for arbitrary rings to the ring $F[\mathtt{X}]$, for polynomials $a, b, n \in F[\mathtt{X}]$, we write $a \equiv b \pmod{n}$ when $n \mid (a-b)$. Because of the division with remainder property for polynomials, we have the analog of Theorem 2.1:

Theorem 17.12. *Let $n \in F[\mathtt{X}]$ be a non-zero polynomial. For every $a \in F[\mathtt{X}]$, there exists a unique $b \in F[\mathtt{X}]$ such that $a \equiv b \pmod{n}$ and $\deg(b) < n$, namely, $b := a \bmod n$.*

For a non-zero $n \in F[\mathtt{X}]$, and $a \in F[\mathtt{X}]$, we say that $a' \in F[\mathtt{X}]$ is a **multiplicative inverse of a modulo n** if $aa' \equiv 1 \pmod{n}$.

All of the results we proved in §2.2 for solving linear congruences over the integers carry over almost identically to polynomials. As such, we do not give proofs of any of the results here. The reader may simply check that the proofs of the corresponding results translate almost directly.

Theorem 17.13. *Let $a, n \in F[\mathtt{X}]$ with $n \neq 0$. Then a has a multiplicative inverse modulo n if and only if a and n are relatively prime.*

Theorem 17.14. *Let $a, n, z, z' \in F[\mathtt{X}]$ with $n \neq 0$. If a is relatively prime to n, then $az \equiv az' \pmod{n}$ if and only if $z \equiv z' \pmod{n}$. More generally, if $d := \gcd(a, n)$, then $az \equiv az' \pmod{n}$ if and only if $z \equiv z' \pmod{n/d}$.*

Theorem 17.15. *Let $a, b, n \in F[\mathtt{X}]$ with $n \neq 0$. If a is relatively prime to n, then the congruence $az \equiv b \pmod{n}$ has a solution z; moreover, any polynomial z' is a solution if and only if $z \equiv z' \pmod{n}$.*

As for integers, this theorem allows us to generalize the "mod" operation as follows: if $n \in F[\mathtt{X}]$ is a non-zero polynomial, and $s \in F(\mathtt{X})$ is a rational function of the form b/a, where $a, b \in F[\mathtt{X}]$, $a \neq 0$, and $\gcd(a, n) = 1$, then $s \bmod n$ denotes the unique polynomial z satisfying

$$az \equiv b \pmod{n} \quad \text{and} \quad \deg(z) < \deg(n).$$

With this notation, we can simply write $a^{-1} \bmod n$ to denote the unique multiplicative inverse of a modulo n with $\deg(a) < \deg(n)$.

Theorem 17.16. *Let $a, b, n \in F[\mathtt{X}]$ with $n \neq 0$, and let $d := \gcd(a, n)$. If $d \mid b$, then the congruence $az \equiv b \pmod{n}$ has a solution z, and any polynomial z' is also a solution if and only if $z \equiv z' \pmod{n/d}$. If $d \nmid b$, then the congruence $az \equiv b \pmod{n}$ has no solution z.*

Theorem 17.17 (Chinese remainder theorem). *Let $n_1, \ldots, n_k \in F[\mathtt{X}]$ be pairwise relatively prime, non-zero polynomials, and let $a_1, \ldots, a_k \in F[\mathtt{X}]$ be arbitrary polynomials. Then there exists a polynomial $z \in F[\mathtt{X}]$ such that*

$$z \equiv a_i \pmod{n_i} \quad (i = 1, \ldots, k).$$

Moreover, any other polynomial $z' \in F[\mathtt{X}]$ is also a solution of these congruences if and only if $z \equiv z' \pmod{n}$, where $n := \prod_{i=1}^{k} n_i$.

Note that the Chinese remainder theorem (with Theorem 17.12) implies that there exists a unique solution $z \in F[\mathtt{X}]$ to the given congruences with $\deg(z) < \deg(n)$.

The Chinese remainder theorem also has a more algebraic interpretation. Define quotient rings $E_i := F[\mathtt{X}]/(n_i)$ for $i = 1, \ldots, k$, which we may naturally view as F-algebras (see Example 17.4), along with the product F-algebra $E := E_1 \times \cdots \times E_k$ (see Example 17.2). The map ρ from $F[\mathtt{X}]$ to E that sends $z \in F[\mathtt{X}]$ to $([z]_{n_1}, \ldots, [z]_{n_k}) \in E$ is an F-algebra homomorphism. The Chinese remainder theorem says that ρ is surjective, and that the kernel of ρ is the ideal of $F[\mathtt{X}]$ generated by n, giving rise to an F-algebra isomorphism of $F[\mathtt{X}]/(n)$ with E.

Let us recall the formula for the solution z (see proof of Theorem 2.8). We have

$$z := \sum_{i=1}^{k} w_i a_i,$$

where

$$w_i := n_i' m_i, \quad n_i' := n/n_i, \quad m_i := (n_i')^{-1} \bmod n_i \quad (i = 1, \ldots, k).$$

Now, let us consider the special case of the Chinese remainder theorem where $a_i \in F$ and $n_i = (\mathtt{X} - b_i)$ with $b_i \in F$, for $i = 1, \ldots, k$. The condition that the n_i are pairwise relatively prime is equivalent to the condition that the b_i are all distinct. A polynomial z satisfies the system of congruences if and only if $z(b_i) = a_i$ for $i = 1, \ldots, k$. Moreover, we have $n_i' = \prod_{j \neq i}(\mathtt{X} - b_j)$, and $m_i = 1/\prod_{j \neq i}(b_i - b_j) \in F$. So we get

$$z = \sum_{i=1}^{k} a_i \frac{\prod_{j \neq i}(\mathtt{X} - b_j)}{\prod_{j \neq i}(b_i - b_j)}.$$

The reader will recognize this as the usual **Lagrange interpolation formula**. Thus, the Chinese remainder theorem for polynomials includes Lagrange interpolation as a special case.

Let us consider this situation from the point of view of vector spaces. Consider the map $\sigma : F[\mathtt{X}]_{<k} \to F^{\times k}$ that sends $z \in F[\mathtt{X}]$ of degree less than k to $(z(b_1), \ldots, z(b_k)) \in F^{\times k}$, where as above, $b_1, \ldots, b_k$ are distinct elements of F. We see that σ is an F-linear map, and by the Chinese remainder theorem, it is bijective. Thus, σ is an F-vector space isomorphism of $F[\mathtt{X}]_{<k}$ with $F^{\times k}$.

We may encode elements of $F[\mathtt{X}]_{<k}$ as row vectors in a natural way, encoding the polynomial $z = \sum_{i=0}^{k-1} z_i \mathtt{X}^i$ as the row vector $(z_0, \ldots, z_{k-1}) \in F^{1\times k}$. With this encoding, we have

$$\sigma(z) = (z_0, \ldots, z_{k-1})V,$$

where V is the $k \times k$ matrix

$$V := \begin{pmatrix} 1 & 1 & & 1 \\ b_1 & b_2 & & b_k \\ \vdots & \vdots & \cdots & \vdots \\ b_1^{k-1} & b_2^{k-1} & \cdots & b_k^{k-1} \end{pmatrix}.$$

The matrix V (well, actually its transpose) is known as a **Vandermonde matrix**. Because σ is an isomorphism, it follows that the matrix V is invertible.

More generally, consider any fixed elements $b_1, \ldots, b_\ell$ of F, where $\ell \leq k$, and consider the F-linear map $\sigma : F[\mathtt{X}]_{<k} \to F^{\times \ell}$ that sends $z \in F[\mathtt{X}]_{<k}$ to $(z(b_1), \ldots, z(b_\ell))$. If $z = \sum_{i=0}^{k-1} z_i \mathtt{X}^i$, then

$$\sigma(z) = (z_0, \ldots, z_{k-1})W,$$

where W is the $k \times \ell$ matrix

$$W := \begin{pmatrix} 1 & 1 & & 1 \\ b_1 & b_2 & & b_\ell \\ \vdots & \vdots & \cdots & \vdots \\ b_1^{k-1} & b_2^{k-1} & \cdots & b_\ell^{k-1} \end{pmatrix}.$$

Now, if $b_i = b_j$ for some $i \neq j$, then the columns of W are linearly dependent, and hence the column rank of W is less than ℓ. Since the column rank of W is equal to its row rank, the dimension of the row space of W is less than ℓ, and hence, σ is not surjective. Conversely, if the b_i are all distinct, then since the submatrix of W consisting of its first ℓ rows is an invertible Vandermonde matrix, we see that the rank of W is equal to ℓ, and hence σ is surjective.

17.5 Polynomial quotient algebras

Throughout this section, F denotes a field.

Let $f \in F[\mathtt{X}]$ be a monic polynomial, and consider the quotient ring $E := F[\mathtt{X}]/(f)$. As discussed in Example 17.4, we may naturally view E as an F-algebra via the map τ that sends $c \in R$ to $[c]_f \in E$. Moreover, if $\deg(f) > 0$, then τ is an embedding of F in $F[\mathtt{X}]/(f)$, and otherwise, if $f = 1$, then E is the trivial ring, and τ maps everything to zero.

Suppose that $\ell := \deg(f) > 0$. Let $\eta := [\mathtt{X}]_f \in E$. Then $E = F[\eta]$, and as an F-vector space, E has dimension ℓ, with $1, \eta, \ldots, \eta^{\ell-1}$ being a basis (see Examples 9.34, 9.43, 14.3, and 14.22). That is, every element of E can be expressed uniquely as $g(\eta)$ for $g \in F[\mathtt{X}]$ of degree less than ℓ.

Now, if f is irreducible, then every polynomial $a \not\equiv 0 \pmod{f}$ is relatively prime to f, and hence invertible modulo f; therefore, it follows that E is a field. Conversely, if f is not irreducible, then E cannot be a field—indeed, if g is a non-trivial factor of f, then $g(\eta)$ is a zero divisor.

If $F = \mathbb{Z}_p$ for a prime number p, and f is irreducible, then we see that E is a finite field of cardinality p^ℓ. In the next chapter, we shall see how one can perform arithmetic in such fields efficiently, and later, we shall also see how to efficiently construct irreducible polynomials of any given degree over a finite field.

Minimal polynomials. Now suppose that E is any F-algebra, and let α be an element of E. Consider the polynomial evaluation map $\rho : F[\mathtt{X}] \to E$ that sends $g \in F[\mathtt{X}]$ to $g(\alpha)$. The kernel of ρ is an ideal of $F[\mathtt{X}]$, and since every ideal of $F[\mathtt{X}]$ is principal, it follows that there exists a polynomial $\phi \in F[\mathtt{X}]$ such that $\ker(\rho)$ is the ideal of $F[\mathtt{X}]$ generated by ϕ; moreover, we can make the choice of ϕ unique by insisting that it is monic or zero. The polynomial ϕ is called the **minimal polynomial of α (over F)**. If $\phi = 0$, then ρ is injective, and hence the image $F[\alpha]$ of ρ is isomorphic (as an F-algebra) to $F[\mathtt{X}]$. Otherwise, $F[\alpha]$ is isomorphic (as an F-algebra) to $F[\mathtt{X}]/(\phi)$; moreover, since any polynomial that is zero at α is a polynomial multiple of ϕ, we see that ϕ is the unique monic polynomial of smallest degree that is zero at α.

If E has finite dimension, say n, as an F-vector space, then any element α of E has a non-zero minimal polynomial. Indeed, the elements $1_E, \alpha, \ldots, \alpha^n$ must be linearly dependent (as must be any $n+1$ vectors in a vector space of dimension n), and hence there exist $c_0, \ldots, c_n \in F$, not all zero, such that

$$c_0 1_E + c_1\alpha + \cdots + c_n\alpha^n = 0_E,$$

and therefore, the non-zero polynomial $g := \sum_i c_i \mathtt{X}^i$ is zero at α.

***Example* 17.8.** The polynomial $\mathtt{X}^2 + 1$ is irreducible over $\mathbb{R}$, since if it were not, it would have a root in $\mathbb{R}$ (see Exercise 17.7), which is clearly impossible, since -1 is not the square of any real number. It follows immediately that $\mathbb{C} = \mathbb{R}[\mathtt{X}]/(\mathtt{X}^2 + 1)$ is a field, without having to explicitly calculate a formula for the inverse of a non-zero complex number. □

***Example* 17.9.** Consider the polynomial $f := \mathtt{X}^4 + \mathtt{X}^3 + 1$ over $\mathbb{Z}_2$. We claim that f is irreducible. It suffices to show that f has no irreducible factors of degree 1 or 2.

If f had a factor of degree 1, then it would have a root; however, $f(0) = 0 + 0 + 1 = 1$ and $f(1) = 1 + 1 + 1 = 1$. So f has no factors of degree 1.

Does f have a factor of degree 2? The polynomials of degree 2 are $\mathtt{X}^2$, $\mathtt{X}^2 + \mathtt{X}$, $\mathtt{X}^2 + 1$, and $\mathtt{X}^2 + \mathtt{X} + 1$. The first and second of these polynomials are divisible by $\mathtt{X}$, and hence not irreducible, while the third has a 1 as a root, and hence is also not irreducible. The last polynomial, $\mathtt{X}^2 + \mathtt{X} + 1$, has no roots, and hence is the only irreducible polynomial of degree 2 over $\mathbb{Z}_2$. So now we may conclude that if f were not irreducible, it would have to be equal to

$$(\mathtt{X}^2 + \mathtt{X} + 1)^2 = \mathtt{X}^4 + 2\mathtt{X}^3 + 3\mathtt{X}^2 + 2\mathtt{X} + 1 = \mathtt{X}^4 + \mathtt{X}^2 + 1,$$

which it is not.

Thus, $E := \mathbb{Z}_2[\mathtt{X}]/(f)$ is a field with $2^4 = 16$ elements. We may think of elements E as bit strings of length 4, where the rule for addition is bit-wise "exclusive-or." The rule for multiplication is more complicated: to multiply two given bit strings, we interpret the bits as coefficients of polynomials (with the left-most bit the coefficient of $\mathtt{X}^3$), multiply the polynomials, reduce the product modulo f, and write down the bit string corresponding to the reduced product polynomial. For example, to multiply `1001` and `0011`, we compute

$$(\mathtt{X}^3 + 1)(\mathtt{X} + 1) = \mathtt{X}^4 + \mathtt{X}^3 + \mathtt{X} + 1,$$

and

$$(\mathtt{X}^4 + \mathtt{X}^3 + \mathtt{X} + 1) \bmod (\mathtt{X}^4 + \mathtt{X}^3 + 1) = \mathtt{X}.$$

Hence, the product of `1001` and `0011` is `0010`.

Theorem 9.16 says that E^* is a cyclic group. Indeed, the element $\eta :=$ `0010` (i.e., $\eta = [\mathtt{X}]_f$) is a generator for E^*, as the following table of powers shows:

i	η^i	i	η^i
1	0010	8	1110
2	0100	9	0101
3	1000	10	1010
4	1001	11	1101
5	1011	12	0011
6	1111	13	0110
7	0111	14	1100
		15	0001

Such a table of powers is sometimes useful for computations in small finite fields such as this one. Given $\alpha, \beta \in E^*$, we can compute $\alpha\beta$ by obtaining (by table lookup) i, j such that $\alpha = \eta^i$ and $\beta = \eta^j$, computing $k := (i + j) \bmod 15$, and then obtaining $\alpha\beta = \eta^k$ (again by table lookup). □

Exercise 17.8. In the field E is Example 17.9, what is the minimal polynomial of $\mathtt{1011}$ over $\mathbb{Z}_2$?

Exercise 17.9. Show that if the factorization of f over $F[\mathtt{X}]$ into irreducibles is as $f = f_1^{e_1} \cdots f_r^{e_r}$, and if $\alpha = [h]_f \in F[\mathtt{X}]/(f)$, then the minimal polynomial ϕ of α over F is $\mathrm{lcm}(\phi_1, \ldots, \phi_r)$, where each ϕ_i is the minimal polynomial of $[h]_{f_i^{e_i}} \in F[\mathtt{X}]/(f_i^{e_i})$ over F.

17.6 General properties of extension fields

We now discuss a few general notions related to extension fields. These are all quite simple applications of the theory developed so far. Recall that if F and E are fields, with F being a subring of E, then E is called an extension field of F. As usual, we shall blur the distinction between a subring and a canonical embedding; that is, if $\tau : F \to E$ is an canonical embedding, we shall simply identify elements of F with their images in E under τ, and in so doing, we may view E as an extension field of F. Usually, the map τ will be clear from context; for example, if $E = F[\mathtt{X}]/(\phi)$ for some irreducible polynomial $\phi \in F[\mathtt{X}]$, then we shall simply say that E is an extension field of F, although strictly speaking, F is embedded in E via the map that sends $a \in F$ to $[a]_\phi \in E$.

Let E be an extension field of a field F. Then E is an F-algebra, and in particular, an F-vector space. If E is a finite dimensional F-vector space, then we say that E is a **finite extension of** F, and $\dim_F(E)$ is called the

degree of the extension, and is denoted $(E : F)$; otherwise, we say that E is an **infinite extension of** F.

An element $\alpha \in E$ is called **algebraic over** F if there exists a non-zero polynomial $f \in F[\mathtt{X}]$ such that $f(\alpha) = 0$; otherwise, α is called **transcendental over** F. If all elements of E are algebraic over F, then we call E an **algebraic extension of** F. From the discussion on minimal polynomials in §17.5, we may immediately state:

Theorem 17.18. *If E is a finite extension of F, then E is also an algebraic extension of F.*

Suppose $\alpha \in E$ is algebraic over F. Let ϕ be its minimal polynomial, so that $F[\mathtt{X}]/(\phi)$ is isomorphic (as an F-algebra) to $F[\alpha]$. Since $F[\alpha]$ is a subring of a field, it must be an integral domain, which implies that ϕ is irreducible, which in turn implies that $F[\alpha]$ is a subfield of E. Moreover, the degree $(F[\alpha] : F)$ is equal to the degree of ϕ, and this number is called the **degree of** α **(over** F**)**. It is clear that if E is finite dimensional, then the degree of α is at most $(E : F)$.

Suppose that $\alpha \in E$ is transcendental over F. Consider the "rational function evaluation map" that sends $f/g \in F(\mathtt{X})$ to $f(\alpha)/g(\alpha) \in E$. Since no non-zero polynomial over F vanishes at α, it is easy to see that this map is well defined, and is in fact an injective F-algebra homomorphism from $F(\mathtt{X})$ into E. The image is denoted $F(\alpha)$, and this is clearly a subfield of E containing F and α, and it is plain to see that it is the smallest such subfield. It is also clear that $F(\alpha)$ has infinite dimension over F, since it contains an isomorphic copy of the infinite dimensional vector space $F[\mathtt{X}]$.

More generally, for any $\alpha \in E$, algebraic or transcendental, we can define $F(\alpha)$ to be the set consisting of all elements of the form $f(\alpha)/g(\alpha) \in E$, where $f, g \in F[\mathtt{X}]$ and $g(\alpha) \neq 0$. It is clear that $F(\alpha)$ is a field, and indeed, it is the smallest subfield of E containing F and α. If α is algebraic, then $F(\alpha) = F[\alpha]$, and is isomorphic (as an F-algebra) to $F[\mathtt{X}]/(\phi)$, where ϕ is the minimal polynomial of α over F; otherwise, if α is transcendental, then $F(\alpha)$ is isomorphic (as an F-algebra) to the rational function field $F(\mathtt{X})$.

***Example* 17.10.** If $f \in F[\mathtt{X}]$ is monic and irreducible, $E = F[\mathtt{X}]/(f)$, and $\eta := [\mathtt{X}]_f \in E$, then η is algebraic over F, its minimal polynomial over F is f, and its degree over F is equal to $\deg(f)$. Also, we have $E = F[\eta]$, and any element $\alpha \in E$ is algebraic of degree at most $\deg(f)$. □

Exercise 17.10. In the field E is Example 17.9, find all the elements of degree 2 over $\mathbb{Z}_2$.

EXERCISE 17.11. Show that if E is a finite extension of F, with a basis $\alpha_1, \ldots, \alpha_n$ over F, and K is a finite extension of E, with a basis $\beta_1, \ldots, \beta_m$ over E, then

$$\alpha_i \beta_j \quad (i = 1, \ldots, n;\ j = 1, \ldots, m)$$

is a basis for K over F, and hence K is a finite extension of F and

$$(K : F) = (K : E)(E : F).$$

EXERCISE 17.12. Show that if E is an algebraic extension of F, and K is an algebraic extension of E, then K is an algebraic extension of F.

EXERCISE 17.13. Let E be an extension of F. Show that the set of all elements in E that are algebraic over F is a subfield of E containing F.

We close this section with a discussion of a **splitting field** — a finite extension of the coefficient field in which a given polynomial splits completely into linear factors. As the next theorem shows, splitting fields always exist.

Theorem 17.19. *Let F be a field, and $f \in F[\mathtt{X}]$ a monic polynomial of degree ℓ. Then there exists a finite extension K of F in which f factors as*

$$f = (\mathtt{X} - \alpha_1)(\mathtt{X} - \alpha_2) \cdots (\mathtt{X} - \alpha_\ell),$$

with $\alpha_1, \ldots, \alpha_\ell \in K$.

Proof. We prove the existence of K by induction on the degree ℓ of f. If $\ell = 0$, then the theorem is trivially true. Otherwise, let g be an irreducible factor of f, and set $E := F[\mathtt{X}]/(g)$, so that $\alpha := [\mathtt{X}]_g$ is a root of g, and hence of f, in E. So over the extension field E, f factors as

$$f = (\mathtt{X} - \alpha)h,$$

where $h \in E[\mathtt{X}]$ is a polynomial of degree $\ell - 1$. Applying the induction hypothesis, there exists a finite extension K of E such that h splits into linear factors over K. Thus, over K, f splits into linear factors, and by Exercise 17.11, K is a finite extension of F. □

17.7 Formal power series and Laurent series

We discuss generalizations of polynomials that allow an infinite number of non-zero coefficients. Although we are mainly interested in the case where the coefficients come from a field F, we develop the basic theory for general rings R.

17.7.1 Formal power series

The ring $R[\![\mathtt{X}]\!]$ of **formal power series over** R consists of all formal expressions of the form

$$a = a_0 + a_1\mathtt{X} + a_2\mathtt{X}^2 + \cdots ,$$

where $a_0, a_1, a_2, \ldots \in R$. Unlike ordinary polynomials, we allow an infinite number of non-zero coefficients. We may write such a formal power series as

$$a = \sum_{i=0}^{\infty} a_i\mathtt{X}^i.$$

The rules for addition and multiplication of formal power series are *exactly* the same as for polynomials. Indeed, the formulas (9.1) and (9.2) in §9.2 for addition and multiplication may be applied directly—all of the relevant sums are finite, and so everything is well defined.

We shall not attempt to interpret a formal power series as a function, and therefore, "convergence" issues shall simply not arise.

Clearly, $R[\![\mathtt{X}]\!]$ contains $R[\mathtt{X}]$ as a subring. Let us consider the group of units of $R[\![\mathtt{X}]\!]$.

Theorem 17.20. *Let* $a = \sum_{i=0}^{\infty} a_i\mathtt{X}^i \in R[\![\mathtt{X}]\!]$. *Then* $a \in (R[\![\mathtt{X}]\!])^*$ *if and only if* $a_0 \in R^*$.

Proof. If a_0 is not a unit, then it is clear that a is not a unit, since the constant term of a product formal power series is equal to the product of the constant terms.

Conversely, if a_0 is a unit, we show how to define the coefficients of the inverse $b = \sum_{i=0}^{\infty} b_i\mathtt{X}^i$ of a. Let $ab = c = \sum_{i=0}^{\infty} c_i\mathtt{X}^i$. We want $c = 1$, meaning that $c_0 = 1$ and $c_i = 0$ for all $i > 0$. Now, $c_0 = a_0b_0$, so we set $b_0 := a_0^{-1}$. Next, we have $c_1 = a_0b_1 + a_1b_0$, so we set $b_1 := -a_1b_0 \cdot a_0^{-1}$. Next, we have $c_2 = a_0b_2 + a_1b_1 + a_2b_0$, so we set $b_2 := -(a_1b_1 + a_2b_0) \cdot a_0^{-1}$. Continuing in this way, we see that if we define $b_i := -(a_1b_{i-1} + \cdots + a_ib_0) \cdot a_0^{-1}$ for $i \geq 1$, then $ab = 1$. □

***Example* 17.11.** In the ring $R[\![\mathtt{X}]\!]$, the multiplicative inverse of $1 - \mathtt{X}$ is $\sum_{i=0}^{\infty} \mathtt{X}^i$. □

Exercise 17.14. For a field F, show that any non-zero ideal of $F[\![\mathtt{X}]\!]$ is of the form $(\mathtt{X}^m)$ for some uniquely determined integer $m \geq 0$.

17.7.2 Formal Laurent series

One may generalize formal power series to allow a finite number of negative powers of $\mathtt{X}$. The ring $R((\mathtt{X}))$ of **formal Laurent series over** R consists of all formal expressions of the form

$$a = a_m\mathtt{X}^m + a_{m+1}\mathtt{X}^{m+1} + \cdots ,$$

where m is allowed to be any integer (possibly negative), and $a_m, a_{m+1}, \ldots \in R$. Thus, elements of $R((\mathtt{X}))$ may have an infinite number of terms involving positive powers of $\mathtt{X}$, but only a finite number of terms involving negative powers of $\mathtt{X}$. We may write such a formal Laurent series as

$$a = \sum_{i=m}^{\infty} a_i\mathtt{X}^i.$$

The rules for addition and multiplication of formal Laurent series are just as one would expect: if

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

then

$$a + b := \sum_{i=m}^{\infty} (a_i + b_i)\mathtt{X}^i, \tag{17.6}$$

and

$$a \cdot b := \sum_{i=2m}^{\infty} \Big(\sum_{k=m}^{i-m} a_k b_{i-k} \Big) \mathtt{X}^i. \tag{17.7}$$

We leave it to the reader to verify that $R((\mathtt{X}))$ is a ring containing $R[[\mathtt{X}]]$.

Theorem 17.21. *If D is an integral domain, then $D((\mathtt{X}))$ is an integral domain.*

Proof. Let $a = \sum_{i=m}^{\infty} a_i\mathtt{X}^i$ and $b = \sum_{i=n}^{\infty} b_i\mathtt{X}^i$, where $a_m \neq 0$ and $b_n \neq 0$. Then $ab = \sum_{i=m+n}^{\infty} c_i$, where $c_{m+n} = a_m b_n \neq 0$. □

Theorem 17.22. *Let $a \in R((\mathtt{X}))$, and suppose that $a \neq 0$ and $a = \sum_{i=m}^{\infty} a_i\mathtt{X}^i$ with $a_m \in R^*$. Then a has a multiplicative inverse in $R((\mathtt{X}))$.*

Proof. We can write $a = \mathtt{X}^m b$, where b is a formal power series whose constant term is a unit, and hence there is a formal power series c such that $bc = 1$. Thus, $\mathtt{X}^{-m}c$ is the multiplicative inverse of a in $R((\mathtt{X}))$. □

As an immediate corollary, we have:

Theorem 17.23. *If F is a field, then $F((\mathtt{X}))$ is a field.*

Exercise 17.15. Show that for a field F, $F((\mathtt{X}))$ is the field of fractions of $F[[\mathtt{X}]]$; that is, there is no proper subfield of $F((\mathtt{X}))$ that contains $F[[\mathtt{X}]]$.

17.7.3 Reversed formal Laurent series

While formal Laurent series are useful in some situations, in many others, it is more useful and natural to consider **reversed formal Laurent series over** R. These are formal expressions of the form

$$a = \sum_{i=-\infty}^{m} a_i \mathtt{X}^i,$$

where $a_m, a_{m-1}, \ldots \in R$. Thus, in a reversed formal Laurent series, we allow an infinite number of terms involving negative powers of $\mathtt{X}$, but only a finite number of terms involving positive powers of $\mathtt{X}$.

The rules for addition and multiplication of reversed formal Laurent series are just as one would expect: if

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

then

$$a + b := \sum_{i=-\infty}^{m} (a_i + b_i)\mathtt{X}^i, \tag{17.8}$$

and

$$a \cdot b := \sum_{i=-\infty}^{2m} \Bigg(\sum_{k=i-m}^{m} a_k b_{i-k} \Bigg) \mathtt{X}^i. \tag{17.9}$$

The ring of all reversed formal Laurent series is denoted $R((\mathtt{X}^{-1}))$, and as the notation suggests, the map that sends $\mathtt{X}$ to $\mathtt{X}^{-1}$ (and acts as the identity on R) is an isomorphism of $R((\mathtt{X}))$ with $R((\mathtt{X}^{-1}))$.

Now, for any $a = \sum_{i=-\infty}^{m} a_i \mathtt{X}^i \in R((\mathtt{X}^{-1}))$ with $a_m \neq 0$, let us define the **degree of** a, denoted $\deg(a)$, to be the value m, and the **leading coefficient of** a, denoted $\mathrm{lc}(a)$, to be the value a_m. As for ordinary polynomials, we define the degree of 0 to be $-\infty$, and the leading coefficient of 0 to be 0. Note that if a happens to be a polynomial, then these definitions of degree and leading coefficient agree with that for ordinary polynomials.

Theorem 17.24. *For $a, b \in R((\mathtt{X}^{-1}))$, we have $\deg(ab) \leq \deg(a) + \deg(b)$, where equality holds unless both $\mathrm{lc}(a)$ and $\mathrm{lc}(b)$ are zero divisors. Furthermore, if $b \neq 0$ and $\mathrm{lc}(b)$ is a unit, then b is a unit, and we have $\deg(ab^{-1}) = \deg(a) - \deg(b)$.*

Proof. Exercise. □

It is also natural to define a **floor function** for reversed formal Laurent series: for $a \in R((\mathtt{X}^{-1}))$ with $a = \sum_{i=-\infty}^{m} a_i \mathtt{X}^i$, we define

$$\lfloor a \rfloor := \sum_{i=0}^{m} a_i \mathtt{X}^i \in R[\mathtt{X}];$$

that is, we compute the floor function by simply throwing away all terms involving negative powers of $\mathtt{X}$.

Now, let $a, b \in R[\mathtt{X}]$ with $b \neq 0$ and $\mathrm{lc}(b)$ a unit, and using the usual division with remainder property for polynomials, write $a = bq + r$, where $q, r \in R[\mathtt{X}]$ with $\deg(r) < \deg(b)$. Let b^{-1} denote the multiplicative inverse of b in $R((\mathtt{X}^{-1}))$. It is not too hard to see that $\lfloor ab^{-1} \rfloor = q$; indeed, multiplying the equation $a = bq+r$ by b^{-1}, we obtain $ab^{-1} = q+rb^{-1}$, and $\deg(rb^{-1}) < 0$, from which it follows that $\lfloor ab^{-1} \rfloor = q$.

Let F be a field. Since $F((\mathtt{X}^{-1}))$ is isomorphic to $F((\mathtt{X}))$, and the latter is a field, it follows that $F((\mathtt{X}^{-1}))$ is a field. Now, $F((\mathtt{X}^{-1}))$ contains $F[\mathtt{X}]$ as a subring, and hence contains (an isomorphic copy) of $F(\mathtt{X})$. Just as $F(\mathtt{X})$ corresponds to the field of rational numbers, $F((\mathtt{X}^{-1}))$ corresponds to the field real numbers. Indeed, we can think of real numbers as decimal numbers with a finite number of digits to the left of the decimal point and an infinite number to the right, and reversed formal Laurent series have a similar "syntactic" structure. In many ways, this syntactic similarity between the real numbers and reversed formal Laurent series is more than just superficial.

Exercise 17.16. Write down the rule for determining the multiplicative inverse of an element of $R((\mathtt{X}^{-1}))$ whose leading coefficient is a unit in R.

Exercise 17.17. Let F be a field of characteristic other than 2. Show that a non-zero $z \in F((\mathtt{X}^{-1}))$ has a square-root in $z \in F((\mathtt{X}^{-1}))$ if and only if $\deg(z)$ is even and $\mathrm{lc}(z)$ has a square-root in F.

Exercise 17.18. Let R be a ring, and let $\alpha \in R$. Show that the multiplicative inverse of $\mathtt{X} - \alpha$ in $R((\mathtt{X}^{-1}))$ is $\sum_{j=1}^{\infty} \alpha^{j-1}\mathtt{X}^{-j}$.

EXERCISE 17.19. Let R be an arbitrary ring, let $\alpha_1, \ldots, \alpha_\ell \in R$, and let

$$f := (\mathtt{X} - \alpha_1)(\mathtt{X} - \alpha_2) \cdots (\mathtt{X} - \alpha_\ell) \in R[\mathtt{X}].$$

For $j \geq 0$, define the "power sum"

$$s_j := \sum_{i=1}^{\ell} \alpha_i^j.$$

Show that in the ring $R((\mathtt{X}^{-1}))$, we have

$$\frac{\mathbf{D}(f)}{f} = \sum_{i=1}^{\ell} \frac{1}{(\mathtt{X} - \alpha_i)} = \sum_{j=1}^{\infty} s_{j-1}\mathtt{X}^{-j},$$

where $\mathbf{D}(f)$ is the formal derivative of f.

EXERCISE 17.20. Continuing with the previous exercise, derive **Newton's identities**, which state that if $f = \mathtt{X}^\ell + f_1\mathtt{X}^{\ell-1} + \cdots + f_\ell$, with $f_1, \ldots, f_\ell \in R$, then

$$\begin{aligned}
s_1 + f_1 &= 0 \\
s_2 + f_1 s_1 + 2f_2 &= 0 \\
s_3 + f_1 s_2 + f_2 s_1 + 3f_3 &= 0 \\
&\vdots \\
s_\ell + f_1 s_{\ell-1} + \cdots + f_{\ell-1}s_1 + \ell f_\ell &= 0 \\
s_{j+\ell} + f_1 s_{j+\ell-1} + \cdots + f_{\ell-1}s_{j+1} + f_\ell s_j &= 0 \quad (j \geq 1).
\end{aligned}$$

17.8 Unique factorization domains $(*)$

As we have seen, both the integers and the ring $F[\mathtt{X}]$ of polynomials over a field enjoy a unique factorization property. These are special cases of a more general phenomenon, which we explore here.

Throughout this section, D denotes an integral domain.

We call $a, b \in D$ **associate** if $a = ub$ for some $u \in D^*$. Equivalently, a and b are associate if and only if $a \mid b$ and $b \mid a$. A non-zero element $p \in D$ is called **irreducible** if it is not a unit, and all divisors of p are associate to 1 or p. Equivalently, a non-zero, non-unit $p \in D$ is irreducible if and only if it cannot be expressed as $p = ab$ where neither a nor b are units.

Definition 17.25. *We call D a* ***unique factorization domain (UFD)*** *if*

(i) every non-zero element of D that is not a unit can be written as a product of irreducibles in D, and

(ii) such a factorization into irreducibles is unique up to associates and the order in which the factors appear.

Another way to state part (ii) of the above definition is that if $p_1 \cdots p_r$ and $p'_1 \cdots p'_s$ are two factorizations of some element as a product of irreducibles, then $r = s$, and there exists a permutation π on the indices $\{1, \ldots, r\}$ such that p_i and $p'_{\pi(i)}$ are associate.

As we have seen, both $\mathbb{Z}$ and $F[\mathtt{X}]$ are UFDs. In both of those cases, we chose to single out a distinguished irreducible element among all those associate to any given irreducible: for $\mathbb{Z}$, we always chose p to be positive, and for $F[\mathtt{X}]$, we chose p to be monic. For any specific unique factorization domain D, there may be such a natural choice, but in the general case, there will not be (see Exercise 17.21 below).

***Example* 17.12.** Having already seen two examples of UFDs, it is perhaps a good idea to look at an example of an integral domain that is not a UFD. Consider the subring $\mathbb{Z}[\sqrt{-3}]$ of the complex numbers, which consists of all complex numbers of the form $a + b\sqrt{-3}$, where $a, b \in \mathbb{Z}$. As this is a subring of the field $\mathbb{C}$, it is an integral domain (one may also view $\mathbb{Z}[\sqrt{-3}]$ as the quotient ring $\mathbb{Z}[\mathtt{X}]/(\mathtt{X}^2 + 3)$).

Let us first determine the units in $\mathbb{Z}[\sqrt{-3}]$. For $a, b \in \mathbb{Z}$, we have $N(a + b\sqrt{-3}) = a^2 + 3b^2$, where N is the usual norm map on $\mathbb{C}$ (see Example 9.5). If $\alpha \in \mathbb{Z}[\sqrt{-3}]$ is a unit, then there exists $\alpha' \in \mathbb{Z}[\sqrt{-3}]$ such that $\alpha\alpha' = 1$. Taking norms, we obtain

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

Since the norm of an element of $\mathbb{Z}[\sqrt{-3}]$ is a non-negative integer, this implies that $N(\alpha) = 1$. If $\alpha = a+b\sqrt{-3}$, with $a, b \in \mathbb{Z}$, then $N(\alpha) = a^2+3b^2$, and it is clear that $N(\alpha) = 1$ if and only if $\alpha = \pm 1$. We conclude that the only units in $\mathbb{Z}[\sqrt{-3}]$ are ± 1.

Now consider the following two factorizations of 4 in $\mathbb{Z}[\sqrt{-3}]$:

$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}). \tag{17.10}$$

We claim that 2 is irreducible. For suppose, say, that $2 = \alpha\alpha'$, for $\alpha, \alpha' \in \mathbb{Z}[\sqrt{-3}]$, with neither a unit. Taking norms, we have $4 = N(2) = N(\alpha)N(\alpha')$, and therefore, $N(\alpha) = N(\alpha') = 2$—but this is impossible, since there are no integers a and b such that $a^2+3b^2=2$. By the same reasoning, since $N(1+\sqrt{-3})=N(1-\sqrt{-3})=4$, we see that $1+\sqrt{-3}$ and $1-\sqrt{-3}$ are both irreducible. Further, it is clear that 2 is not associate to either $1+\sqrt{-3}$ or $1-\sqrt{-3}$, and so the two factorizations of 4 in (17.10) are fundamentally different. □

For $a, b \in D$, we call $d \in D$ a **common divisor** of a and b if $d \mid a$ and $d \mid b$; moreover, we call such a d a **greatest common divisor** of a and b if all other common divisors of a and b divide d. We say that a and b are **relatively prime** if the only common divisors of a and b are units. It is immediate from the definition of a greatest common divisor that it is unique, up to multiplication by units, if it exists at all. Unlike in the case of $\mathbb{Z}$ and $F[\mathtt{X}]$, in the general setting, greatest common divisors need not exist; moreover, even when they do, we shall not attempt to "normalize" greatest common divisors, and we shall speak only of "a" greatest common divisor, rather than "the" greatest common divisor.

Just as for integers and polynomials, we can generalize the notion of a greatest common divisor in an arbitrary integral domain D from two to any number of elements of D, and we can also define a **least common multiple** of any number of elements as well.

Although these greatest common divisors and least common multiples need not exist in an arbitrary integral domain D, if D is a UFD, they will always exist. The existence question easily reduces to the question of the existence of a greatest common divisor and least common multiple of a and b, where a and b are non-zero elements of D. So assuming that D is a UFD, we may write

$$a = u\prod_{i=1}^{r} p_i^{e_i} \quad \text{and} \quad b = v\prod_{i=1}^{r} p_i^{f_i},$$

where u and v are units, $p_1, \ldots, p_r$ are non-associate irreducibles, and the e_i and f_i are non-negative integers, and it is easily seen that

$$\prod_{i=1}^{r} p^{\min(e_i,f_i)}$$

is a greatest common divisor of a and b, while

$$\prod_{i=1}^{r} p^{\max(e_i,f_i)}$$

is a least common multiple of a and b.

It is also evident that in a UFD D, if $c \mid ab$ and c and a are relatively prime, then $c \mid b$. In particular, if p is irreducible and $p \mid ab$, then $p \mid a$ or $p \mid b$. From this, we see that if p is irreducible, then the quotient ring D/pD is an integral domain, and so the ideal pD is a prime ideal (see discussion above Exercise 9.28).

In a general integral domain D, we say that an element $p \in D$ is **prime** if for all $a, b \in D$, $p \mid ab$ implies $p \mid a$ or $p \mid b$ (which is equivalent to saying that the ideal pD is prime). Thus, if D is a UFD, then all irreducibles are primes; however, in a general integral domain, this may not be the case. Here are a couple of simple but useful facts whose proofs we leave to the reader.

Theorem 17.26. *Any prime element in D is irreducible.*

Proof. Exercise. □

Theorem 17.27. *Suppose D satisfies part (i) of Definition 17.25. Also, suppose that all irreducibles in D are prime. Then D is a UFD.*

Proof. Exercise. □

Exercise 17.21. (a) Show that the "is associate to" relation is an equivalence relation.

(b) Consider an equivalence class C induced by the "is associate to" relation. Show that if C contains an irreducible element, then all elements of C are irreducible.

(c) Suppose that for every equivalence class C that contains irreducibles, we choose one element of C, and call it a **distinguished irreducible**. Show that D is a UFD if and only if every non-zero element of D can be expressed as $u \cdot p_1^{e_1} \cdots p_r^{e_r}$, where u is a unit, $p_1, \ldots, p_r$ are distinguished irreducibles, and this expression is unique up to a reordering of the p_i.

Exercise 17.22. Show that the ring $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Exercise 17.23. Let D be a UFD and F its field of fractions. Show that

(a) every element $x \in F$ can be expressed as $x = a/b$, where $a, b \in D$ are relatively prime, and

(b) that if $x = a/b$ for $a, b \in D$ relatively prime, then for any other $a', b' \in D$ with $x = a'/b'$, we have $a' = ca$ and $b' = cb$ for some $c \in D$.

Exercise 17.24. Let D be a UFD and let $p \in D$ be irreducible. Show that there is no prime ideal Q of D with $\{0_D\} \subsetneq Q \subsetneq pD$.

17.8.1 Unique factorization in Euclidean and principal ideal domains

Our proofs of the unique factorization property in both $\mathbb{Z}$ and $F[\mathtt{X}]$ hinged on the division with remainder property for these rings. This notion can be generalized, as follows.

Definition 17.28. *D is said to be a* ***Euclidean domain*** *if there is a "size function" S mapping the non-zero elements of D to the set of non-negative integers, such that for $a, b \in D$ with $b \neq 0$, there exist $q, r \in D$, with the property that $a = bq + r$ and either $r = 0$ or $S(r) < S(b)$.*

***Example* 17.13.** Both $\mathbb{Z}$ and $F[\mathtt{X}]$ are Euclidean domains. In $\mathbb{Z}$, we can take the ordinary absolute value function $|\cdot|$ as a size function, and for $F[\mathtt{X}]$, the function $\deg(\cdot)$ will do. □

***Example* 17.14.** Recall again the ring

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$$

of Gaussian integers from Example 9.22. Let us show that this is a Euclidean domain, using the usual norm map N on complex numbers (see Example 9.5) for the size function. Let $\alpha, \beta \in \mathbb{Z}[i]$, with $\beta \neq 0$. We want to show the existence of $\xi, \rho \in \mathbb{Z}[i]$ such that $\alpha = \beta\xi + \rho$, where $N(\rho) < N(\beta)$. Suppose that in the field $\mathbb{C}$, we compute $\alpha\beta^{-1} = r + si$, where $r, s \in \mathbb{Q}$. Let m, n be integers such that $|m - r| \leq 1/2$ and $|n - s| \leq 1/2$—such integers m and n always exist, but may not be uniquely determined. Set $\xi := m + ni \in \mathbb{Z}[i]$ and $\rho := \alpha - \beta\xi$. Then we have

$$\alpha\beta^{-1} = \xi + \delta,$$

where $\delta \in \mathbb{C}$ with $N(\delta) \leq 1/4 + 1/4 = 1/2$, and

$$\rho = \alpha - \beta\xi = \alpha - \beta(\alpha\beta^{-1} - \delta) = \delta\beta,$$

and hence

$$N(\rho) = N(\delta\beta) = N(\delta)N(\beta) \leq \frac{1}{2}N(\beta). \quad \square$$

Theorem 17.29. *If D is a Euclidean domain and I is an ideal of D, then there exists $d \in D$ such that $I = dD$.*

Proof. If $I = \{0\}$, then $d = 0$ does the job, so let us assume that $I \neq \{0\}$. Let d be an non-zero element of I such that $S(d)$ is minimal, where S is a size function that makes D into a Euclidean domain. We claim that $I = dD$.

It will suffice to show that for all $c \in I$, we have $d \mid c$. Now, we know that there exists $q, r \in D$ such that $c = qd + r$, where either $r = 0$ or $S(r) < S(d)$. If $r = 0$, we are done; otherwise, r is a non-zero element of I with $S(r) < S(d)$, contradicting the minimality of $S(d)$. □

Recall that an ideal of the form $I = dD$ is called a principal ideal. If all ideals of D are principal, then D is called a **principal ideal domain (PID)**. Theorem 17.29 says that any Euclidean domain is a PID.

PIDs enjoy many nice properties, including:

Theorem 17.30. *If D is a PID, then D is a UFD.*

For the rings $\mathbb{Z}$ and $F[\mathtt{X}]$, the proof of part (i) of Definition 17.25 was a quite straightforward induction argument (as it also would be for any Euclidean domain). For a general PID, however, this requires a different sort of argument. We begin with the following fact:

Theorem 17.31. *If D is a PID, and $I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals of D, then there exists an integer k such that $I_k = I_{k+1} = \cdots$.*

Proof. Let $I := \bigcup_{i=1}^{\infty} I_i$. It is easy to see that I is an ideal. Thus, $I = dD$ for some $d \in D$. But $d \in \bigcup_{i=1}^{\infty} I_i$ implies that $d \in I_k$ for some k, which shows that $I = dD \subseteq I_k$. It follows that $I = I_k = I_{k+1} = \cdots$. □

We can now prove the existence part of Theorem 17.30:

Theorem 17.32. *If D is a PID, then every non-zero, non-unit element of D can be expressed as a product of irreducibles in D.*

Proof. Let $n \in D$, $n \neq 0$, and n not a unit. If n is irreducible, we are done. Otherwise, we can write $n = ab$, where neither a nor b are units. As ideals, we have $nD \subsetneq aD$ and $nD \subsetneq bD$. If we continue this process recursively, building up a "factorization tree" where n is at the root, a and b are the children of n, and so on, then the recursion must stop, since any infinite path in the tree would give rise to a chain of ideals

$$nD = I_1 \subsetneq I_2 \subsetneq \cdots ,$$

contradicting Theorem 17.31. □

The proof of the uniqueness part of Theorem 17.30 is essentially the same as for proofs we gave for $\mathbb{Z}$ and $F[\mathtt{X}]$.

Analogous to Theorems 1.6 and 17.8, we have:

Theorem 17.33. *Let D be a PID. For any $a, b \in D$, there exists a greatest common divisor d of a and b, and moreover, $aD + bD = dD$.*

Proof. Exercise. □

As an immediate consequence of the previous theorem, we see that in a PID D, for all $a, b \in D$ with greatest common divisor d, there exist $s, t \in D$ such that $as + bt = d$; moreover, $a, b \in D$ are relatively prime if and only if there exist $s, t \in D$ such that $as + bt = 1$.

Analogous to Theorems 1.7 and 17.9, we have:

Theorem 17.34. *Let D be a PID. For $a, b, c \in D$ such that $c \mid ab$ and a and c are relatively prime, we have $c \mid b$.*

Proof. Exercise. □

Analogous to Theorems 1.8 and 17.10, we have:

Theorem 17.35. *Let D be a PID. Let $p \in D$ be irreducible, and let $a, b \in D$. Then $p \mid ab$ implies that $p \mid a$ or $p \mid b$. That is, all irreducibles in D are prime.*

Proof. Exercise. □

Theorem 17.30 now follows immediately from Theorems 17.32, 17.35, and 17.27.

Exercise 17.25. Show that $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain.

Exercise 17.26. Consider the polynomial

$$\mathtt{X}^3 - 1 = (\mathtt{X} - 1)(\mathtt{X}^2 + \mathtt{X} + 1).$$

Over $\mathbb{C}$, the roots of $\mathtt{X}^3 - 1$ are $1, (-1 \pm \sqrt{-3})/2$. Let $\omega := (-1 + \sqrt{-3})/2$, and note that $\omega^2 = -1 - \omega = (-1 - \sqrt{-3})/2$, and $\omega^3 = 1$.

(a) Show that the ring $\mathbb{Z}[\omega]$ consists of all elements of the form $a + b\omega$, where $a, b \in \mathbb{Z}$, and is an integral domain. This ring is called the ring of **Eisenstein integers**.

(b) Show that the only units in $\mathbb{Z}[\omega]$ are ± 1, $\pm\omega$, and $\pm\omega^2$.

(c) Show that $\mathbb{Z}[\omega]$ is a Euclidean domain.

Exercise 17.27. Show that in a PID, all non-zero prime ideals are maximal.

Recall that for a complex number $\alpha = a + bi$, with $a, b \in \mathbb{R}$, the norm of α was defined as $N(\alpha) = \alpha\bar{\alpha} = a^2 + b^2$ (see Example 9.5). There are other measures of the "size" of a complex number that are useful. The **absolute value** of α is defined as $|\alpha| := \sqrt{N(\alpha)} = \sqrt{a^2 + b^2}$. The **max norm** of α is defined as $M(\alpha) := \max\{|a|, |b|\}$.

EXERCISE 17.28. Let $\alpha, \beta \in \mathbb{C}$. Prove the following statements.

(a) $|\alpha\beta| = |\alpha||\beta|$.

(b) $|\alpha + \beta| \leq |\alpha| + |\beta|$.

(c) $N(\alpha + \beta) \leq 2(N(\alpha) + N(\beta))$.

(d) $M(\alpha) \leq |\alpha| \leq \sqrt{2}M(\alpha)$.

The following exercises develop algorithms for computing with Gaussian integers. We shall assume that for computational purposes, a Gaussian integer $\alpha = a + bi$, with $a, b \in \mathbb{Z}$, is represented as the pair of integers (a, b).

EXERCISE 17.29. Let $\alpha, \beta \in \mathbb{Z}[i]$.

(a) Show how to compute $M(\alpha)$ in time $O(\text{len}(M(\alpha)))$ and $N(\alpha)$ in time $O(\text{len}(M(\alpha))^2)$.

(b) Show how to compute $\alpha + \beta$ in time $O(\text{len}(M(\alpha)) + \text{len}(M(\beta)))$.

(c) Show how to compute $\alpha \cdot \beta$ in time $O(\text{len}(M(\alpha)) \cdot \text{len}(M(\beta)))$.

(d) Assuming $\beta \neq 0$, show how to compute $\xi, \rho \in \mathbb{Z}[i]$ such that $\alpha = \beta\xi + \rho$, $N(\rho) \leq \frac{1}{2}N(\beta)$, and $N(\xi) \leq 4N(\alpha)/N(\beta)$. Your algorithm should run in time $O(\text{len}(M(\alpha)) \cdot \text{len}(M(\beta)))$. Hint: see Example 17.14; also, to achieve the stated running time bound, your algorithm should first test if $M(\beta) \geq 2M(\alpha)$.

EXERCISE 17.30. Using the division with remainder algorithm from part (d) of the previous exercise, adapt the Euclidean algorithm for (ordinary) integers to work with Gaussian integers. On inputs $\alpha, \beta \in \mathbb{Z}[i]$, your algorithm should compute a greatest common divisor $\delta \in \mathbb{Z}[i]$ of α and β in time $O(\ell^3)$, where $\ell := \max\{\text{len}(M(\alpha)), \text{len}(M(\beta))\}$.

EXERCISE 17.31. Extend the algorithm of the previous exercise, so that it computes $\sigma, \tau \in \mathbb{Z}[i]$ such that $\alpha\sigma + \beta\tau = \delta$. Your algorithm should run in time $O(\ell^3)$, and it should also be the case that $\text{len}(M(\sigma))$ and $\text{len}(M(\tau))$ are $O(\ell)$.

The algorithms in the previous two exercises for computing greatest common divisors in $\mathbb{Z}[i]$ run in time cubic in the length of their input, whereas the corresponding algorithms for $\mathbb{Z}$ run in time quadratic in the length of their input. This is essentially because the running time of the algorithm for division with remainder discussed in Exercise 17.29 is insensitive to the size of the quotient.

To get a quadratic-time algorithm for computing greatest common divisors in $\mathbb{Z}[i]$, in the following exercises we shall develop an analog of the binary gcd algorithm for $\mathbb{Z}$.

EXERCISE 17.32. Let $\pi := 1 + i \in \mathbb{Z}[i]$.

(a) Show that $2 = \pi\bar{\pi} = -i\pi^2$, that $N(\pi) = 2$, and that π is irreducible in $\mathbb{Z}[i]$.

(b) Let $\alpha \in \mathbb{Z}[i]$, with $\alpha = a + bi$ for $a, b \in \mathbb{Z}$. Show that $\pi \mid \alpha$ if and only if $a - b$ is even, in which case

$$\frac{\alpha}{\pi} = \frac{a+b}{2} + \frac{b-a}{2}i.$$

(c) Show that for any $\alpha \in \mathbb{Z}[i]$, we have $\alpha \equiv 0 \pmod{\pi}$ or $\alpha \equiv 1 \pmod{\pi}$.

(d) Show that the quotient ring $\mathbb{Z}[i]/\pi\mathbb{Z}[i]$ is isomorphic to the ring $\mathbb{Z}_2$.

(e) Show that for any $\alpha \in \mathbb{Z}[i]$ with $\alpha \equiv 1 \pmod{\pi}$, there exists a unique $\epsilon \in \{\pm 1, \pm i\}$ such that $\alpha \equiv \epsilon \pmod{2\pi}$.

(f) Show that for any $\alpha, \beta \in \mathbb{Z}[i]$ with $\alpha \equiv \beta \equiv 1 \pmod{\pi}$, there exists a unique $\epsilon \in \{\pm 1, \pm i\}$ such that $\alpha \equiv \epsilon\beta \pmod{2\pi}$.

EXERCISE 17.33. We now present a "$(1+i)$-ary gcd algorithm" for Gaussian integers. Let $\pi := 1 + i \in \mathbb{Z}[i]$. The algorithm takes non-zero $\alpha, \beta \in \mathbb{Z}[i]$ as input, and runs as follows:

```
      $\rho \leftarrow \alpha,\ \rho' \leftarrow \beta,\ e \leftarrow 0$
      while $\pi \mid \rho$ and $\pi \mid \rho'$ do $\rho \leftarrow \rho/\pi,\ \rho' \leftarrow \rho'/\pi,\ e \leftarrow e+1$
      repeat
          while $\pi \mid \rho$ do $\rho \leftarrow \rho/\pi$
          while $\pi \mid \rho'$ do $\rho' \leftarrow \rho'/\pi$
          if $M(\rho') < M(\rho)$ then $(\rho, \rho') \leftarrow (\rho', \rho)$
          determine $\epsilon \in \{\pm 1, \pm i\}$ such that $\rho' \equiv \epsilon\rho \pmod{2\pi}$
(∗)       $\rho' \leftarrow \rho' - \epsilon\rho$
      until $\rho' = 0$
      $\delta \leftarrow \pi^e \cdot \rho$
      output $\delta$
```

Show that this algorithm correctly computes a greatest common divisor of α and β, and can be implemented so as to run in time $O(\ell^2)$, where $\ell := \max(\mathrm{len}(M(\alpha)), \mathrm{len}(M(\beta)))$. Hint: to analyze the running time, for $i = 1, 2, \ldots$, let v_i (respectively, v_i') denote the value of $|\rho\rho'|$ just before (respectively, after) the execution of the line marked (∗) in loop iteration i, and show that

$$v_i' \leq (1 + \sqrt{2})v_i \quad \text{and} \quad v_{i+1} \leq v_i'/2\sqrt{2}.$$

EXERCISE 17.34. Extend the algorithm of the previous exercise, so that it computes $\sigma, \tau \in \mathbb{Z}[i]$ such that $\alpha\sigma + \beta\tau = \delta$. Your algorithm should run in time $O(\ell^2)$, and it should also be the case that $\operatorname{len}(M(\sigma))$ and $\operatorname{len}(M(\tau))$ are $O(\ell)$. Hint: adapt the algorithm in Exercise 4.2.

Exercise 17.35. In Exercise 17.32, we saw that 2 factors as $-i\pi^2$ in $\mathbb{Z}[i]$, where $\pi := 1 + i$ is irreducible. This exercise examines the factorization in $\mathbb{Z}[i]$ of prime numbers $p > 2$.

(a) Suppose -1 is not congruent to the square of any integer modulo p. Show that p is irreducible in $\mathbb{Z}[i]$.

(b) Suppose that $c^2 \equiv -1 \pmod{p}$ for some $c \in \mathbb{Z}$. Let $\gamma := c + i \in \mathbb{Z}[i]$ and let δ be a greatest common divisor in $\mathbb{Z}[i]$ of γ and p. Show that $p = \delta\bar{\delta}$, and that δ and $\bar{\delta}$ are non-associate, irreducible elements of $\mathbb{Z}[i]$.

17.8.2 Unique factorization in $D[\mathtt{X}]$

In this section, we prove the following:

Theorem 17.36. *If D is a UFD, then so is $D[\mathtt{X}]$.*

This theorem implies, for example, that $\mathbb{Z}[\mathtt{X}]$ is a UFD. Applying the theorem inductively, one also sees that for any field F, the ring $F[\mathtt{X}_1, \ldots, \mathtt{X}_n]$ of multi-variate polynomials over F is also a UFD.

We begin with some simple observations. First, recall that for an integral domain D, $D[\mathtt{X}]$ is an integral domain, and the units in $D[\mathtt{X}]$ are precisely the units in D. Second, it is easy to see that an element of D is irreducible in D if and only if it is irreducible in $D[\mathtt{X}]$. Third, for $c \in D$ and $f = \sum_i a_i\mathtt{X}^i \in D[\mathtt{X}]$, we have $c \mid f$ if and only if $c \mid a_i$ for all i.

We call a non-zero polynomial $f \in D[\mathtt{X}]$ **primitive** if the only elements in D that divide f are units. If D is a UFD, then given any non-zero polynomial $f \in D[\mathtt{X}]$, we can write it as $f = cf'$, where $c \in D$ and $f' \in D[\mathtt{X}]$ is a primitive polynomial: just take c to be a greatest common divisor of all the coefficients of f.

It is easy to prove the existence part of Theorem 17.36:

Theorem 17.37. *Let D be a UFD. Any non-zero, non-unit element of $D[\mathtt{X}]$ can be expressed as a product of irreducibles in $D[\mathtt{X}]$.*

Proof. Let f be a non-zero, non-unit polynomial in $D[\mathtt{X}]$. If f is a constant, then because D is a UFD, it factors into irreducibles in D. So assume f is not constant. If f is not primitive, we can write $f = cf'$, where c is a non-zero, non-unit in D, and f' is a primitive, non-constant polynomial in $D[\mathtt{X}]$. Again, as D is a UFD, c factors into irreducibles in D.

From the above discussion, it suffices to prove the theorem for non-constant, primitive polynomials $f \in D[\mathtt{X}]$. If f is itself irreducible, we are done. Otherwise, then we can write $f = gh$, where $g, h \in D[\mathtt{X}]$ and neither g nor h are units. Further, by the assumption that f is a primitive, non-constant polynomial, both g and h must also be primitive, non-constant polynomials; in particular, both g and h have degree strictly less than $\deg(f)$, and the theorem follows by induction on degree. □

The uniqueness part of Theorem 17.36 is (as usual) more difficult. We begin with the following fact:

Theorem 17.38. *Let D be a UFD, let p be an irreducible in D, and let $f, g \in D[\mathtt{X}]$. Then $p \mid fg$ implies $p \mid f$ or $p \mid g$.*

Proof. Consider the quotient ring D/pD, which is an integral domain (because D is a UFD), and the corresponding ring of polynomials $(D/pD)[\mathtt{X}]$, which is also an integral domain. Consider the natural map from $D[\mathtt{X}]$ to $(D/pD)[\mathtt{X}]$ that sends $a \in D[\mathtt{X}]$ to the polynomial $\bar{a} \in (D/pD)[\mathtt{X}]$ obtained by mapping each coefficient of a to its residue class modulo p. If $p \mid fg$, then we have

$$0 = \overline{fg} = \bar{f}\bar{g},$$

and since $(D/pD)[\mathtt{X}]$ is an integral domain, it follows that $\bar{f} = 0$ or $\bar{g} = 0$, which means that $p \mid f$ or $p \mid g$. □

Theorem 17.39. *Let D be a UFD. The product of two primitive polynomials in $D[\mathtt{X}]$ is also primitive.*

Proof. Let $f, g \in D[\mathtt{X}]$ be primitive polynomials, and let $h := fg$. If h is not primitive, then $m \mid h$ for some non-zero, non-unit $m \in D$, and as D is a UFD, there is some irreducible element $p \in D$ that divides m, and therefore, divides h as well. By Theorem 17.38, it follows that $p \mid f$ or $p \mid g$, which implies that either f is not primitive or g is not primitive. □

Suppose that D is a UFD and that F is its field of fractions. Any non-zero polynomial $f \in F[\mathtt{X}]$ can always be written as $f = (c/d)f'$, where $c, d \in D$, with $d \neq 0$, and $f' \in D[\mathtt{X}]$ is primitive. To see this, clear the denominators of the coefficients of f, writing $df = f''$, where $0 \neq d \in D$ and $f'' \in D[\mathtt{X}]$. Then take c to be a greatest common divisor of the coefficients of f'', so that $f'' = cf'$, where $f' \in D[\mathtt{X}]$ is primitive. Then we have $f = (c/d)f'$, as required. Of course, we may assume that c and d are relatively prime—if not, we may divide c and d by a greatest common divisor.

As a consequence of the previous theorem, we have:

Theorem 17.40. *Let D be a UFD and let F be its field of fractions. Let $f, g \in D[\mathtt{X}]$ and $h \in F[\mathtt{X}]$ be non-zero polynomials such that $f = gh$ and g is primitive. Then $h \in D[\mathtt{X}]$.*

Proof. Write $h = (c/d)h'$, where $c, d \in D$ and $h' \in D[\mathtt{X}]$ is primitive. Let us assume that c and d are relatively prime. Then we have

$$d \cdot f = c \cdot gh'. \tag{17.11}$$

We claim that $d \in D^*$. To see this, note that (17.11) implies that $d \mid (c \cdot gh')$, and the assumption that c and d are relatively prime implies that $d \mid gh'$. But by Theorem 17.39, gh' is primitive, from which it follows that d is a unit. That proves the claim.

It follows that $c/d \in D$, and hence $h = (c/d)h' \in D[\mathtt{X}]$. □

Theorem 17.41. *Let D be a UFD and F its field of fractions. If $f \in D[\mathtt{X}]$ with $\deg(f) > 0$ is irreducible, then f is also irreducible in $F[\mathtt{X}]$.*

Proof. Suppose that f is not irreducible in $F[\mathtt{X}]$, so that $f = gh$ for non-constant polynomials $g, h \in F[\mathtt{X}]$, both of degree strictly less than that of f. We may write $g = (c/d)g'$, where $c, d \in D$ and $g' \in D[\mathtt{X}]$ is primitive. Set $h' := (c/d)h$, so that $f = gh = g'h'$. By Theorem 17.40, we have $h' \in D[\mathtt{X}]$, and this shows that f is not irreducible in $D[\mathtt{X}]$. □

Theorem 17.42. *Let D be a UFD. Let $f \in D[\mathtt{X}]$ with $\deg(f) > 0$ be irreducible, and let $g, h \in D[\mathtt{X}]$. If f divides gh in $D[\mathtt{X}]$, then f divides either g or h in $D[\mathtt{X}]$.*

Proof. Suppose that $f \in D[\mathtt{X}]$ with $\deg(f) > 0$ is irreducible. This implies that f is a primitive polynomial. By Theorem 17.41, f is irreducible in $F[\mathtt{X}]$, where F is the field of fractions of D. Suppose f divides gh in $D[\mathtt{X}]$. Then because $F[\mathtt{X}]$ is a UFD, f divides either g or h in $F[\mathtt{X}]$. But Theorem 17.40 implies that f divides either g or h in $D[\mathtt{X}]$. □

Theorem 17.36 now follows immediately from Theorems 17.37, 17.38, and 17.42, together with Theorem 17.27.

In the proof of Theorem 17.36, there is a clear connection between factorization in $D[\mathtt{X}]$ and $F[\mathtt{X}]$, where F is the field of fractions of D. We should perhaps make this connection more explicit. Suppose $f \in D[\mathtt{X}]$ factors into irreducibles in $D[\mathtt{X}]$ as

$$f = c_1^{a_1} \cdots c_r^{a_r} h_1^{b_1} \cdots h_s^{b_s}.$$

where the c_i are non-associate, irreducible constants, and the h_i are non-associate, irreducible, non-constant polynomials (and in particular, primitive). By Theorem 17.41, the h_i are irreducible in $F[\mathtt{X}]$. Moreover, by Theorem 17.40, the h_i are non-associate in $F[\mathtt{X}]$. Therefore, in $F[\mathtt{X}]$, f factors as

$$f = ch_1^{b_1} \cdots h_s^{b_s},$$

where $c := c_1^{a_1} \cdots c_r^{a_r}$ is a unit in F, and the h_i are non-associate irreducible polynomials in $F[\mathtt{X}]$.

***Example* 17.15.** It is important to keep in mind the distinction between factorization in $D[\mathtt{X}]$ and $F[\mathtt{X}]$. Consider the polynomial $2\mathtt{X}^2 - 2 \in \mathbb{Z}[\mathtt{X}]$. Over $\mathbb{Z}[\mathtt{X}]$, this polynomial factors as $2(\mathtt{X}-1)(\mathtt{X}+1)$, where each of these three factors are irreducible in $\mathbb{Z}[\mathtt{X}]$. Over $\mathbb{Q}[\mathtt{X}]$, this polynomial has two irreducible factors, namely, $\mathtt{X}-1$ and $\mathtt{X}+1$. □

The following theorem provides a useful criterion for establishing that a polynomial is irreducible.

Theorem 17.43 (Eisenstein's criterion). *Let D be a UFD and F its field of fractions. Let $f = f_n\mathtt{X}^n + f_{n-1}\mathtt{X}^{n-1} + \cdots + f_0 \in D[\mathtt{X}]$. If there exists an irreducible $p \in D$ such that*

$$p \nmid f_n,\ p \mid f_{n-1},\ \cdots,\ p \mid f_0,\ p^2 \nmid f_0,$$

then f is irreducible over F.

Proof. Let f be as above, and suppose it were not irreducible in $F[\mathtt{X}]$. Then by Theorem 17.41, we could write $f = gh$, where $g, h \in D[\mathtt{X}]$, both of degree strictly less than that of f. Let us write

$$g = g_r\mathtt{X}^r + \cdots + g_0 \quad \text{and} \quad h = h_s\mathtt{X}^s + \cdots + h_0,$$

where $g_r \neq 0$ and $h_s \neq 0$, so that $0 < r < n$ and $0 < s < n$. Now, since $f_n = g_r h_s$, and $p \nmid f_n$, it follows that $p \nmid g_r$ and $p \nmid h_s$. Further, since $f_0 = g_0 h_0$, and $p \mid f_0$ but $p^2 \nmid f_0$, it follows that p divides one of g_0 or h_0, but not both—for concreteness, let us assume that $p \mid g_0$ but $p \nmid h_0$. Also, let t be the smallest positive integer such that $p \nmid g_t$—note that $0 < t \le r < n$.

Now consider the natural map that sends $c \in D$ to $\bar{c} \in D/pD$, which we can extend coefficient-wise to the map that sends $a \in D[\mathtt{X}]$ to $\bar{a} \in (D/pD)[\mathtt{X}]$. Because D is a UFD and p is irreducible, both D/pD and $(D/pD)[\mathtt{X}]$ are integral domains. Since $f = gh$, we have

$$\bar{f}_n\mathtt{X}^n = \bar{f} = \bar{g}\bar{h} = (\bar{g}_r\mathtt{X}^r + \cdots + \bar{g}_t\mathtt{X}^t)(\bar{h}_s\mathtt{X}^s + \cdots + \bar{h}_0). \tag{17.12}$$

But notice that when we multiply out the two polynomials on the right-hand side of (17.12), the coefficient of $\mathtt{X}^t$ is $\bar{g}_t\bar{h}_0 \neq 0$, and as $t < n$, this clearly contradicts the fact that the coefficient of $\mathtt{X}^t$ in the polynomial on the left-hand side of (17.12) is zero. □

As an application of Eisenstein's criterion, we have:

Theorem 17.44. *For any prime number q, the qth cyclotomic polynomial*

$$\Phi_q := \frac{\mathtt{X}^q - 1}{\mathtt{X} - 1} = \mathtt{X}^{q-1} + \mathtt{X}^{q-2} + \cdots + 1$$

is irreducible over $\mathbb{Q}$.

Proof. Let

$$f := \Phi_q\big[\ \mathtt{X} + 1\ \big] = \frac{(\mathtt{X}+1)^q - 1}{(\mathtt{X}+1) - 1}.$$

It is easy to see that

$$f = \sum_{i=0}^{q-1} a_i \mathtt{X}_i, \quad \text{where} \quad a_i = \binom{q}{i+1} \quad (i = 0, \ldots, q-1).$$

Thus, $a_{q-1} = 1$, $a_0 = q$, and for $0 < i < q - 1$, we have $q \mid a_i$ (see Exercise 1.12). Theorem 17.43 therefore applies, and we conclude that f is irreducible over $\mathbb{Q}$. It follows that Φ_q is irreducible over $\mathbb{Q}$, since if $\Phi_q = gh$ were a non-trivial factorization of Φ_q, then $f = \Phi_q\big[\ \mathtt{X}+1\ \big] = g\big[\ \mathtt{X}+1\ \big] \cdot h\big[\ \mathtt{X}+1\ \big]$ would be a non-trivial factorization of f. □

Exercise 17.36. Show that neither $\mathbb{Z}[\mathtt{X}]$ nor $F[\mathtt{X}, \mathtt{Y}]$ (where F is a field) are PIDs (even though they are UFDs).

Exercise 17.37. Let $f \in \mathbb{Z}[\mathtt{X}]$ be a monic polynomial. Show that if f has a root $\alpha \in \mathbb{Q}$, then $\alpha \in \mathbb{Z}$, and α divides the constant term of f.

Exercise 17.38. Let a be a non-zero, square-free integer, with $a \notin \{\pm 1\}$. For integer $n \geq 1$, show that the polynomial $\mathtt{X}^n - a$ is irreducible in $\mathbb{Q}[\mathtt{X}]$.

Exercise 17.39. Show that the polynomial $\mathtt{X}^4 + 1$ is irreducible in $\mathbb{Q}[\mathtt{X}]$.

Exercise 17.40. Let F be a field, and consider the ring of bivariate polynomials $F[\mathtt{X}, \mathtt{Y}]$. Show that in this ring, the polynomial $\mathtt{X}^2 + \mathtt{Y}^2 - 1$ is irreducible, provided F does not have characteristic 2. What happens if F has characteristic 2?

Exercise 17.41. Design and analyze an efficient algorithm for the following problem. The input is a pair of polynomials $a, b \in \mathbb{Z}[\mathtt{X}]$, along with their greatest common divisor d in the ring $\mathbb{Q}[\mathtt{X}]$. The output is the greatest common divisor of a and b the ring $\mathbb{Z}[\mathtt{X}]$.

Exercise 17.42. Let $a, b \in \mathbb{Z}[\mathtt{X}]$ be non-zero polynomials with $d := \gcd(a, b) \in \mathbb{Z}[\mathtt{X}]$. Show that for any prime p not dividing $\mathrm{lc}(a)\,\mathrm{lc}(b)$, we have $\bar{d} \mid \gcd(\bar{a}, \bar{b})$, and except for finitely many primes p, we have $\bar{d} = \gcd(\bar{a}, \bar{b})$. Here, $\bar{d}$, $\bar{a}$, and $\bar{b}$ denote the images of d, a, and b in $\mathbb{Z}_p[\mathtt{X}]$.

Exercise 17.43. Let F be a field, and let $f, g \in F[\mathtt{X}, \mathtt{Y}]$. Define $V(f, g) := \{(x, y) \in F \times F : f(x, y) = g(x, y) = 0_F\}$. Show that if f and g are relatively prime, then $V(f, g)$ is a finite set. Hint: consider the rings $F(\mathtt{X})[\mathtt{Y}]$ and $F(\mathtt{Y})[\mathtt{X}]$.

17.9 Notes

The "$(1 + i)$-ary gcd algorithm" in Exercise 17.33 for computing greatest common divisors of Gaussian integers is based on algorithms in Weilert [100] and Damgård and Frandsen [31]. The latter paper also develops a corresponding algorithm for Eisenstein integers (see Exercise 17.26). Weilert [101] presents an asymptotically fast algorithm that computes the greatest common divisor of ℓ-bit Gaussian integers in time $O(\ell^{1+o(1)})$.