Rings and Modules

MAT 3143

Fall 2020

Alistair Savage

Department of Mathematics and Statistics

University of Ottawa

Contents

Preface

These notes are aimed at students in the course *Rings and Modules* (MAT 3143) at the University of Ottawa. This is a first course in ring and module theory. In this course, we study the general definition of a ring and the types of maps that we allow between them, before turning our attention to the important example of polynomials rings. We then discuss classes of rings that have some additional nice properties (e.g. euclidean domains, principal ideal domains, and unique factorization domains). We then discuss modules, which are generalizations of vector spaces, where we allow scalars to lie in a ring.

Notation: In this course $\mathbb{N} = \mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \dots\}$ denotes the set of nonnegative integers.

Acknowledgement: Portions of these notes are based on handwritten notes of Erhard Neher, Hadi Salmasian, and Damien Roy. Other portions follow the text Abstract algebra: theory and applications by Tom Judson, which is the recommended text for the course. Sections 4.8 and 4.9 follow Abstract: abstract and concrete by Frederick Goodman. Many exercises in these notes are borrowed from these textbooks, as well as from *Introduction to abstract algebra* by W. Keith Nicholson.

Alistair Savage

Course website: https://alistairsavage.ca/mat3143

Chapter 1

Rings

In this chapter we introduce the main object of this course. We start with the basic definition of a ring, give several important examples, and deduce some important properties. We then turn our attention to integral domains and fields, two important types of rings. Next, we discuss the important concept of an ideal and the related notion of quotient rings. Finally, we conclude with a discussion of ring homomorphisms and state the important first isomorphism theorem. A reference for the material in this chapter is [Jud19, Ch. 16].

1.1 Examples and basic properties

Definition 1.1.1 (Ring)**.** A *ring* is a nonempty set R with two binary operations, usually written as (and called) *addition* and *multiplication*, satisfying the following axioms.

- **(R1)** $a + b = b + a$ for all $a, b \in R$.
- **(R2)** $(a + b) + c = a + (b + c)$ for all $a, b, c \in R$.
- **(R3)** There exists an element $0 \in R$ such that $a + 0 = a$ for all $a \in R$.
- **(R4)** For every $a \in R$, there exists an element $-a \in R$ such that $a + (-a) = 0$.
- **(R5)** $(ab)c = a(bc)$ for all $a, b, c \in R$.
- **(R6)** There exists an element $1 \in R$ such that $1a = a1 = a$ for all $a \in R$.
- **(R7)** $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ for all $a, b, c \in R$.

A ring R is said to be *commutative* if, in addition,

- **(R8)** $ab = ba$ for all $a, b \in R$.

When we wish to specify the ring, we sometimes write 0_R and 1_R for the elements 0 and 1.

Sometimes condition (R6) is omitted from the definition of a ring and one refers to a *ring with unity (or identity)* to specify that condition (R6) also holds. However, in this course, we will always assume that our rings have a unity and use the term *general ring* for objects that satisfy all the axioms of a ring other than (R6). Axioms (R1)–(R4) are equivalent to $(R, +)$ being an abelian group. Axioms (R5)–(R6) imply that $(R, \cdot)$ is a monoid. Thus, the element 1, called the *unity*, or *identity element*, of R, is unique. The zero element 0 is also unique. See Exercise 1.1.1.

Remark 1.1.2. *Nonassociative rings* are also an important area of study. These are objects for which we do not require (R5) to hold. (A better name might be "not necessarily associative rings".) *Lie algebras* are a well studied class of nonassociative rings.

Example 1.1.3. Each of $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{C}$ is a commutative ring.

Example 1.1.4. Let n be a positive integer and let $\mathbb{Z}_n = \{\bar{0}, \bar{1}, \ldots, \overline{n-1}\}$ with addition and multiplication performed modulo n. Then $\mathbb{Z}_n$ is a commutative ring.

Example 1.1.5 (Matrices). The set $\mathrm{Mat}_n(\mathbb{Q})$ of all $n \times n$ matrices with rational entries is a ring under matrix addition and multiplication. If $n \geq 2$, this ring is noncommutative. More generally, if R is a ring, then $\mathrm{Mat}_n(R)$ is also a ring (with the usual rules for matrix addition and multiplication).

Example 1.1.6 (Polynomial rings). For any ring R, we have the ring

$$R[x] = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m : a_0, \ldots, a_m \in R\},$$

called the *ring of polynomials with coefficients in* R. Here x is an indeterminate (and addition and multiplication of polynomials is "formal"). We will discuss polynomial rings in further detail in the next chapter.

Example 1.1.7 (Function rings). If X is a nonempty set, then the set $\mathcal{F}(X, \mathbb{R})$ of real valued functions $f\colon X \to \mathbb{R}$ is a commutative ring under pointwise addition and multiplication.

Example 1.1.8 (Direct product of rings). If $R_1, \ldots, R_n$ are rings, then their *direct product* is the cartesian product $R_1 \times \cdots \times R_n$ with the operations

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n),$$
$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = (a_1 b_1, \ldots, a_n b_n).$$

Example 1.1.9 (The zero ring). The smallest ring is the *zero ring* $R = \{0\}$. A ring R is the zero ring (i.e. has only one element, its zero element) if and only if $1_R = 0_R$. See Exercise 1.1.2. Since the zero ring is not very interesting, we usually assume that rings are nonzero.

Example 1.1.10. The set $2\mathbb{Z}$ of even integers, with the usual addition and multiplication, is a general ring that is not a ring. Another such example is the set of all 3×3 real matrices whose bottom row is zero, with usual addition and multiplication of matrices.

Definition 1.1.11 (Unit, multiplicative inverse)**.** Let R be a ring. An element $a \in R$ is called a *unit* if there exists an element $b \in R$ such that $ab = ba = 1$. The element b is called the *multiplicative inverse* of a. The set of units of R is denoted $R^\times$. (In some references, the set of units is denoted R^*. We use the notation $R^\times$ to avoid confusion with the dual of a vector space.)

Proposition 1.1.12 (Uniqueness of multiplicative inverses)**.** *If $a \in R$ has a multiplicative inverse, then this inverse is unique.*

Proof. If b and c are both multiplicative inverses of a, then

$$b = b1 = b(ac) = (ba)c = 1c = c. \qquad \square$$

Proposition 1.1.13. *For every ring R, the set $R^\times$ is a group under multiplication.*

Proof. The proof of this proposition is left as Exercise 1.1.5. $\square$

From now on, we will call $R^\times$ the *group of units* of R.

Suppose R is a ring. If $m \in \mathbb{Z}$ and $a \in R$, then we define

$$ma := \begin{cases} \underbrace{a + a + \cdots + a}_{m \text{ summands}} & \text{if } m > 0, \\ 0 & \text{if } m = 0, \\ \underbrace{-a + (-a) + \cdots + (-a)}_{|m| \text{ summands}} & \text{if } m < 0. \end{cases} \tag{1.1}$$

If $m \in \mathbb{N}$, then we define

$$a^m := \underbrace{a \cdot a \cdots a}_{m \text{ factors}}.$$

Here we interpret $a^0 = 1$. If a is a unit, then a^m is defined for all $m \in \mathbb{Z}$. For negative m, we define

$$a^m := \underbrace{a^{-1} \cdot a^{-1} \cdots a^{-1}}_{|m| \text{ factors}}.$$

It is easy to show (see Exercise 1.1.6) that

$$(m+n)a = ma + na, \quad m(na) = (mn)a \quad \text{for all } m, n \in \mathbb{Z},\ a \in R, \tag{1.2}$$

$$a^{m+n} = a^m a^n, \quad (a^m)^n = a^{mn} \quad \text{for all } m, n \in \mathbb{N},\ a \in R. \tag{1.3}$$

If a is a unit, then (1.3) holds for all $m, n \in \mathbb{Z}$.

Theorem 1.1.14. *Let R be a ring and $r, s \in R$. Then*

(a) $r0 = 0 = 0r$,

(b) $(-r)s = r(-s) = -(rs)$,

(c) $(-r)(-s) = rs$, *and*

(d) $(mr)(ns) = (mn)(rs)$ *for all* $m, n \in \mathbb{Z}$.

(Here $0 = 0_R$, as opposed to the integer zero.)

Proof. For $r \in R$, we have $r0 = r(0 + 0) = r0 + r0$. Adding $-r0$ to both sides gives $r0 = 0$. The proof of the relation $0r = 0$ is similar. The remainder of the relations are left as an exercise (or see [Jud19, Prop. 16.8]). $\square$

Definition 1.1.15 (Subtraction)**.** If r and s are elements of a ring R, their *difference* is defined to be

$$r - s := r + (-s).$$

In this way, we can define *subtraction* in a ring.

Definition 1.1.16 (Characteristic)**.** If R is any ring, the *characteristic* of R, denoted $\operatorname{char} R$, is defined to be the order of 1_R in $(R, +)$ if this order is finite and zero if this order is infinite.

Example 1.1.17. We have

$$\operatorname{char}\mathbb{Z} = 0, \quad \operatorname{char}\mathbb{R} = 0, \quad \operatorname{char}\mathbb{C} = 0, \quad \operatorname{char}\mathbb{Z}_n = n, \quad \operatorname{char}(\text{zero ring}) = 1.$$

Remark 1.1.18. Note that if the ring R is finite (i.e. $|R| < \infty$), then $\operatorname{char} R > 0$.

Lemma 1.1.19. *The characteristic of a ring R is the exponent of the ring's additive group. That is, it is the smallest positive integer n such that $na = 0$ for all $a \in R$.*

Proof. It suffices to show that $n1 = 0$ if and only if, for all $a \in R$, we have $na = 0$. Clearly the latter condition implies the former (just take $a = 1$). Now, if $n1 = 0$, then, for all $a \in R$, we have $na = n(1a) = (n1)a = 0a = 0$. □

Definition 1.1.20 (Idempotent, nilpotent)**.** An element $a \in R$ is called an *idempotent* if $a^2 = a$. It is called *nilpotent* if $a^n = 0$ for some positive integer n.

Examples 1.1.21. In any ring R, the elements 0 and 1 are idempotents and 0 is nilpotent. In $\mathrm{Mat}_2(\mathbb{R})$ we have other idempotents:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \ldots .$$

In $\mathrm{Mat}_2(\mathbb{R})$ we also have many nilpotent elements:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix}, \ldots .$$

Definition 1.1.22 (Subring)**.** A subset $S \subseteq R$ of a ring R is called a *subring* of R if it is itself a ring with the same operations (and the same unity) as R.

Proposition 1.1.23 (Subring Test)**.** *A subset $S \subseteq R$ of a ring R is a subring of R if*

(SR1) *$0 \in S$ and $1 \in S$,*

(SR2) *If $s, t \in S$, then $s + t$, st and $-s$ are all in S.*

Alternatively, S is a subring of R if

(SR1′) *$1 \in S$,*

(SR2′) *$rs \in S$ for all $r, s \in S$, and*

(SR3′) *$r - s \in S$ for all $r, s \in S$.*

Proof. The proof of this proposition is left as Exercise 1.1.7. □

Examples 1.1.24. (a) $\mathbb{Z}$ is a subring of $\mathbb{R}$.

(b) $\mathrm{Mat}_2(\mathbb{Z})$ is a subring of $\mathrm{Mat}_2(\mathbb{Q})$.

(c) $\begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix} := \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} : x, y, z \in \mathbb{R} \right\}$ is a subring of $\mathrm{Mat}_2(\mathbb{R})$.

(d) $2\mathbb{Z}$ is *not* a subring of $\mathbb{Z}$.

(e) The *gaussian integers* $\mathbb{Z}(i) := \{a + bi : a, b \in \mathbb{Z}\}$ form a subring of $\mathbb{C}$.

(f) If $a, b \in \mathbb{R}$ with $a < b$, then the continuous real valued functions on the interval $[a, b]$ form a subring of $\mathcal{F}([a, b], \mathbb{R})$.

Example 1.1.25. Let us find $\mathbb{Z}(i)^\times$. We have

$$(a+bi)(c+di) = 1 \implies (a-bi)(c-di) = \overline{(a+bi)(c+di)} = \bar{1} = 1 \implies (a^2+b^2)(c^2+d^2) = 1.$$

Since $a^2 + b^2, c^2 + d^2 \in \mathbb{N}$, this implies that $a^2 + b^2 = 1$. Thus, either $a = 0$, $b = \pm 1$ or $a = \pm 1$, $b = 0$. Therefore $\mathbb{Z}(i)^\times = \{\pm 1, \pm i\}$.

Definition 1.1.26 (Center)**.** The *center* of a ring R is defined to be

$$Z(R) := \{r \in R : rs = sr \text{ for all } s \in R\}.$$

Example 1.1.27. A ring R is commutative if and only if $Z(R) = R$.

Lemma 1.1.28. *The center $Z(R)$ of a ring R is a subring of R.*

Proof. The proof of this lemma is left as Exercise 1.1.10. □

Later, in Section 1.4, we will discuss the notion of a ring homomorphism. However, it is useful to give here the definition of the special case of a ring isomorphism (see also Definition 1.4.10).

Definition 1.1.29 (Isomorphic rings)**.** Two rings R, S are said to be *isomorphic*, and we write $R \cong S$, if there exists a map $\sigma\colon R \to S$ such that

(a) σ is bijective,

(b) $\sigma(a + b) = \sigma(a) + \sigma(b)$ for all $a, b \in R$, and

(c) $\sigma(ab) = \sigma(a)\sigma(b)$ for all $a, b \in R$.

The map σ is called an *isomorphism*.

Remark 1.1.30. Note that if $\sigma\colon R \to S$ is an isomorphism of rings, then we have the following:

(a) σ is an isomorphism of the corresponding additive groups. In particular, $\sigma(0_R) = 0_S$.

(b) $\sigma(1_R) = 1_S$. This can be seen as follows. For any $s \in S$, there exists $r \in R$ such that $\sigma(r) = s$. Thus $s\sigma(1_R) = \sigma(r)\sigma(1_R) = \sigma(r1_R) = \sigma(r) = s$. Similarly $\sigma(1_R)s = s$. Since $s \in S$ was arbitrary, this implies that $\sigma(1_R)$ is the unity of S.

Example 1.1.31. Let

$$R_1 = \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix} = \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} : a, b, c \in \mathbb{R} \right\} \quad \text{and} \quad R_2 = \begin{bmatrix} \mathbb{R} & 0 \\ \mathbb{R} & \mathbb{R} \end{bmatrix} = \left\{ \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} : a, b, c \in \mathbb{R} \right\}.$$

Consider the map $\sigma\colon R_1 \to R_2$ given by

$$\sigma\left(\begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} c & 0 \\ b & a \end{bmatrix}$$

It is easy to see that $\sigma^2 = \mathrm{id}$. Thus σ is invertible and hence bijective. If $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, we also have

$$\begin{gathered} \sigma(A+B) = M(A+B)M^{-1} = MAM^{-1} + MBM^{-1} = \sigma(A) + \sigma(B), \\ \sigma(A)\sigma(B) = (MAM^{-1})(MBM^{-1}) = MABM^{-1} = \sigma(AB). \end{gathered}$$

Thus σ is an isomorphism and so $R_1 \cong R_2$.

Note that the map $A \mapsto A^T$ is *not* a ring isomorphism since $(AB)^T \neq A^T B^T$ in general.

Exercises.

1.1.1. Prove that the identity element in any monoid (hence in any group) is unique.

1.1.2. Show that a ring R is the zero ring (i.e. $R = \{0\}$) if and only if $0_R = 1_R$.

1.1.3. Explain why each of the following is not a ring.

(a) $\mathbb{N} = \{0, 1, 2, 3, \dots, \}$ under usual addition and multiplication.

(b) $2\mathbb{Z}$.

(c) The set of all mappings $f\colon \mathbb{R} \to \mathbb{R}$ under pointwise addition, but with multiplication given by composition.

1.1.4. Suppose R is a ring. The *opposite ring* R^{op} of R is the same set as R, with the same addition, but with multiplication given by $r \cdot s = sr$. (Here $\cdot$ denotes the multiplication in R^{op} and juxtaposition denotes the multiplication in R.) Prove that R^{op} is a ring.

1.1.5. Prove Proposition 1.1.13.

1.1.6. Prove (1.2) and (1.3).

1.1.7. Prove Proposition 1.1.23.

1.1.8 ([Nic12, Ex. 3.1.3]). For each of the following, show that S is a subring of R.

(a) $S = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{R},\ a + c = b + d \right\}$, $R = \mathrm{Mat}_2(\mathbb{R})$.

(b) $S = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in \mathbb{R} \right\}$, $R = \mathrm{Mat}_2(\mathbb{R})$.

(c) $S = \left\{ \begin{bmatrix} a & 0 & b \\ 0 & c & d \\ 0 & 0 & a \end{bmatrix} : a, b, c, d \in \mathbb{R} \right\}$, $R = \mathrm{Mat}_3(\mathbb{R})$.

(d) $S = \left\{ \begin{bmatrix} a & 2b \\ b & a \end{bmatrix} : a, b \in \mathbb{R} \right\}$, $R = \mathrm{Mat}_2(\mathbb{R})$.

1.1.9 ([Nic12, Ex. 3.1.4]). If S and T are subrings of R, show that $S \cap T$ is a subring of R. Is it necessarily true that $S + T = \{s + t : s \in S,\ t \in T\}$ is a subring of R?

1.1.10. Prove Lemma 1.1.28.

1.1.11. Find the center of $\mathrm{Mat}_2(\mathbb{R})$.

1.1.12. Show that if R and S are rings, then $(R \times S)^\times = R^\times \times S^\times$.

1.1.13 ([Nic12, Ex. 3.1.5]). Suppose X is a nonempty subset of a ring R. The *centralizer* of X in R is defined to be

$$C(X) = \{c \in R : cx = xc \text{ for all } x \in X\}.$$

Prove that $C(X)$ is a subring of R.

1.1.14 ([Nic12, Ex. 3.1.10]). Suppose a and b are elements of a ring.

(a) If $ab + ba = 1$ and $a^3 = a$, prove that $a^2 = 1$.

(b) If $ab = a$ and $ba = b$, prove that $a^2 = a$ and $b^2 = b$.

1.1.15 ([Nic12, Ex. 3.1.11]). Show that 0 is the only nilpotent in a ring R if and only if $a^2 = 0$ implies $a = 0$.

1.1.16 ([Nic12, Ex. 3.1.14]). Given r and s in a ring R, show that $1 + rs$ is a unit if and only if $1 + sr$ is a unit. *Hint*: $s(1 + rs) = (1 + sr)s$.

1.1.17. Find all the rings of characteristic one.

1.1.18 ([Nic12, Ex. 3.1.18]). Find the characteristics of the following rings.

(a) $\mathbb{Z}_n \times \mathbb{Z}_m$.

(b) $\mathrm{Mat}_2(\mathbb{Z}_n)$.

(c) $\mathbb{Z} \times \mathbb{Z}_n$.

1.1.19 ([Nic12, Ex. 3.1.20]). If $ua = au$, where u is a unit and a is a nilpotent, show that $u + a$ is a unit.

1.1.20 ([Nic12, Ex. 3.1.28]). For each of the following rings, find all the units, nilpotents, and idempotents:

(a) $\mathbb{Z}$

(b) $\mathbb{Z}_{24}$

(c) $\mathrm{Mat}_2(\mathbb{Z}_2)$

(d) $\begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix}$

1.1.21 ([Nic12, Ex. 3.1.36]). For each of the following pairs of rings, show that the two rings are not isomorphic.

(a) $\mathbb{R}$ and $\mathbb{C}$.

(b) $\mathbb{Q}$ and $\mathbb{R}$.

(c) $\mathbb{Z}$ and $\mathbb{Q}$.

(d) $\mathbb{Z}_8$ and $\mathbb{Z}_4 \times \mathbb{Z}_2$.

1.2 Integral domains and fields

We now consider certain special types of rings that we will examine in further detail later in the course.

Definition 1.2.1 (Zero divisor, integral domain, division ring, field)**.** Suppose R is a ring. A nonzero element $r \in R$ is said to be a *zero divisor* if there exists a nonzero $s \in R$ such that $rs = 0$ or $sr = 0$. A nonzero ring is said to be a *domain* if it has no zero divisors. A commutative domain is called an *integral domain*. If every nonzero element in a (not necessarily commutative) nonzero ring is a unit, then R is called a *division ring* (or *skew field*). A commutative division ring is called a *field*.

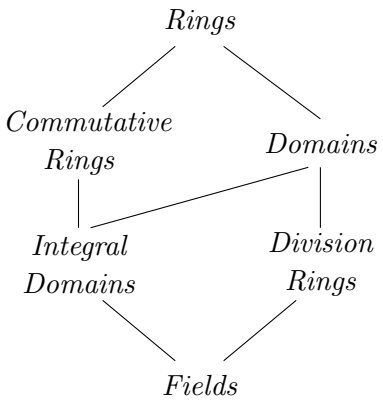

Figure 1.1: Types of rings

Example 1.2.2. The rings $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ are fields. The ring $\mathbb{Z}$ is an integral domain, but not a field.

Example 1.2.3. The ring $\mathbb{Z}(i)$ of gaussian integers is an integral domain (Exercise 1.2.1) but not a field, since $\mathbb{Z}(i)^\times = \{\pm 1, \pm i\}$ (see Example 1.1.25).

Example 1.2.4. The ring $\mathrm{Mat}_2(\mathbb{R})$ is not a domain since, for example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Proposition 1.2.5. *The ring $\mathbb{Z}_m$ is a field if and only if m is a prime number.*

Proof. Let m be a prime. For every $\bar{a} \in \mathbb{Z}_m$, $\bar{a} \neq 0$, we have $\gcd(a, m) = 1$ and so there exist $x, y \in \mathbb{Z}$ such that $ax + my = 1$. Thus $\bar{a}\bar{x} = \bar{1}$ in $\mathbb{Z}_m$. Thus every nonzero element of $\mathbb{Z}_m$ is a unit.

Now suppose m is not prime. Then $m = m_1 m_2$ for some $1 < m_1, m_2 < m$. In particular, $\bar{m}_1, \bar{m}_2 \neq 0$, but $\bar{m}_1 \bar{m}_2 = \bar{m} = \bar{0}$. Therefore m_1 and m_2 are not units of $\mathbb{Z}_m$. □

Proposition 1.2.6. *Let R be a ring such that $|R| = p$, where p is a prime number. Then R is isomorphic to $\mathbb{Z}_p$. In particular, R is a field.*

Proof. By Lagrange's Theorem, the additive group $\langle 1 \rangle$ generated by the unity is equal to all of R and this must be isomorphic to the *group* $\mathbb{Z}_p$ (since this is the only group of order p when p is a prime number). Define $\sigma \colon \mathbb{Z}_p \to R$ by $\sigma(\bar{m}) = m1_R$. Then we have the following:

- σ is injective since, for $a, b \in \mathbb{Z}$,

  $$\sigma(\bar{a}) = \sigma(\bar{b}) \implies a1 = b1 \implies (a-b)1 = 0 \implies p|(a-b) \implies \bar{a} = \bar{b} \text{ in } \mathbb{Z}_p,$$

  where the third implication follows from the fact that $(a-b)1 = 0$ in the additive group $(R, +) \cong \mathbb{Z}_p$ (isomorphism of groups) if and only if $p|(a-b)$. (Note that we *cannot* simply conclude from $a1 = b1$ that $a = b$, since the expressions $a1$ and $b1$ do not represent products in the ring R. Rather they are of the form (1.1).)
- σ is bijective because σ is injective and $|R| = |\mathbb{Z}_p|$.
- We have
  $$\sigma(\bar{a} + \bar{b}) = \sigma\left(\overline{a+b}\right) = (a+b)1 = a1 + b1 = \sigma(\bar{a}) + \sigma(\bar{b}).$$
- $\sigma(\overline{ab}) = \sigma(\bar{a})\sigma(\bar{b})$. The proof of this fact is similar to the above and is left as an exercise. □

Proposition 1.2.7. *Let R be a ring. Then the following statements are equivalent:*

(a) R is a domain.

(b) If $ab = ac$ in R and $a \neq 0$, then $b = c$.

(c) If $ba = ca$ in R and $a \neq 0$, then $b = c$.

Proof. (a) $\implies$ (b). Suppose R is a domain and $ab = ac$ in R. Then $a(b-c) = ab - ac = 0$. Since $a \neq 0$ and R is a domain, we must have $b - c = 0$, hence $b = c$.

(b) $\implies$ (a). Suppose R satisfies (b) and that $ab = 0$ in R. Then we have $ab = a0$. Thus, if $a \neq 0$, we must have $b = 0$ by (b). Hence R is a domain.

We have proven the equivalence of (a) and (b). The proof that (a) and (c) are equivalent is similar. □

Examples 1.2.8. (a) Every division ring is a domain.

(b) Every field is an integral domain.

(c) If R is an (integral) domain and S is a subring of R, then S is an (integral) domain.

(d) $\mathbb{Z}(i)$ is an integral domain (since it is a subring of $\mathbb{C}$).

Example 1.2.9. The subring $\mathbb{Q}(\sqrt{2}) := \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ of the field of complex numbers is a field. Being a subring of $\mathbb{C}$, it is an integral domain. Thus, it remains to show that every nonzero element has a multiplicative inverse. Suppose $x \in \mathbb{Q}(\sqrt{2}) \setminus \{0\}$. Then $x = a + b\sqrt{2}$ with $a \neq 0$ or $b \neq 0$. Then

$$x^{-1} = \frac{1}{a + b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}.$$

Note that $a^2 - 2b^2 \neq 0$ since if $a^2 - 2b^2 = 0$, then $\frac{a^2}{b^2} = 2$, which implies that $\sqrt{2} = \pm\frac{a}{b} \in \mathbb{Q}$, which contradicts the fact that $\sqrt{2}$ is an irrational number.

Definition 1.2.10 (Quaternions)**.** The *quaternions* are the ring

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$$

with multiplication determined by the rules

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik, \quad i^2 = j^2 = k^2 = -1.$$

For $w = a + bi + cj + dk \in \mathbb{H}$, we define $w^* := a - bi - cj - dk$ and $N(w) := a^2 + b^2 + c^2 + d^2$.

For example, we have

$$(3 - 4j)(2i + k) = 6i + 3k - 8ji - 4jk = 6i + 3k + 8k - 4i = 2i + 11k.$$

Lemma 1.2.11. *If $w \in \mathbb{H} \setminus \{0\}$, then $w^{-1} = \frac{1}{N(w)}w^*$.*

Proof. See [Jud19, Example 16.7]. □

By Lemma 1.2.11, the ring of quaternions is a division ring. However, it is not a field since it is not commutative. Thus, the quaternions are an example of a noncommutative division ring.

Proposition 1.2.12. *The characteristic of an integral domain is either zero or a prime number.*

Proof. Let R be an integral domain and suppose that the characteristic of R is $n \neq 0$. If n is not prime, then $n = ab$, with $1 < a < n$ and $1 < b < n$. Then $0 = n1 = (ab)1 = (a1)(b1)$. Since R has no zero divisors, either $a1 = 0$ or $b1 = 0$. Thus the characteristic of R must be less than n, which is a contradiction. Therefore, n must be prime. □

Proposition 1.2.13. *Every finite integral domain is a field.*

Proof. Let R be an integral domain with $|R| = n < \infty$. Suppose $r \in R$ with $r \neq 0$. Then, by Proposition 1.2.7, the function $f\colon R \to R$, $f(x) = rx$ is injective. Since R is finite, this implies that f is bijective. Thus $f(s) = rs = 1$ for some $s \in R$. Hence r is a unit. Since r was an arbitrary nonzero element of R, this implies that R is a field. □

A proof of the following theorem can be found in [Jud19, Th. 16.16]. (We will not use this theorem.)

Theorem 1.2.14 (Wedderburn's little theorem)**.** *Every finite division ring is a field.*

We conclude this section by showing that every integral domain "embeds" into a field. We do this by mimicking the construction of the rational numbers from the integers. We assume for the rest of this section that R is an integral domain.

Let

$$X = \{(r, u) : r, u \in R,\ u \neq 0\}$$

and define a relation on X by

$$(r, u) \equiv (s, v) \iff rv = su.$$

This is an equivalence relation on X. It is easy to see that this relation is reflexive and symmetric. We therefore must show that it is transitive. Suppose that

$$(r, u) \equiv (s, v) \quad \text{and} \quad (s, v) \equiv (t, w).$$

Then, by definition, we have $rv = su$ and $sw = tv$. Therefore,

$$(rw)v = (rv)w = (su)w = u(sw) = utv \qquad (\text{since } R \text{ is commutative}).$$

Since $(s, v) \in X$, we have $v \neq 0$ and so we can cancel v in the above by Proposition 1.2.7. This gives $rw = tu$ and so $(r, u) \equiv (t, w)$. Thus $\equiv$ is an equivalence relation on X. We write $\frac{r}{s}$ for the equivalence class of (r, s). Thus $\frac{r}{s} = \frac{s}{v}$ if and only if $(r, s) \equiv (s, v)$.

We now define

$$Q := \left\{ \frac{r}{u} : r, u \in R,\ u \neq 0 \right\}. \tag{1.4}$$

We define addition and multiplication on Q by

$$\frac{r}{u} + \frac{s}{v} = \frac{rv + su}{uv} \quad \text{and} \quad \frac{r}{u} \cdot \frac{s}{v} = \frac{rs}{uv}.$$

Note that $uv \neq 0$ since $u, v \neq 0$ and R is a domain. Since the quotients above are equivalence classes, we must show that these operations are well defined. We show this for multiplication

and refer the reader to [Jud19, Lem. 18.2] for addition. If $\frac{r}{u} = \frac{r'}{u'}$ and $\frac{s}{v} = \frac{s'}{v'}$, we must show that $\frac{rs}{uv} = \frac{r's'}{u'v'}$. This follows from the fact that

$$(rs)(u'v') = (ru')(sv') = (r'u)(s'v) = (r's')(uv).$$

We leave it as a Exercise 1.2.14 to show that Q is a field with zero $\frac{0}{1}$ and unity $\frac{1}{1}$. The negative of an element $\frac{r}{u}$ is $\frac{-r}{u}$. Furthermore, if $\frac{r}{u}$ in nonzero in Q, then $r \neq 0$. Thus $\frac{u}{r} \in Q$ and we have $\left(\frac{r}{u}\right)^{-1} = \frac{u}{r}$.

Define

$$R' := \left\{ \frac{r}{1} : r \in R \right\}. \tag{1.5}$$

It is easy (Exercise 1.2.15) to verify that R' is a subring of R. Furthermore, it is not hard to check that the map

$$\sigma \colon R \to R', \quad \sigma(r) = \frac{r}{1}, \quad r \in R,$$

is an isomorphism of rings (see [Jud19, Th. 18.4]) and so $R \cong R'$. One usually identifies R with R' via this isomorphism. That is we set $r = \frac{r}{1}$ for all $r \in R$. In this way, we view R as a subring of Q. We call Q the *field of fractions* (or *field of quotients*) of the integral domain R.

Example 1.2.15. The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$. The field of fractions of $\mathbb{R}[x]$ is called the field of *rational functions*.

Exercises.

1.2.1. Show that the ring of gaussian integers is an integral domain.

1.2.2. Show that $\mathbb{Z}_m^\times = \{\bar{k} \in \mathbb{Z}_m : \gcd(k, m) = 1\}$.

1.2.3 ([Nic12, Ex. 3.2.1]). Find all the roots of $x^2 + 3x - 4$ in the rings $\mathbb{Z}$, $\mathbb{Z}_6$, and $\mathbb{Z}_4$.

1.2.4 ([Nic12, Ex. 3.2.2]). Suppose p is a prime and define

$$\mathbb{Z}_{(p)} = \left\{ \frac{n}{m} \in \mathbb{Q} : p \text{ does not divide } m \right\}.$$

Show that $\mathbb{Z}_{(p)}$ is an integral domain and find all its units.

1.2.5 ([Nic12, Ex. 3.2.3]). Determine all the idempotents and nilpotent elements in a domain.

1.2.6 ([Nic12, Ex. 3.2.4]). Suppose R and S are nonzero rings. It is possible for $R \times S$ to be a domain?

1.2.7 ([Nic12, Ex. 3.2.5]). Show that $\mathrm{Mat}_n(R)$, where R is a ring, is never a domain if $n \geq 2$.

1.2.8 ([Nic12, Ex. 3.2.6]). If $a^2 = b^2$ and $a^3 = b^3$ in a domain, show that $a = b$. Now do it for $a^m = b^m$ and $a^n = b^n$ where $m, n > 1$ and $\gcd(m, n) = 1$. *Hint*: $1 = xm + yn$ for some $x, y \in \mathbb{Z}$.

1.2.9 ([Nic12, Ex. 3.2.7]). Suppose that a ring R has no nonzero nilpotent elements. If $ab = 0$ in R, show that $ba = 0$.

1.2.10 ([Nic12, Ex. 3.2.8]). Show that a ring R is a division ring if and only if, for each nonzero $a \in R$, there is a unique element $b \in R$ such that $aba = a$.

1.2.11 ([Nic12, Ex. 3.2.11]). If F is a field with q elements, show that $a^q = a$ for all $a \in F$. *Hint*: Use Lagrange's Theorem.

1.2.12 ([Nic12, Ex. 3.2.14]). Show that there is no field with 6 elements. *Hint*: Use Lagrange's Theorem.

1.2.13 ([Nic12, Ex. 3.2.15]). Show that the center of a division ring is a field.

1.2.14. Show that Q as defined in (1.4) is a field.

1.2.15. Show that R' as defined in (1.5) is a subring of Q.

1.2.16 ([Nic12, Ex. 3.2.28]). Let R be a commutative ring and call $u \in R$ a nonzero-divisor if $ur = 0$, $r \in R$, implies $r = 0$. Let $U \subseteq R$ be a set of nonzero-divisors in R such that $1 \in U$, and $ab \in U$ whenever $a, b \in U$. Generalize the field of fractions construction by showing that a ring of quotients

$$Q = \left\{ \frac{r}{u} : r \in R,\ u \in U \right\}$$

exists. Show further that R can be regarded as a subring of Q and, in this case, that each element of U is a unit in Q and $Q = \{ru^{-1} : r \in R,\ u \in U\}$.

1.3 Ideals and quotient rings

Recall that that if G is a group and $H \subseteq G$, then in order for the quotient set G/H to naturally be a group, we need H to be a *normal* subgroup of G. What is the analogous notion for rings? That is, if R is a ring and $S \subseteq R$, when is R/S naturally a ring?

Example 1.3.1. Let $R = \mathbb{Z} \times \mathbb{Z}$, and $S = \{(x, x) : x \in \mathbb{Z}\}$. It is straightforward to verify that S is a subring of R. Note that $(0, 1) + S = (-1, 0) + S$. If we try to define a multiplication on the set of cosets R/S by multiplying representatives, we have

$$((0, 1) + S)((0, 1) + S) = (0, 1) + S \neq (0, 0) + S = ((0, 1) + S)((-1, 0) + S),$$

which is a contradiction.

Definition 1.3.2 (Ideal)**.** Let R be a ring. An additive subgroup $I \subseteq R$ is called an *ideal* of R if for every $r \in R$, we have $rI \subseteq I$ and $Ir \subseteq I$. The ideal I is said to be *proper* if $I \neq R$.

We now want to define the structure of a ring on the quotient set $R/I := \{a+I : a \in R\}$. (Recall that $a + I = a' + I \iff a - a' \in I$.) We want the addition and multiplication to be given by

$$(a + I) + (b + I) = (a + b) + I \quad \text{and} \quad (a + I)(b + I) = ab + I. \tag{1.6}$$

Theorem 1.3.3. *Let $I \subseteq R$ be an ideal of R. Then R/I, with addition and multiplication defined by (1.6), is a ring. The unity of R/I is $1 + I$ and the zero is $0 + I = I$. If R is commutative, then R/I is also commutative.*

Proof. We must first show that the addition and multiplication are well defined. If $a + I = a' + I$ and $b + I = b' + I$, then $a - a', b - b' \in I$. Thus

$$ab - a'b' = (a - a')b + a'(b - b') \in Ib + a'I \subseteq I$$

and so $ab+I = a'b'+I$. Thus the multiplication is well defined. The proof that the addition is also well defined and that A/I satisfies the axioms of Definition 1.1.1 is left as Exercise 1.3.1. It is clear that R/I is commutative if R is commutative. □

Proposition 1.3.4. *Let I be an ideal of a ring R. Then the following statements are equivalent:*

(a) $1 \in I$.

(b) I *contains a unit.*

(c) $I = R$.

Proof. We leave the proof of this result as Exercise 1.3.2. □

Definition 1.3.5 (Principal ideal)**.** If $a \in Z(R)$, then $Ra = aR$ and this is an ideal of R, often denoted $\langle a \rangle$. It is called the *principal ideal of R generated by a.*

Proposition 1.3.6 (Ideals of $\mathbb{Z}$)**.** *Every ideal of $\mathbb{Z}$ is principal.*

Proof. The zero ideal $\{0\}$ is a principal ideal since $0\mathbb{Z} = \{0\}$. If I is any nonzero ideal in $\mathbb{Z}$, then I must contain some positive integer. Then there exists a least positive integer n in I by the Principle of Well-Ordering. Now let a be any element in I. Using the division algorithm, we know that there exist integers q and r such that

$$a = nq + r$$

where $0 \leq r < n$. This equation tells us that $r = a - nq \in I$, but r must be 0 since n is the least positive element in I. Therefore, $a = nq$ and $I = n\mathbb{Z}$. □

Example 1.3.7. If R is any ring, then $\{0\}$ and R are ideals of R. The corresponding quotient rings are $R/\{0\} \cong R$ and $R/R \cong \{0\}$. The ideal $\{0\}$ is called the *zero ideal* of R.

Definition 1.3.8 (Simple ring)**.** A nonzero ring R is called *simple* if its only ideals are $\{0\}$ and R.

Example 1.3.9. It follows from Proposition 1.3.4 that the only ideals of a division ring R are $\{0\}$ and R. Hence division rings are simple.

Example 1.3.10. Let $R = \mathbb{Z}(i)$ be the ring of gaussian integers. Let $I = \langle 2 + i \rangle$. We wish to describe the ring R/I. Since $i - (-2) = 2 + i \in I$, we have $i + I = (-2) + I$. Thus, for all $m, n \in \mathbb{Z}$,

$$(m + ni) + I = (m + I) + (-2n + I) = (m - 2n) + I.$$

Moreover, $5 = (2 + i)(2 - i) \in (2 + i)R = I$, which implies that $5 + I = I$. Thus the only elements of R/I are $I, 1 + I, 2 + I, 3 + I, 4 + I$. Are all of these elements distinct? Suppose $n, m \in \mathbb{Z}$ and $n + I = m + I$. Thus $(m - n) + I = I$. Then, for some $a, b \in \mathbb{Z}$, we have

$$m - n = (2 + i)(a + bi) = (2a - b) + (a + 2b)i \implies a = -2b \implies m - n = -5b \in 5\mathbb{Z}.$$

Therefore, $R/I \cong \mathbb{Z}_5$ (a field).

Definition 1.3.11 (Prime ideal)**.** An ideal P of a commutative ring R is called a *prime ideal* if $P \neq R$ and

$$r, s \in R,\ rs \in P \implies r \in P \text{ or } s \in P.$$

(See Exercises 1.3.5 and 1.3.20(d) for an indication of why these ideals are called *prime*.)

Examples 1.3.12. (a) $2\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$, but $4\mathbb{Z}$ is not.

(b) $\mathbb{Z} \times \{0\}$ is a prime ideal of $\mathbb{Z} \times \mathbb{Z}$, but $2\mathbb{Z} \times \{0\}$ and $\{0\} \times \{0\}$ are not.

Theorem 1.3.13. *If R is a commutative ring, an ideal $P \neq R$ is prime if and only if R/P is an integral domain.*

Proof. Assume P is prime and $(a + P)(b + P) = ab + P = 0 + P$ in R/P. Then $ab \in P$ and so $a \in P$ or $b \in P$. Thus $a + P = 0 + P$ or $b + P = 0 + P$. Thus R/P is an integral domain.

Now assume that R/P is an integral domain and $ab \in P$. Then $(a+P)(b+P) = ab+P = 0 + P$ and so either $a + P = 0 + P$ or $b + P = 0 + P$. Hence $a \in P$ or $b \in P$. □

Examples 1.3.14. (a) $(\mathbb{Z} \times \mathbb{Z})/(\mathbb{Z} \times \{0\}) \cong \mathbb{Z}$. The ideal $\mathbb{Z} \times \{0\}$ is prime and $\mathbb{Z}$ is an integral domain.

(b) $\mathbb{Z}_n$ is an integral domain if and only if n is prime (see Exercise 1.3.5).

(c) Since, for any ring R, we have $R/\{0\} \cong R$, we have that R is an integral domain if and only if its zero ideal is prime.

Theorem 1.3.15. *Let I be an ideal of a ring R. Then there exists a bijective, inclusion preserving correspondence between the set of ideals of R/I and the set of ideals of R containing I.*

Proof. We split the proof into three steps.

(a) We first show that if $\tilde{A}$ is an ideal of R/I, then

$$A := \{b \in : b + I \in \tilde{A}\}$$

is an ideal of R containing I with $\tilde{A} = A/I$. Since $0 + I \in \tilde{A}$, we have $0 \in A$. If $a, b \in A$, then $a + I \in \tilde{A}$ and $b + I \in \tilde{A}$. Since $\tilde{A}$ is an additive subgroup of R/A, we then have $(a - b) + I = (a + I) - (b + I) \in \tilde{A}$. Hence $a - b \in A$. Thus A is an additive subgroup of R.

Now, if $a \in A$ and $r \in R$, then

$$\begin{aligned} a + I \in \tilde{A} \text{ and } r + I \in R/I &\implies ra + I = (r + I)(a + I) \in \tilde{A} \quad (\text{since } \tilde{A} \text{ in an ideal of } R/I) \\ &\implies ra \in A \quad (\text{by the definition of } A). \end{aligned}$$

Similarly, one can show that $ar \in A$. Thus A is closed under multiplication by elements of R and hence is an ideal of R.

To see that $I \subseteq A$, note that

$$a \in I \implies a + I = 0 + I \in \tilde{A} \implies a \in A \quad (\text{by the definition of } A).$$

It remains to show that $\tilde{A} = A/I$. Since

$$r + I \in \tilde{A} \implies r \in A \implies r + I \in A/I,$$

we have $\tilde{A} \subseteq A/I$. To prove the reverse inclusion, suppose $x \in A/I$. Then there is some $a \in A \subseteq R$ such that $x = a + I$. This implies that $x = a + I \in \tilde{A}$ by the definition of A.

(b) Next we show that if A is an ideal of R containing I then $\tilde{A} := \{a + I : a \in A\}$ is an ideal of R/I and $\tilde{A} = A/I := \{a + I : a \in A\}$ (which is clearly true). Since $0 \in I$, we have $0 + I \in \tilde{A}$.

Suppose $x_1, x_2 \in \tilde{A}$. Then there exist $a_1, a_2 \in A$ such that $x_1 = a_1 + I$ and $x_2 = a_2 + I$. Since A is an ideal of R, we have $a_1 - a_2 \in A$. Then $x_1 - x_2 = (a_1 + I) - (a_2 + I) = (a_1 - a_2) + I \in \tilde{A}$. Thus $\tilde{A}$ is an additive subgroup of R/I.

Now suppose $x \in \tilde{A}$ and $r + I \in R/I$. Then $x = a + I$ for some $a \in A$. Since A is an ideal of R, we have $ra \in I$. Hence $(r + I)x = (r + I)(a + I) = ra + I \in \tilde{A}$. The proof that $(a + I)(r + I) \in \tilde{A}$ is similar. Hence $\tilde{A}$ is an ideal of R/I.

(c) Finally, we show that if A_1 and A_2 are ideals of R containing I, then $A_1 \subseteq A_2$ if and only if $A_1/I \subseteq A_2/I$. That $A_1 \subseteq A_2 \implies A_1/I \subseteq A_2/I$ is obvious. We show the reverse inclusion. Suppose $A_1/I \subseteq A_2/I$. Then

$$\begin{aligned} a_1 \in A_1 &\implies a_1 + I \in A_1/I \\ &\implies \exists\ a_2 \in A_2 \text{ such that } a_1 + I = a_2 + I \\ &\implies a_1 - a_2 \in I \\ &\implies a_1 = a_2 + a \text{ for some } a \in I \\ &\implies a_1 \in A_2 \ (\text{since } a_2 + a \in A_2 + I \subseteq A_2 \text{ because } I \subseteq A_2) \end{aligned}$$ □

Theorem 1.3.16. *A commutative ring is simple if and only if it is a field.*

Proof. Since a field is a commutative division ring by definition, the reverse implication follows from Example 1.3.9. Therefore it suffices to prove the forward implication. Suppose R is a simple commutative ring and let $a \in R \setminus \{0\}$. Then $aR \neq \{0\}$ is an ideal and so $aR = R$. Thus $ab = 1$ for some $b \in R$. □

Definition 1.3.17 (Maximal ideal)**.** An ideal I of a ring R is said to be *maximal* if $I \neq R$ and there is no ideal J such that $I \subsetneq J \subsetneq R$.

Theorem 1.3.18. *Let R be any ring and let I be an ideal of R. Then R/I is simple if and only if I is a maximal ideal.*

Proof. Suppose R/I is simple and A is an ideal of R with $I \subseteq A \subseteq R$. By Theorem 1.3.15, A/I is an ideal of R/I with $I/I \subseteq A/I \subseteq R/I$. Since R/I is simple, we have $A/I = R/I$ or $A/I = I/I$, which implies that $A = R$ or $A = I$.

Now suppose that I is maximal and, towards a contradiction, that R/I is not simple. Then R/I has a nonzero proper ideal of the form A/I. So $I/I \subsetneq A/I \subsetneq R/I$. Then, by Theorem 1.3.15, A is an ideal of R with $I \subsetneq A \subsetneq R$, which is a contradiction. □

Corollary 1.3.19. *An ideal I of a commutative ring R is maximal if and only if R/I is a field.*

Proof. This follows from Theorems 1.3.16 and 1.3.18. □

Corollary 1.3.20. *Every maximal ideal of a commutative ring is a prime ideal.*

Proof. Suppose I is a maximal ideal of a commutative ring R. Then, by Corollary 1.3.19, the quotient R/I is a field. Hence R/I is an integral domain and so I is prime by Theorem 1.3.13. □

Example 1.3.21. We have that $\{0\}$ is a prime ideal of $\mathbb{Z}$ that is not maximal. Similarly, $\{0\} \times \mathbb{Z}$ is an ideal of $\mathbb{Z} \times \mathbb{Z}$ that is prime but not maximal.

Example 1.3.22. By Theorem 1.3.15, the ideals of $\mathbb{Z}/n\mathbb{Z}$ are of the form $A/n\mathbb{Z}$ for some ideal A of $\mathbb{Z}$ such that $n\mathbb{Z} \subseteq A$. By Proposition 1.3.6, $A = m\mathbb{Z}$ for some $m \in \mathbb{Z}$. Since $n\mathbb{Z} \subseteq m\mathbb{Z}$ if and only if $m|n$, the ideals of $\mathbb{Z}/n\mathbb{Z}$ are $m\mathbb{Z}/n\mathbb{Z}$ for $m|n$.

As an explicit example, consider $n = 6$. The bijective correspondence of Theorem 1.3.15 is given explicitly by:

Ideal of $\mathbb{Z}/6\mathbb{Z}$	Ideal of $\mathbb{Z}$
$\mathbb{Z}/6\mathbb{Z} = \langle \bar{1} \rangle$	$\mathbb{Z}$
$\{\bar{0}, \bar{2}, \bar{4}\} = \langle \bar{2} \rangle$	$2\mathbb{Z}$
$\{\bar{0}, \bar{3}\} = \langle \bar{3} \rangle$	$3\mathbb{Z}$
$\{\bar{0}\} = \langle \bar{0} \rangle = \langle \bar{6} \rangle$	$6\mathbb{Z}$

Example 1.3.23. We will discuss polynomial rings in detail in Chapter 2. However, we give here an example involving them. Consider

$$R = \mathbb{Z}[x] = \{a_0 + a_1 x + \cdots + a_k x^k : k \in \mathbb{N},\ a_0, \ldots, a_k \in \mathbb{Z}\},$$
$$I = \{4a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k : k \in \mathbb{N},\ a_0, \ldots, a_k \in \mathbb{Z}\}.$$

Then

$$R/I = \{0 + I, 1 + I, 2 + I, 3 + I\} \cong \mathbb{Z}_4.$$

And, for example, the ideal $\langle \bar{2} \rangle \subseteq \mathbb{Z}_4$ corresponds to the ideal

$$J = \{2a_0 + a_1x + \cdots + a_kx^k : k \in \mathbb{N},\ a_0, \ldots, a_k \in \mathbb{Z}\}.$$

Remark 1.3.24. We have seen (Theorem 1.3.16) that if a commutative ring is simple, then it is a field. For noncommutative rings, this situation is more complicated. There are simple rings that are not division rings. For example, if F is a field, then $\mathrm{Mat}_n(F)$ is a simple ring that is *not* a division ring. This follows from the following result.

Theorem 1.3.25. *If R is a ring and $n \in \mathbb{N}_+$, then every ideal of $\mathrm{Mat}_n(R)$ is of the form $\mathrm{Mat}_n(I)$ where I is an ideal of R.*

We will not prove the above theorem (nor will we use it). The proof can be found, for instance, in [Nic12, Lem. 3.3.3].

Exercises.

1.3.1. Complete the proof of Theorem 1.3.3.

1.3.2. Prove Proposition 1.3.4.

1.3.3 ([Nic12, Ex. 3.3.1]). For each of the following, decide whether A is an ideal of the ring R. Justify your answer.

(a) $R = \mathbb{C}$, $A = \mathbb{Z}$.

(b) $R = \mathbb{Z} \times \mathbb{Z}$, $A = \{(k, k) : k \in \mathbb{Z}\}$.

(c) $R = \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix}$, $A = \begin{bmatrix} 0 & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix}$.

(d) $R = \begin{bmatrix} \mathbb{Z} & \mathbb{Z} \\ 0 & \mathbb{Z} \end{bmatrix}$, $A = \begin{bmatrix} \mathbb{Z} & 2\mathbb{Z} \\ 0 & \mathbb{Z} \end{bmatrix}$.

(e) $R = \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix}$, $A = \begin{bmatrix} \mathbb{Z} & \mathbb{R} \\ 0 & \mathbb{Z} \end{bmatrix}$.

(f) $R = \mathbb{Z}(i)$, $A = \{n + ni : n \in \mathbb{Z}\}$.

1.3.4 ([Nic12, Ex. 3.3.2]). Suppose S is a ring. If $R = \begin{bmatrix} S & S \\ 0 & S \end{bmatrix}$ and $A = \begin{bmatrix} 0 & S \\ 0 & 0 \end{bmatrix}$, show that A is an ideal of R and describe the cosets in R/A.

1.3.5. Show that $n\mathbb{Z}$, $n \in \mathbb{N}$, is a prime ideal of $\mathbb{Z}$ if and only if n is either zero or a prime number.

1.3.6. Suppose R is a ring and $m \in \mathbb{Z}$. Show that $mR = \{mr : R \in R\}$ and $A_m = \{r \in R : mr = 0\}$ are ideals of R.

1.3.7. Prove that the maximal ideals of $\mathbb{Z}$ are precisely the ideals $p\mathbb{Z}$ where p is a prime number.

1.3.8 ([Nic12, Ex. 3.3.5]). Suppose R and S are rings.

(a) If A is an ideal of R and B is an ideal of S, show that $A \times B$ is an ideal of $R \times S$.

(b) Show that every ideal $\mathcal{A}$ of $R \times S$ is of the form $\mathcal{A} = A \times B$ for some ideal A of R and some ideal B of S. *Hint*: $A = \{a \in R : (a, 0) \in \mathcal{A}\}$.

(c) Show that the maximal ideals of $R \times S$ are either of the form $A \times S$, where A is a maximal ideal of R, or of the form $R \times B$, where B is a maximal ideal of S.

1.3.9 ([Nic12, Ex. 3.3.6]). If A is an ideal of R, show that $\mathrm{Mat}_2(A)$ is an ideal of $\mathrm{Mat}_2(R)$.

1.3.10 ([Nic12, Ex. 3.3.8]). If A and B are ideals of a ring R such that $A \cap B = 0$, show that $ab = 0 = ba$ for all $a \in A$ and $b \in B$.

1.3.11 ([Nic12, Ex. 3.3.9]). Let $R = \mathbb{Z}(i)$ be the ring of gaussian integers. For each of the following, find the number of elements in the factor ring R/A and describe the cosets:

(a) $A = Ri$

(b) $A = R(1 - i)$

(c) $A = R(1 + 2i)$

(d) $A = R(1 + 3i)$

Hint: $(1 + 2i)(1 - i) = 3 + i$ and $(1 + 3i)(1 - 3i) = 10$.

1.3.12 ([Nic12, Ex. 3.3.10]). If R is a simple ring, show that $Z(R)$ is a field. Show that the converse is not true by considering $R = \begin{bmatrix} F & F \\ 0 & F \end{bmatrix}$ where F is a field.

1.3.13 ([Nic12, Ex. 3.3.12]). If $X \subseteq R$ is a nonempty subset of a commutative ring R, define the *annihilator* of X to be $\mathrm{ann}(X) = \{a \in R : ax = 0 \text{ for all } x \in X\}$.

(a) Show that $\mathrm{ann}(X)$ is an ideal of R.

(b) If $X \subseteq Y$, show that $\mathrm{ann}(Y) \subseteq \mathrm{ann}(X)$.

(c) Show that $\mathrm{ann}(X \cup Y) = \mathrm{ann}(X) \cap \mathrm{ann}(Y)$.

(d) Show that $X \subseteq \mathrm{ann}\big(\mathrm{ann}(X)\big)$.

(e) Show that $\mathrm{ann}(X) = \mathrm{ann}\Big(\mathrm{ann}\big(\mathrm{ann}(X)\big)\Big)$.

1.3.14 ([Nic12, Ex. 3.3.13]). Give an example of a ring R and an ideal A such that R/A is commutative, but R is not.

1.3.15 ([Nic12, Ex. 3.3.14]). If X and Y are additive subgroups of R, define $X + Y = \{x + y : x \in X,\ y \in Y\}$.

(a) Show that $X + Y$ is an additive subgroup of R that contains both X and Y.

(b) If A and B are ideals of R, show that $A + B$ is an ideal of R.

(c) If A is an ideal of R and S is a subring of R, show that $A + S$ is a subring of R.

1.3.16 ([Nic12, Ex. 3.3.15]). If A is an ideal of R, show that $A \cap S$ is an ideal of S for all subrings S of R.

1.3.17 ([Nic12, Ex. 3.3.16]). If A is an ideal of R, show that R/A is commutative if and only if $rs - sr \in A$ for all $r, s \in R$.

1.3.18 ([Nic12, Ex. 3.3.17]). Let $Z = Z(R)$ denote the center of a ring R.

(a) When is Z an ideal of R? Justify your answer.

(b) If R is simple, show that Z is a field.

(c) If R/Z is cyclic as an additive group, show that R is commutative.

1.3.19 ([Nic12, Ex. 3.3.24]). An additive subgroup L of R is called a *left ideal* if $Ra \subseteq L$ for all $a \in L$. Show that R is a division ring if and only if 0 and R are the only left ideals of R (extending Theorem 1.3.16). *Hint*: Ra is a left ideal for each $a \in R$.

1.3.20 ([Nic12, Ex. 3.3.25]). Let R be a commutative ring. Write $a|b$ if $b = ra$ for some $r \in R$.

(a) Show that $Rab \subseteq Ra \cap Rb$ for all $a, b \in R$.

(b) If $Ra + Rb = R$ (see Exercise 1.3.15), show that $Rab = Ra \cap Rb$.

(c) Show that $u \in R$ is a unit if and only if $Ru = R$.

(d) Suppose $p \in R$ and $Rp \neq R$. Show that Rp is a prime ideal if and only $p|ab$ implies $p|a$ or $p|b$.

(e) If R is an integral domain, show that $Ra = Rb$ if and only if $a = ub$ for some unit $u \in R$.

1.3.21 ([Nic12, Ex. 3.3.26]). Let A, B, and C be ideals of R and define

$$AB = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n : a_i \in A,\ b_i \in B,\ n \geq 1\}.$$

(a) Show that AB is an ideal of R and $AB \subseteq A \cap B$.

(b) Show that $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$. (See Exercise 1.3.15.)

(c) Show that $AR = A = RA$.

(d) Show that $A(BC) = (AB)C$.

1.3.22 ([Nic12, Ex. 3.3.27]). If $a \in R$, let

$$RaR = \{r_1as_1 + r_2as_2 + \cdots + r_nas_n : r_i, s_i \in R,\ n \geq 1\}.$$

Show that RaR is an ideal of R containing a and that it is contained in any such ideal.

1.3.23. Let R be a commutative ring. The ring R is said to be *noetherian* if it satisfies the *ascending chain condition* on ideals: for every ascending chain $I_1 \subseteq I_2 \subseteq \cdots$ of ideals in R, there exists a positive integer N such that $I_n = I_N$ for all $n \geq N$. An ideal I of R is said to be *finitely generated* if there exist elements $a_1, \ldots, a_m \in I$ (called *generators* of I) such that every element of I can be written as $a_1 r_1 + \cdots + a_m r_m$ for some $r_1, \ldots, r_m \in R$. Prove that R is noetherian if and only if every ideal of R is finitely generated.

1.3.24 (Bonus problem). Prove that every nonzero ring has a maximal ideal.

1.4 Homomorphisms

We now discuss the natural maps between rings, together with some of their properties.

Definition 1.4.1 (Ring homomorphism)**.** Let R and S be rings. A mapping $f\colon R \to S$ is called a *ring homomorphism* if it satisfies the following properties.

(RH1) $f(1_R) = 1_S$.

(RH2) $f(a+b) = f(a) + f(b)$ for all $a, b \in R$.

(RH3) $f(ab) = f(a)f(b)$ for all $a, b \in R$.

Examples 1.4.2. (a) $f\colon \mathbb{Z} \to \mathbb{Q}$, $f(x) = x$ is a ring homomorphism.

(b) $f\colon \mathbb{Z} \to \mathbb{Z}_n$, $f(x) = \bar{x}$ is a ring homomorphism.

(c) $f\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, $f(x, y) = x$ is a ring homomorphism.

(d) $f\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, $f(x, y) = x + y$ is *not* a ring homomorphism since, for example,

$$f((1,1)(1,2)) = f(1,2) = 1 + 2 = 3 \quad \text{but} \quad f(1,1)f(1,2) = (1+1)(1+2) = 6.$$

Remark 1.4.3. If $f\colon R \to S$ is a ring homomorphism, then f is a homomorphism of additive groups:

$$f(0_R) = f(0_R + 0_R) = f(0_R) + f(0_R) \implies f(0_R) = 0_S,$$
$$f(a) + f(-a) = f(a + (-a)) = f(0_R) = 0_S \implies f(-a) = -f(a).$$

Here we only used axiom (RH2).

Axioms (RH2) and (RH3) are not enough to conclude that $f(1_R) = 1_S$ as the following example illustrates.

Example 1.4.4. The mapping $f\colon \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ given by $f(x) = (x, 0)$ satisfies (RH2) and (RH3), but not (RH1). Some references call mappings satisfying (RH2) and (RH3) but not necessarily (RH1) *general ring homomorphisms*.

Proposition 1.4.5. *Let $f\colon R \to S$ be a ring homomorphism. Then the following hold.*

(a) For every $m \in \mathbb{Z}$ and $r \in R$, we have $f(mr) = mf(r)$.

(b) For every $m \in \mathbb{N}$ and $r \in R$, we have $f(r^m) = f(r)^m$.

(c) *If $u \in R^\times$, then $f(u) \in S^\times$ and, for every $m \in \mathbb{Z}$, we have $f(u^m) = f(u)^m$. In particular, $f(u^{-1}) = f(u)^{-1}$.*

Proof. The proof of this result is left as Exercise 1.4.1. □

Example 1.4.6. We will show that the equation $m^3 - 6n^3 = 3$ has no solutions for $m, n \in \mathbb{Z}$. Consider the mapping $f\colon \mathbb{Z} \to \mathbb{Z}_7$ given by $f(x) = \bar{x}$. If $m^3 - 6n^3 = 3$, then $f(m)^3 + f(n)^3 = \bar{3}$ in $\mathbb{Z}_7$. But, by explicitly considering all possible values, one can see that, for $a \in \mathbb{Z}_7$, we have $a^3 \in \{\bar{0}, \bar{1}, \bar{6}\}$. We compute

$$\bar{0} + \bar{0} = \bar{0},\ \bar{0} + \bar{1} = \bar{1},\ \bar{0} + \bar{6} = \bar{6},\ \bar{1} + \bar{1} = \bar{2}, \bar{1} + \bar{6} = \bar{0},\ \bar{6} + \bar{6} = \bar{5}.$$

Thus, the equation $\bar{m}^3 - 6\bar{n}^3 = \bar{3}$ does not have any solutions in $\mathbb{Z}_7$. It follows that $m^3 - 6n^3 = 3$ cannot have any solutions in $\mathbb{Z}$.

Example 1.4.7 (Frobenius homomorphism). Let R be a commutative ring with $p = \operatorname{char} R$ a prime number. Then the map

$$f\colon R \to R, \quad f(r) = r^p,$$

is a ring homomorphism, called the *Frobenius homomorphism*. See Exercise 1.4.3.

Definition 1.4.8 (Kernel, image)**.** Let $f\colon R \to S$ be a ring homomorphism. The *kernel* of f is $\ker f := \{x \in R : f(x) = 0\}$ and the *image* of f is $\operatorname{im} f = f(R) := \{f(r) : r \in R\}$.

Remark 1.4.9. A ring homomorphism f is one-to-one if and only if $\ker f = \{0\}$. This follows from the corresponding fact about group homomorphisms since f is a homomorphism of additive groups. By definition, f is surjective (or onto) if and only if $f(R) = S$.

Definition 1.4.10 (Monomorphism, epimorphism, isomorphism, automorphism)**.** An injective ring homomorphism is called a *monomorphism*. A surjective ring homomorphism is called an *epimorphism*. A ring homomorphism that is both a monomorphism and an epimorphism is called an *isomorphism*. If there exists an isomorphism from a ring R to a ring S then we say that R and S are *isomorphic* and write $R \cong S$. (Note that this definition agrees with Definition 1.1.29.) A homomorphism (resp. isomorphism) from a ring to itself is called an *endomorphism* (resp. *automorphism*).

Example 1.4.11. The map $\sigma\colon \mathbb{C} \to \mathbb{C}$ given by $\sigma(z) = \bar{z}$ is an automorphism of $\mathbb{C}$.

Example 1.4.12 (Inner automorphism). If R is a ring and $u \in R^\times$, then

$$\sigma_u\colon R \to R, \quad \sigma_u(r) = uru^{-1},\ r \in R,$$

is an automorphism of R called an *inner automorphism*. We leave the proof of this as Exercise 1.4.4. Also see Example 1.1.31.

Theorem 1.4.13. *Let $f\colon R \to S$ be a ring homomorphism. Then:*

(a) *$f(R)$ is a subring of S.*

(b) $\ker f$ *is an ideal of* R.

Proof. (a) Since $f(1_R) = 1_S$, we have $1_S \in f(R)$. Since $f(0_R) = 0_S$, we have $0_S \in f(R)$. Now suppose $r, s \in f(R)$. Then there exists $s', t' \in R$ such that $f(s') = s$ and $f(t') = t$. Thus

$$\begin{aligned} -s = -f(s') = f(-s') \in f(R), \\ s + t = f(s') + f(t') = f(s' + t') \in f(R), \\ st = f(s')f(t') = f(s't') \in f(R). \end{aligned}$$

(b) We know (from MAT 2143) that $\ker f$ is an additive subgroup of R (since f is a homomorphism of additive groups). Now,

$$r \in R,\ x \in \ker f \implies f(rx) = f(r)f(x) = f(r)0_S = 0_S \implies rx \in \ker f.$$

Similarly, one can show that $r \in R, x \in \ker f$ implies $xr \in \ker f$. Thus $\ker f$ is an ideal of R. □

Example 1.4.14. If I is an ideal of R, then $f\colon R \to R/I$, $f(x) = x + I$, is a ring epimorphism and $\ker f = I$.

Theorem 1.4.15 (First isomorphism theorem for rings)**.** *Let* $f\colon R \to S$ *be a ring homomorphism. Then the map*

$$\bar{f}\colon R/(\ker f) \to \operatorname{im} f, \quad \bar{f}(r + \ker f) = f(r)$$

is a ring isomorphism. In particular, $R/(\ker f) \cong \operatorname{im} f$.

Proof. Let $I = \ker f$. Then $\bar{f}$ is well-defined since

$$\begin{aligned} r + I = s + I \iff r - s \in I \iff f(r - s) = 0 \iff f(r) - f(s) = 0 \\ \iff f(r) = f(s) \iff \bar{f}(r + I) = \bar{f}(s + I). \end{aligned}$$

The reverse implications above also show that $\bar{f}$ is injective. It is clear that $\bar{f}$ is surjective and it is straightforward to verify that f is a ring homomorphism. □

Remark 1.4.16. Theorem 1.4.15 is called the *first* isomorphism theorem because there are also second and third isomorphism theorems (see Exercises 1.4.25 and 1.4.25).

Examples 1.4.17. (a) Consider $f\colon \mathbb{Z} \to \mathbb{Z}_n$, $f(x) = \bar{x}$. Then $\ker f = n\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

(b) Consider $f\colon \mathbb{Z} \to \mathbb{Z}_2 \times \mathbb{Z}_2$, $f(x) = (\bar{x}, \bar{x})$. Then $f(\mathbb{Z}) = \{(\bar{0}, \bar{0}), (\bar{1}, \bar{1})\} \cong \mathbb{Z}_2$ and $\ker f = 2\mathbb{Z}$. Thus $\mathbb{Z}/2\mathbb{Z} \cong f(\mathbb{Z}) \cong \mathbb{Z}_2$.

(c) Consider $f\colon \mathbb{Z}_4 \to \mathbb{Z}_{10}$, $f(x) = 5x$. Then $\ker f = \{\bar{0}, \bar{2}\} \subseteq \mathbb{Z}_4$ and $\operatorname{im} f = \{\bar{0}, \bar{5}\} \subseteq \mathbb{Z}_{10}$. We have $\operatorname{im} f \cong \mathbb{Z}_2$ and so $\mathbb{Z}_4/(\ker f) \cong \mathbb{Z}_2$. (Note here that f is not actually a ring homomorphism, since it does not send the unity to the unity, but it is a general ring homomorphism.)

Example 1.4.18. Let m and n be positive integers and let $I = n\mathbb{Z}_{mn} = \{n\bar{a} : \bar{a} \in \mathbb{Z}_{mn}\}$. Then the map $f\colon \mathbb{Z}_{mn} \to \mathbb{Z}_n$, $f(\bar{x}) = \bar{x}$, is a ring homomorphism with $\operatorname{im} f = \mathbb{Z}_n$ and $\ker f = I$. Thus $\mathbb{Z}_{mn}/I \cong \mathbb{Z}_n$.

Remark 1.4.19. If I and J are ideals of a ring R, then

$$I \cap J,$$
$$IJ := \{r \in R : r = a_1b_1 + \cdots + a_kb_k,\ a_1, \ldots, a_k \in I,\ b_1, \ldots, b_k \in J\}, \quad \text{and}$$
$$I + J := \{r \in R : r = a + b,\ a \in I,\ b \in J\}$$

are also ideals of R. See Exercises 1.3.15(b) and 1.3.21(a).

Theorem 1.4.20. *If R is any ring, then $\mathbb{Z}1_R = \{k1_R : k \in \mathbb{Z}\}$ is a subring of R contained in the center of R. Furthermore, we have the following.*

(a) If $n = \operatorname{char} R > 0$, then $\mathbb{Z}1_R \cong \mathbb{Z}_n$.

(b) If $\operatorname{char} R = 0$, then $\mathbb{Z}1_R \cong \mathbb{Z}$.

Proof. Define a map

$$\theta\colon \mathbb{Z} \to R, \quad \theta(k) = k1_R \text{ for all } k \in \mathbb{Z}.$$

By Proposition 1.4.5, θ is a ring homomorphism. Thus $\mathbb{Z}1_R = \theta(\mathbb{Z})$ is a subring of R by Theorem 1.4.13. Furthermore, one can verify (exercise 1.4.11) that $\mathbb{Z}1_R$ is contained in the center of R.

We have $\ker\theta = \{k \in \mathbb{Z} : k1_R = 0\}$. If $n = \operatorname{char} R > 0$, then $\ker\theta = n\mathbb{Z}$ by Lemma 1.1.19. Thus $\mathbb{Z}1_R = \theta(\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$ by the first isomorphism theorem (Theorem 1.4.15). If $\operatorname{char} R = 0$, then $\ker\theta = \{0\}$ and the result again follows by the first isomorphism theorem. □

Lemma 1.4.21. *If R and S are rings and $I \subseteq R$ and $J \subseteq S$ are ideals, then $I \times J$ is an ideal of $R \times S$ and we have $(R \times S)/(I \times J) \cong (R/I) \times (S/J)$.*

Proof. The map $f\colon R \times S \to (R/I) \times (S/J)$ given by $f(r, s) = (r + I, s + J)$ is a ring homomorphism with $\ker f = I \times J$ and $\operatorname{im} f = (R/I) \times (S/J)$ (exercise). The result then follows from the first isomorphism theorem (Theorem 1.4.15). □

Lemma 1.4.22. *Let R and S be rings. Then every ideal of $R \times S$ is of the form $A \times B$ where $A \subseteq R$ and $B \subseteq S$ are ideals.*

Proof. Let $I \subseteq R \times S$ be an ideal. Let

$$A = \{r \in R : (r, 0) \in I\}, \quad B = \{s \in S : (0, s) \in I\}.$$

It is straightforward (exercise) to show that A is an ideal of R and B is an ideal of S. If $(r, s) \in I$, then $(r, 0) = (r, s)(1, 0) \in I$ and $(0, s) = (r, s)(0, 1) \in I$. Thus $r \in A$ and $s \in B$. Hence $I \subseteq A \times B$. Now suppose $(a, b) \in A \times B$. Then, by the definition of A and B, we have $(a, 0) \in I$ and $(0, b) \in I$. Hence $(a, b) = (a, 0) + (0, b) \in I$ (since I is closed under addition). Therefore $A \times B \subseteq I$ and so $I = A \times B$. □

Theorem 1.4.23 (Chinese Remainder Theorem)**.** *Let R be a ring and let A, B be two ideals of R such that $A + B = R$. Then $R/(A \cap B) \cong (R/A) \times (R/B)$.*

Proof. Consider the map

$$f\colon R \to (R/A) \times (R/B), \quad f(r) = (r + A, r + B).$$

For $r, s \in R$, we have

$$f(1) = (1 + A, 1 + B) = 1_{(R/A)\times(R/B)},$$
$$f(r + s) = (r + s + A, r + s + B) = (r + A, r + B) + (s + A, s + B) = f(r) + f(s)$$

and

$$f(rs) = (rs + A, rs + B) = ((r + A)(s + A), (r + B)(s + B))$$
$$= (r + A, r + B)(s + A, s + B) = f(r)f(s).$$

Thus f is a ring homomorphism.

We next show that f is surjective. Since $R = A + B$, we have $1 = a + b$ for some $a \in A$ and $b \in B$. Now choose $r_1, r_2 \in R$ and set $r = r_1 b + r_2 a$. Then

$$r_1 - r = r_1 - (r_1 b + r_2 a) = r_1(1 - b) - r_2 a = r_1 a - r_2 a \in A \implies r_1 + A = r + A,$$
$$r_2 - r = r_2 - (r_1 b + r_2 a) = r_2(1 - a) - r_1 b = r_2 b - r_1 b \in B \implies r_2 + B = r + B.$$

Thus $f(r) = (r + A, r + B) = (r_1 + A, r_2 + B)$. Since $r_1, r_2 \in R$ were arbitrary, this implies that f is surjective.

Finally,

$$f(r) = (0 + A, 0 + B) \iff (r + A = 0 + A \text{ and } r + B = 0 + B)$$
$$\iff (r \in A \text{ and } r \in B) \iff r \in A \cap B.$$

So $\ker f = A \cap B$. The result then follows from the first isomorphism theorem (Theorem 1.4.15). □

Example 1.4.24. Since $\gcd(5, 6) = 1$, we have $5\mathbb{Z} + 6\mathbb{Z} = \mathbb{Z}$. We also have $5\mathbb{Z} \cap 6\mathbb{Z} = 30\mathbb{Z}$. Thus, by the Chinese Remainder Theorem (Theorem 1.4.23), we have $\mathbb{Z}_{30} \cong \mathbb{Z}_5 \times \mathbb{Z}_6$.

More generally, we have the following result.

Lemma 1.4.25. *If $\gcd(m, n) = 1$, then $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$.*

Proof. Since $\gcd(m, n) = 1$, we have $ma + nb = 1$ for some $a, b \in \mathbb{Z}$. Hence $m\mathbb{Z} + n\mathbb{Z} = \mathbb{Z}$. Also, we have $m\mathbb{Z} \cap n\mathbb{Z} = mn\mathbb{Z}$ since if $m|x$ and $n|x$, then $mn|x$ (since m and n are relatively prime). Thus the result follows from the Chinese Remainder Theorem (Theorem 1.4.23). □

Example 1.4.26. How many units does $\mathbb{Z}_{345}$ have? Since $345 = 3 \cdot 5 \cdot 23$, we have $\mathbb{Z}_{345} \cong \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_{23}$ by Lemma 1.4.25. Thus $\mathbb{Z}_{345}$ has $2 \cdot 4 \cdot 22 = 176$ units (see Exercise 1.1.12).

Example 1.4.27. Little Mary wants to take the farm's eggs to the market. She tries to put them evenly in two baskets, but one is left out. She tries three baskets, but one is still left out. Four baskets, one is left out. Five baskets. Aha! It works! How many eggs does Little Mary have?

Let $x \in \mathbb{N}$ be the number of eggs. Then we have

$$\begin{aligned} \bar{x} &= \bar{1} \text{ in } \mathbb{Z}_2, \\ \bar{x} &= \bar{1} \text{ in } \mathbb{Z}_3, \\ \bar{x} &= \bar{1} \text{ in } \mathbb{Z}_4, \\ \bar{x} &= \bar{0} \text{ in } \mathbb{Z}_5. \end{aligned}$$

First note that the third equation implies the first. So we ignore the first equation. Now, we know that $\mathbb{Z}_{12} \cong \mathbb{Z}_3 \times \mathbb{Z}_4$ by Lemma 1.4.25. Under this isomorphism $\bar{x}$ (in $\mathbb{Z}_{12}$) is mapped to $(\bar{1}, \bar{1}) \in \mathbb{Z}_3 \times \mathbb{Z}_4$. Thus $\bar{x} = \bar{1}$ in $\mathbb{Z}_{12}$.

Now, again by Lemma 1.4.25, we have $\mathbb{Z}_{60} \cong \mathbb{Z}_{12} \times \mathbb{Z}_5 \cong \mathbb{Z}_{60}/(12\mathbb{Z}_{60}) \times \mathbb{Z}_{60}/(5\mathbb{Z}_{60})$. Under this isomorphism $\bar{x}$ (in $\mathbb{Z}_{60}$) is mapped to $(\bar{1}, \bar{0}) \in \mathbb{Z}_{12} \times \mathbb{Z}_5$. Checking all the elements of $\mathbb{Z}_{60}$ which map to $\bar{1} \in \mathbb{Z}_{12}$, i.e. $\bar{1}, \overline{13}, \overline{25}, \overline{37}, \overline{49}$, we see that $\overline{25}$ reduces to $\bar{0} \in \mathbb{Z}_5$. Thus $\bar{x} = \overline{25}$ in $\mathbb{Z}_{60}$. So Mary has $25 + 60k$, $k \in \mathbb{Z}_{\geq 0}$, eggs.

Exercises.

Throughout these exercises, R is a ring.

1.4.1. Prove Proposition 1.4.5.

1.4.2 ([Nic12, Ex. 3.4.1]). For each of the following determine whether the map is a ring homomorphism. Justify your answer.

(a) $\theta\colon \mathbb{Z}_3 \to \mathbb{Z}_{12}$, where $\theta(r) = 4r$.

(b) $\theta\colon \mathbb{Z}_4 \to \mathbb{Z}_{12}$, where $\theta(r) = 3r$.

(c) $\theta\colon R \times R \to R$, where $\theta(r, s) = r + s$.

(d) $\theta\colon R \times R \to R$, where $\theta(r, s) = rs$.

(e) $\theta\colon \mathcal{F}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$, where $\theta(f) = f(1)$.

1.4.3. Show that the Frobenius homomorphism (see Example 1.4.7) is indeed a homomorphism. *Hint:* Use the binomial formula: $(a + b)^p = \sum_{i=0}^{p} \binom{p}{i} a^i b^{p-i}$.

1.4.4. Prove that the map defined in Example 1.4.12 is a ring automorphism.

1.4.5 ([Nic12, Ex. 3.4.2]). Let $\theta\colon R \to S$ be a general ring homomorphism, where R and S are rings.

(a) If θ is surjective, prove that θ is a ring homomorphism.

(b) If S is a domain and $\theta(1) \neq 0$, prove that θ is a ring homomorphism.

1.4.6 ([Nic12, Ex. 3.4.3]). Show that a general ring homomorphism $\theta\colon \mathbb{Z} \to \mathbb{Z}$ is either a ring homomorphism or $\theta(k) = 0$ for all $k \in \mathbb{Z}$.

1.4.7 ([Nic12, Ex. 3.4.4]). Determine all ring homomorphisms and general ring homomorphisms $\mathbb{Z}_{12} \to \mathbb{Z}_6$.

1.4.8 ([Nic12, Ex. 3.4.5]). If $\theta\colon R \to S$ is a surjective ring homomorphism, show that $\theta(Z(R)) \subseteq Z(S)$. Give an example of a (not necessarily surjective) ring homomorphism $\theta\colon R \to S$ where $\theta(Z(R)) \neq Z(S)$.

1.4.9 ([Nic12, Ex. 3.4.6]). If $\theta\colon R \to S$ is a ring homomorphism and $\operatorname{char} R > 0$, show that $\operatorname{char} S$ divides $\operatorname{char} R$.

1.4.10. Show that the composition of two ring homomorphisms is a ring homomorphism.

1.4.11. Complete the proof of Theorem 1.4.20 by showing that $\mathbb{Z}1_R$ is contained in the center of R.

1.4.12. If f is the map defined in the proof of Lemma 1.4.21, show that $\ker f = I \times J$ and $\operatorname{im} f = (R/I) \times (S/J)$.

1.4.13. Show that A and B, as defined in the proof of Lemma 1.4.22, are ideals of R and S respectively.

1.4.14 ([Nic12, Ex. 3.4.9]). Describe the homomorphic images of a division ring. More precisely, if R is a division ring, what possible rings can be the image of a ring homomorphism with domain R?

1.4.15 ([Nic12, Ex. 3.4.11–3.4.14]). (a) Show that $x^3 - 8x^2 + 5x + 3 = 0$ has no solution $x \in \mathbb{Z}$.

(b) Show that $m^3 + 14n^3 = 12$ has no solution in $\mathbb{Z}$.

(c) Show that $7m^2 + 11n^2 = 9$ has no solution in $\mathbb{Z}$.

(d) Show that $n^3 + (n+1)^3 + (n+2)^3 = k^2 + 1$ has no solution in $\mathbb{Z}$.

1.4.16. Prove that the inverse of a ring isomorphism is also a ring isomorphism.

1.4.17. Show that the set $\operatorname{Aut} R$ of all ring automorphisms of R is a group under the operation of composition.

1.4.18. Show that isomorphism is an equivalence relation on the class of all rings.

1.4.19 ([Nic12, Ex. 3.4.19]). Let $\theta\colon R \to S$ be an onto ring homomorphism.

(a) If A is an ideal of R, show that $\theta(A) = \{\theta(a) : a \in A\}$ is an ideal of S.

(b) If also $\ker\theta \subseteq A$, show that $R/A \cong S/\theta(A)$. *Hint:* Use the first isomorphism theorem where $\alpha\colon R \to S/\theta(A)$ is defined by $\alpha(r) = \theta(r) + \theta(A)$ for all $r \in R$.

1.4.20 ([Nic12, Ex. 3.4.20]). If $n > 0$ in $\mathbb{Z}$, describe all the ideals of $\mathbb{Z}$ that contain $n\mathbb{Z}$.

1.4.21 ([Nic12, Ex. 3.4.21]). Show that there is no ring homomorphism $\mathbb{C} \to \mathbb{R}$.

1.4.22 ([Nic12, Ex. 3.4.22]). Let $\theta\colon R \to S$ be a ring homomorphism. If $\theta(R)$ and $\ker\theta$ both contain no nonzero nilpotent elements show that the same is true of R.

1.4.23 ([Nic12, Ex. 3.4.27]). Let $R = \begin{bmatrix} S & S \\ 0 & S \end{bmatrix}$ be the 2×2 upper triangular matrix ring over a ring S. Show that $A = \begin{bmatrix} 0 & S \\ 0 & 0 \end{bmatrix}$ is an ideal of R and $R/A \cong S \times S$.

1.4.24 ([Nic12, Ex. 3.4.31]). Let $R = S \times T$, where S and T are rings, and write $\bar{S} = \{(s, 0) : s \in S\}$. Show that $\bar{S}$ is an ideal of R, and that $R/\bar{S} \cong T$ and $\bar{S} \cong S$ as rings. What is the unity of $\bar{S}$?

1.4.25. Prove the *second isomorphism theorem*: If A is an ideal of R and S is a subring of R, then

(a) $S + A$ is a subring of R,

(b) A and $S \cap A$ are ideals of $S + A$ and S, respectively, and

(c) $(S + A)/A \cong S/(S \cap A)$ as rings.

1.4.26. Prove the *third isomorphism theorem*: If $A \subseteq B \subseteq R$, where A and B are ideals of R, then $B/A = \{b + A : b \in B\}$ is an ideal of R/A and $(R/A)/(B/A) \cong R/B$ as rings.

1.4.27 ([Nic12, Ex. 3.4.37]). Show that $\mathbb{Z}_m \times \mathbb{Z}_n$ has a subring isomorphic to $\mathbb{Z}_t$, where t is the least common multiple of m and n.

1.4.28 (Bonus problem). Let $f\colon \mathbb{R} \to \mathbb{R}$ be a ring homomorphism. Prove that $f(x) = x$ for all $x \in \mathbb{R}$. In other words, the identity map is the only ring homomorphism from $\mathbb{R}$ to $\mathbb{R}$.

1.4.29 (Bonus problem). Prove that if $f\colon \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{m \text{ times}} \to \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}}$ is a ring isomorphism, then $m = n$.

Chapter 2

Polynomials

In this chapter we turn our attention to a particularly important class of rings: polynomial rings. We first introduce polynomial rings in full generality and deduce some of their fundamental properties. We then focus on polynomials whose coefficients lie in a field. We discuss the factorization of such polynomials and the structure of the quotient rings of these polynomial rings. A reference for the material in this chapter is [Jud19, Ch. 17].

2.1 Polynomial rings

Definition 2.1.1 (Indeterminate)**.** If $R \subseteq S$ are rings, then an element $x \in S$ is called an *indeterminate* over R if

$$a_0 + a_1 x + \cdots + a_n x^n = 0,\ a_i \in R \implies a_i = 0\ \forall\ i.$$

Lemma 2.1.2. *Given a ring R, there exists a ring S satisfying the following:*

(a) $R \subseteq S$.

(b) There exists $x \in S$ such that x is an indeterminate over R.

(c) We have $xa = ax$ for all $a \in R$.

Sketch of proof. The proof of this lemma is somewhat technical. We will only give an outline here. The details can be found in [Nic12, §4.6] or, under the assumption that R is commutative, in [Roy, Prop. 4.2].

Let S be the set of all sequences in R. That is, S is the set of all functions $\mathbb{N} \to R$. We often denote an element $\alpha \in S$ by $(\alpha(0), \alpha(1), \alpha(2), \dots)$. We define a ring structure on S by setting

$$(\alpha + \beta)(k) = \alpha(k) + \beta(k),$$
$$(\alpha\beta)(k) = \sum_{\ell=0}^{k} \alpha(\ell)\beta(k-\ell).$$

Then R is a subring of S if we identify $a = (a, 0, 0, \dots)$ for $a \in R$. This proves (a).

Now define $x = (0, 1, 0, 0, \dots)$. Then we can show that

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = (a_0, a_1, \dots, a_n, 0, 0, \dots)$$

for all $a_0, \dots, a_n \in R$. This proves (b). Since $ax = (0, a, 0, \dots) = xa$ for all $a \in R$, we also have (c). □

Definition 2.1.3 (Ring of polynomials)**.** Let R be a ring and let S be as in Lemma 2.1.2. Then

$$R[x] = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n : n \geq 0,\ a_i \in R\ \forall\ i\}$$

is a subring of S called the *ring of polynomials* over R. A *polynomial* over R is an element

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n, \quad n \in \mathbb{N},\ a_0, \dots, a_n \in R,$$

of $R[x]$. The a_i's are called the *coefficients* of the polynomial $f(x)$. We adopt the convention that $a_m = 0$ for $m > n$. We sometimes write f instead of $f(x)$.

Remark 2.1.4. It follows from Definition 2.1.1 that two polynomials $f(x) = a_0 + \cdots + a_k x^k$ and $g(x) = b_0 + \cdots + b_\ell x^\ell$ are *equal* if and only if $a_i = b_i$ for $i \in \mathbb{N}$. Note that this is *not* the same as equality of *polynomial functions*. For example, consider $f(x) = 0, g(x) = x^2 + x \in \mathbb{Z}_2[x]$. These polynomials are *not* equal, but the functions $\mathbb{Z}_2 \to \mathbb{Z}_2$ given by $a \mapsto f(a)$ and $a \mapsto g(a)$ *are* equal (they are both the zero function). Thus, we consider polynomials to be abstract expressions and *not* functions.

It also follows that $R[x]$ is the subring of S (from Lemma 2.1.2) consisting of sequences that are eventually zero. That is, the elements of $R[x]$ are sequences $\alpha \in S$ for which there exists $N \in \mathbb{N}$ with $\alpha(k) = 0$ for all $k > N$.

It follows from the above that if R is a ring, the addition and multiplication on $R[x]$ is computed as follows. If

$$f(x) = a_0 + \cdots + a_k x^k \quad \text{and} \quad g(x) = b_0 + \cdots + b_k x^k$$

(note that we can write $f(x)$ and $g(x)$ in this form, with the same highest power x^k of x by letting some of the coefficients be zero if necessary), then

$$f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_k + b_k)x^k, \quad \text{and}$$
$$f(x)g(x) = \sum_{i=0}^{2k} \left(\sum_{j=0}^{i} a_j b_{i-j} x^i \right) = a_0 b_0 + (a_0 b_1 + a_1 b_0)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \cdots .$$

Example 2.1.5. We have

$$1 - 3x + x^2 \in \mathbb{Z}[x], \quad 2x - 3 \in \mathbb{Z}[x], \quad \text{but} \quad 1 - \frac{1}{x} \notin \mathbb{Z}[x].$$

Also

$$\begin{aligned}(1 - 3x + x^2)(2x - 3) &= (1 - 3x + x^2)(-3 + 2x) \\ &= -3 + (2 + 9)x + (-6 - 3)x^2 + 2x^3 = -3 + 11x - 9x^2 + 2x^3.\end{aligned}$$

Example 2.1.6. We have

$$\bar{1} - \bar{2}x,\ \bar{1} + \bar{2}x^2 \in \mathbb{Z}_4[x]$$

and

$$(\bar{1} - \bar{2}x) + (\bar{1} + \bar{2}x^2) = (\bar{1} + \bar{1}) - \bar{2}x - \bar{2}x^2 = \bar{2} - \bar{2}x + \bar{2}x^2 = \bar{2} + \bar{2}x + \bar{2}x^2,$$
$$\bar{2}(\bar{1} + \bar{2}x^2) = \bar{2} + \bar{0}x^2 = \bar{2},$$
$$(\bar{1} - \bar{2}x)(\bar{1} + \bar{2}x^2) = \bar{1} + (-\bar{2})x + \bar{2}x^2 - \bar{0}x^3 = \bar{1} + \bar{2}x + \bar{2}x^2.$$

Definition 2.1.7 (Degree, leading coefficient, constant coefficient, monic polynomial)**.** Let $f(x) = a_0 + a_1x + \cdots + a_kx^k \in R[x]$.

- If $a_k \neq 0$, then the *degree* of $f(x)$ is $\deg f(x) = k$ and a_k is the *leading coefficient* of $f(x)$.
- a_0 is called the *constant coefficient* of $f(x)$.
- If $a_k = 1$, then $f(x)$ is called a *monic polynomial*.

Theorem 2.1.8. *Let R be a ring. Then*

(a) the center of $R[x]$ is $Z(R)[x]$,

(b) $x \in Z(R[x])$, and

(c) if R is commutative, then $R[x]$ is also commutative.

Proof. We leave the proof of this theorem as Exercise 2.1.4 □

The zero element of $R[x]$ is the *zero polynomial* $0 = 0 + 0x = 0 + 0x + 0x^2 = \cdots$. A polynomial is called a *constant polynomial* if it is of the form $f(x) = a_0 = a_0 + 0x = a_0 + 0x + 0x^2 = \cdots$ for some $a_0 \in R$. Note that R is the subring of $R[x]$ consisting of constant polynomials.

Theorem 2.1.9. *Let R be a ring. If $f(x), g(x) \in R[x] \setminus \{0\}$, then we have the following.*

(a) $\deg(f(x) + g(x)) \leq \max\{\deg f(x), \deg g(x)\}$.

(b) $\deg(f(x)g(x)) \leq \deg f(x) + \deg g(x)$.

(c) If R is a domain, then $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$.

(d) If R is a domain, then $R[x]$ is also a domain.

(e) If R is a domain, then the only units of $R[x]$ are the units of R.

(f) If the leading coefficient of either $f(x)$ or $g(x)$ is a unit in R, then $f(x)g(x) \neq 0$ in $R[x]$ and $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$.

Proof. The proof of this theorem is left as Exercise 2.1.5. □

Example 2.1.10. To see that the inequalities of Theorem 2.1.9 can be strict when R is not a domain, consider $\bar{1} + \bar{2}x, \bar{2}x \in \mathbb{Z}_4[x]$. Then $(\bar{1} + \bar{2}x)(\bar{2}x) = \bar{2}x$ and so

$$\deg(\bar{1} + \bar{2}x)(\bar{2}x) = 1 < 2 = \deg(\bar{1} + \bar{2}x) + \deg(\bar{2}x).$$

Also,

$$\deg((\bar{1} + \bar{2}x) + (\bar{2}x)) = \deg(\bar{1}) = 0 < 1 = \max\{\deg(\bar{1} + \bar{2}x), \deg(\bar{2}x)\}.$$

Example 2.1.11. Let $I \subseteq \mathbb{Z}[x]$ be the ideal generated by x, i.e., $I = \langle x \rangle = x\mathbb{Z}[x]$. So

$$I = \{xf(x) : f(x) \in \mathbb{Z}[x]\} = \{a_1x + \cdots + a_kx^k : k \geq 1,\ a_1, \ldots, a_k \in \mathbb{Z}\}.$$

Consider the ring homomorphism $\varphi\colon \mathbb{Z}[x] \to \mathbb{Z}$ given by $\varphi(a_0 + \ldots + a_kx^k) = a_0$. It is a straightforward exercise (Exercise 2.1.6) to see that φ is a surjective ring homomorphism with $\ker\varphi = I$. Thus, by the first isomorphism theorem (Theorem 1.4.15), we have $\mathbb{Z}[x]/I \cong \mathbb{Z}$.

Theorem 2.1.12 (The Division Algorithm)**.** *Let R be a ring, $f(x), g(x) \in R[x]$, $f(x) \neq 0$, and suppose the leading coefficient of $f(x)$ is a unit in R. Then there exist unique polynomials $q(x), r(x) \in R[x]$ such that*

(a) $g(x) = q(x)f(x) + r(x)$ and

(b) either $r(x) = 0$ or $\deg r(x) < \deg f(x)$.

Proof. We first prove the existence result by induction on $\deg g(x)$. If $g(x) = 0$ or $\deg g(x) < \deg f(x)$, then we can take $q(x) = 0$ and $r(x) = g(x)$. So assume $m \geq n$, where $m = \deg g(x)$ and $n = \deg f(x)$. Write

$$f(x) = ux^n + ax^{n-1} + \cdots, \qquad g(x) = bx^m + cx^{m-1} + \cdots,$$

where u is a unit by hypothesis. Let

$$\begin{aligned} g_1(x) &= g(x) - bu^{-1}x^{m-n}f(x) \\ &= (bx^m + cx^{m-1} + \cdots) - bu^{-1}x^{m-n}(ux^n + ax^{n-1} + \cdots) \\ &= 0x^m + (c - bu^{-1}a)x^{m-1} + \cdots. \end{aligned}$$

Then either $g_1(x) = 0$ or $\deg g_1(x) < m$. Thus, by the inductive hypothesis, there exist polynomials $q_1(x)$ and $r(x)$ such that $g_1(x) = q_1(x)f(x) + r(x)$ with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$. Then

$$g(x) = g_1(x) + bu^{-1}x^{m-n}f(x) = (q_1(x) + bu^{-1}x^{m-n})f(x) + r(x),$$

which completes the proof of the inductive step.

It remains to prove uniqueness. Suppose that we have

$$q_1(x)f(x) + r_1(x) = g(x) = q_2(x)f(x) + r_2(x)$$

with $r_i(x) = 0$ or $\deg r_i(x) < \deg f(x)$ for $i = 1, 2$. Then $r_1(x) - r_2(x) = (q_2(x) - q_1(x))f(x)$. If $q_2(x) - q_1(x) \neq 0$, then, since the leading coefficient of $f(x)$ is a unit (see Theorem 2.1.9(f)), we have $(q_2(x) - q_1(x))f(x) \neq 0$ and

$$\deg(r_1(x) - r_2(x)) = \deg[(q_2(x) - q_1(x))f(x)] = \deg(q_2(x) - q_1(x)) + \deg f(x).$$

But this implies that $\deg(r_1(x) - r_2(x)) \geq \deg f(x)$, which is a contradiction. Thus $q_2(x) - q_1(x) = 0$ and so $r_1(x) - r_2(x) = (q_2(x) - q_1(x))f(x) = 0$. This completes the proof of uniqueness. □

Remark 2.1.13. Note that if R is in fact a field, then the leading coefficient of *any* nonzero polynomial is a unit in R. Then Theorem 2.1.12 says that $R[x]$ is a *euclidean domain* with *euclidean function* $f(x) \mapsto \deg f(x)$.

Example 2.1.14. Let's divide $x^3 - x^2 + 2x - 3$ by $x - 2$.

$$\begin{array}{rr|l}
 & x^2 + x + 4 & \\
\hline
x-2 & x^3 - x^2 + 2x - 3 & \\
 & x^3 - 2x^2 & \\
\hline
 & x^2 + 2x - 3 & \\
 & x^2 - 2x & \\
\hline
 & 4x - 3 & \\
 & 4x - 8 & \\
\hline
 & 5 &
\end{array}$$

Hence,

$$\underbrace{x^3 - x^2 + 2x - 3}_{g(x)} = \underbrace{(x^2 + x + 4)}_{q(x)}\underbrace{(x - 2)}_{f(x)} + \underbrace{5}_{r(x)}.$$

Example 2.1.15. Divide $x^3 + \bar{2}$ by $\bar{2}x + \bar{2}$ in $\mathbb{Z}_3[x]$.

$$\begin{array}{rr|l}
 & \bar{2}x^2 + x + \bar{2} & \\
\hline
\bar{2}x+\bar{2} & x^3 + \bar{0}x^2 + \bar{0}x + \bar{2} & \\
 & x^3 + x^2 & \\
\hline
 & -x^2 + \bar{2} & \\
 & \bar{2}x^2 + \bar{2}x & \\
\hline
 & -\bar{2}x + \bar{2} & \\
 & x + \bar{1} & \\
\hline
 & \bar{1} &
\end{array}$$

Thus

$$\underbrace{x^3 + \bar{2}}_{g(x)} = \underbrace{\bar{2}x^2 + x + \bar{2}}_{q(x)}\underbrace{(\bar{2}x + \bar{2})}_{f(x)} + \underbrace{\bar{1}}_{r(x)}.$$

Can we always evaluate polynomials by specializing the indeterminate? We are used to doing this in calculus, but in fact we must be careful!

Example 2.1.16. Consider the polynomial $f(x) = x^3 - 2x - 4 \in \mathbb{Z}[x]$. We have $f(2) = 0$, which implies that $x - 2$ is a factor of $f(x)$. In fact, we have $f(x) = (x^2 + 2x + 2)(x - 2)$.

Example 2.1.17. Suppose R is a ring and $a, b \in R$ with $ab \neq ba$. Let $f(x) = (x - a)(x - b) \in R[x]$. Then $f(a) = (a - a)(b - a) = 0(b - a) = 0$. But we also have $f(x) = x^2 - ax - bx + ab$ and so $f(a) = a^2 - a^2 - ba + ab = -ba + ab \neq 0$. So "evaluation" of $f(x)$ at a does not seem to be well-defined. The problem is that x lies in the center of $R[x]$ and we are trying to construct a (surjective) ring homomorphism that maps x to something that is *not* in the center of R.

Theorem 2.1.18 (Evaluation Theorem)**.** *Let R be a ring and $a \in Z(R)$. Then the map $\varphi_a \colon R[x] \to R$, $\varphi_a(f(x)) = f(a)$ is a surjective ring homomorphism.*

Proof. Recall that we can view elements of R as constant polynomials in $R[x]$. For any $r \in R$ we have $\varphi_a(r) = r$. Hence φ_a is surjective.

Let $f(x) = \sum_{i=0}^{k} a_i x^i$ and $g(x) = \sum_{i=0}^{k} b_i x^i$ be arbitrary elements of $R[x]$. Then we have

$$\begin{aligned}\varphi_a(f(x)+g(x)) = \varphi_a\left(\sum (a_i+b_i)x^i\right) &= \sum (a_i+b_i)a^i \\ &= \sum a_i a^i + \sum b_i a^i = \varphi_a(f(x)) + \varphi_a(g(x),\end{aligned}$$

$$\begin{aligned}\varphi_a(f(x)g(x)) = \varphi_a\left(\sum_{i=0}^{2k}\left(\sum_{j=0}^{i} a_j b_{i-j} x^i\right)\right) &= \sum_{i=1}^{2k}\left(\sum_{j=0}^{i} a_0 b_{i-j} a^i\right) \\ &= \left(\sum_{i=0}^{k} a_i a^i\right)\left(\sum_{i=0}^{k} b_i a^i\right) = \varphi_a(f(x))\varphi_a(g(x)),\end{aligned}$$

$$\varphi_a(1_{R[x]}) = \varphi(1_R) = 1_R. \qquad \square$$

The map φ_a of Theorem 2.1.18 is called *evaluation* at a. The problem in Example 2.1.17 was that we tried to evaluate a polynomial at an element that was *not* in the center of the coefficient ring R.

Example 2.1.19. Let R be a ring. Then

$$\varphi\colon R[x] \to R, \quad \varphi(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = a_0 + a_1 + a_2 + \cdots + a_n,$$

is a surjective ring homomorphism since $\varphi = \varphi_1$.

Theorem 2.1.20 (Remainder Theorem)**.** *Let R be a commutative ring and $f(x) \in R[x]$. Then, for every $a \in R$, the remainder of the division of $f(x)$ by $x - a$ is $f(a)$.*

Proof. By Theorem 2.1.12, we can write $f(x) = q(x)(x-a) + r(x)$ where $r(x) = 0$ or $\deg r(x) < 1$. Thus we have that $r(x)$ is a constant polynomial. That is, $r(x) = r$ for some $r \in R$. Then

$$f(a) = \varphi_a(f(x)) = \varphi_a(q(x))\varphi_a(x-a) + \varphi_a(r(x)) = 0 + r(a) = r. \qquad \square$$

Theorem 2.1.21 (Factor Theorem)**.** *Let R be a commutative ring and $f(x) \in R[x]$. Then $a \in R$ satisfies $f(a) = 0$ if and only if $f(x) = (x-a)q(x)$ for some $q(x) \in R[x]$.*

Proof. Clearly, $f(a) = 0$ if $f(x) = (x-a)g(x)$ for some $g(x) \in R[x]$. The converse follows immediately from the Remainder Theorem (Theorem 2.1.20). $\square$

Corollary 2.1.22. *Let R be a commutative ring and $\varphi_a\colon R[x] \to R$, $\varphi_a(f(x)) = f(a)$. Then $\ker \varphi_a = \langle x-a \rangle = (x-a)R[x]$ and $R[x]/\langle x-a\rangle \cong R$.*

Proof. We have

$$\begin{aligned} f(x) \in \ker \varphi_a &\iff f(a) = 0 \\ &\iff f(x) = q(x)(x-a) + 0 \text{ for some } q(x) \in R[x] \\ &\iff f(x) \in \langle x-a \rangle. \end{aligned}$$

The second assertion then follows from the first isomorphism theorem (Theorem 1.4.15). $\square$

Definition 2.1.23 (Root of a polynomial)**.** Let R be a commutative ring and $f(x) \in R[x] \setminus \{0\}$. An element $a \in R$ is called a *root* of $f(x)$ if $f(a) = 0$.

Example 2.1.24. The roots of $x^2 - \bar{1} \in \mathbb{Z}_8[x]$ are $\bar{1}, \bar{3}, \bar{5}, \bar{7}$.

Definition 2.1.25 (Multiplicity of a root)**.** Let R be a commutative ring and $f(x) \in R[x] \setminus \{0\}$. If $a \in R$ is a root of $f(x)$, we say a has *multiplicity* $m \geq 0$ if $f(x) = (x-a)^m g(x)$ with $g(a) \neq 0$.

Example 2.1.26. Let's find the multiplicity of the root $\bar{3}$ for $f(x) = x^3 + \bar{3}x + \bar{4} \in \mathbb{Z}_5[x]$. We have $f(\bar{3}) = \bar{0}$. Dividing $f(x)$ by $x - \bar{3}$ gives $f(x) = (x-\bar{3})g(x)$ with $g(x) = x^2 + \bar{3}x + \bar{2}$. Since $g(\bar{3}) = \bar{0}$, we divide again to obtain $g(x) = (x-\bar{3})(x+\bar{1})$. Thus we have $f(x) = (x-\bar{3})^2(x+\bar{1})$. Since $\bar{3}$ is not a root of $x + \bar{1}$, we see that the multiplicity of the root $\bar{3}$ is 2.

Theorem 2.1.27. *Let R be an integral domain and $f(x) \in R[x] \setminus \{0\}$. If $n = \deg f(x)$, then $f(x)$ has at most n roots.*

Proof. We prove the result by induction on n, the cases $n = 0, 1$ being straightforward. Suppose $\deg f(x) = n+1$. Then, if $f(a) = 0$, we have $f(x) = (x-a)g(x)$ with $\deg g(x) = n$. If f has more than $n+1$ roots, then we have $f(a_1) = f(a_2) = \cdots = f(a_{n+1}) = 0$ for some distinct $a_1, \ldots, a_{n+1}$, none of which is equal to a. Then, for $i = 1, \ldots, n+1$, we have $0 = f(a_i) = (a_i - a)g(a_i)$ and $a_i - a \neq 0$, which implies that $g(a_i) = 0$. Thus $g(x)$ has at least $n+1$ roots, which contradicts the inductive hypothesis since $\deg g(x) = n$. □

Theorem 2.1.28 (Rational Roots Theorem)**.** *Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[x]$ with $a_0 \neq 0$, $a_n \neq 0$. If $r \in \mathbb{Q}$ and $f(r) = 0$, then $r = \frac{c}{d}$ with $c | a_0$ and $d | a_n$.*

Proof. Write $r = \frac{c}{d}$ with $\gcd(c, d) = 1$. Then

$$0 = f(r) = a_0 + a_1\left(\frac{c}{d}\right) + \cdots + a_n\left(\frac{c}{d}\right)^n \implies a_0 d^n + a_1 d^{n-1} c + \cdots + a_n c^n = 0.$$

Thus

$$a_0 d^n = -c(a_1 d^{n-1} + a_2 d^{n-2} c + \cdots a_n c^{n-1}) \quad \text{and}$$
$$a_n c^n = -d(a_0 d^{n-1} + a_1 d^{n-2} c + \cdots + a_n c^{n-1}).$$

Since c and d are relatively prime, this implies that $c | a_0$ and $d | a_n$. □

Example 2.1.29. Although we already know that $\sqrt{2}$ is irrational, we can give another proof based on the Rational Roots Theorem (Theorem 2.1.28). Consider the polynomial $x^2 - 2$. If it had a rational root, it must be one of $\pm 2, \pm 1$. But one easily checks that none of these is a root of $x^2 - 2$. Thus, $x^2 - 2$ has no rational roots. Since $\sqrt{2}$ is a root of $x^2 - 2$, it is not rational.

Example 2.1.30. If $f(x) \in \mathbb{Z}[x] \setminus \{0\}$ is monic, then any rational root of $f(x)$ is an integer.

Exercises.

2.1.1. Compute $(f+g)(x)$ and $(fg)(x)$, where

$$f(x) = 2 + 3x + x^2,\ g(x) = 1 + x^2 + x^3 + 4x^3 \in \mathbb{Z}_5[x].$$

2.1.2 ([Nic12, Ex. 4.1.2]). (a) Compute $(1+x)^5$ in $\mathbb{Z}_5[x]$.

(b) Compute $(1+x)^7$ in $\mathbb{Z}_7[x]$.

(c) Show that $(1+x)^p = 1 + x^p$ in $\mathbb{Z}_p[x]$ if p is a prime number.

2.1.3 ([Nic12, Ex. 4.1.3]). (a) How many polynomials of degree 3 are there in $\mathbb{Z}_5[x]$?

(b) How many monic polynomials of degree 3 are there in $\mathbb{Z}_5[x]$?

2.1.4. Prove Theorem 2.1.8.

2.1.5. Prove Theorem 2.1.9.

2.1.6. Show that the map φ of Example 2.1.11 is a surjective ring homomorphism with $\ker\varphi = I$.

2.1.7. Give an example of a commutative ring R and a polynomial $f(x) \in R[x]\setminus\{0\}$ of degree n with more than n distinct roots.

2.1.8. Use the Rational Roots Theorem (Theorem 2.1.28) to show that $\sqrt[3]{5}$ is irrational.

2.1.9 ([Nic12, Ex. 4.1.4]). (a) Find all roots of $(x-4)(x-5)$ in $\mathbb{Z}_6$ and in $\mathbb{Z}_7$.

(b) Find all roots of $x^3 - x$ in $\mathbb{Z}_6$ and in $\mathbb{Z}_4$.

2.1.10 ([Nic12, Ex. 4.1.5(a)]). Find the number of roots of $x^2 - x$ in $\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$, any integral domain, and $\mathbb{Z}_6$.

2.1.11 ([Nic12, Ex. 4.1.8]). Let R be a subring of a ring S, let $f(x) \neq 0$ and $g(x)$ be polynomials in $R[x]$, and assume that the leading coefficient of $f(x)$ is a unit in R. If $f(x)$ divides $g(x)$ in $S[x]$, show that $f(x)$ divides $g(x)$ in $R[x]$. *Hint:* The Division Algorithm.

2.1.12 ([Nic12, Ex. 4.1.9]). Show that $R[x]$ and R have the same characteristic for any ring R.

2.1.13 ([Nic12, Ex. 4.1.11]). If a is a nonzero root of $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$, show that a^{-1} (if it exists) is a root of $g(x) = a_n + a_{n-1}x + \cdots + a_1x^{n-1} + a_0x^n$. Assume that the coefficient ring R is commutative.

2.1.14 ([Nic12, Ex. 4.1.12]). If a, b, and c are real and $f(x) = x^2 - (a+c)x + (ac - b^2)$, show that every complex root of f is real.

2.1.15 ([Nic12, Ex. 4.1.13]). Divide $x^3 - 4x + 5$ by $2x + 1$ in $\mathbb{Q}[x]$. Why is it impossible in $\mathbb{Z}[x]$?

2.1.16 ([Nic12, Ex. 4.1.14]). In each case, write $g(x) = q(x)f(x)+r(x)$ in $R[x]$, where $r(x) = 0$ or $\deg r(x) < \deg f(x)$.

(a) $g(x) = x^5 + 4x^4 + x^3 + 5x^2 + x + 2$, $f = x^2 + x + 1$, $R = \mathbb{Z}_6$.

(b) $g(x) = x^3 + x^2 + 3x + 2$, $f = 3x + 1$, $R = \mathbb{Z}_8$.

(c) $g(x) = 3x^3 + 2x^2 - 8x + 1$, $f = x^2 + 2$, $R = \mathbb{Q}$.

2.1.17 ([Nic12, Ex. 4.1.15]). Which of $x-1$, $x+1$, and $x-2$ is a factor of $x^4-2x^3-x^2+3x-2$ in $\mathbb{Z}[x]$?

2.1.18 ([Nic12, Ex. 4.1.16(a)]). For which primes p is $x - 1$ a factor of $f(x) = 3x^4 + 5x^3 + 2x^2 + x + 4$ in $\mathbb{Z}_p[x]$?

2.1.19 ([Nic12, Ex. 4.1.17]). In each case, factor $f(x)$ into linear factors in $F[x]$.

(a) $f(x) = x^4 + 12$, $F = \mathbb{Z}_{13}$.

(b) $f(x) = x^3 - x^2 + x - 1$, $F = \mathbb{Z}_5$.

2.1.20 ([Nic12, Ex. 4.1.18]). Let $a \neq 0$ in a field F. Determine the integers $n \geq 1$ such that $x + a$ is a factor of $x^n + a^n$ in $F[x]$. In these cases, write down the factorization.

2.1.21 ([Nic12, Ex. 4.1.19]). If F is a field, let u, v, and w be distinct roots of $f(x) = x^3 + ax^2 + bx + c$ in $F[x]$. Show that $a = -(u + v + w)$, $b = uv + uw + vw$, and $c = -uvw$.

2.1.22 ([Nic12, Ex. 4.1.21]). (a) Show that $\mathbb{Z}_4[x]$ has infinitely many units and infinitely many nilpotents.

(b) Find a polynomial in $\mathbb{Z}_4[x]$ that is neither a unit nor a nilpotent.

2.1.23 ([Nic12, Ex. 4.1.23]). In each case determine the multiplicity of a as a root in $f(x) \in R[x]$.

(a) $f(x) = x^3 - 2x^2 - 4x + 3$, $a = 3$, $R = \mathbb{Z}_6$.

(b) $f = x^5 + 2x^4 + x^3 - x^2 + 2x - 1$, $a = 1$, $R = \mathbb{Z}_4$.

2.1.24 ([Nic12, Ex. 4.1.25]). In each case find all rational roots of $f(x)$ and factor $f(x)$ as much as possible in $\mathbb{Q}[x]$.

(a) $f(x) = 4x^4 + x^3 - 3x^2 + 4x - 3$

(b) $f(x) = x^4 - x^3 - x^2 - x - 2$

(c) $f(x) = x^4 + x^3 + 3x^2 + 2x + 2$

2.1.25 ([Nic12, Ex. 4.1.26]). Show that $\sqrt[n]{m}$ is not rational unless $m = k^n$ for some integer k.

2.1.26 ([Nic12, Ex. 4.1.27]). If f is a monic polynomial in $\mathbb{Z}[x]$, show that the only rational roots (if any) are integers.

2.1.27. (a) Suppose R is an infinite integral domain. Show that if $f(x), g(x) \in R[x]$ satisfy $f(a) = g(a)$ for all $a \in R$, then $f(x) = g(x)$. In other words, two polynomials are equal if and only if they correspond to the same polynomial function.

(b) Suppose R is a finite commutative ring. Show that there exists a polynomial $f(x) \in R[x] \setminus \{0\}$ such that $f(a) = 0$ for all $a \in R$.

2.1.28. Let G be the multiplicative group of positive rational numbers. Show that G is isomorphic as a group to $(\mathbb{Z}[x], +)$.

2.1.29 (Bonus). Let R be a commutative ring. Prove that $f(x) \in R[x]$, $f(x) = a_0 + \cdots + a_n x^n$ is a nilpotent element of $R[x]$ if and only if $a_0, \ldots, a_n$ are nilpotent elements of R.

2.2 Factorization of polynomials over a field

We start this section with a few examples.

Example 2.2.1. Let's factor $f(x) = 3x^3 - x^2 - x - 4 \in \mathbb{Q}[x]$ as much as possible. By the Rational Roots Theorem (Theorem 2.1.28), the only possible rational roots are of the form $\frac{c}{d}$ where $c|4$ and $d|3$. We see that $x = \frac{4}{3}$ is a root. Polynomial division then gives that $f(x) = \left(x - \frac{4}{3}\right)(3x^2 + 3x + 3) = (3x - 4)(x^2 + x + 1)$. Again, using the Rational Roots Theorem, we can see that $x^2 + x + 1$ has no rational roots and so we can factor no further.

Example 2.2.2. Let's factor $f(x) = x^4 + 25x + 24 \in \mathbb{Q}[x]$ as much as possible. Since $f(-1) = 0$, we know that $x + 1$ is a factor of $f(x)$. Polynomial division then gives $f(x) = (x + 1)(x^3 - x^2 + x + 24)$. By the Rational Roots Theorem (Theorem 2.1.28), any rational roots of $x^3 - x^2 + x + 24$ must be in the set $\{\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 8, \pm 12, \pm 24\}$. But none of these is a root and therefore $f(x)$ cannot be factored any further (since if it factored, at least one factor would need to be of degree one).

Example 2.2.3. Let's factor $f(x) = 2x^5 + x^4 + 4x^3 + 2x^2 + 2x + 1 \in \mathbb{Q}[x]$ as much as possible. Since $f(-1/2) = 0$, we get $f(x) = (2x + 1)(x^4 + 2x^2 + 1) = (2x + 1)(x^2 + 1)^2$. Since $x^2 + 1$ has no rational (or even real) roots, we cannot factor $f(x)$ any further (in $\mathbb{Q}[x]$).

Definition 2.2.4 (Irreducible polynomial)**.** Let F be a field. A polynomial $f(x) \in F[x]$ is called *irreducible* over F if

(a) $f(x) \neq 0$ and $\deg f(x) \geq 1$ (i.e. $f(x)$ is a nonconstant polynomial), and

(b) if $f(x) = p(x)q(x)$ in $F[x]$, then either $\deg p(x) = 0$ or $\deg q(x) = 0$.

A nonzero polynomial of positive degree is called *reducible* (or we say that it *factors*) if it is not irreducible.

Example 2.2.5. If $\deg f(x) = 1$ then $f(x)$ is irreducible.

Example 2.2.6. If $f(x)$ is irreducible, then, for every $a \neq 0$, the polynomial $af(x)$ is also irreducible.

Theorem 2.2.7. *Let F be a field and consider $f(x) \in F[x] \setminus \{0\}$ such that $\deg f(x) \geq 2$.*

(a) If $f(x)$ is irreducible, then it does not have any roots in F.

(b) Assume $\deg f(x) \leq 3$. Then $f(x)$ is irreducible if and only if $f(x)$ does not have any roots in F.

Proof. (a) If $f(x)$ has a root $a \in F$, then $f(x) = (x-a)q(x)$ for some $q(x) \in F[x]$ by the Factor Theorem (Theorem 2.1.21). Since $\deg f(x) \geq 2$, this means that $f(x)$ is not irreducible, contradicting the hypothesis. Thus $f(x)$ has no root in F.

(b) Assume $f(x)$ has no root in F. If $f(x)$ is reducible, then $f(x) = p(x)q(x)$ and so $p(x)$ and $q(x)$ have no roots in F. Thus $\deg p(x) \neq 1$ and $\deg q(x) \neq 1$. But this contradicts the fact that $\deg p(x) + \deg q(x) = \deg f(x)$ is equal to 2 or 3. Thus $f(x)$ is irreducible. The converse follows from part (a). □

Example 2.2.8. The polynomial $x^2 - 2$ is irreducible over $\mathbb{Q}$ but reducible over $\mathbb{R}$ since it has roots in $\mathbb{R}$ but not in $\mathbb{Q}$.

Example 2.2.9. The polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ but reducible in $\mathbb{C}[x]$.

Example 2.2.10. The polynomial $x^3 + 2x^2 + x + 2$ is irreducible in $\mathbb{Z}_5[x]$ because it has no roots.

Example 2.2.11. The polynomial $x^2 - 2$ is reducible in $\mathbb{R}[x]$ but irreducible in $\mathbb{Z}_3[x]$. The polynomial $x^2 + 2$ is irreducible in $\mathbb{R}[x]$ but reducible in $\mathbb{Z}_3[x]$ (in particular, $x^2 + 2 = (x-1)(x+1)$ in $\mathbb{Z}_3[x]$).

Theorem 2.2.12 (Fundamental Theorem of Algebra)**.** *If $f(x) \in \mathbb{C}[x]$ is a nonconstant polynomial, then $f(x)$ has a root in $\mathbb{C}$.*

We will accept the Fundamental Theorem of Algebra without proof. For those interested in reading the proof, it can be found in [Jud19, Th. 23.33].

Corollary 2.2.13. *Suppose $f(x) \in \mathbb{C}[x]$ with $\deg f(x) \geq 1$. If we write $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, $a_n \neq 0$, then $f(x)$ factors as $f(x) = a_n(x-z_1)(x-z_2)\cdots(x-z_n)$, where $z_1, \ldots, z_n$ are the roots of $f(x)$, counted according to multiplicity. In particular, the only irreducible polynomials in $\mathbb{C}[x]$ are linear.*

Proof. The proof is by induction on the degree of $f(x)$, using the Fundamental Theorem of Algebra (Theorem 2.2.12). The details are left as Exercise 2.2.1. □

Theorem 2.2.14. *If $f(x) \in \mathbb{R}[x]$ is a nonconstant polynomial, then it factors (in $\mathbb{R}[x]$) as a product of polynomials of degree at most two. In particular, the irreducible polynomials in $\mathbb{R}[x]$ are either linear or quadratic.*

Proof. The proof is by induction on $\deg f(x)$. The case $\deg f(x) \leq 2$ is immediate. Assume the result holds for polynomials of degree $\leq n$ and consider $f(x)$ with $\deg f(x) = n+1$. There are two cases.

(a) If $f(z) = 0$ for some $z \in \mathbb{R}$, then $f(x) = (x-z)g(x)$ and $\deg g(x) = n$. The result then follows by the inductive hypothesis applied to $g(x)$.

(b) Otherwise, $f(x)$ has no real roots. By the Fundamental Theorem of Algebra (Theorem 2.2.12), it has a complex root z. Then $f(z) = 0$ and so $f(\bar{z}) = \overline{f(z)} = \bar{0} = 0$. (Here $\bar{z}$ denotes the complex conjugate of z.) Therefore,

$$f(x) = (x-z)(x-\bar{z})g(x) = \underbrace{(x^2 - (z+\bar{z})x + z\bar{z})}_{\in \mathbb{R}[x]} g(x),$$

and the result again follows by applying the inductive hypothesis to $g(x)$. □

If $\alpha\colon R \to S$ is a ring homomorphism, then we have an induced ring homomorphism

$$R[x] \to S[x], \quad \sum a_i x^i \mapsto \sum \alpha(a_i)x^i, \quad a_i \in \mathbb{R}.$$

In particular, if p is a prime number, taking α to be usual ring homomorphism $\mathbb{Z} \to \mathbb{Z}_p$, $\alpha(x) = \bar{x}$, gives a ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}_p[x]$ called *reduction modulo p*.

Theorem 2.2.15 (Gauss' Lemma)**.** *Let $f(x) = g(x)h(x)$ in $\mathbb{Z}[x]$. If a prime $p \in \mathbb{Z}$ divides every coefficient of $f(x)$, then either p divides every coefficient of $g(x)$ or p divides every coefficient of $h(x)$.*

Proof. Consider the reduction modulo p, $\varphi\colon \mathbb{Z}[x] \to \mathbb{Z}_p[x]$, $\varphi(\sum a_i x^i) = \sum \bar{a}_i x^i$. We let $\bar{f}(x) = \varphi(f(x))$. If a prime number p divides every coefficient of $f(x)$, then $\bar{f}(x) = 0$ in $\mathbb{Z}_p[x]$. Thus $0 = \bar{g}(x)\bar{h}(x)$. Since $\mathbb{Z}_p$ is a field, $\mathbb{Z}_p[x]$ is an integral domain by Theorem 2.1.9(d). Thus $\bar{g}(x) = 0$ or $\bar{h}(x) = 0$. But this implies that either every coefficient of $g(x)$ is zero in $\mathbb{Z}_p$ or every coefficient of $h(x)$ is zero in $\mathbb{Z}_p$. □

Definition 2.2.16 (Proper factorization)**.** Let $f(x) \in \mathbb{Z}[x]$ be a nonconstant polynomial. A *proper factorization* of $f(x)$ is a way of writing $f(x) = g(x)h(x)$, where $g(x), h(x) \in \mathbb{Z}[x]$ with $\deg g(x), \deg h(x) > 0$.

Theorem 2.2.17. *Suppose $f(x)$ is a nonconstant polynomial in $\mathbb{Z}[x]$.*

(a) If $f(x) = g(x)h(x)$ with $g(x), h(x) \in \mathbb{Q}[x]$, then $f(x) = g_0(x)h_0(x)$ for some $g_0(x), h_0(x) \in \mathbb{Z}[x]$ with $\deg g_0(x) = \deg g(x)$ and $\deg h_0(x) = \deg h(x)$.

(b) The polynomial $f(x)$ is irreducible in $\mathbb{Q}[x]$ if and only if it has no proper factorization in $\mathbb{Z}[x]$.

Proof. (a) Let a and b be the least common multiples of the denominators of the coefficients of $g(x)$ and $h(x)$, respectively. Then $g_1(x) := ag(x) \in \mathbb{Z}[x]$ and $h_1(x) := bh(x) \in \mathbb{Z}[x]$. Thus, we have

$$abf(x) = g_1(x)h_1(x)$$

in $\mathbb{Z}[x]$. Suppose p is a prime number dividing ab. Then, by Gauss' Lemma (Theorem 2.2.15), either p divides all the coefficients of $g_1(x)$ or it divides all the coefficients of $h_1(x)$. Thus we can cancel p to give

$$\frac{ab}{p} f(x) = g_2(x)h_2(x)$$

in $\mathbb{Z}[x]$. Continuing in this manner, we can cancel every prime factor of ab to obtain a factorization

$$f(x) = g_k(x)h_k(x)$$

in $\mathbb{Z}[x]$. The result then follows since $\deg g(x) = \deg g_1(x) = \cdots = \deg g_k(x)$ and similarly $\deg h(x) = \deg h_k(x)$.

(b) If $f(x)$ is irreducible in $\mathbb{Q}[x]$, then it has no proper factorization in $\mathbb{Z}[x]$, since such a factorization would also be a factorization in $\mathbb{Q}[x]$. The converse follows from part (a). □

Theorem 2.2.18 (Modular Irreducibility Test)**.** *Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$, $n \geq 1$, $a_n \neq 0$. Suppose that there exists a prime number p such that*

(a) $p \nmid a_n$ and

(b) the reduction of $f(x)$ modulo p is irreducible in $\mathbb{Z}_p[x]$.

Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof. The first condition tells us that $\deg \bar{f}(x) = \deg f(x)$. Suppose $f(x)$ is reducible in $\mathbb{Q}[x]$. Then there is a proper factorization $f(x) = g(x)h(x)$ in $\mathbb{Z}[x]$ by Theorem 2.2.17. Then

$$\deg \bar{g}(x) \leq \deg g(x) < \deg f(x) = \deg \bar{f}(x).$$

Similarly, $\deg \bar{h}(x) < \deg \bar{f}(x)$. Since $\bar{f}(x) = \bar{g}(x)\bar{h}(x)$, this contradicts the irreducibility of $\bar{f}(x)$ in $\mathbb{Z}_p[x]$. □

Example 2.2.19. We show that $f(x) = 4x^4 - 3x^2 - 2x - 1$ is irreducible in $\mathbb{Q}[x]$. We apply Theorem 2.2.18 with $p = 3$. Suppose $\bar{f}(x) = x^4 + x + \bar{2} = \bar{g}(x)\bar{h}(x)$ in $\mathbb{Z}_3[x]$. Then $\deg \bar{g}(x) + \deg \bar{h}(x) = 4$. Assume, without loss of generality, that $\deg \bar{g}(x) \leq \deg \bar{h}(x)$. Then $\deg \bar{g}(x) \leq 2$.

It is not possible to have $\deg \bar{g}(x) = 1$, since $x^4 + x + \bar{2}$ does not have a root in $\mathbb{Z}_3$. So it remains to consider the case that $x^4 + x + \bar{2}$ factors as a product of two quadratics in $\mathbb{Z}_3[x]$. So assume that in $\mathbb{Z}_3[x]$ we have

$$x^4 + x + \bar{2} = (x^2 + ax + b)(x^2 + cx + d).$$

This implies that

$$bd = \bar{2}, \quad ad + bc = \bar{1}, \quad b + d + ac = \bar{0}, \quad a + c = \bar{0}. \tag{2.1}$$

The first equation implies that $b = \bar{2}d^{-1}$ and the fourth implies that $a = -c$. Substituting into the third equation then gives

$$\bar{2}d^{-1} + d - c^2 = 0 \implies \bar{2} + d^2 = c^2 d. \tag{2.2}$$

On the other hand, substituting into the second equation of (2.1) gives

$$ad + \bar{2}cd^{-1} = \bar{1} \implies ad^2 + \bar{2}c = d \implies (\bar{2} - d^2)c = d.$$

Since $bd = \bar{2}$, we have $d \neq \bar{0}$, and so the above implies that $c \neq \bar{0}$. Thus c is equal to $\bar{1}$ or $\bar{2}$. In either case we have $c^2 = \bar{1}$ and so (2.2) becomes $2 + d^2 = d$. Hence $d^2 - d + \bar{2} = \bar{0}$. But this equation has no roots in $\mathbb{Z}_3$. This contradiction completes the proof.

Theorem 2.2.20 (Eisenstein Criterion)**.** *Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$, $n \geq 1$, $a_n \neq 0$. Suppose that there exists a prime $p \in \mathbb{Z}$ such that*

(a) p divides each of $a_0, \ldots, a_{n-1}$,

(b) p does not divide a_n, and

(c) p^2 does not divide a_0.

Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof. Suppose, towards a contradiction, that $f(x)$ is not irreducible in $\mathbb{Q}[x]$. Then, by Theorem 2.2.17, we have a proper factorization $f(x) = g(x)h(x)$ in $\mathbb{Z}[x]$. Let

$$g(x) = b_0 + b_1x + \cdots + b_sx^s \quad \text{and} \quad h(x) = c_0 + c_1x + \cdots + c_tx^t,$$

with $b_s, c_t \neq 0$.

- Since $p|b_0c_0$, but $p^2 \nmid a_0 = b_0c_0$, we have that p divides exactly one of b_0 or c_0. Without loss of generality, assume that $p|b_0$ but $p \nmid c_0$.
- Since $p \nmid a_n = b_sc_t$, we have $p \nmid b_s$.
- Let k be the smallest integer such that $p \nmid b_k$. So $0 < k \leq s < n$.
- Equating the coefficients of x^k in $f(x) = g(x)h(x)$ gives

  $$a_k = b_kc_0 + b_{k-1}c_1 + \cdots + b_0c_k.$$

  We know that p divides a_k by assumption. Also, by our choice of k, p divides every term in the sum after the first. Thus p also divides b_kc_0. Since p is prime, it must therefore divide either b_k or c_0, a contradiction. □

Example 2.2.21. The polynomial $2x^7 - 5x^4 + 10x^2 - 15$ is irreducible in $\mathbb{Q}[x]$. This can be seen by applying the Eisenstein Criterion (Theorem 2.2.20) with $p = 5$.

Example 2.2.22. The polynomial $f(x) = x^4 + x^3 + x^2 + x + 1$ is irreducible in $\mathbb{Q}[x]$. To see this, first note that $f(x)(x - 1) = x^5 - 1$. Thus,

$$f(x + 1) = \frac{(x + 1)^5 - 1}{(x + 1) - 1} = \frac{(x + 1)^5 - 1}{x} = x^4 + 5x^3 + 10x^2 + 10x + 5.$$

Then we see that $f(x + 1)$ does not factor by the Eisenstein Criterion (Theorem 2.2.20) with $p = 5$. Hence $f(x)$ does not factor either, since if $f(x) = g(x)h(x)$, then $f(x + 1) = g(x + 1)h(x + 1)$.

Example 2.2.23. If p is a prime number, then $x^{p-1}+x^{p-2}+\cdots+x+1$ is called a *cyclotomic polynomial*. Using the argument of Example 2.2.22, one can show that this polynomial is irreducible.

Theorem 2.2.24. *Let F be a field and $f(x), g(x) \in F[x] \setminus \{0\}$. Then there exists a unique polynomial $d(x) \in F[x] \setminus \{0\}$ such that*

(a) $d(x)$ is monic,

(b) $d(x)|f(x)$ and $d(x)|g(x)$,

(c) if $h(x) \in F[x]$ satisfies $h(x)|f(x)$ and $h(x)|g(x)$, then $h(x)|d(x)$,

(d) there exists $u(x), v(x) \in F[x]$ such that $d(x) = u(x)f(x) + v(x)g(x)$.

Proof. Let $d(x)$ be a monic element of $\{u(x)f(x)+v(x)g(x) : u(x), v(x) \in F[x]\}$ of smallest degree. Then (a), (c) and (d) are obvious. To see (b), divide $f(x)$ by $d(x)$ to obtain $f(x) = q(x)d(x) + r(x)$ with $r(x) = 0$ or $\deg r(x) < \deg d(x)$. Then, writing $d(x) = u(x)f(x) + v(x)g(x)$ for some $u(x), v(x) \in F[x]$, we have

$$r(x) = f(x) - q(x)d(x) = f(x)(1 - q(x)u(x)) + g(x)(-q(x)v(x)).$$

Thus, by our choice of $d(x)$, we cannot have $\deg r(x) < \deg d(x)$. Thus $r(x) = 0$, which proves that $d(x)|f(x)$. The proof that $d(x)|g(x)$ is analogous. Thus (b) is proved.

It remains to show the uniqueness assertion. Suppose $d_1(x)$ and $d_2(x)$ both satisfy the given conditions. Then, $d_1(x)|d_2(x)$ and $d_2(x)|d_1(x)$. Hence

$$d_1(x) = d_2(x)k_1(x) \text{ and } d_2(x) = d_1(x)k_2(x) \text{ for some } k_1(x), k_2(x) \in F[x].$$

Hence $d_1(x) = d_1(x)k_1(x)k_2(x)$. Thus $k_1(x)k_2(x) = 1$, which implies that $k_1(x)$ and $k_2(x)$ are constant polynomials. But since $d_1(x)$ and $d_2(x)$ are both monic, we must have $k_1(x) = k_2(x) = 1$. Thus $d_1(x) = d_2(x)$. □

Definition 2.2.25 (Greatest common divisor of polynomials)**.** The polynomial $d(x)$ whose existence is asserted in Theorem 2.2.24 is called the *greatest common divisor* of $f(x)$ and $g(x)$ and is denoted $\gcd(f(x), g(x))$.

Algorithm 2.2.26 (Euclidean Algorithm)**.** Let $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Set $a_1(x) = f(x)$ and $a_2(x) = g(x)$. By repeated polynomial division, we have

$$\begin{cases} a_1(x) = q_1(x)a_2(x) + a_3(x) & \text{with } a_3(x) \neq 0 \text{ and } \deg a_3(x) < \deg a_2(x), \\ a_2(x) = q_2(x)a_3(x) + a_4(x) & \text{with } a_4(x) \neq 0 \text{ and } \deg a_4(x) < \deg a_3(x), \\ \quad\cdots & \cdots \\ a_{k-2}(x) = q_{k-2}(x)a_{k-1}(x) + a_k(x) & \text{with } a_k(x) \neq 0 \text{ and } \deg a_k(x) < \deg a_{k-1}(x), \\ a_{k-1}(x) = q_{k-1}(x)a_k(x). & \end{cases}$$

Then $\gcd(f(x), g(x)) = \frac{1}{c}a_k(x)$, where c is the leading coefficient of $a_k(x)$. A proof of this result is outlined in Exercise 2.2.3.

Example 2.2.27. In $\mathbb{Q}[x]$, we have $\gcd(x^2+1, 2x-1) = 1$ and $\gcd(x^4-2x^2+1, 2x^2-2) = x^2-1$.

Proposition 2.2.28. *Let F be a field, $p(x) \in F[x]$ an irreducible polynomial, and $f_1(x), \ldots, f_n(x) \in F[x]$. If $p(x)|f_1(x)\cdots f_n(x)$, then $p(x)|f_i(x)$ for some $1 \leq i \leq n$.*

Proof. We prove the result by induction. The case $n = 1$ is obvious. Now consider the case $n = 2$. Suppose $p(x)$ divides $f_1(x)f_2(x)$. Let $d(x) = \gcd(p(x), f_1(x))$. Then $d(x)|p(x)$ and thus, since $p(x)$ is irreducible, we have $\deg d(x) = 0$ (so $d(x) = 1$) or $\deg d(x) = \deg p(x)$. If $\deg d(x) = \deg p(x)$, then $p(x) = ad(x)$ for some nonzero $a \in F$ and so $p(x)$ divides $f_1(x)$ since $d(x)$ does. On the other hand, if $d(x) = 1$, then we have

$$u(x)p(x) + v(x)f_1(x) = 1 \text{ for some } u(x), v(x) \in F[x]$$

by Theorem 2.2.24. Thus $u(x)p(x)f_2(x) + v(x)f_1(x)f_2(x) = f_2(x)$. Therefore $p(x)$ divides $f_2(x)$ since it divides $f_1(x)f_2(x)$. Thus the result holds for $n = 2$.

If $n > 2$, then $p(x)|f_1(x)f_2(x)\cdots f_n(x)$ implies $p(x)|f_1(x)$ or $p(x)|f_2(x)\cdots f_n(x)$ (by the $n = 2$ case). The result then follows by the inductive hypothesis. □

Corollary 2.2.29. *If F is a field and $p(x) \in F[x]$ is irreducible, then $\langle p(x)\rangle$ is a prime ideal of $F[x]$.*

Proof. Suppose $p(x)$ is irreducible and $f(x)g(x) \in \langle p(x)\rangle$. Then $p(x)|f(x)g(x)$. Hence, by Proposition 2.2.28, $p(x)$ divides $f(x)$ or $g(x)$, which implies that $f(x) \in \langle p(x)\rangle$ or $g(x) \in \langle p(x)\rangle$. □

Example 2.2.30. We have that $\langle x^2 + x + 1\rangle$ is a prime ideal of $\mathbb{Z}_2[x]$. We will see later that $\langle p(x)\rangle$ is in fact a maximal ideal of $F[x]$ whenever $p(x)$ is irreducible (see Theorem 2.3.6).

Theorem 2.2.31 (Unique Factorization Theorem)**.** *Let F be a field and $f(x) \in F[x]$ be a nonconstant polynomial. Then*

(a) $f(x) = ap_1(x)\cdots p_m(x)$ where $a \in F^\times$ and $p_1(x), \ldots, p_m(x) \in F[x]$ are irreducible and monic, and

(b) the factorization of part (a) is unique up to a reordering of the $p_i(x)$'s.

Proof. (a) We prove the result by induction on $n = \deg f(x)$. The case $n = 1$ is obvious since degree one polynomials are irreducible. Assume the result holds for n and consider a polynomial $f(x)$ of degree $n + 1$. If $f(x)$ is irreducible, we are done. Otherwise, we can factor $f(x) = g(x)h(x)$ with $\deg g(x), \deg h(x) \leq n$. The result then follows from the inductive hypothesis.

(b) Suppose we have two factorizations

$$f(x) = ap_1(x)\cdots p_m(x) = bq_1(x)\cdots q_\ell(x), \tag{2.3}$$

with $a, b \in F^\times$ and $p_1(x), \ldots, p_m(x), q_1(x), \cdots, q_\ell(x)$ monic irreducible. Equating the leading terms, we see that $a = b$. Now, $p_1(x)|q_1(x)\cdots q_\ell(x)$, and so $p_1(x)|q_i(x)$ for some $1 \leq i \leq \ell$

by Proposition 2.2.28. Thus $p_1(x) = q_i(x)$ since both are monic and irreducible. Canceling $p_1(x)$ and $q_i(x)$ in (2.3) gives

$$p_2(x) \cdots p_m(x) = q_1(x) \cdots q_{i-1}(x) q_i(x) \cdots q_\ell(x).$$

The theorem follows by repeating the above argument. □

Exercises.

2.2.1. Prove Corollary 2.2.13.

2.2.2. Show that if p is a prime number, then the cyclotomic polynomial $x^{p-1}+x^{p-2}+\cdots+x+1$ is irreducible. See Example 2.2.23.

2.2.3. Let $f(x), g(x) \in F[x]$, with $g(x) \neq 0$.

(a) Show that Algorithm 2.2.26 terminates after a finite number of steps.

(b) In the notation of this algorithm, show that $\gcd(a_i(x), a_{i+1}(x)) = \gcd(a_{i+1}(x), a_{i+2}(x))$ for all $i = 1, \ldots, k-2$.

(c) In the same notation, show that $\gcd(a_{k-1}(x), a_k(x)) = \frac{1}{c} a_k(x)$.

(d) Conclude that $\gcd(f(x), g(x)) = \frac{1}{c} a_k(x)$.

2.2.4 ([Nic12, Ex. 4.2.1]). (a) If F is a field and $a \in F$, $a \neq 0$, show that a divides f for every $f \in F[x]$.

(b) If $p \in F[x]$ divides f for every $f \in F[x]$, show that $p \in F$, $p \neq 0$.

2.2.5 ([Nic12, Ex. 4.2.2]). If $f, g \in F[x]$ and F is a field, consider the following statements:

(a) $f = ag$ for some $0 \neq a \in F$;

(b) f and g have the same roots in F.

Prove that (a) $\implies$ (b). Does (b) $\implies$ (a)? Justify your answer.

2.2.6 ([Nic12, Ex. 4.2.3]). In each case, explain why f is reducible over any field.

(a) $f = x^3 - 2x^2 + 3x - 2$

(b) $f = x^3 + x^2 + 4$

2.2.7 ([Nic12, Ex. 4.2.4]). In each case determine whether the polynomial is irreducible. Give reasons.

(a) $x^3 + 5$ in $\mathbb{Z}_7[x]$

(b) $x^2 - 2$ in $\mathbb{R}[x]$

(c) $x^2 + 11$ in $\mathbb{C}[x]$

(d) $x^3 - 4$ in $\mathbb{Z}_{11}[x]$

(e) $x^3 + x + 1$ in $\mathbb{Z}_5[x]$

(f) $x^2 + x + 1$ in $\mathbb{Z}_{17}[x]$

2.2.8 ([Nic12, Ex. 4.2.5]). In each case, determine whether the polynomial is irreducible over each of the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_5$, and $\mathbb{Z}_7$.

(a) $x^2 - 3$

(b) $x^2 + x + 1$

(c) $x^3 + x + 1$

(d) $x^3 - 2$

2.2.9 ([Nic12, Ex. 4.2.6]). Let R be an integral domain and let $f \in R[x]$ be monic. If f factors in $R[x]$, show that it has a factorization $f = gh$, where g and h are both monic (and of positive degree).

2.2.10 ([Nic12, Ex. 4.2.8]). (a) If $x^2 + ax + b$ has roots u and v in a field F, show that $b = uv$ and $a = -(u + v)$.

(b) Show that $1 + i$ is a root of $x^2 + (1 - 2i)x - (3 + i) \in \mathbb{C}[x]$. Find the other root.

2.2.11 ([Nic12, Ex. 4.2.9]). Show that an odd degree polynomial in $\mathbb{R}[x]$ has a real root. (Requires calculus.)

2.2.12 ([Nic12, Ex. 4.2.10]). Find all monic irreducible cubics in $\mathbb{Z}_2[x]$.

2.2.13 ([Nic12, Ex. 4.2.12]). Let p be a monic quartic in $\mathbb{Z}_2[x]$. Show that p is irreducible in $\mathbb{Z}_2[x]$ if and only if

(a) p has no root in $\mathbb{Z}_2$, and

(b) $p \neq x^4 + x^2 + 1$.

Hint: Exercise 2.2.8.

2.2.14 ([Nic12, Ex. 4.2.13]). Show that a monic quintic p in $\mathbb{Z}_2[x]$ in irreducible if and only if

(a) p has no root in $\mathbb{Z}_2$, and

(b) p is neither $x^5 + x^4 + 1$ nor $x^5 + x + 1$.

Hint: Exercise 2.2.12.

2.2.15 ([Nic12, Ex. 4.2.18]). In each case, factor f as a product of irreducible polynomials in $F[x]$.

(a) $f = 3x^4 + 2$, $F = \mathbb{Z}_5$

(b) $f = 3x^4 + 2$, $F = \mathbb{Z}_{11}$

(c) $f = x^3 + 2x^2 + 2x + 1$, $F = \mathbb{Z}_7$

(d) $f = x^3 + 2x^2 + 2x + 1$, $F = \mathbb{Z}_3$

(e) $f = x^4 - x^2 + x - 1$, $F = \mathbb{Z}_{13}$.

(f) $f = x^4 - x^2 + x - 1$, $F = \mathbb{Z}_{17}$.

2.2.16 ([Nic12, Ex. 4.2.19]). Factor $x^5 + x^4 + 1$ as a product of irreducible polynomials in $\mathbb{Z}_2[x]$.

2.2.17 ([Nic12, Ex. 4.2.20]). Factor $x^5 + x^2 - x + 1$ as a product of irreducible polynomials in $\mathbb{Z}_3[x]$.

2.2.18 ([Nic12, Ex. 4.2.21]). Show that each polynomial is irreducible in $\mathbb{Q}[x]$.

(a) $3x^3 + 5x^2 + x + 2$

(b) $5x^3 + 2x + 3$

(c) $x^3 + 9x^2 + x + 6$

(d) $x^3 + x^2 + 10x + 8$

2.2.19 ([Nic12, Ex. 4.2.22]). Show that each polynomial is irreducible in $\mathbb{Q}[x]$.

(a) $x^5 + 6x^4 + 12x + 15$

(b) $4x^5 + 28x^4 + 7x^3 - 28x^2 + 14$

2.2.20 ([Nic12, Ex. 4.2.24]). Show that $f(x) = x^4 + 4x^3 + 4x^2 + 4x + 5$ is irreducible over $\mathbb{Q}$ by considering $f(x - 1)$.

2.2.21 ([Nic12, Ex. 4.2.29]). Show that $x^n - p$ is irreducible in $\mathbb{Q}[x]$ for all $n \geq 2$ and all primes $p \in \mathbb{Z}$. (Hence $\mathbb{Q}[x]$ has infinitely many irreducible polynomials of every degree ≥ 2.)

2.2.22 ([Nic12, Ex. 4.2.30]). Suppose p is a prime number. Show that $x^p - a$ is reducible in $\mathbb{Z}_p[x]$ for every $a \in \mathbb{Z}_p$.

2.2.23 ([Nic12, Ex. 4.2.31]). Let $F \subseteq K$ be fields and $f, g \in F[x]$.

(a) If f is irreducible in $K[x]$, show that it is irreducible in $F[x]$.

(b) If f and g are relatively prime in $F[x]$ (i.e. their greatest common divisor is 1), show that they are relatively prime in $K[x]$. *Hint*: Theorem 2.2.24(d).

2.2.24 ([Nic12, Ex. 4.2.34]). Let $f = x^3 - 42x^2 + 35x + m$. Show that there are infinitely many integers m for which f is irreducible in $\mathbb{Q}[x]$. *Hint*: Eisenstein Criterion.

2.2.25 ([Nic12, Ex. 4.2.36]). If m and p are integers with p prime, show that $x^4 + mx + p$ is irreducible in $\mathbb{Q}[x]$ if and only if it has no roots in $\mathbb{Q}$.

2.2.26 ([Nic12, Ex. 4.2.39]). In each case compute $d = \gcd(f, g)$, and express it in $F[x]$ as a linear combination of f and g.

(a) $f = x^2 + 2$, $g = x^3 + 4x^2 + x + 1$, $F = \mathbb{Z}_5$

(b) $f = x^2 + 1$, $g = x^5 + x^4 + x^3 + x^2 + x + 1$, $F = \mathbb{Z}_2$

(c) $f = x^2 - x - 2$, $g = x^5 - 4x^3 - 2x^2 + 7x - 6$, $F = \mathbb{Q}$

(d) $f = x^3 + x - 2$, $g = x^5 - x^4 + 2x^2 - x - 1$, $F \in \mathbb{Q}$

2.3 Quotient rings of polynomials over a field

The following result states that, if F is a field, all ideals in $F[x]$ are principal.

Theorem 2.3.1. *Let F be a field and let I be a nonzero ideal of $F[x]$. Then there is a unique monic polynomial $h(x) \in F[x]$ such that $I = \langle h(x) \rangle = h(x)F[x] = \{h(x)g(x) : g(x) \in F[x]\}$.*

Proof. Let I be a nonzero ideal of $F[x]$. Then I contains nonzero polynomials and hence it contains monic polynomials (since it is an ideal, we can multiply any polynomials by the inverse of the leading coefficient to obtain a monic polynomial in I). Among all the monic polynomials in I, choose $h(x)$ of minimal degree. Then $\langle h(x) \rangle \subseteq I$.

Now suppose $f(x) \in I$. By the Euclidean Algorithm (Algorithm 2.2.26), we have $f(x) = q(x)h(x) + r(x)$ for $q(x), r(x) \in F[x]$ with $r(x) = 0$ or $\deg r(x) < \deg h(x)$. Suppose $r(x) \neq 0$ and let a be the leading coefficient of $r(x)$. Then $a^{-1}r(x)$ is monic and

$$a^{-1}r(x) = a^{-1}(h(x) - q(x)f(x)) \in I.$$

But $\deg(a^{-1}r(x)) = \deg r(x) < \deg h(x)$, which contradicts our choice of $h(x)$. Thus $r(x) = 0$ and so $I = \langle h(x) \rangle$.

To prove uniqueness, suppose that $\langle h_1(x) \rangle = \langle h_2(x) \rangle \neq 0$, where $h_1(x)$ and $h_2(x)$ are monic.. Then there exist $f_1(x), f_2(x) \in F[x]$ such that

$$h_1(x) = f_1(x)h_2(x) \quad \text{and} \quad h_2(x) = f_2(x)h_1(x).$$

Thus $h_1(x) = f_1(x)f_2(x)h_1(x)$, which implies that $\deg f_1(x) = \deg f_2(x) = 0$. Thus $f_1(x) \in F$. Since $h_1(x)$ and $h_2(x)$ are both monic, we have $f_1(x) = 1$. So $h_1(x) = h_2(x)$. □

Example 2.3.2. Consider the ring $\mathbb{R}[x]$ and let $I = \langle x^2 + 1 \rangle$. We can divide any $f(x) \in \mathbb{R}[x]$ by $x^2 + 1$ to obtain $f(x) = q(x)(x^2 + 1) + (ax + b)$ for unique $q(x) \in \mathbb{R}[x]$ and $a, b \in \mathbb{R}$. Thus, any element of $\mathbb{R}[x]/I$ can be written uniquely in the form $(ax + b) + I$. We have $(ax + b) + I = (a'x + b') + I$ if and only if $a = a'$ and $b = b'$. Furthermore we have

$$((ax + b) + I) + ((a'x + b') + I) = ((a + a')x + (b + b')) + I,$$
$$((ax + b) + I)((a'x + b') + I) = (aa'x^2 + ab'x + a'bx + bb') + I = ((a'b + ab')x + (bb' - aa')) + I.$$

Thus $\mathbb{R}[x]/I \cong \mathbb{C}$ via the map $ax + b \mapsto ai + b$.

The above example is a special case of the following general theorem. Its proof can be found in [Nic12, §4.3].

Theorem 2.3.3. *Let F be a field, $h(x) \in F[x]$, and $I = \langle h(x) \rangle$. Suppose $n = \deg h(x)$. Then every element of $R = F[x]/I$ has a unique representative of degree $\leq n-1$. Thus the factor ring $R = F[x]/I$ is given by*

$$R \cong \{a_0 + a_1 t + \cdots + a_{n-1} t^{n-1} : a_0, \ldots, a_{n-1} \in F\}$$

(isomorphism of vector spaces).

In Theorem 2.3.3, t corresponds to $x + I$. The addition in the representation of R given in Theorem 2.3.3 is given by adding coefficients. The multiplication is computed by using the fact that $h(t) = 0$. This allows us to express t^n in terms of lower powers of t.

Example 2.3.4. Consider the quotient ring $\mathbb{Z}_2[x]/\langle x^2 \rangle$. By Theorem 2.3.3, we have

$$\mathbb{Z}_2[x]/\langle x^2 \rangle \cong \{at + b : a, b \in \mathbb{Z}_2\} = \{\bar{0}, \bar{1}, t, \bar{1} + t\}.$$

When computing in this ring, we use the fact that $t^2 = h(t) = 0$. Thus we have the following addition and multiplication tables.

$+$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
$\bar{0}$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
$\bar{1}$	$\bar{1}$	$\bar{0}$	$\bar{1}+t$	t
t	t	$\bar{1}+t$	$\bar{0}$	$\bar{1}$
$\bar{1}+t$	$\bar{1}+t$	t	$\bar{1}$	$\bar{0}$

$\times$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
t	$\bar{0}$	t	$\bar{0}$	t
$\bar{1}+t$	$\bar{0}$	$\bar{1}+t$	t	$\bar{1}$

Note that this ring is not a field since, for example, the nonzero element t is not invertible.

Example 2.3.5. Consider the quotient ring $\mathbb{Z}_2[x]/\langle x^2 + x + \bar{1} \rangle$. By Theorem 2.3.3, we have an isomorphism of vector spaces

$$\mathbb{Z}_2[x]/\langle x^2 + x + \bar{1} \rangle \cong \{at + b : a, b \in \mathbb{Z}_2\}.$$

When computing in this ring, we use the fact that $h(t) = 0$ and so $t^2 = -t - \bar{1} = t + \bar{1}$. The multiplication table is therefore as follows:

$\times$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	t	$\bar{1}+t$
t	$\bar{0}$	t	$\bar{1}+t$	$\bar{1}$
$\bar{1}+t$	$\bar{0}$	$\bar{1}+t$	$\bar{1}$	t

This is a field with four elements. (Recall that $\mathbb{Z}_4$ is *not* a field, so this quotient is not isomorphic to $\mathbb{Z}_4$.)

Theorem 2.3.6. *Let F be a field and $h(x) \in F[x]$ with $\deg h(x) \geq 1$. Then the following statements are equivalent:*

(a) $h(x)$ is irreducible.

(b) $\langle h(x) \rangle$ is a prime ideal.

(c) $\langle h(x)\rangle$ is a maximal ideal.

Proof. (a) $\Rightarrow$ (b): This is Corollary 2.2.29.

(b) $\Rightarrow$ (c): Suppose $\langle h(x)\rangle$ is prime but not maximal. Then there exists some ideal I of $F[x]$ such that $\langle h(x)\rangle \subsetneq I \subsetneq F[x]$. Now, we know by Theorem 2.3.1 that $I = \langle p(x)\rangle$ for some $p(x) \in F[x]$. Thus $h(x) = p(x)q(x)$ for some $q(x) \in F[x]$. But $p(x) \notin \langle h(x)\rangle$, and so $q(x) \in \langle h(x)\rangle$, i.e., $q(x) = h(x)r(x)$ for some $r(x) \in F[x]$. This implies that $h(x) = p(x)h(x)r(x)$, and so $\deg p(x) = 0$, which contradicts the fact that $I \neq F[x]$.

(c) $\Rightarrow$ (a): If $h(x)$ is not irreducible, then $h(x) = h_1(x)h_2(x)$ for some $h_1(x), h_2(x) \in F[x]$ with $\deg h_1(x), \deg h_2(x) < \deg h(x)$. Thus $\langle h(x)\rangle \subsetneq \langle h_1(x)\rangle \subsetneq F[x]$, which implies that $\langle h(x)\rangle$ is not maximal. □

One of the important consequences of Theorem 2.3.6 is that fields of order p^m exist for all prime numbers p and positive integers m. We explore this in the next section.

Exercises.

Throughout these exercises, F denotes a field.

2.3.1 ([Nic12, Ex. 4.3.1(b)]). Define

$$A = \{f(x) \in F[x] : \text{the sum of the coefficients of } f(x) \text{ is zero}\}.$$

Find a monic polynomial $h(x) \in F[x]$ such that $A = \langle h(x)\rangle$.

2.3.2 ([Nic12, Ex. 4.3.2]). In each case, describe $R = F[x]/\langle h\rangle$ as in Theorem 2.3.3 and write out the addition and multiplication tables for R.

(a) $h = x^2 + 1$, $F = \mathbb{Z}_2$

(b) $h = x^2 + x$, $F = \mathbb{Z}_2$

(c) $h = x^3 + 1$, $F = \mathbb{Z}_2$

(d) $h = x^2 - 1$, $F = \mathbb{Z}_3$

(e) $h = x^2$, $F = \mathbb{Z}_3$

(f) $h = x^2 - x + 1$, $F = \mathbb{Z}_3$

2.3.3 ([Nic12, Ex. 4.3.3]). Construct a field of order 8 and write down the multiplication table.

2.3.4 ([Nic12, Ex. 4.3.4]). Construct a field of order 9 and write down the multiplication table.

2.3.5 ([Nic12, Ex. 4.3.5(b)]). Construct a field of order 25.

2.3.6 ([Nic12, Ex. 4.3.6]). In each case, determine all the idempotents, nilpotents, and units in $R = F[x]/\langle h\rangle$:

(a) $h = x^2 - x$

(b) $h = x^2$

2.3.7 ([Nic12, Ex. 4.3.7]). In each case, show that r is a unit in $R = F[x]/\langle h\rangle$ and find the inverse. Use the notation of Theorem 2.3.3.

(a) $r = 1 + t^2$, $F = \mathbb{Z}_{11}$, $h = x^3 + 1$

(b) $r = 1 + t - t^2$, $F = \mathbb{Z}_7$, $h = x^3 + x^2 - 1$

2.3.8 ([Nic12, Ex. 4.3.8]). Because $x - a$ is irreducible over the field F, Theorem 2.3.6 and Corollary 1.3.19 imply that $F[x]/\langle x - a\rangle$ is a field. Describe this field. How is it related to F?

2.3.9 ([Nic12, Ex. 4.3.9]). Find a subring of $\mathbb{R}$ isomorphic to $\mathbb{Q}[x]/\langle x^3 - 2\rangle$.

2.3.10 ([Nic12, Ex. 4.3.10]). (a) Show that

$$F[x]/\langle x^2\rangle \cong \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in F \right\}, \text{ a subring of } \mathrm{Mat}_2(F).$$

(b) Show that

$$F[x]/\langle x^3\rangle \cong \left\{ \begin{bmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{bmatrix} : a, b \in F \right\}, \text{ a subring of } \mathrm{Mat}_3(F).$$

(c) Generalize the above results.

2.3.11 ([Nic12, Ex. 4.3.11]). Find a ring isomorphism $F[x]/\langle x^2 - x\rangle \to F \times F$.

2.3.12 ([Nic12, Ex. 4.3.12]). Let $R = F[x]/\langle x^2 - 1\rangle = \{a + bt : a, b \in F,\ t^2 = 1\}$. Show that $a + bt$ is a unit in R if and only if $a^2 \neq b^2$. *Hint*: If $r = a + bt$, let $r^* = a - bt$, and $N(r) = rr^*$. Show that $(rs)^* = r^*s^*$, and hence that $N(rs) = N(r)N(s)$ for all $r, s \in R$.

2.3.13 ([Nic12, Ex. 4.3.18]). Let A denote the set of all polynomials in $\mathbb{Z}[x]$ with even constant term. Show that A is an ideal of $\mathbb{Z}[x]$ that is not principal. (*Note:* You must show that A is an ideal.) This shows that Theorem 2.3.1 fails if we replace F by an integral domain in general.

2.3.14 ([Nic12, Ex. 4.3.21]). (a) Suppose $a, b \in F$ such that $a^2 - 4b$ is not the square of an element of F. Show that $x^2 + ax + b$ is irreducible in $F[x]$.

(b) Show that the converse of (a) holds if $2 \neq 0$ in F.

2.3.15 ([Nic12, Ex. 4.3.22]). Let f and g be nonzero polynomials in $F[x]$.

(a) Show that $A = \{uf + vg : u, v \in F[x]\}$ is an ideal of $F[x]$.

(b) Explain how Theorem 2.3.1 is related to Theorem 2.2.24.

2.3.16 ([Nic12, Ex. 4.3.23]). Polynomials $f_1, f_2, \ldots, f_m$ in $F[x]$ are called *relatively prime* if 1 is the only common monic divisor of all of them in $F[x]$. Show that $f_1, f_2, \ldots, f_m$ are relatively prime if and only if $1 = q_1 f_1 + q_2 f_2 + \cdots + q_m f_m$ for some $q_1, q_2, \ldots, q_m \in F[x]$. *Hint*: Theorem 2.3.1.

2.3.17 ([Nic12, Ex. 4.3.26]). (a) Let $A \neq F[x]$ be an ideal in $F[x]$. If $A \neq 0$, show that A is prime if and only if it is maximal.

(b) What happens if $A = 0$? Justify your answer.

2.3.18 ([Nic12, Ex. 4.3.27]). Let h be a nonconstant monic polynomial in $F[x]$. Show that $F[x]/\langle h\rangle$ has no nonzero nilpotent elements if and only if $h = p_1 p_2 \cdots p_r$, where $p_1, p_2, \ldots, p_r$ are distinct monic irreducible polynomials. *Hint*: Use Theorem 2.2.31.

2.3.19. Use the first isomorphism theorem (Theorem 1.4.15) to give an alternative proof that $\mathbb{R}[x]/\langle x^2 + 1\rangle \cong \mathbb{C}$ (see Exercise 2.3.2).

2.3.20 (Bonus). Prove that $\mathbb{C}[x]/\langle x^2 + 1\rangle \cong \mathbb{C} \times \mathbb{C}$ (as rings).

2.4 Finite fields

In this section, we look at finite fields. You've already seen the fields $\mathbb{Z}_p$, p a prime. We will see here that there are other finite fields. In fact, it is possible to completely classify all finite fields, up to isomorphism.

Example 2.4.1. The polynomial $x^2 + x + 1$ has no root in $\mathbb{Z}_2$ and so it is irreducible by Theorem 2.2.7. Thus

$$\mathbb{Z}_2[x]/\langle x^2 + x + \bar{1}\rangle = \{a + bt : a, b \in \mathbb{Z}_2,\ t^2 + t + \bar{1} = 0\}$$

is a field with four elements. We saw this by direct computation in Example 2.3.5.

The idea behind Example 2.4.1 is the following: If $h(x)$ is a monic irreducible polynomial in $\mathbb{Z}_p[x]$ of degree m, then, by Theorem 2.3.6, the ideal $\langle h(x)\rangle$ is maximal and hence $\mathbb{Z}_p[x]/\langle h(x)\rangle$ is a field by Corollary 1.3.19. Using Theorem 2.3.3, it is not hard to see that this field has p^m elements. It turns out that this procedure produces *all* finite fields.

Suppose F is a finite field. Then its additive group is finite and thus, by Lemma 1.1.19, its characteristic is positive. Thus, by Proposition 1.2.12, its characteristic is a prime number p. Therefore, by Theorem 1.4.20, $\mathbb{Z}_p \subseteq F$ (the map $\iota\colon \mathbb{Z}_p \hookrightarrow F$, $\iota(\bar{k}) = k \cdot 1$ is an injective ring homomorphism). We can thus view F as a vector space over the field $\mathbb{Z}_p$.

Proposition 2.4.2. *If F is a finite field, then $|F| = p^n$ for some n, where $p = \operatorname{char} F$.*

Proof. Since F is a finite field, it is finite-dimensional as a $\mathbb{Z}_p$-vector space. If $n = \dim_{\mathbb{Z}_p} F$, then $F \cong (\mathbb{Z}_p)^n$ as a vector space and the result follows. □

Theorem 2.4.3. *The order of a finite field is a prime power. For every prime power $q = p^n$, there is a unique field F of order q, up to isomorphism. In these fields, every element a satisfies $a^q = a$, and the polynomial $x^q - x$ factors as*

$$x^q - x = \prod_{a \in F} (x - a).$$

The proof of this theorem involves some topics in field theory that are beyond the scope of this course. (They are covered in MAT 4143.) Thus, we give only a sketch of the proof.

Sketch of proof. The first statement is Proposition 2.4.2. The *splitting field* of a polynomial $f(x) \in F[x]$, where F is a field, is the smallest field that contains F and in which $f(x)$ can be written as a product of linear factors. One can show that a splitting field of $f(x)$ always exists, and is unique up to isomorphism. The field of order p^n is the splitting field of the polynomial $x^{p^n} - x$. □

Definition 2.4.4 (Galois field)**.** The unique field with p^n elements is called the *Galois field* of order p^n and is denoted $\mathrm{GF}(p^n)$.

Theorem 2.4.5. *Suppose F is a field. If G is a finite subgroup of $F^\times$, then G is cyclic.*

Proof. Let G be a finite subgroup of $F^\times$ with n elements. By the Fundamental Theorem of Finite Abelian Groups (from MAT 2143),

$$G \cong \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_k^{e_k}}$$

for some (not necessarily distinct) primes p_i such that $n = p_1^{e_1} \cdots p_k^{e_k}$. Let m be the least common multiple of $p_1^{e_1}, \ldots, p_k^{e_k}$. Then G contains an element of order m (the product of generators of the cyclic factors above). Since every α in G satisfies $\alpha^m = 1$, we see that every element of G is a root of $x^m - 1$. Since $x^m - 1$ has at most m roots in F, $n \leq m$. On the other hand, we know that $m \leq |G|$; therefore, $m = n$. Thus, G contains an element of order n and must be cyclic. □

Corollary 2.4.6. *The multiplicative group of all nonzero elements of a finite field is cyclic.*

Definition 2.4.7 (Primitive element, primitive root of unity)**.** If F is a finite field, then a generator for $F^\times$ is called a *primitive element* of F. If a (possibly infinite) field F has a multiplicative subgroup G of order n, then a generator of G (which exists by Theorem 2.4.5) is called a *primitive nth root of unity* in F. For example, $e^{2\pi i/n}$ is a primitive nth root of unity in $\mathbb{C}$ for each $n \geq 2$.

Exercises.

Throughout these exercises, F denotes a field.

2.4.1. Find a primitive element for the following fields:

(a) $\mathbb{Z}_7$

(b) $\mathbb{Z}_{13}$

(c) The field of Example 2.4.1.

2.4.2 ([Nic12, Ex. 6.4.3]). Explain why $\mathbb{Z}_2[x]/\langle p\rangle$ and $\mathbb{Z}_2[x]/\langle q\rangle$ are isomorphic if $p = x^3 + x^2 + 1$ and $q = x^3 + x + 1$.

2.4.3. Let F be a field for which $F^\times$ is a cyclic group. Prove that F is finite. You may assume that $\operatorname{char} F \neq 2$ (although the result also holds when $\operatorname{char} F = 2$).

Chapter 3

Integral domains

In this chapter we discuss integral domains in further detail. In particular, we introduce the notions of unique factorization domains, principal ideal domains and euclidean domains. A reference for the material in this chapter is [Jud19, Ch. 18]. Throughout this chapter, unless otherwise noted, we assume that R is an integral domain.

3.1 Unique factorization domains

Recall that, for $a, b \in R$, we say that a *divides* b, and we write $a|b$, if $b = ac$ for some $c \in R$.

Definition 3.1.1 (Factorization)**.** A *factorization* of an element $a \in R$ is way of writing $a = bc$ with $b, c \in R$. Such a factorization is said to be *trivial* if b or c is a unit in R.

Trivial factorizations are not so interesting. We will consider two factorizations $r = ab$ and $r = (ua)(u^{-1}b)$ for some unit $u \in R^\times$ to be essentially the same.

Example 3.1.2. In $\mathbb{Z}$ we have $8 = 4 \cdot 2 = (-4) \cdot (-2)$. In $\mathbb{R}[x]$ we have

$$(x^3 - 1) = (x - 1)(x^2 + x + 1) = (3x - 3)\left(\frac{1}{3}x^2 + \frac{1}{3}x + \frac{1}{3}\right).$$

Theorem 3.1.3. *Let $a, b \in R$. The following are equivalent:*

(a) $a|b$ *and* $b|a$,

(b) $a = ub$ *for some* $u \in R^\times$,

(c) $\langle a \rangle = \langle b \rangle$.

Proof. The proof of this theorem is left as Exercise 3.1.1. □

Definition 3.1.4 (Associates)**.** Two elements $a, b \in R$ are called *associates* if $a|b$ and $b|a$. When a and b are associates, we write $a \sim b$.

Lemma 3.1.5. *The relation $\sim$ (i.e. being associates) is an equivalence relation.*

Proof. The proof of this lemma is left as Exercise 3.1.2. □

Example 3.1.6. The equivalence classes of $\sim$ in $\mathbb{Z}$ are $\{0\}$, $\{\pm 1\}$ $\{\pm 2\}, \ldots$.

Example 3.1.7. Consider the ring $\mathbb{Z}(\sqrt{3}) := \{a+b\sqrt{3} : a, b \in \mathbb{Z}\}$. Since $(2+\sqrt{3})(2-\sqrt{3}) = 1$, we see that $2+\sqrt{3} \in \mathbb{Z}(\sqrt{3})^\times$. Then, since $\sqrt{3}(2+\sqrt{3}) = 3+2\sqrt{3}$, we see that $\sqrt{3} \sim 3+2\sqrt{3}$.

The following concept is useful when analyzing subrings of $\mathbb{C}$. For a complex number $\omega \in \mathbb{C} \setminus \mathbb{Q}$ with $\omega^2 \in \mathbb{Z}$, define the *norm* of $x = m + n\omega$, $m, n \in \mathbb{Z}$, to be

$$N(x) = N(m + n\omega) = (m + n\omega)(m - n\omega) = m^2 - n^2\omega^2. \tag{3.1}$$

Then $N(xy) = N(x)N(y)$ for $x, y \in \mathbb{Z} + \omega\mathbb{Z} := \{m + n\omega : m, n \in \mathbb{Z}\}$. Note also that $N(x) \in \mathbb{Z}$ for all $x \in \mathbb{Z} + \omega\mathbb{Z}$, and that $N(x) = 0$ if and only if $x = 0$ (exercise 3.1.3).

Example 3.1.8. If ω is purely imaginary (e.g. $\omega = \sqrt{-5}$), then

$$N(x) = \bar{x}x = |x|^2.$$

We have $N(x) \in \mathbb{N}$ for all $x \in \mathbb{Z}(i)$.

Example 3.1.9. Let's find all the units in $\mathbb{Z}(\sqrt{-5}) := \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. We have

$$(a+b\sqrt{-5})(c+d\sqrt{-5}) = 1 \implies N(a+b\sqrt{-5})N(c+d\sqrt{-5}) = 1 \implies (a^2+5b^2)(c^2+5d^2) = 1.$$

Since both $a^2 + 5b^2$ and $c^2 + 5d^2$ are nonnegative integers, we must have $a^2 + 5b^2 = 1$ and so $a = \pm 1$ and $b = 0$. Therefore $\mathbb{Z}(\sqrt{-5})^\times = \{\pm 1\}$. So, the only associates of $z \in \mathbb{Z}(\sqrt{-5})$ are $\pm z$. In contrast, recall that $\mathbb{Z}(i)^\times = \{\pm 1, \pm i\}$ (see Example 1.1.25).

The next definition generalizes the concepts of irreducible and prime elements to arbitrary integral domains.

Definition 3.1.10 (Irreducible, prime)**.** Let $r \in R$. We say that r is *irreducible* if the following conditions are satisfied:

(a) $r \neq 0$,

(b) $r \notin R^\times$,

(c) if $r = ab$, then $a \in R^\times$ or $b \in R^\times$.

We say that r is *prime* if the following conditions are satisfied:

(a) $r \neq 0$,

(b) $r \notin R^\times$,

(c) for every $a, b \in R$, if $r|ab$, then $r|a$ or $r|b$.

Example 3.1.11. Every prime number $p \in \mathbb{Z}$ is both irreducible and prime.

Example 3.1.12. Every irreducible polynomial $f(x) \in F[x]$ (where F is a field) is both irreducible (by definition) and prime (by Proposition 2.2.28).

Theorem 3.1.13. *Any prime element of an integral domain is irreducible.*

Proof. Suppose $r \in R$ is prime and $r = ab$ for some $a, b \in R$. Then $r|ab$ and so $r|a$ or $r|b$. Without loss of generality (i.e. interchanging a and b if necessary), we can assume $r|a$. Then $a = rc$ for some $c \in R$. Then

$$r = ab = rcb \implies 1 = cb \implies b \text{ is a unit},$$

where in the first implication we used the cancelation law in domains (Proposition 1.2.7). □

The converse of the above theorem is false, as the next example illustrates.

Example 3.1.14. We will show that $1 + \sqrt{-5}$ is an irreducible element of $\mathbb{Z}(\sqrt{-5})$ that is not prime.

Suppose that we have a factorization

$$1 + \sqrt{-5} = (a + b\sqrt{-5})(c + d\sqrt{-5}).$$

Then, applying the norm function N of (3.1), we have

$$6 = N\left(1 + \sqrt{-5}\right) = N\left(a + b\sqrt{-5}\right) N\left(c + d\sqrt{-5}\right) = (a^2 + 5b^2)(c^2 + 5d^2).$$

Since each of the factors on the right is a nonnegative integer, we must have (interchanging the two factors if necessary)

- $a^2 + 5b^2 = 1$ and $c^2 + 5d^2 = 6$, in which case $a = \pm 1$, $b = 0$; or
- $a^2 + 5b^2 = 2$ and $c^2 + 5d^2 = 3$, which is impossible.

Thus the factorization is trivial.

To see that $1 + \sqrt{-5}$ is not prime, note that $(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6$. So $1 + \sqrt{-5}$ divides $6 = 3 \cdot 2$. However, we claim that $1 + \sqrt{-5}$ does not divide 2 or 3. If it divided 2, we could find $x \in \mathbb{Z}(\sqrt{-5})$ with

$$2 = (1 + \sqrt{-5})x \implies 4 = N(2) = N\left(1 + \sqrt{-5}\right) N(x) \implies 4 = 6N(x),$$

which is a contradiction since $N(x) \in \mathbb{Z}$. Similarly, if $1 + \sqrt{-5}$ divided 3, we would have $x \in \mathbb{Z}(\sqrt{-5})$ with

$$3 = (1 + \sqrt{-5})x \implies 9 = N(3) = N\left(1 + \sqrt{-5}\right) N(x) \implies 9 = 6N(x),$$

which is again a contradiction. Thus $1 + \sqrt{-5}$ is not prime.

Definition 3.1.15 (Unique factorization domain)**.** An integral domain R is called a *unique factorization domain* (or *UFD*) if it satisfies the following conditions:

(a) If $r \in R$ such that $r \neq 0$ and $r \notin R^\times$, then r can be written as a product of irreducible elements.

(b) If $r \in R$, $r \neq 0$ and $r \notin R^\times$ with $r = p_1 \cdots p_s = q_1 \cdots q_t$ (where the p_i and q_i are irreducibles) then $s = t$ and, after relabeling if necessary, $p_i \sim q_i$ for $i = 1, \ldots, s$.

Example 3.1.16. The ring $\mathbb{Z}$ of integers is a UFD. Also, by Theorem 2.2.31, the ring $F[x]$ is a UFD for any field F.

Lemma 3.1.17. *Every irreducible element in a UFD is prime.*

Proof. Suppose R is a UFD and $r \in R$ is irreducible with $r|ab$ for some $a, b \in R$. Then we have $ab = rs$ for some $s \in R$. Since R is a UFD, we can factor a, b, s as products of irreducibles:

$$a = p_1 \cdots p_k, \quad b = q_1 \cdots q_\ell, \quad s = t_1 \cdots t_n.$$

Thus $p_1 \cdots p_k q_1 \cdots q_\ell = r t_1 \cdots t_n$. By the uniqueness of factorization in a UFD, we have $r \sim p_i$ for some $i = 1, \ldots, k$ or $r \sim q_j$ for some $j = 1, \ldots, \ell$. Thus $r|a$ or $r|b$. □

Example 3.1.18. By Example 3.1.14 and Lemma 3.1.17, we see that $\mathbb{Z}(\sqrt{-5})$ is not a UFD. The factorizations

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

are not equivalent (i.e. the factors on the left are not associates of the factors on the right).

In a UFD, one can use the factorization of an element to find all its divisors (up to associates). Suppose R is a UFD and $a \in R$ is a nonzero nonunit. Then we can write uniquely

$$a \sim p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r},$$

where $a_i \geq 1$ and p_i is a prime in R for $i = 1, \ldots, r$ and the p_i are nonassociates (i.e. $p_i \not\sim p_j$ for $i \neq j$). The uniqueness here means that the primes are uniquely determined up to associates (and reordering) as are the exponents a_i. The divisors of a are then determined uniquely (up to associates). Precisely, we have

$$d|a \iff d \sim p_1^{d_1} p_2^{d_2} \cdots p_r^{d_r}, \text{ for } 0 \leq d_i \leq a_i, \ i = 1, \ldots, r.$$

The proof of this is left as an exercise (or see [Nic12, §5.1, p. 258]).

Definition 3.1.19 (Greatest common divisor, least common multiple)**.** Suppose $s_1, \ldots, s_n \in R$. An element $d \in R$ is called a *greatest common divisor* (*gcd*) of $s_1, \ldots, s_n$, and is denoted $\gcd(s_1, \ldots, s_n)$, if it satisfies the following conditions:

(a) $d|s_i$ for $i = 1, 2, \ldots, n$.

(b) If $r \in R$ satisfies $r|s_i$ for $i = 1, 2, \ldots, n$, then $r|d$.

An element $m \in R$ is called a *least common multiple* (*lcm*) of $s_1, \ldots, s_n$, and is denoted $\operatorname{lcm}(s_1, \ldots, s_n)$ if it satisfies:

(a) $s_i|m$ for $i = 1, 2, \ldots, n$.

(b) If $r \in R$ satisfies $s_i|r$ for $i = 1, 2, \ldots, n$, then $m|r$.

Note that the definition of gcd given above agrees with usual definition of a gcd of integers and Definition 2.2.25 for polynomials, except that we require the gcd in $\mathbb{Z}$ to be positive and the gcd in $F[x]$ to be monic. These extra conditions ensure uniqueness of the gcd. In a

general UFD, we do not have any analogous condition to ensure uniqueness. Thus, in an arbitrary UFD, greatest common divisors (and least common multiples) are defined only up to associates. We write $\gcd(s_1, \dots, s_n)$ and $\operatorname{lcm}(s_1, \dots, s_n)$ to denote *any* gcd and lcm. In particular, we have:

$$\text{if } s_i \sim s_i' \text{ for } i = 1, \dots, n \quad \text{then} \quad \gcd(s_1, \dots, s_n) \sim \gcd(s_1', \dots, s_n').$$

Note that gcd's and lcm's need not exist in an arbitrary integral domain, as the following example illustrates.

Example 3.1.20. Consider the elements x^5, x^6 in the subring $\mathbb{Z}[x^2, x^3]$ of $\mathbb{Z}[x]$ consisting of all finite sums of the form $\sum a_{ij}(x^2)^i(x^3)^j$, $a_{ij} \in \mathbb{Z}$. We see that x^2 and x^3 are both common divisors of x^5 and x^6, but neither of x^2 or x^3 divides the other (in $\mathbb{Z}[x^2, x^3]$). Thus, the gcd of x^5 and x^6 in $\mathbb{Z}[x^2, x^3]$ does not exist.

Proposition 3.1.21. *Gcds and lcms exist in UFDs. Suppose R is a UFD and $a, b, c, \dots$ is a finite list of nonzero elements in R. Let $p_1, \dots, p_r$ be the nonassociated primes dividing one of $a, b, c, \dots$ and write*

$$\begin{aligned} a &\sim p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}, & a_i \in \mathbb{N}, \\ b &\sim p_1^{b_1} p_2^{b_2} \cdots p_r^{b_r}, & b_i \in \mathbb{N}, \\ c &\sim p_1^{c_1} p_2^{c_2} \cdots p_r^{c_r}, & c_i \in \mathbb{N}. \\ &\vdots \end{aligned}$$

For each $i = 1, 2, \dots, r$, let $d_i = \min(a_i, b_i, c_i, \dots)$ and $m_i = \max(a_i, b_i, c_i, \dots)$. Then

$$\gcd(a, b, c, \dots) \sim p_1^{d_1} p_2^{d_2} \cdots p_r^{d_r} \quad \text{and} \quad \operatorname{lcm}(a, b, c, \dots) \sim p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}.$$

Proof. The proof when $R = \mathbb{Z}$ carries over to the general case. The details are left as Exercise 3.1.7. □

Definition 3.1.22 (Ascending chain condition on principal ideals)**.** We say an integral domain R satisfies the *ascending chain condition on principal ideals* (or *ACCP*) if R contains no strictly increasing infinite chain

$$\langle a_1 \rangle \subsetneq \langle a_2 \rangle \subsetneq \langle a_3 \rangle \subsetneq \cdots$$

of principal ideals.

Example 3.1.23. Suppose we have an chain of strictly increasing chain of principal ideals in $\mathbb{Z}$:

$$\{0\} \neq \langle m_1 \rangle \subsetneq \langle m_2 \rangle \subsetneq \langle m_3 \rangle \subsetneq \cdots .$$

Then $|m_1| > |m_2| > |m_3| > \cdots$. Thus the chain must terminate after finitely many ideals. So $\mathbb{Z}$ satisfies the ACCP.

Example 3.1.24. Consider an ascending chain of principal ideals in $F[x]$, where F is a field:

$$\{0\} \neq \langle p_1(x)\rangle \subsetneq \langle p_2(x)\rangle \subsetneq \langle p_3(x)\rangle \subsetneq \cdots .$$

Then $\deg p_1(x) > \deg p_2(x) > \cdots$ and so the chain must be finite. Therefore $F[x]$ satisfies the ACCP.

Example 3.1.25. Let $R = \{n + xf(x) : n \in \mathbb{Z},\ f(x) \in \mathbb{Q}[x]\}$, the set of polynomials in $\mathbb{Q}[x]$ whose constant term is an integer. Since R is a subring of $\mathbb{Q}[x]$, it is an integral domain. However,

$$\langle x\rangle \subsetneq \langle \tfrac{1}{2}x\rangle \subsetneq \langle \tfrac{1}{2^2}x\rangle \subsetneq \cdots$$

is an infinite ascending chain of principal ideals in R. Thus R does not satisfy the ACCP.

Theorem 3.1.26. *Suppose R is an integral domain that satisfies the ACCP. Then every nonzero nonunit in R is a product of irreducibles.*

Proof. Suppose a nonzero nonunit $a \in R$ cannot be written as a product of irreducibles. Then a is not irreducible and so we have

$$a = r_1 a_1, \quad a \not\sim r_1, \quad a \not\sim a_1.$$

Now, at least one of r_1, a_1 cannot be written as a product of irreducibles (since if they both can, then so can a). Without loss of generality, we may assume that a_1 cannot be written as a product of irreducibles. Then we again have

$$a_1 = r_2 a_2, \quad a_1 \not\sim r_2, \quad a_1 \not\sim a_2,$$

where a_2 cannot be written as a product of irreducibles. Continuing in this manner, we have

$$\langle a\rangle \subsetneq \langle a_1\rangle \subsetneq \langle a_2\rangle \subsetneq \cdots .$$

Thus R does not satisfy the ACCP. □

Theorem 3.1.27. *The following are equivalent:*

(a) R is a UFD.

(b) R satisfies the ACCP and $\gcd(a, b)$ exists for all $a, b \in R$.

(c) R satisfies the ACCP and every irreducible element of R is prime.

Proof. (a) ⇒ (b): Assume R is a UFD. Then Proposition 3.1.21 shows that gcds exist in R. Now suppose we have an ascending chain of principal ideals in R:

$$\{0\} \subsetneq \langle a_1\rangle \subsetneq \langle a_2\rangle \subsetneq \cdots .$$

Consider a factorization $a_1 = p_1 \cdots p_m$ of a into irreducibles. Since $\langle a_1\rangle \subsetneq \langle a_2\rangle$, we have $a_2 \sim p_{i_1} \cdots p_{i_k}$, where $\{i_1, \ldots, i_k\} \subsetneq \{1, \ldots, m\}$. Similarly, we have $a_3 \sim p_{j_1} \cdots p_{j_\ell}$, where $\{j_1, \ldots, j_\ell\} \subsetneq \{i_1, \ldots, i_k\}$. Continuing in this manner, we see that the chain is finite. Thus R satisfies the ACCP.

(b) ⇒ (c): Suppose (b) holds. Suppose $p \in R$ is irreducible and $p|ab$ in R. Let $d = \gcd(a, p)$. Then $d|p$ and so $d \sim p$ or $d \sim 1$ (since p is irreducible). If $p \sim d$ then, since $d|a$, we have $p|a$. On the other hand, if $d \sim 1$, then $\gcd(a, p) \sim 1$. We claim that this implies that $\gcd(ab, pb) \sim b$. Assuming this claim, we have $p|b$ since $p|ab$ and $p|pb$. So we have shown that $p|a$ or $p|b$ and so p is prime.

It remains to prove the claim. Let $d' = \gcd(ab, pb)$. We have $b|ab$ and $b|pb$. Thus $b|d'$. Say $d' = bu$. We show that u is a unit. Write $ab = d'x$ for $x \in R$. So $ab = bux$. Thus $a = ux$, since $b \neq 0$. Thus $u|a$. Similarly, $u|p$. So $u| \gcd(a, p) \sim 1$. Hence u is a unit.

(c) ⇒ (a): Suppose (c) holds. By Theorem 3.1.26, every element is a product of irreducibles. So it remains to show the uniqueness of such factorizations. Suppose

$$p_1 p_2 \cdots p_r \sim q_1 q_2 \cdots q_s$$

are distinct factorizations, where the p_i and q_i are irreducibles and $r + s$ is minimal. If $r = 1$, then we have $p_1 \sim q_1 \cdots q_s$. Since p_1 is irreducible, this implies that $s = 1$, which contradicts the assumption that the factorizations are distinct. An analogous argument shows that we cannot have $s = 1$. Thus we may assume that $r, s \geq 2$. Since $p_1|q_1 \cdots q_s$, we must have $p_1|q_j$ for some j since p_1 is prime by assumption. Relabeling if necessary, we may assume that $p_1|q_1$. Since q_1 is irreducible, this implies that $p_1 \sim q_1$. Thus $p_2 \cdots p_r \sim q_2 \cdots q_s$ are distinct factorizations, contradicting the minimality of $r + s$. □

Example 3.1.28. The ring R of Example 3.1.25 is not a UFD since it does not satisfy the ACCP.

The remainder of this section is devoted to the proof of the important result that if R is a UFD, then so is $R[x]$ (see Theorem 3.1.34). Before proving this, we will need to define a few terms and prove some preliminary results.

Definition 3.1.29 (Content, primitive polynomial)**.** Suppose R is a UFD and $f(x) \in R[x] \setminus \{0\}$. Then the *content* of $f(x)$ is the gcd of the nonzero coefficients of f and is denoted $c(f(x))$. We say that $f(x)$ is a *primitive polynomial* if $c(f(x)) \sim 1$.

Example 3.1.30. In $\mathbb{Z}[x]$, $c(12 + 6x + 9x^4) = 3$. However, in $\mathbb{Q}[x]$, $c(12 + 6x + 9x^4) = 1$ (of course, it is also 3, since 1 and 3 are associates in $\mathbb{Q}$). Thus $12 + 6x + 9x^4$ is primitive in $\mathbb{Q}[x]$ but not in $\mathbb{Z}[x]$.

Lemma 3.1.31. *Let R be a UFD and let $f(x) \in R[x] \setminus \{0\}$.*

(a) $f(x)$ can be written as $f(x) = c(f(x))f_1(x)$, where $f_1(x) \in R[x]$ is primitive.

(b) If $a \in R \setminus \{0\}$, then $c(a(f(x))) = ac(f(x))$.

(c) If $f(x)$ is irreducible and $\deg f(x) \geq 1$, then $f(x)$ is primitive.

Proof. The proof of this lemma is left as Exercise 3.1.14. □

The following theorem is a generalization of Theorem 2.2.15

Theorem 3.1.32 (Gauss' Lemma)**.** *Suppose R is a UFD and $f(x), g(x) \in R[x] \setminus \{0\}$. Then*

$$c(f(x)g(x)) = c(f(x))c(g(x)).$$

In particular, the product of primitive polynomials is primitive.

Proof. Write $f = c(f)f_1$ and $g = c(g)g_1$, where $f_1, g_1 \in R[x]$ are primitive. Then

$$c(fg) \sim c(c(f)c(g)f_1g_1) \sim c(f)c(g)c(f_1g_1)$$

by Lemma 3.1.31. Thus, it suffices to prove the result when f and g are primitive.

Let $f = \sum_{i=0}^{m} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$. Suppose that p is a prime dividing the coefficients of fg. Let r be the smallest integer such that $p \nmid a_r$ and s be the smallest integer such that $p \nmid b_s$. The coefficient of x^{r+s} in fg is

$$c_{r+s} = a_0b_{r+s} + a_1b_{r+s-1} + \cdots + a_{r+s-1}b_1 + a_{r+s}b_0.$$

Since p divides $a_0, \ldots, a_{r-1}$ and $b_0, \ldots, b_{s-1}$, p divides every term of c_{r+s} except for the term a_rb_s. However, since $p|c_{r+s}$, either p divides a_r or p divides b_s. This contradiction completes the proof. □

Proposition 3.1.33. *Suppose R is a UFD and let F be its field of quotients. We regard R as a subring of F in the usual way (see Section 1.2). Suppose $p(x)$ is a primitive element of $R(x)$. Then $p(x)$ is irreducible in $R[x]$ if and only if it is irreducible in $F[x]$.*

Proof. First suppose that $p(x)$ is irreducible in $F[x]$. Let $p(x) = f(x)g(x)$ be a factorization in $R[x]$. If $\deg f(x), \deg g(x) \geq 1$, then this is a nontrivial factorization in $F[x]$, which contradicts the fact that $p(x)$ is irreducible in $F[x]$. Thus we can assume, without loss of generality, that $\deg f(x) = 0$. Hence $f(x) = u \in R$. Then, since $p(x)$ is primitive, we have

$$1 \sim c(p(x)) = c(f(x)g(x)) = c(f(x))c(g(x)) = uc(g(x)).$$

Hence u is a unit in R and so the factorization $p(x) = f(x)g(x)$ is trivial in $R[x]$. Hence $p(x)$ is irreducible in $R[x]$.

Now suppose $p(x)$ is primitive and irreducible in $R[x]$ and $p(x) = f(x)g(x)$ in $F[x]$. Let a and b be the products of the denominators of the coefficients of $f(x)$ and $g(x)$ respectively. Then $f_1(x) := af(x)$ and $g_1(x) := bg(x)$ are in $R[x]$ and

$$abp(x) = f_1(x)g_1(x)$$

is a factorization in $R[x]$. Thus, by Gauss' Lemma (Theorem 3.1.32), we have

$$ab \sim abc(p(x)) = c(abp(x)) = c(f_1(x)g_1(x)) \sim c(f_1(x))c(g_1(x)). \tag{3.2}$$

Now write $f_1(x) = c(f_1(x))f_2(x)$ and $g_1(x) = c(g_1(x))g_2(x)$, where $f_2(x)$ and $g_2(x)$ are primitive in $R[x]$. Thus

$$abp(x) = f_1(x)g_1(x) = c(f_1(x))c(g_1(x))f_2(x)g_2(x).$$

Therefore, by (3.2), we have $p(x) \sim f_2(x)g_2(x)$ in $R[x]$. Since $p(x)$ is irreducible in $R[x]$, this implies that either $f_2(x)$ or $g_2(x)$ is a unit in $R[x]$ (hence a unit in R). If $f_2(x) = u \in R^\times$, then

$$af(x) = f_1(x) = c(f_1(x))f_2(x) = c(f_1(x))u \implies f(x) = a^{-1}c(f_1(x))u \in F[x]^\times.$$

Similarly, if $g_2(x) \in R^\times$, then $g(x) \in F[x]^\times$. □

Theorem 3.1.34. *If R is a UFD, then so is $R[x]$.*

Proof. Let $p(x)$ be a nonzero polynomial in $R[x]$. If $p(x)$ is a constant polynomial, then it must have a unique factorization since R is a UFD. Now suppose that $p(x)$ is a polynomial of positive degree in $R[x]$. Let F be the field of fractions of R, and let

$$p(x) = f_1(x)f_2(x)\cdots f_n(x)$$

be a factorization of $p(x)$ in $F[x]$, where each $f_i(x)$ is irreducible in $F[x]$. Choose $a_i \in R$ such that $a_i f_i(x)$ is in $R[x]$ (e.g. take a_i to be the product of the denominators of the coefficients of $f_i(x)$). Then, for $i = 1, \ldots, n$, let $b_i = c(a_i f_i(x)) \in R$, so that $a_i f_i(x) = b_i g_i(x)$, where $g_i(x)$ is a primitive polynomial in $R[x]$. Since $g_i(x) = \frac{a_i}{b_i} f_i(x)$, the polynomials $f_i(x)$ and $g_i(x)$ are associates in $F[x]$. Thus, since $f_i(x)$ is irreducible in $F[x]$, so is $g_i(x)$. Then, by Proposition 3.1.33, each $g_i(x)$ is irreducible in $R[x]$. Now, we have

$$a_1 \cdots a_n p(x) = b_1 \cdots b_n g_1(x) \cdots g_n(x). \tag{3.3}$$

Since $g_1(x)\cdots g_n(x)$ is primitive, we have

$$a_1 \cdots a_n c(p(x)) = c(a_1 \cdots a_n p(x)) = c(b_1 \cdots b_n g_1(x) \cdots g_n(x)) = b_1 \cdots b_n.$$

Thus $a_1 \cdots a_n$ divides $b_1 \cdots b_n$. Therefore, dividing both sides of (3.3) by $a_1 \cdots a_n$, we have $p(x) = a g_1(x) \cdots g_n(x)$, where $a \in R$. Since R is a UFD, we can factor a as $c_1 \cdots c_k$, where each of the c_i's is irreducible in R (hence in $R[x]$). Thus we have proven the existence of factorizations into irreducibles.

We will now show the uniqueness of such factorizations. Let

$$p(x) = a_1 \cdots a_m f_1(x) \cdots f_n(x) = b_1 \cdots b_r g_1(x) \cdots g_s(x)$$

be two factorizations of $p(x)$, where all of the factors are irreducible in $R[x]$. By Proposition 3.1.33, each of the $f_i(x)$'s and $g_i(x)$'s is irreducible in $F[x]$. The a_i's and the b_i's are units in F. Since $F[x]$ is a UFD by Theorem 2.2.31, we have $n = s$.

Now rearrange the $g_i(x)$'s so that $f_i(x)$ and $g_i(x)$ are associates in $F[x]$ for $i = 1, \ldots, n$. Then there exist $c_1, \ldots, c_n$ and $d_1, \ldots, d_n$ in R such that $(c_i/d_i)f_i(x) = g_i(x)$ or, equivalently, $c_i f_i(x) = d_i g_i(x)$. The polynomials $f_i(x)$ and $g_i(x)$ are primitive; hence, c_i and d_i are associates in R. Thus, $a_1 \cdots a_m = u b_1 \cdots b_r$ in R, where u is a unit in R. Since R is a unique factorization domain, $m = s$. Finally, we can reorder the b_i's so that a_i and b_i are associates for each i. This completes the uniqueness part of the proof. □

Corollary 3.1.35. *If R is a UFD, then so is $R[x_1, \ldots, x_n]$*

Proof. This follows by an easy induction using Theorem 3.1.34 and the fact that

$$R[x_1, \ldots, x_n] = (R[x_1, \ldots, x_{n-1}])[x_n].$$ □

Remark 3.1.36 (Noetherian and artinian rings). A commutative ring that satisfies the *ascending chain condition on ideals* (the analogue of the ACCP, but where we do not specify that the ideals are principal) is called a *noetherian ring*. A commutative ring that satisfies the *descending chain condition on ideals* is called an *artinian ring*. Artinian rings are also noetherian (but not vice versa). Noetherian and artinian rings are important classes of rings, but we will not study these classes of rings (per se) in this course. In noncommutative rings, one can talk of left and right noetherian (or artinian) rings.

Exercises.

3.1.1. Prove Theorem 3.1.3.

3.1.2. Prove Lemma 3.1.5.

3.1.3. With $N(x)$ defined as in (3.1), prove that $N(x) = 0$ if and only if $x = 0$.

3.1.4 ([Nic12, Ex. 5.1.3]). In the ring $\mathbb{Z}(i)$ of Gaussian integers, show that

(a) $(2 + i) \sim (1 - 2i)$,

(b) $(1 + 2i) \not\sim (2 + i)$.

3.1.5 ([Nic12, Ex. 5.1.6]). Show that an integral domain is a field if and only if $a \sim b$ for all $a \neq 0 \neq b$.

3.1.6 ([Nic12, Ex. 5.1.8]). Find the units in $\mathbb{Z}(\sqrt{-3})$. *Hint*: Use $N(a + b\sqrt{-3}) = a^2 + 3b^2$ as in Example 3.1.14.

3.1.7. Prove Proposition 3.1.21.

3.1.8 ([Nic12, Ex. 5.1.10]). In each case, determine whether p is irreducible in $\mathbb{Z}(i)$:

(a) $p = 11$

(b) $p = 2 - i$

(c) $p = 5$

(d) $p = 7 - i$

3.1.9 ([Nic12, Ex. 5.1.12]). In each case, determine whether p is irreducible in $\mathbb{Z}(\sqrt{-5})$:

(a) $p = 6 + \sqrt{-5}$

(b) $p = 7$

(c) $p = 29$

(d) $p = 2 - 3\sqrt{-5}$.

3.1.10 ([Nic12, Ex. 5.1.13]). In each case, show that p is irreducible in $\mathbb{Z}(\sqrt{-5})$ but is not a prime.

(a) $p = 2 + \sqrt{-5}$

(b) $p = 1 + 2\sqrt{-5}$

3.1.11 ([Nic12, Ex. 5.1.14]). In each case, determine whether p is irreducible in $\mathbb{Z}(\sqrt{-3})$. *Hint*: Use the norm function $N(m + n\sqrt{-3}) = m^2 + 3n^2$.

(a) $p = 3 + 2\sqrt{-3}$

(b) $p = 2 + 3\sqrt{-3}$

(c) $p = 5$

(d) $p = 7$

3.1.12 ([Nic12, Ex. 5.1.15]). Show that $1 + \sqrt{-3}$ is irreducible in $\mathbb{Z}(\sqrt{-3})$ but is not prime.

3.1.13 ([Nic12, Ex. 5.1.16]). Let $p \sim q$ in the integral domain R.

(a) Show that p is irreducible if and only if q is irreducible.

(b) Show that p is prime if and only if q is prime.

3.1.14. Prove Lemma 3.1.31.

3.1.15 ([Nic12, Ex. 5.1.19]). A commutative ring is said to satisfy the *descending chain condition on principal ideals* (DCCP) if $\langle a_1 \rangle \supseteq \langle a_2 \rangle \supseteq \cdots$ in R implies that $a_n \sim a_{n+1} \sim \cdots$ for some $n \geq 1$. Show that an integral domain R satisfies the DCCP if and only if R is a field.

3.1.16 ([Nic12, Ex. 5.1.23]). If S is a UFD and R is a subring of S, is R necessarily a UFD? Justify your answer.

3.1.17 ([Nic12, Ex. 5.1.27]). Show that

$$\gcd(a, \gcd(b, c)) \sim \gcd(\gcd(a, b), c)$$

whenever all the gcds exist in R. Moreover, show that this common value is $\gcd(a, b, c)$.

3.1.18 ([Nic12, Ex. 5.1.36]). Show that if an integral domain R satisfies the ACCP and $\operatorname{lcm}(a, b)$ exists for all $a, b \neq 0$ in R, then R is a UFD. *Hint:* If $p|ab$, p irreducible, consider $m \sim \operatorname{lcm}(a, p)$. Use the fact that $m|ap$ and $m|ab$. (Of course, you must justify this fact in order to use it.)

3.1.19. Suppose R is an integral domain and $a, b \in R \setminus \{0\}$. Show that $\gcd(a, b)$ exists if and only if $\operatorname{lcm}(a, b)$ exists. Furthermore, show that if they exist, we have $\gcd(a, b) \operatorname{lcm}(a, b) \sim ab$.

3.1.20 (Bonus problem). Let $p(x) \in \mathbb{Z}[x]$. Assume that for every root $\alpha \in \mathbb{C}$ of $p(x)$ we have $|\alpha| = 1$. Prove that $p(x)|(x^m - 1)$ for some $m > 0$.

3.2 Principal ideal domains

Definition 3.2.1 (Principal ideal domain)**.** An integral domain R is called a *principal ideal domain* (*PID*) if every ideal of R is principal (i.e. generated by a single element).

Example 3.2.2. By Proposition 1.3.6 and Theorem 2.3.1, the rings $\mathbb{Z}$ and $F[x]$ (where F is a field) are PIDs. Any field is also a PID since its only ideals are $\langle 0\rangle$ and $\langle 1\rangle$, which are both principal.

Theorem 3.2.3. *Every PID is a UFD.*

Proof. Suppose R is a PID. By Theorem 3.1.27, it suffices to show that every irreducible element of R is prime and that R satisfies the ACCP.

Suppose $p \in R$ is irreducible and $p|ab$ in R. Let I be the ideal $Ra + Rp$. Then, since R is a PID, we have $I = \langle d\rangle$ for some $d \in R$. Thus $d|a$ and $d|p$. Since p is irreducible, this implies that $d \sim p$ or $d \sim 1$. If $d \sim p$, then $p|a$ (since $d|a$). On the other hand, if $d \sim 1$, then $1 \in I$. Thus there exist $r, s \in R$ such that $ra + sp = 1$. Then $rab + spb = b$. Since $p|ab$, this implies that $p|b$. Hence p is prime.

It remains to show that R satisfies the ACCP. Suppose

$$\langle a_1\rangle \subseteq \langle a_2\rangle \subseteq \cdots .$$

Let $A = \bigcup_{i=1}^{\infty}\langle a_i\rangle$. Then A is an ideal of R (exercise). Thus, since R is a PID, we have $A = \langle a\rangle$ for some $a \in R$. Then $a \in \langle a_i\rangle$ for some i. Hence $\langle a_i\rangle = \langle a\rangle = A$ and so $\langle a_n\rangle = A$ for $n \geq i$. So R satisfies the ACCP. □

The converse of Theorem 3.2.3 is false, as the following example illustrates.

Example 3.2.4. The ring $\mathbb{Z}[x]$ is a UFD (by Theorem 3.1.34) but not a PID since, for example, the ideal

$$I = \langle 2, x\rangle := \{a_0 + a_1x + \cdots + a_nx^n : a_0 \in 2\mathbb{Z},\ a_1, \ldots, a_n \in \mathbb{Z}\}$$

is not a principal ideal (Exercise 3.2.2).

Remark 3.2.5. The fact that $\mathbb{Z}$ is a PID, while $\mathbb{Z}[x]$ is not, shows that the analogue of Theorem 3.1.34 for PIDs does not hold.

Theorem 3.2.6. *If R is a PID, then the gcd of any nonzero elements $a_1, \ldots, a_n \in R$ exists and it can be written in the form $d = r_1a_1 + \cdots + r_na_n$ for some $r_1, \ldots, r_n \in R$.*

Proof. Let $I = \langle a_1, \ldots, a_n\rangle := \langle a_1\rangle + \cdots + \langle a_n\rangle$. Since R is a PID, we have $I = \langle d\rangle$ for some $d \in R$. Then $d = r_1a_1 + \cdots + r_na_n$ for some $r_1, \ldots, r_n \in R$ by the definition of I. Since $a_1, \ldots, a_n \in \langle d\rangle$, we have $d|a_i$ for all i. Furthermore, if $c|a_i$ for all i, then $c|(r_1a_1 + \cdots + r_na_n)$ and so $c|d$. Thus $d \sim \gcd(a_1, \ldots, a_n)$. □

Theorem 3.2.7. *Suppose R is a PID and $p \in R$, $p \neq 0$, $p \notin R^\times$. Then the following statements are equivalent:*

(a) p is a prime element.

(b) $\langle p \rangle$ *is a prime ideal.*

(c) $\langle p \rangle$ *is a maximal ideal.*

Proof. (a) $\Rightarrow$ (b): Suppose p is a prime element and $ab \in \langle p \rangle$. Then $p|ab$, which implies that $p|a$ or $p|b$. Thus $a \in \langle p \rangle$ or $b \in \langle p \rangle$. Thus $\langle p \rangle$ is a prime ideal.

(b) $\Rightarrow$ (c): Suppose $\langle p \rangle$ is a prime ideal. Then p is prime since

$$p|ab \implies ab \in \langle p \rangle \implies a \in \langle p \rangle \text{ or } b \in \langle p \rangle \implies p|a \text{ or } p|b.$$

Thus p is irreducible by Theorem 3.1.13. Now suppose $\langle p \rangle \subseteq \langle q \rangle \subsetneq R$. Then $q|p$. Since $\langle q \rangle \neq R$, we have $q \not\sim 1$. Since p is irreducible, this implies that $p \sim q$. Thus $\langle p \rangle = \langle q \rangle$.

(c) $\Rightarrow$ (a): Suppose $\langle p \rangle$ is a maximal ideal. Then it is a prime ideal by Corollary 1.3.20. Hence, by the above, we see that p is prime. □

Example 3.2.8. In $\mathbb{Z}[x]$, the ideal $\langle x \rangle$ is prime but $\langle x \rangle \subsetneq \langle 2, x \rangle \subsetneq \mathbb{Z}[x]$ and so this ideal is not maximal. Thus $\mathbb{Z}[x]$ is not a PID (see Example 3.2.4).

Example 3.2.9. Let F be a field. In $F[x, y]$, the ideal $\langle x \rangle$ is prime, but $\langle x \rangle \subsetneq \langle x, y \rangle \subsetneq F[x, y]$. Thus $F[x, y]$ is not a PID. However, $F[x, y]$ is a UFD by Corollary 3.1.35.

Proposition 3.2.10. *Suppose R is a PID and $a_1, \ldots, a_n \in R \setminus \{0\}$.*

(a) $d \sim \gcd(a_1, \ldots, a_n)$ *if and only if* $\langle a_1 \rangle + \langle a_2 \rangle + \cdots + \langle a_n \rangle = \langle d \rangle$.

(b) $m \sim \operatorname{lcm}(a_1, \ldots, a_n)$ *if and only if* $\langle a_1 \rangle \cap \langle a_2 \rangle \cap \cdots \cap \langle a_n \rangle = \langle m \rangle$.

Proof. Part (a) follows from Theorem 3.2.6. The proof of part (b) is left as Exercise 3.2.9. □

Exercises.

3.2.1. Show that if $I_1 \subseteq I_2 \subseteq \cdots$ is a chain of ideals in a ring R, then $I := \bigcup_{i=1}^{\infty} I_i$ is an ideal of R.

3.2.2. Show that $\langle 2, x \rangle \subseteq \mathbb{Z}[x]$ is not a principal ideal (see Example 3.2.4).

3.2.3 ([Nic12, Ex. 5.2.1]). Is every subring of a PID again a PID? Justify your answer.

3.2.4 ([Nic12, Ex. 5.2.2]). If F is a field, show that $I = \{f(x, y) \in F[x, y] : f(0, 0) = 0\}$ is an ideal of $F[x, y]$ that is not principal. This gives another proof that $F[x, y]$ is a UFD that is not a PID; see Example 3.2.9.

3.2.5 ([Nic12, Ex. 5.2.4]). Is $\mathbb{Z}(\sqrt{-5})$ a PID? Justify your answer.

3.2.6 ([Nic12, Ex. 5.2.5]). If R is a PID and $A \neq 0$ is an ideal of R, show that R/A has a finite number of ideals, all of which are principal.

3.2.7 ([Nic12, Ex. 5.2.6]). (a) Is every prime ideal of a PID maximal? Justify your answer.

(b) Show that every ideal $A \neq R$ in a PID R is contained in a maximal ideal of R.

3.2.8 ([Nic12, Ex. 5.2.8]). Let $p \in \mathbb{Z}$ be a prime number and define

$$\mathbb{Z}_{(p)} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z},\ n \notin p\mathbb{Z} \right\} \subseteq \mathbb{Q}.$$

(a) Show that $\mathbb{Z}_{(p)}$ is an integral domain (called the *localization* of $\mathbb{Z}$ at p) and find the units.

(b) If $A \neq 0$ is an ideal of $\mathbb{Z}_{(p)}$, show that $A = \langle p^k \rangle$, where $k \geq 0$ is the smallest integer such that $p^k \in A$. *Hint*: If $0 \neq m \in \mathbb{Z}$, then $m = p^r d$, where $r \geq 0$ and p does not divide d.

(c) Show that $\mathbb{Z}_{(p)}$ is a PID with exactly one maximal ideal.

3.2.9. Prove Proposition 3.2.10(b).

3.2.10 ([Nic12, Ex. 5.2.10]). Let R be a ring such that $\mathbb{Z} \subseteq R \subseteq \mathbb{Q}$. Show that R is a PID. *Hint:* If I is an ideal of R, consider $A = \mathbb{Z} \cap I$.

3.2.11 ([Nic12, Ex. 5.2.12]). Suppose $\omega \in \mathbb{C}$ satisfies $\omega^2 \in \mathbb{Z}$ and $\omega^2 < 0$. Show that $\mathbb{Z}(\omega) = \{a + b\omega : a, b \in \mathbb{Z}\}$ has finitely many units.

3.2.12 (Bonus problem). Suppose R is an integral domain and $a, b \in R \setminus \{0\}$. Show that gcds exist in R if and only if lcms exist. Furthermore, show that if they exist, we have $\gcd(a, b) \operatorname{lcm}(a, b) \sim ab$ for all $a, b \in R$.

3.3 Euclidean domains

Definition 3.3.1 (Division algorithm, divisor function)**.** If R is an integral domain, we say that R has a *division algorithm* if there exists a map $\nu \colon R \setminus \{0\} \to \mathbb{N}$ (called a *divisor function*) such that, for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ with

$$a = qb + r \text{ and either } r = 0 \text{ or } \nu(r) < \nu(b).$$

Definition 3.3.2 (Euclidean domain)**.** An integral domain R is called a *euclidean domain* if it has a divisor function $\nu \colon R \setminus \{0\} \to \mathbb{N}$ that satisfies the following condition:

$$a, b \in R \setminus \{0\} \implies \nu(ab) \geq \nu(a). \tag{3.4}$$

Such a divisor function is called a *euclidean valuation* (or *euclidean function*).

Remark 3.3.3. The function

$$\nu \colon \mathbb{Z} \setminus \{0\} \to \mathbb{N}, \quad \nu(k) = \begin{cases} k & \text{if } k > 0, \\ 2|k| & \text{if } k < 0. \end{cases}$$

is a divisor function that does not satisfy (3.4).

However, Condition (3.4) is actually superfluous in the following sense. Any integral domain R that has a division algorithm possesses a euclidean valuation (i.e. a divisor function satisfying (3.4)). Indeed, if ν is a divisor function for R, then

$$\nu'(a) = \min_{r \in R \setminus \{0\}} \nu(ra), \quad a \in R,$$

is a euclidean valuation on R. See [Rog71] for details.

Examples 3.3.4. (a) The ring $\mathbb{Z}$ is a euclidean domain with euclidean valuation $\nu\colon \mathbb{Z} \setminus \{0\} \to \mathbb{N}$, $\nu(n) = |n|$.

(b) For any field F, the ring $F[x]$ is a euclidean domain with euclidean valuation $\nu\colon F[x] \setminus \{0\} \to \mathbb{N}$, $\nu(f(x)) = \deg f(x)$ (see Theorem 2.1.12).

Theorem 3.3.5. *Every euclidean domain is a PID (and hence a UFD).*

Proof. Suppose R has a euclidean valuation ν, and let I be an ideal of R. If I is the zero ideal, then it is clearly principal. So we assume $I \neq \{0\}$. Choose $b \in I \setminus \{0\}$ with $\nu(b)$ minimal. We claim that $I = \langle b \rangle$. Since $b \in I$, we have $\langle b \rangle \subseteq I$. So it remains to show that $I \subseteq \langle b \rangle$. Suppose $a \in I$. Then we have

$$a = qb + r \quad \text{where } r = 0 \text{ or } \nu(r) < \nu(b).$$

Then $r = a - bq \in I$ and so, by the minimality of $\nu(b)$, we must have $r = 0$. Thus $a = qb$ and so $a \in \langle b \rangle$. □

Lemma 3.3.6. *The ring $\mathbb{Z}(i)$ of gaussian integers is a euclidean domain, hence a PID and a UFD.*

Proof. (See [Jud19, Example 18.20].) For $a, b \in \mathbb{Z}$, define $\nu(a + bi) = a^2 + b^2 = |a + bi|^2$. We claim that ν is a euclidean valuation on $\mathbb{Z}(i)$.

Let $z, w \in \mathbb{Z}(i) \setminus \{0\}$. Then $\nu(zw) = |zw|^2 = |z|^2|w|^2 = \nu(z)\nu(w)$. Since $\nu(z) \geq 1$ for every nonzero $z \in \mathbb{Z}(i)$, we have $\nu(zw) \geq \nu(w)$.

Next, we must show that for any $z = a + bi$ and $w = c + di$ in $\mathbb{Z}(i)$ with $w \neq 0$, there exist elements q and r in $\mathbb{Z}(i)$ such that $z = qw + r$ with either $r = 0$ or $\nu(r) < \nu(w)$. We can view z and w as elements in $\mathbb{Q}(i) = \{p + qi : p, q \in \mathbb{Q}\}$, the field of fractions of $\mathbb{Z}(i)$. Observe that

$$\begin{aligned} zw^{-1} &= (a + bi)\frac{c - di}{c^2 + d^2} \\ &= \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2} i \\ &= \left(m_1 + \frac{n_1}{c^2 + d^2}\right) + \left(m_2 + \frac{n_2}{c^2 + d^2}\right) i \\ &= (m_1 + m_2 i) + \left(\frac{n_1}{c^2 + d^2} + \frac{n_2}{c^2 + d^2} i\right) \\ &= (m_1 + m_2 i) + (s + ti) \end{aligned}$$

in $\mathbb{Q}(i)$. In the last steps we are writing the real and imaginary parts as an integer plus a proper fraction. That is, we take the closest integer m_i such that the fractional part satisfies $|n_i/(c^2+d^2)| \leq 1/2$. For example, we write

$$\frac{9}{8} = 1 + \frac{1}{8}$$
$$\frac{15}{8} = 2 - \frac{1}{8}.$$

Thus, s and t are the "fractional parts" of $zw^{-1} = (m_1 + m_2 i) + (s + ti)$. We also know that $s^2 + t^2 \leq 1/4 + 1/4 = 1/2$. Multiplying by w, we have

$$z = zw^{-1}w = w(m_1 + m_2 i) + w(s + ti) = qw + r,$$

where $q = m_1 + m_2 i$ and $r = w(s + ti)$. Since z and qw are in $\mathbb{Z}(i)$, r must be in $\mathbb{Z}(i)$. Finally, we need to show that either $r = 0$ or $\nu(r) < \nu(w)$. However,

$$\nu(r) = \nu(w)\nu(s + ti) \leq \frac{1}{2}\nu(w) < \nu(w). \qquad \square$$

Remark 3.3.7. It is not easy to find examples of PIDs that are not euclidean domains, but such rings exist. The first such example was given by T. Motzkin in 1949: $\mathbb{Z}\left(\frac{1}{2}\left(1 + \sqrt{-19}\right)\right)$ is a noneuclidean PID.

Summarizing some of the results proven in this chapter, we have the following chain of class inclusions:

commutative rings $\supsetneq$ integral domains $\supsetneq$ UFDs $\supsetneq$ PIDs $\supsetneq$ euclidean domains $\supsetneq$ fields.

Exercises.

3.3.1 ([Nic12, Ex. 5.2.7]). Show that the following conditions are equivalent for an integral domain R:

(a) R is a field,

(b) $R[x]$ is euclidean,

(c) $R[x]$ is a PID.

3.3.2 ([Nic12, Ex. 5.2.9]). Let $\mathbb{Z}_{(p)}$ be as in Exercise 3.2.8. Show that $\mathbb{Z}_{(p)}$ is a euclidean domain where, for each $a \neq 0$ in R, $\nu(a) = k$ where $\langle a \rangle = \langle p^k \rangle$. Indeed, show that $\nu(ab) = \nu(a) + \nu(b)$ for all $a \neq 0$, $b \neq 0$ in $\mathbb{Z}_{(p)}$ and that, if $a + b \neq 0$, then $\nu(a + b) \geq \min\{\nu(a), \nu(b)\}$.

3.3.3 ([Nic12, Ex. 5.2.15]). (a) Show that $\mathbb{Z}(\sqrt{3})$ is euclidean with $\nu(m + n\sqrt{3}) = |m^2 - 3n^2|$.

(b) If $a = 4 + 5\sqrt{3}$ and $b = 1 + \sqrt{3}$, write $a = qb + r$, where $r = 0$ or $\nu(r) < \nu(b)$.

3.3.4 ([Nic12, Ex. 5.2.16]). Show that $\mathbb{Z}(\sqrt{-3})$ is not euclidean with $\nu(m+n\sqrt{-3}) = m^2+3n^2$. In other words, show that ν is not a euclidean valuation. *Hint:* Try $a = 1 + \sqrt{-3}$ and $b = 2$.

3.3.5 ([Nic12, Ex. 5.2.18]). (a) If F is a field, show that F is euclidean.

(b) If the mapping ν is constant in a euclidean domain R, show that R is a field.

3.3.6 ([Nic12, Ex. 5.2.22]). Assume R is a euclidean domain in which $\nu(a+b) \leq \max\{\nu(a), \nu(b)\}$ whenever a, b, and $a + b$ are nonzero. Show that q and r are uniquely determined in the division algorithm.

3.3.7 ([Nic12, Ex. 5.2.23]). Suppose that a euclidean domain R has a unique maximal ideal P. By Theorem 3.3.5, we can write $P = \langle p \rangle$ for some $p \in R$.

(a) Show that P consists of the nonunits of R; that is, a is a nonunit if and only if $p|a$. *Hint*: Exercise 3.2.7.

(b) Show that every ideal $A \neq \{0\}$ of R has the form $A = \langle p^k \rangle$ for some $k \geq 0$.

Chapter 4

Modules

The notion of a module is a natural extension of that of a vector space. The difference is that the scalars no longer lie in a field but rather in a ring. This permits one to encompass more objects. For example, we will see that an abelian group is a module over the ring $\mathbb{Z}$ of integers. We will also see that if V is a vector space over a field F equipped with a linear operator, then V is naturally a module over the ring $F[x]$ of polynomials in one variable over F. In this chapter, we introduce in a general manner the theory of modules, together with the notions of submodule, quotient module, free module, direct sum, homomorphism, and isomorphism of modules.

4.1 The notion of a module

Definition 4.1.1 (Module)**.** A *module* over a ring R (or, an *R-module*) is an abelian group $(M, +)$, together with a map (called an *action*)

$$R \times M \mapsto M, \quad (r, u) \mapsto ru,$$

such that

(M1) $r(u + v) = ru + rv$,

(M2) $(r + s)v = rv + sv$,

(M3) $r(sv) = (rs)v$,

(M4) $1v = v$,

for all $u, v \in M$, $r, s \in R$. (Above, $1 = 1_R$ is the unit element of the ring R.) We call $(M, +)$ the (underlying) *additive group* of M.

What we've defined in Definition 4.1.1 is actually a *left module*. There is a corresponding notion of *right module*. However, in this course we'll restrict our attention to left modules.

If we compare the definition of a module over the ring R to that of a vector space over a field F, we recognize the same axioms. Since a field is a particular type of commutative ring, we deduce the following.

Example 4.1.2. A module over a field F is simply a vector space over F.

Since M is an abelian group under addition, the order in which we add elements $v_1, \ldots, v_n$ of M does not affect their sum, denoted

$$v_1 + \cdots + v_n \quad \text{or} \quad \sum_{i=1}^{n} v_i.$$

The identity element for $(M, +)$ is denoted 0, or sometimes 0_M when we want to avoid confusion with the zero element of the ring R. It is called the *zero* of M. We define subtraction in M by

$$u - v := u + (-v), \quad u, v \in M,$$

where $-v$ is the additive inverse of v (which exists by our assumption that $(M, +)$ is an abelian group). We also verify, by induction on n, that the axioms of distributivity, (M1) and (M2), imply in general

$$\left(\sum_{i=1}^{n} r_i\right) v = \sum_{i=1}^{n} r_i v \quad \text{and} \quad r \sum_{i=1}^{n} v_i = \sum_{i=1}^{n} r v_i \tag{4.1}$$

for all $r, r_1, \ldots, r_n \in R$ and $v, v_1, \ldots, v_n \in M$.

Lemma 4.1.3. *Let M be a module over a ring R. For all $r \in R$ and all $v \in M$, we have*

(a) $0v = 0$ *and* $r0 = 0$,

(b) $(-r)v = r(-v) = -(rv)$.

Proof. In R, we have $0 + 0 = 0$. Then axiom (M2) gives

$$0v = (0 + 0)v = 0v + 0v.$$

Adding $-(0v)$ to both sides gives $0 = 0v$. More generally, since $r + (-r) = 0$, axiom (M2) also gives

$$0v = (r + (-r))v = rv + (-r)v.$$

Since $0v = 0$, this implies that $(-r)v = -(rv)$.

The relations $r0 = 0$ and $r(-v) = -(rv)$ are proven in a similar manner by using (M1) instead of (M2); see Exercise 4.1.1. □

Suppose that M is a $\mathbb{Z}$-module. Then M is an abelian group under addition and, for every integer $n \geq 1$ and for all $v \in M$, (4.1) gives

$$nv = (\underbrace{1 + \cdots + 1}_{n \text{ summands}})v = \underbrace{1v + \cdots + 1v}_{n \text{ summands}} = \underbrace{v + \cdots + v}_{n \text{ summands}}.$$

Thanks to Lemma 4.1.3, we also find that

$$(-n)v = n(-v) = \underbrace{(-v) + \cdots + (-v)}_{n \text{ summands}} \quad \text{and} \quad 0v = 0.$$

In other words, the action of the integers in M is entirely determined by the addition law in M and corresponds to the natural notion of product by an integer in an abelian group. Conversely, if M is an abelian group, defining an action by $\mathbb{Z}$ as above gives an action that satisfies axioms (M1)–(M3). Since it also clearly satisfies axiom (M4), we conclude the following.

Proposition 4.1.4. *A $\mathbb{Z}$-module is simply an abelian group equipped with the natural multiplication by the integers.*

The following proposition gives a third important example of modules.

Proposition 4.1.5. *Let V be a vector space over a field F and let $T \in \mathrm{End}_F(V)$ be a linear operator on V. Then V is an $F[x]$-module for the addition of V and the action of polynomials given by*

$$p(x)v := p(T)(v)$$

for all $p(x) \in F[x]$ and all $v \in V$.

Proof. Since V is an abelian group under addition, it suffices to verify axioms (M1)–(M4). Let $p(x), q(x) \in F[x]$ and $u, v \in V$. We must show:

- $1v = v$,
- $p(x)(q(x)v) = (p(x)q(x))v$,
- $(p(x) + q(x))v = p(x)v + q(x)v$,
- $p(x)(u + v) = p(x)u + p(x)v$.

Given the definition of the action, this reduces to showing

(a) $I_V(v) = v$,

(b) $p(T)(q(T)v) = (p(T) \circ q(T))v$,

(c) $(p(T) + q(T))v = p(T)v + q(T)v$,

(d) $p(T)(u + v) = p(T)u + p(T)v$,

because the map

$$F[x] \to \mathrm{End}_F(V), \quad p(x) \mapsto p(T),$$

of evaluation at T, being a ring homomorphism, maps 1 to I_V, $p(x)+q(x)$ to $p(T)+q(T)$ and $p(x)q(x)$ to $p(T) \circ q(T)$. Equalities (a) to (c) follow from the definitions of I_V, $p(T) \circ q(T)$ and $p(T)+q(T)$ respectively. Finally, (d) follows from the fact that $p(T)$ is a linear map. □

In the context of Proposition 4.1.5, we also note that, for a constant polynomial $p(x) = a$ with $a \in F$, we have

$$p(x)v = (aI)(v) = av$$

for all $v \in V$. Thus the notation av has the same meaning if we consider a as a polynomial in $F[x]$ or as a scalar of the field F. We can then state the following complement to Proposition 4.1.5.

Proposition 4.1.6. *In the notation of Proposition 4.1.5, the action by the elements of $F[x]$ on V extends the scalar multiplication by the elements of F.*

Example 4.1.7. Let V be a vector space of dimension 2 over $\mathbb{Q}$ equipped with a basis $\mathcal{B} = \{v_1, v_2\}$ and let $T\colon V \to V$ be the linear operator determined by

$$T(v_1) = v_1 - v_2, \quad T(v_2) = v_1.$$

We equip V with the structure of a $\mathbb{Q}[x]$-module associated to the endomorphism T and calculate

$$(2x - 1)v_1 + (x^2 + 3x)v_2.$$

By definition, we have

$$\begin{aligned}(2\,T - I)(v_1) + (T^2 + 3\,T)(v_2) &= 2\,T(v_1) - v_1 + T^2(v_2) + 3\,T(v_2)\\ &= 3\,T(v_1) - v_1 + 3\,T(v_2) \quad (\text{since } T^2(v_2) = T(T(v_2)) = T(v_1))\\ &= 3(v_1 - v_2) - v_1 + 3v_1\\ &= 5v_1 - 3v_2.\end{aligned}$$

Fix an arbitrary ring R. We conclude with four general examples of R-modules. The details of the proofs are left as exercises.

Example 4.1.8 (Left regular module). The ring R is an R-module for its addition law and the action

$$R \times R \to R, \quad (r, v) \mapsto rv$$

given by the product in R. This module is called the *left regular module* for R. Whenever we refer to R as an R-module, we are using this action.

Example 4.1.9. Let $n \in \mathbb{N}_{>0}$. The set

$$R^n = \left\{ \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} : r_1, r_2, \ldots, r_n \in R \right\}$$

of n-tuples of elements of R is an R-module for the operations

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} + \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} r_1 + s_1 \\ r_2 + s_2 \\ \vdots \\ r_n + s_n \end{bmatrix} \quad \text{and} \quad t \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} tr_1 \\ tr_2 \\ \vdots \\ tr_n \end{bmatrix}.$$

For $n = 1$, we recover the natural structure of R-module on $R^1 = R$ described by Example 4.1.8.

Example 4.1.10. Recall from Example 1.1.5 that $\mathrm{Mat}_n(R)$ is a ring, with the usual rules for matrix addition and multiplication. Then R^n is a $\mathrm{Mat}_n(R)$-module, with action given by left matrix multiplication.

Example 4.1.11. Let $M_1, \dots, M_n$ be R-modules. Their cartesian product

$$M_1 \times \cdots \times M_n = \{(v_1, \dots, v_n) : v_1 \in M_1, \dots, v_n \in M_n\}$$

is an R-module for the operations

$$(v_1, \dots, v_n) + (y_1, \dots, y_n) = (v_1 + y_1, \dots, v_n + y_n),$$
$$c(v_1, \dots, v_n) = (cv_1, \dots, cv_n).$$

Remark 4.1.12. If we apply the construction of Example 4.1.11 with $M_1 = \cdots = M_n = R$ for the natural R-module structure on R given by Example 4.1.8, we obtain the structure of an R-module on $R \times \cdots \times R = R^n$ that is essentially that of Example 4.1.9 except that the elements of R^n are presented in rows. For other applications, it is nevertheless more convenient to write the elements of R^n in columns.

Remark 4.1.13. We will sometimes write $M_1 \oplus \cdots \oplus M_n$ instead of $M_1 \times \cdots \times M_n$. See Example 4.5.4.

Exercises.

4.1.1. Complete the proof of Lemma 4.1.3 by showing that $a0 = 0$ and $a(-v) = -(av)$.

4.1.2. Let V be a vector space of dimension 3 over $\mathbb{Z}_5$ equipped with a basis $\mathcal{B} = \{v_1, v_2, v_3\}$ and let $T\colon V \to V$ be a linear operator given by

$$T(v_1) = v_1 + \bar{2}v_2 - v_3, \quad T(v_2) = v_2 + \bar{4}v_3, \quad T(v_3) = v_3.$$

For the corresponding structure of an $\mathbb{Z}_5[x]$-module on V, calculate

(a) xv_1,
(b) $xv_1 - (x + \bar{1})v_2$,
(c) $(\bar{2}x^2 - \bar{3})v_1 + (\bar{3}x)v_2 + (x^2 + \bar{2})v_3$.

4.1.3. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator whose matrix in the standard basis $\mathcal{E} = \{e_1, e_2\}$ is $[T]_{\mathcal{E}} = \begin{bmatrix} 2 & -1 \\ 3 & 0 \end{bmatrix}$. For the associated $\mathbb{R}[x]$-module structure on $\mathbb{R}^2$, calculate

(a) xe_1,
(b) $x^2e_1 + e_2$,
(c) $(x^2 + x + 1)e_1 + (2x - 1)e_2$.

4.1.4. Let $n \in \mathbb{N}_{>0}$ and let R be a ring. Show that R^n is an R-module for the operations defined in Example 4.1.9.

4.1.5. Let $M_1, \ldots, M_n$ be modules over a ring R. Show that their cartesian product $M_1 \times \cdots \times M_n$ is an R-module for the operations defined in Example 4.1.11.

4.1.6. Let M be a module over the ring R and let X be a nonempty set. For all $r \in R$ and every pair of functions $f\colon X \to M$ and $g\colon X \to M$, we define $f+g\colon X \to M$ and $rf\colon X \to M$ by setting

$$(f+g)(x) = f(x) + g(x) \quad \text{and} \quad (rf)(x) = rf(x)$$

for all $x \in X$. Show that the set $\mathcal{F}(X, M)$ of functions from X to M is an R-module for these operations.

4.2 Submodules

Fix an arbitrary module M over an arbitrary ring R.

Definition 4.2.1 (Submodule)**.** An *R-submodule* of M is a subset N of M satisfying the following conditions:

(SM1) $0 \in N$,

(SM2) If $u, v \in N$, then $u + v \in N$.

(SM3) If $r \in R$ and $u \in N$, then $ru \in N$.

Note that condition (SM3) implies that $-v = (-1)v \in N$ for all $v \in N$ which, together with conditions (SM1) and (SM2), implies that an R-submodule of M is, in particular, a subgroup of M.

Conditions (SM2) and (SM3) require that an R-submodule N of M be stable under the addition in M and the action by elements of R on M. By restriction, these operations therefore define an addition on N and an action by elements of R on N. We leave it as an exercise to verify that these operations satisfy the axioms of Definition 4.1.1 required for N to be an R-module.

Proposition 4.2.2. *An R-submodule N of M is itself an R-module for the addition and action by the elements of R restricted from M to N.*

Any R-module M has the zero submodule $\{0\}$ and the submodule M itself. We say a submodule N of M is *proper* if $N \neq M$.

Definition 4.2.3 (Simple module)**.** A nonzero R-module M is *simple* if it has no nonzero proper submodules. In other words, M is simple if $M \neq \{0\}$ and the only submodules of M are $\{0\}$ and M.

Example 4.2.4. Let V be a vector space over a field F. An F-submodule of V is simply a vector subspace of V. The vector space V is simple if and only it is one-dimensional.

Example 4.2.5. Let M be an abelian group. A $\mathbb{Z}$-submodule of M is simply a subgroup of M. An abelian group is a simple $\mathbb{Z}$-module if and only if it is a simple group.

Example 4.2.6. Let R be a ring considered as a module over itself (i.e. the left regular module of Example 4.1.8). An R-submodule of R is simply a left ideal of R.

In the case of a vector space V equipped with an endomorphism, the concept of submodule translates as follows:

Proposition 4.2.7. *Let V be a vector space over a field F and let $T \in \mathrm{End}_F(V)$. For the corresponding structure of a $F[x]$-module on V, a $F[x]$-submodule of V is simply a T-*invariant *vector subspace of V, that is, a vector subspace U of V such that $T(u) \in U$ for all $u \in U$.*

Proof. Since the action by elements of $F[x]$ on V extends the scalar multiplication by the elements of F (see Proposition 4.1.6), an $F[x]$-submodule U of V is necessarily a vector subspace of V. Since it also satisfies $xu \in U$ for all $u \in U$, and $xu = T(u)$ (by definition), we also see that U must be T-invariant.

Conversely, if U is a T-invariant vector subspace of V, we have $T(u) \in U$ for all $u \in U$. By induction on n, we deduce that $T^n(u) \in U$ for every integer $n \geq 1$ and all $u \in U$. Therefore, for every polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n \in F[x]$ and $u \in U$, we have

$$p(x)u = (a_0 I + a_1 T + \cdots + a_n T^n)(u) = a_0 u + a_1 T(u) + \cdots + a_n T^n(u) \in U.$$

Thus U is an $F[x]$-submodule of V: we have just showed that it satisfies condition (SM3) and the other conditions (SM1) and (SM2) are satisfied by virtue of the fact that U is a subspace of V. □

We conclude this section with several valuable general constructions for an arbitrary R-module.

Proposition 4.2.8. *Let S be a subset of an R-module M. The set*

$$\{r_1 u_1 + \cdots + r_k u_k : k \in \mathbb{N},\ r_1, \ldots, r_k \in R,\ u_1, \ldots, u_k \in S\}$$

(we allow the possibility that $k = 0$, in which case the above sum is the zero element of M) is an R-submodule of M that contains S and it is the smallest R-submodule of M with this property (meaning that every R-submodule of M containing S contains this one).

Proof. The proof of this result is left as Exercise 4.2.3. □

The R-submodule from Proposition 4.2.8 is called the R-submodule of M *generated* (or *spanned*) by S. It is denoted

$$\langle S \rangle_R \quad \text{or} \quad RS \quad \text{or} \quad \mathrm{Span}_R S.$$

If $S = \{u_1, \ldots, u_k\}$, we also denote it

$$\langle u_1, \ldots, u_k \rangle_R \quad \text{or} \quad Ru_1 + \cdots + Ru_k \quad \text{or} \quad \mathrm{Span}_R\{u_1, \ldots, u_k\}.$$

Definition 4.2.9 (Cyclic module and finite type module)**.** We say that an R-module M (or an R-submodule of M) is *cyclic* if it is generated by a single element. We say that it is of *finite type* (or is *finitely generated*) if it is generated by a finite number of elements of M.

Example 4.2.10. If $u_1, \dots, u_k$ are elements of an abelian group M, then $\langle u_1, \dots, u_k \rangle_{\mathbb{Z}}$ is simply the subgroup of M generated by $u_1, \dots, u_k$. In particular, M is cyclic as an abelian group if and only if it is cyclic as a $\mathbb{Z}$-module.

Example 4.2.11. Let V be a finite-dimensional vector space over a field F, let $\{v_1, \dots, v_k\}$ be a set of generators of V, and let $T \in \mathrm{End}_F(V)$. Since the structure of a $F[x]$-module on V associated to T extends that of an F-module, we have

$$V = \langle v_1, \dots, v_k \rangle_F \subseteq \langle v_1, \dots, v_k \rangle_{F[x]},$$

hence $V = \langle v_1, \dots, v_k \rangle_{F[x]}$ is also generated by $v_1, \dots, v_k$ as an $F[x]$-module. In particular, V is an $F[x]$-module of finite type.

Proposition 4.2.12. *Let $N_1, \dots, N_k$ be submodules of an R-module M. Their* intersection $N_1 \cap \cdots \cap N_k$ *and their* sum

$$N_1 + \cdots + N_k := \{u_1 + \cdots + u_k : u_1 \in N_1, \dots, u_k \in N_k\}$$

are also R-submodules of M.

Proof for the sum. We first note that, since $0 \in N_i$ for $i = 1, \dots, s$, we have

$$0 = 0 + \cdots + 0 \in N_1 + \cdots + N_k.$$

Let u, u' be arbitrary elements of $N_1 + \cdots + N_k$, and let $r \in R$. We can write

$$u = u_1 + \cdots + u_k \quad \text{and} \quad u' = u'_1 + \cdots + u'_k$$

with $u_1, u'_1 \in N_1, \dots, u_k, u'_k \in N_k$. Thus we have

$$\begin{aligned} u + u' &= (u_1 + u'_1) + \cdots + (u_k + u'_k) \in N_1 + \cdots + N_k, \\ ru &= (ru_1) + \cdots + (ru_k) \in N_1 + \cdots + N_k. \end{aligned}$$

This proves that $N_1 + \cdots + N_k$ is an R-submodule of M. □

We note that the notation $Ru_1 + \cdots + Ru_k$ for the R-submodule $\langle u_1, \dots, u_k \rangle_R$ generated by the elements $u_1, \dots, u_s$ of M is consistent with the notion of a sum of submodules since, for every element u of M, we have

$$Ru = \langle u \rangle_R = \{ru : r \in R\}.$$

When $R = F$ is a field, we recover the usual notions of the intersection and sum of subspaces of a vector space over F. When $R = \mathbb{Z}$, we instead recover the notions of the intersection and sum of subgroups of an abelian group.

Definition 4.2.13 (Annihilator)**.** Suppose M is an R-module and X is a subset of M. The *annihilator* of X is

$$\mathrm{ann}(X) := \{r \in R : ru = 0 \text{ for all } u \in X\}.$$

If $X = \{u_1, \dots, u_k\}$, we sometimes write $\mathrm{ann}(u_1, \dots, u_k)$ instead of $\mathrm{ann}(\{u_1, \dots, u_k\})$.

When we consider R as a module over itself as in Example 4.1.8, the above notion of annihilator coincides with that of Exercise 1.3.13.

Exercises.

4.2.1. Prove Proposition 4.2.2.

4.2.2. Let M be an abelian group. Explain why the $\mathbb{Z}$-submodules of M are simply the subgroups of M. *Hint*: Show that every $\mathbb{Z}$-submodule of M is a subgroup of M and, conversely, that every subgroup of M is a $\mathbb{Z}$-submodule of M.

4.2.3. Prove Proposition 4.2.8.

4.2.4. Let M be a module over a ring R, and let $u_1, \ldots, u_k, v_1, \ldots, v_m \in M$. Set $L = \langle u_1, \ldots, u_k \rangle_R$ and $N = \langle v_1, \ldots, v_m \rangle_R$. Show that $L + N = \langle u_1, \ldots, u_k, v_1, \ldots, v_m \rangle_R$.

4.2.5. Prove that an R-module M is simple if and only if it is generated by any nonzero element. In other words, prove that M is simple if and only if $\langle u \rangle_R = M$ for all $u \in M$, $u \neq 0$.

4.2.6. Let M be a module over a ring R. We say that an element u of M is *torsion* if there exists a nonzero element r of R such that $ru = 0$. We say that M is a *torsion R-module* if all its elements are torsion. We say it is *torsion free* if zero is its only torsion element. Suppose that R is an integral domain.

(a) Show that the set M_{tor} of torsion elements of M is an R-submodule of M.

(b) Show that, if M is of finite type and if all its elements are torsion, then there exists a nonzero element r of R such that $ru = 0$ for all $u \in M$.

4.2.7. Suppose M is an R-module and X is a subset of M.

(a) Show that $\operatorname{ann}(X)$ is a left ideal of R. It follows that, if R is commutative, then $\operatorname{ann}(X)$ is a two-sided ideal of R.

(b) Give an example showing that $\operatorname{ann}(X)$ is not always a two-sided ideal of R.

(c) Show that, if R is commutative, then $\operatorname{ann}(X) = \operatorname{ann}(RX)$.

(d) Give an example where $\operatorname{ann}(X) \neq \operatorname{ann}(RX)$.

4.3 Quotient modules

If N is a submodule of an R-module M, then we can consider the quotient M/N of the corresponding additive groups.

Theorem 4.3.1. *If N is a submodule of an R-module M, then M/N is an R-module with action given by*

$$r(u + N) = ru + N, \quad r \in R,\ u \in M.$$

Proof. We leave it as Exercise 4.3.1 to verify that the action is well defined (i.e. independent of the representative u) and that it satisfies the conditions of Definition 4.1.1. □

Example 4.3.2. If M is an abelian group, then a $\mathbb{Z}$-submodule is simply a subgroup N. (Recall that, for abelian groups, all subgroups are normal.) Then the notion of quotient module coincides with the notion of quotient group.

Example 4.3.3. As in Example 4.2.6, the submodules of the left regular module of R are the left ideals of R. Thus, for any left ideal I of R, we have the quotient R-module R/I.

Example 4.3.4. If V is a vector space over a field F, then F-submodules are simply vector subspaces. For any vector subspace U of V we thus have the quotient vector space V/U.

Exercises.

4.3.1. Prove Theorem 4.3.1.

4.3.2. Suppose that R is an integral domain and that M is an R-module. Recall the definition of the torsion submodule M_{tor} of M (Exercise 4.2.6). Show that M/M_{tor} is torsion free.

4.4 Free modules

Definition 4.4.1 (Linear independence, basis, free module)**.** Let S be a subset of an R-module M.

(a) We say that the elements of S are *linearly independent* over R if the only choice of elements $r_1, \dots, r_n \in R$ and distinct $u_1, \dots, u_n \in S$ satisfying

$$r_1 u_1 + \cdots + r_n u_n = 0$$

is when $r_1 = \cdots = r_n = 0$.

(b) We say that S is a *basis* of M if its elements are linearly independent over R and generate M as an R-module.

(c) We say that M is a *free* R-module if it admits a basis.

By convention, the module $\{0\}$ consisting of a single element is a free R-module that admits the empty set as a basis.

Example 4.4.2. Let $n \in \mathbb{N}_{>0}$. The R-module R^n introduced in Example 4.1.9 admits the basis

$$\left\{ e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Indeed, we have

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = r_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + r_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + r_n \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{4.2}$$

for all $r_1, \ldots, r_n \in R$. Thus $e_1, \ldots, e_n$ generate R^n as an R-module. Furthermore, if $r_1, \ldots, r_n \in R$ satisfy $r_1 e_1 + \cdots + r_n e_n = 0$, then equality (4.2) gives

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and so $r_1 = r_2 = \cdots = r_n = 0$. Hence $e_1, \ldots, e_n$ are also linearly independent over R.

Lemma 4.4.3. *Any basis of a free R-module M is a minimal generating set.*

Proof. Suppose B is a basis of M and that B_0 is a proper subset of B. Let $b \in B \setminus B_0$. If b were contained in $\langle B_0 \rangle_R$, then b could be expressed as an R-linear combination of elements of B_0, contradicting the linear independence of B. Therefore $b \notin \langle B_0 \rangle_R$, and B_0 does not generate M. □

Lemma 4.4.4. *Any basis of a finitely-generated free R-module is finite.*

Proof. Suppose M is an R-module with a (possibly infinite) basis B and a finite generating set S. Each element of S is a linear combination of finitely many elements of B. Since S is finite, it is contained in the span of a finite subset B_0 of B. But then $M = \langle S \rangle_R \subseteq \langle \langle B_0 \rangle_R \rangle_R = \langle B_0 \rangle_R$. Thus B_0 spans M. By Lemma 4.4.3, $B = B_0$. □

In linear algebra, you learned that every vector space has a basis and that all bases have the same cardinality. You then define the *dimension* of a vector space to be the cardinality of any basis. When working with the more general concept of modules, you have to be more careful. First of all, *not* all modules are free, hence not all modules have a basis. See, for example, Exercise 4.4.1. Secondly, even if an R-module M is free, it may have bases of different cardinalites! See, for example, Exercise 4.8.2.

Exercises.

4.4.1. Show that a finite abelian group M admits a basis as a $\mathbb{Z}$-module if and only if $M = \{0\}$ (in which case the empty set is, by convention, a basis of M).

4.4.2. In general, show that, if M is a finite module over an infinite ring R, then M is a free R-module if and only if $M = \{0\}$.

4.4.3. Let V be a finite-dimensional vector space over a field F and let $T \in \mathrm{End}_F(V)$. Show that, for the corresponding structure of a $F[x]$-module, V admits a basis if and only if $V = \{0\}$.

4.4.4. Let m and n be positive integers and let R be a ring. Show that $\mathrm{Mat}_{m\times n}(R)$, the set of $m \times n$ matrices with entries in R, is a free R-module for the usual operations of addition and multiplication by elements of R.

4.5 Direct sum

Definition 4.5.1 (Direct sum)**.** Let $M_1, \ldots, M_k$ be submodules of an R-module M. We say that their sum is *direct* if the only choice of $u_1 \in M_1, \ldots, u_k \in M_k$ such that

$$u_1 + \cdots + u_k = 0$$

is $u_1 = \cdots = u_k = 0$. When this condition is satisfied, we write the sum $M_1 + \cdots + M_k$ as $M_1 \oplus \cdots \oplus M_k$. We say that M is the *direct sum* of $M_1, \ldots, M_k$ if $M = M_1 \oplus \cdots \oplus M_k$.

The following propositions, which generalize the analogous results for direct sums of vector spaces are left as exercises. The first justifies the importance of the notion of direct sum.

Proposition 4.5.2. *Let $M_1, \ldots, M_k$ be submodules of an R-module M. Then we have $M = M_1 \oplus \cdots \oplus M_k$ if and only if, for all $u \in M$, there exists a unique choice of $u_1 \in M_1, \ldots, u_k \in M_k$ such that $u = u_1 + \cdots + u_k$.*

Proposition 4.5.3. *Let $M_1, \ldots, M_k$ be submodules of an R-module M. The following conditions are equivalent:*

(a) the sum of $M_1, \ldots, M_k$ is direct;

(b) $M_i \cap (M_1 + \cdots + \widehat{M_i} + \cdots + M_k) = \{0\}$ for $i = 1, \ldots, k$;

(c) $M_i \cap (M_{i+1} + \cdots + M_k) = \{0\}$ for $i = 1, \ldots, k-1$.

(The notation $\widehat{M_i}$ means that we omit the term M_i from the sum.)

Example 4.5.4. Suppose $M_1, \ldots, M_k$ are R-modules, and consider their cartesian product

$$M = M_1 \times \cdots \times M_k$$

as in Example 4.1.11. For $1 \le i \le k$, let

$$M_i' = \{(0, \ldots, 0, u, 0, \ldots, 0) : u \in M_i\},$$

where the u is in the i-th component. Then M_i' is a submodule of M, and we have an R-module isomorphism

$$M_i \to M_i', \quad u \mapsto (0, \ldots, 0, u, 0, \ldots, 0), \tag{4.3}$$

and

$$M_1 \times \cdots \times M_k = M_1' \oplus \cdots \oplus M_k'.$$

(We will give the precise definition of an R-module isomorphism in Definition 4.6.10.) If we use (4.3) to identify M_i and M_i', we thus have $M = M_1 \oplus \cdots \oplus M_k$. For this reason, we sometimes use the notation $\oplus$ for the cartesian product, and refer to it as "direct sum"; see Remark 4.1.13.

Example 4.5.5. Let V be a vector space over a field F and let $T \in \mathrm{End}_F(V)$. For the corresponding structure of an $F[x]$-module on V, we know that the submodules of V are simply the T-invariants subspaces of V. Since the notion of direct sum only involves addition, to say that V, as an $F[x]$-module, is a direct sum of $F[x]$-submodules $V_1, \ldots, V_k$ is equivalent to saying that V, as a vector space over F, is a direct sum of T-invariant subspaces $V_1, \ldots, V_k$.

Definition 4.5.6 (Indecomposable module)**.** A nonzero R-module M is *indecomposable* if it cannot be written as a direct sum of two nonzero submodules.

Proposition 4.5.7. *Every simple module is indecomposable.*

Proof. We prove the contrapositive. Assume M is not indecomposable. Then $M = M_1 \oplus M_2$ for some nonzero submodules M_1 and M_2 of M. Then M_1 is a nonzero submodule of M that is also proper since $M_1 = M$ would imply that $M_2 = \{0\}$. Hence M is not simple. □

Proposition 4.5.7 gives us many examples of indecomposable modules. For example, any one-dimensional vector space over a field F is an indecomposable F-module (see Example 4.2.4) and any simple abelian group is an indecomposable $\mathbb{Z}$-module (see Example 4.2.5). However, the converse of Proposition 4.5.7 is false, as the following example illustrates.

Example 4.5.8. As a module over itself, $\mathbb{Z}$ is an indecomposable $\mathbb{Z}$-module that is not simple. Indeed, since $\mathbb{Z}$ has many nonzero proper ideals, it is clearly not a simple $\mathbb{Z}$-module. So it remains to show it is indecomposable. Suppose $\mathbb{Z} = M \oplus N$ for some nonzero submodules M, N. As noted in Example 4.2.6, this means that M and N are ideals of $\mathbb{Z}$. By Proposition 1.3.6, we have $M = m\mathbb{Z}$ and $N = n\mathbb{Z}$ for some nonzero $m, n \in \mathbb{Z}$. But then $mn \in M \cap N$, and so the sum $M + N$ is not direct, which is a contradiction.

Exercises.

4.5.1. Prove Proposition 4.5.2.

4.5.2. Let M be an R-module and let $u_1, \ldots, u_n \in M$. Show that $\{u_1, \ldots, u_n\}$ is a basis of M if and only if $M = Ru_1 \oplus \cdots \oplus Ru_n$ and none of the elements $u_1, \ldots, u_n$ is torsion. (See Exercise 4.2.6 for the definition of this term.)

4.5.3. Let F be a field. Show that the quotient $F[x]$-module $F[x]/x^2F[x]$ (see Example 4.3.3) is indecomposable but not simple.

4.5.4. Suppose F is a field. Show that an F-module is simple if and only if it is indecomposable.

4.5.5. An R-module P is *projective* if it is a direct summand of a free module. In other words, P is projective if there exists some other R-module M such that $P \oplus M$ is a free R-module. Clearly every free module is projective. (Just take $M = \{0\}$.) Show that the $\mathrm{Mat}_n(\mathbb{C})$-module $\mathbb{C}^n$ (action by left matrix multiplication) is projective but not free.

4.6 Module homomorphisms

Fix a ring R.

Definition 4.6.1. Let M and N be R-modules. A *homomorphism* of R-modules from M to N is a function $\varphi\colon M \to N$ satisfying the following conditions:

(HM1) $\varphi(u_1+u_2) = \varphi(u_1)+\varphi(u_2)$ for all $u_1, u_2 \in M$ (i.e. φ is a homomorphism of additive groups);

(HM2) $\varphi(ru) = r\varphi(u)$ for all $r \in R$ and $u \in M$.

Example 4.6.2. If $R = F$ is a field, then F-modules M and N are simply vector spaces over F and a homomorphism of F-modules $\varphi\colon M \to N$ is nothing but a linear map from M to N.

Example 4.6.3. We recall that if $R = \mathbb{Z}$, then $\mathbb{Z}$-modules M and N are simply abelian groups equipped with the natural action by the integers. If $\varphi\colon M \to N$ is a homomorphism of $\mathbb{Z}$-modules, condition (HM1) shows that it is a homomorphism of groups. Conversely, if $\varphi\colon M \to N$ is a homomorphism of abelian groups, we have

$$\varphi(nu) = n\varphi(u)$$

for all $n \in \mathbb{Z}$ and all $u \in M$ (exercise), hence φ satisfies conditions (HM1) and (HM2) of Definition 4.6.1. Thus, *a homomorphism of $\mathbb{Z}$-modules is simply a homomorphism of abelian groups.*

Example 4.6.4. If N is a submodule of an R-module M, then the *quotient map*

$$M \to M/N, \quad u \mapsto u + N,$$

is a surjective R-module homomorphism. See Exercise 4.6.1.

Example 4.6.5. Let V be a vector space over a field F and let $T \in \mathrm{End}_F(V)$. For the corresponding structure of an $F[x]$-module on V, a homomorphism of $F[x]$-modules $S\colon V \to V$ is simply a linear map that commutes with T, i.e. such that $S \circ T = T \circ S$ (exercise).

Example 4.6.6. Let $m, n \in \mathbb{N}_{>0}$, and let $P \in \mathrm{Mat}_{m\times n}(R)$. Write the elements of R^m and R^n as row matrices:

$$R^m = \left\{ \begin{bmatrix} r_1 & r_2 & \cdots & r_m \end{bmatrix} : r_1, \ldots, r_m \in R \right\}.$$

Then, for all $u, u' \in R^m$ and $r \in R$, we have:

$$(u+u')P = uP + u'P \quad \text{and} \quad (ru)P = r(uP),$$

Thus the function

$$R^m \to R^n, \quad u \mapsto uP,$$

is a homomorphism of R-modules.

The following result shows that we obtain in this way all the homomorphism of R-modules from R^m to R^n.

Proposition 4.6.7. *Let $m, n \in \mathbb{N}_{>0}$ and let $\varphi\colon R^m \to R^n$ be a homomorphism of R-modules. There exists a unique matrix $P \in \mathrm{Mat}_{m\times n}(R)$ such that*

$$\varphi(u) = uP \quad \textit{for all } u \in R^m. \tag{4.4}$$

Proof. Let $\{e_1, \dots, e_m\}$ be the standard basis of R^m defined in Example 4.4.2 (except that we use row matrices instead of column matrices) and let P be the $m \times n$ matrix whose rows are $L_1 := \varphi(e_1), \dots, L_m := \varphi(e_m)$. For every m-tuple $u = \begin{bmatrix} r_1 & \cdots & r_m \end{bmatrix} \in R^m$, we have:

$$\varphi(u) = \varphi(r_1e_1 + \cdots + r_me_m) = r_1L_1 + \cdots + r_mL_m = \begin{bmatrix} r_1 & \cdots & r_m \end{bmatrix} \begin{bmatrix} L_1 \\ \vdots \\ L_M \end{bmatrix} = uP.$$

Thus P satisfies Condition (4.4). It is the only matrix that satisfies this condition since if a matrix $P' \in \mathrm{Mat}_{m\times n}(R)$ also satisfies $\varphi(u) = uP'$ for all $u \in R^m$, then its j-th row is $e_jP' = \varphi(e_j) = L_j$ for all $j = 1, \dots, m$. □

Corollary 4.6.8. *Consider R as a module over itself R, as in Example 4.1.8. Every $s \in R$ determines an R-module homomorphism*

$$\rho_s\colon R \to R, \quad \rho_s(r) = rs.$$

Furthemore, every *R-module homomorphism $R \to R$ is of the form ρ_s for a unique element $s \in R$.*

Proof. This is the case $m = n = 1$ of Proposition 4.6.7. □

We continue the general theory with the following result.

Proposition 4.6.9. *Let $\varphi\colon M \to N$ and $\psi\colon N \to P$ be homomorphisms of R-modules.*

(a) The composite $\psi \circ \varphi\colon M \to P$ is also a homomorphism of R-modules.

(b) If $\varphi\colon M \to N$ is bijective, its inverse $\varphi^{-1}\colon N \to M$ is a homomorphism of R-modules.

Proof. We leave the proof of (a) as an exercise. To prove (b), suppose that φ is bijective. Let $v, v' \in N$ and $r \in R$. We set

$$u := \varphi^{-1}(v) \quad \text{and} \quad u' := \varphi^{-1}(v').$$

Since $u, u' \in M$ and φ is a homomorphism of R-modules, we have

$$\varphi(u + u') = \varphi(u) + \varphi(u') = v + v' \quad \text{and} \quad \varphi(ru) = r\varphi(u) = rv.$$

Therefore,

$$\varphi^{-1}(v + v') = u + u' = \varphi^{-1}(v) + \varphi^{-1}(v') \quad \text{and} \quad \varphi^{-1}(rv) = ru = r\varphi^{-1}(v).$$

This proves that $\varphi^{-1}\colon N \to M$ is indeed a homomorphism of R-modules. □

Definition 4.6.10 (Isomorphism)**.** A bijective homomorphism of R-modules is called an *isomorphism* of R-modules. We say that an R-module M is *isomorphic* to an R-module N if there exists an isomorphism of R-modules $\varphi\colon M \to N$. We then write $M \simeq N$.

In light of these definitions, Proposition 4.6.9 has the following immediate consequence:

Corollary 4.6.11. *Let $\varphi\colon M \to N$ and $\psi\colon N \to P$ be isomorphisms of R-modules. Then $\psi \circ \varphi\colon M \to P$ and $\varphi^{-1}\colon N \to M$ are also isomorphisms of R-modules.*

We deduce from this that isomorphism is an equivalence relation on the class of R-modules. Furthermore, the set of isomorphisms of R-modules from an R-module M to itself, called *automorphisms* of M, form a subgroup of the group of bijections from M to itself. We call it the *group of automorphisms* of M.

Definition 4.6.12. Let $\varphi\colon M \to N$ be a homomorphism of R-modules. The *kernel* of φ is

$$\ker(\varphi) := \{u \in M : \varphi(u) = 0\}$$

and the *image* of φ is

$$\operatorname{im}(\varphi) := \{\varphi(u) : u \in M\}.$$

The kernel and image of a homomorphism of R-modules $\varphi\colon M \to N$ are therefore simply the kernel and image of φ considered as a homomorphism of abelian groups. The proof of the following result is left as an exercise.

Proposition 4.6.13. *Let $\varphi\colon M \to N$ be a homomorphism of R-modules.*

(a) $\ker(\varphi)$ *is an R-submodule of M.*

(b) $\operatorname{im}(\varphi)$ *is an R-submodule of N.*

(c) The homomorphism φ is injective if and only if $\ker(\varphi) = \{0\}$*. It is surjective if and only if* $\operatorname{im}(\varphi) = N$*.*

Example 4.6.14. Let M be an R-module and let $r \in Z(R)$ (see Definition 1.1.26). The function

$$\varphi\colon M \to M, \quad u \to ru,$$

is a homomorphism of R-modules (exercise). Its kernel is

$$M_r := \{u \in M : ru = 0\}$$

and its image is

$$rM := \{ru : u \in M\}.$$

Thus, by Proposition 4.6.13, M_r and rM are both R-submodules of M.

We say that an R-module M is a *homomorphic image* of an R-module N if there is a surjective R-module homomorphism $N \to M$.

Proposition 4.6.15. *Every finitely-generated R-module is the homomorphic image of a finitely-generated free R-module.*

Proof. Suppose M has a finite generating set $v_1, \ldots, v_n$. We leave it as a straightforward exercise (Exercise 4.6.8) to verify that

$$\varphi\colon R^n \to M, \quad \varphi(r_1, \ldots, r_n) \mapsto \sum_{i=1}^{n} r_i v_i. \tag{4.5}$$

is a surjective R-module homomorphism. □

We conclude this section with the following result.

Proposition 4.6.16. *Let $\varphi\colon M \to N$ be a homomorphism of R-modules. Suppose that M is free with basis $\{u_1, \ldots, u_n\}$. Then φ is an isomorphism if and only if $\{\varphi(u_1), \ldots, \varphi(u_n)\}$ is a basis of N.*

Proof. We will show more precisely that

(a) φ is injective if and only if $\varphi(u_1), \ldots, \varphi(u_n)$ are linearly independent over R, and

(b) φ is surjective if and only if $\varphi(u_1), \ldots, \varphi(u_n)$ generate N.

Therefore, φ is bijective if and only if $\{\varphi(u_1), \ldots, \varphi(u_n)\}$ is a basis of N.

To prove (a), we use the fact that φ is injective if and only if $\ker(\varphi) = \{0\}$ (Proposition 4.6.13(c)). We note that, for all $r_1, \ldots, r_n \in R$, we have

$$\begin{aligned} r_1\varphi(u_1) + \cdots + r_n\varphi(u_n) = 0 &\iff \varphi(r_1u_1 + \cdots + r_nu_n) = 0 \\ &\iff r_1u_1 + \cdots + r_nu_n \in \ker(\varphi). \end{aligned} \tag{4.6}$$

If $\ker(\varphi) = \{0\}$, the last condition becomes $r_1u_1 + \cdots + r_nu_n = 0$ which implies $r_1 = \cdots = r_n = 0$ since $u_1, \ldots, u_n$ are linearly independent over R. In this case, we conclude that $\varphi(u_1), \ldots, \varphi(u_n)$ are linearly independent over R. Conversely, if $\varphi(u_1), \ldots, \varphi(u_n)$ are linearly independent, the equivalences (4.6) tell us that the only choice of $r_1, \ldots, r_n \in R$

such that $r_1u_1 + \cdots + r_nu_n \in \ker(\varphi)$ is $r_1 = \cdots = r_n = 0$. Since $u_1, \ldots, u_n$ generate M, this implies that $\ker(\varphi) = \{0\}$. Thus $\ker(\varphi) = \{0\}$ if and only if $\varphi(u_1), \ldots, \varphi(u_n)$ are linearly independent.

To prove (b), we use the fact that φ is surjective if and only if $\operatorname{im}(\varphi) = N$. Since $M = \langle u_1, \ldots, u_n \rangle_R$, we have

$$\begin{aligned}\operatorname{im}(\varphi) &= \{\varphi(r_1u_1 + \cdots + r_nu_n) : r_1, \ldots, r_n \in R\} \\ &= \{r_1\varphi(u_1) + \cdots + r_n\varphi(u_n) : r_1, \ldots, r_n \in R\} \\ &= \langle \varphi(u_1), \ldots, \varphi(u_n) \rangle_R.\end{aligned}$$

Thus $\operatorname{im}(\varphi) = N$ if and only if $\varphi(u_1), \ldots, \varphi(u_n)$ generate N. □

Exercises.

4.6.1. Prove that the map of Example 4.6.4 is an R-module homomorphism.

4.6.2. Prove Proposition 4.6.9(a).

4.6.3. Prove Proposition 4.6.13.

4.6.4. Prove that the map φ of Example 4.6.14 is a homomorphism of R-modules.

4.6.5. Let $\varphi\colon M \to M'$ be a homomorphism of R-modules and let N be an R-submodule of M. We define the *image* of N under φ to be

$$\varphi(N) := \{\varphi(u) : u \in N\}.$$

Show that

(a) $\varphi(N)$ is an R-submodule of M';

(b) if $N = \langle u_1, \ldots, u_m \rangle_R$, then $\varphi(N) = \langle \varphi(u_1), \ldots, \varphi(u_m) \rangle_R$;

(c) if N admits a basis $\{u_1, \ldots, u_m\}$ and if φ is injective, then $\{\varphi(u_1), \ldots, \varphi(u_m)\}$ is a basis of $\varphi(N)$.

4.6.6. Let $\varphi\colon M \to M'$ be a homomorphism of R-modules and let N' be an R-submodule of M'. We define the *inverse image* of N' under φ to be

$$\varphi^{-1}(N') := \{u \in M : \varphi(u) \in N'\}.$$

(This notation does not imply that φ^{-1} exists.) Show that $\varphi^{-1}(N')$ is an R-submodule of M.

4.6.7. Let M be a free R-module with basis $\{u_1, \ldots, u_n\}$, let N be an arbitrary R-module, and let $v_1, \ldots, v_n$ be arbitrary elements of N. Show that there exists a unique homomorphism of R-modules $\varphi\colon M \to N$ such that $\varphi(u_i) = v_i$ for $i = 1, \ldots, n$.

4.6.8. Show that (4.5) is a surjective R-module homomorphism. *Hint*: Use Exercise 4.6.7.

4.7 Isomorphism theorems

In this section we prove some important isomorphism theorems for modules. Throughout this section, R is an arbitrary ring.

Recall, from Proposition 4.6.13, that if $\varphi\colon M \to N$ is an R-module homomorphism, then $\ker(\varphi)$ is a submodule of M and $\operatorname{im}(\varphi)$ is a submodule of N.

Theorem 4.7.1 (First isomorphism theorem for modules)**.** *Let M and N be R-modules, and let $\varphi\colon M \to N$ be an R-module homomorphism. Then we have an isomorphism of R-modules*

$$\operatorname{im}(\varphi) \cong M/\ker(\varphi).$$

Proof. Consider the map

$$\bar{\varphi}\colon M/\ker(\varphi) \to \operatorname{im}(\varphi), \quad \bar{\varphi}(u + \ker(\varphi)) = \varphi(u).$$

By the first isomorphism theorem for groups (MAT 2143), this map is well-defined and is an isomorphism of additive groups. It remains to verify that it is an R-module homomorphism. We leave this straightforward verification as Exercise 4.7.1. □

Recall, from Proposition 4.2.12, that sums and intersections of R-modules are again R-modules.

Theorem 4.7.2 (Second isomorphism theorem for modules)**.** *Suppose M is an R-module, and let S and T be submodules of M. Then the quotient modules $(S+T)/T$ and $S/(S\cap T)$ are isomorphic.*

Proof. Consider the map

$$\varphi\colon S \to (S+T)/T, \quad u \mapsto u + T.$$

Since φ is the composition of the R-module inclusion $S \hookrightarrow S+T$ and the projection map $S+T \to (S+T)/T$ (see Example 4.6.4), it is a homomorphism of R-modules. For $u \in S$, we have

$$\varphi(u) = 0 \iff u + T = 0 + T \iff u \in T.$$

Hence $\ker(\varphi) = S \cap T$. Furthermore, for $u \in S$ and $v \in T$, we have

$$(u+v) + T = u + T = \varphi(u) \in \operatorname{im}(\varphi).$$

Thus φ is surjective. The result now follows from the first isomorphism theorem for modules (Theorem 4.7.1). □

Theorem 4.7.3 (Third isomorphism theorem for modules)**.** *If $N \subseteq L \subseteq M$ are R-modules (i.e. L is a submodule of M and N is a submodule of L), then*

$$M/L \cong (M/N)/(L/N).$$

Proof. Consider the map

$$\varphi\colon M/N \to M/L, \quad \varphi(u+N) = u+L.$$

This map is well-defined since

$$u+N = v+N \implies u-v \in N \subseteq L \implies \varphi(u+N) = u+L = v+L = \varphi(v+N).$$

It is straightforward to verify that φ is a homomorphism of R-modules. Since

$$\varphi(u+N) = u+L = 0+L \iff u \in L,$$

we have $\ker(\varphi) = L/N$. We also clearly have $\mathrm{im}(\varphi) = M/L$. Then the result follows from the first isomorphism theorem for modules (Theorem 4.7.1). □

When $R = F$ is a field, the above theorems yield important isomorphism theorems for vector spaces that you may not have seen in your linear algebra courses.

Example 4.7.4. Let $\mathcal{F}(\mathbb{R})$ be the $\mathbb{R}$-vector space of all functions $\mathbb{R} \to \mathbb{R}$. Let

$$W = \{f \in \mathcal{F}(\mathbb{R}) : f(0) = 0\}.$$

Now, consider the map

$$\varphi\colon \mathcal{F}(\mathbb{R}) \to \mathbb{R}, \quad \varphi(f) = f(0).$$

It is not hard to check that φ is a linear map (i.e. a homomorphism of $\mathbb{R}$-modules). It is clearly surjective, and we have $W = \ker(f)$. It follows that W is a subspace of $\mathcal{F}(\mathbb{R})$. The first isomorphism theorem for modules (Theorem 4.7.1) implies that

$$\mathcal{F}(\mathbb{R})/W \cong \mathbb{R}$$

as $\mathbb{R}$-vector spaces. This would be hard to prove without the first isomorphism theorem, since the space $\mathcal{F}(\mathbb{R})$ is huge. (For instance, it does not even have a countable basis.)

Exercises.

4.7.1. Prove that the map $\bar{\varphi}$ in the proof of Theorem 4.7.1 is an R-module homomorphism.

4.7.2. Prove that the map φ of Example 4.7.4 is a surjective linear map.

4.7.3. Suppose that $M = M_1 \oplus M_2$ for submodules M_1, M_2 of an R-module M. Show that $M/M_1 \cong M_2$ as R-modules.

4.7.4. Recall that a function $\mathbb{R} \to \mathbb{R}$ is *odd* if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. Consider the following subspaces of $\mathcal{F}(\mathbb{R})$ (notation from Example 4.7.4):

$$\begin{aligned} U &= \{f \in \mathcal{F}(\mathbb{R}) : f \text{ is continuous}\}, \\ V &= \{f \in \mathcal{F}(\mathbb{R}) : f \text{ is odd}\}, \\ W &= \{f \in \mathcal{F}(\mathbb{R}) : f \text{ is continuous } \textit{and} \text{ odd}\}, \\ X &= \{f \in \mathcal{F}(\mathbb{R}) : f \text{ is the sum of an odd function and a continuous function}\}. \end{aligned}$$

Prove that $X/V \cong U/W$.

4.7.5. Suppose $m, n \in \mathbb{Z}$, $m, n \geq 2$. Let

$$M = \{ma + mn\mathbb{Z} : a \in \mathbb{Z}\} \subseteq \mathbb{Z}_{mn}.$$

(a) Verify that M is a subgroup of $\mathbb{Z}_{mn}$.

(b) Show that $\mathbb{Z}_m \cong \mathbb{Z}_{mn}/M$ as $\mathbb{Z}$-modules (i.e. as abelian groups).

4.7.6. Suppose M is a cyclic R-module with generator u.

(a) Show that $M \cong R/\operatorname{ann}(u)$ as R-modules.

(b) Suppose R is commutative. Show that $M \cong R/\operatorname{ann}(M)$ as R-modules.

4.8 Modules over commutative rings

Throughout this section R is a *commutative ring*. At some point we will also start assuming that R is a principal ideal domain.

Suppose M is an R-module. Then we can denote the entries of

$$M^n := \underbrace{M \times M \times \cdots \times M}_{n \text{ factors}}$$

as $1 \times n$ matrices with entries in M, which we write as row matrices $(v_1, \ldots, v_n) := \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, $v_1, \ldots, v_n \in M$. If $C \in \operatorname{Mat}_{n \times s}(R)$, then right multiplication by C gives an R-module homomorphism from M^n to M^s. Precisely, if $C = [c_{i,j}]$, then

$$(v_1, \ldots, v_n)C = \left(\sum_{i=1}^{n} c_{i,1} v_i, \ldots, \sum_{i=1}^{n} c_{i,s} v_i \right).$$

If $B \in \operatorname{Mat}_{s \times t}(R)$, then right multiplication by CB corresponds to the composition of right multiplication by C followed by right multiplication by B:

$$(v_1, \ldots, v_n)(CB) = ((v_1, \ldots, v_n)C)B.$$

If $\{v_1, \ldots, v_n\}$ is linearly independent over R and $(v_1, \ldots, v_n)C = 0$, then C is the zero matrix. See Exercise 4.8.1.

You learned in linear algebra that any two bases of a finite-dimensional vector space have the same cardinality (i.e. same number of elements). It turns out that the same thing is true for finitely-generated free modules over a commutative ring.

Lemma 4.8.1. *Let R be a commutative ring. Any two bases of a finitely-generated free R-module have the same cardinality.*

Proof. By Lemma 4.4.4, we know that any basis of a finitely-generated R-module is finite. Let M be an R-module with basis $\{v_1, \ldots, v_n\}$ and a spanning set $\{w_1, \ldots, w_m\}$. We will show that $m \geq n$. The lemma follows from this since any basis is a spanning set.

For each $j = 1, \ldots, m$, we can write w_j uniquely as an R-linear combination of the basis elements v_i:

$$w_j = a_{1,j}v_1 + a_{2,j}v_2 + \ldots a_{n,j}v_n. \tag{4.7}$$

Let $A \in \mathrm{Mat}_{n\times m}(R)$ be the matrix $A = [a_{i,j}]$. Then the m equations (4.7) can be written as a single matrix equation:

$$(v_1, \ldots, v_n)A = (w_1, \ldots, w_m). \tag{4.8}$$

Since $\{w_1, \ldots, w_m\}$ spans M, we can also write each v_j as an R-linear combination of the elements w_i. As above, this means we have a matrix $B \in \mathrm{Mat}_{m\times n}(R)$ such that

$$(w_1, \ldots, w_m)B = (v_1, \ldots, v_n). \tag{4.9}$$

Combining (4.8) and (4.9), we have

$$(v_1, \ldots, v_n)AB = (w_1, \ldots, w_m)B = (v_1, \ldots, v_n),$$

which is equivalent to

$$(v_1, \ldots, v_n)(AB - I_n) = 0,$$

where I_n is the $n \times n$ identity matrix. Since the set $\{v_1, \ldots, v_n\}$ is linearly independent, this forces $AB = I_n$.

Now suppose, towards a contradiction, that $m < n$. Then we augment A by appending $n - m$ columns of zeros on the righthand side to obtain a matrix $A' \in \mathrm{Mat}_{n\times n}(R)$. Similarly, we add $n - m$ rows of zeros to the bottom of B to obtain a matrix $B' \in \mathrm{Mat}_{n\times n}(R)$. Then we have $A'B' = AB = I_n$. Taking determinants,[1] we get

$$1 = \det(I_n) = \det(A'B') = \det(A')\det(B').$$

But $\det(A') = 0$, since the matrix A' has a column of zeros. This contradiction implies that $m \geq n$, as desired. □

Warning! Over a noncommutative ring, a free module can have two bases of different cardinalities! See Exercise 4.8.2.

Definition 4.8.2 (Rank of a module)**.** Let R be a commutative ring. The *rank* of a finitely-generated free R-module is the cardinality of any basis.

Note that it follows from the definition that the zero R-module is free of rank zero, with the empty set as a basis.

[1] Technically speaking, we have not discussed the properties of determinants for matrices with entries in an arbitrary commutative ring R. But the proof you saw in linear algebra shows that the determinant, computed in the usual way, has the multiplicative property used here.

Exercises.

4.8.1. Suppose R is a commutative ring, M is an R-module, and $A \in \mathrm{Mat}_{n\times m}(R)$. Show that if $\{v_1, \ldots, v_n\}$ is a linearly independent subset of M and $(v_1, \ldots, v_n)A = 0$, then $A = 0$.

4.8.2. Let R denote the set of infinite-by-infinite matrices $[a_{i,j}]_{i,j\in\mathbb{Z}}$, $a_{i,j} \in \mathbb{C}$, such that each row and each column has only finitely many nonzero entries. Show that R is a non-commutative ring. (The finiteness assumption is crucial here.) Show also that $R \cong R \times R$ as R-modules. In particular, this implies that R has a basis with one element *and* a basis with two elements. So it does not have a well-defined rank!

4.8.3. Let V and W be free modules over a commutative ring R, with ordered bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, respectively. Let $\varphi\colon V \to W$ be an R-module homomorphism. Let $A = [a_{i,j}] \in \mathrm{Mat}_{m\times n}(R)$ be the matrix determined by

$$\varphi(v_j) = \sum_{i=1}^{m} a_{i,j} w_i.$$

Show that for any element $\sum_{j=1}^{n} r_j v_j$ of V, we have

$$\varphi\left(\sum_{j=1}^{n} r_j v_j\right) = (w_1, \ldots, w_m)A \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}.$$

4.8.4. Let R be a commutative ring and let N be an R-module. Suppose that $\{w_1, \ldots, w_k\}$ generates N as an R-module. Let $D \in \mathrm{Mat}_k(R)$ and define $\{w'_1, \ldots, w'_k\}$ by

$$(w'_1, \ldots, w'_k) = (w_1, \ldots, w_k)D.$$

Show that if D is invertible in $\mathrm{Mat}_k(R)$, then $\{w'_1, \ldots, w'_k\}$ generates N.

4.9 Modules over PIDs

Throughout this section, R is a principal ideal domain.

Lemma 4.9.1. *Let M be a free module of finite rank n over a principal ideal domain R. Any submodule of M has a generating set with no more than n elements.*

Proof. We prove the result by induction on n. If $n = 1$, then $M \cong R$. In this case R-submodules correspond to ideals of R, which are generated by a single element, since R is a PID. Thus, the base case holds.

Now suppose M has rank $n > 1$ and that the assertion holds for free modules of rank less than n. Let $\{v_1, \ldots, v_n\}$ be a basis for M, and let $M' = \langle v_1, \ldots, v_{n-1}\rangle_R$. Let N be a

submodule of M, and let $N' = N \cap M'$. By the induction hypothesis, N' has a generating set with no more than $n-1$ elements.

Every element $u \in N$ can be written uniquely in the form

$$u \in \sum_{i=1}^{n} \alpha_i(u) v_i, \quad \alpha_i(u) \in R.$$

The map

$$\alpha_n \colon N \to R,$$

is an R-module homomorphism. If $\operatorname{im}(\alpha_n) = \{0\}$, then $N = N'$, and so N is generated by no more than $n-1$ elements. Otherwise, $\operatorname{im}(\alpha_n)$ is a nonzero ideal of R, hence of the form dR for some nonzero $d \in R$. Choose $h \in N$ such that $\alpha_n(h) = d$.

If $u \in N$, then $\alpha_n(u) = rd$ for some $r \in R$. Then $y = u - rh$ satisfies $\alpha_n(y) = 0$, and so $y \in N \cap M' = N'$. Thus $u = y + rh \in N' + Rh$. Hence $N = Rh + N'$.

Since N' has a generating set with no more than $n-1$ elements, N is generated by no more than n elements. □

Corollary 4.9.2. *If M is a finitely-generated module over a principal ideal domain, then every submodule of M is finitely generated.*

Proof. Suppose M has a finite generating set $v_1, \ldots, v_n$. Consider the R-module homomorphism

$$\varphi \colon R^n \to M, \quad \varphi(r_1, \ldots, r_n) \mapsto \sum_{i=1}^{n} r_i v_i,$$

of Proposition 4.6.15. Let N be a submodule of M. By Lemma 4.9.1, $\varphi^{-1}(N)$ is generated by a set S with no more than n elements. Then $\varphi(S)$ is a generating subset of N of cardinality at most n. □

You learned in linear algebra that if U is a subspace of an n-dimensional vector space V, then you can find a basis $\{v_1, \ldots, v_n\}$ of V such that $\{v_1, \ldots, v_m\}$ is a basis of U for some $m \leq n$. However, this is *not* true for general modules, even for free modules over a PID. For example, as a $\mathbb{Z}$-modules, $\mathbb{Z}$ is free with basis $\{1\}$ or $\{-1\}$. Consider the submodule $2\mathbb{Z}$. This is free with basis $\{2\}$ or $\{-2\}$. Thus, there is no subset of a basis of $\mathbb{Z}$ that is a basis of $2\mathbb{Z}$.

The following theorem is the proper generalization of this property of bases of free modules and their submodules.

Theorem 4.9.3. *Suppose R is a principal ideal domain. Let M be a free R-module of rank m, and let N be a submodule of M. Then there exists a basis $\{v_1, \ldots, v_m\}$ of M and elements $a_1, \ldots, a_n \in R$, $n \leq m$, such that a_i divides a_j if $i \leq j$ and $\{a_1 v_1, \ldots, a_n v_n\}$ is a basis of N. In particular, N is a free R-module of rank n.*

The proof of Theorem 4.9.3 is not conceptually difficult, but it is somewhat lengthy, and so we will not cover it in this course. It involves the type of row and column operations that you learned in linear algebra, except that you must take extra care since you are working over an arbitrary PID. For details of the proof, see the proof of [Goo15, Th. 8.4.2].

Lemma 4.9.4. *Let $A_1, \dots, A_n$ be R-modules, and let $B_i \subseteq A_i$ be submodules. Then*

$$(A_1 \oplus \cdots \oplus A_n)/(B_1 \oplus \cdots \oplus B_n) \cong A_1/B_1 \oplus \cdots \oplus A_n/B_n.$$

Proof. The R-module homomorphism

$$\varphi\colon A_1 \oplus \cdots \oplus A_n \to A_1/B_1 \oplus \cdots \oplus A_n/B_n, \quad (a_1, \dots, a_n) \mapsto (a_1 + B_1, \dots, a_n + B_n).$$

is surjective, with $\ker(\varphi) = B_1 \oplus \cdots \oplus B_n$. Thus the result follows from Theorem 4.7.1. □

Theorem 4.9.5 (Structure theorem for finitely-generated modules over a PID)**.** *Let R be a principle ideal domain, and let M be a nonzero finitely-generated R-module.*

(a) The module M is a direct sum of cyclic modules,

$$M \cong R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_k) \oplus R^n,$$

where the a_i are nonzero, nonunit elements of R, and a_i divides a_j for $i \leq j$.

(b) The decomposition in part (a) is unique in the following sense: Suppose

$$M \cong R/(b_1) \oplus R/(b_2) \oplus \cdots \oplus R/(b_l) \oplus R^m,$$

where the b_i are nonzero, nonunit elements of R, and b_i divides b_j for $i \leq j$. Then $k = l$, $m = n$, and $(a_i) = (b_i)$ for all i.

Proof. We will prove part (a); the proof of part (b) can be found in [Goo15, §8.5]. Let $\{v_1, \dots, v_m\}$ be a set of generators of M of minimal cardinality. By Proposition 4.6.15, there is a free module L of rank m and a surjective R-module homomorphism

$$\varphi\colon L \to M.$$

Let $N = \ker(\varphi)$. By Theorem 4.9.3, N is free of rank $k \leq m$, and there exists a basis $\{v_1, \dots, v_m\}$ of L and nonzero elements $a_1, \dots, a_k$ of R such that $\{a_1 v_1, \dots, a_k v_k\}$ is a basis of N and a_i divides a_j for $i \leq j$. Therefore

$$\begin{aligned} M \cong L/N &= (Rv_1 \oplus \cdots \oplus Rv_m)/(Ra_1 v_1 \oplus \cdots \oplus Ra_k v_k) \\ &\cong Rv_1/Ra_1 v_1 \oplus \cdots \oplus Ra_k v_k \oplus Rv_{k+1} \oplus \cdots \oplus Rv_m. \end{aligned} \tag{4.10}$$

Now, the map

$$R \to Rv_i/Ra_i v_i, \quad r \mapsto rv_i + Ra_i v_i,$$

is a surjective R-module homomorphism with kernel (a_i). Thus $Rv_i/Ra_i v_i \cong R/(a_i)$ by Theorem 4.7.1. Therefore, (4.10) gives

$$M \cong R/(a_1) \oplus \cdots \oplus R/(a_k) \oplus R^{m-k}.$$

If some a_i were invertible, then $R/(a_i)$ would be the zero module, and hence could be dropped from the sum. But then M would be generated by fewer than m elements, contradicting the minimality of m. □

As an immediate consequence of Theorem 4.9.5, we recover one of the important theorems you saw in MAT 2143.

Theorem 4.9.6 (Structure theorem for finitely-generated abelian groups)**.** *Let G be a finitely-generated abelian group. Then G is a direct product of cyclic groups,*

$$G \cong \mathbb{Z}/a_1\mathbb{Z} \times \mathbb{Z}/a_2\mathbb{Z} \times \cdots \times \mathbb{Z}/a_k\mathbb{Z} \times \mathbb{Z}^n,$$

where the $a_i \geq 2$, and a_i divides a_j for $i \leq j$. The a_i and the values k and n are uniquely determined by G.

Proof. This follows from Theorem 4.9.5, with $R = \mathbb{Z}$. □

If F is an algebraically closed field, then Theorem 4.9.5 applied to the ring $R = F[x]$ leads to the *Jordan canonical form* of a linear transformation of a finite-dimensional vector space over F. For details, see [Goo15, §8.7]. The Jordan canonical form is covered in MAT 3341.

Exercises.

Recall the definition of *torsion* from Exercise 4.2.6.

4.9.1. Let R be an integral domain with a non-principal ideal J.

(a) Show that J is torsion-free as an R-module.

(b) Show that any two elements of J are linearly dependent.

(c) Show that J is not a free R-module.

4.9.2. Suppose R is a principal ideal domain. Use Theorem 4.9.5 to show that, if M is a finitely generated R-module, then M/M_{tor} is a free R-module.

4.9.3. Show that $M = \mathbb{Q}/\mathbb{Z}$ is a torsion $\mathbb{Z}$-module, that M is not finitely generated, and that $\text{ann}(M) = \{0\}$.

Index

Bibliography

[Goo15] Frederick M. Goodman. *Abstract: abstract and concrete*. 2.6 edition, 2015. URL: http://homepage.math.uiowa.edu/~goodman/algebrabook.dir/algebrabook.html.

[Jud19] Thomas W. Judson. *Abstract algebra*. 2019 annual edition, 2019. URL: http://abstract.ups.edu/index.html.

[Nic12] W. Keith Nicholson. *Introduction to abstract algebra*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, fourth edition, 2012.

[Rog71] Kenneth Rogers. The axioms for Euclidean domains. *Amer. Math. Monthly*, 78:1127–1128, 1971. URL: https://doi.org/10.2307/2316324.

[Roy] Damien Roy. MAT 3141 – Linear Algebra II, Lecture notes (translated by A. Savage). URL: http://alistairsavage.ca/mat3141/notes/MAT3141-LinearAlgebraII.pdf.