Ring Theory

MAT 3143

Winter 2018

Alistair Savage

Department of Mathematics and Statistics

University of Ottawa

Contents

Preface **3**

1 Rings **4**
1.1 Examples and basic properties 4
1.2 Integral domains and fields 11
1.3 Ideals and quotient rings 16
1.4 Homomorphisms 23

2 Polynomials **32**
2.1 Polynomial rings 32
2.2 Factorization of polynomials over a field 39
2.3 Quotient rings of polynomials over a field 46

3 Integral domains **49**
3.1 Unique factorization domains 49
3.2 Principal ideal domains 58
3.3 Euclidean domains 60

4 Fields **62**
4.1 Brief review of vector spaces 62
4.2 Field extensions 64
4.3 Splitting fields 71
4.4 Finite fields 76

Index **80**

Preface

These notes are aimed at students in the course *Ring Theory* (MAT 3143) at the University of Ottawa. This is a first course in ring theory (except that students may have seen some basic ring theory near the end of MAT 2143/2543). 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 also spend some time studying fields in more depth than we've seen in previous courses. For example, we examine the ideas of field extensions and splitting fields.

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 and Hadi Salmasian. Other portions follow the text Abstract Algebra: Theory and Applications by Tom Judson, which is the recommended text for the course.

Alistair Savage

Course website: http://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 [Jud12, 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 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 $M_n(\mathbb{Q})$ 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 $M_n(R)$ 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

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

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 group 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

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

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

$$(m+n)a = ma + na, \quad m(na) = (mn)a \quad \text{for all } m, n \in \mathbb{Z},\ a \in R, \tag{1.1}$$

$$a^{m+n} = a^m a^n, \quad (a^m)^n = a^{mn} \quad \text{for all } m, n \in \mathbb{N},\ a \in R. \tag{1.2}$$

If a is a unit, then (1.2) 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 [Jud12, Prop. 16.1]). $\square$

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 $M_2(\mathbb{R})$ 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 $M_2(\mathbb{R})$ 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) $M_2(\mathbb{Z})$ is a subring of $M_2(\mathbb{Q})$.

(c) $\begin{bmatrix} \mathbb{R} & \mathbb{R} \\ 0 & \mathbb{R} \end{bmatrix} := \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} \;\middle|\; x, y, z \in \mathbb{R} \right\}$ is a subring of $M_2(\mathbb{R})$.

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

(e) The *gaussian integers* $\mathbb{Z}(i) := \{a + bi \mid a, b \in \mathbb{Z}\}$ 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

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

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 $s\sigma(1_R) = \sigma(r)\sigma(1_S) = \sigma(r1_S) = \sigma(r) = s$. 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

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

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

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

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 (1.1) and (1.2).

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) $S = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; a, b, c, d \in \mathbb{R},\ a + c = b + d \right\}$, $R = M_2(\mathbb{R})$.

(b) $S = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$, $R = M_2(\mathbb{R})$.

(c) $S = \left\{ \begin{bmatrix} a & 0 & b \\ 0 & c & d \\ 0 & 0 & a \end{bmatrix} \;\middle|\; a, b, c, d \in \mathbb{R} \right\}$, $R = M_3(\mathbb{R})$.

(d) $S = \left\{ \begin{bmatrix} a & 2b \\ b & a \end{bmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$, $R = M_2(\mathbb{R})$.

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 $S + T = \{s + t \mid s \in S,\ t \in T\}$ is a subring of R?

1.1.10. Prove Lemma 1.1.28.

1.1.11. Find the center of $M_2(\mathbb{R})$.

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

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

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) $M_2(\mathbb{Z}_n)$.

(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) $M_2(\mathbb{Z}_2)$

(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*.

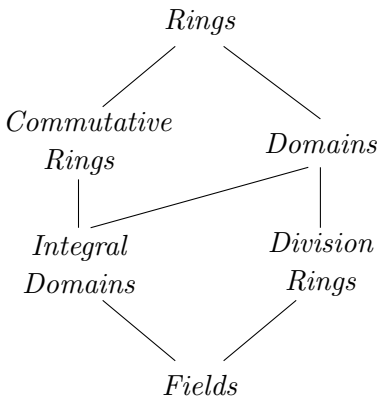


Figure 1.1: Types of rings

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 $M_2(\mathbb{R})$ 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
$$\sigma(\bar{a}) = \sigma(\bar{b}) \implies a1 = b1 \implies (a-b)1 = 0 \implies p|(a-b) \implies \bar{a} = \bar{b} \text{ in } \mathbb{Z}_p,$$
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 $p|(a-b)$.
- 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 $\mathbb{Q}(\sqrt{2}) := \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ 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

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

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 [Jud12, Example 6]. □

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 [Jud12, Th. 16.16]. (We will not use this theorem.)

Theorem 1.2.14 (Wedderburn's 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

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

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

$$Q := \left\{ \frac{r}{u} \;\middle|\; r, u \in R,\ u \neq 0 \right\}. \tag{1.3}$$

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 [Jud12, 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

$$R' := \left\{ \frac{r}{1} \mid r \in R \right\}. \tag{1.4}$$

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 [Jud12, 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 $\mathbb{Z}_m^\times = \{k \in \mathbb{Z}_m \mid \gcd(k, m) = 1\}$.

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

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

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 nilpotents in a domain.

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

1.2.7 ([Nic12, Ex. 3.2.5]). Show that $M_n(R)$, 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 $\gcd(m, n) = 1$. *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 (1.3) is a field.

1.2.15. Show that R' as defined in (1.4) 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

$$Q = \left\{ \frac{r}{u} \;\middle|\; r \in R,\ u \in U \right\}$$

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 $Q = \{ru^{-1} \mid r \in R,\ u \in U\}$.

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 $S = \{(x, x) \mid x \in \mathbb{Z}\}$. 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 $R/I := \{a + I \mid a \in R\}$. (Recall that $a + I = a' + I \iff a - a' \in I$.) We want the addition and multiplication to be given by

$$(a + I) + (b + I) = (a + b) + I \quad \text{and} \quad (a + I)(b + I) = ab + I. \tag{1.5}$$

Theorem 1.3.3. *Let $I \subseteq R$ be an ideal of R. Then R/I, with addition and multiplication defined by (1.5), 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. □

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

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

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 $r + I \in A/I$. Then there is some $r' \in R$ such that $r + I = r' + I$ and $r' \in A$. This implies that $r + I = r' + I \in \tilde{A}$ by the definition of A.

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

If $a_1 + I, a_2 + I \in \tilde{A}$, then $a_1 + I = a_1' + I$ and $a_2 + I = a_2' + I$ for some $a_2, a_2' \in A$. Thus $a_1 - a_1' \in I$ and $a_2 - a_2' \in I$. Hence

$$a_1 - a_2 = (a_1 - a_1') - (a_2 - a_2') + (a_1' - a_2') \in I + I + A \subseteq A$$

(since $I \subseteq A$). Thus $(a_1 + I) - (a_2 + I) = (a_1 - a_2) + I \in \tilde{A}$. Thus $\tilde{A}$ is an additive subgroup of R/I.

Now suppose $a + I \in \tilde{A}$ and $r + I \in R/I$. Then $a + I = a' + I$ for some $a' \in A$. Thus $a - a' \in I$, which implies that $ra - ra' = r(a - a') \in I$. Hence $ra = ra' + I \subseteq A + I \subseteq A$. Hence $(r + I)(a + I) = ra + I \in \tilde{A}$. 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. If R/I is simple, then for every ideal $I \subseteq A \subseteq R$ 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. Consider

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

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

$$J = \{2a_0 + a_1 x + \cdots + a_k x^k \mid k \in \mathbb{N},\ a_0, \ldots, a_k \in \mathbb{Z}\}.$$

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 $\mathbb{F}$ is a field, then $M_n(\mathbb{F})$ 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 $M_n(R)$ is of the form $M_n(I)$ 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) $R = \mathbb{Z} \times \mathbb{Z}$, $A = \{(k, k) \mid k \in \mathbb{Z}\}$.

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

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 $n\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$ if and only if n is either zero or a prime number.

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

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

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*: $A = \{a \in R \mid (a, 0) \in \mathcal{A}\}$.

(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 $M_2(A)$ is an ideal of $M_2(R)$.

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 $R = \begin{bmatrix} \mathbb{F} & \mathbb{F} \\ 0 & \mathbb{F} \end{bmatrix}$ where $\mathbb{F}$ 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 $\operatorname{ann}(X) = \{a \in R \mid ax = 0 \text{ for all } x \in X\}$.

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

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

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

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

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

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 $X + Y = \{x + y \mid x \in X,\ y \in Y\}$.

(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) 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

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

(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

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

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

1.3.23 (Bonus problem). Prove that every 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,

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

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 equaiton $\bar{m}^3 - 6\bar{n}^3 = \bar{3}$ 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 $\ker f := \{x \in R \mid f(x) = 0\}$ and the *image* of f is $\operatorname{im} f = f(R) := \{f(r) \mid r \in R\}$.

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)**.** *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. □

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 $I = n\mathbb{Z}_{mn} = \{n\bar{a} \mid \bar{a} \in \mathbb{Z}_{mn}\}$. Then the map $f : \mathbb{Z}_{mn} \to \mathbb{Z}_n$, $f(\bar{x}) = \bar{x}$, 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

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

are also ideals of R.

Theorem 1.4.20. *If R is any ring, then $\mathbb{Z}1_R = \{k1_R \mid k \in \mathbb{Z}\}$ 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) that $\mathbb{Z}1_R$ is contained in the center of R.

We have $\ker\theta = \{k \in \mathbb{Z} \mid k1_R = 0\}$. 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

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

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

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

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 $a, b \in \mathbb{Z}$. Then $\bar{m}\bar{a} + \bar{n}\bar{b} = \bar{1}$ (where the bar denotes the residue modulo mn). Hence $m\mathbb{Z}_{mn} + n\mathbb{Z}_{mn} = \mathbb{Z}_{mn}$. Also, we have $m\mathbb{Z}_{mn} \cap n\mathbb{Z}_{mn} = \{0\}$ since if $m|x$ and $n|x$, then $mn|x$ (since m and n are relatively prime). Finally, we have $\mathbb{Z}_{mn}/m\mathbb{Z}_{mn} \cong \mathbb{Z}_m$ and $\mathbb{Z}_{mn}/n\mathbb{Z}_{mn} \cong \mathbb{Z}_n$. The result then 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

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

First note that the fourth 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 $25 + 60k$, $k \in \mathbb{Z}$, 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 $\theta(r, s) = rs$.

(d) $\theta\colon \mathcal{F}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$, 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. Furthermore, suppose θ is surjective, 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 an onto ring homomorphism, show that $\theta(Z(R)) \subseteq Z(S)$. Given an example showing this need not be equality.

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 $\theta(A) = \{\theta(a) \mid a \in A\}$ 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 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 $\bar{S} = \{(s, 0) \mid s \in S\}$. 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 $B/A = \{b + A \mid b \in B\}$ 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 $m = n$.

1.4.30 (Bonus problem). Let R be a finite commutative ring with $n = |R| < \infty$. Assume that R is not a field. Prove that R has at least $\sqrt{n-1}$ non-units.

1.4.31 (Bonus problem). Let $R = \mathcal{C}([0,1]) := \{f \in \mathcal{F}([0,1], \mathbb{R}) \mid f \text{ is continuous}\}$. One easily checks that this is a ring.

(a) Prove that, for every $a \in [0,1]$, the set $I_a := \{f \in R \mid f(a) = 0\}$ is a maximal ideal of R.

(b) Conversely, prove that every maximal ideal of R is of the form I_a for some $a \in [0,1]$.

(c) Prove that part (b) is false for the ring $\mathcal{C}(\mathbb{R}) := \{f \in \mathcal{F}(\mathbb{R}) \mid f \text{ is continuous}\}$.

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 the these polynomial rings. A reference for the material in this chapter is [Jud12, 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

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

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 $m > n$.

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

$$(1-3x+x^2)(2x-3) = (1-3x+x^2)(-3+2x) = -3+(2+9)x+(-6-3)x^2+2x^3 = -3+11x-9x^2+2x^3.$$

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*.

We leave the proof of the following theorem as an exercise.

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.

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 $\deg(f(x)g(x)) = \deg f(x) + g(x)$.

Proof. Exercise. □

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

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

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 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

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

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 $x - 2$.

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

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 $\mathbb{Z}_3[x]$.

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

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. 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

$$\varphi_a(f(x)+g(x)) = \varphi_a\left(\sum (a_i+b_i)x^i\right) = \sum (a_i+b_i)a^i$$
$$= \sum a_i a^i + \sum b_i a^i = \varphi_a(f(x)) + \varphi_a(g(x),$$

$$\varphi_a(f(x)g(x)) = \varphi_a\left(\sum_{i=0}^{2k}\left(\sum_{j=0}^{i} a_j b_{i-j} x^i\right)\right) = \sum_{i=1}^{2k}\left(\sum_{j=0}^{i} a_0 b_{i-j} a^i\right)$$
$$= \left(\sum_{i=0}^{k} a_i a^i\right)\left(\sum_{i=0}^{k} b_i a^i\right) = \varphi_a(f(x))\varphi_a(g(x)),$$
$$\varphi_a(1_{R[x]}) = \varphi(1_R) = 1_R.$$

□

The map φ_a of Theorem 2.1.18 is called *evaluation* of 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

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

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 $\deg rx < 1$. Thus we have that is a constant polynomial. That is, $r(x) = r$ for some $r \in R$. Then

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

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). □

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

$$f(x) \in \ker \varphi_a \iff f(a) = 0 \iff f(x) = q(x)(x-a) + 0 \iff f(x) \in \langle x-a\rangle.$$

The second assertion then follows from the First Isomorphism Theorem (Theorem 1.4.15). □

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 $x^2 - \bar{1} \in \mathbb{Z}_8$ are $1, 3, 5, 7$.

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 $a_1, \ldots, a_{n+1} \neq a$. 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 $n+1$ 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 $a_0, a_n \neq 0$. 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

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

Thus

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

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. Prove Theorem 2.1.8.

2.1.2. Prove Theorem 2.1.9.

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

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

2.1.5. 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.

2.2 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 $\mathbb{F}$ be a field. A polynomial $f(x) \in \mathbb{F}[x]$ is called *irreducible* over $\mathbb{F}$ 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 $\mathbb{F}[x]$, 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 $\mathbb{F}$ be a field and consider $f(x) \in \mathbb{F}[x] \setminus \{0\}$ such that $\deg f(x) \geq 2$.*

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

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

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

(b) Assume $p(x)$ has no root in $\mathbb{F}$. If $p(x)$ is reducible, then $p(x) = f(x)g(x)$ and so $f(x)$ and $g(x)$ have no roots in $\mathbb{F}$. Thus $\deg f(x) \neq 1$ and $\deg g(x) \neq 1$. But this contradicts the fact that $\deg f(x) + \deg g(x) = \deg p(x)$ is equal to 2 or 3. Thus $p(x)$ 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 $x^3 + \bar{2}x^2 + x + \bar{2}$ 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.

Corollary 2.2.13. *Suppose $f(x) \in \mathbb{C}[x]$ with $\deg f(x) \geq 1$. If we write $f(x) = a_0 + a_1x + \cdots + a_nx^n$, $a_n \neq 0$, 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 $f(x)$. 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 an exercise. □

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 $g(x)$. □

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 ring homomorphism $\varphi\colon \mathbb{Z}[x] \to \mathbb{Z}_p[x]$, $\varphi(\sum a_i x^i) = \sum \bar{a}_i x^i$ (called *reduction modulo p*). 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(e). Thus $\bar{f}(x) = 0$ 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. 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 $\mathbb{Q}[x]$ and hence no proper factorization in $\mathbb{Z}[x]$. The converse follows from part (a). □

If $\varphi\colon R \to S$ 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 φ to be usual ring homomorphism $\mathbb{Z} \to \mathbb{Z}_p$, $\varphi(x) = \bar{x}$, gives a ring homomorphism $\mathbb{Z}[x] \to \mathbb{Z}_p[x]$ called *reduction modulo p*.

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

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

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

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

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 $d^2 - d + \bar{2} = \bar{2}$. 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 $0 < k < s \leq n$.
- 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.

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

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

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

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

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

Proof. Let $d(x)$ be a monic element of $\{u(x)f(x) + v(x)g(x) \mid u(x), v(x) \in \mathbb{F}[x]\}$ 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

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

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 $d_1(x) = d_2(x)k_1(x) = d_1(x)k_2(x)k_1(x)$ for some $k_1(x), k_2(x) \in \mathbb{F}[x]$. 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 much 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 $f(x), g(x) \in \mathbb{F}[x]$ with $g(x) \neq 0$. Set $a_1(x) = f(x)$ and $a_2(x) = g(x)$. By repeated polynomial division, we have

$$\begin{cases} a_1(x) = q_1(x)a_2(x) + a_3(x) & \text{with } a_3(x) \neq 0 \text{ and } \deg a_3 9x) < \deg a_2(x), \\ a_2(x) = q_2(x)a_3(x) + a_4(x) & \text{with } a_4(x) \neq 0 \text{ and } \deg a_4(x) < \varphi a_3(x), \\ \quad \cdots & \cdots \\ a_{k-2}(x) = q_{k-2}(x)a_{k-1}(x) + a_k(x) & \text{with } a_k(x) \neq 0 \text{ and } \deg a_k(x) < \deg a_{k-1}(x), \\ a_{k-1}(x) = q_{k-1}(x)a_k(x). & \end{cases}$$

Then $\gcd(f(x), g(x)) = a_k(x)$. 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* $\mathbb{F}$ *be a field,* $p(x) \in \mathbb{F}[x]$ *an irreducible polynomial, and* $f_1(x), \ldots, f_n(x) \in \mathbb{F}[x]$. *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 $a \in \mathbb{F}$ and so $p(x)$ divides $f_1(x)$ since $d(x)$ does. On the other hand, if $d(x) = 1$, then we have $u(x)p(x) + v(x)f_1(x) = 1$ for some $u(x), v(x) \in \mathbb{F}[x]$ 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). Then result then follows by the inductive hypothesis. □

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

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 $\mathbb{F}[x]$ whenever $p(x)$ is irreducible (see Theorem 2.3.6).

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

(a) $f(x) = ap_1(x) \cdots p_m(x)$ where $a \in \mathbb{F}^\times$ and $p_1(x), \ldots, p_m(x) \in \mathbb{F}[x]$ 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 $a, b \in \mathbb{F}^\times$ 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 $f(x), g(x) \in \mathbb{F}[x]$, 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 $\gcd(a_{k-1}(x), a_k(x)) = a_k(x)$.

(d) Conclude that $\gcd(f(x), g(x)) = a_k(x)$.

2.2.4 (Bonus problem). (a) 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.

(b) Let $\theta : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ be a map such that

(i) for all $c \in \mathbb{R}$, $\theta\left(\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}\right) = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$,

(ii) for all $A, B \in M_2(\mathbb{R})$, we have $\theta(AB) = \theta(A)\theta(B)$,

(iii) for all $A, B \in M_2(\mathbb{R})$ and $a, b \in \mathbb{R}$, we have $\theta(aA + bB) = a\theta(A) + b\theta(B)$.

Prove that $\theta(X) = PXP^{-1}$ for some invertible $P \in M_2(\mathbb{R})$.

2.3 Quotient rings of polynomials over a field

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

Proof. Let I be a nonzero ideal of $\mathbb{F}[x]$. 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 $q(x), r(x) \in \mathbb{F}[x]$ 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 $I = \langle h(x) \rangle$. □

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

$$R \cong \{a_0 + a_1 t + \cdots + a_{n-1} t^{n-1} \mid a_0, \ldots, a_{n-1} \in \mathbb{F}\}$$

(isomorphism of vector spaces).

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

$$\mathbb{Z}_2[x]/\langle x^2 \rangle \cong \{at + b \mid a, b \in \mathbb{Z}_2\} = \{\bar{0}, \bar{1}, t, \bar{1} + t\}.$$

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

$$\mathbb{Z}_2[x]/\langle x^2 + x + \bar{1} \rangle \cong \{a + bt \mid a, b \in \mathbb{Z}_2\}.$$

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 $\mathbb{F}$ be a field and $h(x) \in \mathbb{F}[x]$ 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 $\mathbb{F}[x]$ such that $\langle h(x)\rangle \subsetneq I \subsetneq \mathbb{F}[x]$. Now, we know by Theorem 2.3.1 that $I = \langle p(x)\rangle$ for some $p(x) \in \mathbb{F}[x]$. Thus $h(x) = p(x)q(x)$ for some $q(x) \in \mathbb{F}[x]$. 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 $r(x) \in \mathbb{F}[x]$. This implies that $h(x) = p(x)h(x)r(x)$, and so $\deg p(x) = 0$, which contradicts the fact that $I \neq \mathbb{F}[x]$.

(c) $\Rightarrow$ (a): If $h(x)$ is not irreducible, then $h(x) = h_1(x)h_2(x)$ for some $h_1(x), h_2(x) \in \mathbb{F}[x]$ with $\deg h_1(x), \deg h_2(x) < \deg h(x)$. Thus $\langle h(x)\rangle \subsetneq \langle h_1(x)\rangle \subsetneq \mathbb{F}[x]$, 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. More precisely, it can be shown that a monic irreducible polynomial $h(x)$ of degree m exists in $\mathbb{Z}_p[x]$ for every such p and 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. It is straightforward to verify that this field has p^m elements.

Example 2.3.7. The polynomial $x^2 + x + 1$ has no root in $\mathbb{Z}_2$ and so it is irreducible by Theorem 2.2.7. Thus

$$\mathbb{Z}_2[x]/\langle x^2 + x + \bar{1}\rangle = \{a + bt \mid a, b \in \mathbb{Z}_2,\ t^2 + t + \bar{1} = 0\}$$

is a field with four elements.

Exercises.

2.3.1. Prove Theorem 2.3.1. *Hint:* Consider a monic element $f(x) \in I$ of smallest degree and follow the approach of Proposition 1.3.6.

2.3.2. 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).

2.3.3 (Bonus Problem). Prove that $\mathbb{C}[x]/\langle x^2 + 1\rangle \cong \mathbb{C} \times \mathbb{C}$ (as rings).

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 [Jud12, 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 $a \mid b$ 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 a or b 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) $a \mid b$ *and* $b \mid a$,

(b) $a = ub$ *for some* $u \in R^\times$,

(c) $\langle a \rangle = \langle b \rangle$.

Proof. Exercise. □

Definition 3.1.4 (Associates)**.** Two elements $a, b \in R$ are called *associates* if $a \mid b$ and $b \mid a$. 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. Exercise. □

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 $\mathbb{Z}(\sqrt{3}) := \{a+b\sqrt{3} \mid a, b \in \mathbb{Z}\}$. 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 $\sqrt{3} \sim 3+2\sqrt{3}$.

Example 3.1.8. Let's find all the units in $\mathbb{Z}(\sqrt{-5}) := \{a + b\sqrt{-5} \mid a, b \in \mathbb{C}\}$. We have

$$(a+b\sqrt{-5})(c+d\sqrt{-5}) = 1 \implies (a-b\sqrt{-5})(c-d\sqrt{-5}) = 1 \implies (a^2+5b^2)(c^2+5d^2) = 1.$$

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 3.1.9 (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 $r \mid ab$, then $r \mid a$ or $r \mid b$.

Example 3.1.10. Every prime number $p \in \mathbb{Z}$ is both irreducible and prime.

Example 3.1.11. Every irreducible polynomial $f(x) \in \mathbb{F}[x]$ (where $\mathbb{F}$ is a field) is both irreducible (by definition) and prime (by Proposition 2.2.28).

Theorem 3.1.12. *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 $r \mid ab$ and so $r \mid a$ or $r \mid b$. Without loss of generality (i.e. interchanging a and b if necessary), we can assume $r \mid a$. 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 3.1.13. We will show that $1+\sqrt{-5}$ is an irreducible element of $\mathbb{Z}(\sqrt{-5})$ that is not prime.

To see that it is irreducible, suppose that we have a factorization

$$1+\sqrt{-5} = (a+b\sqrt{-5})(c+d\sqrt{-5}).$$

Then, taking complex conjugates, we have

$$1 - \sqrt{-5} = (a - b\sqrt{-5})(c - d\sqrt{-5}).$$

Multiplying the sides of this equation by the corresponding sides of the previous one gives

$$6 = (a^2 + 5b^2)(c^2 + 5d^2).$$

Since each of the factors on the right is positive, we must have (interchanging the two factors if necessary)

- $a^2 + 5b^2 = 1$ and $c^2 + 5d^2 = 6$, in which case $a = 1$, $b = 0$; 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 $(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6 = 3 \cdot 2$. However, we claim that $1 + \sqrt{-5}$ does not divide 2 or 3. If it divided 2, we could find $x, y \in \mathbb{Z}$ with

$$2 = (1 + \sqrt{-5})(x + y\sqrt{-5}) \implies 2 = (1 - \sqrt{-5})(x - y\sqrt{-5}) \implies 4 = 6(x^2 + 5y^2),$$

which is a contradiction. Similarly, if $1 + \sqrt{-5}$ divided 3, we would have $x, y \in \mathbb{Z}$ with

$$3 = (1 + \sqrt{-5})(x + y\sqrt{-5}) \implies 3 = (1 - \sqrt{-5})(x - y\sqrt{-5}) \implies 9 = 6(x^2 + 5y^2),$$

which is again a contradiction. Thus $1 + \sqrt{-5}$ is not prime.

Definition 3.1.14 (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 3.1.15. The ring $\mathbb{Z}$ of integers is a UFD. Also, by Theorem 2.2.31, the ring $\mathbb{F}[x]$ is a UFD for any field $\mathbb{F}$.

Lemma 3.1.16. *Every irreducible element in a UFD is prime.*

Proof. Suppose R is a UFD and $r \in R$ is irreducible with $r \mid ab$ for some $a, b \in R$. Then we have $ab = rs$ for some $s \in RR$. 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 $r \mid a$ or $r \mid b$. □

Example 3.1.17. By Example 3.1.13 and Lemma 3.1.16, we see that $\mathbb{Z}(\sqrt{-5})$ is not a UFD.

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

$$d \mid a \iff d \sim p_1^{d_1} p_2^{d_2} \cdots p_r^{d_r}, \text{ for } 0 \leq d_i \leq a_i, \ i = 1, \ldots, r.$$

The proof of this is left as an exercise (or see [Nic12, §5.1, p. 258]).

Definition 3.1.18 (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) $d \mid s_i$ for $i = 1, 2, \ldots, n$.

(b) If $r \in R$ satisfies $r \mid s_i$ for $i = 1, 2, \ldots, n$, then $r \mid d$.

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) $s_i \mid m$ for $i = 1, 2, \ldots, n$.

(b) If $r \in R$ satisfies $s_i \mid r$ for $i = 1, 2, \ldots, n$, then $m \mid r$.

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 $\mathbb{F}[x]$ 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 $\gcd(s_1, \ldots, s_n)$ and $\operatorname{lcm}(s_1, \ldots, s_n)$ to denote *any* gcd and lcm. In particular, we have:

$$\text{if } s_i \sim s_i' \text{ for } i = 1, \ldots, n \quad \text{then} \quad \gcd(s_1, \ldots, s_n) \sim \gcd(s_1', \ldots, s_n').$$

Note that gcd's and lcm's need not exist in an arbitrary integral domain, as the following example illustrates.

Example 3.1.19. 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 3.1.20. *Gcds and lcms exists in UFDs. Suppose R is a UFD and $a, b, c, \ldots$ is a finite list of nonzero elements in R. Let $p_1, \ldots, p_r$ be the nonassociated primes dividing one of $a, b, c, \ldots$ and write*

$$a \sim p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}, \qquad a_i \in \mathbb{N},$$

$$\begin{aligned} b &\sim p_1^{b_1} p_2^{a_2} \cdots p_r^{b_r}, \qquad b_i \in \mathbb{N}, \\ c &\sim p_1^{b_1} p_2^{a_2} \cdots p_r^{c_r}, \qquad c_i \in \mathbb{N}. \\ &\vdots \end{aligned}$$

For each $i = 1, 2, \ldots, r$, *let* $d_i = \min(a_i, b_i, c_i, \ldots)$ *and* $m_i = \max(a_i, b_i, c_i, \ldots)$. *Then*

$$\gcd(a, b, c, \ldots) \sim p_1^{d_1} p_2^{d_2} \cdots p_r^{d_r} \quad \textit{and} \quad \operatorname{lcm}(a, b, c, \ldots) \sim p_1^{m_1} p_2^{m_2} \ldots p_r^{m_r}.$$

Proof. The proof when $R = \mathbb{Z}$ carries over to the general case. The details are left as an exercise. □

Definition 3.1.21 (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

$$\langle a_1 \rangle \subsetneq \langle a_2 \rangle \subsetneq \langle a_3 \rangle \cdots$$

of principal ideals.

Example 3.1.22. Suppose we have an chain of strictly increasing chain of principal ideals in $\mathbb{Z}$:

$$\{0\} \neq \langle m_1 \rangle \subsetneq \langle m_2 \rangle \subsetneq m_3 \rangle \subsetneq \cdots .$$

Then $|m_1| > |m_2| > |m_3| > \cdots$. Thus the chain must terminate after finitely many ideals. So $\mathbb{Z}$ satisfies the ACCP.

Example 3.1.23. Consider an ascending chain of principal ideals in $\mathbb{F}[x]$, where $\mathbb{F}$ is a field:

$$\{0\} \neq \langle p_1(x) \rangle \subsetneq \langle p_2(x) \rangle \subsetneq p_3(x) \rangle \subsetneq \cdots .$$

Then $\deg p_1(x) > \deg p_2(x) > \cdots$ and so the chain must be finite. Therefore $\mathbb{F}[x]$ satisfies the ACCP.

Example 3.1.24. Let $R = \{n + xf(x) \mid n \in \mathbb{Z},\ f(x) \in \mathbb{Q}[x]\}$, 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 3.1.25. *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 3.1.26. *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 3.1.20 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 $\{j_1, \ldots, j_\ell\} \subsetneq \{i_1, \ldots, i_k\}$. It follows that the chain is finite. Thus R satisfies the ACCP.

(b) ⇒ (c): Suppose (b) holds. Suppose $p \in R$ is irreducible and $p \mid ab$ in R. Let $d = \gcd(a, p)$. Then $d \mid p$ and so $d \sim p$ or $d \sim 1$ (since p is irreducible). If $p \sim d$ then, since $d \mid a$, we have $p \mid a$. 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 $p \mid b$ since $p \mid ab$ and $p \mid pb$. So we have shown that $p \mid a$ or $p \mid b$ and so p is prime.

It remains to prove the claim. Let $d' = \gcd(ab, pb)$. We have $b \mid ab$ and $b \mid pb$. Thus $b \mid d'$. 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 $u \mid a$. Similarly, $u \mid p$. So $u \mid \gcd(a, p) \sim 1$. Hence u is a unit.

(c) ⇒ (a): Suppose (c) holds. By Theorem 3.1.25, 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 $p_1 = q_1 \cdots q_s$. 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 $p_1 \mid q_1 \cdots q_s$, we must have $p_1 \mid q_j$ for some j since p_1 is prime by assumption. Relabeling if necessary, we may assume that $p_1 \mid q_1$. 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 3.1.27. The ring R of Example 3.1.24 is not a UFD since it does not satisfy the ACCP.

In 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 3.1.33). Before proving this, we will need to define a few terms and prove some preliminary results.

Definition 3.1.28 (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 3.1.29. 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 3.1.30. *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. Exercise. □

The following theorem is a generalization of Theorem 2.2.15

Theorem 3.1.31 (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 $f(x) = c(f(x))f_1(x)$ and $g(x) = c(g(x))g_1(x)$, where $f_1(x)$, $g_1(x)$ are primitive. Then

$$c(f(x)g(x)) \sim c(c(f(x))c(g(x))f_1(x)g_1(x)) \sim c(f(x))c(g(x))c(f_1(x)g_1(x))$$

by Lemma 3.1.30. Thus, it suffices to prove the result when $f(x)$ and $g(x)$ are primitive.

Let $f(x) = \sum_{i=0}^{m} a_i x^i$ and $g(x) = \sum_{i=0}^{n} b_i x^i$. Suppose that p is a prime dividing the coefficients of $f(x)g(x)$. 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 $f(x)g(x)$ is

$$c_{r+s} = a_0 b_{r+s} + a_1 b_{r+s-1} + \cdots + a_{r+s-1} b_1 + a_{r+s} b_0.$$

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 $a_r b_s$. However, since $p \mid c_{r+s}$, either p divides a_r or p divides b_s. This contradiction completes the proof. □

Proposition 3.1.32. *Suppose R is a UFD and let $\mathbb{F}$ be its field of quotients. We regard R as a subring of $\mathbb{F}$ 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 $\mathbb{F}[x]$.*

Proof. First suppose that $p(x)$ is irreducible in $\mathbb{F}[x]$. 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 $\mathbb{F}[x]$, which contradicts the fact that $p(x)$ is irreducible in $\mathbb{F}[x]$. 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 $p(x) = g(x)h(x)$ in $\mathbb{F}[x]$. Let a and b be the products of the denominators of the coefficients of $g(x)$ and $h(x)$ respectively. Then $g_1(x) := ag(x)$ and $h_1(x) := bh(x)$ are in $R[x]$ and

$$abp(x) = g_1(x)h_1(x)$$

is a factorization in $R[x]$. Thus, by Gauss' Lemma (Theorem 3.1.31), we have

$$ab \sim abc(p(x)) = c(abp(x)) = c(g_1(x)h_1(x)) \sim c(g_1(x))c(h_1(x)). \tag{3.1}$$

Now write $g_1(x) = c(g_1(x))g_2(x)$ and $h_1(x) = c(h_1(x))h_2(x)$, where $g_2(x)$ and $h_2(x)$ are primitive in $R[x]$. Thus

$$abp(x) = g_1(x)h_1(x) = c(g_1(x))c(h_1(x))g_2(x)h_2(x).$$

Therefore, by (3.1), we have $p(x) \sim g_2(x)h_2(x)$ in $R[x]$. Since $p(x)$ is irreducible in $R[x]$, this implies that either $g_2(x)$ or $h_2(x)$ is a unit in $R[x]$ (hence a unit in R). If $g_2(x) = u \in R^\times$, then

$$ag(x) = g_1(x) = c(g_1(x))g_2(x) = c(g_1(x))u \implies g(x) = a^{-1}c(g_1(x))u \in \mathbb{F}[x]^\times.$$

Similarly, if $h_2(x) \in R^\times$, then $h(x) \in \mathbb{F}[x]^\times$. □

Theorem 3.1.33. *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 $\mathbb{F}$ 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 $\mathbb{F}[x]$, where each $f_i(x)$ is irreducible in $\mathbb{F}[x]$. 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 $\mathbb{F}[x]$. Thus, since $f_i(x)$ is irreducible in $\mathbb{F}[x]$, so is $g_i(x)$. Then, by Proposition 3.1.32, each $g_i(x)$ is irreducible in $R[x]$. Now, we have

$$a_1 \cdots a_n p(x) = b_1 \cdots b_n g_1(x) \cdots g_n(x). \tag{3.2}$$

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 (3.2) by $a_1\cdots a_n$, we have $p(x) = ag_1(x)\cdots g_n(x)$, 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 3.1.32, each of the $f_i(x)$'s and $g_i(x)$'s is irreducible in $\mathbb{F}[x]$. The a_i's and the b_i's are units in $\mathbb{F}$. Since $\mathbb{F}[x]$ is a PID, it is a UFD; therefore, $n = s$. Now rearrange the $g_i(x)$'s so that $f_i(x)$ and $g_i(x)$ are associates 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 $c_i f_i(x) = d_i g_i(x)$. The polynomials $f_i(x)$ and $g_i(x)$ are primitive; hence, c_i and d_i are associates in R. Thus, $a_1\cdots a_m = ub_1\cdots b_r$ 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 3.1.34. *If R is a UFD, then so is $R[x_1,\ldots,x_n]$*

Proof. This follows by an easy induction using Theorem 3.1.33 and the fact that

$$R[x_1,\ldots,x_n] = (R[x_1,\ldots,x_{n-1}])[x_n].$$ □

Remark 3.1.35 (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. Prove Proposition 3.1.20.

3.1.4. Prove Lemma 3.1.30.

3.1.5. 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 $\gcd(a, b) \operatorname{lcm}(a, b) \sim ab$.

3.1.6 (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 $p(x) \mid (x^m - 1)$ 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 $\mathbb{F}[x]$ (where $\mathbb{F}$ 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 3.1.26, 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 $p \mid ab$ 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 $d \mid a$ and $d \mid p$. Since p is irreducible, this implies that $d \sim p$ or $d \sim 1$. If $d \sim p$, then $I = \langle p \rangle$ and so $p \mid a$ (since $a \in I$). 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 $p \mid ab$, this implies that $p \mid b$. 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 3.1.33) but not a PID since, for example, the ideal

$$I = \langle 2, x \rangle := \{a_0 + a_1 x + \cdots + a_n x^n \mid a_0 \in 2\mathbb{Z},\ a_1, \ldots, a_n \in \mathbb{N}\}$$

is not a principal ideal (exercise).

Remark 3.2.5. The fact that $\mathbb{Z}$ is a PID but $\mathbb{Z}[x]$ is not shows that the analogue of Theorem 3.1.33 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 $d = r_1 a_1 + \cdots + r_n a_n$ 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 $d = r_1 a_1 + \cdots + r_n a_n$ 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 $d \mid a_i$ for all i. Furthermore, if $c \mid a_i$ for all i, then $c \mid (r_1 a_1 + \cdots + r_n a_n)$ and so $c \mid d$. 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 $p \mid ab$, which implies that $p \mid a$ or $p \mid b$. 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

$$p \mid ab \implies ab \in \langle p \rangle \implies a \in \langle p \rangle \text{ or } b \in \langle p \rangle \implies p \mid a \text{ or } p \mid b.$$

Thus p is irreducible by Theorem 3.1.12. Now suppose $\langle p \rangle \subseteq \langle q \rangle \subsetneq R$. Then $q \mid p$. 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 $\mathbb{F}$ be a field. In $\mathbb{F}[x, y]$, the ideal $\langle x \rangle$ is prime, but $\langle x \rangle \subsetneq \langle x, y \rangle \subsetneq \mathbb{F}[x, y]$. Thus $\mathbb{F}[x, y]$ is not a PID. However, $\mathbb{F}[x, y]$ is a UFD by Corollary 3.1.34.

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 $\langle a_1 \rangle + \langle a_1 \rangle + \cdots + \langle a_n \rangle = \langle d \rangle$.

(b) $m \sim \operatorname{lcm}(a_1, \ldots, a_n)$ if and only if $\langle a_1 \rangle \cap \langle a_1 \rangle \cap \cdots \cap \langle a_n \rangle = \langle m \rangle$.

Proof. Part (a) follows from Theorem 3.2.6. The proof of part (b) is left as an exercise. □

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. Prove Proposition 3.2.10(b).

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:

$$a,b\in R\setminus\{0\} \implies \nu(ab)\geq\nu(a). \tag{3.3}$$

Such a divisor function is called a *euclidean valuation* (or *euclidean function*).

Remark 3.3.3. Condition (3.3) is actually superfluous in the following sense. Any domain R that has a division algorithm possesses a euclidean valuation (i.e. a divisor function satisfying (3.3)). 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.

Examples 3.3.4. (a) The ring $\mathbb{Z}$ is a euclidean domain with euclidean valuation $\nu\colon\mathbb{Z}\to\mathbb{N}$, $\nu(n)=|n|$.

(b) For any field $\mathbb{F}$, the ring $\mathbb{F}[x]$ is a euclidean domain with divisor function $\nu\colon\mathbb{F}[x]\setminus\{0\}\to\mathbb{N}$, $\nu(f(x))=\deg f(x)$ (see Theorem 2.1.12).

Theorem 3.3.5. *Every euclidean domain is a PID (and hence a UFD).*

Proof. Suppose R has a divisor function 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

$$a=qb+r \quad \text{where } r=0 \text{ or } \nu(r)<\nu(a).$$

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 [Jud12, §18.2, Example 8].) 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

$$\begin{aligned} \frac{9}{8} &= 1 + \frac{1}{8} \\ \frac{15}{8} &= 2 - \frac{1}{8}. \end{aligned}$$

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:

$$\text{commutative rings} \supsetneq \text{integral domains} \supsetneq \text{UFDs} \supsetneq \text{PIDs} \supsetneq \text{euclidean domains} \supsetneq \text{fields}.$$

Chapter 4

Fields

In this chapter we delve into the theory of fields. One of our motivating ideas will be roots of polynomials. We have seen many examples of polynomials $f(x) \in \mathbb{F}[x]$, where $\mathbb{F}$ is a field, that do not factor completely into linear factors. However, we will see that it is always possible to enlarge the field $\mathbb{F}$ so that $f(x)$ does completely factor. Furthermore, we will see that, up to isomorphism, a unique minimal such field exists, called the *splitting field* of $f(x)$. This construction will also allow us to completely classify all finite fields. A reference for the material in this chapter is [Jud12, Ch. 21].

4.1 Brief review of vector spaces

In this section we give a very brief review of the definition of a vector space over a field and some important results on vector spaces. For more details, see [Sav]. Through this section, $\mathbb{F}$ will denote a field.

Definition 4.1.1 (Vector space)**.** A *vector space* over $\mathbb{F}$ is an abelian group V (written additively and whose operation is called *vector addition*) and a *scalar multiplication* $\mathbb{F} \times V \to V$, $(a, v) \mapsto av$ such that

- $a(v + w) = av + aw$ for all $a \in \mathbb{F}$ and $v, w \in V$.
- $(a + b)v = av + bv$ for all $a, b \in \mathbb{F}$ and $v \in V$.
- $a(bv) = (ab)v$ for all $a, b \in \mathbb{F}$ and $v \in V$.
- $1v = v$ for all $v \in V$.

Example 4.1.2. The set $\mathbb{F}^n := \{(a_1, \ldots, a_n) \mid a_i \in \mathbb{F} \ \forall \ i\}$ is a vector space over $\mathbb{F}$ with the componentwise vector addition and the usual scalar multiplication.

Example 4.1.3. Any ring R such that $\mathbb{F}$ is a subring of R is vector space over $\mathbb{F}$. For example, $\mathbb{C}$ is a vector space over $\mathbb{R}$ (and over $\mathbb{Q}$).

Definition 4.1.4 (Subspace, linear combination, linear dependence, span, basis)**.** Suppose V is a subspace over $\mathbb{F}$. If $W \subseteq V$ and W is also a vector space over $\mathbb{F}$ (under the restricted operations), then W is a called a *subspace* of V.

If $v_1, \ldots, v_n \in V$, then a *linear combination* of $v_1, \ldots, v_n$ is an element of V of the form $c_1 v_1 + \cdots c_n v_n$, where $c_1, \ldots, c_n \in \mathbb{F}$.

The vectors $v_1, \ldots, v_n \in V$ are called *linearly dependent* if there exist $c_1, \ldots, c_n \in \mathbb{F}$, with at least one of the c_i nonzero, such that $c_1 v_1 + \cdots + c_n v_n = 0$. If $v_1, \ldots, v_n$ are not linearly dependent, then they are called *linearly independent*.

If $v_1, \ldots, v_n \in V$, then $\mathrm{Span}\{v_1, \ldots, v_n\} := \{c_1 v_1 + \cdots + c_n v_n \mid c_1, \ldots, c_n \in \mathbb{F}\}$. We write $\mathrm{Span}_{\mathbb{F}}$ when we wish to emphasize the field in question. We say V is *spanned* (or *generated*) by $v_1, \ldots, v_n$ if $V = \mathrm{Span}\{v_1, \ldots, v_n\}$.

We say $v_1, \ldots, v_n$ is a *basis* for V (over $\mathbb{F}$) if

(a) $V = \mathrm{Span}_{\mathbb{F}}\{v_1, \ldots, v_n\}$, and

(b) $v_1, \ldots, v_n$ are linearly independent.

Theorem 4.1.5. *Let V be a vector space over $\mathbb{F}$. Suppose that*

(a) $V = \mathrm{Span}\{v_1, \ldots, v_n\}$ for some $v_1, \ldots, v_n \in V$, and

(b) the vectors $w_1, \ldots, w_m \in V$ are linearly independent.

Then $m \leq n$.

Corollary 4.1.6. *If $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are two bases of a vector space V over $\mathbb{F}$, then $m = n$.*

Definition 4.1.7 (Dimension)**.** A vector space V over $\mathbb{F}$ is called *finite dimensional* if it has a finite basis. The number of elements of such a basis is called the *dimension* of V (over $\mathbb{F}$) and is denoted $\dim V$, or $\dim_{\mathbb{F}} V$ when we wish to emphasize the field $\mathbb{F}$. If V does not have a finite basis, we say it is *infinite dimensional*.

Example 4.1.8. The vectors $1, i$ form a basis of $\mathbb{C}$ over $\mathbb{R}$. Thus $\dim_{\mathbb{R}} \mathbb{C} = 2$. The vector 1 is a basis of $\mathbb{C}$ over $\mathbb{C}$, so $\dim_{\mathbb{C}} \mathbb{C} = 1$. The space $\mathbb{C}[x]$ is infinite dimensional over $\mathbb{C}$ with basis $1, x, x^2, x^3, \ldots$.

Theorem 4.1.9. *Let $V \neq \{0\}$ be a finite dimensional vector space.*

(a) Every linearly independent subset of V can be extended (i.e. enlarged) to a basis.

(b) Every spanning set of V has a subset which is a basis of V.

Exercises.

4.1.1 (Bonus problem). Show that $\mathbb{R}$ is infinite dimensional as a vector space over $\mathbb{Q}$.

4.2 Field extensions

We know many examples of fields that are subfields of larger fields. In this section, we will examine these inclusions of fields and some of their important properties.

Definition 4.2.1 (Field extension, finite extension)**.** A *field extension* is a pair of fields $\mathbb{F} \subseteq \mathbb{E}$. We say that $\mathbb{F}$ is a *subfield* of $\mathbb{E}$ and that $\mathbb{E}$ is an *extension* of $\mathbb{F}$. If $\mathbb{E}$, as a vector space over $\mathbb{F}$, is finite dimensional, then $\mathbb{E}$ is called a *finite extension* of $\mathbb{F}$, and $\dim_{\mathbb{F}} \mathbb{E}$ is denoted by $[\mathbb{E} : \mathbb{F}]$.

Example 4.2.2. We have that $\mathbb{R} \subseteq \mathbb{C}$ is a finite extension, but $\mathbb{Q} \subseteq \mathbb{R}$ is not a finite extension (see Exercise 4.1.1).

Theorem 4.2.3. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a finite extension and $[\mathbb{E} : \mathbb{F}] = n$. If $u \in \mathbb{E}$, then there exists a nonzero polynomial $f(x) \in \mathbb{F}[x]$ such that $f(u) = 0$ and $\deg f(x) \leq n$.*

Proof. Consider the elements $1, u, u^2, \ldots, u^n \in \mathbb{E}$. Since $\dim_{\mathbb{F}} \mathbb{E} = n$, this list of $n + 1$ elements must be linearly dependent by Theorem 4.1.5. Thus, there exist $a_0, a_1, \ldots, a_n$, with at least one $a_i \neq 0$, such that $a_0 + a_1 u + \cdots + a_n u^n = 0$. □

Definition 4.2.4 (Algebraic, transcendental)**.** Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. An element $u \in \mathbb{E}$ is called *algebraic over* $\mathbb{F}$ if $f(u) = 0$ for some nonzero polynomial $f(x) \in \mathbb{F}[x]$. If $u \in \mathbb{E}$ is not algebraic over $\mathbb{F}$, then it is called *transcendental over* $\mathbb{F}$. If every element of $\mathbb{F}$ is algebraic over $\mathbb{E}$, then we say $\mathbb{F} \subseteq \mathbb{E}$ is an *algebraic extension*.

Example 4.2.5. We see that $\mathbb{R} \subseteq \mathbb{C}$ is an algebraic extension since $a + bi$, $a, b \in \mathbb{R}$, is the root of $x^2 - 2ax + a^2 + b^2$.

Example 4.2.6. (a) The real number $\sqrt[3]{2}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^3 - 2$.

(b) The real number $\sqrt{5 + \sqrt{2}}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^4 - 10x^2 + 23$.

(c) The real number $\sqrt{2} - \sqrt{3}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^4 - 10x^2 + 1$.

It is natural ask if $\mathbb{Q} \subseteq \mathbb{R}$ is an algebraic extension. The answer is no. This follows from the next theorem.

Theorem 4.2.7 (Hermite 1873, Lindemann 1882)**.** *The real numbers e and π are transcendental over $\mathbb{Q}$.*

Remark 4.2.8. Since the set of polynomials with integer (or rational) coefficients is countable and each polynomial has a finite number of roots, the set of real numbers that are algebraic over $\mathbb{Q}$ is countable. Therefore, there are uncountably many real numbers that are transcendental over $\mathbb{Q}$.

Definition 4.2.9 (Field generated by elements)**.** If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension and $u_1, \ldots, u_n \in \mathbb{E}$, then

$$\mathbb{F}(u_1, \ldots, u_n) := \bigcap_{\substack{\mathbb{K} \text{ field} \\ \mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E} \\ u_1, \ldots, u_n \in \mathbb{K}}} \mathbb{K}.$$

In other words, $\mathbb{F}(u_1, \dots, u_n)$ is the smallest subfield of $\mathbb{E}$ containing $\mathbb{F}$ and $u_1, \dots, u_n$. It is called the field *generated by* $u_1, \dots, u_n$ *over* $\mathbb{F}$.

Example 4.2.10. We have $\mathbb{R}(i) = \mathbb{C}$ and $\mathbb{Q}(i, -i) = \mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$.

Example 4.2.11. We have $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ (as a subfield of $\mathbb{R}$ or $\mathbb{C}$) since the right hand side is contained in the left hand side and the right hand side is a field. Thus this notation matches our previous notation (see Example 1.2.9).

It follows from Theorem 4.2.3 that all finite extensions are algebraic. If $F \subseteq \mathbb{E}$ is an algebraic extension, is it necessarily true that $[\mathbb{E} : \mathbb{F}] < \infty$? The answer is no since, for example, $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \dots)$ is an algebraic extension, but is not a finite extension (see Exercise 4.2.4).

Lemma 4.2.12. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension and $u, v \in \mathbb{E}$. Then $\mathbb{F}(u)(v) = \mathbb{F}(u, v)$.*

Proof. We have

$$\mathbb{F}(u)(v) = \bigcap_{\substack{\mathbb{F}(u) \subseteq \mathbb{K} \subseteq \mathbb{E} \\ v \in \mathbb{K}}} \mathbb{K} \quad \text{and} \quad \mathbb{F}(u, v) = \bigcap_{\substack{\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E} \\ u, v \in \mathbb{K}}} \mathbb{K}.$$

Now, if $\mathbb{K}$ satisfies $\mathbb{F}(u) \subseteq \mathbb{K} \subseteq \mathbb{E}$ and $v \in \mathbb{K}$, then it satisfies $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ and $u, v \in \mathbb{K}$. So $\mathbb{F}(u)(v) \supseteq \mathbb{F}(u, v)$. On the other hand, if $\mathbb{K}$ satisfies $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ and $u, v \in \mathbb{K}$, then $\mathbb{F}(u) \subseteq \mathbb{K}$ and $v \in \mathbb{K}$. So $\mathbb{F}(u)(v) \subseteq \mathbb{F}(u, v)$. □

Corollary 4.2.13. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension and $u_1, \dots, u_n \in \mathbb{E}$. Then for every $1 \leq k \leq n - 1$, we have $\mathbb{F}(u_1, \dots, u_k)(u_{k+1}, \dots, u_n) = \mathbb{F}(u_1, \dots, u_n)$.*

Proof. This follows from Lemma 4.2.12 by induction. □

Remark 4.2.14. Corollary 4.2.13 implies that $\mathbb{F}(u_1, \dots, u_n)$ can be constructed via a chain of simple extensions:

$$\mathbb{F}(u_1, \dots, u_n) = \mathbb{F}(u_1)(u_2) \cdots (u_n).$$

Definition 4.2.15 (Minimal polynomial, degree)**.** Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension and $u \in \mathbb{E}$ be algebraic over $\mathbb{F}$. The *minimal polynomial* of u over $\mathbb{F}$ is the nonzero monic polynomial $m(x) \in \mathbb{F}[x]$ of smallest degree such that $m(u) = 0$. The *degree* of u over $\mathbb{F}$, denoted $\deg_{\mathbb{F}}(u)$, is the degree of $m(x)$.

Examples 4.2.16. (a) Consider the field extension $\mathbb{R} \subseteq \mathbb{C}$ and the element $i = \sqrt{-1} \in \mathbb{C}$. Then $m(x) = x^2 + 1$ and $\deg_{\mathbb{R}}(i) = 2$.

(b) Consider the field extension $\mathbb{Q} \subseteq \mathbb{R}$ and the element $\sqrt[3]{2} \in \mathbb{R}$. Then $m(x) = x^3 - 2$ and $\deg_{\mathbb{Q}}(m(x)) = 3$. How do we know that $m(x)$ is the monic polynomial of *minimal* degree such that $m(\sqrt[3]{2}) = 0$? Suppose $f(x) \in \mathbb{Q}[x]$ with $f(\sqrt[3]{2}) = 0$. Let $d(x) = \gcd(f(x), x^3 - 2)$ (in $\mathbb{Q}[x]$). Then

$$d(x) = a(x)f(x) + b(x)(x^3 - 2) \quad \text{for some } a(x), b(x) \in \mathbb{Q}[x].$$

Thus $d(\sqrt[3]{2}) = 0$ and so $d(x) \neq 1$. Since $d(x) \mid (x^3 - 2)$ and $x^3 - 2$ is irreducible by the Eisenstein Criterion, we therefore have $d(x) = x^3 - 2$. Thus $(x^3 - 2) \mid f(x)$.

(c) Suppose $\mathbb{F} \subseteq \mathbb{E}$ is a field extension and $u \in \mathbb{F}$. Then the minimal polynomial of u is $x - u$.

Theorem 4.2.17. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension and $u \in \mathbb{E}$ be algebraic over $\mathbb{F}$ with minimal polynomial $m(x) \in \mathbb{F}[x]$. Then the following statements hold:*

(a) $m(x)$ is irreducible in $\mathbb{F}[x]$.

(b) For every $f(x) \in \mathbb{F}[x]$, we have $f(u) = 0 \iff m(x) \mid f(x)$.

(c) $m(x)$ is uniquely determined by u.

Proof. (a) Suppose $m(x) = a(x)b(x)$ in $\mathbb{F}[x]$. Then $0 = m(u) = a(u)b(u)$ and so $a(u) = 0$ or $b(u) = 0$. Suppose, without loss of generality, that $a(u) = 0$. Then, by the minimality of $\deg m(x)$, we must have $\deg a(x) = \deg m(x)$. Hence $b(x) \in \mathbb{F}^\times$ (since it is nonzero and has degree zero). So $m(x)$ is irreducible.

(b) Clearly, if $m(x) \mid f(x)$, then $f(u) = 0$. Now suppose that $f(u) = 0$. Let $d(x) = \gcd(f(x), m(x))$. Then $d(x)$ is an $\mathbb{F}[x]$-linear combination of $f(x)$ and $m(x)$. So $d(u) = 0$, hence $d(x) \neq 1$. Thus $d(x) = m(x)$ (since $d(x)$ is a monic divisor of $m(x)$, which is irreducible by the above). Therefore $m(x) = d(x) \mid f(x)$.

(c) Suppose $m'(x)$ is another minimal polynomial for u. Then, by the above, we have $m(x) \mid m'(x)$ and $m'(x) \mid m(x)$. Since $m(x)$ and $m'(x)$ are both monic, this implies that $m(x) = m'(x)$.

□

Example 4.2.18. Let's find the minimal polynomial of $u = \sqrt{1 + \sqrt{3}}$ over $\mathbb{Q}$. We have

$$(u^2 - 1)^2 = 3 \implies u^4 - 2u^2 - 2 = 0.$$

Thus the minimial polynomial of u divides $u^4 - 2u^2 - 2$ by Theorem 4.2.17(b). Since $x^4 - 2x^2 - 2$ is irreducible (apply the Eisenstein Criterion with $p = 2$), it must be the minimal polynomial of u.

Theorem 4.2.19. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension, $u \in \mathbb{E}$ be algebraic over $\mathbb{F}$ and $d = \deg_{\mathbb{F}}(u)$.*

(a) $\mathbb{F}(u) = \{a_0 + a_1 u + \cdots + a_{d-1}u^{d-1} \mid a_0, \ldots, a_{d-1} \in \mathbb{F}\}$.

(b) $1, u, \ldots, u^{d-1}$ is a basis of $\mathbb{F}(u)$ over $\mathbb{F}$. In particular, $[\mathbb{F}(u) : \mathbb{F}] = \deg_{\mathbb{F}}(u)$.

(c) $\mathbb{F}(u) \cong \mathbb{F}[x]/\langle m(x) \rangle$, where $m(x)$ is the minimal polynomial of u over $\mathbb{F}$.

Proof. Set $\theta \colon \mathbb{F}[x] \to \mathbb{E}$, $\theta(f(x)) = f(u)$. Then θ is a ring homomorphism since it is the composition of the ring homomorphisms

$$\mathbb{F}[x] \hookrightarrow \mathbb{E}[x] \xrightarrow{\varphi_u} \mathbb{E},$$

where φ_u is the evaluation map at u. Since

$$\theta(f(x)) = 0 \iff f(u) = 0 \iff m(x) \mid f(x) \iff f(x) \in \langle m(x) \rangle,$$

we have $\ker\theta = \langle m(x)\rangle$. We also see that

$$\operatorname{im}\theta = \{a_0 + a_1u + \cdots + a_{d-1}u^{d-1} \mid a_i \in \mathbb{F}\} \tag{4.1}$$

since

$$f(x) \in \mathbb{F}[x] \implies f(x) = m(x)q(x) + r(x)$$

for some $r(x) = a_0 + a_1x + \cdots + a_{d-1}x^{d-1}$, $a_i \in \mathbb{F}$. Thus $f(u) = r(u) = a_0 + \cdots + a_{d-1}u^{d-1}$.

By the First Isomorphism Theorem (Theorem 1.4.15), we have $\operatorname{im}\theta \cong \mathbb{F}[x]/\langle m(x)\rangle$. Since $m(x)$ is irreducible and $\mathbb{F}[x]$ is a PID, the ideal $\langle m(x)\rangle$ is maximal by Lemma 3.1.16 and Theorem 3.2.7. Thus $\operatorname{im}\theta$ is a field by Corollary 1.3.19.

We claim that $\operatorname{im}\theta = \mathbb{F}(u)$. We have $u \in \operatorname{im}\theta$ and $\mathbb{F} \subseteq \operatorname{im}\theta$, so $\mathbb{F}(u) \subseteq \operatorname{im}\theta$. By (4.1), we also have $\operatorname{im}\theta \subseteq \mathbb{K}$ for every field $\mathbb{K}$ such that $\mathbb{F} \subseteq \mathbb{K} \subseteq \mathbb{E}$ with $u \in \mathbb{K}$. Hence $\operatorname{im}\theta \subseteq \mathbb{F}(u)$.

Finally, if $1, u, \ldots, u^{d-1}$ were linearly dependent, then we would have $\deg m(x) < d$, which is a contradiction. Hence $1, u, \ldots, u^{d-1}$ is a basis of $\mathbb{F}(u)$ over $\mathbb{F}$. □

Example 4.2.20. Let $u = 2 - i$. Let's describe the multiplication in $\mathbb{Q}(u)$. We have

$$(u-2)^2 = (-i)^2 = -1 \implies u^2 - 4u + 5 = 0.$$

Let $m(x) = x^2 - 4x + 5$. Since $m(x)$ has degree two and has no rational roots, it is irreducible over $\mathbb{Q}$. Thus $m(x)$ is the minimal polynomial of u over $\mathbb{Q}$. Therefore, by Theorem 4.2.19, we have

$$\mathbb{Q}(u) = \{a + bu \mid a, b \in \mathbb{Q}\}.$$

Since $u^2 = 4u - 5$, the multiplication in $\mathbb{Q}(u)$ is given by

$$(a_1 + b_1u)(a_2 + b_2u) = a_1a_2 + (a_1b_2 + a_2b_1)u + b_1b_2u^2 = (a_1a_2 - 5b_1b_2) + (a_1b_2 + a_2b_1 + 4b_1b_2)u.$$

Corollary 4.2.21. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. If $u, v \in \mathbb{E}$ have the same minimal polynomial, then $\mathbb{F}(u) \cong \mathbb{F}(v)$.*

Example 4.2.22. We have $\mathbb{Q}(\sqrt[3]{2}) = \{a_0 + a_1\sqrt[3]{2} + a_2\sqrt[3]{4} \mid a_0, a_1, a_2 \in \mathbb{Q}\}$ and $\mathbb{Q}(\sqrt[3]{2}) \cong \mathbb{Q}[x]/\langle x^3 - 2\rangle$. Furthermore, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. We also have $\mathbb{Q}(\zeta\sqrt[3]{2}) \cong \mathbb{Q}(\sqrt[3]{2})$ for any $\zeta \in \mathbb{C}$ satisfying $\zeta^3 = 1$, since $\sqrt[3]{2}$ and $\zeta\sqrt[3]{2}$ have the same minimal polynomial.

Example 4.2.23. Note that $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[n]{2}) \subseteq \mathbb{R}$ and the minimal polynomial of $\sqrt[n]{2}$ over $\mathbb{Q}$ is $x^n - 2$ (using the Eisenstein Criterion with $p = 2$). Thus $[\mathbb{R} : \mathbb{Q}]$ is not finite (see Exercise 4.1.1).

Theorem 4.2.24 (Multiplication Theorem)**.** *Let $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ be field extensions.*

(a) If $e_1, \ldots, e_m$ is a basis of $\mathbb{E}$ over $\mathbb{F}$ and $k_1, \ldots, k_n$ is a basis of $\mathbb{K}$ over $\mathbb{E}$, then $\{e_ik_j\}_{1\le i\le m, 1\le j\le n}$ is a basis of $\mathbb{K}$ over $\mathbb{F}$.

(b) $[\mathbb{K} : \mathbb{F}] < \infty$ if and only if $[\mathbb{K} : \mathbb{E}] < \infty$ and $[\mathbb{E} : \mathbb{F}] < \infty$.

(c) If $[\mathbb{K} : \mathbb{F}] < \infty$, then $[\mathbb{K} : \mathbb{F}] = [\mathbb{K} : \mathbb{E}][\mathbb{E} : \mathbb{F}]$.

Proof. (a) Let $v \in \mathbb{K}$. Then

$$v = \sum_{i=1}^{n} c_i k_i,\ c_i \in \mathbb{E} \implies v = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} d_{ij} e_j \right) k_i = \sum_{\substack{1 \le i \le n \\ 1 \le j \le m}} d_{ij} e_j k_i,\ d_{ij} \in \mathbb{F}.$$

Thus $\{e_i k_j\}_{1 \le i \le m,\, 1 \le j \le n}$ spans $\mathbb{K}$ over $\mathbb{F}$. Since

$$\sum_{i,j} f_{ij} e_i k_j = 0 \implies \sum_{j} \left(\sum_{i} f_{ij} e_i \right) k_j = 0 \implies \sum_{i} f_{ij} e_i = 0 \ \forall\ j \implies f_{ij} = 0 \ \forall\ i, j,$$

we see that the set $\{e_i k_j\}_{1 \le i \le m,\, 1 \le j \le n}$ is linearly independent over $\mathbb{F}$.

(b) If $[\mathbb{K} : \mathbb{F}] < \infty$, then $[\mathbb{K} : \mathbb{E}] < \infty$ since any basis for $\mathbb{K}$ over $\mathbb{F}$ is a spanning set for $\mathbb{K}$ over $\mathbb{E}$. Also $[\mathbb{E} : \mathbb{F}] < \infty$ since $\mathbb{E}$ is a subspace of $\mathbb{K}$ (over $\mathbb{F}$). The reverse implication follows from part (a).

(c) This follows from part (a). □

Corollary 4.2.25. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension, let $u \in \mathbb{E}$ be algebraic over $\mathbb{F}$, and let $v \in \mathbb{F}(u)$. Then v is also algebraic over $\mathbb{F}$ and $\deg_{\mathbb{F}}(v) \mid \deg_{\mathbb{F}}(u)$.*

Proof. We have $\mathbb{F} \subseteq \mathbb{F}(v) \subseteq \mathbb{F}(u)$. Since $[\mathbb{F}(u) : \mathbb{F}]$ is finite, so is $[\mathbb{F}(v) : \mathbb{F}]$. Hence v is algebraic by Theorem 4.2.3. Thus

$$\deg_{\mathbb{F}}(u) = [\mathbb{F}(u) : \mathbb{F}] = [\mathbb{F}(u) : \mathbb{F}(v)][\mathbb{F}(v) : \mathbb{F}] = [\mathbb{F}(u) : \mathbb{F}(v)] \deg_{\mathbb{F}}(v).$$ □

Example 4.2.26. Let $u = \sqrt[3]{2}$. Then $\mathbb{Q} \subseteq \mathbb{Q}(u^2) \subseteq \mathbb{Q}(u)$ and so, by Corollary 4.2.25, we have $[\mathbb{Q}(u^2) : \mathbb{Q}] \mid 3$. Since $\mathbb{Q}(u^2) \neq \mathbb{Q}$, we must have $[\mathbb{Q}(u^2) : \mathbb{Q}] = 3$ and so $[\mathbb{Q}(u) : \mathbb{Q}(u^2)] = 1$. Thus $\mathbb{Q}(u) = \mathbb{Q}(u^2)$.

Note, however, that if $u = \sqrt[4]{2}$, then $\mathbb{Q}(u^2) \subsetneq \mathbb{Q}(u)$ since $[\mathbb{Q}(u) : \mathbb{Q}] = 4$, but $[\mathbb{Q}(u^2) : \mathbb{Q}] = 2$.

Lemma 4.2.27. *Let $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. If $u, v \in \mathbb{E}$ are algebraic over $\mathbb{F}$, then uv and $u + v$ are algebraic over $\mathbb{F}$.*

Proof. We have $\mathbb{F} \subseteq \mathbb{F}(u) \subseteq \mathbb{F}(u, v) = \mathbb{F}(u)(v)$. Since v is algebraic over $\mathbb{F}$, it is also algebraic over $\mathbb{F}(u)$. Thus $[\mathbb{F}(u, v) : \mathbb{F}(u)] < \infty$. We also have $[\mathbb{F}(u) : \mathbb{F}] < \infty$. Therefore, by the Multiplication Theorem (Theorem 4.2.24), we have that $[\mathbb{F}(u, v) : \mathbb{F}] < \infty$ (since u is algebraic over $\mathbb{F}$). Therefore, by Theorem 4.2.3, any element of $\mathbb{F}(u, v)$ is algebraic over $\mathbb{F}$. In particular, $u + v$ and uv are algebraic over $\mathbb{F}$. □

Example 4.2.28. Let's find the minimal polynomial of $u = \sqrt{2} + \sqrt{5}$ over $\mathbb{Q}$. First note that

$$u^2 = 7 + 2\sqrt{10} \implies u^2 - 7 = 2\sqrt{10} \implies (u^2 - 7)^2 = 40 \implies u^4 - 14u^2 + 9 = 0.$$

Thus, if we set $f(x) = x^4 - 14u^2 + 9$, we have $f(u) = 0$. Is this the monic polynomial of *minimal* degree with this property?

Note that

$$(\sqrt{2}+\sqrt{5})^3 = 17\sqrt{2}+11\sqrt{5}.$$

Thus,

$$\sqrt{2} = \frac{1}{6}\Big((\sqrt{2}+\sqrt{5})^3 - 11(\sqrt{2}+\sqrt{5})\Big) \in \mathbb{Q}(\sqrt{2}+\sqrt{5}).$$

Therefore we also have $\sqrt{5} = (\sqrt{2}+\sqrt{5}) - \sqrt{2} \in \mathbb{Q}(\sqrt{2}+\sqrt{5})$. It follows that $\mathbb{Q}(\sqrt{2},\sqrt{5}) \subseteq \mathbb{Q}(\sqrt{2}+\sqrt{5})$. Since the reverse inclusion is obvious, we have $\mathbb{Q}(\sqrt{2}+\sqrt{5}) = \mathbb{Q}(\sqrt{2},\sqrt{5})$.

Note that $\deg_{\mathbb{Q}}(\sqrt{2}+\sqrt{5}) = [\mathbb{Q}(\sqrt{2}+\sqrt{5}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2},\sqrt{5}) : \mathbb{Q}]$. Now, $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ (since x^2-2 is irreducible). We claim that $[\mathbb{Q}(\sqrt{2},\sqrt{5}) : \mathbb{Q}(\sqrt{2})] = 2$. Note that $\sqrt{5}$ is a root of $x^2 - 5 \in \mathbb{Q}(\sqrt{2})[x]$, so the minimal polynomial of $\sqrt{5}$ over $\mathbb{Q}(\sqrt{2})$ divides x^2-5. It suffices to show that this polynomial is irreducible. Since it is of degree two, it is enough to show that it has no roots. But, for $a, b \in \mathbb{Q}$, we have

$$(a+b\sqrt{2})^2 = 5 \implies a^2+2b^2+2ab\sqrt{2} = 5 \implies \frac{a^2+2b^2-5}{2ab} = \sqrt{2}$$

(or $a = 0$ or $b = 0$, which would contradict the fact that $\sqrt{5}$ is irrational) which contradicts the fact that $\sqrt{2}$ is irrational. Therefore we have

$$\deg_{\mathbb{Q}}(\sqrt{2}+\sqrt{5}) = [\mathbb{Q}(\sqrt{2}+\sqrt{5}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}+\sqrt{5}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2 \cdot 2 = 4.$$

So $f(x)$ is indeed minimal.

Remark 4.2.29. In Example 4.2.28, we showed that $\mathbb{Q}(\sqrt{2}+\sqrt{5}) = \mathbb{Q}(\sqrt{2},\sqrt{5})$. In fact, one can show that if $\mathbb{F}$ is a field of characteristic zero and u, v are algebraic over $\mathbb{F}$, then $\mathbb{F}(u,v) = \mathbb{F}(w)$ for some $w \in \mathbb{F}(u,v)$. This is called the *Primitive Element Theorem* (see Exercise 4.2.5).

Theorem 4.2.30. *A field extension $\mathbb{F} \subseteq \mathbb{E}$ is finite if and only if $\mathbb{E} = \mathbb{F}(u_1, \ldots, u_n)$, where each $u_i \in \mathbb{E}$ is algebraic over $\mathbb{F}$.*

Proof. First suppose that $\mathbb{E} = \mathbb{F}(u_1, \ldots, u_n)$, where each $u_i \in \mathbb{E}$ is algebraic over $\mathbb{F}$. Consider the chain of extensions

$$\mathbb{F} \subseteq \mathbb{F}(u_1) \subseteq \mathbb{F}(u_1, u_2) \subseteq \cdots \subseteq \mathbb{F}(u_1, \ldots, u_n).$$

For $i = 1, \ldots, n$, the element u_i is algebraic over $\mathbb{F}(u_1, \ldots, u_{i-1})$ (we interpret this to be $\mathbb{F}$ if $i = 1$) since it is algebraic over $\mathbb{F}$. Thus $[\mathbb{F}(u_1, \ldots, u_n) : \mathbb{F}(u_1, \ldots, u_{n-1})]$ is finite. It then follows from the Multiplication Theorem (Theorem 4.2.24) that $[\mathbb{E} : \mathbb{F}]$ is finite.

Now suppose that $\mathbb{F} \subseteq \mathbb{E}$ is a finite extension. We prove the result by induction on $d = [\mathbb{F} : \mathbb{E}]$. If $d = 1$, then $\mathbb{F} = \mathbb{E}$ and we are done. So assume $d > 1$. Choose $u \in \mathbb{E} \setminus \mathbb{F}$. Then $[\mathbb{F}(u) : \mathbb{F}] > 1$. Thus, by the Multiplication Theorem (Theorem 4.2.24), we have

$$[\mathbb{E} : \mathbb{F}(u)] = \frac{[\mathbb{E} : \mathbb{F}]}{[\mathbb{F}(u) : \mathbb{F}]} < [\mathbb{E} : \mathbb{F}].$$

Thus, by the inductive hypothesis, we have $E = \mathbb{F}(u)(u_1, \ldots, u_n) = \mathbb{F}(u, u_1, \ldots, u_n)$. The elements $u, u_1, \ldots, u_n$ are algebraic over $\mathbb{F}$ by Theorem 4.2.3. □

Theorem 4.2.31. *If $\mathbb{F} \subseteq \mathbb{E} \subseteq \mathbb{K}$ are fields, then $\mathbb{F} \subseteq \mathbb{K}$ is an algebraic extension if and only if both $\mathbb{F} \subseteq \mathbb{E}$ and $\mathbb{E} \subseteq \mathbb{K}$ are algebraic extensions.*

Proof. If $\mathbb{F} \subseteq \mathbb{K}$ is an algebraic extension, then all elements of $\mathbb{K}$ are algebraic over $\mathbb{F}$, hence over $\mathbb{E}$. Also, all elements of $\mathbb{E}$, being also elements of $\mathbb{K}$, are algebraic over $\mathbb{F}$. So $\mathbb{F} \subseteq \mathbb{E}$ and $\mathbb{E} \subseteq \mathbb{K}$ are both algebraic extensions.

Now assume that $\mathbb{F} \subseteq \mathbb{E}$ and $\mathbb{E} \subseteq \mathbb{K}$ are algebraic extensions. Let $u \in \mathbb{K}$. Since $\mathbb{E} \subseteq \mathbb{K}$ is algebraic, there exists a nonconstant $f(x) \in \mathbb{E}[x]$ such that $f(u) = 0$. Let $f(x) = a_0 + a_1x + \cdots + a_mx^m$ (so $a_0, \ldots, a_m \in \mathbb{E}$) and consider the chain of field extensions

$$\mathbb{F} \subseteq \mathbb{F}(a_0) \subseteq \mathbb{F}(a_0, a_1) \subseteq \cdots \subseteq \mathbb{F}(a_0, \ldots, a_m) \subseteq \mathbb{F}(a_0, \ldots, a_m, u).$$

Then we have:

- $[\mathbb{F}(a_0, \ldots, a_m, u) : \mathbb{F}(a_0, \ldots, a_m)] \leq \deg f(x) < \infty$.
- Since $a_0 \in \mathbb{E}$, we know that a_0 is algebraic over $\mathbb{F}$, and so $[\mathbb{F}(a_0) : \mathbb{F}] < \infty$.
- Similarly, a_1 is algebraic over $\mathbb{F}$ and so $[\mathbb{F}(a_0, a_1) : \mathbb{F}(a_0)] \leq [\mathbb{F}(a_1) : \mathbb{F}] < \infty$.
- $\vdots$
- a_m is algebraic over $\mathbb{F}$, and so $[\mathbb{F}(a_0, \ldots, a_m) : \mathbb{F}(a_0, \ldots, a_{m-1})] \leq [\mathbb{F}(a_m) : \mathbb{F}] < \infty$.

Thus $[\mathbb{F}(a_0, \ldots, a_m, u) : \mathbb{F}] < \infty$. Therefore, by Theorem 4.2.3, u is algebraic over $\mathbb{F}$. □

Remark 4.2.32. The extension $\mathbb{Q} \subseteq \mathbb{Q}(\pi)$ is not algebraic.

Example 4.2.33. Let's find the minimal polynomial of $\sqrt[4]{2}$ over $\mathbb{Q}(\sqrt{2})$. Let $f(x) = x^2 - \sqrt{2} \in \mathbb{Q}(\sqrt{2})[x]$. Then $f(\sqrt[4]{2}) = 0$. Now consider

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[4]{2}).$$

We have $[\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}] = 4$ since $x^4 - 2$ is irreducible over $\mathbb{Q}$. Also, $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ since $x^2 - 2$ is irreducible over $\mathbb{Q}$. Thus $[\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}(\sqrt{2})] = 2$. Now, we clearly have $\mathbb{Q}(\sqrt[4]{2}) = \mathbb{Q}(\sqrt{2})(\sqrt[4]{2})$. By Theorem 4.2.19(b), we thus have $\deg_{\mathbb{Q}(\sqrt{2})}(\sqrt[4]{2}) = 2$. Thus $f(x)$ is the minimal polynomial of $\sqrt[4]{2}$ over $\mathbb{Q}(\sqrt{2})$.

Remark 4.2.34. It follows from the above example that $x^2 - \sqrt{2}$ is irreducible in $\mathbb{Q}(\sqrt{2})[x]$.

Exercises.

4.2.1. Show that $\sqrt{3 + \sqrt{5}}$ is algebraic over $\mathbb{Q}$.

4.2.2. Show that the intersection in Definition 4.2.9 is in fact a field. More generally, show that if $\mathbb{F}$ is a field, then the intersection of an arbitrary collection of subfields of $\mathbb{F}$ is a field.

4.2.3. Find the minimal polynomial of $u = \sqrt{3+\sqrt{6}}$ over $\mathbb{Q}$.

4.2.4 (Bonus problem). Show that $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2}, \sqrt[3]{2}, \sqrt[4]{2}, \dots)$ is an algebraic extension, but is not a finite extension.

4.2.5 (Bonus problem). Assume that $\operatorname{char} \mathbb{F} = 0$. Let $\mathbb{F} \subseteq \mathbb{E}$ be an algebraic extension and $\alpha_1, \dots, \alpha_n \in \mathbb{E}$. Prove that there exists a $\beta \in \mathbb{E}$ such that $\mathbb{F}(\alpha_1, \dots, \alpha_n) = \mathbb{F}(\beta)$.

4.3 Splitting fields

We know that, if $\mathbb{F}$ is a field, then a polynomial $f(x) \in \mathbb{F}[x]$ may not have any roots in $\mathbb{F}$, but may have roots in some extension $\mathbb{E}$ of $\mathbb{F}$. In this section, we will discuss this idea in further detail.

Theorem 4.3.1 (Kronecker's Theorem)**.** *If $\mathbb{F}$ is any field and $f(x) \in \mathbb{F}[x]$ is a nonconstant polynomial, then there exists an extension $\mathbb{F} \subseteq \mathbb{E}$ such that $f(x)$ has a root in $\mathbb{E}$.*

Proof. Let $p(x) \in \mathbb{F}[x]$ be irreducible such that $p(x) \mid f(x)$ (note that such a $p(x)$ exists since $\mathbb{F}[x]$ is a UFD). Then set $\mathbb{E} = \mathbb{F}[t]/\langle p(t)\rangle$. Since $p(t)$ is irreducible and $\mathbb{F}[t]$ is a PID, the ideal $\langle p(t)\rangle$ is maximal by Lemma 3.1.16 and Theorem 3.2.7. Thus $\mathbb{E}$ is a field by Corollary 1.3.19. We also have $\mathbb{F} \subseteq \mathbb{E}$ (viewing the elements of $\mathbb{F}$ as the classes of the constant polynomials). Since $f(t) \in \langle p(t)\rangle$, we have $f(t) = 0$ in $\mathbb{E}$. □

Example 4.3.2. The polynomial $f(x) = x^3 + x + 1 \in \mathbb{Z}_2[x]$ does not have a root in $\mathbb{Z}_2$. Since $f(x)$ has degree three, this implies that $f(x)$ is irreducible. Then $f(x)$ has a root in the field

$$\mathbb{E} = \mathbb{Z}_2[t]/\langle f(t)\rangle = \{a_0 + a_1 s + a_2 s^2 \mid a_0, a_1, a_2 \in \mathbb{Z}_2\},$$

where s denotes the image of t in $\mathbb{E}$. In fact, $f(x)$ factors completely in $\mathbb{E}[x]$:

$$f(x) = (x+s)(x^2 + sx + (1+s^2)) = (x+s)(x+s^2)(x+s+s^2).$$

Definition 4.3.3 (Splitting field)**.** Let $\mathbb{F}$ be a field and $f(x) \in \mathbb{F}[x]$ be a nonconstant polynomial. An extensions $\mathbb{F} \subseteq \mathbb{E}$ is called a *splitting field* of $f(x)$ over $\mathbb{F}$ if

(a) $f(x) = a(x-u_1)\cdots(x-u_n)$, $a, u_1, \dots, u_n \in \mathbb{E}$, and

(b) $\mathbb{E} = \mathbb{F}(u_1, \dots, u_n)$.

When (a) holds, we say that $f(x)$ *splits over* $\mathbb{E}$ (or *in* $\mathbb{E}[x]$).

Example 4.3.4. In Example 4.3.2, $\mathbb{E}$ is a splitting field of $x^3 + x + 1 \in \mathbb{Z}_2[x]$ over $\mathbb{Z}_2$

Example 4.3.5. We have that $\mathbb{Q}(\sqrt{2})$ is a splitting field of $x^2 - 2 \in \mathbb{Q}[x]$ over $\mathbb{Q}$, whereas $\mathbb{Q}(\sqrt[4]{2})$ is not.

Example 4.3.6. If $f(x) = x - a$, $a \in \mathbb{F}$, then $\mathbb{F}$ is a splitting field of $f(x)$ over $\mathbb{F}$.

Example 4.3.7. We have that $\mathbb{Q}(\sqrt{-1})$ is a splitting field of x^2+1 over $\mathbb{Q}$, but not of x^2+2. However, $\mathbb{R}(\sqrt{-1})$ is a splitting field of x^2+1 over $\mathbb{R}$ and is also a splitting field of x^2+2 over $\mathbb{R}$.

Theorem 4.3.8. *Let $\mathbb{F}$ be a field and $f(x) \in \mathbb{F}[x]$ be a nonconstant polynomial. Then there exists a splitting field $\mathbb{E}$ of $f(x)$ such that $[\mathbb{E}:\mathbb{F}] \leq n!$, where $n = \deg f(x)$.*

Proof. We prove the result by induction on n. If $n = 1$, then $\mathbb{E} = \mathbb{F}$ and we are done. Now assume $n > 1$. Set $\mathbb{E}_1 = \mathbb{F}[t]/\langle p(t)\rangle$ where $p(x)$ is an irreducible polynomial dividing $f(x)$. So $f(x)$ has a root in $\mathbb{E}$ and so we have

$$f(x) = (x-a)g(x) \text{ for some } g(x) \in \mathbb{E}_1[x],\ a \in \mathbb{E}_1.$$

Since $\deg g(x) \leq n-1$, by the induction hypothesis, $g(x)$ has a splitting field $\mathbb{E}$ such that $[\mathbb{E}:\mathbb{E}_1] \leq (n-1)!$. Since $[\mathbb{E}_1:\mathbb{F}] = \deg p(x) \leq \deg f(x) = n$, we have

$$[\mathbb{E}:\mathbb{F}] = [\mathbb{E}:\mathbb{E}_1][\mathbb{E}_1:\mathbb{F}] \leq n\cdot(n-1)! = n!. \qquad \square$$

Example 4.3.9. Let's find a splitting field $\mathbb{E}$ of $f(x) = x^4 - 2x^2 - 3 \in \mathbb{Q}[x]$ and find $[\mathbb{E}:\mathbb{Q}]$. Since $f(x) = (x^2-3)(x^2+1)$, we have that $\mathbb{E} = \mathbb{Q}(\sqrt{3},\sqrt{-1})$ is a splitting field. Consider the field extensions

$$\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{-1}) \subseteq \mathbb{Q}(\sqrt{-1},\sqrt{3}) = \mathbb{E}.$$

We have $[\mathbb{Q}(\sqrt{-1}):\mathbb{Q}] = 2$ since x^2+1 is irreducible in $\mathbb{Q}[x]$. Also, $[\mathbb{Q}(\sqrt{-1},\sqrt{3}):\mathbb{Q}(\sqrt{-1})] = 2$ since x^2-3 is irreducible in $\mathbb{Q}(\sqrt{-1})[x]$ (since $\sqrt{3} \notin \mathbb{Q}(\sqrt{-1})$). So $[\mathbb{E}:\mathbb{Q}] = 4$.

Example 4.3.10. Let's find a splitting field $\mathbb{E}$ of $f(x) = x^3 - 5 \in \mathbb{Q}[x]$ and find $[\mathbb{E}:\mathbb{Q}]$. We have

$$x^3 - 5 = (x - \sqrt[3]{5})(x - \omega\sqrt[3]{5})(x - \omega^2\sqrt[3]{5}), \quad \omega = e^{2\pi i/3}.$$

So $\mathbb{E} = \mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5}, \omega^2\sqrt[3]{5}) = \mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5})$. Consider the extensions $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{5}) \subseteq \mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5})$. We have

$$[\mathbb{Q}(\sqrt[3]{5}):\mathbb{Q}] = \deg(x^3-5) = 3.$$

Since $\mathbb{Q}(\sqrt[3]{5}) \subsetneq \mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5})$, we have

$$2 \leq [\mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5}) : \mathbb{Q}(\sqrt[3]{5})] = \deg m(x), \tag{4.2}$$

where $m(x)$ is the minimal polynomial of $\omega\sqrt[3]{5}$ in $\mathbb{Q}(\sqrt[3]{5})[x]$. Now, $\omega\sqrt[3]{5}$ is a root of x^3-5. Recalling the difference of cubes factorization $x^3 - a^3 = (x-a)(x^2+ax+a^2)$, we have

$$x^3 - 5 = (x - \sqrt[3]{5})(x^2 + \sqrt[3]{5}x + (\sqrt[3]{5})^2).$$

Since $\omega\sqrt[3]{5}$ is not a root of the first factor, it is a factor of the second. Thus, the degree of the minimal polynomial is at most two. Since, by (4.2), it is at least two, it must be equal to two (and $m(x) = x^2 + \sqrt[3]{5}x + (\sqrt[3]{5})^2$). Thus

$$[\mathbb{E}:\mathbb{Q}] = [\mathbb{Q}(\sqrt[3]{5}, \omega\sqrt[3]{5}) : \mathbb{Q}(\sqrt[3]{5})][\mathbb{Q}(\sqrt[3]{5}):\mathbb{Q}] = 2\cdot 3 = 6.$$

The above example illustrates the fact that the degree of a splitting field of a polynomial (over the base field) can be strictly greater than the degree of the polynomial.

Example 4.3.11. Let's find a splitting field $\mathbb{E} \supseteq \mathbb{Q}$ for $f(x) = x^3 - 1$ and find $[\mathbb{E} : \mathbb{Q}]$. We have

$$x^3 - 1 = (x-1)(x-\omega)(x-\omega^2), \quad \omega = e^{2\pi i/3}.$$

Thus $\mathbb{E} = \mathbb{Q}(1, \omega, \omega^2) = \mathbb{Q}(\omega)$. Now $x^3 - 1 = (x-1)(x^2+x+1)$. Since ω is not a root of the first factor, it is a root of the second. Because the roots of $x^2 + x + 1$ are not rational, this polynomial is irreducible over $\mathbb{Q}$. Thus $x^2 + x + 1$ is the minimal polynomial of ω over $\mathbb{Q}$. Hence $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$.

A natural question to ask is whether or not splitting fields are unique. Our next goal in this section is to show that they are unique up to isomorphism.

Suppose $\mathbb{F} \subseteq R$ and $\bar{\mathbb{F}} \subseteq \bar{R}$, where $\mathbb{F}$ and $\bar{\mathbb{F}}$ are fields that are subrings of the rings R and $\bar{R}$, respectively. If $\sigma\colon \mathbb{F} \to \bar{\mathbb{F}}$ is a ring homomorphism, then a ring homomorphism $\hat{\sigma}\colon R \to \bar{R}$ is said to *extend* σ if $\hat{\sigma}(a) = \sigma(a)$ for all $a \in \mathbb{F}$. In other words, the following diagram commutes (where the vertical maps are inclusions):

$$\begin{array}{ccc} R & \xrightarrow{\hat{\sigma}} & \bar{R} \\ \uparrow & & \uparrow \\ \mathbb{F} & \xrightarrow{\sigma} & \bar{\mathbb{F}} \end{array}$$

Example 4.3.12. Suppose $\sigma\colon \mathbb{F} \to \bar{\mathbb{F}}$ is an isomorphism of fields. For $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{F}[x]$, define $f^\sigma(x) := \sigma(a_0) + \sigma(a_1)x + \cdots + \sigma(a_n)x^n \in \bar{\mathbb{F}}[x]$. Then the mapping $\mathbb{F}[x] \to \bar{\mathbb{F}}[x]$, $f(x) \mapsto f^\sigma(x)$ is a ring isomorphism that extends σ (exercise).

Suppose $p(x)$ is a monic irreducible polynomial in $\mathbb{F}[x]$. Then $p^\sigma(x)$ is monic in $\bar{\mathbb{F}}[x]$ (since $\sigma(1) = 1$) and is irreducible (exercise). Consider the composition of ring homomorphisms

$$\mathbb{F}[x] \xrightarrow{f(x) \mapsto f^\sigma(x)} \bar{\mathbb{F}}[x] \twoheadrightarrow \mathbb{F}[x]/\langle p^\sigma(x) \rangle,$$

where the second map is the quotient homomorphism. It is easy to show that the kernel of this composition is $\langle p(x) \rangle$. Thus, by the First Isomorphism Theorem (Theorem 1.4.15), we have an induced ring isomorphism

$$\varphi\colon \mathbb{F}[x]/\langle p(x) \rangle \to \bar{\mathbb{F}}[x]/\langle p^\sigma(x) \rangle, \quad \varphi(f(x) + \langle p(x) \rangle) = f^\sigma(x) + \langle p^\sigma(x) \rangle \text{ for } f(x) \in \mathbb{F}[x]. \tag{4.3}$$

Theorem 4.3.13. *Suppose $\sigma : \mathbb{F} \to \bar{\mathbb{F}}$ is an isomorphism of fields and $p(x) \in \mathbb{F}[x]$ is monic and irreducible. Let u be a root of $p(x)$ in an extension field $\mathbb{E} \supseteq \mathbb{F}$ and let v be a root of $p^\sigma(x)$ in an extension field $\bar{\mathbb{E}} \supseteq \bar{\mathbb{F}}$. Then there is a unique isomorphism*

$$\mathbb{F}(u) \to \bar{\mathbb{F}}(v), \quad f(u) \mapsto f^\sigma(v), \ f \in \mathbb{F}[x],$$

that extends σ and maps u to v.

Proof. Since $p(x)$ is monic and irreducible, it is the minimal polynomial of u over $\mathbb{F}$. As in the proof of Theorem 4.2.19, we have a ring isomorphism $\mathbb{F}(u) \cong \mathbb{F}[x]/\langle p(x) \rangle$ given by

$f(u) \mapsto f(x) + \langle p(x)\rangle$. Similarly, $\bar{\mathbb{F}}[x]/\langle p^\sigma(x)\rangle \cong \mathbb{F}(v)$. Composing these isomorphisms with the isomorphism (4.3), we have

$$\begin{array}{ccccccc} \mathbb{F}(u) & \to & \mathbb{F}[x]/\langle p(x)\rangle & \to & \bar{\mathbb{F}}[x]/\langle p^\sigma(x)\rangle & \to & \bar{\mathbb{F}}(v), \\ f(u) & \mapsto & f(x) + \langle p(x)\rangle & \mapsto & f^\sigma(x) + \langle p^\sigma(x)\rangle & \mapsto & f^\sigma(v). \end{array}$$

Thus the composite map $\mathbb{F}(u) \to \bar{\mathbb{F}}(v)$, $f(u) \mapsto f^\sigma(v)$ is an isomorphism. Taking $f(x) = x$, we see that this map sends u to v. Considering elements of $\mathbb{F}$ as a constant polynomials, we also see that the map extends σ. □

Example 4.3.14. Let's consider Theorem 4.3.13 in the special case that $\mathbb{F} = \bar{\mathbb{F}}$. Let $p(x)$ be a monic irreducible polynomials of degree n and let u and v be two roots of $p(x)$ in extension fields $\mathbb{E} \supseteq \mathbb{F}$ and $\bar{\mathbb{E}} \supseteq \mathbb{F}$ respectively. Then the map

$$\hat{\sigma}\colon \mathbb{F}(u) \to \mathbb{F}(v), \quad \hat{\sigma}(a_0 + a_1 u + \cdots + a_{n-1}u^{n-1}) = a_0 + a_1 v + \cdots + a_{n-1}v^{n-1},$$

is an isomorphism that *fixes* $\mathbb{F}$ (i.e. $\hat{\sigma}(a) = a$ for all $a \in \mathbb{F}$) and maps u to v. For example $\mathbb{Q}(\sqrt[3]{5}) \cong \mathbb{Q}(e^{2\pi i/3}\sqrt[3]{5})$ since $\sqrt[3]{5}$ and $e^{2\pi i/3}\sqrt[3]{5}$ are both roots of the irreducible polynomial $x^3 - 5 \in \mathbb{Q}[x]$.

The following theorem shows that splitting fields are unique.

Theorem 4.3.15. *Suppose $\sigma\colon \mathbb{F} \to \bar{\mathbb{F}}$ is an isomorphism of fields and let $f(x) \in \mathbb{F}[x]$ be a nonconstant polynomial. If $\mathbb{E} \supseteq \mathbb{F}$ is a splitting field for $f(x)$ and $\bar{\mathbb{E}} \supseteq \bar{\mathbb{F}}$ is a splitting field for $f^\sigma(x)$, then there is an isomorphism $\mathbb{E} \to \bar{\mathbb{E}}$ that extends σ.*

Proof. We prove the result by induction on $n = \deg f(x) = \deg f^\sigma(x)$. If $n = 1$, then $\mathbb{E} = \mathbb{F}$ and $\bar{\mathbb{E}} = \bar{\mathbb{F}}$, so σ itself has the desired properties.

Now assume $n > 1$. Let $p(x)$ be a monic irreducible divisor of $f(x)$. Let $u \in \mathbb{E}$ be a root of $f(x)$ and let $v \in \bar{\mathbb{E}}$ be a root of $p^\sigma(x)$. By Theorem 4.3.13, σ extends to an isomorphism $\tau\colon \mathbb{F}(u) \to \bar{\mathbb{F}}(v)$ such that $\tau(u) = v$.

$$\begin{array}{ccc} \mathbb{E} & \xrightarrow{\psi} & \bar{\mathbb{E}} \\ \uparrow & & \uparrow \\ \mathbb{F}(u) & \xrightarrow{\tau} & \bar{F}(v) \\ \uparrow & & \uparrow \\ \mathbb{F} & \xrightarrow{\sigma} & \bar{\mathbb{F}} \end{array}$$

Write

$$f(x) = a(x-u)g(x), \quad g(x) \in \mathbb{F}(u)[x],\ \deg g(x) = n-1.$$

Then we have

$$f^\sigma(x) = f^\tau(x) = \sigma(a)(x-v)g^\tau(x), \quad g^\tau(x) \in \bar{\mathbb{F}}(v)[x],\ \deg g^\tau(x) = n-1.$$

It follows that $\bar{\mathbb{E}}$ is a splitting field for $g^\tau(x)$ (since it is a splitting field for $f^\sigma(x)$). We also have that $\mathbb{E}$ is a splitting field for $g(x)$ (since it is a splitting field for $f(x)$). Thus, by the induction hypothesis, there is an isomorphism $\psi : \mathbb{E} \to \bar{\mathbb{E}}$ that extends τ and hence extends σ. □

Recall that the Fundamental Theorem of Algebra states that every polynomial in $\mathbb{C}[x]$ splits. In the remainder of this section, we discuss this property of fields.

Theorem 4.3.16. *If $\mathbb{F}$ is a field, then the following conditions are equivalent:*

(a) Every nonconstant polynomial in $\mathbb{F}[x]$ has a root in $\mathbb{F}$.

(b) Every irreducible polynomial in $\mathbb{F}[x]$ has degree one.

(c) Every nonconstant polynomial in $\mathbb{F}[x]$ splits in $\mathbb{F}[x]$.

(d) If $\mathbb{F} \subseteq \mathbb{E}$ is an algebraic extension, then $\mathbb{E} = \mathbb{F}$.

Proof. (a) $\Rightarrow$ (b): This follows immediately from the Factor Theorem (Theorem 2.1.21).

(b) $\Rightarrow$ (c): Since $\mathbb{F}[x]$ is a UFD, every polynomial is a product of irreducibles, hence a product of degree one polynomials by (b).

(c) $\Rightarrow$ (d): If $u \in \mathbb{E}$, let $f(x) \in \mathbb{F}[x]$ be a nonzero polynomial such that $f(u) = 0$ (such a polynomial exists since $\mathbb{F} \subseteq \mathbb{E}$ is an algebraic extension). Then $f(x)$ in nonconstant and so, by (c),

$$f(x) = a(x - b_1)(x - b_2) \cdots (x - b_n) \text{ for some } a, b_1, \ldots, b_n \in \mathbb{F}.$$

Since $f(u) = 0$, we have $u = b_i$ for some i. Thus $u \in \mathbb{F}$.

(d) $\Rightarrow$ (a): Suppose $f(x) \in \mathbb{F}[x]$ is nonconstant. By Theorem 4.3.1, we can find a root u of $f(x)$ in some extension field $\mathbb{E}$. Thus $\mathbb{F} \subseteq \mathbb{F}(u)$ is an algebraic extension (since it is finite by Theorem 4.2.19). Thus $\mathbb{F}(u) = \mathbb{F}$ by (d) and so $u \in \mathbb{F}$. □

Definition 4.3.17 (Algebraically closed)**.** A field is *algebraically closed* if it satisfies the conditions of Theorem 4.3.16.

Example 4.3.18. The field $\mathbb{C}$ of complex numbers is algebraically closed.

Proposition 4.3.19. *Suppose $\mathbb{F} \subseteq \mathbb{E}$ is a field extension. Let*

$$A = \{u \in \mathbb{E} \mid u \text{ is algebraic over } \mathbb{F}\}.$$

Then A is a subfield of $\mathbb{E}$ and so is an algebraic extension of $\mathbb{F}$.

Proof. For any $u, v \in A$, the field $\mathbb{F}(u, v)$ is a finite field extension of $\mathbb{F}$ by Theorem 4.2.30. Thus, all elements of $\mathbb{F}(u, v)$ are algebraic over $\mathbb{F}$ by Theorem 4.2.3. So $\mathbb{F}(u, v) \subseteq A$. In particular, $u + v, uv, -u \in A$. Furthermore, if $u \neq 0$, then $u^{-1} \in A$. So A is a field. □

Definition 4.3.20 (Algebraic closure)**.** If $\mathbb{F} \subseteq \mathbb{E}$ is a field extension, then the field A in Proposition 4.3.19 is called the *algebraic closure* of $\mathbb{F}$ in $\mathbb{E}$.

If $\mathbb{F}$ is a field, a field extension $\mathbb{F} \subseteq \mathbb{E}$ is called an *algebraic closure* of $\mathbb{F}$ if $\mathbb{E}$ is an algebraic extension of $\mathbb{F}$ and $\mathbb{E}$ is algebraically closed. (Note the subtle difference in terminology – here the closure is not *in* any larger field.)

Proposition 4.3.21. *Suppose $\mathbb{F} \subseteq \mathbb{E}$ is a field extension and $\mathbb{E}$ is algebraically closed. Then the algebraic closure of $\mathbb{F}$ in $\mathbb{E}$ is itself algebraically closed.*

Proof. Let $A = \{u \in \mathbb{E} \mid u \text{ is algebraic over } \mathbb{F}\}$ be the algebraic closure of $\mathbb{F}$ in $\mathbb{E}$. Suppose

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

is a nonconstant polynomial in $A[x]$. Then $f(x) \in \mathbb{E}[x]$ and so $f(x)$ has a root u in $\mathbb{E}$. Let $\mathbb{K} = \mathbb{F}(a_0, a_1, \ldots, a_n)$. Then $[\mathbb{K} : \mathbb{F}]$ is a finite extension by Theorem 4.2.30, since each $a_i \in A$, and so is algebraic over $\mathbb{F}$. Furthermore $[\mathbb{K}(u) : \mathbb{K}]$ is finite because u is algebraic over $\mathbb{K}$. Thus $[\mathbb{K}(u) : \mathbb{F}]$ is finite and hence algebraic. Since $u \in \mathbb{K}(u)$, it follows that u is algebraic over $\mathbb{F}$. Thus every noncontant polynomial in $A[x]$ has a root in A. So A is algebraically closed. □

Definition 4.3.22 (Field of algebraic numbers)**.** The *field of algebraic numbers* is

$$\mathbb{A} = \{u \in \mathbb{C} \mid u \text{ is algebraic over } \mathbb{Q}\}.$$

By Proposition 4.3.21, $\mathbb{A}$ is an algebraic closure of $\mathbb{Q}$.

Example 4.3.23. We see from Example 4.2.23 that $\mathbb{A}$ is an algebraic extension of $\mathbb{Q}$ that is not finite.

Theorem 4.3.24. *Every field $\mathbb{F}$ has an algebraic closure $\mathbb{E} \supseteq \mathbb{F}$. Furthermore, if $\bar{\mathbb{E}} \supseteq \mathbb{F}$ is another algebraic closure, then there is an isomorphism $\sigma\colon \mathbb{E} \to \bar{\mathbb{E}}$ that fixes $\mathbb{F}$.*

The proof of Theorem 4.3.24 requires Zorn's Lemma and will be omitted.

Exercises.

4.3.1. Prove that the map $f(x) \mapsto f^\sigma(x)$ defined in Example 4.3.12 is a ring isomorphism.

4.4 Finite fields

In this final section, we look at finite fields. In particular, we will completely classify them up to isomorphism.

Suppose $\mathbb{F}$ 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, $\mathbb{Z}_p \subseteq \mathbb{F}$ (the map $\iota : \mathbb{Z}_p \hookrightarrow \mathbb{F}$, $\iota(\bar{k}) = k \cdot 1$ is an injective ring homomorphism).

Theorem 4.4.1. *If $\mathbb{F}$ is a finite field, then $|\mathbb{F}| = p^n$ for some n, where $p = \operatorname{char} \mathbb{F}$.*

Proof. Since $\mathbb{F}$ is a finite field, $[\mathbb{F} : \mathbb{Z}_p]$ is finite. If $n = [\mathbb{F} : \mathbb{Z}_p]$, then $\mathbb{F} \cong (\mathbb{Z}_p)^n$ as a vector space and the result follows. □

Theorem 4.4.2. *Let p be a prime number and let $\mathbb{F} \supseteq \mathbb{Z}_p$ be a splitting field of $f(x) = x^{p^n} - x$. Then $|\mathbb{F}| = p^n$.*

Proof. Let $S = \{a \in \mathbb{F} \mid a^{p^n} - a = 0\}$. We will prove that $\mathbb{Z}_p \subseteq S$ and S is a field (hence $S = \mathbb{F}$). Fermat's Little Theorem implies that for any integer k, $k^p \cong k \mod p$. Thus, for all $\bar{k} \in \mathbb{Z}_p$, we have

$$\bar{k}^{p^n} = \left(\bar{k}^{p^{n-1}}\right)^p = \bar{k}^{p^{n-1}} = \bar{k}^{p^{n-2}} = \cdots = k.$$

Therefore $\mathbb{Z}_p \subseteq S$.

Now,

$$a, b \in S \implies (a+b)^p = \sum_{i=0}^{p} \binom{p}{i} a^i b^{p-i} = a^p + b^p.$$

Therefore

$$(a+b)^{p^n} = ((a+b)^p)^{p^{n-1}} = (a^p + b^p)^{p^{n-1}} = \left(a^{p^2} + b^{p^2}\right)^{p^{n-2}} = \cdots = a^{p^n} + b^{p^n} = a + b.$$

Thus $a + b \in S$. So S is closed under addition. Also

$$a \in S \implies (-a)^{p^n} = (-1)^{p^n} a^{p^n} = (-1)^{p^n} a = -a,$$

since if $p > 2$, then p is odd, and if $p = 2$, then $-1 = 1$.

We also have

$$a, b \in S \implies (ab)^{p^n} = a^{p^n} b^{p^n} = ab \implies ab \in S,$$

and

$$a \in S \implies \left(\frac{1}{a}\right)^{p^n} = \frac{1}{a^{p^n}} = \frac{1}{a} \implies \frac{1}{a} \in S.$$

Thus S is a field and so $S = \mathbb{F}$.

It remains to find $|\mathbb{F}| = |S|$. In $\mathbb{F}[x]$ we have

$$f(x) = (x - a_1)^{m_1}(x - a_2)^{m_2} \cdots (x - a_r)^{m_r}, \quad r = |S|,\ a_1, \ldots, a_r \text{ distinct.}$$

If we can show that $m_1 = m_2 = \cdots = m_r = 1$, then it will follow that $r = p^n$. Assume, towards a contradiction, that $m_i > 1$ for some i. Then $f(x) = (x - a_i)^2 g(x)$ for some $g(x) \in \mathbb{F}[x]$. Thus $f(a_i) = 0$ and

$$f'(x) = 2(x - a_i)g(x) + (x - a_i)^2 g'(x)$$

(using the usual rules of differentiation: the derivative is $\mathbb{F}$-linear and the derivative of x^n is nx^{n-1}). Thus $f'(a_i) = 0$. However, since $f(x) = x^{p^n} - x$, we have $f'(x) = -1$, which is never zero. This contradiction completes the proof. □

Theorem 4.4.3. *If $\mathbb{F}$ is a field with p^n elements, where $p = \operatorname{char} \mathbb{F}$, then $\mathbb{F}$ is a splitting field of $f(x) = x^{p^n} - x$.*

Proof. We have $|\mathbb{F}^\times| = p^n - 1$. Thus, by Lagrange's Theorem, we have $a^{p^n - 1} = 1$ for all $a \in \mathbb{F}^\times$. Thus $a^{p^n} = a$ for all $a \in \mathbb{F}$ and so every element of $\mathbb{F}$ is a root of $f(x)$.

Since $\deg f(x) = p^n$, $f(x)$ has at most p^n roots in $\mathbb{F}$. Thus

$$p^n = |\mathbb{F}| \leq \ \#\text{ of roots of } f(x) \leq p^n.$$

Therefore, $\mathbb{F}$ is the set of *all* roots of $f(x)$. In particular, it is a splitting field of $f(x)$. □

Corollary 4.4.4. *If p is a prime number and $n \geq 1$, then, up to isomorphism, there exists a unique field with p^n elements.*

Proof. This follows from Theorems 4.4.3 and 4.3.15. □

Definition 4.4.5 (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 4.4.6. *(a) If $\mathbb{K} \subseteq \mathrm{GF}(p^n)$ is a subfield, then $\mathbb{K} \cong \mathrm{GF}(p^m)$, where $m \mid n$.*

(b) If $m \mid n$, then $\mathrm{GF}(p^n)$ has a unique subfield of p^m elements (which is therefore isomorphic to $\mathrm{GF}(p^m)$).

Proof. (a) We have

$$[\mathrm{GF}(p^n) : \mathbb{K}][\mathbb{K} : \mathbb{Z}_p] = [\mathrm{GF}(p^n) : \mathbb{Z}_p] = n \implies [\mathbb{K} : \mathbb{Z}_p] \mid n.$$

(b) If $\mathbb{K} \subseteq \mathrm{GF}(p^n)$ is a subfield with $|\mathbb{K}| = p^m$, then $\mathbb{K}$ is the splitting field of $x^{p^m} - x$ by Theorem 4.4.3 and hence is unique. It remains to prove existence. Note that

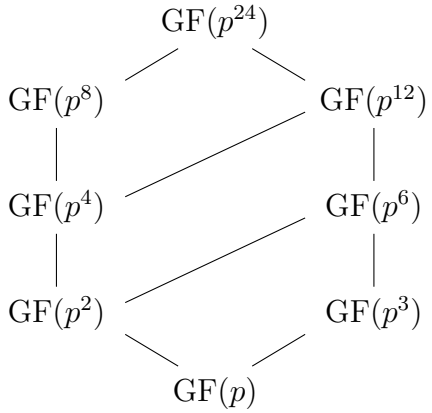
$$a^{p^m} = a \implies a^{p^{2m}} = (a^{p^m})^{p^m} = a^{p^m} = a.$$

Continuing in this manner, we see that $a^{p^n} = a$. Thus $x^{p^m} - x$ splits in $\mathrm{GF}(p^n)$. Set

$$\mathbb{E} = \{u \in \mathrm{GF}(p^n) \mid u \text{ is a root of } x^{p^m} - x\}.$$

Then $|\mathbb{E}| = p^m$ because the roots of $x^{p^m} - x$ are distinct. Furthermore, $\mathbb{E}$ is a field as in the proof of Theorem 4.4.2. □

Example 4.4.7. The lattice of subfields of $\mathrm{GF}(p^{24})$ is as follows:



Theorem 4.4.8. *Suppose $\mathbb{F}$ is a field. If G is a finite subgroup of $\mathbb{F}^\times$, then G is cyclic.*

Proof. Let G be a finite subgroup of $\mathbb{F}^\times$ 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 $\mathbb{F}$, $n \leq m$. 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 4.4.9. *The multiplicative group of all nonzero elements of a finite field is cyclic.*

Definition 4.4.10 (Primitive element, primitive root of unity)**.** If $\mathbb{F}$ is a finite field, then a generator for $\mathbb{F}^\times$ is called a *primitive element* of $\mathbb{F}$. If a (possibly infinite) field $\mathbb{F}$ has a multiplicative subgroup G of order n, then a generator of G (which exists by Theorem 4.4.8) is called a *primitive nth root of unity* in $\mathbb{F}$. For example, $e^{2\pi i/n}$ is a primitive nth root of unity in $\mathbb{C}$ for each $n \geq 2$.

Index

ACCP, 53
algebraic closure, 75
algebraic element, 64
algebraic extension, 64
algebraic numbers, 76
algebraically closed, 75
ann(X), 22
annihilator, 22
artinian ring, 57
ascending chain condition on ideals, 57
ascending chain condition on principal ideals, 53
associates, 49
automorphism, 25
    inner, 25

basis, 63

$c(f(x))$, 55
center, 8
centralizer, 10
characteristic, 7
Chinese Remainder Theorem, 27
coefficient, 33
commutative ring, 4
constant coefficient, 34
constant polynomial, 34
content, 55
cyclotomic polynomial, 43, 45

degree
    of a polynomial, 34
    of an element in a field extension, 65
descending chain condition on ideals, 57
difference of elements in a ring, 6
dimension, 63
direct product of rings, 5
division, 49
    of polynomials, 35
Division Algorithm, 35
division algorithm, 60
division ring, 11
divisor function, 60
domain, 11

Eisenstein Criterion, 43
endomorphism, 25
epimorphism, 25
equality of polynomials, 33
euclidean domain, 36, 60
euclidean function, 36, 60
euclidean valuation, 60
evaluation, 37
Evaluation Theorem, 36
extending a ring homomorphism, 73
extension, 64

$\mathcal{F}(X, \mathbb{R})$, 5
Factor Theorem, 37
factorization, 49
    proper, 41
    trivial, 49
field, 11
field extension, 64
field generated by elements, 65
field of fractions, 15
field of quotients, 15
finite dimensional, 63
finite extension, 64
First Isomorphism Theorem, 26
Frobenius homomorphism, 24
Fundamental Theorem of Algebra, 75

Galois field, 78
Gauss' Lemma, 41, 55
gaussian integers, 8

are a euclidean domain, 60
gcd, 52
general ring, 4
general ring homomorphism, 24
GF(p^n), 78
greatest common divisor, 52
of polynomials, 44
group of units, 6

homomorphism
of general rings, 24
of rings, 23

ideal, 16
zero, 17
idempotent, 7
identity element, 4
image, 25
indeterminate, 32
infinite dimensional, 63
inner automorphism, 25
integral domain, 11
irreducible, 50
polynomial, 39
isomorphic rings, 8, 25
isomorphism, 8, 25

kernel, 25
Kronecker's Theorem, 71

lcm, 52
leading coefficient, 34
least common multiple, 52
left ideal, 23
Lie algebra, 5
linear combination, 63
linearly dependent, 63
linearly independent, 63

maximal ideal, 19
minimal polynomial, 65
Modular Irreducibility Test, 42
monic polynomial, 34
monomorpihsm, 25
Multiplication Theorem, 67
multiplicative inverse, 5
multiplicity of a root, 38

$\mathbb{N}$, 3
nilpotent, 7
noetherian ring, 57
nonassociative ring, 5

opposite ring, 9

PID, 58
polynomial, 33
function, 33
polynomial ring, 5
prime element, 50
prime ideal, 18
primitive element, 79
Primitive Element Theorem, 69
primitive polynomial, 55
primitive root of unity, 79
principal ideal, 17
principal ideal domain, 58
proper factorization, 41
proper ideal, 16

quaternions, 13

$R^\times$, 5
rational function, 15
Rational Roots Theorem, 38
reducible polynomial, 39
reduction modulo p, 42
Remainder Theorem, 37
ring, 4
automorphism, 25
endomorphism, 25
epimorphism, 25
homomorphism, 23
isomorphism, 25
monomorphism, 25
of polynomials, 33
ring isomorphism, 8
root of a polynomial, 38

scalar multiplication, 62
Second Isomorphism Theorem, 30
simple ring, 17
skew field, 11

span, 63
split, 71
splitting field, 71
subfield, 64
subring, 7
Subring Test, 7
subspace, 62
subtraction in a ring, 6

Third Isomorphism Theorem, 30
transcendental element, 64

UFD, 51
unique factorization domain, 51
unit, 5
unity, 4

vector addition, 62
vector space, 62

$\mathbb{Z}_n$, 5
zero divisor, 11
zero ideal, 17
zero polynomial, 34
zero ring, 5

Bibliography

[Jud12] Thomas W. Judson. *Abstract algebra*. 2012 annual edition, 2012. Available at http://abstract.ups.edu/index.html.

[Nic12] W. Keith Nicholson. *Introduction to abstract algebra*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, fourth edition, 2012.

[Roy] Damien Roy. MAT 3143 – Linear Algebra II, Lecture notes (translated by A. Savage). Available at http://alistairsavage.ca/mat3141/notes/MAT3141-LinearAlgebraII.pdf.

[Sav] Alistair Savage. MAT 2141 – Linear Algebra I, Lecture notes. Available at http://alistairsavage.ca/mat2141/notes/MAT2141-LinearAlgebraI.pdf.