# Introduction to Higher Mathematics
# Unit #5: Abstract Algebra

Joseph H. Silverman


Version Date: February 14, 2018

# Contents



# Introduction to Abstract Algebra

The overall theme of this unit is algebraic structures in mathematics. Roughly speaking, an algebraic structure consists of a set of objects and a set of rules that let you manipulate the objects. Here are some examples that will be familiar to you:

*Example* 0.1. The objects are the numbers $1, 2, 3, \ldots$. You already know two ways to manipulate these objects, namely addition $a + b$ and multiplication $a \cdot b$

*Example* 0.2. The objects are triangles in the plane, and we can be manipulate them by translation and by rotation and by reflection.

*Example* 0.3. The objects are functions $f : \mathbb{R} \to \mathbb{R}$, and we can manipulate them by addtion $f(x)+g(x)$, by multiplication $f(x)\cdot g(x)$, and also by composition $f\big(g(x)\big)$.

Our primary goal is to take examples of this sort and generalize them, or in mathematical terminology, *axiomatize them*. To do this, we strip away everything that is not essential and reduce down to an abstract description consisting of a set with operations (such as addition and multiplication) that are required to satisfy certain rules, also known as *axioms*.[1]

In this unit's four chapters we will study four different types of objects and their associated rules:

**Chapter 1** Groups

**Chapter 2** Rings

**Chapter 3** Vector Spaces

**Chapter 4** Fields

Although groups, rings, fields, and vector spaces are not the same, the four chapters share common themes. In each chapter we use axioms to describe objects having an algebraic structure, and we study maps between these objects that preserve the structure. Roughly speaking, each chapter is organized as follows, although the order may vary slightly from chapter to chapter:

[1] Axioms are also sometimes called "laws". For example, you're probably familiar with the "commutative law" for addition, which says that $a + b = b + a$. But this isn't really a law, debated and approved by a legislative body! Instead, addition is a rule that explains how to combine two numbers and get a third number, and the "commmutative law" is a property that we impose on the "addition rule".

- Give an example of a certain type of algebraic structure
- Give a formal definition, using axioms, of the algebraic structure.
- Proof of a basic property directly from the definitions.
- Discuss what a map must do to "preserve the algebraic structure."
- Give additional examples.
- Investigate and prove a deeper property.
- As time permits, discuss sub-objects and quotient objects.

**A Note on the Role of Definitions, Axioms, and Proofs in Higher Mathematics:** Since at least the time of Euclid, circa 300 BC,, the ultimate test of mathematical rigor lies in the construction of proofs of mathematical statements. Without getting into deep matters of philosophy, a proof is a sequence of steps that starts with a known fact and ends with the desired final statement. Each step is required to follow logically from a combination of one or more of the following:[2]

- Steps in the proof that have already been completed.
- Statements that have previously been proven.
- Axioms, which are statements that are assumed to be true.
- Definitions, which describe the properties possessed by objects.

**Mini-Remark 1**. **Further Remarks about Definitions**: There is nothing magical about a definition, and in principle there are no restrictions on what may be defined. For example, I might define a *zyglx* to be a purple pig with wings. I could then potentially that definition to prove that zyglxes are able to fly, since they have wings. Is ths useful? No, since as far as I am aware, there is nothing in the real world to which I could apply the "Zyglx Theory." So although definitions are, to some extent, arbitrary, the usefulness of a definition is determined by its applicability to a range of (realistic) situations. We will see many examples of such definitions, including especially the definitions of *groups*, *rings*, *fields*, and *vector spaces*. The primary goal of theoretical mathematics, and likewise of this course, is to formulate and prove interesting mathematical statements, which in our case means statements about groups, rings, etc. And the only way to get started is to have a first understanding of the definitions of the objects that we want to study. This is why understanding and applying definitions is a crucial part of modern mathematics, and why you should spend time studying definitions when they're introduced and using definitions when you're trying to prove things.

**Mini-Remark 2**. **Further Remarks about Axioms**: In Greek mathematics, axioms were viewed as statement that are so self-evident, they must be true. The modern viewpoint is that in principle, one is free to use any set of axioms that one wants. However, not all axiom systems are created equal. The best and most interesting axiom systems are those that start with very few axioms and allow one to prove a very large number of useful and interesting and beautiful statements. The axioms for geometry that appear in Euclid's work are an example. But one of thos axioms, the so-called *parallel postulate*, led to a revolution in mathematics.

[2] Axioms and definitions are discussed further in the Mini-Remarks in this section.

This axiom says that given a line $L$ in the plane and a point $P$ not lying on $L$, there is exactly one line $L'$ that contains $P$ and does not intersect $L$. Seems reasonable, but maybe not entirely self-evident, so mathematicians spent centuries trying to prove that it follows from Euclid's other axioms. All failed. Then, in the 19th century, it was discovered that if one changes the parallel postulate by replacing the words "exactly one line" with "infinitely many lines," or with "no lines," then one gets geometries that are as valid as Euclid's. These so-called non-Euclidean geometries have many uses in modern mathematics and physics, and indeed it is likely that the universe in which we live is actually a "no lines" space!

**Important Notes for Math 760: Unit 5:** These notes contain more material than we will have time to cover in class. The "Mini-Notes" are there for you to read as an aid to understanding. The extra "Optional" sections are there as an invitation for you to explore additional aspects of abstract algebra. And before you ask, the answer is no, the material in the mini-notes and the optional sections will not be on the exam!

| | | | |
|---|---|---|---|
| Ch. 1 | Th | 03/01/18 | Groups |
| Ch. 1 | Tu | 03/06/18 | Groups |
| Ch. 2 | Th | 03/08/18 | Rings |
| Ch. 2 | Tu | 03/13/18 | Rings |
| Ch. 3 | Th | 03/15/18 | Vector Spaces |
| Ch. 3 | Tu | 03/20/18 | Vector Spaces |
| Ch. 4 | Th | 03/22/18 | Fields |
| Ch. 4 | Tu | 04/03/18 | Fields |
| — | Th | 04/05/18 | Unit 5 Exam |

**Schedule for Math 760 Unit 5: Abstract Algebra (Spring 2018)**

# Chapter 1

# Groups

## 1.1 Introduction to Groups

We start with a simple question. What are the different ways that we can rearrange the list of numbers $1, 2, 3, 4$? For example, we could send $1$ to $2$, send $2$ to $3$, send $3$ to $4$, and send $4$ to $1$. This is conveniently illustrated by the picture

$$1 \to 2, \qquad 2 \to 3, \qquad 3 \to 4, \qquad 4 \to 1. \tag{1.1}$$

Another way to rearrange them would be swap $1$ and $2$ and swap $3$ and $4$, illustrated by

$$1 \to 2, \qquad 2 \to 1, \qquad 3 \to 4, \qquad 4 \to 3. \tag{1.2}$$

The mathematical word for such a rearrangement is a *permutation*, so we have just described two different permutations of the set $\{1, 2, 3, 4\}$. A permutation of the set $\{1, 2, 3, 4\}$ is described by a rule that assigns to each element of the set $\{1, 2, 3, 4\}$ an element of the same set $\{1, 2, 3, 4\}$, with the added proviso that we don't use any element twice.

**Mini-Remark 3**. How many permutations are there of the set $\{1, 2, 3, 4\}$? We can assign $1$ to any of $1, 2, 3, 4$, so there are $4$ choices for $1$, then we can assign $2$ to any of the remaining $3$ values, after which we can assign $3$ to either of the remaining $2$ values, and finally we have to assign $4$ to the last remaining value. Thus there are $4 \cdot 3 \cdot 2 \cdot 1$, i.e., $24$, different permuations of $\{1, 2, 3, 4\}$. More generally, Exercise #1.1 asks you to compute how many permutations there are of the set $\{1, 2, \ldots, n\}$.

**Mini-Remark 4**. If we have two permuations of $\{1, 2, 3, 4\}$, we can "compose" them by doing first one, and then the other. So for example, if we let $\sigma$ be the permuation described in (1.1) and we let $\tau$ be the permuation described in (1.2), then $\sigma \circ \tau$ is the permuation having the following effect on the set $\{1, 2, 3, 4\}$:

$$1 \xrightarrow{\tau} 2 \xrightarrow{\sigma} 3, \qquad 2 \xrightarrow{\tau} 1 \xrightarrow{\sigma} 2, \qquad 3 \xrightarrow{\tau} 4 \xrightarrow{\sigma} 1, \qquad 4 \xrightarrow{\tau} 3 \xrightarrow{\sigma} 4.$$

An interesting observation is that if we compose $\sigma$ and $\tau$ in the other order, we get a different permutation. Thus

$$1 \xrightarrow{\sigma} 2 \xrightarrow{\tau} 1, \qquad 2 \xrightarrow{\sigma} 3 \xrightarrow{\tau} 4, \qquad 3 \xrightarrow{\sigma} 4 \xrightarrow{\tau} 3, \qquad 4 \xrightarrow{\sigma} 1 \xrightarrow{\tau} 2.$$

In general, a *permutation* of a set $X$ is a rule that "mixes up" the elements of $X$. Our first goal is to give a precise mathematical meaning to the notion of "a rule that mixes up a set."

We already have a mathematical name for "rules" that tell us how to take elements of a set $X$ and assign them to elements of a set $Y$. These rules are called *functions with domain $X$ and range $Y$*. So a permutation on a set $X$ is a function whose domain and range are both the same set $X$, but with some added conditions to ensure that every image element comes from exactly one domain element.

**Definition.** A *permutation* of a set $X$ is a bijective function[1] whose domain and range are $X$. In other words, a permuation of $X$ is a function

$$\pi : X \longrightarrow X$$

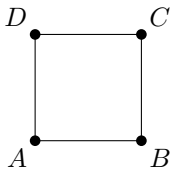
having the following property: For every element $x \in X$ there is exactly one element $x' \in X$ satisfying $\pi(x') = x$. This allows us to define the *inverse of* $\pi$ to be the permutation

$$\pi^{-1} : X \longrightarrow X$$

determined by the rule that $\pi^{-1}(x)$ is equal to the unique element $x' \in X$ such that $\pi(x') = x$. Finally, we define the *identity permuation of* $X$ to be the identity map,

$$e : X \longrightarrow X, \quad e(x) = x \quad \text{for all } x \in X.$$

*Example* 1.1 (**Symmetries of a Square**). Next we consider a rigid square whose vertices are labeled $A, B, C, D$ as in the following picture:

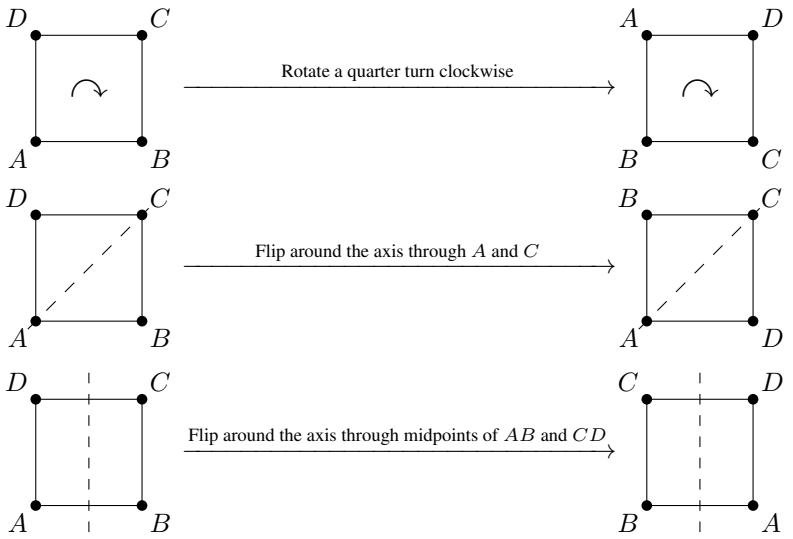



Suppose that we pick up the square and rotate or flip it[2] in some way, then put it back down. Here are three examples:

[1] Recall from Set Theory that a function $\phi : S \to T$ is *injective* if for every $t \in T$ there is at most one $s \in S$ satisfying $\phi(s) = t$, and that $\phi$ is *surjective* if for every $t \in T$ there is at least one $s \in S$ satisfying $\phi(s) = t$. A function that is both injective and surjective is said to be *bijective*. We also recall that another name for injective functions is *one-to-one*, and another name for surjective functions is *onto*.

[2] Many mathematicians prefer to say that we *reflect* the square, instead of using the more action-packed word *flip*.




The rotation and flips described in these pictures are permutations of the set $\{A, B, C, D\}$. Explicitly, if we call them Rot, $\text{Flip}_1$, and $\text{Flip}_2$,

| Rot | $\text{Flip}_1$ | $\text{Flip}_2$ |
|---|---|---|
| $A \to D$ | $A \to A$ | $A \to B$ |
| $B \to A$ | $B \to D$ | $B \to A$ |
| $C \to B$ | $C \to C$ | $C \to D$ |
| $D \to C$ | $D \to B$ | $D \to C$ |

But not all permuations of $\{A, B, C, D\}$ are permitted, since we're not allowed to bend or break the sides of the square. For example, there is no way to pick up the square and put it back down so that

$$A \to A, \quad B \to B, \quad C \to D, \quad D \to C,$$

without bending or breaking its sides. So the collection of symmetries of the square includes only some of the permutations of the set $\{A, B, C, D\}$. We leave it to you to check that among the $24$ permutations of $\{A, B, C, D\}$, there are exactly $8$ that are valid symmetries of the square.

## 1.2 Abstract Groups

**Definition.** A *group* consists of a set $G$ together with a composition law

$$\begin{aligned} G \times G &\longrightarrow G, \\ (g_1, g_2) &\longmapsto g_1 \cdot g_2, \end{aligned}$$

satisfying the following axioms:

(a) (*Identity Axiom*) There is an element $e \in G$ such that

$$e \cdot g = g \cdot e = g \quad \text{for all } g \in G.$$

The element $e$ is called the *identity element of* $G$.

(b) (*Inverse Axiom*) For every $g \in G$ there is an element $h \in G$ such that

$$g \cdot h = h \cdot g = e.$$

The element $h$ is denoted $g^{-1}$ and is called the *inverse of* $g$.

(c) (*Associative Law*) For all $g_1, g_2, g_3 \in G$, the *associative law* holds, that is,

$$g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3.$$

(d) (*Commutative Law*) If in addition it is true that

$$g_1 \cdot g_2 = g_2 \cdot g_1 \quad \text{for all } g_1, g_2 \in G,$$

then $G$ is said to be *commutative* or *abelian*.[3]

*Remark* 1.2. The key attribute of a group is that it includes a "rule" or "operation" or "law" (satisfying three axioms) for combining two elements of the group to create a third element. Depending on the context, you may find the group law being called "addition" or "multiplication" or "composition," but assigning a name to the group law is simply a linguistic convenience,[4] and if you prefer, you may make up some other name, say "xzyglpqz," for the group law in your favorite group.[5]

There are various basic properties of groups that follow from three group axioms. We list some of them here, prove one, and leave the others as exercises.

**Proposition 1.3** (Basic Propreties of Groups)**.** *Let $G$ be a group.*

(a) *$G$ has exactly one identity element.*

(b) *Each element of $G$ has exactly one inverse.*

(c) *Let $g, h \in G$. Then $(g \cdot h)^{-1} = h^{-1} \cdot g^{-1}$.*

(d) *Let $g \in G$. Then $(g^{-1})^{-1} = g$.*

*Proof.* We prove one part, and leave the others as exercises

(b) Let $g \in G$, and suppose that $h_1, h_2 \in G$ are both inverses for $g$. Then

$$\begin{aligned} h_1 &= h_1 \cdot e && \text{since } e \text{ is the identity element,} \\ &= h_1 \cdot (g \cdot h_2) && \text{since } h_2 \text{ is an inverse of } g, \\ &= (h_1 \cdot g) \cdot h_2 && \text{associative law,} \\ &= e \cdot h_2 && \text{since } h_1 \text{ is an inverse of } g, \\ &= h_2 && \text{since } e \text{ is the identity element.} \end{aligned}$$

This completess the proof. □

[3] The word "abelian" comes from Niels Henrik Abel (1802–1829), a Norwegian mathematician famous for many discoveries, including a proof that it is impossible to solve a general quintic equation using radicals. The Abel prize for mathematics, modeled after the Nobel prizes and awarded annually since 2002, is named in his honor.

[4] Or, as Juliet actually said to Romeo, "a group law by any other name would smell as sweet."

[5] Although in practice, people generally don't call a group law "addition" unless it is commutative.

**Definition.** The *order* of a group $G$, which we denote by $\#G$, is the cardinality of the set of elements of $G$, e.g., if $G$ is finite, it is simply the number of elements in $G$.[6]

**Definition.** Let $G$ be a group, and let $g \in G$. The *order* of the element $g$ is the smallest integer $n \ge 1$ with the property that $g^n = e$. If no such $n$ exists, then we say that $g$ has infinite order.

**Proposition 1.4.** *Let $G$ be a group, let $g \in G$, and let $n \ge 1$ be an integer such that $g^n = e$. Then the order of $g$ divides $n$.*

*Proof.* Let $m$ be the order of $g$, so $m$ is the smallest positive integer satisfying $g^m = e$. Dividing $n$ by $m$ yields a quotient and remainder

$$n = mq + r \quad \text{with } 0 \le r < m.$$

We use this equality and the fact that $g^n = g^m = e$ to compute

$$e = g^n = g^{mq+r} = (g^m)^q \cdot g^r = e^q \cdot g^r = g^r.$$

Thus $g^r = e$ and $0 \le r < m$. But by definition, the smallest postive power of $g$ that equals $e$ is $g^m$. Therefore $r = 0$ and $n = mq$, which shows that $m$, which is the order of $g$, divides $n$. □

## 1.3 Interesting Examples of Groups

In Section 1.1 we saw a couple of groups. It's time to expand our repertoire.

*Example* 1.5 (**Group of Integers and Integers Modulo $m$**). The set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is a group if we use addition as the group law. It is an example of an *infinite group*, that is, a group having infinitely many elements. On the other hand, if we try to use multiplication as the group law, then $\mathbb{Z}$ is not a group. Do you see why not? The set $\mathbb{Z}/m\mathbb{Z}$ of integers modulo $m$ forms a group with addition as the group law. It is a finite group of order $m$.

*Example* 1.6 (**Additive Group of Real, Rational, and Complex Numbers**). The set of real numbers $\mathbb{R}$ with addition is an infinite group, as is the set of rational numbers $\mathbb{Q}$ and the set of complex numbers $\mathbb{C}$.

*Example* 1.7 (**Multiplicative Group of Real Numbers**). The set of *non-zero* real numbers forms a group with mulitplication as the group law. The set of *positive* real numbers also forms a group using multiplication.

**Definition.** A group $G$ is a *cyclic group* if there is an element $g \in G$ with the property that

$$G = \{\ldots, g^{-3}, g^{-2}, g^{-1}, e, g, g^2, g^3, \ldots\}.$$

(Here $g^{-k}$ is shorthand for the $k$-fold product $g^{-1} \cdot g^{-1} \cdots g^{-1}$.) The element $g$ a *generator of $G$*, but note that cyclic groups may have more than one generator.

[6] Other common notations for the order of a group, or more generally, for the cardinality of a set, include $o(G)$ and $|G|$.

*Example* 1.8 (**Cyclic Groups**). We have already seen some examples of cyclic groups. The group of integers $(\mathbb{Z},+)$ is an infinite cyclic group whose generators are $1$ and $-1$. The group $(\mathbb{Z}/m\mathbb{Z},+)$ of integers modulo $m$ is a finite cyclic group of order $m$ whose generators are precisely the elements $a \bmod m$ such that $\gcd(a,m)=1$; see Exercise 1.6.

In general, for $n \ge 1$, we create an abstract cyclic group order $n$, which we denote $\mathcal{C}_n$, by taking the set

$$\mathcal{C}_n = \{g_0, g_1, g_2, \ldots, g_{n-1}\}$$

and using the composition rule

$$g_i \cdot g_j = \begin{cases} g_{i+j} & \text{if } i+j < n, \\ g_{i+j-n} & \text{if } i+j \ge n. \end{cases}$$

The identity element of $\mathcal{C}_n$ is the element $g_0$, and the inverse of an element $g_i$ is the element $g_{n-i}$, except that the inveres of $g_0$ is $g_0$. We note that $\mathcal{C}_n$ is an abelian group, since $g_{i+j} = g_{j+i}$.

*Example* 1.9 (**Permutation Groups**). Let $X$ be a set. We recall that a *permutation of* $X$ is a bijective function

$$\pi : X \longrightarrow X.$$

The *symmetry group of* $X$ is the collection of all of permuations of $X$, with the group law being composition of permuations. It is denoted $\mathcal{S}_X$. In the special case that $X = \{1,2,\ldots,n\}$ consists of the integers from $1$ to $n$, we write $\mathcal{S}_n$. We saw in Section 1.1 that the group $\mathcal{S}_4$ has order $24$, and that it is nonabelian, since we described elements $\sigma, \tau \in \mathcal{S}_4$ with the property that $\sigma\tau \ne \tau\sigma$. Exercise 1.1 asks you to compute the order of the group $\mathcal{S}_n$.

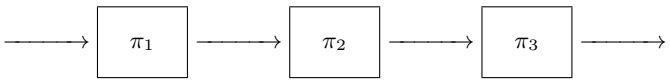
**Mini-Remark 5**. The identity element for the group $\mathcal{S}_X$ is the identity map $\pi_0(x) = x$, while the inverse of an element $\pi \in \mathcal{S}_X$ is the inverse map $\pi^{-1}$, which exists because $\pi$ is bijective. But why does composition of permutations satisfy the associative law? The composition of two permutations $\pi_1, \pi_2 \in \mathcal{S}_X$ is defined by the formula

$$(\pi_1 \circ \pi_2)(x) = \pi_1\big(\pi_2(x)\big).$$

So we can formally compute

$$\begin{aligned} \big((\pi_1 \circ \pi_2) \circ \pi_3\big)(x) &= (\pi_1 \circ \pi_2)\big(\pi_3(x)\big) = \pi_1\big(\pi_2\big(\pi_3(x)\big)\big), \\ \big(\pi_1 \circ (\pi_2 \circ \pi_3)\big)(x) &= \pi_1\big((\pi_2 \circ \pi_3)(x)\big) = \pi_1\big(\pi_2\big(\pi_3(x)\big)\big). \end{aligned}$$

Alternatively, you may prefer to view a permutation as a function that takes in a value and spits out a value. Then the composition $\pi_1 \circ \pi_2 \circ \pi_3$, regardless of how you group the functions, is illustrated by the following picture:



*Example* 1.10 (**Matrix Groups**). Many of you will have learned how to multiply 2-by-2 matrices,

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix}. \tag{1.3}$$

The following set of 2-by-2 matrices ,

$$\mathrm{GL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R},\ ad - bc \neq 0. \right\},$$

is a group using matrix multiplication as the group operation; see Exercise 1.11.

*Example* 1.11 (**Dihedral Groups**). Let $P$ be a regular $n$-gon, with vertices labeled $1, 2, \ldots, n$. Figure 1.1 illustrates the case $n = 6$. Just as we did with the square in Section 1.1, we can permute the vertices $\{1, 2, \ldots, n\}$ of $P$ by lifting up the $n$-gon, rotating and/or flipping it, and then putting it back down where it originally was. The group of all such permutations of the $n$-gon is called the *$n$'th dihedral group* and is denoted $\mathcal{D}_n$. There are exactly $n$ rotations (if we treat no movement as the trivial rotation) and exactly $n$ flips, so $\mathcal{D}_n$ is a group of order $2n$; see Exercise 1.8.

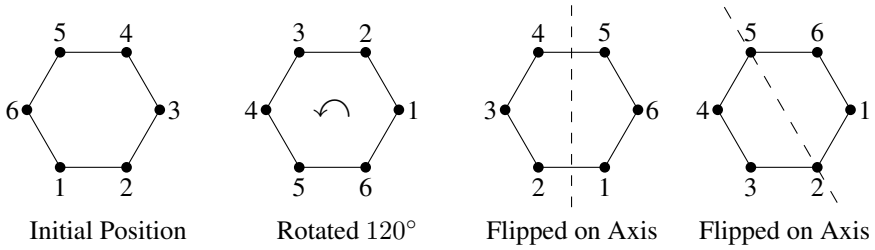



Figure 1.1: A rotation and two flips of a regular $n$-gon with $n = 6$

*Example* 1.12 (**Quaternion Group**). The *quaternion group* $\mathcal{Q}$ is a non-commutative group with eight elements,

$$\mathcal{Q} = \{\pm 1, \pm \boldsymbol{i}, \pm \boldsymbol{j}, \pm \boldsymbol{k}\}.$$

The pluses and minuses work as usual with $(-1)^2 = 1$. The rules for multiplying the quantities $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are determined by the formulas

$$\boldsymbol{i} \cdot \boldsymbol{i} = -1, \quad \boldsymbol{j} \cdot \boldsymbol{j} = -1, \quad \boldsymbol{k} \cdot \boldsymbol{k} = -1, \quad \boldsymbol{i} \cdot \boldsymbol{j} \cdot \boldsymbol{k} = -1.$$

From these rules one can prove, for example, that $\boldsymbol{j} \cdot \boldsymbol{i} = -\boldsymbol{i} \cdot \boldsymbol{j}$. See Exercise 1.13.

## 1.4 Group Homomorphisms

Suppose that $G$ and $G'$ are groups, and suppose that $\phi$ is a function

$$\phi : G \longrightarrow G'$$

from the elements of $G$ to the elements of $G'$. There are many such functions, but since $G$ and $G'$ are groups, we want to concentrate on functions $\phi$ that respect the "*group-i-ness*" of $G$ and $G'$.

**Question**: What makes a group a group?
**Answer**: Groups have a composition law and identity elements and inverses.

So we should require that the function $\phi : G \to G'$ have the following properties:

- $\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2)$ for all $g_1, g_2 \in G$.
- $\phi(e) = e'$, where $e$ and $e'$ are, respectively, the identtity elements on $G$ and $G'$.
- $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G$.

**Important Observation**: Did you notice that the two "dots" in the formula

$$\begin{array}{c} \phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) \\ \uparrow \qquad\qquad \uparrow \\ \boxed{G \text{ group law}} \; \boxed{G' \text{ group law}} \end{array} \tag{1.4}$$

are not the same dot?!! That's because the dot in $\phi(g_1 \cdot g_2)$ means that $g_1$ and $g_2$ are being combined using the composition law on the group $G$, while the dot in $\phi(g_1) \cdot \phi(g_2)$ means that $\phi(g_1)$ and $\phi(g_2)$ are being combined using the composition law on the group $G'$. So the formula (1.4) says that $\phi$ cleverly intertwines the group laws on $G$ and $G'$. It turns out that this is enough to force the other two properties to be true, which leads to the following fundamenatl definition.

**Definition.** Let $G$ and $G'$ be groups. A *homomorphism from $G$ to $G'$* is a function $\phi : G \to G'$ satisfying

$$\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) \quad \text{for all} \quad g_1, g_2 \in G.$$

We'll now check that this is enough to get the other two properties that we want.

**Proposition 1.13.** *Let $\phi : G \to G'$ be a homomorphism of groups.*
(a) *Let $e \in G$ be the identity element of $G$. Then $\phi(e)$ is the identity element of $G'$.*
(b) *Let $g \in G$. Then $\phi(g^{-1})$ is the inverse of $\phi(g)$.*

*Proof.* (a) We use the fact that $e \cdot e = e$ and the fact that $\phi$ is a homomorphism to compute

$$\phi(e) = \phi(e \cdot e) = \phi(e) \cdot \phi(e). \tag{1.5}$$

We now apply $\phi(e)^{-1}$ to both sides to obtain

$$\begin{aligned}
e' &= \phi(e)\cdot\phi(e)^{-1} && \text{since } \phi(e)^{-1} \text{ is the inverse of } \phi(e),\\
&= \big(\phi(e)\cdot\phi(e)\big)\cdot\phi(e)^{-1} && \text{using (1.5)},\\
&= \phi(e)\cdot\big(\phi(e)\cdot\phi(e)^{-1}\big) && \text{associative law},\\
&= \phi(e)\cdot e' && \text{since } \phi(e)^{-1} \text{ is the inverse of } \phi(e),\\
&= \phi(e) && \text{since } e' \text{ is the identity element of } G'.
\end{aligned}$$

(b) We need to show that $\phi(g^{-1})$ has the property to be the inverse of $\phi(g)$. So we compute

$$\begin{aligned}
\phi(g^{-1})\cdot\phi(g) &= \phi(g^{-1}\cdot g) && \text{since } \phi \text{ is a homomorphism},\\
&= \phi(e) && \text{since } g^{-1} \text{ is the inverse of } g,\\
&= e' && \text{from what we proved in (a)}.
\end{aligned}$$

The proof that $\phi(g)\cdot\phi(g^{-1}) = e'$ is similar, which completes the proof that $\phi(g^{-1})$ is the inverse of $\phi(g)$. □

*Example* 1.14. Recall from Example 1.11 the the dihedral group $\mathcal{D}_n$ is the collections of rotations and flips of an $n$-sided polygon. It is a group with $2n$ elements, half of which are rotations. We can define a homomorphism from $\mathcal{D}_n$ to the two-element group $\{\pm1\}$ by the rule

$$\phi : \mathcal{D}_n \longrightarrow \{\pm1\}, \quad \phi(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is a rotation,}\\ -1 & \text{if } \sigma \text{ is a flip.}\end{cases}$$

In order to check that $\phi$ is a homomorphism, one needs to check

$$\begin{aligned}
\text{Rotation}\circ\text{Rotation} &= \text{Rotation}, \quad \text{Rotation}\circ\text{Flip} = \text{Flip},\\
\text{Flip}\circ\text{Rotation} &= \text{Flip}, \quad \text{Flip}\circ\text{Flip} = \text{Rotation},
\end{aligned} \tag{1.6}$$

a task which we leave to you (Exercise 1.16).

*Example* 1.15. For any integers $n \ge m \ge 1$, there is an injective homomorphism

$$\phi : \mathcal{S}_m \longrightarrow \mathcal{S}_n \quad \text{given by the rule} \quad \phi(\pi)(k) = \begin{cases} \pi(k) & \text{if } 1 \le k \le m,\\ k & \text{if } m < k \le n.\end{cases}$$

In other words, if $\pi$ is a permutation of $\{1, 2, \ldots, m\}$, then we view $\pi$ as a permutation of $\{1, 2, \ldots, n\}$ by letting it permute $1, 2, \ldots, m$ and having it fix $m+1, m+2, \ldots, n$.

*Example* 1.16. You already know a very important group homomorphism, namely the logarithm function (to any base), which gives a homomorphism

$$\log : \{\text{positive real numbers with } \times\} \longrightarrow \{\text{real numbers with } +\}.$$

The logarithm function is a homomorphism because it converts multiplication to addition,[7]

$$\log(ab) = \log(a) + \log(b).$$

**Definition.** Two groups $G_1$ and $G_2$ are said to be *isomorphic* if there is a bijective homomorphism

$$\phi : G_1 \longrightarrow G_2.$$

The map $\phi$ is called an *isomorphism* from $G_1$ to $G_2$. Isomorphic groups are really the same group, but their elements have been given different names.[8]

*Example* 1.17. The groups $\mathcal{C}_2$ and $\mathcal{S}_2$ are isomorphic, as are the group $\mathcal{D}_3$ and $\mathcal{S}_3$. If $p$ is a prime number, then every group with exactly $p$ elements is isomorphic to $\mathcal{C}_p$. The logarithm map (Example 1.16) is an isomorphism from the group of positive real numbers with multiplication to the group of real numbers with addition. You will get to prove these assertions in the exercises.

## 1.5 Subgroups, Cosets, and Lagrange's Theorem

A guiding principle in mathematics when attempting to analyze a complicated object may be summarized by the following three steps:

**Step 1 (Deconstruction):** Break your object into smaller and simpler pieces.

**Step 2 (Analysis):** Analyze the smaller, simpler pieces.

**Step 3 (Reconstruction):** Fit the pieces back together.

For a group $G$, a natural way to form a "smaller and simpler piece" is by taking subsets $H$ that are themselves groups. This prompts the following definition.

**Definition.** Let $G$ be a group. A *subgroup of* $G$ is a subset $H \subset G$ that is itself a group using $G$'s group law. Explicitly, $H$ needs to satisfy:[9]

(i) For every $h_1, h_2 \in H$, the product $h_1 \cdot h_2$ is in $H$.

(ii) The identity element $e$ is in $H$.

(iii) For every $h \in H$, the inverse $h^{-1}$ is in $H$.

Note that since $H$ uses the same group law as $G$, the elements of $H$ automatically satisfy the associative law, so we do not need to add that as a requirement. If $H$ is finite, we define the *order of* $H$ to be the number of elements in $H$.

[7] Logarithms were discovered by John Napier (1550–1617). Back in "ancient days" when computations were done by hand, tables of logarithm were used extensively to speed numerical calculations in astronomy, engineering, and physics.

[8] The group you are about to study is true. Only the names of its elements have been changed to protect the innocent. . . . Cue *Dragnet* theme.

[9] In order to prove that a subset $H$ is a subgroup, it suffices to check that $H \neq \emptyset$ and that for every $h_1, h_2 \in H$, the element $h_1 h_2^{-1}$ is in $H$. See Exercise 1.20.

*Example* 1.18. Every group $G$ has at least two subgroups, namely the *trivial subgroup* $\{e\}$ consisting of only the identity element, and the entire group $G$. Most groups other subgroups; see Exercise 1.22.

*Example* 1.19. Let $G$ be a group, and let $g \in G$. Then the *cyclic subgroup of $G$ generatedy by $g$*, denoted $\langle g \rangle$, is the set

$$\langle g \rangle = \{\ldots, g^{-3}, g^{-2}, g^{-1}, e, g, g^2, g^3, \ldots\}.$$

If $g$ has order $n$, then

$$\langle g \rangle = \{e, g, g^2, g^3, \ldots, g^{n-1}\}$$

is isomorphic to the cyclic group $\mathcal{C}_n$, while if $g$ has infinite order, then $\langle g \rangle$ is isomorphic to $\mathbb{Z}$.

Every group homomorphism has an associated subgroup, called its kernel, which can be used to give a convenient criterion for checking if the homomorphism is injective.

**Definition.** Let $\phi : G \to G'$ be a group homomorphism. The *kernel of $\phi$* is the set of element of $G$ that are sent to the identity element of $G'$,

$$\ker(\phi) = \big\{g \in G : \phi(g) = e'\big\}.$$

**Proposition 1.20.** *Let $\phi : G \to G'$ be a group homomorphism.*
(a) $\ker(\phi)$ *is a subgroup of $G$.*
(b) $\phi$ *is injective if and only if* $\ker(\phi) = \{e\}$.

*Proof.* (a) Proposition 1.13(a) says $\phi(e) = e'$, so $e \in \ker(\phi)$. Next let $g_1, g_2 \in \ker(\phi)$. Then the homomorphism property of $\phi$ gives $\phi(g_1 \cdot g_2) = \phi(g_1) \cdot \phi(g_2) = e' \cdot e'$, so $g_1 \cdot g_2 \in \ker(\phi)$. Finally, for $g \in \ker(\phi)$, Proposition 1.13(b) says $\phi(g^{-1}) = \phi(g)^{-1}) = e'^{-}1 = e'$, so $g^{-1} \in \ker(\phi)$. This completes the proof that $\ker(\phi)$ is a subgroup of $G$.
(b) We know from Proposition 1.13(a) that $e \in \ker(\phi)$. If $\phi$ injective, then there is at most one element $g \in G$ satisfying $\phi(g) = e'$, so we must have $\ker(\phi) = \{e\}$

Next we suppose that $\ker(\phi) = \{e\}$. Let $g_1, g_2 \in G$ satisfy $\phi(g_1) = \phi(g_2)$. Again using the homomorphism property and Proposition 1.13(b), we find that

$$\phi(g_1 \cdot g_2^{-1}) = \phi(g_1) \cdot \phi(g_2^{-1}) = \phi(g_1) \cdot \phi(g_2)^{-1} = e' \cdot e'^{-1} = e'.$$

Thus $g_1 \cdot g_2^{-1} \in \ker(\phi) = \{e\}$, so $g_1 = g_2$. This proves that $\phi$ is injective. □

*Example* 1.21. Let $d \in \mathbb{Z}$, then we can form a subgroup of $\mathbb{Z}$ using the multiples of $d$,

$$d\mathbb{Z} = \{dn : n \in \mathbb{Z}\}.$$

*Example* 1.22. The set of rotations in the dihedral group $\mathcal{D}_n$ is a subgroup of $\mathcal{D}_n$.

*Example* 1.23. The set of elements of the symmetric group $\mathcal{S}_n$ that fix $n$ form a subgroup of $\mathcal{S}_n$. This subgroup is naturaly isomorphic to $\mathcal{S}_{n-1}$, since its elements are the permutations of $1, 2, \ldots, n-1$.

We are going to use a subgroup $H$ of $G$ to break $G$ into pieces that are called cosets of $H$.

**Definition.** Let $G$ be a group, and let $H \subset G$ be a subgroup of $G$. For each $g \in G$, the (left) *coset of $H$ attached to $g$* is the set

$$gH = \{gh : h \in H\}.$$

In other words, $gH$ is the set that we get when we multiply $g$ by every element of $H$.

We now prove several properties of cosets which help explain why they're important.

**Proposition 1.24.** *Let $G$ be a finite group, and let $H \subset G$ be a subgroup of $G$.*
(a) *Every element of $G$ is in some coset of $H$.*
(b) *Every coset of $H$ has the same number of elements.*
(c) *Let $g_1, g_2 \in G$. Then the cosets $g_1H$ and $g_2H$ satisfy either*

$$g_1H = g_1H \quad \textit{or} \quad g_1H \cap g_2H = \emptyset.$$

*Using set-theoretic terminology, this says that cosets of $H$ are either equal or disjoint*

*Proof.* (a) This is easy. Let $g \in G$. The subgroup $H$ contains the identity element $e$, so the coset $gH$ contains $g \cdot e = g$.
(b) Let $g_1, g_2 \in G$. We claim that

$$F : g_1H \longrightarrow g_2H, \quad F(g) = g_2g_1^{-1}g$$

is a well-defined bijective map from $g_1H$ to $g_2H$. We first check that $F$ is well-defined. Let $g \in g_1H$, so $g = g_1h$ for some $h \in H$. Then

$$F(g) = g_2g_1^{-1}g = g_2g_1^{-1}g_1h = g_2h \in g_2H,$$

Next we check that $F$ is injective. Suppose that $g, g' \in g_1H$ and $F(g) = F(g')$. This means that $g_2g_1^{-1}g = g_2g_1^{-1}g'$. Multiplying by $g_1g_2^{-1}$ on left gives

$$(g_1g_2^{-1})(g_2g_1^{-1}g) = (g_1g_2^{-1})(g_2g_1^{-1}g'), \quad \text{and canceling yields} \quad g = g'.$$

Finally, we check that $F$ is surjective. Let $g \in g_2H$, so $g = g_2h$ for some $h \in H$. Then $g_1h \in g_1H$, and $F(g_1h) = g_2g_1^{-1}g_1h = g_2h = g$.
(c) If $g_1H \cap g_2H = \emptyset$, we are done, so assume the two cosets are not disjoint. This means we can find elements $h_1, h_2 \in H$ satisfying $g_1h_1 = g_2h_2$. We rewrite this as $g_1 = g_2h_2h_1^{-1}$. Now take any element $g \in g_1H$. We need to show that $g$ is also in $g_2H$. We write $g$ as $g = g_1h$ for some $h \in H$. Then

$$g = g_1h = g_2h_2h_1^{-1}h \in g_2H,$$

since the assumption that $H$ is a subgroup ensures that the product $h_2h_1^{-1}h$ is in $H$. This shows that every element of $g_1H$ is in $g_2H$, and a similar argument shows the reverse inclusion. Alternatively, we can use the fact from (b) that $g_1H$ and $g_2H$ have the same number of elements, so if one is a subset of the other, they must be equal. □

We are now going to use the properties of cosets proven in Proposition 1.24 to derive a fundamental divisibility property for the orders of subgroups.

**Theorem 1.25** (Lagrange's Theorem)**.** *Let $G$ be a finite group and let $H$ be a subgroup of $G$. Then the order of $H$ divides the order of $G$.*

*Proof.* We start by choosing elements $g_1, \ldots, g_k \in G$ so that $g_1H, \ldots, g_kH$ is a list of all of the different cosets of $H$. Proposition 1.24(a) tells us that every element of $G$ is in some coset of $H$, so $G$ is the equal to the union

$$G = g_1H \cup g_2H \cup \cdots \cup g_kH. \tag{1.7}$$

On the other hand, Proposition 1.24(c) tells us that distinct cosets have no elements in common, i.e., if $i \neq j$, then $g_iH \cap g_jH = \emptyset$. Thus the union in (1.7) is a disjoint union, so the number of elements in $G$ is the sum of the number of elements in the cosets,

$$\#G = \#g_1H + \#g_2H + \cdots + \#g_kH. \tag{1.8}$$

We next invoke Proposition 1.24(b), which tells us that every coset has the same number of elements, so in particular, $\#g_iH = \#eH = \#H$. Using this fact in (1.8) yields

$$\#G = k\#H.$$

Thus the order of $G$ is a multiple of the order of $H$, which completes the proof of Lagrange's theorem. □

**Corollary 1.26.** *Let $G$ be a finite group and let $g \in G$. Then the order of $g$ divides the order of $G$.*

*Proof.* The order of the subgroup $\langle g \rangle$ generated by $G$ is equal to the order of the element $g$, and Theorem 1.25 tells us that the order of $\langle g \rangle$ divides the order of $G$. □

We give one application of Lagrange's theorem. It marks the starting line of a long and ongoing mathematical journey that strives to classify finite groups according to their orders.

**Proposition 1.27.** *Let $p$ be a prime, and let $G$ be a finite group of order $p$. Then $G$ is isomorphic to the cyclic group $\mathcal{C}_p$.*

*Proof.* Since $p \geq 2$, we know that $G$ contains more than just the identity element, so we choose some non-identity element $g \in G$. Lagrange's theorem (Theorem 1.25) tells us that the order of the subgroup $\langle g \rangle$ generated by $g$ divides the order of $G$. But $\#G = p$ is prime, so $\#\langle g \rangle$ equals 1 or $p$, and we know that it doesn't equal 1, since $\langle g \rangle$ contains $e$ and $g$. Hence $\#\langle g \rangle = p = \#G$. Thus the subgroup has the same number of elemeents as the full group, so they are equal, $G = \langle g \rangle$. Writing the cyclic group $\mathcal{C}_n$ as $\mathcal{C}_n = \{g_0, g_1, g_2, \ldots, g_{n-1}\}$, with group law as described in Example 1.8, we obtain an isomorphism

$$\mathcal{C}_n \longrightarrow G, \quad g_i \longmapsto g^i.$$

This completes the proof of the proposition. □

**Mini-Remark 6**. The vast theory of finite groups includes many fascinating, and frequently unexpected, results whose proofs are unfortunately beyond the scope of these notes. To whet your appetite for studying more group theory, we state two such theorems.

**Theorem 1.28.** *Let $p$ be a prime number, and let $G$ be a group of order $p^2$. Then $G$ is an abelian group.*

On the other hand, we know that there exist non-abelian groups of order $p^3$. For example, the dihedral group $\mathcal{D}_4$ (Example 1.11) and the quaternion group $\mathcal{Q}$ (Example 1.12) are non-abelian groups of order 8. The next result is an important partial converse to Lagrange's theorem.

**Theorem 1.29** (Sylow's Theorem)**.** *Let $G$ be a group, let $p$ be a prime, and suppose that $p^n$ divides $\#G$ for some power $n \ge 1$. Then $G$ has a subgroup of order $p^n$.*

One might hope, more generally, that if $d$ is any number that divides the order of $G$, then $G$ has a subgroup of order $d$. Unfortunately, this is not true, although we have not yet seen a group that is a counterexample.

## 1.6 Normal Subgroups and Quotient Groups (*Optional*)

**This Section is Under Construction**

Let $\phi : G \to G'$ be a homomorphism of groups. We proved in Proposition 1.20 that $\ker(\phi)$ is a subgroup of $G$. It actually is a special sort of subgroup, which prompts the following definition.

**Definition.** Let $G$ be a group, let $H \subset G$ be a subgroup, and let $g \in G$. The *$g$-conjugate of $H$* is the subgroup

$$g^{-1}Hg = \{g^{-1}hg : g \in G\}.$$

We say that $H$ is a *normal subgroup of $G$* if it satisfies

$$g^{-1}Hg = H \quad \text{for every } g \in G.$$

**Proposition 1.30.** *Let $G$ be a group, let $H \subset G$ be a subgroup, and let $g \in G$.*
(a) *The conjugate $g^{-1}Hg$ is a subgroup of $G$*
(b) *The map $H \to g^{-1}Hg$ defined by $h \mapsto g^{-1}hg$ is an group isomorphism.*

*Proof.* □

**Proposition 1.31.** *Let $\phi : G \to G'$ be a homomorphism of groups. Then $\ker(\phi)$ is a normal subgroup of $G$.*

*Proof.* We already know from Proposition 1.20(a) that $\ker(\phi)$ is a subgroup of $G$. Let $h \in \ker(\phi)$ and $g \in G$. Then

$$\begin{aligned}
\phi(g^{-1} \cdot h \cdot g) &= \phi(g^{-1}) \cdot \phi(h) \cdot \phi(g) && \text{homomorphism property of } \phi,\\
&= \phi(g)^{-1} \cdot \phi(h) \cdot \phi(g) && \text{Proposition 1.13(b)},\\
&= \phi(g)^{-1} \cdot \phi(g) && \text{since } h \in \ker(\phi),\\
&= e'.
\end{aligned}$$

Hence $g^{-1} \cdot h \cdot g \in \ker(\phi)$. We have proven that this is true for all $h \in \ker(\phi)$ and all $g \in G$, which completes the proof that $\ker(\phi)$ is a normal subgroup of $G$. □

We now want to turn Proposition 1.31 on its head and use a given normal subgroup $H \subset G$ to create a group $G'$ and a group homomorphism $\phi : G \to G'$ with the property that $\ker(\phi) = H$.

Recall from Section 1.5 that if $H$ is a subgroup of $G$ and $g \in G$, then the corresponding left coset of $H$ is the set

$$gH = \{gh : h \in H\}.$$

We would like to take the collection of cosets of $H$ and make it into a group. It is convenient to have a notation for the set of cosets.

**Definition.** Let $G$ be a group, and let $H$ be a subgroup of $G$. We denote the set of cosets of $G$ by

$$G/H = \{\text{cosets of } H\}.$$

There is a natural way try to make $G/H$ into a group, namely by defining a group law on cosets via the rule

$$g_1H \cdot g_2H = g_1g_2H. \tag{1.9}$$

But there is a serious potential problem that our notation conceals. The issue is that although every coset of $H$ has the form $gH$, there are lots of choices for $g$ that give the same coset. Indeed, if $h \in H$ is any element of $H$, then $hH = H$, so $ghH = gH$.[10] So in (1.9), if we choose different elements $g_1$ and $g_2$ of $G$ that give the same cosets, how do we know that we get the same product coset? The answer is that in general, we do not get the same product. ☹ However, if $H$ is a *normal* subgroup of $G$, then turning darkness to light ☺, we do get the same product coset, as we now verify.

**Lemma 1.32.** *Let $G$ be a group, and let $H$ be a normal subgroup of $G$. Let $g_1, g_1', g_2, g_2' \in G$ be elements of $G$ satisfying*

$$g_1'H = g_1H \quad \textit{and} \quad g_2'H = g_2H.$$

*Then*

$$g_1'g_2'H = g_1g_2H.$$

[10] Exercise 1.21 says that the converse is also true, i.e., if $g_1H = g_2H$, then there is an $h \in H$ such that $g_1 = g_2h$.

*Proof.* The assumption that $g_1'H = g_1H$ implies that there is an $h_1 \in H$ such that $g_1' = g_1h_1$. (This assertion is part of Exercise 1.21, but it is very easy. Here is the short proof: $g_1' = g_1' \cdot e \in g_1'H = g_1H$.) Similarly the assumption that $g_2'H = g_2H$ implies that there is an $h_2 \in H$ such that $g_2' = g_2h_2$.

Let $g_1'g_2'h$ be an element of $g_1'g_2'H$. We want to show that $g_1'g_2'h$ is in $g_1g_2H$. To do this, we compute

$$\begin{aligned}
g_1'g_2'h &= g_1h_1g_2h_2h && \text{since } g_1' = g_1h_1 \text{ and } g_2' = g_2h_2,\\
&= g_1(g_2g_2^{-1})h_1g_2h_2h && \text{inserting } g_2g_2^{-1} = e \text{ doesn't change the value,}\\
&= g_1g_2(g_2^{-1}h_1g_2)h_2h && \text{associative law of group multiplication,}\\
&\in g_1g_2H && \text{the normality of } H \text{ tells us that } g_2^{-1}h_1g_2 \in H, \text{ so}\\
& && g_2^{-1}h_1g_2 \cdot h_2 \cdot h \text{ is a product of three elements of}\\
& && H, \text{ and thus is in } H.
\end{aligned}$$

Since this is true for every $h \in H$, we have proven that

$$g_1'g_2'H \subseteq g_1g_2H.$$

Reversing the roles of $g_1, g_2$ and $g_1', g_2'$ gives the opposite inclusion. This completes the proof that $g_1'g_2'H = g_1g_2H$. □

The content of Lemma 1.32 is that the multiplication rule $g_1H \cdot g_2H = g_1g_2H$ on cosets of $H$ is well-defined provided we take $H$ to be a normal subgroup of $G$. The following properties of coset multiplication then follow directly from the corresponding properties of the group operation on $G$:

$$\begin{aligned}
eH \cdot gH = gH \cdot eH = gH,\\
gH \cdot g^{-1}H = g^{-1}H \cdot gH = eH,\\
(g_1H \cdot g_2H) \cdot g_3H = g_1H \cdot (g_2H \cdot g_3H).
\end{aligned}$$

We have proven the first part the following important theorem.

**Theorem 1.33.** *Let $G$ be a group, and let $H$ be a normal subgroup of $G$.*

(a) *The collection of cosets $G/H$ is a group via the well-defined group operation*

$$g_1H \cdot g_2H = g_1g_2H. \tag{1.10}$$

(b) *The map*

$$\phi : G \longrightarrow G/H, \quad \phi(g) = gH,$$

*is a homomorphism whose kernel is* $\ker(\phi) = H$.

(c) *Let*

$$\psi : G \longrightarrow G'$$

*be a homomorphism with the property that $H \subset \ker(\psi)$. Then there is a unique homomorphism*

$$\lambda : G/H \longrightarrow G' \quad \textit{satisfying} \quad \lambda(gH) = \psi(g).$$

*Proof.* (a) The fact that the group operation (1.10) is well-defined is exactly what Lemma 1.32 says, and as we noted earlier, the group axioms for $G/H$ follow immediately from the groups axioms for $G$.
(b) In order to check that $\phi$ is a homomorphism, we compute

$$\phi(g_1)\phi(g_2) = g_1H \cdot g_2H = g_1g_2H = \phi(g_1g_2).$$

The kernel of $\phi$ is

$$\ker(\phi) = \big\{g \in G : \phi(g) = eH\big\} = \big\{g \in G : gH = H\big\} = H.$$

(c) We would like to define $\lambda : G/H \to G'$ by the following three-step algorithm:

(1) Let $C \in G/H$ be a coset.

(2) Choose some $g \in G$ with $C = gH$

(3) Define $\lambda(C)$ to be $\psi(g)$.

However, there is a potential problem, since there are usually lots of choices for $g$ in Step (2). So we need to prove the following assertion:

$$\text{If} \quad g'H = gH, \quad \text{then} \quad \psi(g') = \psi(g). \tag{1.11}$$

The assumption that $g'H = gH$ means that $g' = gh$ for some $h \in H$. This allows us to compute

$$\begin{aligned}
\psi(g') &= \psi(gh) && \text{since } g' = gh,\\
&= \psi(g) \cdot \psi(h) && \text{since } \psi \text{ is a group homomorphism,}\\
&= \psi(g) \cdot e' && \text{since } h \in H \text{ and } H \subset \ker(\psi),\\
&= \psi(g).
\end{aligned}$$

This proves assertion (1.11), so our three-step algorithm gives a well-defined map $\lambda : G/H \to G'$. And now that we know that $\lambda$ is well-defined, it's easy to check that it is a homomorphism,

$$\lambda(g_1g_2H) = \psi(g_1g_2) = \psi(g_1) \cdot \psi(g_2) = \lambda(g_1H) \cdot \lambda(g_2H).$$

Finally for a given homomorphism $\psi$, it is clear that there is only one map $\lambda$ satisfying $\psi(g) = \lambda(gH)$, since this equality completely determines the values of $\lambda$ in terms of the values of $\psi$. □

## 1.7 Cayley's Theorem (*Optional*)

**This Section is Under Construction**

**Theorem 1.34** (Cayley's Theorem)**.** *Every (finite) group is isomorphic to a subgroup of some symmetric group.*

# Exercises

## Section 1.1. Introduction to Groups

**1.1.** Let $n$ be a positive integer, and let $G$ be the group of permutations of the set $\{1, 2, \ldots, n\}$. Prove that $G$ is a finite group, and give a formula for the order of $G$.

**1.2.** (a) Let $S$ be a finite set, and let $\phi : S \to S$ be a function. Prove that the following are equivalent:

(i) $\phi$ is injective. (ii) $\phi$ is surjective. (iii) $\phi$ is bijective.

(b) Give an example of an infinite set $S$ and a function $\phi : S \to S$ such that $\phi$ is injective, but is not surjective.

(c) Give an example of an infinite set $S$ and a function $\phi : S \to S$ such that $\phi$ is surjective, but is not injective.

**1.3.** Figure 1.2 shows various rotations and flips of a square. Fill in the boxes with the correct vertex labels for the indicated operations.

## Section 1.2. Abstract Groups

**1.4.** Let $G$ be a group. In this exercise you will prove the remaining parts of Proposition 1.3. Be sure to justify each step using the group axioms or by reference to a previously proven fact.

(a) $G$ has exactly one identity element.

(b) Let $g, h \in G$. Then $(g \cdot h)^{-1} = h^{-1} \cdot g^{-1}$.

(c) Let $g \in G$. Then $(g^{-1})^{-1} = g$.

**1.5.** Let $G$ be a group, let $g$ and $h$ be elements of $G$, and suppose that $g$ has order $n$ and $h$ has order $m$.

(a) If $G$ is an abelian group and if $\gcd(m, n) = 1$, prove that the order of $gh$ is $mn$.

(b) Give an example of an abelian group showing that (a) need not be true if $\gcd(m, n) > 1$.

(c) Give an example a nonabelian group showing that (a) need not be true even if we retain the requirement that $\gcd(m, n) = 1$.

## Section 1.3. Interesting Examples of Groups

**1.6.** Let $G$ be a finite cyclic group of order $n$, and let $g$ be a generator of $G$. Prove that $g^k$ is a generator of $G$ if and only if $\gcd(k, n) = 1$.

**1.7.** Figure 1.3 shows a hexagon in its initial and three subsequent positions. It thus illustrates four elements $e, r_1, f_1, f_2$ of the dihedral group $\mathcal{D}_6$, where $e$ is the identiy element, $r_1$ is a rotation, and $f_1$ and $f_2$ are flips about the indicated axes. We mention that the flips are the same as those given in Figure 1.1, but the rotation is different.

(a) Write down and give names to the other 8 ways to rotate and/or flip the hexagon. The 12 pictures illustrate the 12 elements of the dihedral group $\mathcal{D}_6$.

(b) What is the smallest power of each of $r_1$, $f_1$, and $f_2$ that is equal to the identity transformation $e$?

(c) Write down the hexagon configurations that correspond to the compositions $r_1 f_1$, $f_1 r_1$, $r_1 f_2$, and $f_2 r_1$. Does $r_1$ commute with $f_1$ or $f_2$.

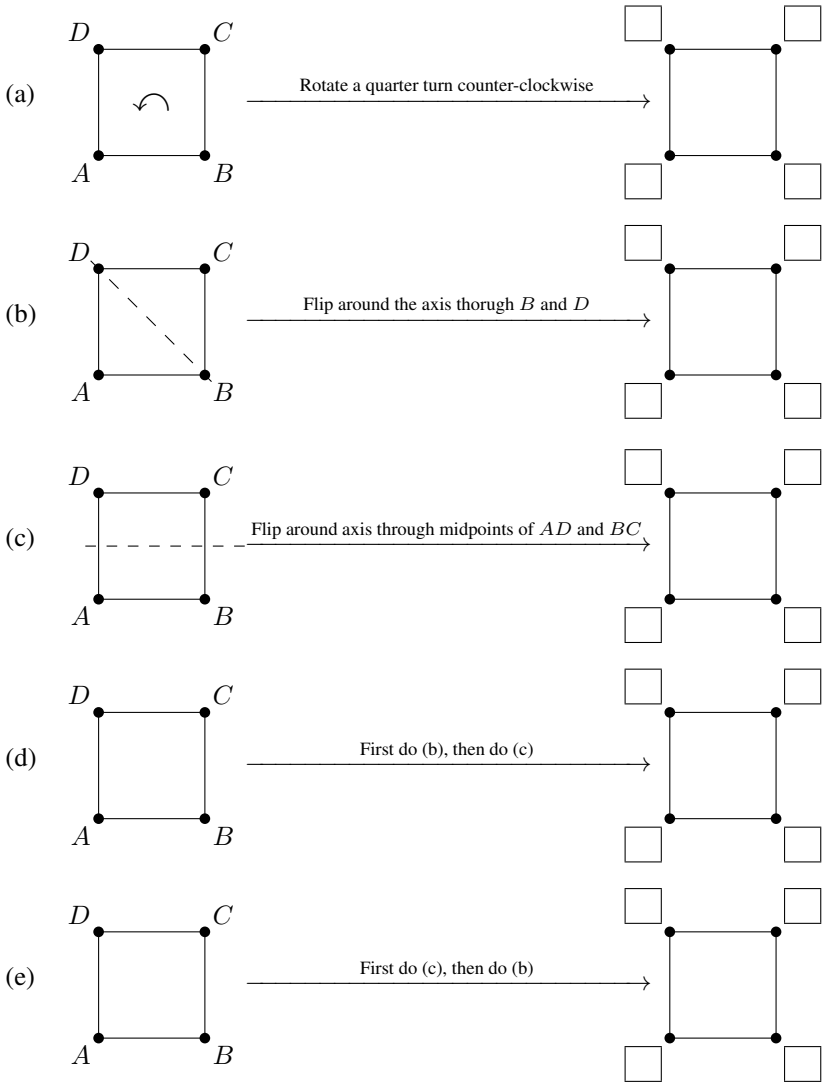



Figure 1.2: Motions of a Square for Exercise 1.3

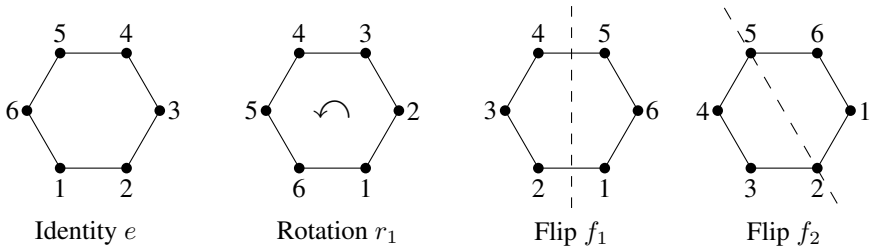



Figure 1.3: A rotation and two flips of a regular hexagon

(d) Write down the hexagon configurations that correspond to the $f_1 f_2$ and $f_2 f_1$. Show that each of them is equal to a power of the rotation $r_1$, i.e., the compositions of these two flips gives a rotation.
(e) Prove that every rotation is equal to some power of $r_1$.
(f) Prove that every flip is equal the composition of $f_1$ and some power of $r_1$.
(g) Using (e) and (f), prove that the entire group $\mathcal{D}_6$ consists of the 12 elemenets

$$\{f_1^i r_1^j : 0 \le i \le 1 \text{ and } 0 \le j \le 5\}.$$

(h) Express $r_1 f_1$ in the form $f_1^i r_1^j$.
(i) More generally, describe one or more formulas that explain how to write the product $(f_1^k r_1^\ell)(f_1^m r_1^n)$ in the form $f_1^i r_1^j$.

**1.8.** Prove that the dihedral group $\mathcal{D}_n$, as described in Example 1.11, has exactly $2n$ elements.

**1.9.** (a) Let $\mathbb{Q}^*$ be the set of non-zero rational numbers, with the group law being multiplication. Prove that $\mathbb{Q}^*$ is a group.
(b) Let $p$ be a prime number. Prove that the non-zero elements of $\mathbb{Z}/p\mathbb{Z}$ form a group using multiplication as the group law.
(c) Let $m \ge 4$ be an integer that is not a prime number. Prove that the non-zero elements of $\mathbb{Z}/m\mathbb{Z}$ do not form a group using multiplication as the group law. (Try it first with $m = 4$ and $m = 6$ to see what's going on.)
(d) Let $m \ge 2$ be any integer, and define

$$(\mathbb{Z}/m\mathbb{Z})^* = \big\{a \in \mathbb{Z}/m\mathbb{Z} : a \not\equiv 0 \pmod m\big\}.$$

Prove that $(\mathbb{Z}/m\mathbb{Z})^*$ forms a group using multiplication as the group law.

**1.10.** Let $\mathbb{C}$ be the set of complex numbers, that is, the set of numbers of the form $x + yi$, where $x, y \in \mathbb{R}$ and $i^2 = -1$.
(a) We make $\mathbb{C}$ into a group using addition. What is the identity element? What is the inverse of an element $z \in \mathbb{C}$?
(b) Let $\mathbb{C}^*$ be the set of non-zero complex numbers. We make $\mathbb{C}^*$ into a group using multiplication. What is the identity element? What is the inverse of an element $z \in \mathbb{C}^*$? Be sure to write you answers as a real number added to $i$ times another real number.

**1.11.** (a) Let

$$\mathrm{GL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R},\ ad - bc \ne 0 \right\}.$$

be the indicated set of 2-by-2 matrices, with composition law being matrix multiplication, as described in Example 1.10. Prove that $\mathrm{GL}_2(\mathbb{R})$ is a group.

(b) Let $\mathrm{SL}_2(\mathbb{R})$ be the set of 2-by-2 matrices

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{R},\ ad-bc=1 \right\}.$$

Prove that $\mathrm{SL}_2(\mathbb{R})$ is a group, where the group law is again matrix multiplication.

(c) This part is for those who have studied $n$-dimensional linear algebra. Fix an integer $n \ge 1$. Generalize (a) and (b) by proving that each of the following sets of $n$-by-$n$ matrices is a group using matrix multiplication for the group law:

$$\mathrm{GL}_n(\mathbb{R}) = \left\{ n\text{-by-}n \text{ matrices } A \text{ with real entries satisfying } \det(A) \ne 0 \right\},$$
$$\mathrm{SL}_n(\mathbb{R}) = \left\{ n\text{-by-}n \text{ matrices } A \text{ with real entries satisfying } \det(A) = 1 \right\}.$$

The group $\mathrm{GL}_n(\mathbb{R})$ is called the *general linear group*, and the group $\mathrm{SL}_n(\mathbb{R})$ is called the *special linear group*.

**1.12.** Let $\mathrm{GL}_2(\mathbb{R})$ be the general linear group as described in Example 1.10 and Exercise 1.11(a). Prove or disprove that each of the following subsets of $\mathrm{GL}_2(\mathbb{R})$ is a subgroup of $\mathrm{GL}_2(\mathbb{R})$. In the case of non-subgroups, indicate which of the subgroup conditions fail.

(a) $\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) : ad-bc=2 \right\}$.

(b) $\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) : ad-bc \in \{-1,1\} \right\}$.

(c) $\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) : c=0 \right\}$.

(d) $\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) : d=0 \right\}$.

(e) $\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R}) : a=d=1 \text{ and } c=0 \right\}$.

**1.13.** Let $\mathcal{Q} = \{\pm 1, \pm \boldsymbol{i}, \pm \boldsymbol{j}, \pm \boldsymbol{k}\}$ be the group of quaternions as describe in Example 1.12. We claimed there that the group law for $\mathcal{Q}$ is determined by the formulas

$$\boldsymbol{i}\cdot\boldsymbol{i} = -1, \qquad \boldsymbol{j}\cdot\boldsymbol{j} = -1, \qquad \boldsymbol{k}\cdot\boldsymbol{k} = -1, \qquad \boldsymbol{i}\cdot\boldsymbol{j}\cdot\boldsymbol{k} = -1.$$

Use these formulas to prove the following formulas, which completely determine the group operations on $\mathcal{Q}$:

$$\begin{aligned} &\boldsymbol{i}\cdot\boldsymbol{j} = \boldsymbol{k}, && \boldsymbol{j}\cdot\boldsymbol{k}, = \boldsymbol{i} && \boldsymbol{k}\cdot\boldsymbol{i} = \boldsymbol{j}, \\ &\boldsymbol{j}\cdot\boldsymbol{i} = -\boldsymbol{k}, && \boldsymbol{k}\cdot\boldsymbol{j}, = -\boldsymbol{i} && \boldsymbol{i}\cdot\boldsymbol{k} = -\boldsymbol{j}. \end{aligned}$$

**1.14.** We can form groups of matrices whose entries are in any algebraic system where we can add, subtract, and multply. For example, let $m \ge 2$ be an integer, and define

$$\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}/m\mathbb{Z},\ ad-bc=1 \right\}.$$

Prove that matrix multiplication (1.3) makes $\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})$ into a non-commutative group.

## Section 1.4. Group Homomorphisms

**1.15.** Recall that two groups $G_1$ and $G_2$ are said to be *isomorphic* if there is a bijective homomorphism

$$\phi : G_1 \longrightarrow G_2.$$

The fact that $\phi$ is bijective means that the inverse map $\phi^{-1} : G_2 \to G_1$ exists. Prove that $\phi^{-1}$ is a homomorphism from $G_2$ to $G_1$.

**1.16.** Complete the proof that the map $\mathcal{D}_n \to \{\pm 1\}$ in Example 1.14 is a homomorphism by showing that compositions of rotations and flips satisfy the rules shown in equation 1.6.

**1.17.** In this exercise, $\mathcal{C}_n$ is a cyclic group of order $n$, $\mathcal{D}_n$ is the $n$'th dihedral group, and $\mathcal{S}_n$ is the $n$'th symmetric group.

(a) Prove that $\mathcal{C}_2$ and $\mathcal{S}_2$ are isomorphic.

(b) Prove that $\mathcal{D}_3$ is isomorphic to $\mathcal{S}_3$.

(c) Let $m \ge 3$. Prove that for every $n$, the groups $\mathcal{C}_m$ and $\mathcal{S}_n$ are not isomorphic.

(d) Prove that for every $n \ge 4$, the groups $\mathcal{D}_n$ and $\mathcal{S}_n$ are not isomorphic.

(e) More generally, let $m \ge 4$ and let $n \ge 4$. Prove the groups $\mathcal{D}_m$ and $\mathcal{S}_n$ are not isomorphic.

(f) The dihedral group $\mathcal{D}_4$ (Example 1.11) and the quaternion group $\mathcal{Q}$ (Example 1.12) are non-abelian groups of order 8. Prove that they are not isomorphic. (*Hint*. How many elements of order 2 and order 4 are there in $\mathcal{D}_4$ and $\mathcal{Q}$?)

**1.18.** Let $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ be the group that we defined in Exercise 1.14.

(a) Prove that $\#\,\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z}) = 6$.

(b) Prove that $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ is isomorphic to the symmetric group $\mathcal{S}_3$. (*Hint*. Show that the matrices in $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ permute the vectors in the set $\big\{(1,0),(0,1),(1,1)\big\}$, where the coordinates of the vectors are viewed as numbers modulo 2.)

## Section 1.5. Subgroups, Cosets, and Lagrange's Theorem

**1.19.** Let $G$ be a cyclic group of order $n$, and let $d$ be an integer that divides $n$. Prove that $G$ has a subgroup of order $d$.

**1.20.** Let $G$ be a group, and let $H \subset G$ be a subset of $G$. Prove that $H$ is a subgroup if and only if it has the following two properties:

(1) $H \ne \emptyset$

(2) For every $h_1, h_2 \in H$, the product $h_1 \cdot h_2^{-1}$ is in $H$.

**1.21.** This exercise explains when two elements of $G$ determine the same coset of $H$. Let $G$ be a group, let $H$ be a subgroup of $G$, and let $g_1, g_2 \in G$. Prove that the following three statements are equivalent:

(a) (1) $g_1 H = g_2 H$.

(b) (2) There is an element $h \in H$ such that $g_1 = g_2 h$.

(c) (3) $g_2^{-1} g_1 \in H$.

**1.22.** Let $G$ be a finite group whose only subgroups are $\{e\}$ and $G$. Prove that $G$ is a cyclic group whose order is a prime.

**1.23.** Let $G$ be a group and $H \subset G$ a subgroup. The *index of $H$ in $G$*, which is denoted by $(G : H)$, is the quantity $\#G/\#H$.

(a) Prove that $(G : H)$ is the number of distinct cosets of $H$.

(b) Suppose that $K \subset H$ is a subgroup of $H$, so we may also view $K$ as a subgroup of $G$. In other words, $K \subset H \subset G$ is a chain of subgroups. Prove the *Index Multiplication Rule*

$$(G : K) = (G : H)(H : K).$$

(*Hint*. Count cosets.)

Section 1.6. Normal Subgroups and Quotient Groups (*Optional*)

Section 1.7. Cayley's Theorem (*Optional*)

# Chapter 2

# Rings

## 2.1 Introduction to Rings

When we introduced groups in Chapter 1, they were probably unfamiliar to most of you. In this chapter we introduce another fundamental type of algebraic object, called a *ring*. The good news is that you are already familiar with many rings. Here are some examples:

- The integers $\mathbb{Z}$ are a ring.
- The rational numbers $\mathbb{Q}$ and the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ are rings. (They are a special type of ring, called a field, but that's a topic for a later chapter!)
- The set of mod $m$ integers $\mathbb{Z}/m\mathbb{Z}$ that you studied in the Number Theory Unit forms a ring.

What do these examples have in common? They each have two operations, addition and multiplication. These operations, individually, satisfy some axioms, and the two operations interact via one further axiom, the all-powerful distributive law,

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

In general, a ring is a set with two operations satisfying a bunch of axioms that are modeled after the properties satisfied by addition and multiplication of integers.

## 2.2 Abstract Rings and Ring Homomorphisms

**Definition.** A *ring* $R$ is a set with two operations, generally called *addition* and *multiplication* and written

$$\underbrace{a+b}_{\text{addition}} \quad \text{and} \quad \underbrace{a \cdot b \text{ or } ab}_{\text{multiplication}},$$

satisfying the following axioms:

(1) The set $R$ with its addition law $+$ is an abelian group. The identity element of this group is denoted $0$ or $0_R$.
(2) The set $R$ with its multiplication law $\cdot$ is almost a group, but its elements are not required to have inverses.[1] Explicitly, the multiplication law of a ring satisfies:
   - There is an element $1_R \in R$ satisfying[2]
$$1_R \cdot a = a \cdot 1_R = a \quad \text{for all } a \in R.$$
   - The associative law holds,
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c \quad \text{for all } a, b, c \in R.$$
(3) [Distributive Law] For all $a, b, c \in R$ we have
$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{and} \quad (b + c) \cdot a = b \cdot a + c \cdot a.$$
(4) If further $a \cdot b = b \cdot a$ for all $a, b \in R$, then the ring is said to be *commutative*.

Your long experience with the ring of integers $\mathbb{Z}$ might lead you to assume that various "obvious" formulas are true in every ring. For example, the formulas
$$0_R \cdot a = 0_R \quad \text{and} \quad (-a) \cdot (-b) = a \cdot b$$
must be true, right? But why should they be true? The definition of $0_R$ is as the identity element for *addition*, i.e., $a + 0_R = 0_R + a = a$ for every $a \in R$, so why should that tell us anything about $0_R$ when we switch to *multiplication*? Similarly, the defintion of $-a$ is as the element that gives $0_R$ when it is *added* to $a$, which seems to tell us very little about the *product* of $-a$ with other elements of $R$. The only hope of proving multiplication properties for $0_R$ and $-a$ lies in the distributive law, which intertwines addition and multiplication. Study closely the use of the distributive law in the following proof that $0_R \cdot a = 0_R$.

**Proposition 2.1.** *Let $R$ be a ring.*
(a) $0_R \cdot a = 0_R$ *for all $a \in R$.*
(b) $(-a) \cdot (-b) = a \cdot b$ *for all $a, b \in R$. In particular, we have $(-1_R) \cdot a = -a$.*

*Proof.* (a) We start with $1_R = 1_R + 0_R$, which is true because $0_R$ is the identity for addition. We multiply both sides by $a$ and compute

$$\begin{aligned}
a &= a \cdot 1_R && \text{since } 1_R \text{ is the identity for multiplication,} \\
&= a \cdot (1_R + 0_R) && \text{since } 0_R \text{ is the identity for addition,} \\
&= a \cdot 1_R + a \cdot 0_R && \text{distributive law,} \\
&= a + a \cdot 0_R && \text{since } 1_R \text{ is the identity for multiplication.}
\end{aligned}$$

[1] For those who are interested, algebraic objects that are like groups except that not every element needs to have an inverse also have a name; they are called *monoids*.

[2] To avoid the trivial ring consisting of a single element, we also include the requirement that $1_R \neq 0_R$.

We now "subtract" $a$ from both sides. But this one time we will spell out every detail so that you can see how the different ring axioms come into play:

$$\begin{aligned}
0_R &= (-a) + a && \text{definition of inverse for addition,}\\
&= (-a) + (a + a \cdot 0_R) && \text{from our earlier calculation,}\\
&= \big((-a) + a\big) + a \cdot 0_R && \text{associativity of addition,}\\
&= 0_R + a \cdot 0_R && \text{definition of inverse for addition,}\\
&= a \cdot 0_R && \text{since } 0_R \text{ is the identity for addition.}
\end{aligned}$$

(b) We leave this part for you to do; see Exercise 2.1. □

Just as we did with groups, we want to look at maps

$$\phi : R \to R'$$

between rings that respects the "*ring-i-ness*" of $R$ and $R'$. Rings are characterized by their addition and multiplication laws, leading to the following definition.

**Definition.** Let $R$ and $R'$ be rings. A *ring homomorphism from $R$ to $R'$* is a function $\phi : R \to R'$ satisfying

$$\begin{aligned}
\phi(1_R) &= 1_{R'},\\
\phi(a+b) &= \phi(a) + \phi(b) \quad \text{for all } a, b \in R,\\
\phi(a \cdot b) &= \phi(a) \cdot \phi(b) \quad \text{for all } a, b \in R.
\end{aligned}$$

The *kernel of $\phi$* is the set of elements that is sent to 0,,

$$\ker(\phi) = \big\{a \in R : \phi(a) = 0\big\}.$$

(The zero here is, of course, the zero element in $R'$.)

In the next section, after we have a few more examples of ring, we will give some examples of ring homomorphisms.

*Remark* 2.2. The axiom $\phi(1_R) = 1_{R'}$ is included to rule out the boring and trivial map $\phi(a) = 0_{R'}$ that sends every $a \in R$ to zero.

## 2.3 Interesting Examples of Rings

Four of the rings that we described in Section 2.1 fit one into another, sort of like Russian stacking dolls:

$$\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

We say that $\mathbb{Z}$ is a *subring* of $\mathbb{Q}$, and similarly for the others. The fifth ring that we mentioned in the introduction is $\mathbb{Z}/m\mathbb{Z}$, the ring of integers modulo $m$. The ring $\mathbb{Z}/m\mathbb{Z}$ is not a subring of $\mathbb{C}$, but there is a very natural homomorphism

$$\phi : \mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}, \quad \phi(a) = a \bmod m,$$

called naturally enough the *reduction mod $m$ homomorphism*. This homomorphism sends an integer to its congruence class modulo $m$, and its kernel is the set of all multiples of $m$. The fact that $\phi$ is a homomorphism means checking that reduction modulo $m$ behaves well for addition and multiplication, facts that you saw in the Number Theory Unit.

*Example* 2.3 (**Gaussian Integers $\mathbb{Z}[i]$**). Here is another interesting subring of $\mathbb{C}$. It is called the *ring of Gaussian integers*.

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

The quantity $i$ is, as usual, a symbol that represents a square root of $-1$. Addition and multiplication of elements in $\mathbb{Z}[i]$ follow the usual rules for adding and multiplying complex numbers,

$$\begin{aligned}(a_1 + b_1 i) + (a_2 + b_2 i) &= (a_1 + a_2) + (b_1 + b_2)i,\\ (a_1 + b_1 i) \cdot (a_2 + b_2 i) &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i.\end{aligned}$$

Note that if we allowed $a$ and $b$ to be real numbers, then we would get the entire ring of complex numbers; but we're restricting $a$ and $b$ to be integers.

*Example* 2.4 (**Polynomial Rings $R[x]$**). Polynomial rings are a way to create bigger (and better?) rings from an old ring. Thus for any commutative ring $R$, we use $R$ to build the *ring of polynomials over $R$*,

$$R[x] = \left\{\begin{array}{l}\text{polynomials } a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \text{ of all}\\ \text{degrees with coefficients } a_0, a_1, \ldots, a_d \in R\end{array}\right\}.$$

You've undoubtedly seen polynomials whose coefficients are real numbers, but the rules that you learned to add and multiply polynomials work with coefficients in any commutative ring. Indeed, the rule for multiplying polynomials is forced on you by the distributive law. Here's a simple example:

$$\begin{aligned}&(a_0 + a_1 x + a_2 x^2) \cdot (b_0 + b_1 x)\\ &\qquad = a_0 \cdot (b_0 + b_1 x) + a_1 x \cdot (b_0 + b_1 x) + a_2 x^2 \cdot (b_0 + b_1 x)\\ &\qquad = (a_0 b_0 + a_0 b_1 x) + (a_1 b_0 x + a_1 b_1 x^2) + (a_2 b_0 x^2 + a_2 b_1 x^3)\\ &\qquad = a_0 b_0 + (a_1 b_1 + a_1 b_0)x + (a_1 b_1 + a_2 b_0)x^2 + (a_2 b_1)x^3.\end{aligned}$$

Let

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \in R[x]$$

be a polynomial. Then for any element $c \in R$, we can *evaluate $f$ at $c$* simply by substituting $c$ for $x$. Thus

$$f(c) = a_0 + a_1 c + a_2 c^2 + \cdots + a_d c^d \in R.$$

In other classes you probably took a polynomial $f(x)$ and evaluated it at lots of different values. In other words, you viewed $f(x)$ as defining a function $f : R \to R$. This function is almost never a ring homomorphism!

We are going to take a different approach. We choose one particular element $c \in R$ use it to define a function from the ring of polynomials $R[x]$ to the ring $R$. We denote this function $E_c$ and call it the *evalaution at c map*. It is defined exactly as its name suggests,

$$E_c : R[x] \longrightarrow R, \quad E_c(f) = f(c).$$

The evaluation by $c$ map is a ring homomorphism, as you will verify in Exercise 2.7, and its kernel is exactly the set of polynomials that have a factor of $x - c$.

*Example* 2.5 (**Ring of Quarternions** $\mathbb{H}$). We next describe a famous non-commutative ring, called the *ring of quaternions*,[3]

$$\mathbb{H} = \{a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} : a, b, c, d \in \mathbb{R}\}.$$

The quantities $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are three different square roots of $-1$, and although we specify that they commute with elements of $\mathbb{R}$, they do not commute with one another. More precisely, the rule for multipying two quaternions is to use the distributative law to reduce to multiplying pairs of $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, and then applying the following rules:

$$\boldsymbol{i}^2 = -1, \quad \boldsymbol{j}^2 = -1, \quad \boldsymbol{k}^2 = -1, \quad \boldsymbol{i} \cdot \boldsymbol{j} = \boldsymbol{k}, \quad \boldsymbol{j} \cdot \boldsymbol{k} = \boldsymbol{i}, \quad \boldsymbol{k} \cdot \boldsymbol{i} = \boldsymbol{j}.$$

To see that $\mathbb{H}$ is a noncommutative ring, we compute

$$-\boldsymbol{j} \cdot \boldsymbol{i} = \boldsymbol{j} \cdot (-1) \cdot \boldsymbol{i} = \boldsymbol{j} \cdot \boldsymbol{k}^2 \cdot \boldsymbol{i} = (\boldsymbol{j} \cdot \boldsymbol{k}) \cdot (\boldsymbol{k} \cdot \boldsymbol{i}) = \boldsymbol{i} \cdot \boldsymbol{j}.$$

Thus $\boldsymbol{j} \cdot \boldsymbol{i} = -\boldsymbol{i} \cdot \boldsymbol{j}$, and one can similarly check that $\boldsymbol{k} \cdot \boldsymbol{i} = -\boldsymbol{i} \cdot \boldsymbol{k}$ and $\boldsymbol{k} \cdot \boldsymbol{j} = -\boldsymbol{j} \cdot \boldsymbol{k}$.

The ring of quaternions $\mathbb{H}$ played an important role in the development of modern mathematics and physics because it satisfies the so-called *cancelation law*. Thus you know that if $\alpha$ and $\beta$ are real numbers satisfying $\alpha \cdot \beta = 0$, then either $\alpha = 0$ or $\beta = 0$, and similarly when $\alpha$ and $\beta$ are complex numbers. It turns out that the same is true if $\alpha$ and $\beta$ are quaternions! See Exercise 2.14.

*Example* 2.6 (**Matrix Rings**). There are also rings whose elements are matrices. Let

$$M_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\}$$

denote the set of 2-by-2 matrices with real entries. We add matrices by adding their corresponding entries, and we multiply matrices using matrix multiplication.[4] With these operations, $M_2(\mathbb{R})$ is a non-commutative ring. However, it does not satisfy the cancelation law, since

[3]The letter $\mathbb{H}$ to denote the ring of quaternions is in honor of William Hamilton, who first described them in 1843.

[4]See (1.3) in Example 1.10 for the formula for matrix multiplication.

$$\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}0&0\\0&1\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix}$$

shows that the product of two non-zero elements may equal $0$.

There are lots of interesting homomorphisms to matrix rings. For example, the map

$$\mathbb{C}\hookrightarrow M_2(\mathbb{R}),\quad x+yi\longmapsto\begin{pmatrix}x&y\\-y&x\end{pmatrix},$$

is an injective ring homomorphism. You will prove this fact in Exercise 2.8.

## 2.4 Some Important Properties of Rings

Some rings, such as $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, have the special property that every non-zero element has a multiplicative inverse. These types of rings are so important that they have a special name all their own.

**Definition.** A *field* is a commutative ring $R$ with the property that every non-zero element of $R$ has a multiplicative inverse. In other words, for every $a\in R$ with $a\neq 0$ there is a $b\in R$ satisfying $ab=1$.

We will use fields in Chapter 3 as the basic building block for the theory of Vector Spaces, after which we will spend an entire chapter (Chapter 2) studying properties of fields and field extensions.

*Example* 2.7. In addition to the already mentioned fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, for every prime $p$, the ring $\mathbb{Z}/p\mathbb{Z}$ is a field . It is an example of a *finite field*, and is frequently denoted $\mathbb{F}_p$ to emphasize its "fieldiness." The fact that $\mathbb{F}_p$ is a field follows from the Number Theory Unit; or see Exercise 2.16.

We also know lots of rings that are not fields, for example $\mathbb{Z}$, $\mathbb{Z}[i]$, and $R[x]$. But these rings do have the following nice property, which is very useful for solving equations.

**Cancellation Property**: Let $R$ be a commutative ring and suppose that $a,b,c\in R$ with $a\neq 0$. Then

$$ab=ac\quad\Longleftrightarrow\quad b=c.$$

**Definition.** Let $R$ be a ring. An element $a\in R$ is called a *zero divisor* if $a\neq 0$ and there is some $b\in R$ such that $ab=0$. The ring $R$ is an *integral domain* if it has no zero divisors. Equivalently, the ring $R$ is an integral domain if the only way to get $ab=0$ is to have either $a=0$ or $b=0$.

It is easy to check that every field is an integral domain, and that a ring $R$ is an integral domain if and only if it has the cancellation property; see Exercises 2.16 and 2.17. It is also the case that every integral domain is a subring of a field. The smallest such field, which we discuss in Section 2.8, is called the *field of fractions of* $R$.

## 2.5 Ideals and Quotient Rings

Do you recall how we constructed the ring $\mathbb{Z}/m\mathbb{Z}$ of integer modulo $m$ starting from the ring $\mathbb{Z}$? We simply pretended that that two integers $a$ and $b$ are "the same" if their difference $a - b$ is a multiple of $m$. In fancier language, we defined an equivalence relation on $\mathbb{Z}$ by the rule

$$a \text{ is equivalent to } b \text{ if } a - b \text{ is a multiple of } m,$$

and we then defined $\mathbb{Z}/m\mathbb{Z}$ to be the set of equivalence classes. Of course, it takes some work to check that addition and multiplication of equivalence classes makes sense.

Our goal in this section is to generalize this important construction to arbitrary (commutative) rings. The first step is the generalize the concept of being a "multiple of $m$."

**Definition.** Let $R$ be a commutative ring. An *ideal* of $R$ is a non-empty subset $I \subseteq R$ with the following two properties:

- If $a \in I$ and $b \in I$, then $a + b \in I$.
- If $a \in I$ and $r \in R$, then $ra \in I$.

One way to create an ideal is to start with some element of $R$ and take all of its multiples.

**Definition.** Let $R$ be a commutative ring and let $c \in R$. The *principal ideal generated by* $c$, denoted $cR$ or $(c)$, is the set of all multiples of $c$,

$$cR = \{rc : r \in R\}.$$

We let you verify that $cR$ is an ideal; see Exercise 2.19.

In some rings, such as $\mathbb{Z}$ and $\mathbb{Z}[i]$ and $\mathbb{R}[x]$, every ideal is a principal ideal, although it requires real work to prove that these assertions are valid. On the other hand, there are rings such as $\mathbb{Z}[x]$ that have non-prinicipal ideals; see Exercise 2.25.

We now create a quotient ring $R/I$ by identifying pairs of elements of $R$ if their difference is in $I$, just as we did when we defined $\mathbb{Z}/m\mathbb{Z}$. We note that for a given $a \in R$, the set of $b \in R$ that are equivalent to $a$ consists of the set of $b$ such that $b - a \in I$, or equivalently, such that $b$ is in the set that is naturally denoted by $a + I$. This prompts the following definitions.

**Definition.** Let $R$ be a commutative ring, and let $I$ be an ideal of $R$. Then for each element $a \in R$, the *coset of* $a$ is the set

$$a + I = \{a + c : c \in I\}.$$

We note that $a$ is an element of its coset, since $0 \in I$. If $a, b \in R$ satisfy $b - a \in I$, then people sometimes write

$$b \equiv a \pmod{I}$$

and say that "$b$ is congruent to $a$ modulo $I$." Given two cosets $a+I$ and $b+I$, we define their sum and product by the formulas

$$(a+I)+(b+I)=(a+b)+I, \qquad (a+I)\cdot(b+I)=(a\cdot b)+I,$$

and we denote the collection of distinct cosets by $R/I$.

We now check that our definitions of addition and multiplication of cosets makes sense, and that they turn the collection of cosets into a ring.

**Proposition 2.8.** *Let $R$ be a commutative ring, and let $I$ be an ideal of $R$.*

(a) *Let $a+I$ and $a'+I$ be two cosets. Then $a'+I=a+I$ if and only if $a'-a\in I$.*

(b) *Addition and multiplication of cosets is well-defined, in the sense that it doesn't matter which element of the coset we use in the definition.*

(c) *Addition and multiplication of cosets turns $R/I$ into a commutative ring.*

*Proof.* We prove that multiplication is well-defined, and leave the rest of the proof to you; see Exercise 2.21. Let $a,a',b,b'\in R$ be elements whose cosets satisfy $a'+I=a+I$ and $b'+I=b+I$. We need to prove that $ab+I$ is equal to $a'b'+I$.

The assumption that $a+I=a'+I$ means that there is some $c\in I$ such that $a'=a+c$, and similarly the assumption that $b+I=b'+I$ means that there is some $d\in I$ such that $b'=b+d$. It follows that

$$a'b'=(a+c)(b+d)=ab+\underbrace{ad+bc+cd}_{\text{This is in } I,\text{ since } a,\,b\,\in\, I}.$$

Hence $a'b'-ab\in I$, so from (a), the cosets $a'b'+I$ and $ab+I$ are equal. □

Ideals and homomorphisms are closely related, as shown by our next result.

**Proposition 2.9.** *Let $R$ be a commutative ring.*

(a) *Let $I$ be an ideal of $R$. Then the map*

$$R\longrightarrow R/I, \quad a\longmapsto a+R,$$

*that sends an element to its coset is a ring homomorphism whose kernel is $I$.*

(b) *Let $\phi:R\to R'$ be a ring homomorphism.*

(i) *The kernel of $\phi$ is an ideal of $R$.*

(ii) *The homomorphism $\phi$ is injective if and only if $\ker(\phi)=(0)$.*

(iii) *Writing $I_\phi=\ker(\phi)$ for convenience, there is a well-defined injective ring homomorphism*

$$\bar\phi: R/I_\phi\longrightarrow R' \quad \textit{defined by} \quad \bar\phi(a+I_\phi)=\phi(a).$$

*Proof.* We prove (b), and leave (a) as an exercise; see Exercise 2.22. Our first goal is to prove that $\ker(\phi)$ is an ideal. Let $a,b\in\ker(\phi)$. Then

$$\phi(a+b)=\phi(a)+\phi(b)=0+0=0,$$

so $a+b \in \ker(\phi)$. Next let $a \in \ker(\phi)$ and $r \in R$. Then

$$\phi(ra) = \phi(r) \cdot \phi(a) = \phi(r) \cdot 0 = 0,$$

so $ra \in \ker(\phi)$. This completes the proof of (i) that $\ker(\phi)$ is an ideal.

Next suppose that $\ker(\phi) = (0)$, and that $\phi(a) = \phi(b)$ for some $a, b \in R$. Then $\phi(a-b) = 0$, so $a-b \in \ker(\phi)$, and hence $a-b = 0$. This proves the $\phi$ is injective.

Conversely, suppose that $\phi$ is injective, and let $a \in \ker(\phi)$. Then $0 = \phi(a) = \phi(0)$, so the injectivity of $\phi$ implies that $a = 0$. This proves that $\ker(\phi) = (0)$, which completes the proof of (ii).

For (iii), we first want to show that the map $\bar{\phi}$ is well-defined. So suppose that $a' + I_\phi = a + I_\phi$ are two ways of writing the same coset. We need to show that $\phi(a') = \phi(a)$. The assumption that $a' + I_\phi = a + I_\phi$ means that $a' = a + b$ for some $b \in I_\phi$. then

$$\phi(a') = \phi(a+b) = \phi(a) + \phi(b) = \phi(a) + 0 = \phi(a).$$

This shows that $\bar{\phi}$ is well-defined. Next, the fact that $\bar{\phi}$ is a ring homomorphism follows directly from the assumption that $\phi$ is a ring homomorphism. Finally, to see that $\bar{\phi}$ is injective, we observe that

$$\begin{aligned} \bar{\phi}(a' + I_\phi) = \bar{\phi}(a + I_\phi) \quad &\iff \quad \phi(a') = \phi(a) \quad \iff \quad \phi(a'-a) = 0 \\ &\iff \quad a' - a \in I_\phi \quad \iff \quad a' + I_\phi = a + I_\phi. \end{aligned}$$

This completes the proof of Proposition 2.9(b). □

## 2.6 Prime Ideals and Maximal Ideals

You have seen the importance of prime numbers when we studied the Number Theory. Recall that an integer $p$ is prime if its only (positive) divisors are $1$ and $p$. An important property of prime numbers is that if $p$ is prime and $p$ divides a product $ab$, then either $p$ divides $a$ or $p$ divides $b$. We can rephrase this divisibility property using ideals: if a product $ab$ is in the ideal $p\mathbb{Z}$, then either $a \in p\mathbb{Z}$ or $b \in p\mathbb{Z}$. This version is the right way to generalize the notion of primes to arbitrary rings.[5]

**Definition.** Let $R$ be a commutative ring. An ideal $I$ of $R$ is a *prime ideal* if $I \neq R$ and if whenever a product of elements $ab \in I$, then either $a \in I$ or $b \in I$.

We observe that if $I$ is a prime ideal, then it also has the following property:

$$a \notin I \text{ and } b \notin I \quad \Longrightarrow \quad ab \notin I.$$

This statement is the contrapositive of, hence logically equivalent to, the stated definition of prime ideal.

[5]There is also an analogue of the "no non-trivial factors" definition to arbitrary rings. Such elements are called *irreducible*. Unique factorization and irreducibility are discussed in Section 2.7.

*Example* 2.10. Let $m \neq 0$ be an integer. The ideal $m\mathbb{Z}$ is a prime ideal if and only if $|m|$ is a prime number in the usual sense.

*Example* 2.11. Let $F$ be a field. For every $a, b \in F$ with $a \neq 0$, the principal ideal $(ax+b)F[x]$ is a prime ideal. For every $a, b, c \in F$ such that $a \neq 0$ and $b^2 - ac$ is not equal to the square of an element of $F$, the principal ideal $(ax^2+bx+c)F[x]$ is a prime ideal. See Exercise 2.26.

The largest possible ideal in a ring $R$ is the entire ring itself. The ideals that are as large as possible without being all of $R$ play an important role.

**Definition.** Let $R$ be a commutative ring. An ideal $I$ is called a *maximal ideal* if $I \neq R$ and if there are no ideal properly contained between $I$ and $R$. In other words, if $J$ is an ideal and $I \subseteq J \subseteq R$, then either $J = I$ or $J = R$.

*Example* 2.12. Let $p \in \mathbb{Z}$ be a prime number. Then the ideal $p\mathbb{Z}$ is not only a prime ideal, it is also a maximal ideal.

*Example* 2.13. In the ring $\mathbb{Z}[x]$ of polynomials with integer coefficients, the principal ideals $2\mathbb{Z}[x]$ and $x\mathbb{Z}[x]$ are prime ideals, but they are not maximal ideals, since they are contained in non-principal maximal the ideal

$$\big\{2a(x) + xb(x) : a(x), b(x) \in \mathbb{Z}[x]\big\}.$$

See Exercise 2.25.

Just as prime numbers in $\mathbb{Z}$ form the basic building blocks for all numbers, the prime and maximal ideals of a ring $R$ are, in some sense, the basic building blocks underlying the algebraic (and geometric!) structure of $R$. On the other hand, integral domains and fields are two particularly nice kinds of rings. These facts help to explain why the next result is so important.

**Theorem 2.14.** *Let $R$ be a commutative ring, and let $I$ an ideal with $I \neq R$.*
(a) *$I$ is a prime ideal if and only if the quotient ring $R/I$ is an integral domain.*
(b) *$I$ is a maximal ideal if and only if the quotient ring $R/I$ is a field.*

*Proof.* This theorem consists of two if-and-only-if statements, so there are really four statements that need to be proven.

(a) $\boxed{I = \text{Prime Ideal} \quad \Longrightarrow \quad R/I = \text{Integral Domain}}$

Let $a + I$ and $b + I$ be elements of $R/I$ whose product is zero, i.e.,

$$(a + I) \cdot (b + I) = 0 + I.$$

This means that $ab + I = 0 + I$, so $ab \in I$. The assumption that $I$ is a prime ideal tells us that either $a \in I$ or $b \in I$, which means that either $a + I = I$ or $b + I = I$. Thus at least one of $a + i$ or $b + I$ is equal to $0 + I$, which completes the proof that $R/I$ is an integral domain.

(a) $\boxed{R/I = \text{Integral Domain} \quad \Longrightarrow \quad I = \text{Prime Ideal}}$

Suppose that $a, b \in R$ satisfy $ab \in I$. Then

$$(a+I)\cdot(b+I) = ab+I = I,$$

so the product of $a+I$ and $b+I$ is zero in the quotient ring $R/I$. We are assuming that $R/I$ is an integral domain, so we conclude that either $a+I=I$ or $b+I=I$. These in turn imply that either $a\in I$ or $b\in I$, which completes the proof that $I$ is a prime ideal.

(b) $\boxed{I = \text{Maximal Ideal} \implies R/I = \text{Field}}$

Let $a+I$ be a non-zero element of $R/I$, which means that $a\notin I$. We look at the ideal

$$J = \{ar+b : r\in R \text{ and } b\in I\}.$$

(We leave it to you to check that $J$ is an ideal; see Exercise 2.24.) Taking the elements of $J$ with $r=0$ shows that $I\subset J$, while takeing $r=1$ and $b=0$ shows that $a\in J$. We know that $a\notin I$, so $J$ is strictly larger than $I$; in symbols, $I\subsetneq J\subseteq R$. We are assuming that $I$ is a maximal ideal, so by definition this forces $J=R$. In particular, we have $1\in J$. Thus there exists some $c\in R$ and some $b\in I$ such that $1=ac+b$. In terms of elements of $R/I$, using the fact that $b+I=I$, we find

$$1+I = (ab+b)+I = ac+I = (a+I)\cdot(b+I).$$

Hence $a+I$ has a multiplicative inverse in $R/I$, and we've proven that this is true for all non-zero elements of $R/I$, hence $R/I$ is a field.

(b) $\boxed{R/I = \text{Field} \implies I = \text{Maximal Ideal}}$

Let $J$ be an ideal satisfying $I\subseteq J\subseteq R$. If $J=I$, we're done, so we assume that $J\neq I$. This means that we can find some element $a\in R$ with $a\notin I$. Then the coset $a+I\neq I$, so $a+I$ is a non-zero element of the quotient ring $R/I$. We are assuming that $R/I$ is a field, so $a+I$ has a multiplicative inverse, say $c+I$. This means that

$$1+I = (a+I)\cdot(c+I) = ac+I,$$

so there is an element $b\in I$ such that $1=ac+b$. But $a\in J$, so $ac\in J$, while $b\in I\subset J$, and thus the quantity $ac+b$ is in the ideal $J$. This proves that $1\in J$, but then for every $r\in R$ we have $r=r\cdot 1\in J$. Hence $J=R$, which completes the proof that $I$ is a maximal ideal. □

The strength of Theorem 2.14 is illustrated by the slick proof of the following corollary.

**Corollary 2.15.** *Every maximal ideal is a prime ideal.*[6]

*Proof.* It is easy to check that a field is an integral domain; see Exercise 2.15. Then we can apply Theorem 2.14,

$$I \text{ maximal} \implies R/I \text{ field} \implies R/I \text{ integral domain} \implies I \text{ prime}.$$

This completes the proof of the corollary. □

[6]The converse is not true, i.e., there may exist prime ideals that are not maximal; see Example 2.13.

**Mini-Remark 7**. It would be nice to know that every ring has at least one maximal ideal. It turns out that this assertion is yet another statement that is equivalent to the axiom of choice!

## 2.7 Irreducibility and Factorization (*Optional*)

**This Section is Under Construction**

In this section we discuss irreducibility, which is in some ways a more direct analogue of the definition of a prime number in $\mathbb{Z}$. However, the unique factorization property for $\mathbb{Z}$ that you proved in the Number Theory Unit fails for many rings. This is one reason that mathematicians study prime ideals, as we did in Section 2.6. For example there is a very nice type of ring called a Dedekind domain in which every ideal factors uniquely into a product of prime ideals.

**Definition.** Let $R$ be a commutative ring. An element $a \in R$ is called a *unit* if it has a multiplicative inverse. The set of units of $R$ is denoted $R^*$. It forms a group with group law being multiplication; see Exercises 2.12 and 2.13.

We note that if $u \in R^*$ is a unit, then we can factor any element $a \in R$ as $a = u^{-1} \cdot u \cdot a$. This generalizes the fact that in $\mathbb{Z}$, every integer factors as $a = 1 \cdot a$ and as $a = (-1) \cdot (-a)$. These are the only trivial factorizations, reflecting that fact that $\mathbb{Z}^* = \{\pm 1\}$.

**Definition.** Let $R$ be a commutative ring. An element $a \in R$ is said to be *irreducible* if $a$ is not a unit and if the only way to factor $a$ as $a = bc$ is to take either $b$ or $c$ to be a unit.[7]

When you studied unique factorization of integers in the Number Theory Unit, it was necessary to take some care to define just what uniqueness means. For example, the integer $12$ appears to have many "factorizations" as a product of primes,

$$12 = 2 \cdot 2 \cdot 3 = 2 \cdot 3 \cdot 2 = 1 \cdot 2 \cdot 2 \cdot 3 = 1 \cdot 2 \cdot 1 \cdot 2 \cdot 3.$$

And if we allow negative numbers, there are also factorizations such as $(-2) \cdot 2 \cdot (-3)$. But all of these factorizations are intrinsically the same, they are composed of two copies of 2, one copy of 3, and some copies of the units $1$ and $-1$.

In order to discuss unique factorization in more general rings, we need to account for the ambiguities arising from re-ordering the factors and/or including extra unit factors. These difficulties explain the complications in the following definition.

**Definition.** A domain $R$ is called a *unique factorization domain* (UFD) if it has the following two properties:

[7] Note that if $u \in R^*$ is a unit, then we can "factor" any element $a \in R$ as $a = u \cdot (u^{-1}a)$, but this sort of factorization is not very interesting.

(a) Let $a \in R$ be a non-zero element that is not a unit. Then $a$ can be factored in the form
$$a = b_1 \cdot b_2 \cdots b_n$$
using irreducible elements $b_1, b_2, \ldots, b_n \in R$.

(b) Suppose that
$$a = c_1 \cdot c_2 \cdots c_m$$
is some other factorization of $a$ using irreducible elements $c_1, c_2, \ldots, c_m \in R$. Then $m = n$, and after rearranging $c_1, \ldots, c_n$, there are units $u_1, \ldots, u_n \in R^*$ so that[8]
$$c_i = u_i b_i \quad \text{for all } 1 \le i \le n.$$

*Example* 2.16. We already know that $\mathbb{Z}$ is a UFD. Other examples of UFDs include the ring $\mathbb{Z}[i]$ of Gaussian integers (Example 2.3) and polynomial rings $F[x]$ with coefficients in a field $F$ (Example 2.4). These rings actually have stronger properties, namely they are *Principal Ideal Domains*, which means that all of their ideals are principal, and they are *Euclidean Domains*, which means that they have a "division-with-remainder" procedure. It turns out that every Euclidean domain is a principal ideal domain, and that every principal ideal domain is a unique factorization domain, but this is material beyond what we want to discuss in this section. However, they help to illuminate why our next result is interesting, since it gives examples of UFDs that do not have these stronger properties.

**Theorem 2.17.** *Let $F$ be a field, and let $F[x_1, \ldots, x_n]$ be the ring of polyonomials in $n$ variables with coefficients taken from the field $F$. Then $F[x_1, \ldots, x_n]$ is a unique factorization domain.*

*Proof.* Unfortunately, the proof of this important result is beyond the scope of these notes. We can only point out that a key idea is to prove a stronger statement: if $R$ is a UFD, then the polynomial ring $R[x]$ is also a UFD. The stated result then follows by induction, by adding the variables $x_1, \ldots, x_n$ one at a time. The stronger result also implies, for example, that the polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ is a UFD. □

*Example* 2.18. It seems worthwhile to describe a ring that is not a UFD. We consider the ring[9]
$$R = \mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\}.$$
It is a subring of the field of complex numbers. The number $4 \in R$ can be factored in $R$ as
$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$
We claim that $2$, $1 + \sqrt{-3}$, and $1 - \sqrt{-3}$ are irreducible elements of $R$. To see that $2$ is irreducible, we first prove, by contradiction, that $2$ is not a unit in the ring $R$. So suppose that $2$ is a unit. That means we can find some $a + b\sqrt{-3}$ so that

[8] If you prefer to be formal, what we should say is that there is a permutation $\pi \in \mathcal{S}_n$ so that for all $1 \le i \le n$ we have $c_i = u_i b_{\pi(i)}$.

[9] See Exercise 2.5 for some properties of this ring, which we will use in our analysis.

$$2 \cdot (a + b\sqrt{-3}) = 1.$$

This implies that $2a + 2b\sqrt{-3} = 1$, so $2a = 1$ and $b = 0$. But $\frac{1}{2} \notin R$, which proves that 2 is not a unit. We next suppose that 2 factors as

$$2 = (a + b\sqrt{-3})(c + d\sqrt{-3}), \tag{2.1}$$

and our goal is to prove that one of the factors is a unit in $R$. Applying Exercise 2.5(b) to (2.1), we see that

$$2 = (a - b\sqrt{-3})(c - d\sqrt{-3}).$$

Multipying these two equations yields

$$4 = (a^2 + 3b^2)(c^2 + 3d^2),$$

and now we're working with integers. The only ways to factor $4$ as a product of two positive integers are $1 \cdot 4$ and $2 \cdot 2$ and $4 \cdot 1$. But $a^2 + 3b^2$ cannot equal 2, so either

$$a^2 + 3b^2 = 4 \text{ and } c^2 = 3d^2 = 1 \qquad \text{or} \qquad a^2 + 3b^2 = 1 \text{ and } c^2 = 3d^2 = 4.$$

The first case yields $c = \pm 1$ and $d = 0$, while the second case yields $a = \pm 1$ and $b = 0$. Hence one of the factors in our factorization (2.1) of 2 is $\pm 1$, and $\pm 1$ are units. This completes the proof that 2 is irreducible. A similar analysis, which we leave as an exercise (Exercise 2.29), shows that $1 \pm \sqrt{-3}$ are irreducible, and hence that in $R$, the number $4$ has two different factorizations into irreducibles.

## 2.8 Field of Fractions (*Optional*)

**This Section is Under Construction**

Our goal in this section is to prove that every integral domain $R$ is a subring of a field, and that there is a smallest such field $F$. The idea of the proof is to construct $F$ from $R$ just as one constructs $\mathbb{Q}$ from $\mathbb{Z}$.

**Theorem 2.19.** *Let $R$ be an integral domain. There exists a field $F$, called the* field of fractions of *$R$, with the following properties*:

(i) *The ring $R$ is a subring of a field $F$.*

(ii) *If $R$ is also a subring of some other field $K$, then there is a unique injective homomorphism $F \hookrightarrow K$ that takes $R$ to itself by the identity map.*

*Proof.* As noted above, the idea is to construct $F$ from $R$ in the same way that one constructs the field of rational numbers $\mathbb{Q}$ from the integral domain $\mathbb{Z}$. But we need to be careful, since even for $\mathbb{Q}$, it is important to keep in mind that different looking fractions such as $\frac{1}{2}$, $\frac{2}{4}$, and $\frac{23}{46}$ all represent the same rational number. . . . □

# Exercises

## Section 2.2. Abstract Rings and Ring Homomorphisms

**2.1.** Let $R$ be a ring, and let $a, b, \in R$. Prove that

$$(-a) \cdot (-b) = a \cdot b.$$

Be sure to justify each step of your proof by using either a definition or a ring axiom. This is Proposition 2.1(b).

**2.2.** Let $R$ be a ring.

(a) Suppose that the map

$$f : R \longrightarrow R, \quad f(a) = a^2,$$

is a ring homomorphism. Prove that $1_R + 1_R = 0_R$. In less fancy notation, prove that $2 = 0$ in the ring $R$.

(b) Let $p$ be a prime. Suppose that $R$ is a commutative ring in which $p = 0$. Prove that the map

$$f : R \longrightarrow R, \quad f(a) = a^p,$$

is a ring homomorphism. (*Hint*. Use the binomial theorem.)

## Section 2.3. Interesting Examples of Rings

**2.3.** Let $m \geq 1$ be an integer, and define a map

$$\phi : \mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}, \quad \phi(a) = a \bmod m.$$

In other words, the map $\phi$ sends an integer to its congruence class modulo $m$. Prove that $\phi$ is a ring homomorphism.

**2.4.** (a) Let 7 and 11 be elements of the ring $\mathbb{Z}/17\mathbb{Z}$. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

(b) Let $2 + 4x$ and $1 + 4x + 3x^2$ be elements of the polynomial ring $(\mathbb{Z}/7\mathbb{Z})[x]$. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

(c) Let $\alpha = 3 + 2i$ and $\beta = 2 - 3i$ be elements of the ring of Gaussian integers $\mathbb{Z}[i]$. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

(d) Let $\alpha = 3 + 2x - x^2$ and $\beta = 2 - 3x + x^2$ be elements of the polyomial ring $\mathbb{Z}[x]$. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

(e) Let $R = \mathbb{Z}[i]$ be the ring of Gaussian integers, and let $\alpha = (1 + i) + (2 - i)x - x^2$ and $\beta = (2 + i) + (1 + 3i)x$ be elements of the polynomial ring $R[x]$. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

(f) Let $\alpha = 1 + 2\boldsymbol{i} - \boldsymbol{j} + \boldsymbol{k}$ and $\beta = 2 - \boldsymbol{i} + 3\boldsymbol{j} - \boldsymbol{k}$ be elements of the ring $\mathbb{H}$ of quaternions. Compute $\alpha + \beta$ and $\alpha \cdot \beta$.

**2.5.** We have already seen the ring of Gaussian integers $\mathbb{Z}[i]$. More generally, for any integer $D$ that is not the square of an integer,[10] we can form a ring

$$\mathbb{Z}[\sqrt{D}] = \{a + b\sqrt{D} : a, b \in \mathbb{Z}\}.$$

If $D > 0$, then $\mathbb{Z}[\sqrt{D}]$ is a subring of $\mathbb{R}$, while if $D < 0$, then in any case it is a subring of $\mathbb{C}$.

[10] We rule out the case the $D$ is a square, because if $D = d^2$, then $\mathbb{Z}[\sqrt{D}] = \mathbb{Z}[d] = \mathbb{Z}$, so we don't get an interesting new ring.

(a) Let $\alpha = 2 + 3\sqrt{5}$ and $\beta = 1 - 2\sqrt{5}$ be elements of $\mathbb{Z}[\sqrt{5}]$. Compute the quantities

$$\alpha + \beta, \quad \alpha \cdot \beta, \quad \alpha^2.$$

(b) Prove that the map

$$\phi : \mathbb{Z}[\sqrt{D}] \longrightarrow \mathbb{Z}[\sqrt{D}], \quad \phi(a + b\sqrt{D}) = a - b\sqrt{D}$$

is a ring homomorphism. (For notational convenience, for $\alpha = a + b\sqrt{D} \in \mathbb{Z}[\sqrt{D}]$, people often write $\overline{\alpha} = a - b\sqrt{D}$, similar to the notation for complex conjugation.)

(c) With notation as in (b), prove that

$$\alpha \cdot \overline{\alpha} \in \mathbb{Z} \quad \text{for every } \alpha \in \mathbb{Z}[\sqrt{D}].$$

**2.6.** Let $\rho$ be the complex number $\rho = \frac{-1+i\sqrt{3}}{2} \in \mathbb{C}$, and let

$$\mathbb{Z}[\rho] = \{a + b\rho : a, b, \in \mathbb{Z}\}.$$

(a) Prove that $\mathbb{Z}[\rho]$ is a subring of $\mathbb{C}$. (The key here is to prove that if you add or multiply two elements of $\mathbb{Z}[\rho]$, you get back an element of $\mathbb{Z}[\rho]$.)

(b) Prove that $\rho^3 = 1$. Thus $\rho$ is a cube root of unity.

(c) Prove that the polynomial $X^3 - 1$ factors as

$$X^3 - 1 = (X - 1)(X - \rho)(X - \rho^2).$$

**2.7.** Let $R$ be a commutative ring, let $c \in R$, and let $E_c : R[x] \to R$ be the evaluation map $E_c(f) = f(c)$.

(a) Prove that $E_c$ is a ring homomorphism.

(b) Prove that $E_c(f) = 0$ if and only if there is a polynomial $g(x) \in R[x]$ satisfying $f(x) = (x - c)g(x)$, i.e., prove that $\ker(E_c)$ is the set of multiples of $x - c$.

**2.8.** Prove that the map

$$\mathbb{C} \hookrightarrow M_2(\mathbb{R}), \quad x + yi \longmapsto \begin{pmatrix} x & y \\ -y & x \end{pmatrix},$$

as discussed in Example 2.6, is an injective ring homomorphism.

**2.9.** For any ring $R$, let

$$M_2(R) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in R \right\}$$

be the set of 2-by-2 matrices with entries in $R$. Define addition by adding the corresponding entries, and define multiplication as described by (1.3) in Example 1.10.

(a) Prove that $M_2(R)$ is a ring.

(b) Prove that $M_2(R)$ is non-commutative.

(c) Find non-zero elements $A, B \in M_2(R)$ such that $AB = 0$. (In the terminology of Section 2.4, the elements $A$ and $B$ are zero divisors, and $M_2(R)$ is not an integral domain.)

(d) Prove that a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has a multiplicative inverse if and only if $ad - bc$ has a multiplicative inverse in $R$.

(e) For those who have taken a class in linear algebra, generalize (a), (b), and (c) to $M_n(R)$, the set of $n$-by-$n$ matrices with entries in $R$.

**2.10.** Let $R$ be a commutative ring. We consider the ring of polynomials in two variables[11] with coefficients in $R$,

$$R[x,y] = \{a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \cdots : a_{ij} \in R\}.$$

In other words, an element of $R[x,y]$ is a sum of the form

$$f(x,y) = \sum_{k=0}^{n} \sum_{i=0}^{k} a_{i,k-i} x^i y^{k-i}.$$

(a) Let $f(x,y) = 3 + 2x - y + x^2 + xy$ and $g(x,y) = 1 - x + 3y - xy + 2y^2$ be elements of the ring $\mathbb{Z}[x,y]$. Compute $f+g$ and $f \cdot g$.
(b) Same question as in (a), except suppose that $f$ and $g$ are in the ring $(\mathbb{Z}/4\mathbb{Z})[x,y]$.
(c) For $b, c \in R$, define an evaluation map

$$E_{b,c} : R[x,y] \longrightarrow R, \quad E_{b,c}\big(f(x,y)\big) = f(b,c).$$

Prove that $E_{b,c}$ is a ring homomorphism.

**2.11.** Let $R$ be a commutative ring, and let $f(x) =\in R[x]$ be a polynomial with coefficients in $R$. We define the *formal derivative* $f'(x)$ *of* $f(x)$ by writing $f(x)$ as

$$f(x) = \sum_{k=0}^{n} a_k x^k \quad \text{and setting} \quad f'(x) = \sum_{k=0}^{n} k a_k x^{k-1}.$$

Note that there is no limit being taken, so the formal derivatie makes sense even if, for example, $R$ is a ring such as $\mathbb{Z}/m\mathbb{Z}$. It also means that when doing this exericse, you'll need to directly use the definition of $f'(x)$, since you can't rely on the proofs from calculus.
(a) Let $f(x), g(x) \in R[x]$. Prove that $(f+g)'(x) = f'(x) + g'(x)$.
(b) Let $f(x), g(x) \in R[x]$. Prove that $(f \cdot g)'(x) = f(x)g'(x) + g(x)f'(x)$.
(c) Let $f(x), g(x) \in R[x]$. Prove that the formal derivative of $f\big(g(x)\big)$ is $f'\big(g(x)\big)g'(x)$.

**2.12.** Let $R$ be a commutative ring. The *group of units of* $R$ is the subset $R^*$ of $R$ defined by

$$R^* = \{a \in R : \text{there is some } b \in R \text{ satisfying } ab = 1\}.$$

Prove that $R^*$ is a group, where we use multiplication for the group law.

**2.13.** For a commutative ring $R$, we let $R^*$ denote the group of units of $R$ as defined in Exercise 2.12.
(a) Prove that $\mathbb{Z}^* = \{-1, 1\}$.
(b) Prove that $\mathbb{Q}^* = \{a \in \mathbb{Q} : a \neq 0\}$.
(c) Prove that $\mathbb{Z}[i]^* = \{-1, 1, i, -i\}$.
(d) Consider the ring $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$. Prove that $1 + \sqrt{2} \in \mathbb{Z}[\sqrt{2}]^*$. Prove that the powers of $1 + \sqrt{2}$, that is, the numbers $(1 + \sqrt{2})^n$ for $n = 1, 2, 3, \ldots$, are all different, and use that fact to deduce that $\mathbb{Z}[\sqrt{2}]^*$ has infinitely many elements.
(e) Prove that $\mathbb{R}[x]^* = \mathbb{R}^*$, i.e., the only polynomials in $\mathbb{R}[x]$ that have multiplicative inverses are the non-zero constants.

[11] We leave it to you to generalize to polynomials in more variables, if you want.

(f) Prove that $1 + 2x$ is a unit in the ring $(\mathbb{Z}/4\mathbb{Z})[x]$ of polynomials with coefficients in the ring $\mathbb{Z}/4\mathbb{Z}$. (*Challenge*: Describe the complete unit group $(\mathbb{Z}/4\mathbb{Z})[x]^*$.)

**2.14.** For a quaternion $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \in \mathbb{H}$, we let $\overline{\alpha} = a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}$.
(a) Prove that $\alpha\overline{\alpha} \in \mathbb{R}$.
(b) Prove that $\alpha\overline{\alpha} = 0$ if and only if $\alpha = 0$.
(c) Suppose that $\alpha, \beta \in \mathbb{H}$ and that $\alpha\beta = 0$. Prove that either $\alpha = 0$ or $\beta = 0$.

## Section 2.4. Some Important Properties of Rings

**2.15.** Let $R$ be a field. Prove that $R$ is an integral domain, i.e., prove that $R$ does not have any zero divisors.

**2.16.** Let $m$ be a positive integer.
(a) Prove that $\mathbb{Z}/m\mathbb{Z}$ is an integral domain if and only if $m$ is prime.
(b) Prove that $\mathbb{Z}/m\mathbb{Z}$ is a field if and only if $m$ is prime.

**2.17.** Let $R$ be a ring. Prove that $R$ is an integral domain if and only if it has the cancellation property.

**2.18.** Let $R$ be a commutative ring.
(a) Prove that there is exactly one integral domain $R$ such that the map

$$f : R \longrightarrow R, \quad f(a) = a^6,$$

is a ring homomorphism. (You'll need to use the fact that $1_R \neq 0_R$.)
(b) Find all integral domains $R$ such that the map

$$f : R \longrightarrow R, \quad f(a) = a^{15},$$

is a ring homomorphism.
(c) For each of parts (a) and (b), find at least one ring that is not an integral domain for which the indicated map is a ring homomorphism.
(d) Let $p$ and $q$ be distinct primes. Characterize all integral domains $R$ for which the map $f(a) = a^{pq}$ is a ring homomorphism.

## Section 2.5. Ideals and Quotient Rings

**2.19.** Let $R$ be a commutative ring and let $c \in R$. Prove that $cR = \{rc : r \in R\}$ is an ideal of $R$.

**2.20.** Let $R$ be a commutative ring. Prove that $R$ is a field if and only if its only ideals are the zero ideal $(0)$ and the entire ring $R$.

**2.21.** Prove the remaining parts of Proposition 2.8. Let $R$ be a commutative ring, and let $I$ be an ideal of $R$.
(a) Let $a + I$ and $a' + I$ be two cosets. Prove that $a + I = a' + I$ if and only if $a - a' \in I$.
(b) Prove that addition of cosets is well-defined. definition.
(c) Prove that addition and multiplication of cosets turns $R/I$ into a commutative ring.

**2.22.** Let $R$ be a commutative ring, and let $I$ be an ideal of $R$. Prove that the map

$$R \longrightarrow R/I, \quad a \longmapsto a + R$$

that sends an element to its coset is a ring homomorphism whose kernel is $I$. This is Proposition 2.9(a).

**2.23.** Let $I$ be the principal ideal of $\mathbb{R}[x]$ generated by the polynomial $x^2+1$. Prove that the map

$$\phi : \mathbb{R}[x]/I \longrightarrow \mathbb{C}, \quad \phi\big(f(x)+I\big) = f(i),$$

is a well-defined isomorphism, where $i = \sqrt{-1}$ as usual. This shows how one can use ring theory to abstractly construct the complex numbers from the real numbers. We will super-generalize this example in Section 4.6. (*Hint*. One way to do this exercise is to write out all the grubby details, but it is easier to apply Proposition 2.9 to the evaluation homomorphism $E_i : \mathbb{R}[x] \to \mathbb{C}$.)

**2.24.** Let $R$ be a commutative ring and let $I$ and $J$ be ideals of $R$.

(a) Prove that the *ideal sum*

$$I + J = \{a+b : a \in I \text{ and } b \in J\}$$

is an ideal of $R$.

(b) Give an example to show that the set of products $\{ab : a \in I \text{ and } b \in J\}$ need not be an ideal. (*Hint*. If $I$ or $J$ is a principal ideal, then this set will be an ideal, so you'll need to use some non-principal ideals.)

(c) The *ideal product* of two ideals is defined to be

$$IJ = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n : n \ge 1 \text{ and } a_1,\ldots,a_n \in I \text{ and } b_1,\ldots,b_n \in J\}.$$

Prove that $IJ$ is an ideal of $R$.

## Section 2.6. Prime Ideals and Maximal Ideals

**2.25.** Let $I$ be the following subset of the ring $\mathbb{Z}[x]$ of polynomials having integer coefficients:

$$I = \big\{2a(x) + xb(x) : a(x), b(x) \in \mathbb{Z}[x]\big\}.$$

(a) Prove that $I$ is an ideal of $\mathbb{Z}[x]$.

(b) Prove that $I \ne \mathbb{Z}[x]$.

(c) Prove that $I$ is not a principal ideal, i.e., prove that there does not exist a polynomial $c(x) \in \mathbb{Z}[x]$ such that $I = c(x)\mathbb{Z}[x]$.

(d) Prove that $I$ is a maximal ideal of $\mathbb{Z}[x]$.

**2.26.** (a) Let $m \ne 0$ be an integer. Prove that the ideal $m\mathbb{Z}$ is a prime ideal if and only if $|m|$ is a prime number in the usual sense of primes in $\mathbb{Z}$.

(b) Let $F$ be a field, and let $a, b \in F$ with $a \ne 0$. Prove that the principal ideal $(ax+b)F[x]$ is a prime ideal of the polynomial ring $F[x]$.

(c) Again let $F$ be a field, and let $a, b, c \in F$ be elements with $a \ne 0$ and $b^2 - ac$ not equal to the square of an element of $F$. Prove that the principal ideal $(ax^2+bx+c)F[x]$ is a prime ideal of the polynomial ring $F[x]$.

**2.27.** Let $R$ be a ring, let $b, c \in R$, and let $E_{b,c} : R[x,y] \to R$ be the evaluation homomorphism described in Exericse 2.10.

(a) If $R$ is an integral domain, prove that $\ker(E_{b,c})$ is a prime ideal of of $R[x,y]$.

(b) If $R$ is a field, prove that $\ker(E_{b,c})$ is a maximal ideal of of $R[x,y]$.

(*Hint*. Use Proposition 2.9 and Theorem 2.14.)

## Section 2.7. Irreducibility and Factorization (*Optional*)

**2.28.** Let $R$ be a ring, and let $a, b \in R$.

(a) Assume that $R$ is a domain. Prove that the principal ideals $aR$ and $bR$ are equal if and only if there is a unit $u \in R^*$ satisfying $b = au$.

(b) Is (a) true for rings that are not domains? Either prove that it is true, or give an example of a ring for which it is not true.

**2.29.** Let $R = \mathbb{Z}[\sqrt{-3}]$ be the ring that we studied in Example 2.18.

(a) Prove that $\pm 1$ are the only units in $R$.

(b) Prove that $1 + \sqrt{-3}$ and $1 - \sqrt{-3}$ are irreducible elements of $R$. (*Hint*. After you prove that $1 + \sqrt{-3}$ is irreducible, you can use Exercise 2.5(b) to easily prove that $1 - \sqrt{-3}$ is also irreducible.)

## Section 2.8. Field of Fractions (*Optional*)

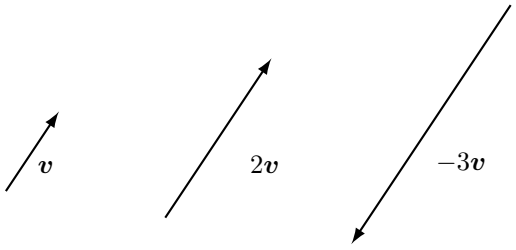
# Chapter 3

# Vector Spaces

## 3.1 Introduction to Vector Spaces

You have probably already studied vectors in the plane, possibly in the guise of arrows and possibly as pairs of real numbers, where $(a, b)$ denotes the arrow with tail at $(0, 0)$ and tip at $(a, b)$.

What sorts of operations can we perform on vectors. One thing that we can do is multiply a vector $\boldsymbol{v}$,[1] by a number $c$. For example, the vector $2\boldsymbol{v}$ points in the same direction as $\boldsymbol{v}$ and is twice as long, while $-3\boldsymbol{v}$ points in the opposite direction to $\boldsymbol{v}$ and is three times as long:



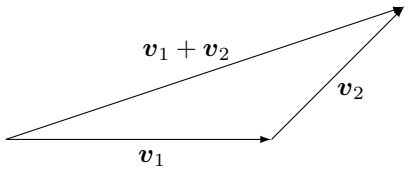

Writing $\boldsymbol{v} = (a, b)$ using coordinates, multipying $\boldsymbol{v}$ by the number $c$ is described by the simple algebraic formula

$$c\boldsymbol{v} = (ca, cb). \tag{3.1}$$

You may also have learned how to add vectors by putting them tip-to-tail and drawing the third side of the triangle, as in the following picture:

[1] In this book, we denote vectors using boldface type. You can't do this in handwritten homework, so instead you should put a little arrow over the letter, thus $\vec{v}$.




In terms of coordinates, addition of two vectors in the plane is given by the formula

$$\boldsymbol{v}_1 + \boldsymbol{v}_2 = (a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2). \tag{3.2}$$

Thinking abstractly, we observe that vectors allow two sorts of operations:

- We can multiply a vector by a number. This is called *scalar multiplication*.
- We can add two vectors. This is called *vector addition*.

And these operations are related by various sorts of formulas, for example there is a distributive law $c(\boldsymbol{v}_1 + \boldsymbol{v}_2) = c\boldsymbol{v}_1 + c\boldsymbol{v}_2$.

So now it's time to take the space of vectors in the plane and "jazz it up"[2] into a general mathematical construction, similar to the way that we started with $\mathbb{Z}$ and ended up with Rings.

## 3.2 Vector Spaces and Linear Transformations

The numbers (scalars) that we used for the "arrow-vectors" in Section 3.1 are real numbers. If we instead look at vectors that are pairs $(a, b)$ of numbers, and if we add and multiply by $c$ them using the rules (3.1) and (3.2), then we could take $a, b, c$ to be some other sort number. For example, we could take $a, b, c$ to be complex numbers. More generally, we can use any sort of "numbers" that allow us to add, subtract, multiply, and divide. As we discussed in Section 2.4, these algebraic objects are called *fields*. They are commutative rings, but they have the special property that every (non-zero) element has a multiplicative inverse.

**Definition.** A *field* is a commutative ring $F$ with the property that for every $a \in F$ with $a \neq 0$ there is a $b \in F$ satisfying $ab = 1$.

*Example* 3.1. You are already familiar with lots of fields, including $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, known respectively as the fields of rational numbers, real numbers, and complex numbers. For every prime $p$, the ring of integers modulo $p$ is a finite field, denoted $\mathbb{F}_p$ or $\mathbb{Z}/p\mathbb{Z}$. In Chapter 4 we will see lots of other fields.

In this chapter we fix a field and use it as a basic building block in the definition of a vector space. Subsequently, in Chapter 4, we will use vector spaces as a fundamental tool to study fields and field extensions.

[2] As you know by now, the formal mathematical term for "jazz it up" is "axiomatize".

**Definition.** Let $F$ be a field. A *vector space with field of scalars* $F$, or alternatively an *$F$-vector space*, is an abelian group $V$ with a rule for multiplying a a vector $\boldsymbol{v} \in V$ by a scalar $c \in F$ to obtain a new vector $c\boldsymbol{v} \in V$. Vector addition and scalar multiplication are required to satisfy the following axioms:

(1) [Identity Law]

$$1\boldsymbol{v} = \boldsymbol{v} \text{ for all } \boldsymbol{v} \in V.$$

(2) [Distributive Law #1]

$$c(\boldsymbol{v}_1 + \boldsymbol{v}_2) = c\boldsymbol{v}_1 + c\boldsymbol{v}_2 \text{ for all } \boldsymbol{v}_1, \boldsymbol{v}_2 \in V \text{ and all } c \in F.$$

(3) [Distributive Law #2]

$$(c_1 + c_2)\boldsymbol{v} = c_1\boldsymbol{v} + c_2\boldsymbol{v} \text{ for all } \boldsymbol{v} \in V \text{ and all } c_1, c_2 \in F.$$

The identity element of $V$ is called the *zero vector* and is denoted by $\mathbf{0}$. It should not be confused with $0 \in F$, which is the zero element of the field $F$.

Just as with the axiomatic definitions of groups and rings, there are many basic facts about vector spaces that can be proven directly from the definitions. We list a couple of them here, but leave the proofs for you as an exercise.

**Proposition 3.2.** *Let $V$ be an $F$-vector space.*

(a) $0\boldsymbol{v} = \mathbf{0}$ *for all* $\boldsymbol{v} \in V$.

(b) $(-1)\boldsymbol{v} + \boldsymbol{v} = \mathbf{0}$ *for all* $\boldsymbol{v} \in V$.

*Proof.* See Exercise 3.1. □

The operations that characterize a vector space are vector addition and scalar multiplication, so we focus on maps between vector spaces that respect these operations.

**Definition.** Let $F$ be a field, and let $V$ and $W$ be $F$-vector spaces. A *linear transformation* from $V$ to $W$ is a function

$$L : V \longrightarrow W$$

satisfying:

$$L(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1L(\boldsymbol{v}_1) + c_2L(\boldsymbol{v}_2) \quad \text{for all } \boldsymbol{v}_1, \boldsymbol{v}_2 \in V \text{ and all } c_1, c_2 \in F.$$

In the next section we will provide many examples of vector spaces and linear transformations.

## 3.3 Interesting Examples of Vector Spaces

We already discussed how to define vectors in the plane as pairs of real numbers $(a, b)$, and similarly vectors in 3-dimensional space where we live[3] can be specified by triples $(a, b, c)$ of real numbers. More generally, we can form a vector space using the set of $n$-tuples with coordinates in any field.

[3]Or maybe not, since according to modern physics, we may be living in a 26-dimensional space, but most of the dimensions are packed up so tightely that we can't see them!

*Example* 3.3. Let $F$ be a field, and let $n \ge 1$ be an integer. Then $F^n$ is the $F$-vector space whose vectors are $n$-tuples of elements of $F$,

$$F^n = \big\{(a_1, a_2, \dots, a_n) : a_1, \dots, a_n \in F\big\}.$$

Vector addition and scalar multiplication are done coordinate-by-coordinate,

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n), \tag{3.3}$$

$$c(a_1, a_2, \dots, a_n) = (ca_1, ca_2, \dots, ca_n). \tag{3.4}$$

We leave it to you to check the vector space axioms; see Exercise 3.6. Among the interesting special cases of this construction are $\mathbb{R}^n$, $\mathbb{C}^n$, and $\mathbb{F}_p^n$. Notice that the vector space $\mathbb{F}_p^n$ is finite; it contains exactly $p^n$ different vectors.

*Example* 3.4. The maps

$$\begin{aligned} L(a_1, a_2) &= (3a_1 - 5a_2, 2a_1 + 3a_2), \\ L'(a_1, a_2, a_3) &= (3a_1 - 5a_2 + 2a_3, 2a_1 + 3a_2 - 7a_3), \end{aligned}$$

are, respectively, linear transformations $L : \mathbb{R}^2 \to \mathbb{R}^2$ and $L' : \mathbb{R}^3 \to \mathbb{R}^2$. See Theorem 3.17 for a super-generalization of these examples.

*Example* 3.5. The set of polynomials $F[x]$ with coefficients in a field $F$ is an $F$-vector space, where we add polynomials and multiply polynomials by scalars in the usual way. For any $c \in F$, the evaluation map

$$E_c : F[x] \longrightarrow F, \quad E_c\big(f(x)\big) = f(c),$$

is a linear transformation. More generally, for any list of values $c_1, \dots, c_n \in F$, we can define a linear transformation by the formula

$$E_{\boldsymbol{c}} : F[x] \longrightarrow F^n, \quad E_{\boldsymbol{c}}\big(f(x)\big) = \big(f(c_1), \cdots, f(c_n)\big).$$

*Example* 3.6. This example is for students who have taken calculus. Let

$$\begin{aligned} V &= \{\text{functions } f : \mathbb{R} \to \mathbb{R}\}, \\ V^{\text{cont}} &= \{\text{continuous functions } f : \mathbb{R} \to \mathbb{R}\}, \\ V^{\text{diff}} &= \{\text{differentiable functions } f : \mathbb{R} \to \mathbb{R}\}. \end{aligned}$$

These are $\mathbb{R}$-vector spaces, where we add functions and multiply them by scalars as usual,

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (cf)(x) = cf(x).$$

Differentiation is a linear transformation

$$D : V^{\text{diff}} \longrightarrow V, \quad D\big(f(x)\big) = f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Similarly, for any $a \in \mathbb{R}$, integration is a linear transformation

$$I_a : V^{\text{cont}} \longrightarrow V^{\text{diff}}, \quad I_a\big(f(x)\big) = \int_a^x f(t)\, dt.$$

The Fundamental Theorem of Calculus is then summarized by the two formulas

$$I_a \circ D\big(f(x)\big) = f(x) - f(a) \quad \text{and} \quad D \circ I_a\big(f(x)\big) = f(x).$$

## 3.4 Bases and Dimension

It is very convenient that every vector in $\mathbb{R}^2$ can be uniquely expressed using the two vectors $\boldsymbol{e}_1 = (1,0)$ and $\boldsymbol{e}_2 = (0,1)$. More precisely, a vector $\boldsymbol{v} = (a,b) \in \mathbb{R}^2$ can be written as

$$(a,b) = a(1,0) + b(0,1) = a\boldsymbol{e}_1 + b\boldsymbol{e}_2,$$

and the coefficients $a$ and $b$ uniquely identify the vector $\boldsymbol{v}$. A similar construction works for $\mathbb{R}^n$, and indeed for $F^n$, where $F$ is any field. As we've done many times before, we now axiomitize the idea of a collection of vectors in $V$ that can be used to uniquely express every vector $V$.

**Definition.** Let $V$ be an $F$-vector space. A *finite*[4] *basis for* $V$ is a finite set of vectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$ with the following property:

> Every vector $\boldsymbol{v} \in V$ can be written in the form
>
> $$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n$$
>
> for exactly one choice of scalars $a_1, \ldots, a_n \in F$.

An expression of the form $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$ is called a *linear combination of* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

*Example* 3.7. Let $F$ be a field.The *standard basis*[5] for $F^n$ is the collection of vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$, where

$$\boldsymbol{e}_k = (0,0,\ldots,0,\underset{\substack{\uparrow\\ k\text{'th coordinate}}}{1},0,\ldots,0,0).$$

We note that a vector $\boldsymbol{v} = (a_1, \ldots, a_n) \in F^n$ can be written as a sum

$$\boldsymbol{v} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + \cdots + a_n\boldsymbol{e}_n,$$

and that the coefficients of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are uniquely determined by $\boldsymbol{v}$. The coefficients $a_1, \ldots, a_n$ are called the *coordinates of* $\boldsymbol{v}$ *for the standard basis of* $F^n$.

*Example* 3.8. Consider the three vectors $\boldsymbol{v}_1 = (1,0)$, $\boldsymbol{v}_2 = (0,1)$, and $\boldsymbol{v}_3 = (1,1)$ in $\mathbb{R}^2$. Every vector in $\mathbb{R}^2$ is a linear combination of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$, but in many different ways, so $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is not a basis for $\mathbb{R}^2$. For example we can write the vector $\boldsymbol{v} = (5,3)$ as

$$\boldsymbol{v} = 5\boldsymbol{v}_1 + 3\boldsymbol{v}_2 = 3\boldsymbol{v}_1 + \boldsymbol{v}_2 + 2\boldsymbol{v}_3 = 7\boldsymbol{v}_1 + 5\boldsymbol{v}_2 - 2\boldsymbol{v}_3.$$

How can we tell if a given set of vectors is a basis? This is answered by a proposition that depends on two important concepts.

[4]Not every vector space has a finite basis. See Remark 3.15 for a discussion of infinite bases and Exercise 3.13 for an example of an infinite dimensional vector space.

[5]As the terminology suggests, we will see shortly that there are many other bases for $F^n$.

**Definition.** Let $V$ be an $F$-vector space, and let $\mathcal{A} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a set of vectors in $V$.

(1) The set $\mathcal{A}$ *spans* $V$ if every vector in $V$ is a linear combination of the vectors in $\mathcal{A}$, i.e., if for every vector $\boldsymbol{v} \in V$ we can find scalars $a_1, \ldots, a_n \in F$ so that

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n.$$

(2) The set $\mathcal{A}$ is *linearly indpeendent* if the only scalars $a_1, \ldots, a_n \in F$ that make

$$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n = \mathbf{0} \quad \text{are} \quad a_1 = a_2 = \cdots = a_n = 0.$$

**Proposition 3.9.** *Let $V$ be an $F$-vector space, and let $\mathcal{A} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a set of vectors in $V$. Then $\mathcal{A}$ is a basis for $V$ if and only if $\mathcal{A}$ both spans $V$ and is linearly independent.*

*Proof.* Suppose that $\mathcal{A}$ is a basis. If we write

$$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n = \mathbf{0} = 0 \cdot \boldsymbol{v}_1 + 0 \cdot \boldsymbol{v}_2 + \cdots + 0 \cdot \boldsymbol{v}_n$$

for some scalars $a_1, \ldots, a_n$, then the uniqueness of the coefficients tells us that $a_1 = a_2 = \cdots = a_n$. Hence $\mathcal{A}$ is linearly independent.

Suppose now that $\mathcal{A}$ spans and is linearly independent. The fact that $\mathcal{A}$ spans means that every vector $\boldsymbol{v} \in V$ is a linear combination of the vectors in $\mathcal{A}$, so it remains to prove that the coefficients are uniquely determined by $\boldsymbol{v}$. So suppose that we have

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_n\boldsymbol{v}_n \quad \text{and} \quad \boldsymbol{v} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_n\boldsymbol{v}_n.$$

Then

$$\mathbf{0} = \boldsymbol{v} - \boldsymbol{v} = (a_1 - b_1)\boldsymbol{v}_1 + (a_2 - b_2)\boldsymbol{v}_2 + \cdots + (a_n - b_n)\boldsymbol{v}_n.$$

The definition of linear independence tells us that every scalar coefficient is $0$, so $a_i = b_i$ for every $1 \le i \le n$. □

The next theorem tells us that if we are given a finite spanning set, then we can find a subset that's a basis.

**Theorem 3.10.** *Let $V$ be an $F$-vector space, and let $\mathcal{A} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a finite set of vectors in $V$ that spans $V$. Then there is a subset of $\mathcal{A}$ that is a basis for $V$.*

*Proof.* Among the subsets of $\mathcal{A}$ that are spanning sets, we take one containing the smallest number of vectors. Relabeling the vectors in $\mathcal{A}$, we may assume that this "smallest" subset, which we denote by $\mathcal{B}$, is $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. Thus $\mathcal{B}$ is a spanning set with the property that no subset of $\mathcal{B}$ is a spanning set. We claim that $\mathcal{B}$ is a basis.

In view of Proposition 3.9, it suffices to prove that $\mathcal{B}$ is a linearly independent set. So we suppose that $\mathcal{B}$ is not linearly independent and deduce a contradiction. Our assumption means that we can find scalars $a_1, \ldots, a_n \in F$, not all $0$, such that

$$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_m\boldsymbol{v}_m = \boldsymbol{0}. \tag{3.5}$$

Again relabeling the vectors if necessary, we may assume that $a_1 \neq 0$. Dividing (3.5) by $a_1$, and to make notation easier letting $b_i = -a_i/a_1$, we have

$$\boldsymbol{v}_1 = b_2\boldsymbol{v}_2 + \cdots + b_m\boldsymbol{v}_m. \tag{3.6}$$

We claim that $\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is a spanning set, which will contradict the choice of $\mathcal{B}$ as having the smallest number of elements of any spanning subset of $\mathcal{A}$.

To check that $\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is a spanning set, we start with an arbitrary vector $\boldsymbol{v} \in V$ . The fact that $\mathcal{B}$ is a spanning set means that we can find scalars $c_1, \ldots, c_m$ so that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m. \tag{3.7}$$

Then

$$\begin{aligned}
\boldsymbol{v} &= c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m && \text{from (3.7),}\\
&= c_1(b_2\boldsymbol{v}_2 + \cdots + b_m\boldsymbol{v}_m) + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m && \text{from (3.6),}\\
&= (c_1b_2 + c_2)\boldsymbol{v}_2 + \cdots + (c_1b_m + c_m)\boldsymbol{v}_m.
\end{aligned}$$

This shows that $\boldsymbol{v}$ is a linear combination of $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$, and hence $\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is a spanning set. This contradiction concludes the proof of Theorem 3.10. □

The next result is of fundamental importance in studying vector spaces and for applying the theory of vector spaces to other areas of mathematics.

**Theorem 3.11.** *Let $V$ be a vector space that has a finite basis. Then every basis for $V$ has exactly the same number of elements.*

*Proof.* Let

$$\mathcal{A} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \quad \text{and} \quad \mathcal{B} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$$

be two bases for $V$, where by switching $\mathcal{A}$ and $\mathcal{B}$ if necessary, we may assume that $m \leq n$. The idea of the proof is to show that we can replace $m$ carefully chosen vectors in $\mathcal{A}$ with the vectors in $\mathcal{B}$ and still have a basis.

We want to do this one vector at a time, so what we will do is prove that the following statement holds for all integers $k$ between 0 and $m$. The proof will be by induction on $k$.

> **Statement($k$)**: It is possible to relabel the vectors in $\mathcal{A}$ so that the set of vectors
>
> $$\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k, \boldsymbol{v}_{k+1}, \boldsymbol{v}_{k+2}, \ldots, \boldsymbol{v}_n\} \quad \text{is a basis for } V.$$

We start by observing that **Statement(0)** simply says that $\mathcal{A}$ is a basis, which we know. So that gets the induction started.

We next assume that **Statement($k$)** is true for some $k$ satisfying $0 \leq k < m$, and our goal is to prove that **Statement($k+1$)** is true. The assumption that

**Statement($k$)** is true says that possibly after relabeling the vecors in $\mathcal{A}$, the set of vectors

$$\mathcal{C} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k, \boldsymbol{v}_{k+1}, \boldsymbol{v}_{k+2}, \ldots, \boldsymbol{v}_n\}$$

is a baiss for $V$. In particular, we can write $\boldsymbol{w}_{k+1}$ as a linear combination of the vectors in $\mathcal{C}$, say

$$\boldsymbol{w}_{k+1} = a_1\boldsymbol{w}_1 + \cdots + a_k\boldsymbol{w}_k + a_{k+1}\boldsymbol{v}_{k+1} + \cdots + a_n\boldsymbol{v}_n. \tag{3.8}$$

We note that it is not possible for all of the coefficients $a_{k+1}, \ldots, a_n$ to vanish, since if they did, then we would have

$$a_1\boldsymbol{w}_1 + \cdots + a_k\boldsymbol{w}_k - \boldsymbol{w}_{k+1} = 0,$$

contradicting the fact that $\mathcal{B}$ is a linearly independent set. Hence at least one of $a_{k+1}, \ldots, a_n$ is non-zero, and relabeling the vectors in $\mathcal{A}$, we may assume that $a_{k+1} \neq 0$. This allows us to divide (3.8) by $a_{k+1} \neq 0$ and solve for $\boldsymbol{v}_{k+1}$, thus

$$\boldsymbol{v}_{k+1} = -\frac{a_1}{a_{k+1}}\boldsymbol{w}_1 - \cdots - \frac{a_k}{a_{k+1}}\boldsymbol{w}_k - \boldsymbol{w}_{k+1} - \frac{a_{k+2}}{a_{k+1}}\boldsymbol{v}_{k+2} - \cdots - \frac{a_n}{a_{k+1}}\boldsymbol{v}_n.$$

To ease notation, we are going to let $b_i = -a_i/a_{k+1}$, which allows us to rewrite this as

$$\boldsymbol{v}_{k+1} = b_1\boldsymbol{w}_1 + \cdots + b_k\boldsymbol{w}_k + b_{k+1}\boldsymbol{w}_{k+1} + b_{k+2}\boldsymbol{v}_{k+2} + \cdots + b_n\boldsymbol{v}_n, \tag{3.9}$$

where we note that $b_{k+1} = -1$.

We claim that

$$\mathcal{D} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k, \boldsymbol{w}_{k+1}, \boldsymbol{v}_{k+2}, \ldots, \boldsymbol{v}_n\}$$

is a basis for $V$, which will complete the proof that **Statement($k+1$)** is true. According to Proposition 3.9, we need to prove that $\mathcal{D}$ spans $V$ and is linearly independent.

We start by showing that $\mathcal{D}$ spans. Let $\boldsymbol{v} \in V$. We want to write $\boldsymbol{v}$ as a linear combination of the vectors in $\mathcal{D}$. We first use the induction assumption that $\mathcal{C}$ is a basis for $V$ to write $\boldsymbol{v}$ as a linear combination of the vectors in $\mathcal{C}$, say

$$\boldsymbol{v} = c_1\boldsymbol{w}_1 + \cdots + c_k\boldsymbol{w}_k + c_{k+1}\boldsymbol{v}_{k+1} + c_{k+2}\boldsymbol{v}_{k+2} + \cdots + c_n\boldsymbol{v}_n.$$

We use (3.9) to eliminate $\boldsymbol{v}_{k+1}$ and then a bit of algebra to compute

$$\begin{aligned}
\boldsymbol{v} &= c_1\boldsymbol{w}_1 + \cdots + c_k\boldsymbol{w}_k \\
&\qquad + c_{k+1}(b_1\boldsymbol{w}_1 + \cdots + b_k\boldsymbol{w}_k + b_{k+1}\boldsymbol{w}_{k+1} + b_{k+2}\boldsymbol{v}_{k+2} + \cdots + b_n\boldsymbol{v}_n) \\
&\qquad + c_{k+2}\boldsymbol{v}_{k+2} + \cdots + c_n\boldsymbol{v}_n \\
&= (c_1 + c_{k+1}b_1)\boldsymbol{w}_1 + \cdots + (c_k + c_{k+1}b_k)\boldsymbol{w}_k + c_{k+1}\boldsymbol{w}_{k+1} \\
&\qquad + (c_{k+1}b_{k+2})\boldsymbol{v}_{k+2} + \cdots + (c_{k+1}b_n)\boldsymbol{v}_n.
\end{aligned}$$

This expresses $\boldsymbol{v}$ as a linear combination of the vectors in $\mathcal{D}$, which completes the proof that $\mathcal{D}$ spans $V$.

Finally, we want to show that $\mathcal{D}$ is a linearly independent set. So suppose that we have a linear combination of the elements of $\mathcal{D}$ that sum to $\mathbf{0}$, say

$$d_1\boldsymbol{w}_1 + \cdots + d_k\boldsymbol{w}_k + d_{k+1}\boldsymbol{w}_{k+1} + d_{k+2}\boldsymbol{v}_{k+2} + \cdots + d_n\boldsymbol{v}_n = \mathbf{0} \tag{3.10}$$

for some scalars $d_1, \ldots, d_n \in F$. We need to show that all of the $d_i$ vanish. Using the fact that $b_{k+1} = -1$, we rewrite (3.9) as

$$\boldsymbol{w}_{k+1} = b_1\boldsymbol{w}_1 + \cdots + b_k\boldsymbol{w}_k + b_{k+1}\boldsymbol{v}_{k+1} + b_{k+2}\boldsymbol{v}_{k+2} + \cdots + b_n\boldsymbol{v}_n. \tag{3.11}$$

We use (3.11) to eliminate $\boldsymbol{w}_{k+1}$ from (3.10) and do a little algebra, thus

$$\begin{aligned}
\mathbf{0} &= d_1\boldsymbol{w}_1 + \cdots + d_k\boldsymbol{w}_k + d_{k+1}\boldsymbol{w}_{k+1} + d_{k+2}\boldsymbol{v}_{k+2} + \cdots + d_n\boldsymbol{v}_n \quad \text{from (3.10),} \\
&= d_1\boldsymbol{w}_1 + \cdots + d_k\boldsymbol{w}_k \\
&\qquad + d_{k+1}(b_1\boldsymbol{w}_1 + \cdots + b_k\boldsymbol{w}_k + b_{k+1}\boldsymbol{v}_{k+1} + b_{k+2}\boldsymbol{v}_{k+2} + \cdots + b_n\boldsymbol{v}_n) \\
&\qquad + d_{k+2}\boldsymbol{v}_{k+2} + \cdots + d_n\boldsymbol{v}_n \quad \text{using (3.11) to eliminate } \boldsymbol{w}_{k+1}, \\
&= (d_1 + d_{k+1}b_1)\boldsymbol{w}_1 + \cdots + (d_k + d_{k+1}b_k)\boldsymbol{w}_k + d_{k+1}b_{k+1}\boldsymbol{v}_{k+1} \\
&\qquad + (d_{k+1}b_{k+2} + d_{k+2})\boldsymbol{v}_{k+2} + \cdots + (d_{k+1}b_n + d_n)\boldsymbol{v}_n.
\end{aligned} \tag{3.12}$$

Our induction assumption that $\mathcal{C}$ is a linearly independent set implies that all of the coefficients of (3.12) vanish. Looking first at the coefficients of $\boldsymbol{v}_{k+1}$, we find that $d_{k+1}b_{k+1} = 0$. But $b_{k+1} = -1$, so $d_{k+1} = 0$. Substituting $d_{k+1} = 0$ into (3.12) yields

$$\mathbf{0} = d_1\boldsymbol{w}_1 + \cdots + d_k\boldsymbol{w}_k + d_{k+2}\boldsymbol{v}_{k+2} + \cdots + d_n\boldsymbol{v}_n.$$

The vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k, \boldsymbol{v}_{k+2}, \ldots, \boldsymbol{v}_n$ are in the set $\mathcal{C}$ that we know is a linearly independent set, which allows us to conclude that $d_1 = \cdots = d_k = d_{k+2} = \cdots = d_n = 0$. But we already proved that $d_{k+1} = 0$, so we have shown that all of the coefficients of (3.10) vanish. This completes the proof that $\mathcal{D}$ is a a linearly independent set, and thus that it is a basis.

We have now proven that for all $0 \le k < m$, if **Statement($k$)** is true, then **Statement($k+1$)** is also true.We conclude by induction that **Statement($k$)** is true for all $0 \le k \le m$. In particular, **Statement($m$)** is true, so possibly after relabeling the vectors in our original basis $\mathcal{A}$, we have shown that the set

$$\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m, \boldsymbol{v}_{m+1}, \ldots, \boldsymbol{v}_n\} \tag{3.13}$$

is a basis for $V$. But we also know that $\mathcal{B} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}$ is a basis for $V$.

Suppose that $n \ge m+1$. Then the fact that $\mathcal{B}$ is a basis means that we can write $\boldsymbol{v}_{m+1}$ as a linear combination of the vectors in $\mathcal{B}$, say

$$\boldsymbol{v}_{m+1} = e_1\boldsymbol{w}_1 + \cdots + e_m\boldsymbol{w}_m.$$

This allows us to write $\mathbf{0}$ as $e_1\boldsymbol{w}_1 + \cdots + e_m\boldsymbol{w}_m - \boldsymbol{v}_{m+1}$, which is a non-trivial linear combination of the vectors in (3.13). Hence (3.13) is not linearly independent,

so it's not a basis. This contradiction proves that $n < m + 1$. Since we started with the assumption that $m \le n$, and since $m$ and $n$ are integers, this completes the proof that $m = n$. □

**Definition.** Let $V$ be a vector space that has a finite basis. The *dimension of* $V$ is the number of vectors in a basis of $V$. This is a well-defined quantity, since Theorem 3.11 tells us that every basis has the same number of elements. The dimension of $V$ is denoted $\dim(V)$, or sometimes $\dim_F(V)$ if we want to specify the field of scalars. A vector space that does not have a finite basis is said to be *infinite dimensional*.

*Example* 3.12. The vector space $F^n$ has dimension $n$. The set of standard basis vectors $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$, as described in Example 3.7, is a basis.

*Example* 3.13. For any $n \ge 0$, the set of polynomials

$$\big\{ f(x) \in F[x] : \deg(f) \le n \big\}$$

is an $F$-vector space of dimension $n + 1$. The set $\{1, x, x^2, \ldots, x^n\}$ is a basis.

*Example* 3.14. The $F$-vector space $F[x]$ consisting of all polynomials is an infinite dimensional vector space. See Exercise 3.13.

*Remark* 3.15. Does every vector space have a basis? Before answering this question, we need to figure out what it means for an infinite set $\mathcal{B}$ to be a basis. Here is the standard definition. We say that $\mathcal{B}$ is a basis for $V$ if every element of $V$ can be written in exactly one way as a linear combination of some finite subset of $\mathcal{B}$. It is important that our linear combinations have only finitely many terms, since in general there is no way to compute infinite sums, which would require some sort of limiting process. With this defintion, the assertion that every vector space has a basis is one of the many statements that are equivalent to the Axiom of Choice, which you studied in Unit #1.

## 3.5 Linear Transformations and Matrices (*Optional*)

**This Section is Under Construction**

In this section we discuss how linear transformations may be described using matrices, and vice versa. The key to doing this is to choose a basis for the domain and the range.

**Definition.** Let $V$ and $W$ be finite dimensional $F$-vector spaces, and let

$$\mathcal{B}_V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \quad \text{and} \quad \mathcal{B}_W = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}$$

be bases for $V$ and $W$, respectively. Let $L : V \to W$ be a linear transformation. Then we can write the images of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ uniquely as linear combinations of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m$, say

$$\begin{aligned} L(\boldsymbol{v}_1) &= a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + \cdots + a_{m1}\boldsymbol{w}_m, \\ L(\boldsymbol{v}_2) &= a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + \cdots + a_{m2}\boldsymbol{w}_m, \\ &\vdots \qquad\qquad\qquad \vdots \\ L(\boldsymbol{v}_n) &= a_{1n}\boldsymbol{w}_1 + a_{2n}\boldsymbol{w}_2 + \cdots + a_{mn}\boldsymbol{w}_m, \end{aligned} \tag{3.14}$$

for a list of scalars $a_{11}, \ldots, a_{mn} \in F$. It is standard practice to arrange these scalars into an $m$-by-$n$ array called a the *matrix of $L$ relative to the bases $\mathcal{B}_V$ and $\mathcal{B}_W$*,

$$M_{L,\mathcal{B}_V,\mathcal{B}_W} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{3.15}$$

A matrix having $m$ rows and $n$ columns is called an *$m$-by-$n$ matrix*.

*Example* 3.16. Let $L : F^n \to F^m$ be a linear transforamtion. Then the matrix of $L$ relative to the standard bases of $F^n$ and $F^m$ is often just called the matrix of $L$. For example, the matrices of the linear transformations $L : F^2 \to F^2$ and $L' : F^2 \to F^3$ in Example 3.4 are

$$M_L = \begin{pmatrix} 3 & 2 \\ -5 & 3 \end{pmatrix} \quad \text{and} \quad M_{L'} = \begin{pmatrix} 3 & 2 \\ -5 & 3 \\ 2 & -7 \end{pmatrix}$$

**Theorem 3.17.** *Let $V$ and $W$ be finite-dimensional $F$-vector spaces and let $\mathcal{B}_V$ and $\mathcal{B}_W$ be bases, respectively, for $V$ and $W$.*

(a) *Let $L : V \to W$ be a linear transformation. Then the matrix $M_{L,\mathcal{B}_V,\mathcal{B}_W}$ of $L$ relative to the two given bases, as given by* (3.15)*, completely determines $L$.*

(b) *Conversely, if $A$ is an $m$-by-$n$ matrix with entries $a_{ij}$ in $F$ for $1 \le i \le m$ and $1 \le j \le n$, then there is a unique linear transformation $L_A : V \to W$ whose matrix relative to the two given bases satisfies $M_{L,\mathcal{B}_V,\mathcal{B}_W} = A$.*

*Proof.* (a) Let $\mathcal{B}_V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ and $\mathcal{B}_W = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}$ be the given bases, so the entries of $M_{L,\mathcal{B}_V,\mathcal{B}_W}$ are determined by (3.14). Let $\boldsymbol{v} \in V$. Since $\mathcal{B}_V$ is a basis, we can write $\boldsymbol{v}$ as a linear combination of the vectors in $\mathcal{B}_V$, say

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n. \tag{3.16}$$

We compute

$$\begin{aligned}
L(\boldsymbol{v}) &= L\biggl(\sum_{j=1}^{n} c_j \boldsymbol{v}_j\biggr) && \text{from (3.16),}\\
&= \sum_{j=1}^{n} c_j L(\boldsymbol{v}_j) && \text{since } L \text{ is a linear transformation,}\\
&= \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij}\boldsymbol{w}_i && \text{using (3.14),}\\
&= \sum_{i=1}^{m} \biggl(\sum_{j=1}^{n} c_j a_{ij}\biggr)\boldsymbol{w}_i && \text{switching order of sum.}
\end{aligned}$$

This formula shows how to use the matrix $M_{L,\mathcal{B}_V,\mathcal{B}_W}$ to compute the value of $L(\boldsymbol{v})$ for every vector $\boldsymbol{v} \in V$.
(b) We leave this as an exercise. See Exercise 3.14. □

The next order of buisness is to figure out what happens to the matrix $M_{L,\mathcal{B}_V,\mathcal{B}_W}$ if we choose different bases for $V$ and $W$. . . .

## 3.6 Subspaces and Quotient Spaces (*Optional*)

**This Section is Under Construction**

## 3.7 Inner Products (*Optional*)

**This Section is Under Construction**

# Exercises

### Section 3.2. Vector Spaces and Linear Transformations

**3.1.** Let $V$ be an $F$-vector space. Use the axioms for a vector space to prove the following statements.
(a) $0\boldsymbol{v} = \mathbf{0}$ for all $\boldsymbol{v} \in V$.
(b) $(-1)\boldsymbol{v} + \boldsymbol{v} = \mathbf{0}$ for all $\boldsymbol{v} \in V$.

**3.2.** Let $V$ be a $\mathbb{Q}$-vector space. Let $n \in \mathbb{Q}$ be a positive integer, and let $\boldsymbol{v} \in V$ be a vector.
(a) Prove that

$$n\boldsymbol{v} = \underbrace{\boldsymbol{v} + \boldsymbol{v} + \cdots + \boldsymbol{v}}_{n \text{ copies of } \boldsymbol{v}}.$$

(*Hint*. Induction.)
(b) What happens if $n$ is a negative integer?

**3.3.** Let $F$ be a field, let $V$ and $W$ be $F$-vector spaces, and let $L : V \to W$ be a linear transformation from $V$ to $W$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ and $c_1, \ldots, c_n \in F$. Prove that

$$L(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n) = c_1L(\boldsymbol{v}_1) + c_2L(\boldsymbol{v}_2) + \cdots + c_nL(\boldsymbol{v}_n).$$

**3.4.** Let $V$ be an $F$-vector space, and let

$$L_1 : V \to V \quad \text{and} L_2 : V \to V$$

be linear transformations from $V$ to itself. We define new functions $L_1 + L_2$ and $L_1L_2$ mapping $V$ to $V$ by the following rules:

$$(L_1 + L_2)(\boldsymbol{v}) = L_1(\boldsymbol{v}) + L_2(\boldsymbol{v}) \quad \text{and} \quad (L_1L_2)(\boldsymbol{v}) = L_1\big(L_2(\boldsymbol{v})\big). \tag{3.17}$$

(a) Prove that $L_1 + L_2$ and $L_1L_2$ are linear transformations.

(b) Let $L_3 : V \to V$ be another linear transformation. Prove the following formulas:

(1) $(L_1 + L_2) + L_3 = L_1 + (L_2 + L_3)$.

(2) $(L_1L_2)L_3 = L_1(L_2L_3)$.

(3) $L_1(L_2 + L_3) = L_1L_2 + L_1L_3$ and $(L_1 + L_2)L_3 = L_1L_3 + L_2L_3$.

(c) Prove that the set of linear transformations from $V$ to $V$ is a ring, where addition and multiplication are given by (3.17). What is the identity element of this ring. What is the additive inverse of a linear transformation $L$?

The ring of linear transformations from $V$ to $V$ is called the *endomorphism ring of* $V$ and is denoted $\mathrm{End}(V)$, or sometimes $\mathrm{End}_F(V)$ if one wants to emphasize the field of scalars.

**3.5.** This is a continuation of Exercise 3.4. Let $V$ be an $F$-vector space, and let $L \in \mathrm{End}_F(V)$. Prove that

$$L \text{ is a unit in the ring } \mathrm{End}_F(V) \quad \text{if and only} \quad \text{if } L : V \to V \text{ is an isomorphism.}$$

The group of units in $\mathrm{End}_F(V)$ is called the *General Linear Group of* $V$ and is denoted by $\mathrm{GL}_F(V)$.[6]

## Section 3.3. Interesting Examples of Vector Spaces

**3.6.** Let $F$ be a field, and let $n \ge 1$ be an integer.

(a) Prove that the set of $n$-tuples $F^n$, with the vector addition and scalar multiplication rules (3.3) and (3.4), is an $F$-vector space.

(b) Let $p$ be a prime. Prove that the $\mathbb{F}_p$-vector space $\mathbb{F}_p^n$ has exactly $p^n$ distinct vectors.

**3.7.** Prove that the map $L$ and $L'$ in Example 3.4 are linear transformations.

**3.8.** Let $F$ be a field, and let $c_1, \ldots, c_n \in F$. Prove that the map

$$E_{\boldsymbol{c}} : F[x] \longrightarrow F^n, \quad E_{\boldsymbol{c}}\big(f(x)\big) = \big(f(c_1), \cdots, f(c_n)\big).$$

is a linear transformation.

## Section 3.4. Bases and Dimension

[6] In the special case that $V = F^n$, the group $\mathrm{GL}_F(F^n)$ is frequently denoted by $\mathrm{GL}_n(F)$ or $\mathrm{GL}(n, F)$.

**3.9.** Let $V$ be a finite dimensional vector space, let $\mathcal{A}$ be a finite subset of $V$, and suppose that $\#\mathcal{A} = \dim(V)$. Prove that the following are equivalent:
(1) $\mathcal{A}$ spans $V$.
(2) $\mathcal{A}$ is linearly independent.
(3) $\mathcal{A}$ is a basis for $V$.

**3.10.** Let $V$ be a finite-dimensional $F$-vector space, and let $n = \dim(V)$. Prove that $V$ is isomorphic to $F^n$, the vector space of $n$-tuples discussed in Examples 3.3 and 3.7.

**3.11.** Let $V$ be a finite-dimensional $F$-vector space, let $n = \dim(V)$, and let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be a set of linearly independent vectors in $V$. Prove that $k \le n$, and that if $k < n$, then there exist vectors $\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n \in V$ so that $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is a basis for $V$.

**3.12.** Let $V$ and $W$ be finite-dimensional $F$-vector spaces.
(a) Suppose that there is an injective linear transformation $V \to W$. Prove that $\dim(V) \le \dim(W)$.
(b) Suppose that there is a surjective linear transformation $V \to W$. Prove that $\dim(V) \ge \dim(W)$.

**3.13.** Prove that the $F$-vector space $F[x]$ consisting of all polynomials with coefficients in $F$ is an infinite dimensional vector space, i.e., prove that $F[x]$ does not have a finite basis.

## Section 3.5. Linear Transformations and Matrices (*Optional*)

**3.14.** Let $V$ and $W$ be finite-dimensional $F$-vector spaces and let $\mathcal{B}_V$ and $\mathcal{B}_W$ be bases, respectively, for $V$ and $W$. Let $A$ is an $m$-by-$n$ matrix with entries in $F$. Prove that there is a unique linear transformation $L_A : V \to W$ whose matrix relative to the two given bases satisfies $M_{L,\mathcal{B}_V,\mathcal{B}_W} = A$. (This exercise is asking you to prove Theorem 3.17(b).)

## Section 3.6. Subspaces and Quotient Spaces (*Optional*)

**3.15.** Let $V$ be a finite-dimensional $F$-vector space, and let $W \subseteq V$ be a vector subspace. Prove that $W$ is finite-dimensional and satisfies

$$\dim(W) \le \dim(V).$$

## Section 3.7. Inner Products (*Optional*)

# Chapter 4

# Fields

## 4.1 Introduction to Fields

In Chapter 2 we introduced fields as being those commutative rings having one additional very special property,[1] and in Chapter 3 we built vector spaces using fields as scalars.

**Definition.** A *field* is a commutative ring $F$ with the property that for every $a \in F$ with $a \neq 0$ there is a $b \in F$ satisfying $ab = 1$.

In this chapter we continue the journey by studying fields in their own right. On the way, we investigate how fields fit one within another, learn how to construct new fields from old fields, and describe all fields having a finite number of elements. These finite fields are not just mathematical curiosities; they play a crucial role in many areas of pure and applied mathematics and engineering, including for example signal processing, error correcting codes, and cryptography.

**Mini-Remark 8**. *Some Historical Motivation*: Field theory was originally developed to aid in studying roots of polynomials. It is easy to find the root of a linear polynomnial $ax + b$, and methods for finding the roots of a quadratic polynomial $ax^2 + bx + c$ have been known since antiquity. In the 16'th century, similar formulas involving cube roots and fourth roots were discovered for the roots of cubic and quartic polynomials,[2] and the race was on to find formulas for polynomials of degree 5 and higher. The theory of fields was developed to provide a repository for the roots of the polynomials, and somewhat surprisingly, group theory turned out to be a fundamental tool in proving that there is no formula involving only $n$'th roots of numbers to solve a general equation of degree at least 5. Indeed, field theory and the theory of finite groups both originated primarily to study roots of polynomials.

[1] Fields are not your plain, everyday, commonplace, come-what-may, average, ordinary kind of rings.

[2] The history of these discoveries, and the machinations of the rival mathematicians, makes for a fascinating story, which you can read about by looking up "Cardano's formula."

## 4.2 Abstract Fields and Homomorphisms

We begin by recalling from Section 2.7 the definition of the unit group of a commutative ring $R$; see also Exercises 2.12 and 2.13.

**Definition.** Let $R$ be a commutative ring $R$. The *unit group of* $R$ is the group

$$R^* = \{a \in R : \text{there is a } b \in R \text{ such that } ab = 1\},$$

where the group law is ring multiplication.

With this definition, a succint way to characterize a field is that it is a commutative ring $F$ satisfying

$$F^* = \{a \in F : a \neq 0\}.$$

As for maps between fields, we obviously want them to preserve the properties that make a field a field. In particular, since fields are rings, we want our maps to (at least) be ring homomorphisms. It turns out that that's enough to ensure that multiplicative inverses go to multiplicative inverses. We prove this assertion, as well as the somewhat surprising fact that maps between fields are always injective. Note that this last assertion is definitely not true in general for rings; for example, the reduction mod $m$ homomorphism $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ his highly non-injective.

**Proposition 4.1.** *Let $F$ and $K$ be fields, and let $\phi : F \to K$ be a ring homomorphism.*
(a) *The map $\phi$ is injective.*
(b) *Let $a \in F^*$. Then $\phi(a^{-1}) = \phi(a)^{-1}$.*

*Proof.* (a) According to Theorem 2.9(b-ii), we need to show that $\ker(\phi)$, the kernel of $\phi$, is the zero ideal. We do a proof by contradiction. Suppose that $\ker(\phi)$ is not the zero ideal, and let $a \in \ker(\phi)$ with $a \neq 0$. Then $a$ has an inverse, say $a \cdot b = 1_F$. We use this to compute

$$1_K = \phi(1_F) = \phi(a \cdot b) = \phi(a) \cdot \phi(b) = 0_K \cdot \phi(b) = 0_K.$$

But the axioms for a ring include the fact that $1 \neq 0$, so this contradiction proves that $\ker(\phi)$ consists of only $0_F$.
(b) From (a) we know that $\phi$ maps non-zero elements of $F$ to non-zero elements of $K$. And since we are told that $\phi$ is a ring homomorphism, we know in particular that $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in F$. In particular, we see that the map

$$\phi : F^* \longrightarrow K^*$$

is a group homomorphism from the unit group of $F$ to the unit group of $K$. Then the fact that $\phi$ sends multiplicative inverses to multiplicative inverses is exactly the statement of Proposition 1.13. □

## 4.3 Interesting Examples of Fields

*Example* 4.2 (**The Fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$**). There are three fields with which you're already very familiar, and they fit one within the next, like Matryoshka nesting dolls,

$$\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

*Example* 4.3 (**The Field $\mathbb{Q}(i)$**). The following subset of $\mathbb{C}$ is a field:

$$\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}.$$

The multiplicative inverse of a non-zero element $a+bi$ can be obtained by "rationalizing the denominator,"

$$\frac{1}{a+bi} = \frac{1}{a+bi} \cdot \frac{a-bi}{a-bi} == \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2} i.$$

*Example* 4.4 (**The Field $\mathbb{Q}(\sqrt{2})$**). In a similar fashion, we can use $\sqrt{2}$ to describe a subset of $\mathbb{R}$ that is a field:

$$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

The multiplicative inverse of a non-zero element $a + b\sqrt{2}$ can again be obtained by rationalizing the denominator,

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

The inverse of $a + b\sqrt{2}$ is well defined, because $\sqrt{2}$ is not a rational number, as you proved in the Number Theory Unit. It follows that $a^2 - 2b^2 \neq 0$ for $a, b \in \mathbb{Q}$ as long as $a$ and $b$ are not both $0$.

*Example* 4.5 (**The Finite Field $\mathbb{F}_p$**). In general, the ring $\mathbb{Z}/m\mathbb{Z}$ need not be a field. For example, the ring $\mathbb{Z}/6\mathbb{Z}$ is not a field, since $2$ does not have a multiplicative inverse. However, it follows from the Number Theory Unit that if $p$ is a prime number, then every non-zero element of $\mathbb{Z}/p\mathbb{Z}$ has a multiplicative inverse, and hence $\mathbb{Z}/p\mathbb{Z}$ is a field. It is an example of a *finite field* and is often denoted $\mathbb{F}_p$. It turns out that there are other finite fields. More precisely, it can be shown that for every prime power $p^k$, there is exactly one finite field containing $p^k$ elements. See Exercise 4.5 for a description of a field with $4$ elements.

*Example* 4.6 (**Skew Fields**). A *skew field*, also called a *division ring*, is ring in which every non-zero element has a multiplicative inverse, but we no longer require that the ring be commutative. A famous theorem of Wedderburn says that a finite skew field must be commutative, but there are many interesting non-commutative infinite skew fields. See Exercise 4.2 for an example.

## 4.4 Subfields and Extension Fields

**Definition.** Let $K$ be a field.[3] A *subfield of* $K$ is a subset $F$ of $K$ that is itself a field using the addition and multiplication operations from $K$.

**Definition.** Let $F$ be a field. An *extension field of* $F$ is a field $K$ such that $F$ is a subfield of $K$. One writes that $K/F$ is an extension of fields.

*Example* 4.7. The field $\mathbb{Q}$ is a subfield of the field $\mathbb{R}$, and thus $\mathbb{R}$ is an extension field of $\mathbb{Q}$. The fields $\mathbb{Q}$ and $\mathbb{R}$ are subfields of the field $\mathbb{C}$, and thus $\mathbb{C}$ is an extension field of the fields $\mathbb{Q}$ and $\mathbb{R}$.

The fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{2})$ described in Examples 4.3 and 4.4 are extension fields of $\mathbb{Q}$. The former is a subfield of $\mathbb{C}$, but not of $\mathbb{R}$, while the latter is a subfield of $\mathbb{R}$.

Let $K/F$ be an extension of fields. We observe that we can add elements of $K$, and we can multiply elements of $K$ by elements of $F$, and that with these observations, the field $K$ becomes an $F$-vector space. What we're doing, in essence, is discarding much of the multiplication operation in $K$, and only retaining multiplication of elements of $K$ by elements of $F$. This myopic policy is the key to using tools from linear algebra to study field extensions. We start with an important definition.

**Definition.** Let $K/F$ be an extension of fields. The *degree of* $K$ *over* $F$, denoted $[K : F]$, is the dimension of $K$ when viewed as an $F$-vector space,

$$[K : F] = \dim_F(K).$$

If $[K : F]$ is finite, we say that $K/F$ is a *finite extension*, otherwise we say that $K/F$ is an *infinite extension*.

*Example* 4.8. The fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{2})$ described in Examples 4.3 and 4.4 have degree 2 over $\mathbb{Q}$,

$$\begin{aligned}\big[\mathbb{Q}(i) : \mathbb{Q}\big] &= 2, \quad \text{since } \{1, i\} \text{ is a } \mathbb{Q}\text{-basis for } \mathbb{Q}(i),\\ \big[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}\big] &= 2, \quad \text{since } \{1, \sqrt{2}\} \text{ is a } \mathbb{Q}\text{-basis for } \mathbb{Q}(\sqrt{2}).\end{aligned}$$

Similarly, we have $[\mathbb{C} : \mathbb{R}] = 2$, since $\{1, i\}$ is an $\mathbb{R}$-basis for $\mathbb{C}$. On the other hand, it is not hard to see that $[\mathbb{R} : \mathbb{Q}] = \infty$. Here's a fancy proof by contradiction, although there are certainly other, more perspicacious, proofs. Suppose that $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{R}$ is a finite set of real numbers that formed a basis for $\mathbb{R}$ as a $\mathbb{Q}$-vector space. Then

$$\mathbb{R} = \{c_1\alpha_1 + \cdots + c_n\alpha_n : c_1, \ldots, c_n \in \mathbb{Q}\}.$$

But the set on the right is countable, since its cardinality is the same as the cardinality of the set of $n$-tuples of rational numbers, while you proved in the Logic and Set Theory Unit that $\mathbb{R}$ is uncountable.

[3] A remark on notation. Since the letter $F$ is frequently used to denote a function, the letter $K$ is often used to denote a field, coming from the German word *Körper* for field. The French word for field is *corp*, but it would be confusing to use $C$, which usually denotes a constant, to be a field.

The next theorem has a flavor similar to the index multiplication rule, described in Exercise 1.23, which counts cosets in a chain of groups.

**Theorem 4.9.** *Let $L/K/F$ be extensions of fields, i.e., $L$ is an extension field of $K$, and $K$ is an extension field of $F$, so we may also view $L$ as an extension field of $F$. Then*

$$[L : F] = [L : K] \cdot [K : F], \tag{4.1}$$

*in the sense that one of the following is true*:
(1) *All of the degrees $[L : F]$, $[L : K]$,and $[K : F]$ are finite, and* (4.1) *is true.*
(2) *$[L : F] = \infty$, and either $[L : K] = \infty$ or $[K : F] = \infty$.*

*Proof.* We start with the case that $L/K$ and $K/F$ are finite extensions. This means that we can choose bases,

$$\begin{aligned}\mathcal{A} &= \{\alpha_1, \alpha_2, \ldots, \alpha_m\} = \text{basis for } K \text{ as an } F\text{-vector space,}\\ \mathcal{B} &= \{\beta_1, \beta_2, \ldots, \beta_n\} \ \ = \text{basis for } L \text{ as an } K\text{-vector space.}\end{aligned}$$

We claim that the set

$$\mathcal{C} = \{\alpha_i\beta_j : 1 \le i \le m,\ 1 \le j \le n\}$$

is a basis for $L$ as an $F$-vector space. Assuming this, it is easy to prove (4.1), since

$$\begin{aligned}[L : F] = \dim_F(L) = \#\mathcal{C} = mn &= \#\mathcal{A} \cdot \#\mathcal{B}\\ &= \dim_F(K) \cdot \dim_K(L) = [K : F] \cdot [L : K].\end{aligned}$$

We will prove that $\mathcal{C}$ is a linearly independent set and leave the rest of the proof of Theorem 4.9 to you. So suppose that we have an $F$-linear combination of the elements of $\mathcal{C}$ that sums to 0, say

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\alpha_i\beta_j = 0 \quad \text{with all } c_{ij} \in F.$$

Our goal is to show that all of the $c_{ij}$ are 0. We switch the order of the sums to obtain

$$\sum_{j=1}^{n}\Bigg(\underbrace{\sum_{i=1}^{m} c_{ij}\alpha_i}_{\text{The inner sum is in } K, \text{ since } c_{ij} \in F \text{ and } \alpha_i \in K.}\Bigg)\beta_j = 0. \tag{4.2}$$

This gives a linear combination of $\beta_1, \ldots, \beta_n$ with coefficients in $K$. But $\beta_1, \ldots, \beta_n$ is a basis for $L$ as a $K$-vector space, so in particular it is $K$-linearly independent. It follows that all of the coefficients of the $\beta_j$ in the sum (4.2) vanish,

$$\sum_{i=1}^{m} c_{ij}\alpha_i = 0 \quad \text{for all } 1 \le j \le n. \tag{4.3}$$

But we also know that $\alpha_1, \ldots, \alpha_m$ is a basis for $K$ as an $F$-vector space, and the $F$-linear independence of $\alpha_1, \ldots, \alpha_m$ implies that the $c_{ij}$ coefficients in (4.3), being in $F$, must all vanish. This completes the proof that every $c_{ij} = 0$, and hence that $\mathcal{C}$ is an $F$-linearly independent set.

The other step in proving that $\mathcal{C}$ is a basis for $L$ as an $F$-vector space is to prove that $\mathcal{C}$ is a spanning set. As mentioned earlier, we leave this, and the other parts of the proof of Theorem 4.9, to you. See Exercise 4.6. □

## 4.5 Polynomial Rings

In this section we briefly discuss some properties of polynomials with coefficients in a field $F$. We start with a couple of familiar definitions.

**Definition.** Let $F$ be a field, and let $f(x) \in F[x]$ be a non-zero polynomial. Write $f(x)$ as

$$f(x) = a_0 + a_1 x + \cdots + a_d x^d \quad \text{with } a_d \neq 0.$$

The *degree of* $f$ is

$$\deg(f) = d.$$

By convention, we set $\deg(0) = -\infty$, a quantity that is smaller than every real number. Further, if $a_d = 1$, then we say that $f$ is a *monic polynomial*.

One easily checks that the product of two polynomials $f_1(x), f_2(x) \in F[x]$ satisfies

$$\deg(f_1 f_2) = \deg(f_1) + \deg(f_2).$$

See Exercise 4.7.

An important property of polynomials is the following division-with-remainder formula, which you've undoubtedly seen for polynomials with real coefficients. It is analogous to the division-with-remainder formula for integers that you used in the Number Theory Unit.

**Proposition 4.10** (Division-with-Remainder for Polynomials)**.** *Let $F$ be a field, and let $f(x), g(x) \in F[x]$ be a polynomials with $g(x) \neq 0$. Then there are unique polynomials $q(x), r(x) \in F[x]$ satisfying*

$$f(x) = g(x)q(x) + r(x) \quad \textit{with} \quad \deg(r) < \deg(g). \tag{4.4}$$

*Proof.* The division algorithm for dividing polynomials that you learned in high school actually works for polynomials with coefficients in any field and can be used to compute $q(x)$ and $r(x)$. We give a fancier existence proof that is not algorithmic.

We consider the following set of polynomials:

$$S = \big\{ f(x) - g(x)q(x) : q(x) \in F[x] \big\}.$$

Among the polynomials in $S$, we take a polynomial $r(x) \in X$ having smallest degree, and we let $q(x)$ be the polynomial satisfying $f(x) - g(x)q(x) = r(x)$.

If $\deg(r) < \deg(g)$, we're done. So we assume that $\deg(r) \ge \deg(g)$ adn derive a contradiction.

We write

$$\begin{aligned} g(x) &= a_0 + a_1x + a_2x^2 + \cdots + a_dx^d \quad \text{with } a_d \neq 0, \\ r(x) &= b_0 + b_1x + b_2x^2 + \cdots + b_nx^n \quad \text{with } b_n \neq 0. \end{aligned}$$

Our assumption that $\deg(r) \ge \deg(g)$ means that $n \ge d$. This allows us to consider the polynomial

$$r(x) - b_n a_d^{-1} x^{n-d} g(x).$$

What we've done is canceled the highest degree term in $r(x)$, so

$$\deg\bigl(r(x) - b_n a_d^{-1} x^{n-d} g(x)\bigr) < \deg(r). \tag{4.5}$$

On the other hand, this polynomial is in the set $S$, since it is equal to

$$\begin{aligned} r(x) - b_n a_d^{-1} x^{n-d} g(x) &= \bigl(f(x) - g(x)q(x)\bigr) - b_n a_d^{-1} x^{n-d} g(x) \\ &= f(x) - \bigl(q(x) + b_n a_d^{-1} x^{n-d} g(x)\bigr) g(x) \in S, \end{aligned} \tag{4.6}$$

i.e., it has the correct form to be in $S$. But among the polynomials in $S$, the polynomial $r(x)$ has the absolutely smallest possible degree, so (4.5) and (4.6) are contradictory. This completes our proof by contradiction that there are polynomials $q(x)$ and $r(x)$ satisfying (4.4).

The second assertion of Proposition 4.10 is that for a given $f(x)$ and $g(x)$, there is only one choice for $q(x)$ and $r(x)$. We leave the proof of this uniqueness property to you; see Exercise 4.8. □

## 4.6 Building Extension Fields

In this section we are going to build fields using roots of polynomials. However, we don't know, *a priori*, where those roots might live. So instead we take a field $F$ and a polynomial $f(x) \in F[x]$ and construct a field extension $K/F$ that magically contains a root of $f(x)$. The model for this construction is the abstract construct of the field of complex numbers from the real numbers and the polynomial $x^2 + 1$ described in Exercise 2.23.

**Definition.** Let $F$ be a field. A non-constant polynomial $f(x) \in F[x]$ is said to be *reducible* if there exist non-constant polynomials $g(x), h(x) \in F[x]$ such that $f(x)$ factors as $f(x) = g(x)h(x)$. An *irreducible polynomial* is a non-constant polynomial this is not reducible.

*Example* 4.11. We can list the four quadratic polynomials in $\mathbb{F}_2[x]$, only one of which turns out to be irreducible.

| | |
|---|---|
| $x^2 = x \cdot x$ | $x^2 + 1 = (x+1)^2$ |
| $x^2 + x = x(x+1)$ | $x^2 + x + 1$ Irreducible |

Quadratic Polynomials in $\mathbb{F}_2[x]$

Similarly, there are eight cubic polynomials in $\mathbb{F}_2[x]$, two of which are irreducible.

| | |
|---|---|
| $x^3 = x \cdot x \cdot x$ | $x^3 + 1 = (x+1)(x^2+x+1)$ |
| $x^3 + x = x(x+1)^2$ | $x^3 + x + 1$ Irreducible |
| $x^3 + x^2 = x^2(x+1)$ | $x^3 + x^2 + 1$ Irreducible |
| $x^2 + x^2 + x = x(x^2+x+1)$ | $x^3 + x^2 + x + 1 = (x+1)^2$ |

Cubic Polynomials in $\mathbb{F}_2[x]$

We start with an elementary proposition that characterizes irreducible polynomials of degrees at most 3. But note that this result goes as far as it can, since for example if $f(x)$ and $g(x)$ are irreducible polynoimals of degree 2, then $f(x)g(x)$ is a reducible polynomial of degree $4$, yet has no roots in $F$.

**Proposition 4.12.** *Let $F$ be a field.*

(a) *Every polynomial of degree $1$ is irreducible.*

(b) *A polynomial of degree $2$ is irreducible if and only if it has no roots in $F$.*

(c) *A polynomial of degree $3$ is irreducible if and only if it has no roots in $F$.*

*Proof.* (a) This is trivial. If a linear polynomial $ax + b$ factors as $g(x)h(x)$, then comparing degrees shows that one of $g(x)$ or $h(x)$ has degree $0$, and hence is constant.

(b,c) Let $f(x) \in F[x]$ be a polynomial whose degree satisfies $2 \le \deg(f) \le 3$, and supppose that $f(x)$ is reducible. This means that we can factor $f(x)$ as $f(x) = g(x)h(x)$ using non-constant polynomials $g(x)$ and $h(x)$. Switching $g$ and $h$ if necessary, we may assume that $\deg(h) \ge \deg(g)$. We use this to compute

$$3 \ge \deg(f) = \deg(g) + \deg(h) \ge 2\deg(g).$$

Hence $\deg(g) \le \frac{3}{2}$. But $\deg(g)$ is a positive integer, so we must have $\deg(g) = 1$. Thus $g(x) = ax + b$ for some $a, b \in F$ with $a \ne 0$. Then $a^{-1}b \in F$ is a root of $g(x)$, and hence also a root of $f(x) = g(x)h(x)$.

Conversely, suppose that $f(x)$ has a root $c \in F$. We divide $f(x)$ by $x - c$ to get

$$f(x) = (x-c)q(x) + r(x) \quad \text{with } \deg(r) < \deg(x-c) = 1.$$

Thus $r(x)$ is a constant polynomial, which we denote by $r$. Evaluating $f(x) = (x - c)q(x) + r$ at $x = c$ and using the assumption that $f(c) = 0$ yields $0 = 0 + r$, so $r = 0$. Therefore $f(x) = (x - c)q(x)$. Since $\deg(f) \ge 2$, this is a non-trivial factorization of $f(x)$, which shows that $f(x)$ is reducible. □

**Theorem 4.13.** *Let $F$ be a field and let $f(x) \in F[x]$ be an irreducible polynomial. Then the principal ideal $f(x)F[x]$ generated by $f(x)$ is a maximal ideal.*

*Proof.* The proof of Theorem 4.13 is almost identical to the proof, which you saw in the Number Theory Unit, that if $p \in \mathbb{Z}$ is a prime number, then its principal ideal $p\mathbb{Z}$ is a maximal ideal. More precisely, you proved that $\mathbb{Z}/p\mathbb{Z}$ is a field, but Theorem 2.14 says that this is equivalent to proving that $p\mathbb{Z}$ is a maximal ideal. The key tool in both proofs is division with remainder. Exercise 4.13 sketches a proof of Theorem 4.13 and asks you to fill in the details. □

We can use Theorem 4.13 to construct an extension field that contains the root of a given polynomial; see also Exercise 4.12.

**Theorem 4.14.** *Let $F$ be a field, let $f(x) \in F[x]$ be an irreducible polynomial, let $I_f = f(x)F[x]$ be the principal ideal generated by $f(x)$, and let $K_f = F[x]/I_f$ be the indicated quotient ring.*

(a) *The ring $K_f$ is a field.*

(b) *The field $K_f$ is a finite extension of the field of $F$. Its degree is given by*

$$[K_f : F] = \deg(f).$$

(c) *The polynomial $f(x)$ has a root in the field $K_f$.*

*Proof.* (a) Theorem 4.13 tells us that $I_f$ is a maximal ideal of the ring $F[x]$, and then Theorem 2.14(b) tells us that the quotient ring $F[x]/I_f$ is a field.
(b) Let $d = \deg(f)$, and let

$$\beta = x + I_f = \text{the coset of } x \text{ in the quotient ring } F[x]/I_f.$$

We claim that the set

$$\mathcal{B} = \{1, \beta, \beta^2, \ldots, \beta^{d-1}\}$$

is an $F$-basis for $K_f$.

First we check that $\mathcal{B}$ spans. Let $g(x) + I_f$ be an arbitrary element of $K_f$. We divide the polynomial $g(x)$ by $f(x)$ to get a quotient and a remainder,

$$g(x) = f(x)q(x) + r(x) \quad \text{for some } q(x), r(x) \in F[x] \text{ with } \deg(r) < \deg(f) = d.$$

Since $f(x) \in I_f$, the cosets $g(x)+I_f$ and $r(x)+I_f$ are the same. On the other hand, if we write $r(x)$ as

$$r(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{d-1}x^{d-1} \quad \text{with } a_0, \ldots, a_{d-1} \in F,$$

then we can compute

$$\begin{aligned} g(x) + I_f &= r(x) + I_f \\ &= (a_0 + a_1x + a_2x^2 + \cdots + a_{d-1}x^{d-1}) + I_f \\ &= (a_0 + I_f) + (a_1x + I_f) + (a_2x^2 + I_f) + \cdots + (a_{d-1}x^{d-1} + I_f) \\ &= a_0(1 + I_f) + a_1(x + I_f) + a_2(x + I_f)^2 + \cdots + a_{d-1}(x + I_f)^{d-1} \\ &= a_0 + a_1\beta + a_2\beta^2 + \cdots + a_{d-1}\beta^{d-1}. \end{aligned}$$

This prove that every element of $K_f$ is an $F$-linear combination of the elements in $\mathcal{B}$.

Next we check that $\mathcal{B}$ is $F$-linearly independent. So we suppose that we have a linear combination

$$c_0 + c_1\beta + c_2\beta^2 + \cdots + c_{d-1}\beta^{d-1} = 0$$

for some $c_0, \dots, c_{d-1} \in F$. Our goal is to prove that all of the $c_i$ vanish. Using the definition $\beta = x + I_f$, we compute

$$\begin{aligned}
0 + I_f &= c_0 + c_1\beta + c_2\beta^2 + \cdots + c_{d-1}\beta^{d-1} \\
&= c_0(1 + I_f) + c_1(x + I_f) + c_2(x + I_f)^2 + \cdots + c_{d-1}(x + I_f)^{d-1} \\
&= (c_0 + I_f) + (c_1x + I_f) + (c_2x^2 + I_f) + \cdots + (c_{d-1}x^{d-1} + I_f) \\
&= (c_0 + c_1x + c_2x^2 + \cdots + c_{d-1}x^{d-1}) + I_f.
\end{aligned}$$

This tells us that the polynomial $c_0 + c_1x + c_2x^2 + \cdots + c_{d-1}x^{d-1}$ is in the principal ideal $I_f$ generated by $f(x)$, so there is some polynomial $g(x) \in F[x]$ such that

$$c_0 + c_1x + c_2x^2 + \cdots + c_{d-1}x^{d-1} = f(x)g(x). \tag{4.7}$$

If $g(x) \neq 0$, then we could take the degrees of both sides to obtain the contradiction

$$\begin{aligned}
d - 1 &\ge \deg(c_0 + \cdots + c_{d-1}x^{d-1}) \\
&= \deg\bigl(f(x)g(x)\bigr) = \deg\bigl(f(x)\bigr) + \deg\bigl(g(x)\bigr) = d + \deg\bigl(g(x)\bigr) \ge d.
\end{aligned}$$

Hence $g(x) = 0$, and thus (4.7) becomes an equality

$$c_0 + c_1x + c_2x^2 + \cdots + c_{d-1}x^{d-1} = 0$$

that is true in the polynomial ring $F[x]$. Therefore $c_0 = \cdots = c_{d-1} = 0$, which completes the proof that $\mathcal{B}$ is $F$-linearly independent, and thus is a basis for $K_f/F$. It further proves that

$$[K_f : F] = \#\mathcal{B} = d = \deg(f).$$

(c) It almost seems like cheating, but we claim that the element $\beta = x + I_f \in K_f$ is a root of $f(x)$. To see this, we write $f(x)$ as

$$f(x) = b_0 + b_1x + b_2x^2 + \cdots + b_dx^d.$$

Then

$$\begin{aligned}
f(\beta) &= b_0 + b_1\beta + b_2\beta^2 + \cdots + b_d\beta^d \\
&= b_0(1 + I_f) + b_1(x + I_f) + b_2(x + I_f)^2 + \cdots + b_d(x + I_f)^d \\
&= (b_0 + I_f) + (b_1x + I_f) + (b_2x^2 + I_f) + \cdots + (b_dx^d + I_f) \\
&= b_0 + b_1x + b_2x^2 + \cdots + b_dx^d + I_f \\
&= f(x) + I_f \\
&= 0 + I_f.
\end{aligned}$$

Hence $f(\beta)$ is zero element of the field $K_f$. □

## 4.7 Finite Fields

In this section we apply all of the tools that we've developed in order describe finite fields and to construct finite fields of various prime power orders.

**Proposition 4.15.** *Let $F$ be a finite field.*

(a) *There is a unique prime $p$ with the property that $p = 0$ in $F$, or equivalently, with the property that every $a \in F$ satisfies*

$$\underbrace{a + a + \cdots + a}_{p \text{ terms}} = 0.$$

*The prime $p$ is called the* characteristic of the field $F$.

(b) *The finite field $\mathbb{F}_p$ is a subfield of $F$.*

(c) *The number of elements of $F$ is given by the formula*

$$\#F = p^{[F:\mathbb{F}_p]}.$$

*In particular, the order of every finite field is a power of a prime.*

*Proof.* (a) Let $m = \#F$ be the number of elements of $F$. We write $0_F$ and $1_F$ for the additive and multiplicative identity elements of $F$, so as to distinguish them from the integers $0$ and $1$. Addition makes $F$ into a finite group with $m$ elements, so Lagrange's theorem (Corollary 1.26) says that every element of the group $(F, +)$ has order dividing $m$. In particular, if we add any element to itself $m$ times, we get $0_F$.

We let $n \ge 1$ be the smallest integer with this property. We claim that $n$ is prime. In any case, we let $p$ be a prime dividing $n$ and write $n = pq$, and then our goal is so show that $n = p$. We know that

$$\begin{aligned} 0_F &= \underbrace{1_F + 1_F + \cdots + 1_F}_{n \text{ terms}} \\ &= \underbrace{1_F + 1_F + \cdots + 1_F}_{pq \text{ terms}} \\ &= \Bigl(\underbrace{1_F + 1_F + \cdots + 1_F}_{p \text{ terms}}\Bigr) \cdot \Bigl(\underbrace{1_F + 1_F + \cdots + 1_F}_{q \text{ terms}}\Bigr). \end{aligned}$$

Since $F$ is a field, we know that one of the factors must be $0_F$. If it's the first factor, then for every $\alpha \in F$ we have

$$0_F = \alpha \cdot 0_F = \alpha \cdot \Bigl(\underbrace{1_F + 1_F + \cdots + 1_F}_{p \text{ terms}}\Bigr) = \Bigl(\underbrace{\alpha + \alpha + \cdots + \alpha}_{p \text{ terms}}\Bigr),$$

and similarly if it's the second factor, then for every $\alpha \in F$ we have

$$0_F = \alpha \cdot 0_F = \alpha \cdot \Bigl(\underbrace{1_F + 1_F + \cdots + 1_F}_{q \text{ terms}}\Bigr) = \Bigl(\underbrace{\alpha + \alpha + \cdots + \alpha}_{q \text{ terms}}\Bigr).$$

But $n$ is the smallest positive integer with this property, so since $p \ge 2$, we must have $n = p$ and $q = 1$.
(b) Informally, we can identify $\mathbb{F}_p$ with a subfield of $F$ via the map

$$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} \hookrightarrow F, \quad n + p\mathbb{Z} \longmapsto \underbrace{1_F + 1_F + \cdots + 1_F}_{n \text{ terms}}.$$

More formally, given any ring $R$, there is a unique ring homorphism from $\mathbb{Z}$ to $R$ that take $1_\mathbb{Z}$ to $1_R$, so in particular we have

$$\phi : \mathbb{Z} \longrightarrow F, \quad \phi(n) = \underbrace{1_F + 1_F + \cdots + 1_F}_{n \text{ terms}}.$$

From (a) we see that the kernel of $\phi$ contains the ideal $p\mathbb{Z}$, and $\ker(\phi)$ cannot be all of $\mathbb{Z}$, since $\phi(1) = 1_F \ne 0_F$, so the fact that $p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$ implies that $\ker(\phi) = p\mathbb{Z}$. Then Proposition 2.9 tells us that $\phi$ induces an injective ring homomorphism

$$\mathbb{Z}/p\mathbb{Z} \hookrightarrow F.$$

This gives the desired copy of $\mathbb{F}_p$ in $F$.
(c) From (b) we know that $F$ is an extension field of $\mathbb{F}_p$. And since $F$ has only finitely many elements, its dimension as an $\mathbb{F}_p$-vector space is finite, since there are only finitely many elements to use in forming a basis! Let

$$\mathcal{B} = \{\beta_1, \beta_2, \ldots, \beta_d\}$$

be an $\mathbb{F}_p$-basis for $F$, so by definition, $d = [F : \mathbb{F}_p]$. The definition of basis tells us that $F$ consists of exactly the elements

$$c_1\beta_1 + c_2\beta_2 + \cdots + c_d\beta_d \quad \text{with } c_1, \ldots, c_d \in \mathbb{F}_p,$$

and that different choices of $c_1, \ldots, c_d$ give different elements of $F$. In fancier terms, there is a bijection of sets (and indeed, of $\mathbb{F}_p$-vector spacdes) defined by

$$\mathbb{F}_p^d \longrightarrow F, \quad (c_1, \ldots, c_d) \longmapsto c_1\beta_1 + \cdots + c_d\beta_d.$$

Hence

$$\#F = \#\mathbb{F}_p^d = p^d,$$

which completes the proof of (c). □

The full proof of the next result is unfortunately beyond the scope of these notes. We will prove that it is true up to degree 3, and refer the reader to other sources for the complete proof. We want to stress that the result is by no means obvious. For example, it would not be true if we replaced $\mathbb{F}_p$ with the field $\mathbb{R}$, since every irreducible polynomial in the ring $\mathbb{R}[x]$ has degree 1 or 2.

**Theorem 4.16.** *Let $p$ be a prime and let $d \ge 1$. Then the ring $\mathbb{F}_p[x]$ contains an irreducible polynomial of degree $d$.*

*Proof* (*for $d \le 3$*). For $d = 1$, any degree 1 polynomial will work.

Since we can always multiply a polynomial by a non-zero scalar without affecting its irreducibility, it suffices to look at *monic polynomials*, that is, polynomials whose leading coefficient is 1. We let

$$\begin{aligned}\mathrm{Poly}_d &= \bigl\{\text{monic polynomials in } \mathbb{F}_p[x] \text{ having degree } d\bigr\},\\ \mathrm{Irred}_d &= \bigl\{\text{monic irreducible polynomials in } \mathbb{F}_p[x] \text{ having degree } d\bigr\}\\ \mathrm{Red}_d &= \bigl\{\text{monic reducible polynomials in } \mathbb{F}_p[x] \text{ having degree } d\bigr\}.\end{aligned}$$

We note that $\#\,\mathrm{Poly}_d = p^d$, since a monic polynomial of degree $d$ in $\mathbb{F}_p[x]$ has exactly $d$ coefficients that may be freely chosen from $\mathbb{F}_p$

We start with $d = 2$, so we want to compute $\#\,\mathrm{Irred}_2$, the number of irreducible monic quadratic polynomials. We won't do this directly, but instead will use a lesson learned in the Combinatorics Unit. We will count the number of reducible monic quadratic polynomials, which are the polynomials that we don't want, and subtract this value from the total number of monic quadratic polynomials to get the number that we do want.

Okay, which polynomials of the form $x^2+ax+b$ are reducible. They are precisely the polynomials that factor as

$$x^2 + ax + b = (x - \alpha)(x - \beta) \quad \text{for some } \alpha, \beta \in \mathbb{F}_p.$$

It may look as if there are $p^2$ choices for $(\alpha, \beta)$, but we need to be careful not to double count the resulting polynomials! There are two cases. First, if $\alpha = \beta$, then we get $p$ different polynomials $(x - \alpha)^2$ with $\alpha \in \mathbb{F}_p$. Second, if $\alpha \ne \beta$, then there are $p(p - 1)$ choices for the pair $(\alpha, \beta)$, but since the order of $\alpha$ and $\beta$ does not change the product $(x - \alpha)(x - \beta)$, the number of polynomials that we get is the combinatorial symbol $\binom{p}{2}$, i.e., the number of ways to choose two distinct elements of $\mathbb{F}_p$ if the order doesn't matter. We have shown that

$$\#\,\mathrm{Red}_2 = p + \binom{p}{2} = p + \frac{p(p-1)}{2} = \frac{p^2+p}{2}.$$

Removing these from the totality of all monic quadratic polynomials gives

$$\#\,\mathrm{Irred}_2 = \#\,\mathrm{Poly}_2 - \#\,\mathrm{Red}_2 = p^2 - \frac{p^2+p}{2} = \frac{p^2-p}{2}. \tag{4.8}$$

This last quantity is positive for all $p \ge 2$, which proves that $\mathrm{Irred}_2$ is non-empty.

Next up are cubic polynomials. We want to count how many of them are reducible. There are quite a few possible ways that a monic cubic polynomial can factor. Here's the list, with a count of how many distinct polynomials there are of each type:

| Polynomial | Notes | # of Polys |
|---|---|---|
| $(x-\alpha)^3$ | $\alpha \in \mathbb{F}_p$ | $p$ |
| $(x-\alpha)^2(x-\beta)$ | $\alpha, \beta \in \mathbb{F}_p,\ \alpha \ne \beta$ | $p(p-1)$ |
| $(x-\alpha)(x-\beta)(x-\gamma)$ | $\alpha, \beta, \gamma \in \mathbb{F}_p,\ \alpha, \beta, \gamma$ distinct | $\binom{p}{3}$ |
| $(x-\alpha)(x^2+ax+b)$ | $\alpha, a, b \in \mathbb{F}_p,\ x^2+ax+b$ irreducible | $p \cdot \#\,\mathrm{Irred}_2$ |

We note that for polynomials of the form $(x-\alpha)^2(x-\beta)$, the order of $\alpha$ and $\beta$ does matter, which is why we get $p(p-1)$ of them, and not $\binom{p}{2}$. Adding the last column of the table and using the value $\#\operatorname{Irred}_2 = (p^2-p)/2$ that we computed earlier, we find that

$$\begin{aligned}\#\operatorname{Red}_2 &= p + p(p-1) + \binom{p}{3} + p\cdot\#\operatorname{Irred}_2 \\ &= p^2 + \frac{p(p-1)(p-2)}{6} + p\cdot\frac{p^2-p}{2} \\ &= \frac{2p^3+p}{3}.\end{aligned}$$

Hence

$$\#\operatorname{Irred}_3 = \#\operatorname{Poly}_3 - \#\operatorname{Red}_3 = p^3 - \frac{2p^3+p}{3} = \frac{p^3-p}{3}. \tag{4.9}$$

This proves that $\operatorname{Irred}_3$ is non-empty for all $p \ge 2$. □

*Example* 4.17. Taking $p=2$ in the formulas $\#\operatorname{Irred}_2 = \frac{1}{2}(p^2-p)$ and $\#\operatorname{Irred}_3 = \frac{1}{3}(p^3-p)$ that we derived while proving Theorem 4.16 shows that $\mathbb{F}_2[x]$ has exactly one monic irreducible quadratic polynomial and exactly two monic irreducible cubic polynomials. These values agree with the exhaustive lists that we made in Example 4.11.

We are going to prove the first part of the following fundamental theorem, but for completeness we include the second part, which would generally be proven in a more advanced abstract algebra class.

**Theorem 4.18.** *Let $p$ be a prime and let $d \ge 1$.*
(a) *There exists a field $F$ containing exactly $p^d$ elements.*
(b) *Any two fields containing $p^d$ elements are isomorphic.*

*Proof.* (a) Theorem 4.16 tells us that there is an irreducible polynomial $f(x) \in \mathbb{F}_p[x]$ satisfying $\deg(f) = d$. We let $K_f = \mathbb{F}_p[x]/f(x)\mathbb{F}_p[x]$ be the field described in Theorem 4.14. That theorem tells us that $[K_f : \mathbb{F}_p] = \deg(f) = d$. In particular, we see that $K_f$ has finitely many elements, since if $a_1, \dots, a_d \in K_f$ is a basis for $K_f$ considered as an $\mathbb{F}_p$-vector spaces, then

$$K_f = \{c_1a_1 + \cdots + c_da_d : c_1, \dots, c_d \in \mathbb{F}_p\}.$$

Indeed, we have essentially reproven Proposition 4.15, which says that

$$\#K_f = p^{[K_f:\mathbb{F}_p]} = p^d.$$

This completes the proof of (a), and we refer the reader to any standard abstract algebra text for a proof of (b). □

## Addendum to Section 4.7

*Remark* 4.19. The inclusion/exclusion combinatorics that we used in proving Theorem 4.16 for degrees 2 and 3, where we listed all possible factorizations of a a polynomial of degree $d$, becomes increasingly complicated as $d$ increases. So although one might be able to prove a general formula for $\#\operatorname{Irred}_d$ by this method, the task would be painful. An alternative approach starts with the following interesting identity, which we will not prove:[4]

$$\prod_{\substack{f(x)\in\mathbb{F}_p[x]\\ f \text{ monic irreducible}\\ \deg(f) \text{ divides } n}} f(x) = x^{p^n} - x. \tag{4.10}$$

Taking degrees of both sides of (4.10) gives the formula

$$\sum_{d|n} d \cdot \#\operatorname{Irred}_d = p^n. \tag{4.11}$$

This formula and inclusion/exclusion can be used to compute $\#\operatorname{Irred}_d$. For example, suppose that we take $n$ to be equal to a prime $\ell$. The only divisors of $\ell$ are 1 and $\ell$, so (4.11) becomes

$$\#\operatorname{Irred}_1 + \ell \cdot \#\operatorname{Irred}_\ell = p^\ell.$$

We know $\#\operatorname{Irred}_1 = p$, which yields the formula

$$\#\operatorname{Irred}_\ell = \frac{p^\ell - p}{\ell}.$$

This agrees, as it should, with the formulas (4.8) and (4.9) that we found earlier for $\#\operatorname{Irred}_2$ and $\#\operatorname{Irred}_3$.

# Exercises

### Section 4.2. Abstract Fields and Homomorphisms

**4.1.** Let $F$ be a field, and let $f(x) \in F[x]$ be a non-zero polynomial.

(a) Suppose that $\alpha \in F$ is a root of $f(x)$, i.e., $f(\alpha) = 0$. Prove that there is a polynomial $g(x) \in F[x]$ such that $f(x) = (x - \alpha)g(x)$.

(b) More generally, suppose that $\alpha_1, \ldots, \alpha_n \in F$ are distinct roots of $f(x)$. Prove that there is a polynomial $g(x) \in F[x]$ such that

$$f(x) = (x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)g(x)$$

(*Hint*. Use (a) and induction. But note that somewhere you will need to use the fact that $F$ is a field, since the result need not be true if $F$ is an arbitrary ring. For example, the polynomial $x^2 - 1 \in (\mathbb{Z}/8\mathbb{Z})[x]$ has distinct roots $1, 3, 5, 7 \in \mathbb{Z}/8\mathbb{Z}$.)

[4] You can find a proof of formulas (4.10) and (4.11) in most number theory and algebra textbooks; see for example Theorem 2 of Chapter 7 Section 2 of Ireland–Rosen, *A Classical Introduction to Modern Number Theory*, Springer, 1990.

(c) Use (b) to deduce the following important result.

**Theorem 4.20.** *Let $F$ be a field, and let $f(x) \in F[x]$ be a non-zero polynomial. Then $f(x)$ has as most $\deg(f)$ distinct roots in $F$.*

**4.2.** Prove that the ring of quaternions $\mathbb{H}$ described in Example 2.5 is a skew field, that is, it has the property that every non-zero element has a multiplicative inverse. (*Hint*. You may find Exercise 2.14 helpful.)

## Section 4.3. Interesting Examples of Fields

**4.3.** Prove that each of the following subsets of $\mathbb{R}$ is a field.
(a) $\mathbb{Q}(\sqrt{3}) = \{a + b\sqrt{3} : a, b \in \mathbb{Q}\}$.
(b) $\mathbb{Q}(\sqrt{4}) = \{a + b\sqrt{4} : a, b \in \mathbb{Q}\}$.
(c) $\mathbb{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} : a, b, c \in \mathbb{Q}\}$.

**4.4.** Prove that each of the following rings is not a field.
(a) $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$.
(b) $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$.
(c) $\mathbb{Z}/p^2\mathbb{Z}$, where $p$ is a prime.
(d) $\mathbb{Z}/mn\mathbb{Z}$, where $m \geq 2$ and $n \geq 2$.

**4.5.** Consider the set $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ consisting of $4$ elements. Define an addition law and a multiplication law on $\mathbb{F}_4$ using Table 4.1.
(a) Prove that these rules make $\mathbb{F}_4$ into a field.
(b) Prove that $\mathbb{F}_4$ is not isomorphic to the ring $\mathbb{Z}/4\mathbb{Z}$, although they both have $4$ elements. (*Hint*. Find some property for which $\mathbb{F}_4$ and $\mathbb{Z}/4\mathbb{Z}$ differ.)

| $+$ | $0$ | $1$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| $0$ | $0$ | $1$ | $\alpha$ | $\beta$ |
| $1$ | $1$ | $0$ | $\beta$ | $\alpha$ |
| $\alpha$ | $\alpha$ | $\beta$ | $0$ | $1$ |
| $\beta$ | $\beta$ | $\alpha$ | $1$ | $0$ |

| $\times$ | $0$ | $1$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| $0$ | $0$ | $0$ | $0$ | $0$ |
| $1$ | $0$ | $1$ | $\alpha$ | $\beta$ |
| $\alpha$ | $0$ | $\alpha$ | $\beta$ | $1$ |
| $\beta$ | $0$ | $\beta$ | $1$ | $\alpha$ |

Table 4.1: Addition and Multiplication Tables for $\mathbb{F}_4$

## Section 4.4. Subfields and Extension Fields

**4.6.** Let $L/K/F$ be extensions of fields. This exercise asks you to complete the proof of Theorem 4.9.
(a) Assume that $L/K$ and $K/F$ are finite extensions. During the proof of Theorem 4.9, we defined a set $\mathcal{C} = \{\alpha_i\beta_j : 1 \leq i \leq m,\ 1 \leq j \leq n\}$, claimed that $\mathcal{C}$ is a basis for $L$ as an $F$-vector space, and proved that $\mathcal{C}$ is linearly independent. Prove that $\mathcal{C}$ is a spanning set, thereby completing the proof in this case.
(b) Prove that $L/F$ is an infinite extension if and only if at least one of $L/K$ or $K/F$ is an infinite extension.

## Section 4.5. Polynomial Rings

**4.7.** (a) Let $F$ be a field and let $f_1(x), f_2(x) \in F[x]$. Prove that

$$\deg(f_1 f_2) = \deg(f_1) + \deg(f_2).$$

(b) Let $R$ be the ring $\mathbb{Z}/6\mathbb{Z}$. Find non-constant polynomials $f_1(x), f_2(x) \in R[x]$ for which there is a strict inequality $\deg(f_1 f_2) < \deg(f_1) + \deg(f_2)$.

**4.8.** This excerise asks you to prove the uniqueness of the quotient and remainder appearing in Proposition 4.10. Let $F$ be a field, let $f(x), g(x) \in F[x]$ be a polynomials with $g(x) \neq 0$, and suppose that there are polynomials $q_1(x), q_2(x), r_1(x), r_2(x) \in F[x]$ satisfying

$$\begin{aligned} f(x) &= g(x)q_1(x) + r_1(x) \quad \text{with} \quad \deg(r_1) < \deg(g), \\ f(x) &= g(x)q_2(x) + r_2(x) \quad \text{with} \quad \deg(r_2) < \deg(g). \end{aligned}$$

Prove that $q_1(x) = q_2(x)$ and $r_1(x) = r_2(x)$.

## Section 4.6. Building Extension Fields

**4.9.** This exercise asks you to prove the converse of Theorem 4.13. Let $F$ be a field, and let $f(x) \in F[x]$ be a polynomial with the property that the principal ideal $f(x)F[x]$ is maximal. Prove that $f(x)$ is irreducible in $F[x]$.

**4.10.** This exercise shows that Theorem 4.13 is not true for polynomials in more than one variable. Let $F$ be a field, and let $f(x, y) \in F[x, y]$ be a polynomial in two variables. Prove that the principal ideal $f(x)F[x]$ generated by $f(x, y)$ is not maximal.

**4.11.** In this exercise, you may use the fact that every non-constant polynomial in $\mathbb{C}[x]$ has at least one root in $\mathbb{C}$. The mathematical term for this property is that $\mathbb{C}$ is *algebraically closed*.

(a) Prove that the only irreducible polynomials in $\mathbb{C}[x]$ are linear polynomials.

(b) Let $f(x) \in \mathbb{R}[x]$ and suppose that $a + bi \in \mathbb{C}$ is a root of $f(x)$. Prove that $a - bi$ is also a root of $f(x)$.

(c) Let $f(x) \in \mathbb{R}[x]$ be an irreducible polynomial in $\mathbb{R}[x]$. Prove that $\deg(f) \leq 2$. (*Hint.* Use (b).)

**4.12.** Let $F$ be a field, let $f(x) \in F[x]$ be a possibly reducible non-constant polynomial, and let $d = \deg(f)$.

(a) Prove that there exists a field extension $K/F$ satisfying $[K : F] \leq d$ such that $f(x)$ has a root in $K$.

(b) Prove that there exists a field extension $L/F$ and elements $c \in F$ and $\alpha_1, \ldots, \alpha_d \in L$ such that $f(x)$ factors in $L[x]$ as

$$f(x) = c(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_d).$$

Prove that it is always possible to find such an $L$ that also satisfies

$$[L : F] \leq d!.$$

The field $L$ is called a *splitting field for the polynomial $f(x)$ over the field $F$*.

**4.13.** Let $F$ be a field. In this exercise you will prove that irreducible polynomials in $F[x]$ generate maximal ideals.

(a) Let $f(x), g(x) \in F[x]$ be polynomials, not both 0. Let $I$ be the set[5]

$$I = \big\{ f(x)u(x) + g(x)v(x) : u(x), v(x) \in F[x] \big\}.$$

Among the non-zero polynomials in $I$, let $h(x)$ be one having the smallest possible degree.

(i) Prove that $h(x)$ divides $f(x)$ and $g(x)$. (*Hint*. To show that $h(x)$ divides $f(x)$, write $f(x) = h(x)q(x) + r(x)$ with $\deg(r) < \deg(h)$, and prove that $r(x) \in I$. Then use the minimality of the degree of $h(x)$ to show that $r(x) = 0$.)

(ii) Conversely, prove that if a polynomial divides both $f(x)$ and $g(x)$, then it also divides $h(x)$.

(b) Let $f(x) \in F[x]$ be an irreducible polynomial. Prove that the principal ideal $f(x)F[x]$ generated by $f(x)$ is a maximal ideal. (*Hint*. Let $J$ be an ideal containing $f(x)F[x]$, and suppose that $J$contains a polynomial $g(x)$ not in $f(x)F[x]$. Look at the set $I$ generated by $f(x)$ and $g(x)$ as in (a) and prove that a smallest degree non-zero polynomial $h(x)$ in $I$ must be a constant polynomial.)

## Section 4.7. Finite Fields

**4.14.** Let $F$ be a finite field with $q$ elements.

(a) Prove that every non-zero element of $F$ is a root of the polynomial $x^{q-1} - 1$. (*Hint*. Apply the corollary of Lagrange's theorem, Corollary 1.26, to the group of units $F^*$.)

(b) Prove that every element of $F$ is a root of the polynomial $x^q - x$.

(c) Prove the formula

$$\prod_{\alpha \in F} (x - \alpha) = x^q - x.$$

(*Hint*. Use Theorem 4.20 that appears in Exercise 4.1(c).)

**4.15.** In the proof of Theorem 4.16, we made a table listing all of the possible non-trivial factorizations of polynomials of degree 3 in $\mathbb{F}_p[x]$, counted how many there were of each type of factorization, and summed them to compute $\#\operatorname{Red}_3$.

(a) Make a list of all possible factorizations of polynomials of degree 4 in $\mathbb{F}_p[x]$, and use your list to compute $\#\operatorname{Red}_4$ and $\#\operatorname{Irred}_4$.

(b) Check your answer by using formula (4.11), which we did not prove, to compute $\#\operatorname{Irred}_4$. Is this easier than the method in (a)?

**4.16.** In this exercise, you may use formula (4.11) to compute $\#\operatorname{Irred}_d$ for various values of $d$.

(a) Let $\ell$ be a prime. Compute $\#\operatorname{Irred}_{\ell^2}$ and $\#\operatorname{Irred}_{\ell^3}$, use your answer to guess a formula for $\#\operatorname{Irred}_{\ell^k}$ for all $k \ge 1$, and then use (4.11) and induction to prove that your formula is correct.

(b) Let $\ell$ and $q$ be distinct primes. Compute $\#\operatorname{Irred}_{\ell q}$.

[5] You may want to verify that $I$ is in fact an ideal of the ring $F[x]$.