MA383: Introduction to Modern Algebra

Dylan C. Beck

Acknowledgements

Primarily, the first two chapters of this document were created in the Fall 2022 semester at Baker University. Bearing this in mind, I would like to express my gratitude toward my students in MA383 (Introduction to Modern Algebra) — especially those who assisted in the enhancement of these notes with comments and suggestions, including Sabrena Hayles and Allie Renfro. I am also grateful to Vance Gaffar for his helpful comments and interesting solutions to some of the exercises.

Elsewhere, the bulk of the advanced material from the third chapter on group theory and the fourth chapter on ring theory was written independently from August 2018 to January 2022 at the request of students from the Algebra Qualifying Exam Study Group at the University of Kansas. Exercises in these sections have in part been adopted from the qualifying exams in algebra at the University of Kansas. I would like to thank the graduate students there for their motivation and participation in these study sessions — especially Wayne Ng Kwing King, Enrique Salcido, Neethu Suma-Raveendran, and Christopher Wong. I am also grateful to Souvik Dey and Monalisa Dutta for their assistance with these meetings and their peripheral contributions to these notes.

Contents

Chapter 1

Basic Properties of Sets and Logic

Before we delve into the theory of modern algebra, we must first understand and appreciate that mathematics is a language of its own; thus, it is imperative for us to develop the necessary diction, grammar, and syntax in order for us to effectively communicate. We accomplish this formally via the tools of set theory and the calculus of logic. Even now, these branches of mathematics enjoy an ongoing ubiquity and significance that makes them an active area of research, but we will not trouble ourselves with these subtle complexities. (Explicitly, if it matters to the reader, we will adopt the standard axioms of the Zermelo-Fraenkel set theory with the Axiom of Choice.)

1.1 Sets and Set Operations

We define a **set** X as a collection of like objects, e.g., functions or real numbers. We refer to an arbitrary object x of X as an **element** (or **member**) of X. If x is an element of X, then we write $x \in X$ to denote that "x is an element (or member) of the set X." We may also say in this case that x "belongs to" or "lies in" X, or we may wish to emphasize that X "contains" x. Conversely, if X does not contain y, then we write $y \notin X$ to signify that "y is not an element of X."

If there are "few enough" distinct elements of X, then we can explicitly write down X using pointy braces. For instance, $X = \{1, 2, 3, 4, 5, 6\}$ is a finite set consisting of the first six positive integers. Unfortunately, as the number of members of X increases, such an explicit expression of X becomes cumbersome to write down; instead, we may use **set builder notation** to express a set whose members possess a closed-form. Explicitly, set builder notation exhibits an arbitrary element x of the attendant set X followed by a bar $\mid$ and a list of qualitative information about x, e.g.,

$$X = \{1, 2, 3, 4, 5, 6\} = \{x \mid x \text{ is an integer and } 1 \leq x \leq 6\}.$$

Even more, set builder notation can be used to write down infinite sets. We will henceforth fix the following notation for the natural numbers $\mathbb{Z}_{\geq 0} = \{n \mid n \text{ is a non-negative integer}\}$, the integers $\mathbb{Z} = \{n \mid n \text{ is an integer}\}$, and the rational numbers $\mathbb{Q} = \left\{\frac{a}{b} \mid a \text{ and } b \text{ are integers such that } b \neq 0\right\}$. Using the rational numbers, one can construct the real numbers $\mathbb{R} = \{x \mid x \text{ is a real number}\}$.

Like with the arithmetic of real numbers, there are mathematical operations on sets that allow us, e.g., to compare them; take their differences; and combine them. For instance, every element of $Y = \{1, 2, 3, 4, 5\}$ is also an element of $X = \{1, 2, 3, 4, 5, 6\}$, but the element 6 of X is not contained in Y. We express this by saying that Y is a **proper subset** of X: the additional modifier "proper" is used to indicate that X and Y are not the same set (because they do not have the same members). Put into symbols, we write that $Y \subsetneq X$ whenever it is true that (i.) every element of Y is also an element of X and (ii.) there exists an element of X that is not contained in Y; this can be read as "Y is contained in X, but Y does not equal X." We may also say that Y is "included in" X. One other way to indicate that Y is a (proper) subset of X is by saying that X is a (proper) **superset** of Y, in which case we write that $X \supseteq Y$ (or $X \supsetneq Y$). If we could step through the paper and look at the superset containment $X \supseteq Y$ from the other side, it would be nothing more than $Y \subseteq X$.

We introduce the **relative complement** of Y with respect to X to formalize our previous observation that 6 belongs to X but does not belong to Y. By definition, the relative complement of Y with respect to X is the set consisting of the elements of X that are not elements of Y. We use the symbolic notation $X \setminus Y = \{w \in X \mid w \notin Y\}$ to denote the relative complement of Y with respect to X, e.g., we have that $X \setminus Y = \{6\}$ in our running example. We may view the relative complement of Y with respect to X as the "set difference" of X and Y. Conversely, the two sets X and Y "overlap" in $\{1, 2, 3, 4, 5\}$ because they both contain the elements 1, 2, 3, 4, and 5. We define the **intersection** $X \cap Y = \{w \mid w \in X \text{ and } w \in Y\}$ of the sets X and Y as the set consisting of those elements that belong to both X and Y; in this case, we have that $X \cap Y = \{1, 2, 3, 4, 5\}$.

Consider next the finite sets $V = \{1, 2, 3\}$ and $W = \{4, 5, 6\}$. Because none of the elements of V belong to W and none of the element of W belong to V, the intersection of V and W does not possess any elements; it is empty! Conventionally, this is called the **empty set**, and it is denoted by $\emptyset$. Put another way, our observations thus far in this paragraph can be stated as $V \cap W = \emptyset$. We will soon see that the empty set is a proper subset of every nonempty set. Going back to our discussion of V and W, we remark that the keen reader might have already noticed that $W = X \setminus V$ and $V = X \setminus W$, i.e., every element of X lies in either V or W (but not both because there are no elements that lie in both V and W). We say in this case that the set X is the **union** of the two sets V and W, and we write $X = V \cup W$. Generally, the union of two sets X and Y is the set consisting of all objects that are either an element of X or an element of Y — that is, $X \cup Y = \{w \mid w \in X \text{ or } w \in Y\}$.

If X and Y are any sets, then one can form the **Cartesian product** of X and Y; this is the set that consists of ordered pairs (x, y) for each element $x \in X$ and $y \in Y$, i.e., the Cartesian product of X and Y is the set $X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y$. Observe that the Cartesian product $\mathbb{Z} \times \mathbb{Z}$ of the integers $\mathbb{Z}$ with itself is the collection of integer points in the Cartesian plane $\mathbb{R} \times \mathbb{R}$. We refer to a subset R of the Cartesian product $X \times X$ as a **relation** on X. Every set X admits a relation called the **diagonal** Δ_X of X that consists precisely of the elements of $X \times X$ of the form (x, x). Put another way, the diagonal of X is the relation $\Delta_X = \{(x, x) \mid x \in X\} \subseteq X \times X$.

One important consideration in the arithmetic of sets is the number of elements in a finite set X. For instance, in our previous examples, it is clear that $X = \{1, 2, 3, 4, 5, 6\}$ consists of six elements, but $Y = \{1, 2, 3, 4, 5\}$ possesses five elements. Observe that this immediately distinguishes the sets X and Y. We refer to the number of elements in a finite set X as the **cardinality** of X, denoted by $\#X$ or $|X|$. Like we previously mentioned, we have that $|X| = 6$ and $|Y| = 5$. If X and Y are finite sets with cardinalities $|X|$ and $|Y|$, then the Cartesian product $X \times Y$ has cardinality $|X| \cdot |Y|$ because an element of $X \times Y$ is uniquely determined by the ordered pair (x, y). Cardinality can be defined even for infinite sets, but additional care must be taken in this case, so we will not bother.

Like with real numbers, functions can be defined between arbitrary sets. Explicitly, a **function**

$f : X \to Y$ from a set X to a set Y is merely an assignment of each element $x \in X$ to a unique (not necessarily distinct) element $f(x) \in Y$; the **domain** of $f : X \to Y$ is X, and the **codomain** of f is Y. Out of desire for notational convenience, we may sometimes omit the letter $f : X \to Y$ when defining a function f from a set X to a set Y and simply use an arrow $X \to Y$ to indicate the sets involved and an arrow $x \mapsto f(x)$ to declare the element $f(x) \in Y$ onto which the element $x \in X$ is sent; often, this will become clearer in practice. Every set X possesses a function $\text{id}_X : X \to X$ that is called the **identity function** and defined by $\text{id}_X(x) = x$. If X is a subset of Y, then the **inclusion** $X \subseteq Y$ can be viewed as the function $X \to Y$ that sends $x \mapsto x$, where the symbol x that appears to the left of the arrow $\mapsto$ is viewed as an element of X, and the symbol x that appears to the right of the arrow $\mapsto$ is then viewed as an element of Y. Or in the usual notation, the inclusion may be thought of as the function $f : X \to Y$ defined by $f(x) = x$. Even more, every set X induces a function $\delta_X : X \to X \times X$ that is called the **diagonal function** (of X) and defined by $\delta_X(x) = (x, x)$. By Exercise 1.10.4, the diagonal Δ_X of X is exactly the image of the diagonal function δ_X of X, hence there should be no confusion in terminologies. Conversely, we say that a function $* : X \times X \to X$ that sends $(x_1, x_2) \mapsto x_1 * x_2$ is a **binary operation**; implicit in the definition of a binary operation $*$ is the requirement that $x_1 * x_2$ is an element of X for every pair of elements $x_1, x_2 \in X$. For instance, addition is a binary operation on the real number $\mathbb{R}$.

Each time we define a function $f : X \to Y$, in addition, we implicitly distinguish the collection of elements $y \in Y$ such that $y = f(x)$ for some element $x \in X$; this is called the **image** of X (in Y) with respect to f, and it is denoted by $f(X) = \{y \in Y \mid y = f(x) \text{ for some element } x \in X\}$. Conversely, if W is a subset of Y, then the collection of elements $x \in X$ such that $f(x) \in W$ is the **pre-image** of W (in X) with respect to f. Explicitly, we have that $f^{-1}(W) = \{x \in X \mid f(x) \in W\}$. We note that for any subsets $V \subseteq X$ and $W \subseteq Y$, it is always the case that $V \subseteq f^{-1}(f(V))$ and $f(f^{-1}(W)) \subseteq W$; however, the superset containments $V \supseteq f^{-1}(f(V))$ and $f(f^{-1}(W)) \supseteq W$ do not always hold (cf. Exercise 1.10.7). We introduce two properties of functions that are sufficient to guarantee that these superset inclusions. If $f(x_1) = f(x_2)$ implies that $x_1 = x_2$ for any pair of elements $x_1, x_2 \in X$, then we say that $f : X \to Y$ is **injective**. Essentially, a function $f : X \to Y$ is injective if and only if distinct elements $x_1, x_2 \in X$ induce distinct elements $f(x_1), f(x_2) \in Y$. We will soon verify this formally. Even more, we say that $f : X \to Y$ is **surjective** if for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. One way to think about the surjective property is that every element of Y is "mapped onto" or "covered" by an element of X. If $f : X \to Y$ is both injective and surjective, then we say that f is **bijective**. We may think about a bijection $f : X \to Y$ as a relabelling of the elements of Y using the names of elements of X; in this way, two sets X and Y are "essentially the same" if there exists a bijection $f : X \to Y$.

Proposition 1.1.1. *Let $f : X \to Y$ be any function between any two sets X and Y.*

1.) *If f is injective, then $f^{-1}(f(V)) = V$ for any set $V \subseteq X$.*

2.) *If f is surjective, then $f(f^{-1}(W)) = W$ for any set $W \subseteq Y$.*

Proof. 1.) By Exercise 1.10.7, it suffices to prove that $f^{-1}(f(V)) \subseteq V$. Let x be an arbitrary element of $f^{-1}(f(V))$. By definition of the pre-image $f^{-1}(f(V))$ of $f(V)$, this means that $f(x) \in f(V)$. By definition of the image $f(V)$, we have that $f(x) = f(v)$ for some element $v \in V$. Last, by assumption that f is injective and $V \subseteq X$, we conclude that $x = v$, hence x is an element of V.

(2.) By Exercise 1.10.7, it suffices to prove that $W \subseteq f(f^{-1}(W))$. Let w be any element of W. By assumption that f is surjective and $W \subseteq Y$, there exists an element $x \in X$ such that $w = f(x)$. By definition of the pre-image $f^{-1}(W)$, it follows that $x \in f^{-1}(W)$. By definition of the image $f(f^{-1}(W))$, we conclude that $w = f(x)$ for some element $x \in f^{-1}(W)$ so that $w \in f(f^{-1}(W))$. □

Conversely, if $f^{-1}(f(V)) = V$ holds for any set $V \subseteq X$, then $f : X \to Y$ must be injective; likewise, if $f(f^{-1}(W)) = W$ for any set $W \subseteq Y$, then f must be surjective (cf. Exercise 1.10.8).

1.2 Logic and Truth Tables

We have thus far garnered a working knowledge of set theory, and we have seen some mathematical proofs. We turn our attention next to fleshing out some details regarding the calculus of logic that will soon assist us with writing proofs. We will assume throughout this section that P and Q are **statements**, i.e., P and Q are complete sentences that assert some property or quality that can be unambiguously measured as true or false. For instance, "Every positive whole number is an integer" is an example of a (true) statement; however, "The weather in Kansas City is lovely this time of year" is not a statement because some individuals might think so while others might not.

We will be interested primarily in logical constructions of the form $P \implies Q$, where the double-arrow $\implies$ stands for "implies." Under this convention, the entire expression $P \implies Q$ can be read either as "P implies Q" or "If P, then Q." Unsurprisingly, statements of this form are called **implications**. We refer to P in this construction as the **antecedent** and to Q as the **consequent**. We may deduce the validity of a statement $P \implies Q$ by constructing the following **truth table**.

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

Table 1.1: the truth table for the implication $P \implies Q$

Perhaps the best way to understand the above truth table is by example. For instance, if P is the statement that "3 is an odd number" and Q is the statement that "Madrid is the capital of Spain," then $P \implies Q$ must be true because both P and Q are true statements. On the other hand, if P is false, then the implication $P \implies Q$ is true regardless of the **truth-value** (or verity) of Q; in this case, $P \implies Q$ is called a **vacuous truth**, or it is said to be vacuously true. Essentially, the idea is that P cannot be satisfied because it is false, so the implication must be true: if P is the statement that "17 is larger than 38," then $P \implies Q$ is true regardless of the statement Q. On the other hand, if the statement P is true but the statement Q is false, then the statement $P \implies Q$ must be false because the consequent is false. By example, we can verify this intuition in the case that P is the statement that "3 is an odd number" and Q is the statement that "17 is larger than 38": certainly, the statement $P \implies Q$ is false (read it aloud to convince yourself), hence the verity of the antecedent P has no bearing on $P \implies Q$ because the consequent Q is false.

Unfortunately, in some situations, it is difficult to establish the verity of a statement Q from a statement P that is known to be true. Under these circumstances, it is not possible to determine if the statement $P \implies Q$ is true or false because this depends entirely on whether Q is true or false; however, it is possible in some cases to extract a statement $S(P, Q)$ (depending upon P and Q) that is **logically equivalent** to the implication $P \implies Q$. We say that two statements S and S' are logically equivalent if and only if their values in a truth table are equal. Consequently, if we could demonstrate that the statement $S(P, Q)$ were true, then $P \implies Q$ must be true, as well.

We examine next some different ways to construct new statements from two given statements P and Q. One way to do so is by considering the case that either P or Q is true. Put into symbols, the **disjunction** $P \vee Q$ is the statement "either P or Q," for which the upside-down wedge $\vee$ denotes the connective "or." Crucially, if either P or Q is true, then $P \vee Q$ must also be true. On the other hand, we may also think about when both P and Q are true, which gives rise to the statement "both P and Q" or the **conjunction** $P \wedge Q$; this is true if and only if both P and Q are true, hence if one of P or Q is false, then $P \wedge Q$ is also false. Be careful not to confuse the upside-down wedge $\vee$ (meaning "or") with the right-side up $\wedge$ (meaning "and"). Last, the **negation** $\neg P$ of the statement P is the statement "not P." Observe that the truth-value for $\neg P$ is the opposite of the truth-value of P. Ultimately, we may construct the following truth tables for the above scenarios.

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	$\neg P$	$P \vee \neg P$	$P \wedge \neg P$
T	F	T	F
T	F	T	F
F	T	T	F
F	T	T	F

Table 1.2: the truth tables for the disjunction $P \vee Q$, conjunction $P \wedge Q$, $P \vee \neg P$, and $P \wedge \neg P$

We note that the statement $P \vee \neg P$ ("P or not P") is always true; it is a **tautology**. On the other hand, the statement $P \wedge \neg P$ is always false; it is a **self-contradiction**; this proves the following.

Theorem 1.2.1 (Law of the Excluded Middle)**.** *If P is any statement, then either P or $\neg P$ is true.*

Theorem 1.2.2 (Law of Non-Contradiction)**.** *If P is any statement, then "P and not P" is false.*

We concern ourselves next with the interplay between the disjunction, conjunction, negation, and implication. Often, it is useful in mathematics to determine when a statement $P \implies Q$ is false. Put another way, we wish to determine if P does not imply Q, i.e., if P is not sufficient information from which to deduce the verity of Q. One way to accomplish this is to prove that the consequent Q is false when the antecedent P is true; this is a valid law of inference because the statements $\neg(P \implies Q)$ and $P \wedge \neg Q$ are logically equivalent, as the following truth table bears.

P	Q	$\neg Q$	$P \implies Q$	$\neg(P \implies Q)$	$P \wedge \neg Q$
T	T	F	T	F	F
T	F	T	F	T	T
F	T	F	T	F	F
F	F	T	T	F	F

Table 1.3: the truth table for the negated implication $\neg(P \implies Q)$ and $P \wedge \neg Q$

Both of column $\neg(P \implies Q)$ and $P \wedge \neg Q$ take the same truth-values, hence these statements are logically equivalent. We will also consider the negation of the disjunction $P \vee Q$ ("P or Q") and the negation of the conjunction $P \wedge Q$ ("P and Q"). By Table 1.2, if "P or Q" is not true (i.e., its negation is true), then neither P nor Q can be true. Likewise, by the same table, if "P and Q" is not true (i.e., its negation is true), then either P must not be true or Q must not be true. Collectively, these observations constitute the so-called **De Morgan's Laws** that we prove below.

P	Q	$\neg P$	$\neg Q$	$P \vee Q$	$\neg(P \vee Q)$	$\neg P \wedge \neg Q$	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P \vee \neg Q$
T	T	F	F	T	F	F	T	F	F
T	F	F	T	T	F	F	F	T	T
F	T	T	F	T	F	F	F	T	T
F	F	T	T	F	T	T	F	T	T

Table 1.4: the truth table for $\neg(P \vee Q)$ and $\neg(P \wedge Q)$

Theorem 1.2.3 (De Morgan's Laws). *Let P and Q be any statements.*

1.) $\neg(P \vee Q)$ *is logically equivalent to* $\neg P \wedge \neg Q$.

2.) $\neg(P \wedge Q)$ *is logically equivalent to* $\neg P \vee \neg Q$.

Proof by contraposition is yet another indispensable law of inference we will employ. We say that the **contrapositive** of the implication $P \implies Q$ is the implication $\neg Q \implies \neg P$ formed by taking the implication of the negation of Q and the negation of P. For instance, suppose that P is the statement that "The sun is shining in Kansas City" and Q is the statement that "Bob rides his bike to work." Consider the implication $P \implies Q$ given by the statement, "If the sun is shining in Kansas city, then Bob rides his bike to work"; its contrapositive is the statement, "If Bob does not ride his bike to work, then the sun is not shining in Kansas City." Proof by contraposition exploits that the implications $P \implies Q$ and $\neg Q \implies \neg P$ are logically equivalent, as we can verify below.

P	Q	$P \implies Q$	$\neg Q$	$\neg P$	$\neg Q \implies \neg P$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Table 1.5: the truth table for $P \implies Q$ and its contrapositive $\neg Q \implies \neg P$

Last, the **proof by contradiction** (or *reductio ad absurdum*) rounds out the tools that we will most often use in mathematical proofs. Essentially, the proof by contradiction constitutes a valid law of inference by a combination of the Law of the Excluded Middle (which asserts that either a statement or its negation must be true), the Law of Non-Contradiction (which asserts that a statement and its negation cannot both be true), and Table 1.3 (which asserts that $\neg(P \implies Q)$ and $P \wedge \neg Q$ are logically equivalent): one of the statements $P \implies Q$ or $\neg(P \implies Q)$ must be true, hence if we can establish that $P \wedge \neg Q$ is not true, then it must be the case that $\neg(P \implies Q)$ is not true (because these two statements are logically equivalent) so that $P \implies Q$ is true. For instance, if we define an even number to be a whole number that is divisible by two, then we may appeal to a proof by contradiction to establish that twice any whole number is also even. Explicitly, suppose that P is the statement that "x is a whole number" and Q is the statement that "$2x$ is an even number." If we wish to establish the verity of the implication $P \implies Q$ given by the statement, "If x is a whole number, then $2x$ is an even number," then we may assume to the contrary that P is true (i.e., x is a whole number) and $\neg Q$ is also true (i.e., $2x$ is not an even number); all together, we are assuming $P \land \neg Q$, i.e., "x is a whole number and $2x$ is not an even number." By definition, $2x$ is an even number because it is twice a whole number, so we have arrived at a contradiction — namely, that $2x$ is an even number (by definition) and $2x$ is not an even number (by assumption). Ultimately, the statement $P \land \neg Q$ cannot be true, hence $P \implies Q$ must be true. Generally, a successful proof by contradiction begins by assuming (to the contrary) that P is true and that Q is not true; then, a contradiction of the form (a.) $P \land \neg P$ or (b.) $Q \land \neg Q$ is derived. Observe that if $\neg P$ can be deduced from $\neg Q$ (i.e., (a.) holds) then a proof by contraposition may be simpler than a proof by contradiction; on the other hand, if Q can be deduced from P (i.e., (b.) holds), then a **direct proof** may be simpler than a proof by contradiction. Bear this in mind always.

Given any two statements P and Q, we have already considered the implication $P \implies Q$ and its contrapositive $\neg Q \implies \neg P$; however, we could also consider the implication $Q \implies P$ and its contrapositive $\neg P \implies \neg Q$. We refer to the statement $Q \implies P$ as the **converse** of the implication $P \implies Q$; the statement $\neg P \implies \neg Q$ is the **inverse** of the implication $P \implies Q$. Generally, the implication is not logically equivalent to its converse, as the next truth table shows.

P	Q	$P \implies Q$	$Q \implies P$	$\neg P$	$\neg Q$	$\neg P \implies \neg Q$
T	T	T	T	F	F	T
T	F	F	T	F	T	T
F	T	T	F	T	F	F
F	F	T	T	T	T	T

Table 1.6: the truth table for $P \implies Q$, its converse $Q \implies P$, and its inverse $\neg P \implies \neg Q$

Unsurprisingly, we find that the converse $Q \implies P$ and the inverse $\neg P \implies \neg Q$ are logically equivalent because they are contrapositives of one another; however, the implication and its converse are not logically equivalent, hence the implication and its inverse are not logically equivalent.

If P implies Q, then we say that P is **sufficient** for Q or that Q is **necessary** for P. One can rephrase this by saying that P is sufficient for Q when it is true that Q holds if P holds; equivalently, we may say that Q is necessary for P when it is true that P holds only if P holds, i.e., if Q does not hold, then P does not hold. We note that if P is sufficient for Q (or Q is necessary for P), then it might not be true that P is necessary for Q (or that Q is sufficient for P) because the converse is not logically equivalent to the implication; however, if P is both necessary and sufficient for Q, then we have that $P \implies Q$ and $Q \implies P$ so that $P \iff Q$, i.e., "P if and only if Q." If this holds, then we say that P and Q are **(materially) equivalent** statements. Observe that the material equivalence $P \iff Q$ is logically equivalent to the conjunction $(P \implies Q) \land (Q \implies P)$.

Even more, **logical quantifiers** allow us to symbolically express the concepts of "for all" (or "for every") and "there exists" (or "for at least one" or "for some"). Explicitly, we adopt the **universal quantifier** $\forall$ as the symbolic representation of the phrase "for all" and the **existential quantifier** $\exists$ as the symbolic representation of the phrase "there exists." Using these quantifiers, we may convert statements involving quantities into purely symbolic expressions. For instance, that the sum of any whole number and one is a whole number can be written symbolically as $(\forall n \in \mathbb{Z})(n+1 \in \mathbb{Z})$. On the other hand, there exists a non-negative whole number whose difference with one is negative, i.e., $(\exists n \in \mathbb{Z}_{\geq 0})(n-1 \notin \mathbb{Z}_{\geq 0})$. (Explicitly, the non-negative integer $n = 0$ satisfies this property.) Observe that every real number x admits a **unique** real number y such that $x + y = 0$; using logical quantifiers yields $(\forall x \in \mathbb{R})(\exists! y \in \mathbb{R})(x + y = 0)$ with the **uniqueness quantifier** $\exists!$ signifying both the existence ($\exists$) and uniqueness (!). Put another way, the logical quantifier $\exists!$ denotes that "there exists one and only one" element with the prescribed property.

1.3 Sets and Set Operations, Revisited

Using the calculus of logic, we may deduce further properties of sets and set operations. Before proceeding to any new material, we provide first a reinterpretation of Section 1.1 in the language of Section 1.2. We will assume to this end that X and Y are arbitrary (possibly empty) sets.

- We may view the set membership $x \in X$ as the statement "x is an element of X"; its negation is the statement that "x is not an element of X" (or $x \notin X$ in symbols).

- We have that $X \subseteq Y$ ("X is a subset of Y") if and only if for every element $x \in X$, it is true that $x \in Y$, i.e., if and only if it is true that $(\forall x \in X)(x \in Y)$. Consequently, the empty set $\emptyset$ is a subset of every set: there are no elements in $\emptyset$, hence $(\forall e \in \emptyset)(e \in X)$ is vacuously true!

- If it holds that $X \subseteq Y$ and $(\exists y \in Y)(y \notin X)$ ("there exists an element $y \in Y$ such that $y \notin X$"), then we say that X is a proper subset of Y, and we write $X \subsetneq Y$; otherwise, it must be the case that $Y \subseteq X$, hence X and Y are equal, i.e., we must have that $X = Y$.

- Elements of either X or Y comprise the union $X \cup Y$ of X and Y. Put another way, we have that $X \cup Y$ is the superset of both X and Y for which $(w \in X) \vee (w \in Y)$ is true.

- Elements of both X and Y comprise the intersection $X \cap Y$ of X and Y. Put another way, we have that $X \cap Y$ is the subset of both X and Y for which $(w \in X) \wedge (w \in Y)$ is true.

- Elements in Y but not in X comprise the relative complement $Y \setminus X$ of X with respect to Y. Put another way, we have that $Y \setminus X$ is the subset of Y for which $(y \in Y) \wedge (y \notin X)$ is true.

- We may view the Cartesian product $X \times Y$ of X and Y as the collection of all ordered pairs (x, y) for which the statement $(x \in X) \wedge (y \in Y)$ is true.

We will suppose now that W is an arbitrary set for which the inclusions $X \subseteq W$ and $Y \subseteq W$ hold. We say in this case that W is our **universe**, and we may view all elements of X and Y as elements of W via the aforementioned inclusions. We obtain the following membership laws.

Theorem 1.3.1 (Law of the Excluded Middle for Sets)**.** *For any element $w \in W$, we must have that either $w \in X$ or $w \notin X$, and the analogous statement holds for Y in place of X.*

Theorem 1.3.2 (Law of Non-Contradiction for Sets)**.** *For any element $w \in W$, we cannot have that both $w \in X$ and $w \notin X$, and the analogous statement holds for Y in place of X.*

We omit the proofs of the following facts because they follow immediately from the Law of the Excluded Middle and the Law of Non-Contradiction for the statement P that "$w \in X$." Even more, there are analogous De Morgan's Laws for the relative complements of $X \cup Y$ and $X \cap Y$ in W.

Theorem 1.3.3 (De Morgan's Laws for Sets)**.** *Let $X \subseteq W$ and $Y \subseteq W$ be arbitrary sets.*

1.) *We have that $W \setminus (X \cup Y) = (W \setminus X) \cap (W \setminus Y)$.*

2.) *We have that $W \setminus (X \cap Y) = (W \setminus X) \cup (W \setminus Y)$.*

We leave the proofs of De Morgan's Laws for Sets as Exercise 1.10.14.

Often, we will deal with more sets than simply a pair; in this case, it is easiest to adopt the following notation. Let $X_1, X_2, \dots, X_n$ be arbitrary sets such that $X_i \subseteq W$ for each integer $1 \leq i \leq n$. Each set X_i is **indexed** by a subscript i for distinction. We may consider the union

$$\bigcup_{i=1}^n X_i = X_1 \cup X_2 \cup \dots \cup X_n = \{w \mid w \in X_i \text{ for some integer } 1 \leq i \leq n\}.$$

Once again, we note that the subscript i indicates the set X_i under consideration; the identification $i = 1$ beneath the union symbol indicates that we will begin with $i = 1$; and the superscript n above the union symbol indicates that we will end with $i = n$. Put another way, the elements of $\cup_{i=1}^n X_i$ are precisely those elements $w \in W$ such that $w \in X_i$ for some integer $1 \leq i \leq n$, i.e., it holds that $w \in \cup_{i=1}^n X_i$ if and only if $(\exists i \in \{1, 2, \dots, n\})(w \in X_i)$. We may also consider the intersection

$$\bigcap_{i=1}^n X_i = X_1 \cap X_2 \cap \dots \cap X_n = \{w \mid w \in X_i \text{ for all integers } 1 \leq i \leq n\}.$$

Observe that $w \in \cap_{i=1}^n X_i$ if and only if $(\forall i \in \{1, 2, \dots, n\})(w \in X_i)$. Generally, the following extension of De Morgan's Laws for Sets holds; its proof is left as Exercise 1.10.15.

Proposition 1.3.4. *Let $X_1, X_2, \dots, X_n \subseteq W$ be arbitrary sets.*

1.) *We have that $W \setminus (X_1 \cup X_2 \cup \dots \cup X_n) = (W \setminus X_1) \cap (W \setminus X_2) \cap \dots \cap (W \setminus X_n)$.*

2.) *We have that $W \setminus (X_1 \cap X_2 \cap \dots \cap X_n) = (W \setminus X_1) \cup (W \setminus X_2) \cup \dots \cup (W \setminus X_n)$.*

If $X_i \cap X_j = \emptyset$, then we say that X_i and X_j are **disjoint**. Even more, if the sets $X_1, X_2, \dots, X_n$ satisfy the condition that X_i and X_j are disjoint (i.e., $X_i \cap X_j = \emptyset$) for every pair of integers $1 \leq i < j \leq n$, then we say that $X_1, X_2, \dots, X_n$ are **pairwise disjoint** (or **mutually exclusive**). Observe that if $X_i = \emptyset$ for any integer $1 \leq i \leq n$, then $X_i \cap X_j = \emptyset$ for all integers $1 \leq j \leq n$. Consequently, we may restrict our attention to collections of nonempty pairwise disjoint sets. We say that the collection $\mathcal{P} = \{X_1, X_2, \dots, X_n\}$ forms a **partition** of the set W if and only if

(i.) X_i is nonempty for each integer $1 \leq i \leq n$;

(ii.) $W = X_1 \cup X_2 \cup \dots \cup X_n$; and

(iii.) $X_1, X_2, \dots, X_n$ are pairwise disjoint (i.e., $X_i \cap X_j = \emptyset$ for every pair of integers $1 \leq i < j \leq n$).

We note that every set W admits a trivial partition $\mathcal{W} = \{\{w\} \mid w \in W\}$ via the **singleton** sets $\{w\}$ for each element $w \in W$; however, many sets we will consider throughout this course allow for more interesting partitions. Explicitly, every integer is either odd or even; the quality of being odd or even is called the **parity** of an integer. Consequently, the integers $\mathbb{Z}$ can be partitioned via $\mathcal{P} = \{\mathbb{O}, \mathbb{E}\}$, where $\mathbb{O} = \{n \mid n \text{ is an odd integer}\}$ and $\mathbb{E} = \{n \mid n \text{ is an even integer}\}$.

Generally, a partition of an arbitrary set W need not be finite. Every property of the previous paragraph can be reformulated in the case that the **index set** I is arbitrary. Particularly, we say that an arbitrary collection $\mathcal{P} = \{X_i \mid i \in I\}$ form a partition of W if and only if

(i.) X_i is nonempty for each index $i \in I$;

(ii.) $W = \cup_{i \in I} X_i$; and

(iii.) the sets X_i are pairwise disjoint (i.e., $X_i \cap X_j = \emptyset$ for every pair of distinct indices $i, j \in I$).

1.4 Equivalence Relations and Partial Orders

We will continue to assume that X is an arbitrary set. Recall that a relation on X is by definition a subset R of the Cartesian product $X \times X$. We say that R is **reflexive** if and only if $(x, x) \in R$ for all elements $x \in X$ if and only if R contains the diagonal Δ_X of X (i.e., $R \supseteq \Delta_X$). Even more, if it holds that $(x_2, x_1) \in R$ whenever $(x_1, x_2) \in R$, then R is **symmetric**. Last, if $(x_1, x_2) \in R$ and $(x_2, x_3) \in R$ together imply that $(x_1, x_3) \in R$, then we refer to the relation R as **transitive**. Relations that are reflexive, symmetric, and transitive are distinguished as **equivalence relations**. Every set admits at least one equivalence relation, as our next proposition illustrates.

Proposition 1.4.1. *If X is an any set, the diagonal Δ_X of X is an equivalence relation on X.*

Essentially, as an equivalence relation on X, the diagonal of X captures equality of the elements of X: if $(x_1, x_2) \in \Delta_X$, then we must have that $x_1 = x_2$, and if $x_1 = x_2$, then $(x_1, x_2) \in \Delta_X$.

We shall soon discover that there are many objects on which it is fruitful to consider certain equivalence relations. Classically, the rational numbers $\mathbb{Q}$ are constructed by defining an equivalence relation on the integers $\mathbb{Z}$. Before we prove this, let us try an example of a different flavor.

Example 1.4.2. Consider the collection $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ whose first derivatives $f'(x)$ are continuous for all real numbers x. Let E denote the relation on $\mathcal{C}^1(\mathbb{R})$ defined by $(f, g) \in E$ if and only if $f'(x) = g'(x)$ for all real numbers x. Because E is defined by equality and equality is reflexive, symmetric, and transitive, it follows that E is an equivalence relation on $\mathcal{C}^1(\mathbb{R})$.

Let E denote an equivalence relation on an arbitrary set X. Often, it is convenient to adopt the notation that $x_1 \sim_E x_2$ if and only if $(x_1, x_2) \in E$, in which case we may also say that x_1 and x_2 are **equivalent modulo** E. (We note that this convention is due to Carl Friedrich Gauss; it can be understood as asserting that x_1 and x_2 are "the same except for differences accounted for by E.") We define the **equivalence class** of an element $x_0 \in X$ as the collection of elements $x \in X$ that are equivalent to x_0 modulo E, i.e., $[x_0] = \{x \in X \mid x \sim_E x_0\} = \{x \in X \mid (x, x_0) \in E\}$.

Example 1.4.3. Consider the equivalence relation E defined on the set $\mathcal{C}^1(\mathbb{R})$ of Example 1.4.2. By the Fundamental Theorem of Calculus, if $f'(x) = g'(x)$, then there exists a real number C such that $f(x) = g(x) + C$. Conversely, if $f(x) = g(x) + C$ for some real number C, then $f'(x) = g'(x)$. We conclude that the equivalence classes of $\mathcal{C}^1(\mathbb{R})$ modulo E are given precisely by the sets

$$[g] = \{f \in \mathcal{C}^1(\mathbb{R}) \mid f(x) = g(x) + C \text{ for some real number } C\}.$$

Our next proposition illustrates that a pair of equivalence classes of X modulo E are either equal or disjoint; as a corollary, we obtain a relationship between equivalence relations and partitions.

Proposition 1.4.4. *Let E denote an equivalence relation on an arbitrary set X. Every pair of equivalence classes of X modulo E are either equal or disjoint.*

Proof. Consider any pair $[x_1]$ and $[x_2]$ of equivalence classes of X modulo E. By Exercise 1.10.10, it suffices to prove that $[x_1] = [x_2]$ if they are not disjoint. Consequently, we may assume that there exists an element $x \in [x_1] \cap [x_2]$. By definition, this means that $(x, x_1) \in E$ and $(x, x_2) \in E$. By assumption that E is an equivalence relation, it follows that $(x_1, x) \in E$ by symmetry, hence the transitivity of E implies that $(x_1, x_2) \in E$. Given any element $x_0 \in [x_1]$, we have that $(x_0, x_1) \in E$ implies that $(x_0, x_2) \in E$ by transitivity, hence we conclude that $[x_1] \subseteq [x_2]$. Likewise, the symmetry of E implies that $(x_2, x_1) \in E$, hence the same argument as the previous sentences shows that $[x_2] \subseteq [x_1]$. Combined, the containments $[x_1] \subseteq [x_2]$ and $[x_2] \subseteq [x_1]$ yields that $[x_1] = [x_2]$. □

Corollary 1.4.5. *Let X be an arbitrary set. Every equivalence relation on X induces a partition of X. Conversely, every partition of X induces an equivalence relation on X.*

Proof. By Proposition 1.4.4, if E is an equivalence relation on X, then the collection $\mathcal{P}$ of distinct equivalence classes of X modulo E is pairwise disjoint. Even more, every equivalence class of X modulo E is nonempty because E is reflexive. Last, every element of X belongs to some equivalence class of X modulo E, hence we have that X is the union of its distinct equivalence classes.

Conversely, suppose that $\mathcal{P} = \{X_i \mid i \in I\}$ is a partition of X indexed by I. Consider the relation $E_{\mathcal{P}} = \{(x_1, x_2) \mid x_1, x_2 \in X_i \text{ for some index } i \in I\} \subseteq X \times X$. By definition of a partition, every element $x \in X$ lies in X_i for some index $i \in I$, hence $(x, x) \in E_{\mathcal{P}}$ for every element $x \in X$, i.e., $E_{\mathcal{P}}$ is reflexive. By definition of $E_{\mathcal{P}}$, if $(x_1, x_2) \in E_{\mathcal{P}}$, then $(x_2, x_1) \in E_{\mathcal{P}}$, hence $E_{\mathcal{P}}$ is symmetric. Last, if $(x_1, x_2), (x_2, x_3) \in E_{\mathcal{P}}$, then $x_1, x_2 \in X_i$ and $x_2, x_3 \in X_j$ for some indices $i, j \in I$. By definition of a partition, we have that $X_i \cap X_j = \emptyset$ if and only if i and j are distinct, hence we must have that $i = j$ by assumption that $x_2 \in X_i \cap X_j$. We conclude that $(x_1, x_3) \in X_i$ so that $(x_1, x_3) \in E_{\mathcal{P}}$, i.e., $E_{\mathcal{P}}$ is transitive. Ultimately, this shows that $E_{\mathcal{P}}$ is an equivalence relation on X. □

We say that a relation R on an arbitrary set X is **antisymmetric** if for every pair of elements $x_1, x_2 \in X$, the inclusions $(x_1, x_2) \in R$ and $(x_2, x_1) \in R$ together imply that $x_1 = x_2$. Equivalence relations are defined as reflexive, symmetric, and transitive relations on a set; however, if we replace the requirement of symmetry with the condition of antisymmetry, then we obtain a **partial order**. Explicitly, a partial order P on X is a subset $P \subseteq X \times X$ that is reflexive, antisymmetric, and transitive. Every set admits at least one partial order. Once again, it is simply the diagonal.

Proposition 1.4.6. *If X is any set, the diagonal Δ_X of X is a partial order on X.*

Like with equivalence relations, there are interesting examples of partial orders.

Example 1.4.7. The real numbers $\mathbb{R}$ are partially ordered via the usual less-than-or-equal-to $\leq$.

Example 1.4.8. Divisibility constitutes a partial order on the non-negative integers $\mathbb{Z}_{\geq 0}$. Explicitly, consider the relation $D = \{(a, b) \mid a \text{ divides } b\} \subseteq \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$. Observe that a divides a, hence D is reflexive. Even more, if a divides b and b divides a, then there exist integers m and n such that $b = am$ and $a = bn$; together, these identities yield that $a = bn = amn$. Certainly, if $a = 0$, then $b = 0$, hence we may assume that a is nonzero. Cancelling a factor of a from both sides gives that $mn = 1$, which in turn implies that $m = n = 1$ because a and b are non-negative. Ultimately, this proves that $a = b$, hence D is antisymmetric. Last, if a divides b and b divides c, then a divides c.

Every set admits a partial order, hence every set is a **partially ordered set**; however, there can be many ways to view a set as a partially ordered set. We say that a pair of elements p and q of a partial order P on a set X are **comparable** if it holds that either $(p, q) \in P$ or $(q, p) \in P$; otherwise, the elements p and q are said to be **incomparable**. Every pair of prime integers are incomparable with respect to the partial order of divisibility on the non-negative integers. Conversely, if every pair of elements $p, q \in P$ are comparable, then P is a **total order** on X. Observe that if $Y \subseteq X$, then we may define a partial order $P|_Y = \{(y_1, y_2) \in Y \times Y \mid (y_1, y_2) \in P\}$ on Y by viewing the elements of Y as elements of X. If $P|_Y$ is a total order on $Y \subseteq X$, then we say that Y is a **chain** (with respect to P) in X. We say that an element $x_0 \in X$ is an **upper bound** of Y (with respect to P) if it holds that $(y, x) \in P$ for every element $y \in Y$. We will also say that an element $x_0 \in X$ is **maximal** (with respect to P) if it does not hold that $(x_0, x) \in P$ for any element $x \in X \setminus \{x_0\}$. Our next theorem combines each of these ingredients to comprise one of the most ubiquitous results in mathematics; it will hold vital importance in our study of (two-sided) ideals of unital rings.

Theorem 1.4.9 (Zorn's Lemma). *Let X be an arbitrary set. Let P be a partial order on X. If every chain Y in X has an upper bound in Y, then Y admits a maximal element $y_0 \in Y$.*

1.5 The Principle of Mathematical Induction

One of the most useful proof techniques is the **Principle of Mathematical Induction**. We say that a subset S of real numbers is **hereditary** if it holds that $x + 1 \in S$ whenever we have that $x \in S$. Basically, the Principle of Mathematical Induction is a property of the non-negative integers that asserts that if S is any hereditary subset of non-negative integers such that the smallest element n_0 of S satisfies a statement $P(n_0)$ involving n_0, then every element n of S satisfies the statement $P(n)$. Before we proceed to the definition of the Principle of Mathematical Induction, let us see some examples of properties of integers for which a proof by induction is appropriate.

Example 1.5.1. Consider the positive integer $o(n) = \sum_{i=0}^{n-1}(2i + 1) = 1 + 3 + 5 + \cdots + (2n - 1)$, i.e., the sum of the first n consecutive odd positive integers. We may compute $o(n)$ for small values of n. Explicitly, we have that $o(1) = 1$ and $o(2) = 1 + 3 = 4$ and $o(3) = 1 + 3 + 5 = 9$ and so on.

n	1	2	3	4	5
$o(n)$	1	4	9	16	25

Table 1.7: the sum of first n consecutive odd positive integers

Observe that $o(n) = n^2$ for each integer $1 \leq n \leq 5$. Continuing with the table, we would find that $o(n) = n^2$ for all integers $1 \leq n \leq k$ for any positive integer k. Consequently, we have the following.

Conjecture 1.5.2. If $o(n)$ is defined as in Example 1.5.1, then $o(n) = n^2$ for all integers $n \geq 1$.

Observe that $o(1) = 1 = 1^2$ and $o(n+1) = \sum_{i=0}^{n}(2i+1) = \sum_{i=0}^{n-1}(2i+1)+(2n+1) = o(n)+(2n+1)$, hence if we knew that $o(n) = n^2$, then we could conclude that $o(n+1) = n^2 + 2n + 1 = (n+1)^2$. We will soon return to validate this idea: it is precisely the Principle of Mathematical Induction!

Example 1.5.3. Consider the positive integer $c(n) = \sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n$, i.e., the sum of the first n consecutive positive integers. We may compute $c(n)$ for small values of n. Explicitly, we have that $c(1) = 1$ and $c(2) = 1 + 2 = 3$ and $c(3) = 1 + 2 + 3 = 6$ and so on.

n	1	2	3	4	5
$c(n)$	1	3	6	10	15

Table 1.8: the sum of the first n consecutive positive integers

Even though it is not nearly as obvious as the pattern from Example 1.5.1, one can verify that $c(n) = \frac{n(n+1)}{2}$ for each integer $1 \leq n \leq 5$. Continuing with the table, we would find that $c(n) = \frac{n(n+1)}{2}$ for all integers $1 \leq n \leq k$ for any positive integer k. Consequently, we have the following.

Conjecture 1.5.4. If $c(n)$ is defined as in Example 1.5.3, then $c(n) = \frac{n(n+1)}{2}$ for all integers $n \geq 1$.

Like before, we have that $c(1) = 1 = \frac{1 \cdot 2}{2}$ and $c(n+1) = \sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n+1) = c(n) + (n+1)$, hence if we knew that $c(n) = \frac{n(n+1)}{2}$, then we could conclude that

$$c(n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$

Definition 1.5.5 (Principle of Ordinary Induction)**.** Let $P(n)$ be any statement involving a non-negative integer n. If the following hold, then $P(n)$ holds for all non-negative integers n.

(i.) $P(0)$ is a true statement.

(ii.) $P(k+1)$ is a true statement whenever $P(k)$ is a true statement for some integer $k \geq 1$.

Remark 1.5.6. Be aware that we have taken the Principle of Ordinary Induction as an axiom in our set theory; however, some authors prefer to prove it as a corollary by first *defining* the non-negative integers $\mathbb{Z}_{\geq 0}$ as the intersection of all hereditary subsets of $\mathbb{R}$ that contain 0 (cf. [DW00, Definition 3.5]). Put another way, we may define $\mathbb{Z}_{\geq 0}$ as the intersection of all sets $S \subseteq \mathbb{R}$ such that

(a.) $0 \in S$ and

(b.) if $s \in S$, then $s + 1 \in S$.

Using this axiom, the Principle of Ordinary Induction can be established by proving that the set $S = \{n \in \mathbb{Z}_{\geq 0} \mid P(n) \text{ is a true statement}\}$ is simply $\mathbb{Z}_{\geq 0}$. But this is clear: by definition of S, if $P(0)$ is a true statement, then $0 \in S$; likewise, if $n \in S$, then $P(n)$ is a true statement, hence $P(n+1)$ is a true statement, i.e., $n + 1 \in S$. Combined, these observations illustrate that S is a hereditary subset of $\mathbb{R}$ that contains 0, i.e., $S \supseteq \mathbb{Z}_{\geq 0}$. By definition of S, we have also that $S \subseteq \mathbb{Z}_{\geq 0}$.

By the Principle of Ordinary Induction, we can return to prove Conjectures 1.5.2 and 1.5.4; we leave these as the respective Exercises 1.10.21 and 1.10.22 for the reader. Occasionally, it is desirable to strengthen the hypotheses of the Principle of Ordinary Induction in order to simplify proofs involving induction. Currently, we may view induction as a property of falling dominoes: (a.) if the 0th domino falls and (b.) the nth domino falling causes the $(n+1)$th domino to fall, then all dominoes indexed by the non-negative integers will fall. But suppose that we could knock down all dominoes from the first to the nth domino: this would provide even more power with which to knock down the $(n+1)$th domino! We introduce this as the following.

Definition 1.5.7 (Principle of Complete Induction)**.** Let $P(n)$ be any statement involving a non-negative integer n. If the following hold, then $P(n)$ holds for all non-negative integers n.

(i.) $P(0)$ is a true statement.

(ii.) $P(k+1)$ is a true statement whenever $P(j)$ is a true statement for all integers $1 \leq j \leq k$.

Even though the hypotheses of the Principle of Complete Induction appear to be stronger than the Principle of Ordinary Induction, the two are in fact equivalent to one another (cf. Exercise 1.10.25); together, they are the Principle of Mathematical Induction. Using complete induction, we may obtain another ubiquitous mathematical tool that will prove crucial in our future endeavors.

Theorem 1.5.8 (Well-Ordering Principle)**.** *Every nonempty set of non-negative integers admits a smallest element with respect to the total order $\leq$. Put another way, if $S \subseteq \mathbb{Z}_{\geq 0}$ is a nonempty set, then there exists an element $s_0 \in S$ such that $s_0 \leq s$ for all elements $s \in S$.*

Proof. We will establish the contrapositive, i.e., we will prove that if $S \subseteq \mathbb{Z}_{\geq 0}$ has the property that for every element $s \in S$, there exists an element $s_0 \in S$ such that $s_0 < s$, then S must be empty. Let $P(n)$ be the statement that $n \notin S$. We claim that $P(n)$ holds for all integers $n \geq 0$. We proceed by the Principle of Complete Induction. Observe that if $0 \in S$, then there exists an element $s_0 \in S$ such that $s_0 < 0$. But this is not possible because S consists of non-negative integers. Consequently, we must have that $0 \notin S$, hence $P(0)$ is true. We will assume according to the Principle of Complete Induction that $P(k)$ holds for each integer $1 \leq k \leq n$. By definition, this means that $k \notin S$ for any integer $1 \leq k \leq n$. Observe that if $n+1 \in S$, then there exists an integer $s_0 \in S$ such that $1 \leq s_0 \leq n$. But this is not possible by the hypothesis of our induction. Consequently, we must have that $n+1 \notin S$, i.e., $P(n+1)$ is a true statement whenever $P(k)$ is a true statement for each integer $1 \leq k \leq n$. By the Principle of Complete Mathematical Induction, our proof is complete. □

Conversely, the Well-Ordering Principle implies the Principle of Ordinary Induction, hence it is equivalent to both ordinary induction and complete induction (cf. Exercise 1.10.26).

1.6 The Division Algorithm

Even as early as grade school, we learn the process of dividing one integer by another. Each time we divide an integer a by a nonzero integer b, we obtain an integer q and a non-negative integer r that is strictly smaller than $|b|$ such that $a = qb + r$. Explicitly, we say that a is the **dividend**; b is the **divisor**; q is the **quotient**; and r is the **remainder** of the division. Our aim throughout this section is to establish that this process is well-founded, i.e., the process of division of an integer a by a nonzero integer b always results in *unique* integers q and r such that $a = qb + r$ and $0 \leq r < |b|$. We will also establish an algorithm that will allow us to efficiently find the integers q and r.

Example 1.6.1. Consider the case that $a = 11$ and $b = 2$. One can easily see that $11 = 5 \cdot 2 + 1$, hence the integers $q = 5$ and $r = 1$ satisfy the requirements that $a = qb + r$ and $0 \leq r < |b|$.

Example 1.6.2. Consider the case that $a = -17$ and $b = 6$. One can easily see that $-17 = -3 \cdot 6 + 1$, hence the integers $q = -3$ and $r = 1$ satisfy the requirements that $a = qb + r$ and $0 \leq r < |b|$.

Example 1.6.3. Consider the case that $a = -8$ and $b = -9$. One can easily see that $-8 = 1(-9) + 1$, hence the integers $q = 1$ and $r = 1$ satisfy the requirements that $a = qb + r$ and $0 \leq r < |b|$.

Each of the previous examples can be completed by noticing that the integer multiples of b are completely determined by b. Consequently, we may consider all integer multiples of b that do not exceed a, i.e., we may consider the collection $D(a, b) = \{a - qb \mid q \text{ is an integer and } a \geq qb\}$. Our idea is to find the largest (in absolute value) integer q such that $a \geq qb$; then, the difference $a - qb$ must be non-negative (by assumption) and strictly smaller than b (otherwise, we could increase q). Using this intuition as our guide, let us return to find $D(a, b)$ in our previous examples.

Example 1.6.4. By definition, we have that $D(11, 2) = \{11 - 2q \mid q \text{ is an integer and } 11 \geq 2q\}$. Observe that $11 \geq 2q$ if and only if $q \leq 11/2$, hence the only valid values of q in $D(11, 2)$ are $q \leq 5$. Consequently, we have that $-2q \geq -10$ so that $11 - 2q \geq 1$. By consecutively decreasing the value of $q \leq 5$, we find that $D(11, 2) = \{1, 3, 5, 7, \dots\}$ consists of all odd positive integers.

Example 1.6.5. We have that $D(-17, 6) = \{-17 - 6q \mid q \text{ is an integer and } -17 \geq 6q\}$. Observe that $-17 \geq 6q$ if and only if $q \leq -17/6$, hence the only valid values of q in $D(-17, 6)$ are $q \leq -3$. Consequently, we conclude that $D(-17, 6) = \{-17 - 6q \mid q \leq -3 \text{ is an integer}\} = \{1, 7, 13, 19, \dots\}$.

Example 1.6.6. We have that $D(-8, -9) = \{-8 + 9q \mid q \text{ is an integer and } -8 \geq -9q\}$. Observe that $-8 \geq -9q$ if and only if $q \geq 8/9$, hence the only valid values of q in $D(-8, -9)$ are $q \geq 1$. Consequently, we conclude that $D(-8, -9) = \{-8 + 9q \mid q \geq 1 \text{ is an integer}\} = \{1, 10, 19, 28, \dots\}$.

Generalizing the collection $D(a, b)$ and using the Well-Ordering Principle yields the following.

Theorem 1.6.7 (Division Algorithm)**.** *Let a be any integer, and let b be any nonzero integer. There exist* unique *integers q and r such that $a = qb + r$ and $0 \leq r < |b|$.*

Proof. Consider the collection $D(a, b) = \{a - qb \mid q \text{ is an integer and } a \geq qb\}$. By definition, $D(a, b)$ consists of non-negative integers. Observe that if $a \geq 0$, then $D(a, b)$ is nonempty because we may take $q = 0$ to conclude that $D(a, b)$ contains a. On the other hand, if $a < 0$, then if $b \geq 1$, we conclude that $D(a, b)$ is nonempty because we may take $q = a - 1$ to find that $D(a, b)$ contains $a - qb$ because $a \geq a - 1 \geq (a - 1)b = qb$. Last, if $a < 0$ and $b \leq -1$, then $D(a, b)$ must once again be nonempty because we may take $q = -(a - 1)$ to find that $D(a, b)$ contains $a - qb$ because $a \geq a - 1 \geq -(a - 1)b = qb$. Ultimately, this shows that $D(a, b)$ is a nonempty subset of non-negative integers, hence the Well-Ordering Principle implies that there exists a smallest element $r(a, b) = a - qb$ with respect to the total order $\leq$. Rearranging this identity and rewriting $r(a, b)$ as r yields that $a = qb + r$. Clearly, it follows that $r \geq 0$, hence it suffices to see that $r < |b|$. On the contrary, suppose that $a - bq = r \geq |b|$. Observe that if $b \geq 1$, then $|b| = b$ yields that $a - qb \geq b$ and $a - (q + 1)b \geq 0$. Considering that $a - (q + 1)b$ is smaller than the smallest element $r(a, b) = a - qb$

of $D(a,b)$, we obtain a contradiction. Likewise, if $b \leq -1$, then $|b| = -b$ implies that $a - qb \geq b$ and $a-(q-1)b \geq 0$. Considering that $b \leq -1$, we find that $a-(q-1)b = a - qb + b < a - qb = r(a,b)$. Once again, this contradicts the fact that $r(a,b)$ is the smallest element of $D(a,b)$. Ultimately, we conclude that there exist integers q and r such that $a = qb + r$ and $0 \leq r < |b|$.

We must prove next that these integers are *unique*. We accomplish this by assuming that there exist integers q' and r' such that $a = q'b + r'$ and $0 \leq r' < |b|$. Considering that $a = qb + r$ by the previous paragraph, we conclude that $qb + r = q'b + r'$ so that $b(q-q') = r' - r$. Observe that if $q' = q$, then it is clear that $r' = r$, hence our proof is complete. Consequently, we may assume on the contrary that $q - q'$ is nonzero, hence we must have that $|b| \leq |r'-r|$. Observe that if $r' > r$, then $|r'-r| = r' - r$ implies that $r' \geq |b| + r \geq |b|$ — a contradiction. Likewise, if $r' < r$, then $|r'-r| = r - r'$ implies that $r \geq |b| + r' \geq |b|$ — a contradiction. Either way, we conclude that $r' = r$ so that $b(q-q') = 0$. By hypothesis that b is nonzero, we conclude that $q - q' = 0$ or $q' = q$. □

We have therefore rigorously verified the method of division we have taken for granted since elementary school! Even though the Division Algorithm does not explicitly provide the steps to compute the unique integers q and r such that $a = qb + r$ and $0 \leq r < |b|$, we note that the proof is constructive in the sense that the unique integers q and $0 \leq r < |b|$ can be deduced from the collection $D(a,b) = \{a - qb \mid q \text{ is an integer and } a \geq qb\}$, as we have done in previous examples.

If the Division Algorithm produces a remainder of zero, then we will say that b **divides** a, and we will write that $b \mid a$. Put another way, we have that $b \mid a$ if and only if $a = qb$ for some integer q. If c is any nonzero integer such that $c \mid a$ and $c \mid b$, then we say that c is a **common divisor** of a and b; the **greatest common divisor** of a and b is the unique integer $d = \gcd(a,b)$ such that

(a.) $d \mid a$ and $d \mid b$, i.e., d is a common divisor of a and b and

(b.) if d' is any common divisor of a and b, then $d' \mid d$.

Example 1.6.8. Consider the integers $a = 24$ and $b = 16$. By writing down the prime factorizations of a and b, their greatest common divisor can easily be read off. Observe that $24 = 4 \cdot 6 = 2^3 \cdot 3$ and $16 = 4^2 = 2^4$. Consequently, the greatest common divisor of 24 and 16 is 2^3, i.e., $\gcd(24,16) = 8$.

Generally, for any nonzero integers a and b, we may determine $\gcd(a,b)$ from the prime factorizations of a and b in the same manner as Example 1.6.8 (cf. Exercise 1.10.32).

Certainly, it is possible that $\gcd(a,b) = 1$, e.g., if both a and b are prime numbers. Generalizing this notion, we say that a and b are **relatively prime** if and only if $\gcd(a,b) = 1$. Our next lemma states that $\gcd(a,b)$ can always be realized as an integer-linear combination of a and b.

Lemma 1.6.9 (Bézout's Identity)**.** *If a and b be are nonzero integers, then there exist integers x and y such that $\gcd(a,b) = ax + by$. Even more, $\gcd(a,b)$ divides $av + bw$ for all integers v and w.*

Proof. Consider the collection $L(a,b) = \{ax + by \mid x, y \text{ are integers and } ax + by \geq 1\}$. Considering the sign of a and b, one of the elements $a+b$, $a-b$, $-a+b$, or $-a-b$ must lie in $L(a,b)$, hence it is nonempty. By the Well-Ordering Principle, there exists a smallest element $d(a,b) = ax + by$ with respect to the total order $\leq$. We will establish that $\gcd(a,b) = d(a,b)$.

By the Division Algorithm, there exist unique integers q_a and r_a such that $a = q_a d(a,b) + r_a$ and $0 \leq r_a < d(a,b)$. By rearranging this identity and using that $d(a,b) = ax + by$, we find that

$$r_a = a - q_a d(a,b) = a - q_a(ax + by) = (1 - q_a x)a - (q_a y)b.$$

Observe that if r_a were nonzero, then it would lie in $L(a,b)$ and satisfy $1 \leq r_a < d(a,b)$, but this is impossible because $d(a,b)$ is the smallest element of $L(a,b)$. Consequently, it must be the case that $r_a = 0$. Likewise, the Division Algorithm with b in place of a yields that $d(a,b)$ divides b. Ultimately, this proves that $d(a,b) \mid a$ and $d(a,b) \mid b$, hence $d(a,b)$ is a common divisor of both a and b.

Consider another common divisor d' of a and b. We must prove that $d' \mid d(a,b)$. By assumption, there exist integers q_a and q_b such that $a = q_a d'$ and $b = q_b d'$, from which it follows that

$$d(a,b) = ax + by = (q_a d')x + (q_b d')y = (q_a x + q_b y)d'.$$

By definition, this implies that d' divides $d(a,b)$ so that $\gcd(a,b) = d(a,b) = ax + by$, as desired.

Last, let v and w be any integers. By the previous two paragraphs, there exist integers q_a and q_b such that $a = q_a \gcd(a,b)$ and $b = q_b \gcd(a,b)$, hence $\gcd(a,b)$ divides $av + bw$. □

Corollary 1.6.10. *If a and b are relatively prime, then $ax + by = 1$ for some integers x and y.*

Corollary 1.6.11. *If a and b are nonzero integers, then $\gcd(a,b)$ is unique.*

Proof. By the proof of Bézout's Identity, we conclude that $\gcd(a,b)$ is unique because it is by construction the smallest (with respect to $\leq$) nonzero integer satisfying some property. □

Even though Bézout's Identity guarantees the existence of integers x and y such that we may write $\gcd(a,b) = ax + by$, it does not provide any tools for explicitly finding these integers x and y.

Example 1.6.12. Consider the case that $a = 24$ and $b = 16$. We know already that $\gcd(a,b) = 8$, and it is not difficult to see that $8 = 24 \cdot 1 + 16(-1)$; however, this can also be seen as follows. By the Division Algorithm, we have that $24 = 1 \cdot 16 + 8$, hence we have that $8 = 24 \cdot 1 + 16(-1)$.

Example 1.6.13. Consider the case that $a = 110$ and $b = 24$. Observe that the unique prime factorizations of 110 and 15 are $110 = 10 \cdot 11 = 2 \cdot 5 \cdot 11$ and $24 = 2^3 \cdot 3$, respectively. By Exercise 1.10.32, it follows that $\gcd(110, 15) = 2$. By successively implementing the Division Algorithm, we may find the integers x and y such that $110x + 24y = 2$, as guaranteed to us by Bézout's Identity. Explicitly, we begin by running the Division Algorithm with $a = 110$ and $b = 24$ to find the unique integers q_1 and $0 \leq r_1 < 24$ such that $110 = 24q_1 + r_1$; then, we run the Division Algorithm with 24 and r_1 to produce the unique integers q_2 and $0 \leq r_2 < r_1$ such that $24 = q_2 r_1 + r_2$. Continuing in this manner produces a strictly decreasing sequence $r_1 > r_2 > \cdots > r_n$ of non-negative integers at the nth step; by the Well-Ordering Principle, this sequence must have a least element, hence the process must eventually terminate. Putting this process to the test, we find that

$$\begin{aligned} 110 &= 4 \cdot 24 + 14, \\ 24 &= 1 \cdot 14 + 10, \\ 14 &= 1 \cdot 10 + 4, \text{ and} \\ 10 &= 2 \cdot 4 + 2. \end{aligned}$$

We find the integers x and y such that $110x + 24y = 2$ by unravelling this process in reverse. Explicitly, our last identity yields that $10 - 2 \cdot 4 = 2$; the identity before that yields that $4 = 14 - 1 \cdot 10$, hence we have that $-2 \cdot 14 + 3 \cdot 10 = 10 - 2 \cdot (14 - 1 \cdot 10) = 2$; the identity before $14 = 1 \cdot 10 + 4$

yields that $10 = 24 - 1 \cdot 14$, hence we have that $3 \cdot 24 - 5 \cdot 14 = -2 \cdot 14 + 3 \cdot (24 - 1 \cdot 14) = 2$; and at last, the identity before $24 = 1 \cdot 14 + 10$ yields that $14 = 110 - 4 \cdot 24$, hence we have that

$$110(-5) + 24(23) = 3 \cdot 24 - 5 \cdot (110 - 4 \cdot 24) = 2.$$

Algorithm 1.6.14 (Euclidean Algorithm)**.** Let a and b be any nonzero integers such that $a \geq b$.

1.) Use the Division Algorithm to find integers q_1 and r_1 such that $a = q_1 b + r_1$ and $0 \leq r_1 < |b|$.

2.) Use the Division Algorithm to find integers q_2 and r_2 such that $b = q_2 r_1 + r_2$ and $0 \leq r_2 < r_1$.

3.) Use the Division Algorithm to find integers q_3 and r_3 such that $r_1 = q_3 r_2 + r_3$ and $0 \leq r_3 < r_2$.

4.) Continue in this manner until r_{n+1} divides r_n. By the Well-Ordering Principle, this must eventually occur, and moreover, it must occur in a finite number of steps.

5.) Use the fact that $r_{n-1} = q_{n+1} r_n + r_{n+1}$ to express that $r_{n+1} = r_{n-1} - q_{n+1} r_n$.

6.) Use the fact that $r_{n-2} = q_n r_{n-1} + r_n$ to express that $r_n = r_{n-2} - q_n r_{n-1}$; then, use the fact that $r_{n+1} = r_{n-1} - q_{n+1} r_n$ to express that $r_{n+1} = r_{n-1} - q_{n+1}(r_{n-2} - q_n r_{n-1})$ so that

$$r_{n+1} = (q_n q_{n+1} + 1) r_{n-1} - q_{n+1} r_{n-2}.$$

7.) Continue in this manner to produce integers x and y such that $r_{n+1} = ax + by$.

By Bézout's Identity, we must have that $\gcd(a, b) \leq r_{n+1}$. Conversely, because r_{n+1} divides r_n by step four, it must divide r_k for all integers $1 \leq k \leq n$ by the fifth through seventh steps above. Consequently, by the second step above, we conclude that r_{n+1} must divide b, and by the first step above, we conclude that r_{n+1} must divide a. Ultimately, this shows that r_{n+1} is a common divisor of a and b, hence we must have that r_{n+1} divides $\gcd(a, b)$; in particular, we have that $r_{n+1} = \gcd(a, b)$.

1.7 The Integers Modulo n

We will assume throughout this section that n is a fixed nonzero integer. By the Division Algorithm, for every integer a, there exist unique integers q_a and r_a such that $a = q_a n + r_a$ and $0 \leq r_a < |n|$. Considering that the remainder r_a of the division of a by n is always a non-negative integer, we may assume without loss of generality that n is a positive integer. We will refer to the unique integer r_a as the remainder of a **modulo** n. Our naming convention is justified by the next proposition.

Proposition 1.7.1. *If $\mathbb{Z}$ is the set of integers, then $R_n = \{(a, r) \mid a = qn + r \text{ for some integer } q\}$ is an equivalence relation on $\mathbb{Z}$ with distinct equivalence classes $\{qn + r \mid q \in \mathbb{Z}\}$ for each integer $0 \leq r \leq n - 1$. Explicitly, the equivalence class of a modulo n is given by $[a] = \{qn + r_a \mid q \in \mathbb{Z}\}$.*

Proof. By definition, we must justify that R_n is (i.) reflexive, (ii.) symmetric, and (iii.) transitive.

(i.) Clearly, the pair (a, a) lies in R_n because we may always write $a = 0 \cdot n + a$ for any integer a.

(ii.) We must next show that if $(a, r) \in R_n$, then $(r, a) \in R_n$. By definition of R_n, if we assume that $(a, r) \in R_n$, then there exists an integer q such that $a = qn + r$. Consequently, the integer $-q$ satisfies that $r = -qn + a = (-q)n + a$, and we conclude that $(r, a) \in R_n$.

(iii.) Last, we will assume that $(a, r) \in R_n$ and $(r, s) \in R_n$. By definition of R_n, there exist integers q and q' such that $a = qn + r$ and $r = q'n + s$. Consequently, we have that $(a, s) \in R_n$ because

$$a = qn + r = qn + (q'n + s) = (q + q')n + s,$$

and the sum $q + q'$ of the two integers q and q' is itself an integer.

We have therefore established that R_n is an equivalence relation on $\mathbb{Z}$; the equivalence class of an arbitrary integer a modulo R_n is defined by $[a] = \{r \in \mathbb{Z} \mid a = qn + r \text{ for some integer } q\}$. By the Division Algorithm, for every integer a, there exist unique integers q_a and r_a such that $a = q_a n + r_a$ and $0 \leq r_a \leq n - 1$. Consequently, we have that $r_a \in [a]$. By Proposition 1.4.4, we conclude that $[a] = [r_a] = \{r \in \mathbb{Z} \mid r = -qn + r_a \text{ for some integer } q \in \mathbb{Z}\} = \{qn + r_a \mid q \in \mathbb{Z}\}$, as desired. □

Example 1.7.2. Observe that R_2 is an equivalence relation on $\mathbb{Z}$ whose distinct equivalence classes consist of the even integers $\mathbb{E} = \{2q \mid q \in \mathbb{Z}\}$ and the odd integers $\mathbb{O} = \{2q + 1 \mid q \in \mathbb{Z}\}$.

We will henceforth refer to the collection $\mathbb{Z}_n$ of equivalence classes of $\mathbb{Z}$ modulo R_n as the equivalence classes of $\mathbb{Z}$ **modulo** n. By Proposition 1.7.1, $\mathbb{Z}_n$ consists of exactly n distinct elements. Even more, for any two integers a and b, we have that $[a] = [b]$ if and only if the remainder of a modulo n is equal to the remainder of b modulo n if and only if there exist unique integers q_a, q_b, and r such that $a = q_a n + r$ and $b = q_b + r$ and $0 \leq r \leq n - 1$ if and only if $b - a = (q_b - q_a)n$. Put another way, two integers lie in the same equivalence class modulo n if and only if their difference is divisible by n. Generally, an equivalence relation is merely a set whose elements possess no arithmetic; however, the above observation allows us to deduce that $\mathbb{Z}_n$ (i.e., the set of equivalence classes of $\mathbb{Z}$ modulo n) admits a notion of addition and multiplication, as we demonstrate next.

Proposition 1.7.3. *Let $\mathbb{Z}_n$ denote the set of equivalence classes of the integers modulo n.*

(1.) *If a and b are arbitrary integers, then $[a] + [b] = [a + b]$ is a well-defined operation. Even more, this addition is associative, commutative, and satisfies that $[a] + [0] = [a] = [0] + [a]$.*

(2.) *Every equivalence class $[a]$ of the integers modulo n admits an additive inverse $[-a]$.*

(3.) *If a and b are arbitrary integers, then $[a][b] = [ab]$ is a well-defined operation. Even more, this multiplication is associative, commutative, distributive, and satisfies that $[a][1] = [a] = [1][a]$.*

(4.) *If a is an arbitrary integer, then $[a]$ admits a multiplicative inverse if and only if* $\gcd(a, n) = 1$.

Proof. (1.) We must demonstrate that if $[a_1] = [a_2]$ and $[b_1] = [b_2]$, then $[a_1 + b_1] = [a_2 + b_2]$. By the previous paragraph, if we assume that $[a_1] = [a_2]$ and $[b_1] = [b_2]$, then there exist integers q_a and q_b such that $a_1 - a_2 = q_a n$ and $b_1 - b_2 = q_b n$. Consequently, we have that

$$(a_1 + b_1) - (a_2 + b_2) = (a_1 - a_2) + (b_1 - b_2) = q_a n + q_b n = (q_a + q_b)n,$$

from which we conclude that $[a_1 + b_1] = [a_2 + b_2]$. Considering that integer addition is associative and commutative, our addition defined here is associative and commutative.

(3.) We must demonstrate that if $[a_1] = [a_2]$ and $[b_1] = [b_2]$, then $[a_1 b_1] = [a_2 b_2]$. By the paragraph preceding the proposition statement, if we assume that $[a_1] = [a_2]$ and $[b_1] = [b_2]$, then there exist integers q_a and q_b such that $a_1 - a_2 = q_a n$ and $b_1 - b_2 = q_b n$. Consequently, we have that

$$a_1 b_1 - a_2 b_2 = a_1 b_1 - a_1 b_2 + a_1 b_2 - a_2 b_2 = a_1(b_1 - b_2) + b_2(a_1 - a_2) = q_b a_1 n + q_a b_2 n = (q_b a_1 + q_a b_2)n,$$

from which we conclude that $[a_1 b_1] = [a_2 b_2]$. Considering that integer multiplication is associative and commutative, our multiplication defined here is associative and commutative. Even more, this multiplication is distributive because the first and third parts of the proposition that we have proved thus far establish that $[a]([b] + [c]) = [a][b + c] = [ab + ac] = [ab] + [ac] = [a][b] + [a][c]$.

(4.) By definition of our multiplication, the equivalence class $[a]$ admits a multiplicative inverse $[b]$ if and only if $[a][b] = [1]$ if and only if $[ab] = [1]$ if and only if $ab - 1 = qn$ for some integer q if and only if $ab - qn = 1$ for some integer q if and only if $\gcd(a, n) = 1$ by Bézout's Identity. Consequently, $[a]$ admits a multiplicative inverse if and only if $\gcd(a, n) = 1$, as desired. □

Combined, the operations of addition and multiplication on $\mathbb{Z}_n$ form the **modular arithmetic**.

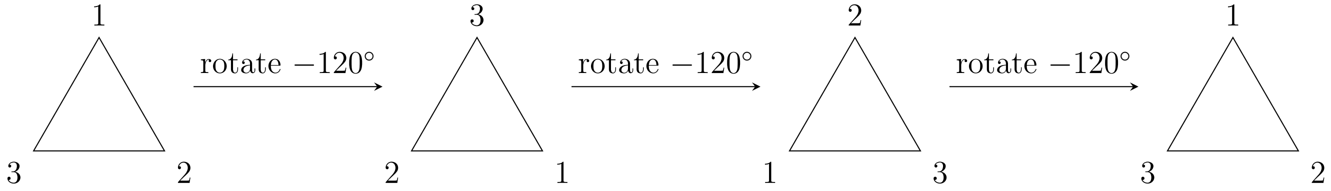
Remark 1.7.4. Going forward, we will adopt the standard notation $b \equiv a \pmod{n}$ ("b is equivalent to a modulo n") in place of our current notation that $[b] = [a]$. Explicitly, we will set $b \equiv a \pmod{n}$ if and only if $n \mid (b - a)$ if and only if $b - a = qn$ for some integer q. Under this identification, observe that $[a] = \{r \in \mathbb{Z} \mid a \equiv r \pmod{n}\}$. One immediate advantage of this notation is that we can perform addition and multiplication modulo n in a natural way: indeed, if $b \equiv a \pmod{n}$, then we have that $b + c \equiv a + c \pmod{n}$ and $bc \equiv ac \pmod{n}$ for all integers c because it holds that $(b+c)-(a+c) = b-a = qn$ and $bc-ac = (b-a)c = (qc)n$ in this case. Even more, Proposition 1.7.3 implies that if $b \equiv a \pmod{n}$ and $d \equiv c \pmod{n}$, then $b+d \equiv a+c \pmod{n}$ and $bd \equiv ac \pmod{n}$.

1.8 Rigid Motions

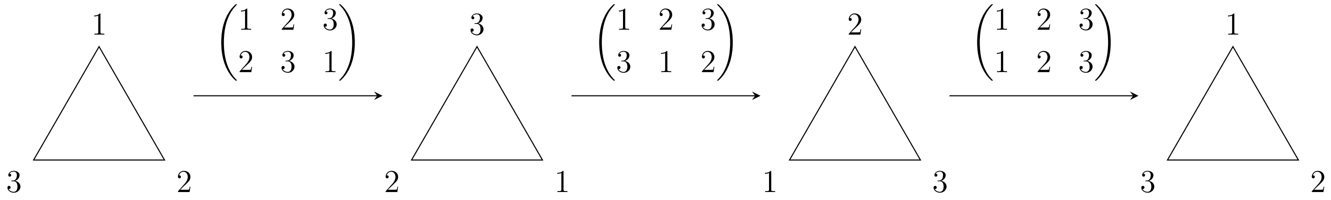
Recall that a **polygon** is a two-dimensional object consisting of straight line segments that intersect to form a closed and bounded region in the plane. Common examples of polygons include triangles, rectangles, and stars. Each of intersection point of a pair of straight line segments is called a **vertex** of the polygon. Particularly, triangles have three vertices; rectangles have four vertices; and stars typically have six vertices. We say that a polygon is **regular** if and only if each of its sides possesses equal length and each (interior) angle formed by the intersection of any two sides has equal measure (in either degrees or radians). Consequently, triangles and rectangles are not necessarily regular polygons; however, equilateral triangles and squares are both examples of regular polygons. We will henceforth refer to a (regular) polygon with n vertices as a (regular) n-**gon**. Under this naming convention, an (equilateral) triangle is a (regular) 3-gon; a (square) rectangle is a (regular) 4-gon; a (regular) pentagon is a (regular) 5-gon; and a (regular) hendecagon is a (regular) 11-gon.

Rigid motions of polygons are those operations that we can perform on polygons without altering the distance between any two vertices of the polygon. For instance, if we have a square in the plane, then we may shift each of the vertices of the square any distance north, south, east, or west without disturbing the distances between any of the vertices of the square; however, we cannot move just one vertex any nonzero distance north, south, east, or west without altering its distance from another vertex. Put another way, **translation** of a polygon is a rigid motion.

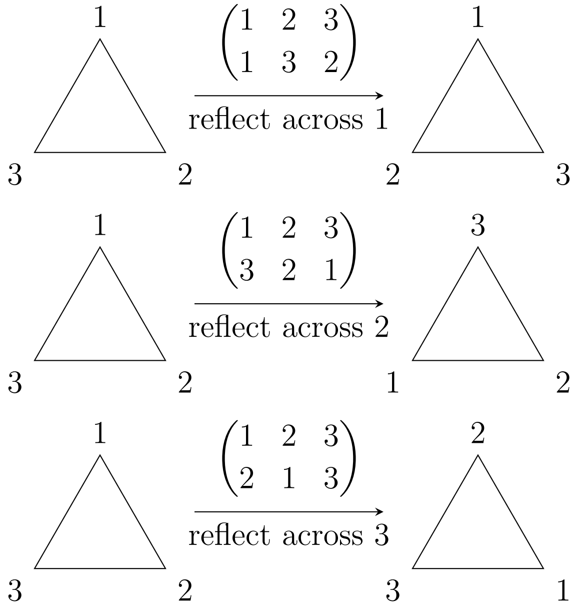
We will fix our attention throughout this section on two specific rigid motions of any regular n-gon. Each of the n vertices of a regular n-gon lies on the circumference of a circle. Consequently, for any integer $1 \leq k \leq n$, a **rotation** of a regular n-gon through an angle of $-360k/n$ degrees produces a copy of the regular n-gon with the ith vertex in place of the $(i+k)$th vertex (modulo n). Pictorially, we may visualize this with the rotations of a regular 3-gon (i.e., an equilateral triangle).



Each rotation is counterclockwise through an angle equal to the common measure of each exterior angle of the n-gon. Consequently, if we perform n rotations, then we wind up with the original arrangement of the vertices of the n-gon. Put another way, the rotations of a regular n-gon through an angle of $-360k/n$ degrees correspond to the **permutations** of the regular n-gon that move vertex i to vertex $i+k$ (modulo n). Explicitly, if we return to our example, we have the following.


1
3 2
$\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}$
3
2 1
$\begin{pmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{pmatrix}$
2
1 3
$\begin{pmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{pmatrix}$
1
3 2

On the other hand, a **reflection** of a regular n-gon through a vertex k is a permutation of the vertices of the regular n-gon that fixes the vertex k and swaps some other vertices (depending upon the parity of n). Going back to our example once more, there are three possible reflections.


1
3 2
$\begin{pmatrix}1 & 2 & 3\\ 1 & 3 & 2\end{pmatrix}$
reflect across 1
1
2 3
1
3 2
$\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{pmatrix}$
reflect across 2
3
1 2
1
3 2
$\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix}$
reflect across 3
2
3 1

Combined, these three rotations and three reflections completely exhaust all possible rotations and reflections of the regular 3-gon because there are only $3! = 6$ permutations of the integers $\{1, 2, 3\}$. Even more, if we execute a rotation followed by a reflection (or vice-versa), then we obtain a permutation of the integers $\{1, 2, 3\}$, hence every sequence of rotations and reflections yields a rotation or a reflection. We will return to this concept soon in our discussion of groups.

1.9 Chapter 1 Overview

A **set** X is a collection of distinct objects called **elements** or **members** of X that possess common properties. Elements of X are written abstractly as the lowercase symbol x. We assume the existence of a set $\emptyset$ that does not possess any elements; it is the **empty set**. Every collection of sets comes equipped with certain operations that allow us to combine; compare; and take differences of sets.

- The **union** of the sets X and Y is the set $X \cup Y = \{w \mid w \in X \text{ or } w \in Y\}$.
- The **intersection** of the sets X and Y is the set $X \cap Y = \{w \mid w \in X \text{ and } w \in Y\}$.
- The **relative complement** of X with respect to Y is the set $Y \setminus X = \{w \in Y \mid w \notin X\}$.

We say that Y is a **subset** of X if every element of Y is an element of X, in which case we write $Y \subseteq X$; if Y is a subset of X and there exists an element of X that is not an element of Y, then Y is a **proper subset** of X, in which case we write $Y \subsetneq X$. Observe that Y is a (proper) subset of X if and only if $X \cap Y = Y$ (and $X \cup Y = X$). By the Law of the Excluded Middle, it is always true that $X = Y \cup (X \setminus Y)$ for any set $Y \subseteq X$. If $Y \subseteq X$ and $X \subseteq Y$, then $X = Y$; otherwise, the sets X and Y are distinct. One other way to distinguish a (finite) set X is by the number of elements X possesses — its **cardinality**, denoted by $\#X$ or $|X|$ when this notation is unambiguous.

We define the **Cartesian product** of two sets X and Y to be the set consisting of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$, i.e., $X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$; a subset R of the Cartesian product $X \times X$ is called a **relation** (on X). Every set X admits a relation called the **diagonal** (of X) and defined by $\Delta_X = \{(x, x) \mid x \in X\}$. Cardinality of sets is multiplicative in the sense that if X and Y are finite sets, then it holds that $|X \times Y| = |X| \cdot |Y|$.

We define a **function** $f : X \to Y$ with **domain** X and **codomain** Y by declaring for each element $x \in X$ a unique (but not necessarily distinct) element $f(x) \in Y$. Every function $f : X \to Y$ induces a subset $f(X) = \{y \in Y \mid y = f(x) \text{ for some element } x \in X\}$ of Y called the **image** of X (in Y) with respect to f. Given any set $W \subseteq Y$, we may also consider the **pre-image** of W (in X) with respect to f, i.e., $f^{-1}(W) = \{x \in X \mid f(x) \in W\}$. We say that $f : X \to Y$ is **injective** if it holds that $f(x_1) = f(x_2)$ implies that $x_1 = x_2$ for any pair of elements $x_1, x_2 \in X$. On the other hand, if for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$, then $f : X \to Y$ is **surjective**. If a function $f : X \to Y$ is both injective and surjective, then it is **bijective**.

We say that a complete sentence P is a **statement** if it asserts something that can be unambiguously measured as true or false. Examples of statements include “3 is an odd number” and “17 is larger than 38”; the first statement is true, but the second statement is false. Using logical connectives, we can form new statements from given statements P and Q. Explicitly, the **implication** $P \implies Q$ is the statement that “P implies Q” (or equivalently, “If P, then Q”); the implication is false if and only if P is true and Q is false. If P is false, then $P \implies Q$ is called a **vacuous truth**. We define the **disjunction** $P \vee Q$ ("P or Q"), the **conjunction** $P \wedge Q$ ("P and Q"), and the **negation** $\neg P$ ("not P"). Observe that the disjunction $P \vee Q$ is true if and only if P is true or Q is true; the conjunction $P \vee Q$ is true if and only if P is true and Q is true; and the negation $\neg P$ takes the opposite truth-value of P. The Law of the Excluded Middle asserts that either P or $\neg P$ must be true, and the Law of Non-Contradiction asserts that P and $\neg P$ cannot both be true.

We use **truth tables** to deduce the verity of a statement $S(P, Q)$ depending upon two statements P and Q. One can construct a truth table for $S(P, Q)$ by writing all possible **truth-values** of P in one column; all possible truth-values of Q in a subsequent column; and the resultant truth-values of the statement $S(P, Q)$ is a third column. Considering that the statements P and Q could themselves depend upon other statements $P_1, \dots, P_n$, truth tables may become quite large when the attendant statements are complicated. Generally, we need 2^n rows and $n + 1$ columns to construct the truth table of a statement $S(P_1, \dots, P_n)$ depending upon n distinct statements $P_1, \dots, P_n$. If two statements S and S' induce the same truth table, then they are **logically equivalent**; in particular, the truth-values of S are exactly the truth-values of S', hence the verity of the statement S can be deduced from the verity of the statement S' (and vice-versa). If the truth-values for S are all true, then S is a **tautology**; if the truth-values for S are all false, then S is a **self-contradiction**.

De Morgan's Laws are two rules of inference that relate disjunction, conjunction, and negation; explicitly, they assert that (1.) $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$ are logically equivalent and (2.) $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ are logically equivalent. We define the **contrapositive** of the implication $P \implies Q$ as the statement $\neg Q \implies \neg P$ ("If not Q, then not P") obtained by taking the implication of the negations of Q and P. **Proof by contraposition** is a law of inference that exploits the fact that the contrapositive is logically equivalent to the implication, i.e., the statements $P \implies Q$ and $\neg Q \implies \neg P$ induce the same truth table (cf. Table 1.5). **Proof by contradiction** is a law of inference that can be deduced from the Law of the Excluded Middle, the Law of Non-Contradiction, and the logical equivalence of the statements $\neg(P \implies Q)$ and $P \wedge \neg Q$ (cf. Table 1.3). We carry out a proof by contradiction by first assuming that P is true and that Q is not true; then, we arrive at a contradiction of the form (a.) $P \wedge \neg P$ or (b.) $Q \wedge \neg Q$. We note that if the former holds (i.e., if $\neg P$ can be deduced from $\neg Q$), then a proof by contraposition may be simpler than a proof by contradiction; on the other hand, if the latter holds (i.e., if Q can be deduced from P), then a **direct proof** may be simpler than a proof by contradiction. But any of the three is valid.

Logical quantifiers allow us to symbolically handle statements involving quantities. We use the **universal quantifier** $\forall$ to express that a statement holds "for all" specified objects, and we use the **existential quantifier** $\exists$ to express "there exists" an object satisfying a given statement. We say that an object satisfying a given statement is **unique** if it is the only object that satisfies the given statement. If there exists one and only one object satisfying a specified condition, then we use the **uniqueness quantifier** $\exists!$ to express its existence ($\exists$) and uniqueness (!).

Using logical quantifiers allows us to conveniently state many properties of sets, e.g., the Law of the Excluded Middle for Sets, Law of Non-Contradiction for Sets, and De Morgan's Laws for Sets. Even more, logical quantifiers enable us to extend De Morgan's Laws for Sets to arbitrary unions and arbitrary intersections of sets. Explicitly, we may consider an arbitrary set I as an **index set** for an arbitrary family of sets $\{X_i \mid i \in I\}$ such that each set X_i is a subset of some set W called our **universe**. By definition, the arbitrary union of these sets is simply $\cup_{i \in I} X_i$; membership of an element $w \in W$ in this arbitrary union is characterized by $w \in \cup_{i \in I} X_i$ if and only if $w \in X_i$ for some index $i \in I$. Likewise, the arbitrary intersection of these sets is $\cap_{i \in I} X_i$ with membership of an element $w \in W$ characterized by $w \in \cap_{i \in I} X_i$ if and only if $w \in X_i$ for all indices $i \in I$. We say that two sets X_i and X_j are **disjoint** if $X_i \cap X_j = \emptyset$; moreover, if $X_i \cap X_j = \emptyset$ for all distinct indices $i, j \in I$, then we say that the sets in $\{X_i \mid i \in I\}$ are **pairwise disjoint** or **mutually exclusive**. We say that the collection $\mathcal{P} = \{X_i \mid i \in I\}$ forms a **partition** of the set W if and only if

(i.) X_i is nonempty for each index $i \in I$;

(ii.) $W = \cup_{i \in I} X_i$; and

(iii.) the sets X_i are pairwise disjoint (i.e., $X_i \cap X_j = \emptyset$ for every pair of distinct indices $i, j \in I$).

If X is an arbitrary set, then a relation on X is a subset R of the Cartesian product $X \times X$. We say that a relation R on X is **reflexive** if and only if $(x, x) \in R$ for all elements $x \in X$; **symmetric** if and only if $(x_1, x_2) \in R$ implies that $(x_2, x_1) \in R$ for all pairs of elements $x_1, x_2 \in X$; **antisymmetric** if and only if $(x_1, x_2) \in R$ and $(x_2, x_1) \in R$ implies that $x_1 = x_2$ for all pairs of elements $x_1, x_2 \in X$; and **transitive** if and only if $(x_1, x_2) \in R$ and $(x_2, x_3) \in R$ together imply that $(x_1, x_3) \in R$ for all triples of elements $x_1, x_2, x_3 \in X$. **Equivalence relations** are precisely the reflexive, symmetric, and transitive relations; **partial orders** are precisely the reflexive, antisymmetric, and transitive relations. Every equivalence relation E on X induces a partition of E via the **equivalence classes** of elements of X. Explicitly, we say that two elements $x_1, x_2 \in X$ are **equivalent modulo** E if and only if $(x_1, x_2) \in E$, in which case we write that $x_1 \sim_E x_2$; the equivalence class of an element $x_0 \in X$ is the collection of elements $x \in X$ that are equivalent to x_0 modulo E, i.e., the equivalence class of x_0 is $[x_0] = \{x \in X \mid x \sim_E x_0\} = \{x \in X \mid (x, x_0) \in E\}$. Every element of X belongs to one and only one equivalence class of X modulo E, hence X is partitioned by the collection of distinct equivalence classes modulo E (cf. Proposition 1.4.4 and Corollary 1.4.5). Every set admits a partial order, hence every set is a **partially ordered set**; however, there can be many ways to view a set as a partially ordered set because there can be many different partial orders on a set. If P is a partial order on a set X, then we say that a pair of elements $p, q \in P$ are **comparable** if either $(p, q) \in P$ or $(q, p) \in P$; otherwise, we say that p and q are **incomparable**. We say that a partial order P on X is a **total order** on X if every pair of elements $p, q \in P$ are comparable. Every partial order P of X induces a partial order on the subsets $Y \subset X$ via $P|_Y = \{(y_1, y_2) \in Y \times Y \mid (y_1, y_2) \in P\}$; if $P|_Y$ is a total order on $Y \subseteq X$, then we say that Y is a **chain** (with respect to P) in X. We say that an element $x_0 \in X$ is an **upper bound** on Y (with respect to P) if $(y, x) \in P$ for every element $y \in Y$. We will also say that an element $x_0 \in X$ is **maximal** (with respect to P) if it does not hold that $(x_0, x) \in P$ for any element $x \in X \setminus \{x_0\}$. Zorn's Lemma asserts that if P is a partial order on an arbitrary set X such that every chain Y in X has an upper bound in Y, then Y admits a maximal element $y_0 \in Y$ (with respect to P). We will make use of this throughout the course.

One of the most useful tools in mathematics is the **Principle of Mathematical Induction**. Collectively, the Principle of Mathematical Induction contains the (equivalent) Principle of Ordinary Induction and the Principle of Complete Induction. Explicitly, the Principle of Ordinary Induction asserts that if $P(n)$ is any statement about a non-negative integer n such that

(1.) $P(0)$ is a true statement and

(2.) $P(k+1)$ is a true statement whenever $P(k)$ is a true statement,

then $P(n)$ is a true statement for all non-negative integers n; the Principle of Complete Induction asserts that if $P(n)$ is any statement about a non-negative integer n such that

(1.) $P(0)$ is a true statement and

(2.) $P(k+1)$ is a true statement whenever $P(1), P(2), \dots, P(k)$ are all true statements,

then $P(n)$ is a true statement for all non-negative integers n. One of the benefits of using complete induction is that its stronger hypotheses allow us more information with which to conveniently write proofs that might otherwise be awkward with ordinary induction (cf. Exercise 1.10.24). Even more, the Principle of Mathematical Induction appears also in the guise of the Well-Ordering Principle for the non-negative integers; this powerful tool guarantees that every nonempty set of non-negative integers admits a smallest element with respect to the total order $\leq$. Put another way, if $S \subseteq \mathbb{Z}_{\geq 0}$ is a nonempty set, then there exists an element $s_0 \in S$ such that $s_0 \leq s$ for all elements $s \in S$.

Using the Well-Ordering Principle, we may rigorously establish that for any integer a and nonzero integer b, there exist unique integers q and r such that $a = qb + r$ and $0 \leq r < |b|$; this fact is known as the Division Algorithm. We refer to the integer a as the **dividend**; b is the **divisor**; q is the **quotient**; and r is the **remainder**. Conventionally, if we obtain a remainder of zero when we divide an integer a by a nonzero integer b, then we say that b **divides** a; in this case, there exists a unique integer q such that $a = qb$, and we use the notation $b \mid a$. If a and b are any integers, then a nonzero integer c is called a **common divisor** of a and b if it holds that $c \mid a$ and $c \mid b$; the **greatest common divisor** of a and b is the unique integer $d = \gcd(a, b)$ such that

(a.) $d \mid a$ and $d \mid b$, i.e., d is a common divisor of a and b and

(b.) if d' is any common divisor of a and b, then $d' \mid d$.

We say that a and b are **relatively prime** if and only if $\gcd(a, b) = 1$. Bézout's Identity asserts that there exist integers x and y such that $\gcd(a, b) = ax + by$; the Euclidean Algorithm is one method from which the integers x and y guaranteed by Bézout's Identity can be obtained.

By the Division Algorithm, for any positive integer n, we may partition the integers $\mathbb{Z}$ into distinct equivalence classes determined by the unique remainder of an integer **modulo** n. Explicitly, we say that two integers a and b are **equivalent modulo** n if and only if $b - a$ is divisible by n; if this is the case, then we write $b \equiv a \pmod{n}$. One can verify that equivalence modulo n induces an equivalence relation R_n on the integers defined by $(a, b) \in R_n$ if and only if $b \equiv a \pmod{n}$; the distinct equivalence classes of $\mathbb{Z}$ modulo R_n are given by $\{qn + r \mid q \in \mathbb{Z}\}$ for each integer $0 \leq r \leq n - 1$; and the collection $\mathbb{Z}_n$ of equivalence classes of $\mathbb{Z}$ modulo n admits operations of addition and multiplication that together comprise the so-called **modular arithmetic**. Explicitly, if $b \equiv a \pmod{n}$ and $d \equiv c \pmod{n}$, then we have that $b + d \equiv a + c \pmod{n}$ and $bd \equiv ac \pmod{n}$.

Polygons are two-dimensional, closed, and bounded objects determined by the intersection of finitely many straight line segments in the plane; the intersection points are called **vertices**. Examples of polygons include triangles, rectangles, and stars. **Regular** polygons have the additional property that their sides possess equal length and each angle at a vertex has equal measure. Equilateral triangles and squares are regular, but most triangles and rectangles are not regular. Generally, an n-**gon** is any polygon with n vertices. **Rigid motions** of a polygon are those operations that can be performed on the polygon without altering the distance between any two of its vertices. Regular n-gons have the property that **rotation** by an angle of $-360k/n$ degrees is a rigid motion for each integer $1 \leq k \leq n$. Likewise, **reflection** of a regular n-gon across any one of its n vertices also constitutes a rigid motion of the regular n-gon. Combined, rotations and reflections of a regular n-gon can be performed in any order to produce another rotation or reflection.

1.10 Chapter 1 Exercises

1.10.1 Sets and Set Operations

Exercise 1.10.1. Consider the sets

- $W = \{1, 2, 3, \dots, 10\}$ of positive integers from 1 to 10;
- $X = \{1, 3, 5, 7, 9\}$ of odd positive integers from 1 to 10;
- $Y = \{2, 4, 6, 8, 10\}$ of even positive integers from 1 to 10;
- $\mathbb{O} = \{n \mid n \text{ is an odd integer}\}$;
- $\mathbb{E} = \{n \mid n \text{ is an even integer}\}$; and
- $\mathbb{Z} = \{n \mid n \text{ is an integer}\}$.

Use the set operations $\subseteq$, $\cup$, $\cap$, and $\setminus$ to describe as many relations among these sets as possible.

Exercise 1.10.2. Let $W, X, Y, \mathbb{O}, \mathbb{E}$, and $\mathbb{Z}_{>0}$ be the sets defined in Exercise 1.10.1.

(a.) Compute the number of elements of $X \times Y$; then, list at least three of them.

(b.) List all elements of the diagonal Δ_X of X.

(c.) Every odd integer can be written as $2k + 1$ for some integer k, and every even integer can be written as 2ℓ for some integer ℓ. Express the sets $\mathbb{O}$ and $\mathbb{E}$ in set-builder notation accordingly.

(d.) Convince yourself that $\mathbb{O}$ and $\mathbb{E}$ have "essentially the same" number of elements; then, find a function $f : \mathbb{O} \to \mathbb{E}$ such that f is injective and f is surjective. Observe that this gives a rigorous justification of the fact that $\mathbb{O}$ and $\mathbb{E}$ have "essentially the same" number of elements.

(e.) Convince yourself that $\mathbb{O}$ and $\mathbb{Z}$ have "essentially the same" number of elements; then, find a function $f : \mathbb{O} \to \mathbb{Z}$ such that f is injective and f is surjective. Conclude from this exercise and the previous one that there are "as many" odd (or even) integers as there are integers.

Exercise 1.10.3. Let W be an arbitrary set. Let $X \subseteq W$ and $Y \subseteq W$ be arbitrary subsets of W.

(a.) Prove that for any subset $Z \subseteq W$ such that $Z \supseteq X$ and $Z \supseteq Y$, it follows that $Z \supseteq X \cup Y$. Conclude that $U = X \cup Y$ is the "smallest" subset of W containing both X and Y.

(b.) Prove that for any subset $Z \subseteq W$ such that $Z \subseteq X$ and $Z \subseteq Y$, it follows that $Z \subseteq X \cap Y$. Conclude that $I = X \cap Y$ is the "largest" subset of W contained in both X and Y.

Consider the relative complement $X' = W \setminus X$ of X in W. We may sometimes refer to X' simply as the **complement** of X if we are dealing only with subsets of W, i.e., if W is our universe.

(c.) Prove that $Y \setminus X = Y \cap X'$. Use part (b.) above to conclude that $C = Y \cap X'$ is the "largest" subset of W that is contained in Y and disjoint from X.

Exercise 1.10.4. Let X be an arbitrary set. Prove that $\Delta_X = \delta_X(X)$, where $\delta_X : X \to X \times X$ is the diagonal function defined by $\delta_X(x) = (x, x)$ and $\Delta_X = \{(x, x) \mid x \in X\}$ is the diagonal of X.

Exercise 1.10.5. Let X and Y be arbitrary finite sets.

(a.) Prove that if $|X| \leq |Y|$, then there exists an injective function $f : X \to Y$.

(b.) Prove that if $|X| \geq |Y|$, then there exists a surjective function $f : X \to Y$.

(c.) Conclude that if $|X| = |Y|$, then there exists a bijective function $f : X \to Y$.

(**Caution:** this is not necessarily true if X and Y are infinite sets.)

(d.) Let $|X| = |Y|$. Prove that a function $f : X \to Y$ is injective if and only if it is surjective.

(**Caution:** this is not necessarily true if X and Y are infinite sets.)

Exercise 1.10.6. Let X and Y be arbitrary sets.

(a.) Prove the converse to part (c.) of Exercise 1.10.5, i.e., establish that if there exists a bijective function $f : X \to Y$, then we must have that $|X| = |Y|$.

(b.) Prove that if there exists a function $f^{-1} : Y \to X$ such that $f^{-1} \circ f = \mathrm{id}_X$, then f is injective.

(c.) Prove that if there exists a function $f^{-1} : Y \to X$ such that $f \circ f^{-1} = \mathrm{id}_Y$, then f is surjective.

Exercise 1.10.7. Let $f : X \to Y$ be any function between any two sets X and Y.

(a.) Prove that $V \subseteq f^{-1}(f(V))$ for any set $V \subseteq X$.

(b.) Exhibit sets $V \subseteq X$ and Y and a function $f : X \to Y$ such that $f^{-1}(f(V)) \not\subseteq V$.

(**Hint:** By Proposition 1.1.1, $f : X \to Y$ cannot be injective.)

(c.) Prove that $f(f^{-1}(W)) \subseteq W$ for any set $W \subseteq Y$.

(d.) Exhibit sets X and $W \subseteq Y$ and a function $f : X \to Y$ such that $W \not\subseteq f(f^{-1}(W))$.

(**Hint:** By Proposition 1.1.1, $f : X \to Y$ cannot be surjective.)

Exercise 1.10.8. Let $f : X \to Y$ be any function between any two sets X and Y.

(a.) Prove that if $f^{-1}(f(V)) = V$ for any set $V \subseteq X$, then f is injective.

(**Hint:** If $f(x_1) = f(x_2)$, then consider the set $V = \{x_1\}$.)

(b.) Prove that if $f(f^{-1}(W)) = W$ for any set $W \subseteq Y$, then f is surjective.

(**Hint:** Consider the set $W = Y$; then, use the definition of $f(f^{-1}(W))$.)

1.10.2 Logic and Truth Tables

Exercise 1.10.9. Let P be the statement that "The sun is shining in Kansas City." Let Q be the statement that "Bob rides his bike to work." Use the letters P and Q and logical connectives such as $\implies$, $\iff$, $\vee$, $\wedge$, and $\neg$ to convert each of the following statements into symbols; then, identify all of the logically equivalent statements, tautologies, and self-contradictions.

(a.) "If the sun is shining in Kansas City, then Bob rides his bike to work."

(b.) "Bob rides his bike to work only if the sun is shining in Kansas City."

(c.) "Either the sun is not shining in Kansas City or Bob rides his bike to work."

(d.) "The sun is shining in Kansas City, and Bob does not ride his bike to work."

(e.) "If the sun is not shining in Kansas City, then Bob does not ride his bike to work."

(f.) "If Bob does not ride his bike to work, then the sun is not shining in Kansas City."

(g.) "Neither the sun is shining in Kansas City nor Bob rides his bike to work."

(h.) "Either the sun is not shining in Kansas City or Bob does not ride his bike to work."

(i.) "The sun is not shining in Kansas City, and Bob does not ride his bike to work."

(j.) "Either Bob rides his bike to work or Bob does not ride his bike to work."

(k.) "The sun is shining in Kansas City, and the sun is not shining in Kansas City."

(l.) "Bob rides his bike to work if and only if the sun is shining in Kansas City."

(m.) "The sun is not shining in Kansas City if and only if Bob does not ride his bike to work."

Exercise 1.10.10. Let P, Q, and R be any statements. Construct a truth table to prove that the statements "If P, then Q or R" and "If P and not Q, then R" are logically equivalent.

Exercise 1.10.11. Let P and Q be any statements. Construct a truth table to prove that the statement "If P or Q but not Q, then P" is a tautology.

Exercise 1.10.12. Use Exercise 1.10.11 to prove that if Bob placed in the top two in a cycling race on Saturday, but he did not place second, then Bob must have placed first.

Exercise 1.10.13. Use a proof by contradiction to prove that if Bob placed in the top two in a cycling race on Saturday, but he did not place second, then Bob must have placed first. Cite any theorems or laws of inference (by name) that you use in your proof.

1.10.3 Sets and Set Operations, Revisited

Exercise 1.10.14. (De Morgan's Laws for Sets) Let $X \subseteq W$ and $Y \subseteq W$ be arbitrary sets.

(a.) Prove that $W \setminus (X \cup Y) = (W \setminus X) \cap (W \setminus Y)$.

(b.) Prove that $W \setminus (X \cap Y) = (W \setminus X) \cup (W \setminus Y)$.

(**Hint:** Define statements P and Q for which $P \vee Q$ is the statement that "$w \in X \cup Y$" and $P \wedge Q$ is the statement that "$w \in X \cap Y$"; then, use De Morgan's Laws to conclude the results.)

Exercise 1.10.15. Let $X_1, X_2, \dots, X_n \subseteq W$ be arbitrary sets.

(a.) Prove that $W \setminus (X_1 \cup X_2 \cup \cdots \cup X_n) = (W \setminus X_1) \cap (W \setminus X_2) \cap \cdots \cap (W \setminus X_n)$.

(b.) Prove that $W \setminus (X_1 \cap X_2 \cap \cdots \cap X_n) = (W \setminus X_1) \cup (W \setminus X_2) \cup \cdots \cup (W \setminus X_n)$.

(**Hint:** De Morgan's Laws for Sets guarantee that if $w \in W$ and $w \notin X_1 \cup X_2 \cup \cdots \cup X_n$, then it must be that $w \notin X_1$ and $w \notin X_2 \cup X_3 \cup \cdots \cup X_n$. Repeat this process finitely many times.)

Exercise 1.10.16. Let $\mathbb{Z}$ denote the set of integers.

(a.) Provide a partition of $\mathbb{Z}$ into three sets.

(**Hint:** By the Division Algorithm, if we divide any integer by 3, what are the only possible remainders? Use this observation to construct a partition of $\mathbb{Z}$ into three sets.)

(b.) Provide a partition of $\mathbb{Z}$ into four sets.

(c.) Provide a partition of $\mathbb{Z}$ into n sets for any positive integer n.

1.10.4 Equivalence Relations and Partial Orders

Exercise 1.10.17. Consider the set W consisting of all words in the English language.

(a.) Prove that $R = \{(v, w) \in W \times W \mid v \text{ and } w \text{ begin with the same letter}\}$ is an equivalence relation on W; then, determine the number of distinct equivalence classes of W modulo R.

(b.) Prove that $R = \{(v, w) \in W \times W \mid v \text{ and } w \text{ have the same number of letters}\}$ is an equivalence relation on W; then, describe the equivalence class of the word "awesome."

Exercise 1.10.18. Let $\mathbb{Z}$ be the set of integers. Prove that $(a, b) \sim (c, d)$ if and only if $ad = bc$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is an equivalence relation. Describe the collection of distinct equivalence classes.

(**Hint:** For the second part of the problem, try replacing the notation (a, b) with a/b, instead.)

Exercise 1.10.19. Let X be an arbitrary set. Consider the collection $S = \{Y \mid Y \subseteq X\}$. Prove that the inclusion $\subseteq$ defines a partial order P on S such that $(Y_1, Y_2) \in P$ if and only if $Y_1 \subseteq Y_2$; then, either prove that P is a total order on S, or provide a counterexample to show that it is not.

Exercise 1.10.20. List the maximal elements of the subset $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$ of the set $\mathbb{Z}_{\geq 0}$ of non-negative integers with respect to the partial order D of divisibility.

(**Hint:** List as many pairs of comparable elements of S as necessary to compute the chains in S with three or four elements; then, use this information deduce the maximal elements of S.)

1.10.5 The Principle of Mathematical Induction

Exercise 1.10.21. Prove Conjecture 1.5.2 using the Principle of Ordinary Induction.

Exercise 1.10.22. Prove Conjecture 1.5.4 using the Principle of Ordinary Induction.

If X is an arbitrary set, then the **power set** of X is the set $P(X) = \{Y \mid Y \subseteq X\}$, i.e., it is the collection of all subsets of X. Explicitly, if $X = \{x, y\}$, then $P(X) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$.

Exercise 1.10.23. Let X be an arbitrary finite set with power set $P(X)$.

(a.) Use ordinary induction on $n = |X|$ to prove that $|P(X)| = 2^{|X|}$.

(b.) Let 2^X denote the collection of all functions $f : X \to X$. Exhibit an explicit bijection between $P(X)$ and 2^X; then, conclude from part (a.) above that $|2^X| = 2^{|X|}$.

One of the most curious objects in mathematics is the sequence $(F_n)_{n \geq 0}$ of **Fibonacci numbers** that are defined recursively by $F_0 = 0$, $F_1 = 1$, and $F_{n+2} = F_{n+1} + F_n$. We refer to F_n as the nth Fibonacci number. Quite astoundingly, the Fibonacci numbers appear abundantly in nature.

Exercise 1.10.24. Let F_n denote the nth Fibonacci number.

(a.) Prove that $F_n < 2^n$ for each integer $n \geq 0$.

(b.) Prove that $F_{n+1}F_{n-1} = F_n^2 + (-1)^n$ for each integer $n \geq 2$.

(c.) Prove that $\gcd(F_n, F_{n+1}) = 1$ for all integers $n \geq 0$.

Exercise 1.10.25. Prove that the Principle of Ordinary Induction and the Principle of Complete Induction are equivalent to one another by completing the following two steps.

(1.) Given any statement $P(n)$ involving a non-negative integer n, let $Q(n)$ be the statement that $P(k)$ holds for all integers $1 \leq k \leq n$. Use the Principle of Ordinary Induction to prove that the statement $Q(n)$ is true for all integers $n \geq 0$, hence $P(n)$ is true for all integers $n \geq 0$. Unravelling this shows that ordinary induction implies complete induction.

(**Hint:** Observe that $Q(0)$ is vacuously true, hence we may assume that $Q(n)$ is true. By definition, this means that $P(k)$ is true for all integers $1 \leq k \leq n$. What about $P(n+1)$?)

(2.) Given any statement $P(n)$ involving a non-negative integer n, let $Q(n)$ be the statement that $P(k)$ holds for some integer $1 \leq k \leq n$. Use the Principle of Complete Induction to prove that the statement $Q(n)$ is true for all integers $n \geq 0$, hence $P(n)$ is true for all integers $n \geq 0$. Unravelling this shows that complete induction implies strong induction.

(**Hint:** Observe that $Q(0)$ is vacuously true, hence we may assume that $Q(k)$ is true for all integers $1 \leq k \leq n$; in particular, $P(1)$ is true. What does this say about $Q(n+1)$?)

Exercise 1.10.26. Prove that the Well-Ordering Principle and the Principle of Ordinary Induction are equivalent to one another by completing the following three steps.

(1.) Prove that 0 is the smallest non-negative integer with respect to $\leq$.

(2.) Prove that if $S \subseteq \mathbb{Z}_{\geq 0}$ satisfies $0 \in S$ and $n + 1 \in S$ whenever $n \in S$, then $\mathbb{Z}_{\geq 0} \subseteq S$.

(3.) Conclude that the Well-Ordering Principle implies the Principle of Ordinary Induction; then, use Exercise 1.10.25 and the proof of the Well-Ordering Principle to conclude that the Principle of Ordinary Induction implies the Well-Ordering Principle.

1.10.6 The Division Algorithm

Exercise 1.10.27. Recall that a positive integer p is **prime** if and only if the only integers that divide p are $\pm p$ and 1. Prove that if a and b are any integers such that $p \mid ab$, then $p \mid a$ or $p \mid b$.

(**Hint:** We may assume that $p \nmid a$ and show that $p \mid b$; now, use Bézout's Identity.)

Exercise 1.10.28. Let a, b, and c be any integers. Prove that if $\gcd(a, b) = 1$ and $a \mid bc$, then $a \mid c$.

Exercise 1.10.29 (Fundamental Theorem of Arithmetic)**.** Let a be a positive integer. Prove that

(a.) there exist (not necessarily distinct) prime numbers $p_1, \dots, p_k$ such that $a = p_1 \cdots p_k$ and

(b.) the primes $p_1, \dots, p_k$ are unique in the sense that if $a = q_1 \dots q_\ell$, then we must have that $\ell = k$ and $\{p_1, \dots, p_k\} = \{q_1, \dots, q_k\}$ (i.e., $q_1, \dots, q_k$ are just a rearrangement of $p_1, \dots, p_k$).

(**Hint:** Consider the collection N of positive integers that do *not* possess such a prime factorization. Use the Well-Ordering Principle to show that if N is nonempty, then there exists a smallest element n with respect to $\leq$. What can be said about the factors of n? Conclude that N must be empty, hence the existence is established. On the matter of uniqueness, proceed by induction on k.)

Exercise 1.10.30. Let a be any positive integer. Prove that there exist distinct prime numbers $p_1, \dots, p_n$ and unique non-negative integers $e_1, \dots, e_n$ such that $a = p_1^{e_1} \dots p_n^{e_n}$.

Given any integers a and b, the **least common multiple** $\operatorname{lcm}(a, b)$ of a and b can be defined in a manner analogous to the greatest common divisor of a and b. Explicitly, we say that an integer m is a **multiple** of a if and only if $a \mid m$. Consequently, m is a **common multiple** of a and b if and only if $a \mid m$ and $b \mid m$; a least common multiple of a and b is an integer $\ell = \operatorname{lcm}(a, b)$ such that

(a.) $a \mid \ell$ and $b \mid \ell$, i.e., ℓ is a common multiple of a and b and

(b.) if ℓ' is any common multiple of a and b, then $\ell \mid \ell'$.

Exercise 1.10.31. Prove that the $\operatorname{lcm}(a, b)$ is unique.

By the Fundamental Theorem of Arithmetic, for any positive integers a and b, there exist prime numbers $p_1, \dots, p_k$ and unique non-negative integers $e_1, \dots, e_k, f_1, \dots, f_k$ such that $a = p_1^{e_1} \cdots p_k^{e_k}$ and $b = p_1^{f_1} \cdots p_k^{f_k}$. Consider these prime factorizations of a and b for the next three exercises.

Exercise 1.10.32. Prove that $\gcd(a, b) = p_1^{\min\{e_1, f_1\}} \cdots p_k^{\min\{e_k, f_k\}}$.

Exercise 1.10.33. Prove that $\operatorname{lcm}(a, b) = p_1^{\max\{e_1, f_1\}} \cdots p_k^{\max\{e_k, f_k\}}$.

Exercise 1.10.34. Conclude from Exercises 1.10.32 and 1.10.33 that $ab = \gcd(a, b) \operatorname{lcm}(a, b)$.

1.10.7 The Integers Modulo n

Exercise 1.10.35. Complete the following using modular arithmetic.

(a.) If $a \equiv 1 \pmod 6$, find the least positive x for which $5a + 4 \equiv x \pmod 6$.

(b.) If $a \equiv 4 \pmod 7$ and $b \equiv 5 \pmod 7$, find the least positive x for which $6a - 3b \equiv x \pmod 7$.

(c.) (Modular Exponentiation) Use the fact that $2^{2022} \equiv 4 \pmod{10}$ to find $2022^{2022} \pmod{10}$.

Exercise 1.10.36. Consider the collection $\mathbb{Z}_n$ of equivalence classes of the integers modulo n. If $ab \equiv 0 \pmod{n}$, must it be true that $a \equiv 0 \pmod{n}$ or $b \equiv 0 \pmod{n}$? Explain.

Exercise 1.10.37. Let p be any prime number. Consider the collection $\mathbb{Z}_p$ of equivalence classes of the integers modulo p. Prove that if $ab \equiv 0 \pmod{p}$, then $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$.

Exercise 1.10.38. Let p be any prime number. Consider the collection $\mathbb{Z}_p$ of equivalence classes of the integers modulo p. Prove that $[a]$ admits a multiplicative inverse if and only if $p \nmid a$.

1.10.8 Rigid Motions

Exercise 1.10.39. Prove that for a regular n-gon, there are at most $n! = n(n-1)(n-2)\cdots 2 \cdot 1$ symmetries corresponding to rotation through an angle or reflection across a vertex.

Exercise 1.10.40. List all permutations of the integers $\{1, 2, 3, 4\}$ corresponding to the rotations and reflections of a regular 4-gon. Conclude that the upper bound of Exercise 1.10.39 can be strict.

(**Caution:** Because there are an even number of vertices of the square, only two of the symmetry-preserving reflections of the square will pass through a pair of vertices; however, there are other symmetry-preserving reflections of the square that do not correspond to reflection about a vertex.)

Exercise 1.10.41. Conjecture a formula for the number of symmetry-preserving rotations and reflections of a regular n-gon; then, prove that your formula holds.

(**Hint:** Use the example of Section 8, your work from Exercise 1.10.40, and possibly an additional example to spot the pattern and deduce a formula; then, use the Fundamental Counting Principle.)

Exercise 1.10.42. Consider the regular 3-gon of Section 1.8 with vertices labelled 1, 2, and 3 in clockwise order. Let ρ_k denote rotation of the regular 3-gon through an angle of $-120k$ degrees. Explicitly, there are three distinct rotations ρ_1, ρ_2, and ρ_3. Let ϕ_k denote reflection of the regular 3-gon across the vertex k. Explicitly, there are three distinct reflections ϕ_1, ϕ_2, and ϕ_3. Given any elements $x, y \in \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$, let yx denote the symmetry obtained by first performing x and subsequently performing y. Explicitly, $\phi_\ell \rho_k$ is the operation of first rotating through an angle of $-120k$ degrees and then reflecting about the vertex ℓ of the original arrangement of the labels 1, 2, and 3. Complete the table below by computing yx according to the rows x and columns y.

$_y\backslash^x$	ρ_1	ρ_2	ρ_3	ϕ_1	ϕ_2	ϕ_3
ρ_1	ρ_2	ρ_3		ϕ_3		
ρ_2						
ρ_3						
ϕ_1	ϕ_2			ρ_3		
ϕ_2						
ϕ_3						

Chapter 2

Group Theory I

Group theory is the study of algebraic structures equipped with associative binary operations that admit distinguished elements called the multiplicative identity and multiplicative inverses. Even though groups are often relatively tame to describe and possess simple arithmetic, their structure can be surprisingly complex. One of the most significant results in group theory is the development of the so-called solvable groups by the French mathematician Évariste Galois. Using the theory of solvable groups, it is possible to demonstrate that there is no analog to the quadratic formula for polynomials of degree greater than or equal to five. Group theory also appears in the study of coding theory, counting, and symmetries and in applications to biology, chemistry, and physics.

2.1 Groups (Definitions and Examples)

We will assume throughout this chapter that G is a nonempty set. Back in Section 1.1, we defined a **binary operation** on G as a function $* : G \times G \to G$ that sends $(g_1, g_2) \mapsto g_1 * g_2$. We say that G is a **group** with respect to $*$ whenever the following properties hold for the pair $(G, *)$.

1.) We have that $g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3$ for all elements $g_1, g_2, g_3 \in G$, i.e., $*$ is associative.

2.) G admits an element $e_G \in G$ such that $e_G * g = g = g * e_G$ for all elements $g \in G$.

3.) Given any element $g \in G$, there exists an element $g^{-1} \in G$ such that $g * g^{-1} = e_G = g^{-1} * g$.

Example 2.1.1. Let $\mathbb{Z}$ be the set of integers. Observe that (1.) addition of integers is associative; (2.) the integer 0 satisfies that $0 + n = n = n + 0$ for all integers n; and (3.) for any integer n, there exists an integer $-n$ such that $n + (-n) = 0 = -n + n$. Consequently, $(\mathbb{Z}, +)$ is a group. Crucially, we use the usual notation of additive inverses in place of the multiplicative notation above.

Example 2.1.2. Consider the collection $\mathbb{Z}_n$ of equivalence classes of integers modulo n. By Proposition 1.7.1, the distinct elements of $\mathbb{Z}_n$ are given by $r \pmod{n}$ for each integer $0 \leq r \leq n - 1$, hence it is nonempty. Using modular arithmetic, we may define an associative binary operation $+_n$ on $\mathbb{Z}_n$. Explicitly, we accomplish this by setting $r_1 \pmod{n} +_n r_2 \pmod{n} = (r_1 + r_2) \pmod{n}$. Of course, we may reduce $r_1 + r_2$ modulo n by computing the least non-negative integer x for which $r_1 + r_2 \equiv x \pmod{n}$; then, we may view $(r_1 + r_2) \pmod{n}$ as $x \pmod{n}$, hence $+_n$ is a binary operation on $\mathbb{Z}_n$. Considering that addition of integers is associative, $+_n$ is associative; the identity element of $\mathbb{Z}_n$ is simply 0 (mod n); and if $1 \leq r \leq n-1$, then the inverse of r (mod n) is simply $(n-r)$ (mod n). Ultimately, this goes to show that $(\mathbb{Z}_n, +_n)$ is a group. Once again, observe that we have used additive notation in place of the multiplicative notation of arbitrary groups.

Example 2.1.3. Consider any regular 3-gon. Let ρ_k denote rotation of the regular 3-gon through an angle of $-120k$ degrees for each integer $1 \leq k \leq 3$. Let ϕ_k denote reflection of the regular 3-gon across the vertex k for each integer $1 \leq k \leq 3$. Consider the set $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of symmetry-preserving rotations and reflections of a regular 3-gon. By Exercise 1.10.42, for any pair of elements $x, y \in D_3$, concatenation yx is an associative binary operation on D_3. Even more, we have that ρ_3 satisfies that $x\rho_3 = x = \rho_3 x$ for all elements $x \in D_3$, hence ρ_3 is the multiplicative identity of D_3; the rotations ρ_1 and ρ_2 are multiplicative inverses of one another; and the reflection ϕ_k is its own multiplicative inverse for each integer $1 \leq k \leq 3$. Consequently, we conclude that D_3 is a group under concatenation: it is typically called the **dihedral group** of order $6 = 2 \cdot 3$.

We say that a group $(G, *)$ is **abelian** if it holds that $g_1 * g_2 = g_2 * g_1$ for all elements $g_1, g_2 \in G$.

Example 2.1.4. Observe that the group $(\mathbb{Z}, +)$ is abelian because addition of integers is commutative. Likewise, for any elements r_1 (mod n) and r_2 (mod n) of $\mathbb{Z}_n$, we have that

$$r_1 \pmod{n} +_n r_2 \pmod{n} = (r_1 + r_2) \pmod{n} = (r_2 + r_1) \pmod{n} = r_2 \pmod{n} +_n r_1 \pmod{n}.$$

Consequently, the group $(\mathbb{Z}_n, +_n)$ is abelian, as well. By Exercise 1.10.42, on the other hand, the group D_3 of Example 2.1.3 is not abelian because we have that $\rho_1\phi_1 = \phi_3 \neq \phi_2 = \phi_1\rho_1$.

Example 2.1.5. Let $\mathbb{R}$ be the set of real numbers. Given any positive integer n, let $\mathbb{R}^{n \times n}$ denote the collection of all $n \times n$ real matrices. Under matrix addition, $\mathbb{R}^{n \times n}$ forms a group: the identity element of $\mathbb{R}^{n \times n}$ is the $n \times n$ zero matrix $O_{n \times n}$, and the inverse of an $n \times n$ real matrix A is the real matrix $-A$ whose (i, j)th entry is simply the (i, j)th entry of A with the opposite sign. Considering that addition of real numbers is commutative, it follows that $(\mathbb{R}^{n \times n}, +)$ is abelian.

Even more, let $\operatorname{GL}(n, \mathbb{R})$ denote the subset of $\mathbb{R}^{n \times n}$ consisting of invertible $n \times n$ matrices. Under matrix multiplication, $\operatorname{GL}(n, \mathbb{R})$ forms a group: the multiplicative identity of $\operatorname{GL}(n, \mathbb{R})$ is the $n \times n$ identity matrix $I_{n \times n}$, and the multiplicative inverse of an invertible $n \times n$ matrix A is A^{-1}. Considering that matrix multiplication is not commutative, $(\operatorname{GL}(n, \mathbb{R}), \cdot)$ is not abelian. We refer to this multiplicative group as the **general linear group** of size n over the field $\mathbb{R}$.

We refer to the cardinality of the underlying set defining a group as the **order** of the group. Observe that the additive group $(\mathbb{Z}_n, +_n)$ of the integers modulo n has order $|\mathbb{Z}_n| = n$, and the dihedral group $(D_3, \circ)$ of order six has order $|D_3| = 6$. On the other hand, the additive groups $(\mathbb{Z}, +)$ of integers and $(\mathbb{R}^{n \times n}, +)$ of real $n \times n$ matrices and the multiplicative group $(\operatorname{GL}(n, \mathbb{R}), \cdot)$ of real invertible $n \times n$ matrices have infinitely many elements, hence they each possess infinite order.

Remark 2.1.6. Unfortunately, even if a nonempty set G admits some associative binary operation $*$, it is not always true that $(G, *)$ is a group. Explicitly, multiplication of integers is an associative binary operation on the integers, and the integer 1 satisfies that $n \cdot 1 = n = 1 \cdot n$ for all integers n; however, the integer 0 admits no multiplicative inverse because it always holds that $n \cdot 0 = 0$, and it does not hold that $0 = 1$. Even if we consider the set $\mathbb{Z}^* = \mathbb{Z} \setminus \{0\}$ with respect to integer multiplication, we do not obtain a group because an integer n admits a multiplicative inverse n^{-1} in $\mathbb{Z}^*$ if and only if $n \cdot n^{-1} = 1$ if and only if $n^{-1} = \frac{1}{n}$ is an integer if and only if $n = \pm 1$.

2.2 Groups (Basic Properties and Subgroups)

We will continue to assume throughout this section that $(G, *)$ is a group, i.e., G is a nonempty set and $* : G \times G \to G$ is an associative binary operation with respect to which

1.) G admits an element $e_G \in G$ such that $e_G * g = g = g * e_G$ for all elements $g \in G$ and

2.) for each element $g \in G$, there exists an element $g^{-1} \in G$ such that $g * g^{-1} = e_G = g^{-1} * g$.

Our primary objective here is to explore some immediate properties and to illuminate the basic structure of groups. We begin by establishing the uniqueness of the identity and inverses.

Proposition 2.2.1. *If $(G, *)$ is a group, then the element e_G of property (1.) above is unique. Even more, for each element $g \in G$, the element $g^{-1} \in G$ of property (2.) above is unique.*

Proof. We must show that if e is any element of G with the property that $e * g = g = g * e$ for all elements $g \in G$, then $e = e_G$. Crucially, if this holds, then $e * e_G = e_G = e_G * e$ by assumption and $e_G * e = e$ by definition of e_G. But this implies that $e = e_G * e = e * e_G = e_G$, as desired.

Likewise, we must show that if h is any element of G with the property that $g * h = e_G = h * g$, then $h = g^{-1}$. Considering that $*$ is associative and $g^{-1} * g = e_G$, it follows that

$$h = e_G * h = (g^{-1} * g) * h = g^{-1} * (g * h) = g^{-1} * e_G = g^{-1}. \qquad \square$$

Consequently, we refer to the element e_G of property (1.) above as the (multiplicative) **identity element** of G and to the element g^{-1} of property (2.) above as the (multiplicative) **inverse** of g. Our next result greatly simplifies the task of finding multiplicative inverses.

Corollary 2.2.2. *If g is an element of a group $(G, *)$ and $g * h = e_G$, then $h * g = e_G$ and $h = g^{-1}$.*

Proof. By Proposition 2.2.1, it suffices to prove that $h * g = e_G$. By hypothesis that $g * h = e_G$, it follows that $(h * g) * (h * g) = h * (g * h) * g = h * e_G * g = h * g$. Consequently, multiplying both sides of the above identity $(h * g) * (h * g) = h * g$ by $(h * g)^{-1}$ yields the result. □

Usually, we will omit the operation $*$ of G and simply use concatenation, e.g., we will write $g_1 * g_2$ as $g_1 g_2$. By definition of a binary operation, for every pair of elements $g_1, g_2 \in G$, the product $g_1 g_2$ lies in G. Consequently, by property (2.) above, $g_1 g_2$ must have a multiplicative inverse.

Proposition 2.2.3. *If G is a group, then $(g_1 g_2)^{-1} = g_2^{-1} g_1^{-1}$ and $(g_1^{-1})^{-1} = g_1$ for all $g_1, g_2 \in G$.*

Proof. By Corollary 2.2.2, it suffices to verify that $(g_1 g_2)(g_2^{-1} g_1^{-1}) = e_G$ and $g_1^{-1} g_1 = e_G$. □

Existence of multiplicative inverses implies that every group possesses the **cancellation property**, i.e., if $g_1 g_2 = g_1 g_3$ for any elements $g_1, g_2, g_3 \in G$, then it must be the case that $g_2 = g_3$. Likewise, an identity $g_1 g_3 = g_2 g_3$ implies that $g_1 = g_2$. Often, we will invoke this property by using the expression "cancel on both sides" instead of saying "multiply both sides by the inverse."

Given any element $g \in (G, *)$ and any positive integer n, we may define the n-fold powers

$$g^n = \underbrace{g * g * \cdots * g}_{n \text{ times}} \text{ and } g^{-n} = \underbrace{g^{-1} * g^{-1} * \cdots * g^{-1}}_{n \text{ times}}$$

of g, and we adopt the convention that $g^0 = e_G$. Under these identifications, we have the following.

Proposition 2.2.4 (Group Exponent Laws)**.** *Let G be a group. Let m and n be any integers.*

1.) *We have that $g^m g^n = g^{m+n}$ for any element $g \in G$.*

2.) *We have that $(g^m)^n = g^{mn}$ for any element $g \in G$.*

3.) *If G is abelian, then $(g_1 g_2)^n = g_1^n g_2^n$ for all elements $g_1, g_2 \in G$.*

We leave the proofs of the Group Exponent Laws as Exercise 2.8.17. Often, we reserve the additive notation for abelian groups, hence in this case, the result is clear because

$$ng = \underbrace{g + g + \cdots + g}_{n \text{ times}} \text{ and } -ng = \underbrace{(-g) + (-g) + \cdots + (-g)}_{n \text{ times}}.$$

Given any nonempty set $H \subseteq G$, we say that H is a **subgroup** of G whenever $(H, *)$ is itself a group. Even more, if H is a nonempty proper subset of G, then $(H, *)$ is called a **proper subgroup** of G in this case. Every group admits a subgroup consisting solely of its identity element $\{e_G\}$; we refer to this as the **trivial subgroup** of G. Generally, though, there are other proper subgroups.

Example 2.2.5. Let $(\mathbb{Z}, +)$ be the abelian group of integers under addition. Given any integer n, consider the collection $n\mathbb{Z} = \{nk \mid k \text{ is an integer}\}$ of integer multiples of n. We can readily verify that $(n\mathbb{Z}, +)$ is a subgroup of $\mathbb{Z}$. Explicitly, the additive identity $0 = n \cdot 0$ lies in $n\mathbb{Z}$, and for any pair of integers k and ℓ, we have that $nk + n\ell = n(k + \ell)$ lies in $n\mathbb{Z}$, hence addition constitutes an associative binary operation on $n\mathbb{Z}$. Observe that the additive inverse of nk is $-nk = n(-k)$.

Example 2.2.6. Consider the dihedral group $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ consisting of the symmetry-preserving rotations and reflections of a regular 3-gon. Observe that $(\rho_1) = \{\rho_1, \rho_2, \rho_3\}$ is a subgroup of D_3 with respect to concatenation. Explicitly, we have that $\rho_j \rho_i = \rho_{i+j \pmod 3}$, hence every element of (ρ_1) has a multiplicative inverse in (ρ_1), and concatenation is an associative binary operation on (ρ_1). Even more, ρ_3 is the multiplicative identity of D_3, so it is the multiplicative identity of (ρ_1).

Example 2.2.7. Consider the general linear group $\operatorname{GL}(n, \mathbb{R})$ of size n over the field $\mathbb{R}$. Considering that $\det(AB) = \det(A)\det(B)$ for all $n \times n$ matrices, it follows that the subset

$$\operatorname{SL}(n, \mathbb{R}) = \{A \in \operatorname{GL}(n, \mathbb{R}) \mid \det(A) = 1\}$$

of $\operatorname{GL}(n, \mathbb{R})$ inherits the associative binary operation of matrix multiplication. By definition, every element of $\operatorname{SL}(n, \mathbb{R})$ has a multiplicative inverse, and the $n \times n$ identity matrix is the multiplicative identity of $\operatorname{SL}(n, \mathbb{R})$, hence it is a subgroup of $\operatorname{GL}(n, \mathbb{R})$ called the **special linear group**.

Remark 2.2.8. We cannot understate the importance of context when discussing the structure of groups and subgroups. Like we mentioned in Remark 2.1.6, a nonempty set with an associative binary operation need not be a group — even if it possesses a multiplicative identity. Likewise, a nonempty subset of a group is not necessarily a subgroup. Crucially, a subgroup must inherit the same binary operation as the group in which it is contained. Observe that the group $(\mathbb{R}^{n \times n}, +)$ of $n \times n$ real matrices contains $\operatorname{GL}(n, \mathbb{R})$ as a subset; however, $\operatorname{GL}(n, \mathbb{R})$ is not a subgroup of $\mathbb{R}^{n \times n}$ because the sum of two invertible matrices is not necessarily invertible. Even more, $\mathbb{R}^{n \times n}$ is not a group with respect to matrix multiplication because not all $n \times n$ matrices are invertible.

Often, it is convenient to use the following proposition and its two corollary to determine when a (nonempty) subset of a group itself constitutes a group with respect to the operation of the group.

Proposition 2.2.9 (Subgroup Test). *Let $(G, *)$ be a group, and let H be any subset of G. We have that $(H, *)$ is a subgroup of G if and only if the following three conditions hold.*

1.) *H contains the identity element e_G of G.*

2.) *We have that $h_1 * h_2 \in H$ for all elements $h_1, h_2 \in H$.*

3.) *We have that $h^{-1} \in H$ for all elements $h \in H$.*

Proof. Certainly, if the above three conditions hold for H, then in order to establish that $(H, *)$ is a group, we need only verify that $*$ is associative. But this holds by viewing H as a subset of G.

Conversely, suppose that $(H, *)$ is a subgroup of G. Condition (2.) holds because H is itself a group, hence it suffices to check that conditions (1.) and (3.) are satisfied. By assumption that H is a group, it admits an identity element e_H. Observe that as elements of G, we have that $e_H e_H = e_H = e_H e_G$. Cancellation on the left yields that $e_H = e_G$, as desired. Last, for all elements $h \in H$, there exists a unique element $h' \in H$ such that $hh' = e_H = h'h$. Considering that $e_H = e_G$, it follows that $hh' = e_G$, hence Proposition 2.2.2 yields that $h' = h^{-1}$ lies in H. □

Corollary 2.2.10 (Two-Step Subgroup Test). *Given a group $(G, *)$ and a nonempty set $H \subseteq G$, we have that $(H, *)$ is a subgroup of G if and only if*

1.) *we have that $h_1 * h_2 \in H$ for all elements $h_1, h_2 \in H$ and*

2.) *we have that $h^{-1} \in H$ for all elements $h \in H$.*

Proof. Clearly, if $(H, *)$ is a subgroup of G, then the stated properties of H must hold. Conversely, if we assume that the second and third conditions of the Subgroup Test hold, then the first condition holds because we have that $e_G = h * h^{-1}$ lies in H for all elements $h \in H$ and H is nonempty. □

Corollary 2.2.11 (One-Step Subgroup Test). *Given a group $(G, *)$ and a nonempty set $H \subseteq G$, we have that $(H, *)$ is a subgroup of G if and only if $h_1 * h_2^{-1} \in H$ for all elements $h_1, h_2 \in H$.*

Proof. Once again, if $(H, *)$ is a subgroup of G, then the stated property of H must hold. Conversely, by the Subgroup Test, it suffices to demonstrate that the following conditions holds.

1.) H contains the identity element e_G of G.

2.) We have that $h_1 * h_2 \in H$ for all elements $h_1, h_2 \in H$.

3.) We have that $h^{-1} \in H$ for all elements $h \in H$.

We verify condition (1.) by noting that $e_G = h_1 h_1^{-1}$ is in H for any element $h_1 \in H$. Consequently, condition (3.) follows because $h^{-1} = e_G h^{-1}$ for all elements $h \in H$ and $e_G \in H$. Last, condition (2.) holds by using Proposition 2.2.3 and condition (3.) to see that $h_1 * h_2 = h_1 * (h_2^{-1})^{-1} \in H$. □

Each of the above corollaries achieves one more step of reduction from the most tedious Subgroup Test; the most common form that we will use is the One-Step Subgroup Test.

Before we conclude this section, we provide an example to motivate the study of subgroups.

Example 2.2.12. We will soon come to see that like the order of a group, the subgroups admitted by a group provide a concrete way to distinguish that group from other groups of the same order. Consider the groups $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ with respect to modular addition. Both of these have order four, but we will demonstrate that they are distinct by showing that $\mathbb{Z}_4$ admits only one non-trivial proper subgroup while $\mathbb{Z}_2 \times \mathbb{Z}_2$ admits three non-trivial proper subgroups. If we drop the modulo n notation for this example, the elements of $\mathbb{Z}_4$ are $\{0, 1, 2, 3\}$, and its **Cayley table** is as follows.

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

By looking for the additive identity 0 in this table, we find that the only non-trivial subgroup of $(\mathbb{Z}_4, +_4)$ is given by $\{0, 2\}$. On the other hand, the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ admits the following Cayley table.

$(+_2, +_2)$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(0,0)$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(1,0)$	$(1,0)$	$(0,0)$	$(1,1)$	$(0,1)$
$(0,1)$	$(0,1)$	$(1,1)$	$(0,0)$	$(1,0)$
$(1,1)$	$(1,1)$	$(0,1)$	$(1,0)$	$(0,0)$

Once again, by looking for the additive identity $(0, 0)$ in this table, we find three non-trivial subgroups: they are $\{(0, 0),(1, 0)\}$, $\{(0, 0),(0, 1)\}$ and $\{(0, 0),(1, 1)\}$. Consequently, $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ are distinct groups of order four; the latter is famously called the **Klein four-group**.

2.3 Cyclic Groups

We begin our stratification of groups based on their structure by studying those groups that are "simplest" in the following sense. Given any group G and any element $g \in G$, we have that g^n lies in G for any integer n. Even more, these elements naturally give rise to a subgroup of G.

Proposition 2.3.1. *Given any group G and any element $g \in G$, the collection $\langle g \rangle = \{g^n \mid n \in \mathbb{Z}\}$ of integer powers of g forms a subgroup of G called the* **cyclic subgroup** *generated by g.*

Proof. Certainly, the set $\langle g \rangle$ is nonempty because it contains $g^0 = e_G$. Even more, for any elements $g^m, g^n \in \langle g \rangle$, we have that $g^m(g^n)^{-1} = g^m g^{-n} = g^{m-n}$ by the Group Exponent Laws so that $g^m(g^n)^{-1}$ lies in $\langle g \rangle$. We conclude by the One-Step Subgroup Test that $\langle g \rangle$ is a subgroup of G. □

Example 2.3.2. Consider the dihedral group $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of order six. Observe that the distinct powers of ρ_1 are given by ρ_1, ρ_1^2, and ρ_1^3. Consequently, we have that $\langle \rho_1 \rangle = \{\rho_1, \rho_1^2, \rho_1^3\}$. Considering that $\rho_1^2 = \rho_2$ and $\rho_1^3 = \rho_3$, this is the subgroup of D_3 consisting of all rotations of the regular 3-gon. On the other hand, for any reflection ϕ_k, we have that ϕ_k^2 does not affect any change, hence it is the identity ρ_3. Put another way, we have that $\langle \phi_k \rangle = \{\phi_k, \rho_3\}$ for each integer $1 \leq k \leq 3$.

Using additive notation, the cyclic subgroup generated by an element g of an abelian group $(G, +)$ is simply $\langle g \rangle = \{ng \mid n \in \mathbb{Z}\}$. We have unwittingly already encountered such groups.

Example 2.3.3. Observe that for any integer n, the cyclic subgroup of $(\mathbb{Z}, +)$ generated by n is given by $n\mathbb{Z} = \{qn \mid q \in \mathbb{Z}\} = \{\dots, -2n, -n, 0, n, 2n, \dots\}$, i.e., the collection of multiples of n.

Remark 2.3.4. If H is a subgroup of G that contains some element $g \in G$, then H contains the cyclic subgroup $\langle g \rangle$ because it contains all powers of g by the second property of the Subgroup Test. Consequently, the cyclic subgroup $\langle g \rangle$ is in this sense the "smallest" subgroup of G that contains g.

We will say that G is a **cyclic group** if it admits an element $g \in G$ such that $G = \langle g \rangle$. Like we have mentioned before, we will say in this case that g is a **generator** of G. By definition, the order of the cyclic subgroup $\langle g \rangle$ is the (possibly infinite) number of distinct elements of $\langle g \rangle$. If it happens that $\langle g \rangle$ is finite, then the **order** of g is the smallest positive integer $r = \operatorname{ord}(g)$ such that $g^r = e_G$. Consequently, the distinct elements of $\langle g \rangle$ are $g^0, g^1, \dots, g^{r-1}$ so that $\operatorname{ord}(g) = \#\langle g \rangle$.

Example 2.3.5. Every nonzero element of the additive group of integers $(\mathbb{Z}, +)$ has infinite order. Even more, every integer n can be written as $n \cdot 1$ or $(-n)(-1)$, hence we have that $\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle$. On the other hand, the subgroups $n\mathbb{Z}$ are proper for all integers n such that $|n| \geq 2$.

Example 2.3.6. Consider the abelian group $(\mathbb{Z}_{10}, +_{10})$ of equivalence classes of $\mathbb{Z}$ modulo 10 with respect to addition modulo 10. Observe that $\langle 5 \rangle = \{0, 5\}$, hence we have that $\operatorname{ord}(5) = 2$. On the other hand, we have that $\langle 1 \rangle = \{0, 1, 2, \dots, 9\} = \langle 9 \rangle$, hence both 1 and 9 generate $(\mathbb{Z}_{10}, +_{10})$.

Example 2.3.7. Consider the dihedral group $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of order six. By Example 2.3.2 and Exercise 1.10.42, every subgroup of D_3 is cyclic, but D_3 is not itself a cyclic group.

Even more, our next propositions illustrate several important properties of cyclic groups.

Proposition 2.3.8. *Every cyclic group is abelian.*

Proof. If G is cyclic, then G admits an element $g \in G$ such that $G = \{g^n \mid n \in \mathbb{Z}\}$. Consequently, for any elements $g_1, g_2 \in G$, there exist integers n_1 and n_2 such that $g_1 = g^{n_1}$ and $g_2 = g^{n_2}$. By the Group Exponent Laws, we conclude that $g_1 g_2 = g^{n_1} g^{n_2} = g^{n_1 + n_2} = g^{n_2 + n_1} = g^{n_2} g^{n_1} = g_2 g_1$. □

Corollary 2.3.9. *Groups that are not abelian are not cyclic.*

Proposition 2.3.10. *Every subgroup of a cyclic group is cyclic.*

Proof. We will assume that G is a cyclic group that is generated by some element $g \in G$. Explicitly, we will write that $G = \{g^n \mid n \in \mathbb{Z}\}$. Consider any subgroup $H \subseteq G$. Certainly, if $H = \{e_G\}$, then H is trivially cyclic. Consequently, we may assume that H admits a non-identity element $h \in H$. By hypothesis that G is cyclic, there exists an integer n such that $h = g^n$. By the Two-Step Subgroup Test and the Group Exponent Laws, we must have that $h^{-1} = (g^n)^{-1} = g^{-n}$ lies in H. Considering that h is not the identity element of G, we must have that $n > 0$ or $-n > 0$, so we may assume without loss of generality that $n > 0$. Ultimately, this analysis reveals that the collection $S = \{i \in \mathbb{Z}_{>0} \mid g^i \in H\}$ is nonempty, hence the Well-Ordering Principle ensures that S admits a smallest element s with respect to $\leq$. We will prove in the next paragraph that $H = \langle g^s \rangle$.

Consider any element $k \in H$. By assumption that G is cyclic, there exists an integer m such that $k = g^m$. By the Division Algorithm, there exist unique integers q and r such that $m = qs + r$ and $0 \leq r < s$. Consequently, we have that $k = g^m = g^{qs+r} = g^{qs} g^r$. By multiplying both sides of this identity (on the left) by g^{-qs}, we find that $g^r = g^{-qs} k$ lies in H. But this is impossible unless $r = 0$ because $0 \leq r < s$ and s is the smallest positive integer such that g^s lies in H. We conclude that $m = qs$, and every element of H can be written as $g^{qs} = (g^s)^q$ for some unique integer q. □

Corollary 2.3.11. *Every subgroup of $(\mathbb{Z}, +)$ is of the form $n\mathbb{Z}$ for some non-negative integer n.*

By paragraph preceding Example 2.3.5, the order of an element g of a group G is the smallest positive integer $r = \operatorname{ord}(g)$ such that $g^r = e_G$. Our next two results demonstrate that the order of a generator of a cyclic group determines the order of all other elements of the group.

Lemma 2.3.12. *Let G be a cyclic group. If g is generates G, then $g^n = e_G$ if and only if $\operatorname{ord}(g) \mid n$.*

Proof. Certainly, if $\operatorname{ord}(g) \mid n$, then $g^n = e_G$ because there exists an integer q such that $n = \operatorname{ord}(g)q$, and the Group Exponent Laws imply that $g^n = g^{\operatorname{ord}(g)q} = (g^{\operatorname{ord}(g)})^q = e_G^q = e_G$. Conversely, by the Division Algorithm, there exist unique integers q and r such $n = \operatorname{ord}(g)q + r$ and $0 \leq r < \operatorname{ord}(g)$. Observe that if r were nonzero, then it would be a smaller positive integer than $\operatorname{ord}(g)$ with the property that $g^r = e_G^q g^r = (g^{\operatorname{ord}(g)})^q = g^{\operatorname{ord}(g)q} g^r = g^{\operatorname{ord}(g)q+r} = g^n = e_G$ — a contradiction. □

Corollary 2.3.13. *If $g \in G$ has finite order and $g^n = e_G$, then $\operatorname{ord}(g) \mid n$.*

Proof. Every element of G generates a cyclic group $\langle g \rangle = \{g^k \mid k \in \mathbb{Z}\}$ of order $\operatorname{ord}(g)$. □

Proposition 2.3.14. *Let G be a cyclic group. If g generates G, then for any integer n, the order of g^n is $\operatorname{ord}(g)/\gcd(n, \operatorname{ord}(g))$.*

Proof. We denote $\operatorname{ord}(g) = d$. By definition, the order of g^n is the smallest positive integer r such that $(g^n)^r = g^{nr} = e_G$ by the Group Exponent Laws. By Corollary 2.3.13, if $g^{nr} = e_G$, then $d \mid nr$, hence r is the smallest positive integer such that $d \mid nr$. Considering that $\gcd(n, d)$ divides both n and d, we seek the smallest positive integer r such that $d/\gcd(n, d)$ divides $nr/\gcd(n, d)$. By Bézout's Identity, the integers $d/\gcd(n, d)$ and $n/\gcd(n, d)$ are relatively prime, hence Exercise 1.10.28 yields that $d/\gcd(n, d)$ divides r so that $r \geq d/\gcd(n, d) > 0$ and $r = d/\gcd(n, d)$. □

We conclude this section with a corollary that determines all generators of a cyclic group.

Corollary 2.3.15. *If G is a cyclic group that is generated by some element $g \in G$, then g^n is a generator of G for all integers n such that $\gcd(n, \operatorname{ord}(g)) = 1$.*

Proof. By Proposition 2.3.14, if n is any integer such that $\gcd(n, \operatorname{ord}(g)) = 1$, then the order of g^n is $\operatorname{ord}(g)$, and this is precisely the order of the entire group G; thus, g^n generates G. □

2.4 Complex Numbers as a Group Under Multiplication

Complex numbers arise naturally out of consideration of the solutions to certain polynomials with real coefficients. Explicitly, if x is a variable, then the quadratic polynomial $x^2 + 1$ does not possess a real root because the square of any real number is non-negative. We may therefore construct a solution i of the polynomial equation $x^2 + 1 = 0$; in this case, it holds that $i^2 = -1$ so that $i = \sqrt{-1}$.

Complex numbers are defined as the collection $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}$. Consequently, we may view i itself as a complex number. We refer to the real number a of the complex number $a + bi$ as the **real part** of $a + bi$, and the real number b is the **imaginary part** of $a + bi$. Complex numbers admit a notion of addition that allow us to view $\mathbb{C}$ as the two-dimensional real vector space $\mathbb{C} = \mathbb{R}\langle 1, i \rangle$. Explicitly, we define $(a + bi) + (c + di) = (a + b) + (c + d)i$ according to usual addition of vectors with respect to a basis. Consequently, the additive identity element of $\mathbb{C}$ is $0 + 0i$. We may also define multiplication of complex numbers by "foiling" the expression

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

We note that multiplication of complex numbers is associative, distributive, and commutative because multiplication of real numbers is associative, distributive, and commutative. Even more, one can readily verify that the multiplicative identity of $\mathbb{C}$ is $1 + 0i$. Last, observe that if a and b are nonzero, then $a+bi$ and $a-bi$ are nonzero, and we have that $(a+bi)(a-bi) = a^2+b^2$. Consequently, it follows that every nonzero complex numbers $a + bi$ admits a multiplicative inverse

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

We summarize the contents of this section thus far in the following.

Proposition 2.4.1. *Let $\mathbb{C}$ denote the complex numbers. Let $\mathbb{C}^\times = \mathbb{C} \setminus \{0 + 0i\}$.*

1.) *We have that $(\mathbb{C}, +)$ is an abelian group under complex addition; the identity is $0 + 0i$.*

2.) *We have that $(\mathbb{C}^\times, \cdot)$ is an abelian group under complex multiplication; the identity is $1 + 0i$.*

We refer to the complex number $a-bi$ as the **complex conjugate** of $a+bi$, and the real number $\sqrt{a^2 + b^2} = (a + bi)(a - bi)$ is the **modulus** of $a + bi$. Often, authors throughout the literature will denote $z = a + bi$; its complex conjugate $\bar{z} = a - bi$; and its modulus $|z| = \sqrt{a^2 + b^2}$.

Proposition 2.4.2. *Let $z = a + bi$ for some nonzero real numbers a and b.*

1.) *We have that $|\bar{z}| = |z|$ and $|z|^2 = z\bar{z}$.*

2.) *We have that $\left|\dfrac{z}{c}\right| = \dfrac{|z|}{|c|}$ for all nonzero real numbers c.*

3.) *We have that $z^{-1} = \dfrac{\bar{z}}{|z|^2}$ and $|z^{-1}| = \dfrac{1}{|z|}$.*

Graphically, complex numbers can be realized via their structure as the two-dimensional real vector space $\mathbb{C} = \mathbb{R}\langle 1, i\rangle \cong \mathbb{R} \times \mathbb{R}$. Consequently, the complex number $a + bi$ may be identified with the point (a, b) in the Cartesian plane whose x-axis corresponds to the real part of a complex number and whose y-axis corresponds to the imaginary part of the complex number. Using the polar coordinates interpretation of the Cartesian plane $\mathbb{R} \times \mathbb{R}$, we obtain the polar form for the complex numbers. Explicitly, any ordered pair of real numbers (a, b) can be written as $a = r\cos\theta$ and $b = r\sin\theta$ for some real numbers $r = \sqrt{a^2 + b^2}$ and $0 \leq \theta < 2\pi$, hence the complex number $a+bi$ can be written as $r(\cos\theta + i\sin\theta)$ such that r is the modulus of $a+bi$ and $0 \leq \theta < 2\pi$. Often, the most convenient way to express the complex number $r(\cos\theta + i\sin\theta)$ is as $r\operatorname{cis}\theta$.

Example 2.4.3. Consider the complex number $\sqrt{2}-i\sqrt{2}$. Observe that the modulus of $\sqrt{2}-i\sqrt{2}$ is $r = \sqrt{2+2} = 2$, hence it suffices to find $0 \leq \theta < 2\pi$. By viewing $\sqrt{2} - i\sqrt{2}$ as the point $(\sqrt{2}, -\sqrt{2})$ in the fourth quadrant of the Cartesian plane, we know that $3\pi/2 < \theta < 2\pi$. Even more, there exists an angle $0 < \phi < \pi/2$ such that $\theta = 2\pi - \phi$ and $\tan\phi = 1$. We conclude that $\phi = \arctan(1) = \pi/4$ so that $\theta = 7\pi/4$. Ultimately, we obtain the polar form $\sqrt{2} - i\sqrt{2} = 2\operatorname{cis}(7\pi/4)$.

Conversely, if we begin with the polar form of a complex number $\sqrt{3}\operatorname{cis}(2\pi/3)$, then unravelling this notation gives that $\sqrt{3}\operatorname{cis}(2\pi/3) = \sqrt{3}(\cos(2\pi/3) + i\sin(2\pi/3)) = -\sqrt{3}/2 + 3i/2$.

Even more, the polar representation can be the most efficient way to multiply complex numbers. We leave the proof of the following proposition as Exercise 2.8.42 for the reader.

Proposition 2.4.4. *We have that* $(r_1 \operatorname{cis} \theta_1)(r_2 \operatorname{cis} \theta_2) = r_1 r_2 \operatorname{cis}(\theta_1 + \theta_2)$.

Corollary 2.4.5 (DeMoivre's Theorem)**.** *We have that* $(r \operatorname{cis} \theta)^n = r^n \operatorname{cis}(n\theta)$ *for all integers* $n \geq 0$.

Proof. By the Principle of Ordinary Induction on $n \geq 0$, this follows from Proposition 2.4.4. □

Corollary 2.4.6. *We have that* $|z_1 z_2| = |z_1| \cdot |z_2|$ *for all complex numbers* z_1 *and* z_2.

Corollary 2.4.7. *We have that* $|z^n| = |z|^n$ *for all complex numbers* z *and all integers* n.

Example 2.4.8. One of the benefits of DeMoivre's Theorem is that it makes quick work of exponentiation of complex numbers that would normally require the Binomial Theorem. Explicitly, if we wish to compute $(\sqrt{2} - i\sqrt{2})^7$, then we simply recognize that $\sqrt{2} - i\sqrt{2} = 2\operatorname{cis}(7\pi/4)$ by Example 2.4.3, and DeMoivre's Formula gives $(\sqrt{2} - i\sqrt{2})^7 = 2^7 \operatorname{cis}(49\pi/4) = 128 \operatorname{cis}(\pi/4) = 64(\sqrt{2} + i\sqrt{2})$.

Consider the set $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. By definition, we have that a complex number z lies in $\mathbb{T}$ if and only if $|z| = 1$ if and only if the Cartesian coordinate representation of z lies in the unit circle. By Corollary 2.4.6, if $|z_1| = 1$ and $|z_2| = 1$, then $|z_1 z_2| = 1$, hence we have that $z_1 z_2$ lies in $\mathbb{T}$ for all elements $z_1, z_2 \in \mathbb{T}$. Even more, if $|z| = 1$, then $|z^{-1}| = 1$ by Proposition 2.4.2, hence z^{-1} lies in $\mathbb{T}$ for all elements $z \in \mathbb{T}$. By the Two-Step Subgroup Test, we conclude the following.

Proposition 2.4.9. *Let* $(\mathbb{C}^\times, \cdot)$ *denote the multiplicative group of complex numbers. We have that* $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ *is a subgroup of* $(\mathbb{C}^\times, \cdot)$ *called the* **circle group**.

Recall that a **root** of a polynomial $a_n x^n + \cdots + a_1 x + a_0$ with complex coefficients $a_0, a_1, \ldots, a_n$ is a complex number z such that $a_n z^n + \cdots + a_1 z + a_0 = 0$. Even though it is a classical theorem of algebra, the following is typically proved using complex analysis. Consequently, we will not attempt in this course to supply any justification ourselves; we will take it for granted.

Theorem 2.4.10 (Fundamental Theorem of Algebra)**.** *Let* n *be a positive integer. Every univariate polynomial of degree* n *with complex coefficients has exactly* n *(not necessarily distinct) roots.*

Consequently, the polynomial equation $z^3 = 1$ has exactly three solutions over the complex numbers. Certainly, one solution is simply $z = 1$; however, the other two solutions have nonzero imaginary part. Explicitly, we may factor $x^3 - 1 = (x-1)(x^2 + x + 1)$ such that $x^2 + x + 1$ has no real roots because the discriminant $b^2 - 4ac$ of the Quadratic Formula is negative. Generally, for any positive integer n, we refer to the roots of the polynomial $x^n - 1$ as the nth **roots of unity**.

Proposition 2.4.11. *If* n *is a positive integer, then the* n*th roots of unity are* $\operatorname{cis}(2k\pi/n)$ *for each integer* $0 \leq k \leq n-1$*; they form a cyclic subgroup of the circle group* $\mathbb{T}$.

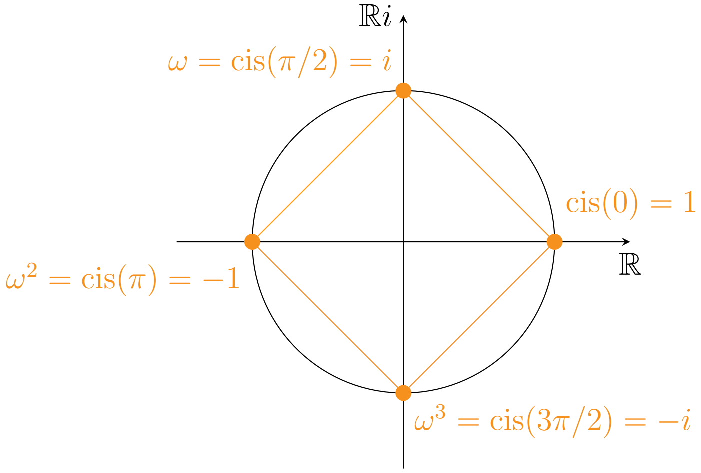
Proof. By the Fundamental Theorem of Algebra, it suffices to prove that $\operatorname{cis}(2k\pi/n)^n = 1$ for each integer $0 \leq k \leq n-1$ because for any pair of integers $0 \leq i < j \leq n-1$, we have that $\operatorname{cis}(2i\pi/n)$ and $\operatorname{cis}(2j\pi/n)$ are distinct. Observe that $\operatorname{cis}(2k\pi/n)^n = \operatorname{cis}(2kn\pi/n) = \operatorname{cis}(2k\pi) = 1$ holds by DeMoivre's Theorem. Last, the nth roots of unity form a cyclic subgroup of the circle group $\mathbb{T}$ once again by DeMoivre's Theorem because $\operatorname{cis}(2k\pi/n) = \operatorname{cis}(2\pi/n)^k$ for each integer $0 \leq k \leq n-1$. □

We refer to a generator of the cyclic subgroup of $\mathbb{T}$ consisting of the nth roots of unity as a **primitive** nth root of unity. By Propositions 2.4.11 and 2.3.15, we obtain the following.

Corollary 2.4.12. *If n is a positive integer, then $\operatorname{cis}(2k\pi/n)$ is a generator for the cyclic subgroup of $\mathbb{T}$ consisting of the nth roots of unity if and only if $\gcd(k, n) = 1$. Put another way, if we denote $\omega = \operatorname{cis}(2\pi/n)$, then $\omega^k = \operatorname{cis}(2k\pi/n)$ generates the nth roots of unity if and only if $\gcd(k, n) = 1$.*

Pictorially, the nth roots of unity consist of n equally-spaced points on the circumference of the unit circle; the distance between any two consecutive nth roots of unity is given by the angle measure of $2\pi/n$ radians; and a primitive nth root of unity is one for which successive rotation by the angle $2k\pi/n$ generates all of the nth roots of unity on the unit circle after $n - 1$ steps.

Example 2.4.13. Below are the fourth roots of unity on the unit circle.


$\mathbb{R}i$
$\omega = \operatorname{cis}(\pi/2) = i$
$\operatorname{cis}(0) = 1$
$\mathbb{R}$
$\omega^2 = \operatorname{cis}(\pi) = -1$
$\omega^3 = \operatorname{cis}(3\pi/2) = -i$

2.5 The Symmetric Group on n Letters

Given a nonempty set X, we may consider the set of bijections from the set X to itself. Conventionally, we use the Fraktur "S" with subscript X to denote this set. Explicitly, we have that $\mathfrak{S}_X = \{f : X \to X \mid f \text{ is injective and surjective}\}$. Certainly, the identity map $\operatorname{id}_X : X \to X$ defined by $\operatorname{id}_X(x) = x$ for every element $x \in X$ is a bijection, hence $\mathfrak{S}_X$ is nonempty. Given any two bijections $f, g : X \to X$, it follows that $f \circ g$ is a bijection from X to itself so that $\mathfrak{S}_X$ is closed with respect to function composition. Composition of functions is associative, so function composition is an associative binary operation on $\mathfrak{S}_X$. Last, for any bijection $f : X \to X$, there exists a unique function $f^{-1} : X \to X$ such that $f \circ f^{-1} = \operatorname{id}_X = f^{-1} \circ f$: indeed, for every element $x \in X$, there exists an element $y \in X$ such that $x = f(y)$ because f is surjective; this element $y \in X$ is unique because f is injective, so we may define $f^{-1}(x) = y$. We conclude therefore that $(\mathfrak{S}_X, \circ)$ is a group. We refer to $\mathfrak{S}_X$ as the **symmetric group on the set** X. Considering that a bijection is by definition a **permutation**, we say that $\mathfrak{S}_X$ the group of permutations of the set X.

Observe that if X is a finite set, then there exists a bijection between X and the set $\{1, 2, \dots, |X|\}$ that maps an element from X uniquely to some element of $\{1, 2, \dots, |X|\}$. Consequently, in order to study the group of permutations of a finite set, we may focus our attention on the permutation groups of the finite sets $[n] = \{1, 2, \dots, n\}$ for all positive integers n. We refer to the group $\mathfrak{S}_{[n]}$ as the **symmetric group on n letters**, and we adopt the shorthand $\mathfrak{S}_n$. Conventionally, the elements of $\mathfrak{S}_n$ are denoted by Greek letters such as sigma σ and tau τ; in particular, the identity function on $\mathfrak{S}_n$ is the Greek letter iota ι. Composition $\sigma \circ \tau$ is typically abbreviated by concatenation $\sigma\tau$, and the product $\sigma\tau$ is read from right to left (and not from left to right), as we are dealing with functions. Our first result concerning the symmetric group on n letters is the following.

Proposition 2.5.1. *We have that* $|\mathfrak{S}_n| = n! = n(n-1)(n-2)\cdots 2 \cdot 1$.

Proof. By definition, the elements of $[n]$ are bijections from $[n]$ to itself. Each bijection $\sigma : [n] \to [n]$ is uniquely determined by the values of $\sigma(1), \sigma(2), \dots, \sigma(n)$. Consequently, we may construct a bijection from $[n]$ to itself by specifying the value $\sigma(i)$ for each of the integers $1 \leq i \leq n$ in turn. Certainly, there are n distinct choices for the value of $\sigma(1)$. Once this value has been specified, there are $n-1$ distinct choices for the value of $\sigma(2)$ that differ from $\sigma(1)$. Once both $\sigma(1)$ and $\sigma(2)$ have been specified, there are $n-2$ distinct choices for the value of $\sigma(3)$ that differ from both $\sigma(1)$ and $\sigma(2)$. Continuing in this manner, there are $n-i+1$ distinct choices for the value of $\sigma(i)$ that differ from $\sigma(1), \sigma(2), \dots, \sigma(i-1)$ for each integer $1 \leq i \leq n$. By the Fundamental Counting Principle, there are $\prod_{i=1}^n (n-i+1) = n(n-1)(n-2)\cdots 2 \cdot 1 = n!$ distinct bijections from $[n]$ to itself. □

By Exercise 1.10.7, every element σ of $\mathfrak{S}_n$ is uniquely determined by $\sigma(1), \sigma(2), \dots, \sigma(n)$, hence we may visualize σ as the following $2 \times n$ array by listing $\sigma(i)$ beneath each integer $1 \leq i \leq n$.

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

Using the notation σ^n to denote the composition $\sigma \circ \cdots \circ \sigma$ of σ with itself n times, we have that $\sigma^2(i) = \sigma \circ \sigma(i) = \sigma(\sigma(i))$ for each integer $1 \leq i \leq n$, so we may build upon this array to list the image $\sigma^2(i)$ of $\sigma(i)$ under σ beneath $\sigma(i)$ for each integer $1 \leq i \leq n$ as follows.

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ \sigma^2(1) & \sigma^2(2) & \cdots & \sigma^2(n) \end{pmatrix}$$

Continuing in this manner, each of the integers $1 \leq i \leq n$ must eventually appear in the ith column twice because the integers $i, \sigma(i), \sigma^2(i), \dots, \sigma^n(i)$ cannot all be distinct. Let r_i denote the first row of the ith column for which it holds that $\sigma^{r_i}(i) = i$, i.e., r_i is the smallest positive integer not exceeding n for which the integers $i, \sigma(i), \dots, \sigma^{r_i - 1}(i)$ are all distinct. Observe that the columns of the resulting array allow us to easily read off the consecutive integers $i, \sigma(i), \sigma^2(i), \dots, \sigma^{r_i-1}(i)$. Considering that $\sigma(\sigma^{r_i-1}(i)) = \sigma^{r_i}(i) = i$, it follows that $i, \sigma(i), \sigma^2(i), \dots, \sigma^{r_i-1}(i)$ constitute a **cycle**; we will refer to the positive integer r_i as the **length** of the cycle $(i, \sigma(i), \sigma^2(i), \dots, \sigma^{r_i-1}(i))$, and we will say that the cycle itself is an r_i-cycle. Cycles of length two are commonly called **transpositions**. By definition, the order of a cycle as an element of the permutation group $\mathfrak{S}_n$ is its length, i.e., if σ is an r_i-cycle, then $\operatorname{ord}(\sigma) = r_i$. Conventionally, cycles are written without commas, but we will use them when convenient. We will also say that two cycles $(a_1, a_2, \dots, a_k)$ and $(b_1, b_2, \dots, b_\ell)$ are **disjoint** if the entries a_i and b_j are pairwise distinct for all pairs of integers $1 \leq i \leq k$ and $1 \leq j \leq \ell$.

Example 2.5.2. We have already encountered the symmetric group $\mathfrak{S}_3$ on three letters in a different guise. Consider the dihedral group $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of order six whose elements ρ_k are the rotations about an angle of $-120k$ degrees and whose elements ϕ_k are the reflections about the vertex k of a regular 3-gon. Going back to the main example of Section 1.8, we have the following.

$$\rho_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \qquad \rho_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \qquad \rho_3 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

$$\phi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \qquad \phi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \qquad \phi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

Considering that $\rho_3(i) = i$ for each integer $1 \leq i \leq 3$, it follows that ρ_3 is the identity element of $\mathfrak{S}_3$. Carrying out the process of the previous paragraph, we obtain the cycles of $\mathfrak{S}_3$. Generally, the identity permutation ι is a cycle of length one; this can be verified here by looking at ρ_3 above. Each of the other above permutations is not a cycle because the entries of some column are distinct. Consequently, we must apply the permutations until each column has a repeated integer.

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

Consequently, the permutation ρ_1 is the 3-cycle (132); the permutation ρ_2 is the 3-cycle (123); the permutation ϕ_1 is the 2-cycle (23); the permutation ϕ_2 is the 2-cycle (13); and the permutation ϕ_3 is the 2-cycle (12). We will explore this phenomenon when we discuss general dihedral groups. By Exercise 1.10.42, we have that $\phi_1 \rho_1 \neq \rho_1 \phi_1$, hence the symmetric group is not necessarily abelian.

Example 2.5.3. Consider the following permutation σ in **two-line notation**.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 8 & 4 & 1 & 7 & 6 & 3 \end{pmatrix}$$

Computing the disjoint cycles of σ amounts to building upon the above array row-by-row until each of the integers $1 \leq i \leq 8$ appears in the ith column twice. Explicitly, we have the following array.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 8 & 4 & 1 & 7 & 6 & 3 \\ 5 & 1 & 3 & 4 & 2 & 6 & 7 & 8 \\ 1 & 2 & 8 & 4 & 5 & 7 & 6 & 3 \end{pmatrix}$$

Consequently, the disjoint cycles of σ are (125), (38), (4), and (67). Even though we have used two-line notation to express most permutations up to this point, it is occasionally most convenient to adopt the **one-line notation** of a permutation σ by specifying all of its disjoint cycles. Explicitly, in one-line notation, we may write $\sigma = (125)(38)(4)(67)$. Cycles are by definition expressed in one-line notation. Even more, our next two propositions demonstrate that it is possible to find the one-line notation of any permutation and that the representation of a permutation as a product of disjoint cycles is unique up to a rearrangement of the non-trivial cycles appearing in the product.

Proposition 2.5.4. *Every permutation can be written as a product of disjoint cycles.*

Proof. Given any permutation σ of $[n]$, observe that the integers $1, \sigma(1), \dots, \sigma^n(1)$ cannot all be distinct. Consequently, there exists an integer $1 \leq r_1 \leq n-1$ such that $\sigma^{r_1}(1) = 1$; the integers $1, \sigma(1), \dots, \sigma^{r_1 - 1}(1)$ are distinct; and $\sigma_1 = (1, \sigma(1), \dots, \sigma^{r_1-1}(1))$ is a cycle of length r_1. Consider the smallest integer i_2 that does not appear as an entry of σ_1. Once again, the integers $i_2, \sigma(i_2), \dots, \sigma^n(i_2)$ cannot all be distinct, so there must be an integer $1 \leq r_2 \leq n-1$ such that $\sigma^{r_2}(i_2) = i_2$. Like before, we obtain a cycle $\sigma_2 = (i_2, \sigma(i_2), \dots, \sigma^{r_2-1}(i_2))$ of length r_2. Crucially, we note that σ_1 and σ_2 are disjoint. Explicitly, if it were the case that $\sigma^i(1) = \sigma^j(i_2)$ for some integers $0 \leq i \leq r_1 - 1$ and $0 \leq j \leq r_2 - 1$, then it would follow that $\sigma^{r_2 - j + i}(1) = \sigma^{r_2}(i_2) = i_2$ so that i_2 appears as an entry of σ_1 — a contradiction. Continuing in this manner, we may construct disjoint cycles $\sigma_1, \sigma_2, \dots, \sigma_k$ such that every element of $[n]$ lies in one and only one cycle and $\sigma = \sigma_1 \cdots \sigma_k$. □

Proposition 2.5.5. *Every pair of disjoint cycles σ and τ commute, i.e., we have that $\sigma\tau = \tau\sigma$.*

Proof. Consider an integer $1 \leq i \leq n$. Crucially, we note that if i is does not appear in the one-line notation of σ, then $\sigma(i) = i$. Considering that σ and τ are disjoint, it follows that $\tau(i)$ cannot be an entry of σ in one-line notation, hence we have that $\sigma\tau(i) = \tau(i) = \tau\sigma(i)$. Consequently, it suffices to consider the case that i appears in the one-line notation of σ. Consider the entry j of σ corresponding to $\sigma(i) = j$ in one-line notation. By assumption that σ and τ are disjoint, neither of the integers i and j can appear as an entry of τ, hence we have that $\tau(i) = i$ and $\tau(j) = j$. We conclude therefore that $\sigma\tau(i) = \sigma(i) = j = \tau(j) = \tau\sigma(i)$. By the Law of the Excluded Middle, every integer $1 \leq i \leq n$ either appears in the one-line notation of σ or not, so our proof is complete. □

Corollary 2.5.6. *Every permutation can be written as a product of disjoint cycles in a manner that is unique up to the arrangement of the disjoint cycles appearing in the product.*

Consequently, we refer to the representation of a permutation σ as a product $\sigma_k \cdots \sigma_2 \sigma_1$ of disjoint cycles as the **cycle decomposition** of σ. Because the order of the disjoint cycles does not matter, we will henceforth simplify the notation to $\sigma = \sigma_1 \cdots \sigma_k$. Later, it will be important to note that if the cycles σ_i have length r_i for each integer $1 \leq i \leq k$, then $r_1 + \cdots + r_k = n$ because each of the integers $1, 2, \dots, n$ appears in one and only one cycle σ_i. We are now able to prove the following.

Proposition 2.5.7. *Let σ be any permutation with cycle decomposition $\sigma_1 \cdots \sigma_k$. Let r_i denote the length of the cycle σ_i. We have that* $\operatorname{ord}(\sigma) = \operatorname{lcm}(r_1, \dots, r_k)$.

Proof. By Proposition 2.5.5, the disjoint cycles $\sigma_1, \dots, \sigma_k$ commute, hence we have that

$$\operatorname{ord}(\sigma) = \operatorname{ord}(\sigma_1 \cdots \sigma_k) = \min\{r \geq 1 \mid (\sigma_1 \cdots \sigma_k)^r = \iota\} = \min\{r \geq 1 \mid \sigma_k^r \cdots \sigma_1^r = \iota\}.$$

We claim that $\sigma_k^r \cdots \sigma_1^r = \iota$ if and only if $\sigma_i^r = \iota$ for each integer $1 \leq i \leq k$. Certainly, if $\sigma_i^r = \iota$ for each integer $1 \leq i \leq k$, then $\sigma_k^r \cdots \sigma_1^r = \iota$. Conversely, if $\sigma_i^r \neq \iota$ for some integer $1 \leq i \leq k$, then $\sigma_k^r \cdots \sigma_1^r \neq \iota$ because the cycles $\sigma_1, \dots, \sigma_k$ are disjoint. Consequently, we conclude that

$$\begin{aligned} \operatorname{ord}(\sigma) &= \min\{r \geq 1 \mid \sigma_i^r = \iota \text{ for each integer } 1 \leq i \leq k\} \\ &= \min\{r \geq 1 \mid \operatorname{ord}(\sigma_i) = r_i \text{ divides } r \text{ for each integer } 1 \leq i \leq k\} = \operatorname{lcm}(r_1, \dots, r_k). \end{aligned}$$

Explicitly, the second equality follows from Corollary 2.3.13, and the third equality follows from the definition of the least common multiple that precedes Exercise 1.10.31. □

Permutations of order two are called **involutions**. By Propositions 2.5.6 and 2.5.7, a permutation is an involution if and only if its cycle decomposition is the product of disjoint transpositions.

Example 2.5.8. Consider the permutation σ of Example 2.5.3 with disjoint cycles (125), (38), (4), and (67). By Proposition 2.5.6, its cycle decomposition is given by $\sigma = (125)(38)(4)(67)$, hence we have that $\operatorname{ord}(\sigma) = \operatorname{lcm}(3, 2, 1, 2) = \operatorname{lcm}(6, 1, 2) = \operatorname{lcm}(6, 2) = 6$ by Proposition 2.5.7.

Corollary 2.5.6 guarantees that every permutation can be written as a product of disjoint cycles uniquely up to the arrangement of the factors. Consequently, if we are handed the cycle decomposition of a permutation, it is natural to ask how to reconstruct its two-line notation representation.

Algorithm 2.5.9. We can reconstruct the two-line notation for any permutation σ from its cycle decomposition $\sigma_1 \cdots \sigma_k$ as follows.

1.) Find the largest positive integer n lying in some cycle σ_i.

2.) Build a $2 \times n$ array with the integers $1, 2, \dots, n$ listed in ascending order in the first row.

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ & & & \end{pmatrix}$$

3.) Begin to fill the space below the integer 1 by first locating the integer 1 in some cycle σ_{i_1}.

4.) If 1 is immediately followed by a right parenthesis, then $\sigma(1)$ is the integer that begins the cycle σ_{i_1}; otherwise, $\sigma(1)$ is the integer that immediately follows 1 in the cycle σ_{i_1}.

5.) Repeat the above two steps until the integers $\sigma(1), \sigma(2), \dots, \sigma(n)$ are all found.

Example 2.5.10. Consider the permutation $\sigma = (135)(48)(276)$. Observe that the largest positive integer n lying in some cycle is $n = 8$. Consequently, we will build the two-line notation of σ from the the 2×8 array with the integers $1, 2, \dots, 8$ listed in ascending order in the first row.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & & & & & & & \end{pmatrix}$$

Observe that 1 lies in the cycle (135); it is immediately followed by 3, hence we have that $\sigma(1) = 3$. Observe that 2 lies in the cycle (276); it is immediately followed by 7, hence we have that $\sigma(2) = 7$. Observe that 3 lies in the cycle (135); it is immediately followed by 5, hence we have that $\sigma(3) = 5$. Observe that 4 lies in the cycle (48); it is immediately followed by 8, hence we have that $\sigma(4) = 8$. Observe that 5 lies in the cycle (135); it is immediately followed by a right parenthesis, hence we have that $\sigma(5) = 1$. Continuing in this manner, we obtain the two-line notation for σ.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 7 & 5 & 8 & 1 & 2 & 6 & 4 \end{pmatrix}$$

Unfortunately, there is no guarantee that we will be handed the cycle decomposition of a permutation; rather, if we are given a product of (not necessary disjoint) cycles, the following algorithm generalizes the method of Algorithm 2.5.9 to find the two-line notation for the resulting permutation.

Algorithm 2.5.11. We reconstruct the two-line notation for any permutation $\sigma = \sigma_1 \cdots \sigma_k$ that is a product of (not necessarily disjoint) cycles $\sigma_1, \dots, \sigma_k$ as follows.

1.) Find the largest positive integer n lying in some cycle σ_i.

2.) Build a $2 \times n$ array with the integers $1, 2, \dots, n$ listed in ascending order in the first row.

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ & & & \end{pmatrix}$$

3.) Begin to fill the space below the integer 1 by first locating the integer 1 in the cycle σ_{i_1} that is farthest to the right among the cycles in the product $\sigma_1 \cdots \sigma_k$.

4.) If 1 is immediately followed by a right parenthesis, then 1 maps to the integer b_{i_1} that begins σ_{i_1}; otherwise, 1 maps to the integer n_{i_1} that immediately follows 1 in σ_{i_1}.

5.) Locate the integer b_{i_1} or n_{i_1} in the cycle that is farthest to the right among the cycles in the product $\sigma_1 \cdots \sigma_{i_1 - 1}$; then, repeat the third step. If $i_1 = 1$, then $\sigma(1) = b_{i_1}$ or $\sigma(1) = n_{i_1}$.

6.) Repeat the third and fourth steps until it is not possible; the last integer found is $\sigma(1)$.

7.) Repeat the above four steps until the integers $\sigma(1), \sigma(2), \dots, \sigma(n)$ are found.

One useful way to think about and to understand the mechanics of this algorithm is that function composition is read from right to left. Considering that each cycle is itself a permutation, in order to find the image of i under the composite function $\sigma_1 \cdots \sigma_k$, we follow the image of i under the successive composite functions σ_k, $\sigma_{k-1}\sigma_k$, etc., up to $\sigma_1 \cdots \sigma_k$. Further, if the integer $\sigma_i(i)$ does not appear in σ_{i+1}, then $\sigma_{i+1}\sigma_i(i) = \sigma_i(i)$, hence we must only consider the cycle farthest to the right that contains the integer under consideration: all cycles that do not contain $\sigma_i(i)$ will fix $\sigma_i(i)$.

Example 2.5.12. We will write the permutation $\sigma = (134)(45)(14)(23)$ of $\mathfrak{S}_5$ in two-line notation. Using the algorithm above, we find that 1 maps to 4; then, 4 maps to 5; and finally, 5 does not appear in any cycle to the left of (45), so it follows that $\sigma(1) = 5$. We find next that 2 maps to 3; then, 3 maps to 4; and there are no permutations to the left of (134), so it follows that $\sigma(2) = 4$. We find next that 3 maps to 2 in the last cycle, and 2 does not appear in any cycle to the left of (23), so it follows that $\sigma(3) = 2$. We find next that 4 maps to 1; then, 1 maps to 3; and there are no permutations to the left of (134), so it follows that $\sigma(4) = 3$. Last, we find that 5 maps to 4; then, 4 maps to 1; and there are no permutations to the left of (134), so it follows that $\sigma(5) = 1$. We conclude therefore that σ can be written in two-line notation as follows.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 2 & 3 & 1 \end{pmatrix}$$

Often, it is advantageous to omit the cycles of length one when describing a permutation via its cycle decomposition. For instance, the permutation $\sigma = (123)$ can be viewed as the 3-cycle

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

in $\mathfrak{S}_3$ or as the permutation τ in $\mathfrak{S}_n$ for any integer $n \geq 3$ that acts as σ on the subset $\{1, 2, 3\}$ and acts as the identity on the subset $\{4, \dots, n\}$. Consequently, a permutation is uniquely determined by its cycle decomposition (excluding 1-cycles) regardless of the symmetric group to which it belongs.

Proposition 2.5.13. *For every integer $n \geq 3$, the symmetric group $\mathfrak{S}_n$ is not abelian.*

Proof. Consider the cycles $\sigma = (12)$ and $\tau = (13)$ in $\mathfrak{S}_3$. By the paragraph above, we may view σ and τ as elements of $\mathfrak{S}_n$ for every integer $n \geq 3$. Considering that $\sigma\tau = (12)(13) = (132)$ is not equal to $\tau\sigma = (13)(12) = (123)$, we conclude that $\mathfrak{S}_n$ is not abelian for any integer $n \geq 3$. □

Computing the inverse of a permutation can be quite tedious; however, if we have a permutation σ written as its cycle decomposition $\sigma = \sigma_1 \cdots \sigma_k$, then the inverse of σ can be obtained as follows. Observe that if σ_i has length r_i, then $\sigma_i \sigma_i^{r_i - 1} = \sigma_i^{r_i} = \iota = \sigma_i^{r_i - 1}\sigma_i$. Consequently, we have that $\sigma_i^{-1} = \sigma_i^{r_i - 1}$. Considering that disjoint cycles commute, we have the following.

Proposition 2.5.14. *Let σ be any permutation with cycle decomposition $\sigma_1 \cdots \sigma_k$ and cycle type $(r_1, \dots, r_k)$. We have that $\sigma^{-1} = \sigma_1^{r_1 - 1} \cdots \sigma_k^{r_k - 1}$.*

Ultimately, Proposition 2.5.14 makes small work of the matter of finding inverses of permutations written in cycle decomposition: observe that $(a_1, \dots, a_k)^{-1} = (a_1, a_k, a_{k-1}, \dots, a_3, a_2)$.

2.6 Dihedral Groups

Previously, in Section 1.8, we considered the rigid motions of a regular n-gon. Explicitly, we noticed that rotation of a regular n-gon through an angle of $-360k/n$ degrees produces a copy of the regular n-gon with the ith vertex in place of the $(i+k)$th vertex (modulo n). Likewise, the reflection of any regular n-gon across a vertex k swaps the labels of the vertices other than k according to the parity of n. Our immediate aim is to establish that the collection D_n of these rotations and reflections of a regular n-gon constitutes a subgroup of order $2n$ of the symmetric group $\mathfrak{S}_n$ on n letters.

Proposition 2.6.1. *Let $n \geq 3$ be an integer. Let D_n denote the set of symmetry-preserving rotations and reflections of a regular n-gon. Every element of D_n can be written as a product of some distinguished elements $r, s \in D_n$ such that r has order n; s has order two; and $sr = r^{n-1}s$.*

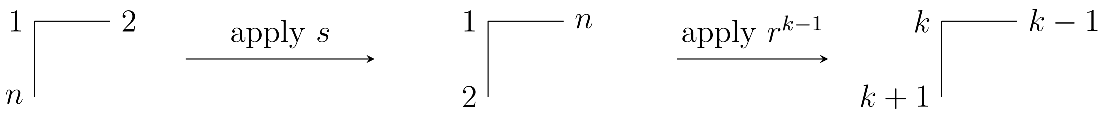
Proof. Consider the rotation r of the regular n-gon through the angle $-360/n$ degrees and the reflection s of the regular n-gon about the vertex 1. Conventionally, we denote by sr the composite function $s \circ r$. Observe that r has order n: indeed, it follows that r^k is the rotation of the regular n-gon through an angle of $-360k/n$ degrees, and the rational numbers $-360k/n$ are distinct for each integer $1 \leq k \leq n$. On the other hand, r^{n+1} is the rotation through the angle $-360 - 360/n$ degrees; this has the same effect as rotating about the angle $-360/n$ degrees, hence we conclude that $r^{n+1} = r$, and the order of r is n. Certainly, the order of s is two because reflection about the vertex 1 twice does not swap any of the vertices, i.e., we have that s^2 is the identity permutation.

We will demonstrate next that every reflection of the regular n-gon can be achieved by performing r and s sequentially in some order. We claim that $r^{k-1}s$ is a distinct reflection of the regular n-gon for each integer $1 \leq k \leq n$. Observe that s has the effect of labelling the vertices $1, 2, \dots, n$ of the regular n-gon counterclockwise (as opposed to the usual clockwise order); then, r^{k-1} replaces vertex

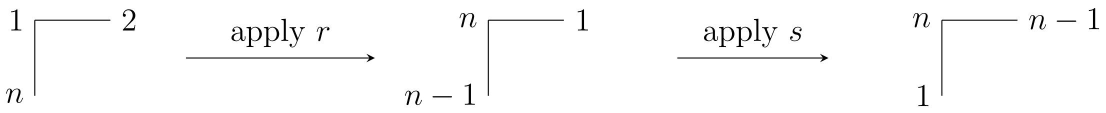
1 with the label k, vertex n with the label $k-1$, and vertex 2 with the label $k+1$ (modulo n). Consequently, we conclude that $r^{k-1}s$ is a distinct reflection for each integer $1 \leq k \leq n$. Considering that there are n reflections of any regular n-gon, they must be precisely $s, rs, r^2s, \dots, r^{n-1}s$.

Last, we attend to sr. Observe that r has the effect of labelling vertex 1 with label n, vertex 2 with label 1, and vertex n with label $n-1$; then, under s, vertex 1 retains the label n, vertex 2 obtains the label $n-1$, and vertex n obtains the label 1. Put another way, we have that $sr = r^{n-1}s$. □

$r^{k-1}s$:



sr:



Proposition 2.6.2. *Let $n \geq 3$ be an integer. Let D_n denote the set of symmetry-preserving rotations and reflections of a regular n-gon. We have that D_n is a subgroup of $\mathfrak{S}_n$ of order $2n$.*

Proof. Every symmetry-preserving rotation or reflection of the regular n-gon can be viewed as a permutation of the integers $1, \dots, n$, hence D_n is a subset of $\mathfrak{S}_n$. By Proposition 2.6.1, the distinct elements of D_n are $r, r^2, \dots, r^n, s, rs, r^2s, \dots, r^{n-1}s$, hence D_n has order $2n$. Even more, every rotation r^k has a multiplicative inverse r^{n-k}, and every reflection r^ks is its own multiplicative inverse. Consequently, by the Two-Step Subgroup Test, it suffices to prove that $xy \in D_n$ for any elements $x, y \in D_n$. Certainly, the product of two rotations is a rotation, hence we may assume that x and y are not both rotations. By Proposition 2.6.1, we may assume first that $x = r^k$ and $y = r^\ell s$ for some integers $1 \leq k, \ell \leq n$. Observe that $xy = r^{k+\ell}s$; by taking the exponent $k+\ell$ modulo n, we conclude that xy lies in D_n. Conversely, if $x = r^ks$ and $y = r^\ell$ for some integers $1 \leq k, \ell \leq n$, we conclude that $xy = r^ksr^\ell = r^kr^{\ell(n-1)}s = r^{\ell(n-1)+k}s$ lies in D_n. Last, if $x = r^ks$ and $y = r^\ell s$ for some integers $1 \leq k, \ell \leq s$, then $xy = r^ksr^\ell s = r^kr^{\ell(n-1)}s^2 = r^{\ell(n-1)+k}$ is in D_n. □

We will henceforth refer to D_n as the **dihedral group** of order $2n$ in light of Proposition 2.6.2. We adopt the convention that the identity of this group is 1; it is obtained from the original arrangement of the n vertices of the regular n-gon in clockwise order by doing nothing.

Example 2.6.3. Consider the dihedral group D_4 of order 8, i.e., the group of symmetry-preserving rotations and reflections of a square. By Proposition 2.6.1, the elements of D_4 are the identity element 1; the rotation r by $-90°$; the rotation r^2 by $-180°$; the rotation r^3 by $-270°$; the reflection s across vertices 1 and 3; the reflection rs across the line perpendicular to side 12; the reflection r^2s across the vertices 2 and 4; and the reflection r^3s across the line perpendicular to side 14.

Considering that every symmetry-preserving rotation and reflection of a square is a bijection from the set $\{1, 2, 3, 4\}$ to itself, we can realize each of the eight elements of D_4 as a permutation of the integers 1, 2, 3, and 4. Explicitly, the following hold in two-line and one-line notation.

$$1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = (1) \qquad s = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} = (24)$$

$$r = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = (1234) \qquad rs = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (12)(34)$$

$$r^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = (13)(24) \qquad r^2s = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = (13)$$

$$r^3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} = (1432) \qquad r^3s = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = (14)(23)$$

We leave the above details for the reader to check pictorially in Exercise 2.8.48.

Example 2.6.4. Consider the dihedral group D_5 of order 10, i.e., the group of symmetry-preserving rotations and reflections of a regular pentagon. By Proposition 2.6.1, the elements of D_5 are the identity element 1; the rotation r by $-72°$; the rotation r^2 by $-144°$; the rotation r^3 by $-216°$; the rotation r^4 by $-288°$; the reflection s across vertex 1; the reflection rs across vertex 4; the reflection r^2s across vertex 2; the reflection r^3s across vertex 5; and the reflection r^4s across vertex 3.

$$\begin{aligned} 1 &= (1) & s &= (24)(35) \\ r &= (12345) & rs &= (12)(35) \\ r^2 &= (13524) & r^2s &= (13)(45) \\ r^3 &= (14253) & r^3s &= (14)(23) \\ r^4 &= (15432) & r^4s &= (15)(24) \end{aligned}$$

We leave the above details for the reader to check pictorially in Exercise 2.8.49.

2.7 Chapter 2 Overview

This section is currently under construction.

2.8 Chapter 2 Exercises

2.8.1 Groups (Definitions and Examples)

Exercise 2.8.1. Explain why $\mathbb{R}$ does not form a group with respect to multiplication.

Exercise 2.8.2. Use the definition of a group to prove that $\{-1, 1\}$ forms an abelian group with respect to multiplication; you may assume that integer multiplication is associative.

Exercise 2.8.3. Let i denote the complex number satisfying that $i^2 = -1$. Use the definition of a group to prove that $\{-1, 1, -i, i\}$ forms an abelian group with respect to multiplication; you may assume that complex multiplication is associative and commutative.

Exercise 2.8.4. Use the definition of a group to prove that $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$ forms an abelian group with respect to multiplication; you may use that real multiplication is associative and commutative.

Exercise 2.8.5. Let G be a group. We say that an element $g \in G$ is **idempotent** if it holds that $g^2 = g$. Prove that the only idempotent element of a group is the identity element e_G.

Exercise 2.8.6. Given a positive integer n, let $\mathbb{Z}_n$ denote the collection of equivalence classes of the integers modulo n. Explain why $\mathbb{Z}_n$ does not form a group with respect to multiplication modulo n.

(**Hint:** Exercise 1.10.36 could be a useful reference here.)

Exercise 2.8.7. Given a prime integer p, let $\mathbb{Z}_p$ denote the collection of equivalence classes of the integers modulo p. Prove that $\mathbb{Z}_p$ forms an abelian group with respect to multiplication modulo p.

(**Hint:** Exercise 1.10.38 could be a useful reference here.)

Exercise 2.8.8. Given a positive integer n, let $\mathbb{Z}_n$ denote the collection of equivalence classes of integers modulo n. Prove that the set $\mathbb{Z}_n^* = \{a \in \mathbb{Z}_n \mid \gcd(a, n) = 1\}$ forms a group with respect to multiplication modulo n; this group is called the **multiplicative group of integers modulo** n.

Exercise 2.8.9. Consider the nonempty set $G = \mathbb{R} \setminus \{-1\}$.

(a.) Prove that $* : G \times G \to G$ defined by $x * y = x + y + xy$ is a binary operation on G.

(b.) Use the definition of a group to prove that $(G, *)$ is an abelian group.

Exercise 2.8.10. Prove that a group G is abelian if $(gh)^2 = g^2h^2$ for all elements $g, h \in G$.

(**Hint:** Compute $(gh)^2$ in two ways; then, compare your results.)

Exercise 2.8.11. Prove that a group G is abelian if $g^2 = e_G$ for every element $g \in G$.

(**Hint:** Use the fact that G is abelian if and only if $ghg^{-1}h^{-1} = e_G$ for all elements $g, h \in G$.)

Exercise 2.8.12. Prove that a group G is abelian if $gh = g^{-1}h^{-1}$ for all elements $g, h \in G$.

(**Hint:** Essentially, this follows as a corollary of Exercise 2.8.11.)

Exercise 2.8.13. Prove that a group G is abelian if $g^3 = e_G$ and $g^4h = hg$ for all elements $g, h \in G$.

Exercise 2.8.14. Prove that any group of order four must be abelian.

2.8.2 Groups (Basic Properties and Subgroups)

Exercise 2.8.15. Use Exercise 2.8.4 and the One-Step Subgroup Test to prove that $\{-1, 1\}$ forms an abelian group with respect to multiplication.

Exercise 2.8.16. Let i denote the complex number satisfying that $i^2 = -1$. Use the One-Step Subgroup Test to prove that $\{-1, 1, -i, i\}$ forms an abelian group with respect to multiplication.

Exercise 2.8.17. Prove that the Group Exponent Laws hold for any group G.

Exercise 2.8.18. Let G be a group. Prove that $Z(G) = \{x \in G \mid gx = xg \text{ for all elements } g \in G\}$ is a subgroup of G called the **center** of G (the notation is derived from the German "das Zentrum").

Exercise 2.8.19. Let G be a group. Prove that the **centralizer** $Z_G(x) = \{g \in G \mid gx = xg\}$ of any element $x \in G$ is a subgroup of G.

Exercise 2.8.20. Let G be a group with a subgroup H. Prove that for any element $g \in G$, the set $gHg^{-1} = \{ghg^{-1} \mid h \in H\} \subseteq G$ is a subgroup of G; this is called the **conjugate** of H by g.

Exercise 2.8.21. Let G be a group. Prove that the **normalizer** $N_G(H) = \{g \in G \mid gHg^{-1} = H\}$ of any subgroup $H \subseteq G$ is itself a subgroup of G.

(**Hint:** Observe that if $gHg^{-1} = H$, then for every element $h \in H$, there exists an element $k \in H$ such that $h = gkg^{-1}$. Conclude that $g^{-1}hg$ lies in H for every element $g \in N_G(H)$.)

Exercise 2.8.22. Let G be a group. Prove that for any subgroups H and K of G, the intersection $H \cap K = \{g \in G \mid g \in H \text{ and } g \in K\}$ is a subgroup of G.

Exercise 2.8.23. Let G be a group. Prove that for any subgroups H and K of G, the collection of products $HK = \{hk \mid h \in H \text{ and } k \in K\}$ is a subgroup of G if and only if $HK = KH$.

2.8.3 Cyclic Groups

Exercise 2.8.24. Explain if the rational numbers $\mathbb{Q}$ form a cyclic group with respect to addition.

Exercise 2.8.25. Let $\mathbb{Z}_p^* = \{a \in \mathbb{Z}_p \mid \gcd(a, p) = 1\}$ denote the multiplicative group of integers modulo a prime number p. Verify that $\mathbb{Z}_p^*$ is a cyclic group of order $p - 1$ for $p = 2, 3, 5, 7$, and 11.

Exercise 2.8.26. Let $\mathbb{Z}_n^*$ denote the multiplicative group of integers modulo n. Prove that there exists a non-identity element $a \in \mathbb{Z}_n^*$ of order two, i.e., such that $a^2 \equiv 1 \pmod{n}$.

(**Hint:** Observe that for any positive integer n, we have that $n^2 - 2n + 1 \equiv 1 \pmod{n}$.)

Exercise 2.8.27. Let $(\mathbb{Z}_p, +)$ denote the additive group of integers modulo a prime number p. Prove that $(\mathbb{Z}_p, +)$ admits no subgroups other than itself and the trivial subgroup.

Euler's totient function is the unique piecewise function $\phi : \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 1}$ defined by $\phi(1) = 1$ and $\phi(n) = \#\{k \mid 1 \leq k \leq n - 1 \text{ and } \gcd(k, n) = 1\}$ for all integers $n \geq 2$. Explicitly, we note that $\phi(n)$ is precisely the number of positive integers not exceeding n that are relatively prime to n.

Exercise 2.8.28 (Euler's Theorem)**.** Prove that $|\mathbb{Z}_n^*| = \phi(n)$ for every positive integer n. Use this to deduce **Euler's Theorem** that $a^{\phi(n)} \equiv 1 \pmod{n}$ for all integers a such that $\gcd(n, a) = 1$.

Exercise 2.8.29 (Fermat's Little Theorem)**.** Prove that $|\mathbb{Z}_p^*| = p - 1$ for every prime integer p. Use this to deduce **Fermat's Little Theorem** that $a^{p-1} \equiv 1 \pmod{n}$ for all integers a such that $p \nmid a$.

Exercise 2.8.30. Let G be a group. Prove that the following statements holds.

(a.) Given any element $g \in G$, we have that $\operatorname{ord}(g^{-1}) = \operatorname{ord}(g)$.

(b.) Given any element $x \in G$, we have that $\operatorname{ord}(gxg^{-1}) = \operatorname{ord}(x)$ for all elements $g \in G$.

(**Hint:** Observe that $(gxg^{-1})^r = (gxg^{-1})(gxg^{-1}) \cdots (gxg^{-1})$ with r factors of gxg^{-1}; then, simplify the blue terms, and collect the orange terms using the Group Exponent Laws.)

(c.) Given any elements $g, h \in G$, we have that $\operatorname{ord}(gh) = \operatorname{ord}(hg)$.

(**Hint:** Observe that $hg = hghh^{-1} = h(gh)h^{-1}$; then, use part (b.) to conclude the result.)

Exercise 2.8.31. Let G be an abelian group. Consider any elements $g, h \in G$ such that $\operatorname{ord}(g) = r$, $\operatorname{ord}(h) = s$, and $\operatorname{ord}(gh) = t$ are all finite.

(a.) Prove that $t \mid rs$.

(b.) Prove that $r \mid st$ and that $s \mid rt$. Conclude that $\dfrac{r}{\gcd(r,s)} \,\Big|\, \dfrac{s}{\gcd(r,s)} t$ and $\dfrac{s}{\gcd(r,s)} \,\Big|\, \dfrac{r}{\gcd(r,s)} t$.

(c.) Prove that $\dfrac{r}{\gcd(r,s)} \,\Big|\, t$ and $\dfrac{s}{\gcd(r,s)} \,\Big|\, t$. Conclude that $\dfrac{rs}{\gcd(r,s)^2} \,\Big|\, t$.

Ultimately, conclude that if $\operatorname{ord}(g)$ and $\operatorname{ord}(h)$ are relatively prime, then $\operatorname{ord}(gh) = \operatorname{ord}(g)\operatorname{ord}(h)$.

(**Hint:** Corollary 2.3.13 yields (a.). On the first part of (b.), to quote the great Lucian Grand, you will need to make the inspired substitution $h^{st} = e_G$ to find that $g^{st} = g^{st}h^{st}$; simplify and use the corollary. Use Exercise 1.10.28 for the first part of (c.) and Exercise 1.10.34 for the second part.)

Exercise 2.8.32. Let G be an abelian group. Let p be a prime number. Prove that for any elements $g, h \in G$ with $\operatorname{ord}(g) = p^m$ and $\operatorname{ord}(h) = p^n$, we have that $\operatorname{ord}(gh) = \max\{m, n\}$.

Exercise 2.8.33. Let G be a cyclic group of order n. Prove that $\operatorname{ord}(x) \mid n$ for all elements $x \in G$.

Exercise 2.8.34. Let G be a cyclic group of order n. Prove that for all positive integers $d \mid n$, there exists a (cyclic) subgroup of G of order d.

Exercise 2.8.35. Let G be an abelian group. Prove that the set $G_T = \{g \in G \mid \operatorname{ord}(g) \text{ is finite}\}$ of elements of G of finite order is a subgroup of G called the **torsion subgroup**.

Exercise 2.8.36. Prove that if a group G is not cyclic, then it admits (at least) one proper non-trivial subgroup. Conclude that if G has no proper non-trivial subgroups, then G is cyclic.

2.8.4 Complex Numbers as a Group Under Multiplication

Exercise 2.8.37. Prove that if $z \in \mathbb{C}^\times$ has finite order, then we must have that $|z| = 1$. Conclude that every nonzero complex number such that $|z| \neq 1$ has infinite order.

Exercise 2.8.38. Find all elements of finite order in the multiplicative group of complex numbers.

Exercise 2.8.39. Graph the fifth roots of unity. List each of them as a complex number of the form $\operatorname{cis}(\theta)$ for some angle $0 \leq \theta < 2\pi$ and in the form $a + bi$ for some nonzero real numbers a and b; then, indicate which of the fifth roots of unity are primitive fifth roots of unity.

Exercise 2.8.40. Graph the sixth roots of unity. List each of them as a complex number of the form $\operatorname{cis}(\theta)$ for some angle $0 \leq \theta < 2\pi$ and in the form $a + bi$ for some nonzero real numbers a and b; then, indicate which of the sixth roots of unity are primitive sixth roots of unity. Compare the results of this exercise with your results from Exercise 2.8.39, and explain the differences.

Exercise 2.8.41. Generally, what shape do the nth roots of unity form in the complex plane? Use this information to deduce when the polynomial $x^n - 1$ has two real roots or only one real root.

Exercise 2.8.42. Prove that $(r_1 \operatorname{cis}\theta_1)(r_2 \operatorname{cis}\theta_2) = r_1 r_2 \operatorname{cis}(\theta_1 + \theta_2)$.

(**Hint:** Use $\cos(\theta_1 + \theta_2) = \cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2$ and $\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1$.)

2.8.5 The Symmetric Group on n Letters

Exercise 2.8.43. Let n be a positive integer. Let $[n] = \{1, 2, \dots, n\}$. Complete the following steps to obtain an alternate proof of Proposition 2.5.5. Use the first step provided as a guide the rest.

(i.) Use words and symbols to define when a bijection $\sigma : [n] \to [n]$ is a **cycle**.

We have that σ is a cycle if and only if there exists a nonempty set $S \subseteq [n]$ for which the restriction $\sigma|_S : S \to S$ of σ to S is a bijection and $\sigma(i) = i$ for all integers $i \in [n] \setminus S$.

(ii.) Use words and symbols to define the entries of the **one-line notation** of σ.

(iii.) Use words and symbols to define when two cycles σ and τ are **disjoint**.

(iv.) Prove that if i does not appear in either the one-line notation of σ or τ, then $\sigma\tau(i) = \tau\sigma(i)$.

(v.) Prove that if i appears in the one-line notation of σ, then it does not appear in the one-line notation of τ. Conclude in this case that $\sigma\tau(i) = \tau\sigma(i)$.

Exercise 2.8.44. Prove that every k-cycle can be written as a product of transpositions. Conclude by Proposition 2.5.4 that every permutation can be written as a product of transpositions.

(**Hint:** Consider the k-cycle $(a_1, \dots, a_k)$. Use the fact that permutations are multiplied right to left, hence if a_i does not appear in the one-line notation of σ, then $\sigma \circ (a_1, a_i)$ sends a_1 to a_i.)

Like integers, permutations possess **parity**. Explicitly, we say that a permutation σ is **even** (or **odd**, respectively) if it can be expressed as a product of an even (or odd, respectively) number of transpositions. We will assume that the identity permutation ι is even (cf. [JB21, Lemma 5.14]).

Exercise 2.8.45. Prove that a permutation σ is either even or odd but not both.

(**Hint:** Observe that if $\sigma = \tau_1 \cdots \tau_m$ for some transpositions $\tau_1, \dots, \tau_m$ and $\sigma = \theta_1 \cdots \theta_n$ for some transpositions $\theta_1, \dots, \theta_n$, then $\iota = \sigma\sigma^{-1} = \tau_1 \cdots \tau_m \theta_n^{-1} \cdots \theta_1^{-1}$.)

Exercise 2.8.46. Prove that a cycle of odd length is even and a cycle of even length is odd.

Exercise 2.8.47. Consider the collection $\mathfrak{A}_n$ of even permutations on n letters.

(a.) Prove that $\mathfrak{A}_n$ is a subgroup of $\mathfrak{S}_n$ called the **alternating group on n letters**.

(b.) Compute the order of the alternating group $\mathfrak{A}_4$ on four letters.

(**Hint:** Every cycle of odd length is even; all else in $\mathfrak{A}_4$ is a product of disjoint transpositions.)

(c.) Use part (b.) above and Lagrange's Theorem to compute the index $[\mathfrak{S}_4 : \mathfrak{A}_4]$ of $\mathfrak{A}_4$ in $\mathfrak{S}_4$.

2.8.6 Dihedral Groups

Exercise 2.8.48. Verify the explanation of Example 2.6.3 by using pictures to illustrate how each of the eight elements $1, r, r^2, r^3, s, rs, r^2s, r^3s$ of D_4 acts on the square.

Exercise 2.8.49. Verify the explanation of Example 2.6.4 by using pictures to illustrate how each of the ten elements $1, r, r^2, r^3, r^4, s, rs, r^2s, r^3s, r^4s$ of D_5 acts on the regular pentagon.

Exercise 2.8.50. Prove that the dihedral group D_n of order $2n$ is not abelian for $n \geq 3$.

(**Hint:** On the contrary, if $rs = sr$, then what can be said about r by Proposition 2.6.1?)

Exercise 2.8.51. Prove that the dihedral group D_n of order $2n$ admits elements x and y of order two such that their product xy has order n. Conclude that the order of a product of two elements of order two can be any positive integer exceeding two.

Exercise 2.8.52. Consider the center $Z(D_n) = \{x \in D_n \mid yx = xy \text{ for all } y \in D_n\}$ of the dihedral group D_n of order $2n$. Complete the following steps to prove that

$$Z(D_n) = \begin{cases} \{1\} & \text{if } n \text{ is odd and} \\ \{1, r^{\frac{n}{2}}\} & \text{if } n \text{ is even.} \end{cases}$$

(i.) Every element of D_n is of the form $r^i s^j$ for some integers $0 \leq i \leq n-1$ and $0 \leq j \leq 1$. By definition, for any element $x \in Z(D_n)$, we must have that $xr = rx$. Conclude that if $x = r^k s$ for some integer $0 \leq k \leq n-1$, then $xr \neq rx$, i.e., x is not in $Z(D_n)$.

(ii.) Use the step (i.) to prove that if $x \in Z(D_n)$, then $x = r^k$ for some integer $0 \leq k \leq n-1$.

(iii.) On the other hand, for any element $x \in Z(D_n)$, we must have that $xs = sx$. By the previous part, we may assume that $x = r^k$ for some integer $0 \leq k \leq n-1$, hence we must have that $r^k s = sr^k$. Use the identity $sr = r^{n-1}s$ to find that if $x \in Z(D_n)$, then $r^k = r^{nk-k}$.

(iv.) Cancelling a factor from both sides of the last identity of part (iii.), we find that $r^{nk-2k} = 1$. By Corollary 2.3.13, conclude that $n \mid (nk-2k)$.

(v.) Observe that if $n \mid (nk-2k)$, then there exists an integer q such that $nk - 2k = nq$. Conclude that $n \mid 2k$, hence we must have that $n = 0$ or $n = 2k$. Ultimately, conclude the desired result.

Exercise 2.8.53. Prove that if $n \geq 4$, then there exists an element $\sigma \in \mathfrak{S}_n$ such that $\sigma \notin D_n$. Conclude that the dihedral group of order $2n$ is a proper subgroup of $\mathfrak{S}_n$ for all integers $n \geq 4$.

(**Hint:** Every element of D_n must do what to the consecutive clockwise vertices n, 1, and 2?)

Chapter 3

Group Theory II

One of the primary objectives of group theory is to solve the classification problem for groups. Explicitly, it is natural to ask the ways in which two groups differ from one another. We define this distinction formally by asking whether two groups are isomorphic. Groups that are isomorphic are "essentially the same" in the sense that the elements of one group can be "renamed" to obtain the elements of the other group, and the binary operations on each group are "compatible" with one another. Ultimately, we will find that isomorphic groups possess the same properties. We devote this chapter to developing the necessary tools and solving the classification problem in some cases.

3.1 Cosets and Lagrange's Theorem

Central to the study of groups is the question of finding all proper non-trivial subgroups of any group. We have already seen in Example 2.2.12 that the abelian groups $(\mathbb{Z}_4, +)$ and $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$ both have order four, but they are distinct from one another as groups because $(\mathbb{Z}_4, +)$ admits only one non-trivial proper subgroup compared to the three non-trivial proper subgroups of $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$. Our aim throughout this section is to prove Lagrange's Theorem; it is a powerful tool in group theory that drastically narrows down the possible subgroups of any group of finite order.

We will assume throughout this section that $(G, *)$ is a group. If $(H, *)$ is a subgroup of G, then the **left coset** of H in G **represented** by an element $g \in G$ is the collection $g * H = \{g * h \mid h \in H\}$ of all products of all elements of H with g on the left. We define right cosets analogously.

Example 3.1.1. Consider the dihedral group $D_3 = \{1, r, r^2, s, rs, r^2s\}$ and its subgroup $H = \{1, s\}$. We obtain the left cosets $1H = H = sH$, $rH = \{r, rs\} = rsH$, and $r^2H = \{r^2, r^2s\} = r^2sH$. We obtain the right cosets $H1 = H = Hs$, $Hr = \{r, r^2s\} = Hr^2s$, and $Hr^2 = \{r^2, rs\} = Hrs$ by using the identity $sr = r^2s$ of Proposition 2.6.1. Observe that $rH \neq Hr$ and $r^2H \neq Hr^2$, hence it is not necessarily true that the left and right cosets with respect to the same representative are equal.

Conversely, it holds that the left and right cosets of the subgroup $K = \{1, r, r^2\}$ in D_3 coincide for each representative. Explicitly, the left cosets $1K = K = rK = r^2K$ coincide with the right cosets $K1 = K = Kr = Kr^2$ and $sK = \{s, rs, r^2s\} = rsK = r^2sK$ and $Ks = \{s, rs, r^2s\} = Krs = Kr^2s$. We will return to this example and discuss this phenomenon in greater detail in Section 3.2.

We note that if G is abelian, then it holds that $g * h = h * g$ for all elements $g \in G$ and $h \in H$, hence the left and right cosets of H in G are equal, and we may refer to them simply as cosets.

Example 3.1.2. Consider the group $(\mathbb{Z}, +)$ and its subgroup $2\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$. Observe that $0 + 2\mathbb{Z} = \{0 + 2n \mid n \in \mathbb{Z}\} = \{2n + 0 \mid n \in \mathbb{Z}\} = 2\mathbb{Z} + 0$ consists of all even integers and $1 + 2\mathbb{Z} = \{1 + 2n \mid n \in \mathbb{Z}\} = \{2n + 1 \mid n \in \mathbb{Z}\} = 2\mathbb{Z} + 1$ consists of all odd integers. Consequently, there are only two left cosets of $2\mathbb{Z}$ in $\mathbb{Z}$, and the left and right cosets of $2\mathbb{Z}$ in $\mathbb{Z}$ coincide.

We provide the following propositions to summarize and generalize our current observations.

Proposition 3.1.3. *Let G be a group with a subgroup H. Given any element $g \in G$, the cosets gH and Hg of H in G represented by g satisfies that $gH = ghH$ and $Hg = Hhg$ for all elements $h \in H$. Put another way, the left and right coset representatives of H in G are not unique.*

Proof. Observe that for every element $h \in H$, we have that $gh = ghe_G$ lies in ghH because H is a subgroup of G, hence we conclude that $gH \subseteq ghH$. Conversely, for any element $h' \in H$, we have that hh' is an element of H so that ghh' lies in gH for all elements $h' \in H$, i.e., $ghH \subseteq gH$. □

Proposition 3.1.4. *Let G be a group. Consider any pair of elements $g_1, g_2 \in G$. If H is a subgroup of G, then the following properties of the left cosets of H in G are equivalent.*

(i.) *We have that $g_1H = g_2H$.*

(ii.) *We have that $g_1H \supseteq g_2H$.*

(iii.) *We have that $g_2 \in g_1H$.*

(iv.) *We have that $g_1^{-1}g_2 \in H$.*

(v.) *We have that $Hg_1^{-1} = Hg_2^{-1}$.*

Even more, the above properties hold for the right cosets of H in G, as well.

Proof. We leave it to the reader as Exercise 3.18.1 to prove directly that the first three implications hold. We will assume that $g_1^{-1}g_2 \in H$. Consequently, for every element $h \in H$, we have that $hg_1^{-1}g_2$ lies in H by assumption that H is a subgroup of G. Put another way, we have that $Hg_1^{-1}g_2 \subseteq H$, from which it follows that $Hg_1^{-1} \subseteq Hg_2^{-1}$. Conversely, we have that $g_2^{-1}g_1 = (g_1^{-1}g_2)^{-1}$, and we conclude as before that $Hg_2^{-1} \subseteq Hg_1^{-1}$, hence the fifth property above holds. Last, if $Hg_1^{-1} = Hg_2^{-1}$ holds, then we claim that $g_1H = g_2H$. By hypothesis, for every element $h_1 \in H$, there exists an element $h_2 \in H$ such that $h_1g_1^{-1} = h_2g_2^{-1}$. By taking the inverses of both sides, we find that $g_1h_1^{-1} = g_2h_2^{-1}$. Consequently, it follows that $g_1h_1 = g_2h_2^{-1}h_1^2$ lies in g_2H by assumption that H is a subgroup. We conclude that $g_1H \subseteq g_2H$; the other inclusion is proved analogously, hence equality holds. □

Given a group G with a subgroup H, we refer to the number $[G : H]$ of distinct left cosets of H in G as the **index** of H in G. We note that it is possible that $[G : H]$ is infinite. Explicitly, the rational numbers $\mathbb{Q}$ form a subgroup of the additive group $(\mathbb{R}, +)$ of real numbers such that $[\mathbb{R} : \mathbb{Q}]$ is infinite (cf. Exercise 3.18.2). Often, we will restrict our attention to the case that there are finitely many left cosets of H in G, hence $[G : H]$ will typically be a positive integer.

Example 3.1.5. Consider the dihedral group $D_3 = \{1, r, r^2, s, rs, r^2s\}$ and its subgroups $H = \{1, s\}$ and $K = \{1, r, r^2\}$ of Example 3.1.1. We established previously that $[G : H] = 3$ and $[G : K] = 2$.

Example 3.1.6. We established in Example 3.1.2 that $[\mathbb{Z} : 2\mathbb{Z}] = 2$ as groups under addition.

Our next proposition illustrates that we do not need to define an analogous term to measure the number of right cosets of H in G; in fact, this is exactly equal to the index of H in G.

Proposition 3.1.7. *Let G be a group. Given any subgroup H of G, the number of right cosets of H in G is equal to the left cosets of H in G, i.e., the index $[G : H]$ of H in G.*

Proof. Once again, by Exercise 1.10.5(d.), it suffices to provide a bijection $f_g : gH \to Hg$ for each element $g \in G$. We claim that such a function is given by the rule $f_g(gH) = Hg^{-1}$. We must first establish that this definition results in a **well-defined** function, i.e., we must demonstrate that if $g_1 H = g_2 H$, then $Hg_1^{-1} = Hg_2^{-1}$. (Essentially, this is the converse of the definition of injective.) But this holds by Proposition 3.1.4. Even more, the same proposition illustrates that if $Hg_1^{-1} = Hg_2^{-1}$, then $g_1 H = g_2 H$, i.e., it holds that f_g is injective. Last, f_g is surjective by construction. □

Before we prove Lagrange's Theorem, we provide two more crucial observations about left cosets.

Lemma 3.1.8. *Let G be a group with a subgroup H. Every left coset of H in G has the cardinality as H. Put another way, for every element $g \in G$, we have that $|gH| = |H|$.*

Proof. By Exercise 1.10.5(d.), it suffices to provide a bijection $f_g : H \to gH$ for each element $g \in G$. We may define one by declaring that $f_g(h) = gh$ for every element $h \in H$. By definition, every element of gH can be written as gh for some element $h \in H$, hence f_g is surjective. Cancellation holds in G, hence $gh_1 = f_g(h_1) = f_g(h_2) = gh_2$ implies that $h_1 = h_2$, i.e., f_g is injective. □

Lemma 3.1.9. *Let G be a group with a subgroup H. We have that $g_1 H \sim g_2 H$ if and only if $g_1 g_2^{-1} \in H$ is an equivalence relation on the left cosets of H in G. Consequently, the left cosets of H in G partition G, i.e., G is the disjoint union of the distinct left cosets of H in G.*

Proof. We must demonstrate that the relation on the left cosets of H in G defined by $g_1 H \sim g_2 H$ if and only if $g_1 g_2^{-1} \in H$ is reflexive, symmetric, and transitive.

1.) By assumption that H is a subgroup of G, we have that $e_G = g_1 g_1^{-1}$ lies in H for all left cosets $g_1 H$ of H in G, hence we have that $g_1 H \sim g_1 H$, and the relation is reflexive.

2.) If $g_1 H \sim g_2 H$, then $g_1 g_2^{-1}$ lies in H. Once again, by hypothesis that H is a subgroup of G, it follows that $g_2 g_1^{-1} = (g_1 g_2^{-1})^{-1} \in H$ so that $g_2 H \sim g_1 H$, i.e., the relation is symmetric.

3.) If $g_1 H \sim g_2 H$ and $g_2 H \sim g_3 H$, then both $g_1 g_2^{-1}$ and $g_2 g_3^{-1}$ lie in H. Consequently, their product $g_1 g_3^{-1} = (g_1 g_2^{-1})(g_2 g_3^{-1})$ lies in H so that $g_1 H \sim g_3 H$, and the relation is transitive.

By Proposition 3.1.4, the inclusion $g_1 g_2^{-1} \in H$ is equivalent to equality of the left cosets $g_1 H = g_2 H$, hence left coset equality is an equivalence relation on the left cosets of H in G. By Corollary 1.4.5, we conclude that the left cosets of H in G partition G: the members of the partition are the disjoint equivalence classes of G modulo this relation, i.e., they are the disjoint left cosets of H in G. □

Theorem 3.1.10 (Lagrange's Theorem)**.** *Given a group G and any subgroup H of G, we have that $|G| = [G : H]|H|$. Put another way, the order of any subgroup H of G must divide the order of G.*

Proof. By Lemma 3.1.9, there exists a bijection between G and the union of $[G : H]$ many disjoint left cosets of H in G. Each of these left cosets of H in G has $|H|$ elements by Lemma 3.1.8. □

Remark 3.1.11. We note that if $[G : H] = n$ is finite, then we can be more explicit about the details of the proof of Lagrange's Theorem. By Lemma 3.1.9, there exist elements $g_1, \dots, g_n \in G$ such that $g_1 H, \dots, g_n H$ are pairwise disjoint and $G = g_1 H \cup \dots \cup g_n H$. Consequently, we have that $|G| = \sum_{i=1}^n |g_i H|$. Lemma 3.1.8 yields that $|g_i H| = |H|$, and there are $[G : H]$ summands.

Corollary 3.1.12. *If G is a finite group with subgroups $H \supseteq K$, then $[G : K] = [G : H][H : K]$.*

Proof. By assumption that K is a subgroup of G and $H \supseteq K$, it follows by the Subgroup Test that K is a subgroup of H. Consequently, if $|G|$ is a positive integer, then $|H|$ and $|K|$ are positive integers, and Lagrange's Theorem yields that $[G : K] = |G|/|K| = (|G|/|H|)(|H|/|K|) = [G : H][H : K]$. □

Like we mentioned at the beginning of this section, Lagrange's Theorem provides a tool with which we may determine the possible subgroups of a group based on the order of the group.

Corollary 3.1.13. *Every group of prime order is cyclic.*

Proof. By Lagrange's Theorem, the order of any non-identity element of a group G of prime order is prime. Consequently, there exists an element $g \in G$ such that $\operatorname{ord}(g) = |G|$, i.e., $G = \langle g \rangle$. □

Corollary 3.1.14. *Every group of prime order is abelian.*

Proof. By Corollary 3.1.13, such a group is cyclic and hence abelian by Proposition 2.3.8. □

Corollary 3.1.15. *If G is a finite group, then $\operatorname{ord}(g)$ divides $|G|$ for every element $g \in G$. Put another way, the order of any element of G divides the order of G.*

Proof. Observe that the order of an element $g \in G$ is exactly the cardinality of the cyclic subgroup $\langle g \rangle$ generated by G. By Lagrange's Theorem, we conclude that $\operatorname{ord}(g)$ divides $|G|$. □

Corollary 3.1.16. *If G is a finite group, then $g^{|G|} = e_G$ for every element $g \in G$.*

Proof. By Corollary 3.1.15, there exists a positive integer q such that $|G| = \operatorname{ord}(g)q$. Consequently, by the Group Exponent Laws, it follows that $g^{|G|} = g^{\operatorname{ord}(g)q} = (g^{\operatorname{ord}(g)})^q = (e_G)^q = e_G$. □

Caution: Lagrange's Theorem states that the order of every subgroup of a finite group divides the order of the group; however, the converse to Lagrange's Theorem is false. Explicitly, there exists a group G and an integer d dividing $|G|$ such that G does not admit a subgroup of order d.

Proposition 3.1.17 (The Converse of Lagrange's Theorem Is False)**.** *The alternating group $\mathfrak{A}_4$ on four letters is a subgroup of the symmetric group $\mathfrak{S}_4$ on four letters of order $12 = 2^2 \cdot 3$. Even more, there is not a subgroup of $\mathfrak{A}_4$ of order $6 = 2 \cdot 3$, hence the converse of Lagrange's Theorem is false.*

Proof. We simplify the clever proof of [Hen19, Example 2.18]. By Exercise 2.8.47, the first sentence of the proposition statement holds. On the contrary, we will assume that there exists a subgroup H of $\mathfrak{A}_4$ of order six. By Lagrange's Theorem, we have that $12 = |\mathfrak{A}_4| = [\mathfrak{A}_4 : H]|H| = 6[\mathfrak{A}_4 : H]$, from which it follows that $[\mathfrak{A}_4 : H] = 2$. Consequently, the only cosets of H in $\mathfrak{A}_4$ are H itself and $\mathfrak{A}_4 \setminus H$ by Remark 3.1.11. By Proposition 3.1.4, we conclude that for every element $\sigma \in \mathfrak{A}_4$, we have that $\sigma^2 H = H$, i.e., $\sigma^2 \in H$: indeed, we must have that either $\sigma^2 H = H$ or $\sigma H = H$, and the latter implies the former. We claim moreover that if σ is a 3-cycle, then σ belongs to H. Given any 3-cycle σ, observe that $\sigma = \sigma^4 = (\sigma^2)^2$ lies in H because σ^2 lies in H. We note that there are $4!/3 = 8$ 3-cycles in $\mathfrak{A}_4$, hence the order of H is at least eight — a contradiction. □

3.2 Quotient Groups and Normal Subgroups

Let G be a group. Given any subgroup H of G, we denote by G/H the collection of left cosets of H in G, i.e., we have that $G/H = \{gH \mid g \in G\}$ and $gH = \{gh \mid h \in H\}$ for all elements $g \in G$.

Proposition 3.2.1. *If G is a group and H is a subgroup of G, then the following are equivalent.*

(i.) *G/H is a group with respect to the operation $(g_1H)(g_2H) = g_1g_2H$.*

(ii.) *We have that $gH = Hg$ for all elements $g \in G$.*

(iii.) *We have that $gH \subseteq Hg$ for all elements $g \in G$.*

(iv.) *We have that $ghg^{-1} \in H$ for all elements $g \in G$ and $h \in H$.*

Proof. We will assume first that G/H is a group with respect to the operation $(g_1H)(g_2H) = g_1g_2H$. Explicitly, the product of two left cosets g_1H and g_2H results in a left coset of H. By definition, for all elements $g_1, g_2 \in G$ and all elements $h_1, h_2 \in H$, we must have that $g_1h_1g_2h_2$ is an element of H. We claim that $gH = Hg$ for all elements $g \in G$. Given any element $h \in H$, by assumption, there exists an element $k \in H$ such that $ghg^{-1}e_G = k$. Consequently, we find that $gh = kg$ so that $gH \subseteq Hg$. Conversely, for every element $h \in H$, there exists an element $k \in H$ such that $g^{-1}hge_G = k$. We conclude therefore that $Hg \subseteq gH$, hence their equality holds.

Certainly, if $gH = Hg$ for all elements $g \in G$, then $gH \subseteq Hg$ for all elements $g \in G$. Even more, if $gH \subseteq Hg$ for all elements $g \in G$, then for every element $h \in H$, there exists an element $h' \in H$ such that $gh = h'g$, hence we have that $ghg^{-1} = h'$ lies in H for all elements $g \in G$ and $h \in H$.

Last, if ghg^{-1} lies in H for all elements $g \in G$ and $h \in H$, then we will demonstrate that G/H is a group with respect to the operation $(g_1H)(g_2H) = g_1g_2H$. Crucially, this operation is clearly associative; the identity element of G/H is the left coset e_GH; and the inverse of a left coset gH is the left coset $g^{-1}H$; however, we have not demonstrated that this is a binary operation on G/H. Explicitly, we must ensure that for any pair of coset representatives $g_1H = g_3H$ and $g_2H = g_4H$, we have that $g_1g_2H = g_3g_4H$. By Proposition 3.1.4, it suffices to prove that $(g_3g_4)^{-1}g_1g_2 \in H$. Considering that $g_1H = g_3H$, it follows that $g_3^{-1}g_1$ lies in H, hence we have that $(g_3g_4)^{-1}g_1g_2 = g_4^{-1}g_3^{-1}g_1g_2$ is of the form $g_4^{-1}hg_2$ for some element $h \in H$. Likewise, we have that $g_4^{-1}g_2$ lies in H by assumption that $g_2H = g_4H$. Our original hypothesis that ghg^{-1} lies in H for all elements $g \in G$ and $h \in H$ yields that $(g_3g_4)^{-1}g_1g_2 = g_4^{-1}hg_2 = (g_4^{-1}hg_4)(g_4^{-1}g_2)$ lies in H. □

We say that H is a **normal** subgroup of G if any of the above conditions of Proposition 3.2.1 holds for H; we denote this situation by $H \trianglelefteq G$. Often, if H is a subgroup of G, then we it is most convenient to write $H \leq G$ in place of the relatively cumbersome "H is a subgroup of G," hence the notation for normal subgroups is a specialization of this notation for subgroups. Even more, we will say that G/H is the **quotient group**, and we will refer to G/H as "G modulo H."

Corollary 3.2.2. *If G is a group and H is a subgroup of G, then G/H is a group of order $[G : H]$ with respect to the operation $(g_1H)(g_2H) = g_1g_2H$ if and only if H is a normal subgroup of G.*

Example 3.2.3. Consider the dihedral group $D_3 = \{1, r, r^2, s, rs, r^2s\}$ of order six and its cyclic subgroup $K = \{1, r, r^2\}$. By Example 3.1.1, we have that $xK = Kx$ for every element $x \in D_3$.

Consequently, it follows by Proposition 3.2.1 that K is a normal subgroup of D_3, i.e., $K \trianglelefteq D_3$. Even more, there are two distinct cosets of K in D_3 — namely, they are K and sK — hence the quotient group D_3/K has two distinct elements K and sK satisfying that $(sK)(sK) = s^2K = K$.

Proposition 3.2.4. *Every subgroup of an abelian group is normal.*

Proof. Let H be any subgroup of an abelian group G. Observe that for every element $g \in G$ and every element $h \in H$, we have that $gh = hg$, i.e., it holds that ghg^{-1} lies in H. □

Example 3.2.5. Consider the abelian group $(\mathbb{Z}, +)$ of the integers under addition and its normal subgroup $2\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$. By Proposition 3.2.4, $2\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$, hence we may form the quotient group $\mathbb{Z}/2\mathbb{Z}$. By Example 3.1.2, $(\mathbb{Z}/2\mathbb{Z}, +)$ consists of the two distinct cosets $0+2\mathbb{Z}$ and $1+2\mathbb{Z}$ satisfying that $(0+2\mathbb{Z})+(1+2\mathbb{Z}) = 1+2\mathbb{Z}$ and $(1+2\mathbb{Z})+(1+2\mathbb{Z}) = 2+2\mathbb{Z} = 0+2\mathbb{Z}$.

Generally, for any positive integer n, we may consider the subgroup $n\mathbb{Z} = \{qn \mid q \in \mathbb{Z}\}$ of $\mathbb{Z}$. Considering that $\mathbb{Z}$ is abelian, Proposition 3.2.4 yields that $n\mathbb{Z}$ is normal, hence Corollary 3.2.2 implies that $\mathbb{Z}/n\mathbb{Z}$ is a group with respect to the operation $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$. Consequently, for any integer $k \in \mathbb{Z}$, we have that $k(1+n\mathbb{Z}) = k+n\mathbb{Z}$, hence $\mathbb{Z}/n\mathbb{Z}$ is a cyclic group of order n: the coset $1 + n\mathbb{Z}$ generates $\mathbb{Z}/n\mathbb{Z}$, and the cosets $0 + n\mathbb{Z}, 1 + n\mathbb{Z}, \dots, (n-1) + n\mathbb{Z}$ are distinct. We will soon establish that $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}_n$ are "indistinguishable" groups under addition.

We demonstrate next that the quotient groups inherits some properties of the original group.

Proposition 3.2.6. *Let G be a group. Let H be a normal subgroup of G.*

1.) *If G is cyclic, then G/H is cyclic. Explicitly, if $G = \langle g \rangle$, then $G/H = \langle gH \rangle$.*

2.) *If G is abelian, then G/H is abelian.*

Proof. (1.) By definition, if G is cyclic, then there exists an element $g \in G$ such that every element of G can be written as g^n for some integer n. Consequently, for any coset xH of G/H, there exists an integer n such that $x = g^n$ and $xH = g^nH = (gH)^n$. We conclude that G/H is cyclic.

(2.) By definition, if G is abelian, then $g_1g_2 = g_2g_1$ for all elements $g_1, g_2 \in G$. Consequently, for all cosets g_1H, g_2H of H in G, it follows that $(g_1H)(g_2H) = g_1g_2H = g_2g_1H = (g_2H)(g_1H)$. □

3.3 Group Homomorphisms

Given a pair of groups $(G, *)$ and $(H, \star)$, we say that a function $\varphi : G \to H$ is a **group homomorphism** if and only if $\varphi(g_1 * g_2) = \varphi(g_1) \star \varphi(g_2)$ for all elements $g_1, g_2 \in G$. Put another way, a group homomorphism is a function between groups for which the binary operations of the two groups are compatible in the sense that the image of a product of two elements in the domain is the product of the images of the elements in the codomain. Let us try a few examples before we discuss further.

Example 3.3.1. Consider the group $(\mathbb{Z}, +)$ of integers under addition. Given any integer n, we may define a function $\varphi_n : \mathbb{Z} \to \mathbb{Z}$ by $\varphi_n(m) = mn$. Observe that for any pair of integers ℓ and m, we have that $\varphi_n(\ell+m) = n(\ell+m) = \ell n+mn = \varphi_n(\ell)+\varphi_n(m)$, hence φ is a group homomorphism. Even more, because the domain and codomain of φ are equal, we say that φ is an **endomorphism**.

Example 3.3.2. Consider the group $(\mathbb{Z}/n\mathbb{Z}, +)$ of integers modulo a positive integer n and the multiplicative group $(G, \cdot)$ of the nth roots of unity. By definition of the integers modulo n, every element of $\mathbb{Z}/n\mathbb{Z}$ is of the form $k + n\mathbb{Z}$ for some integer $1 \leq k \leq n$. By definition of the nth roots of unity, every element of G is of the form $\operatorname{cis}(2\pi k/n) = \cos(2\pi k/n) + i \sin(2\pi k/n)$ for some integer $1 \leq k \leq n$, where i is the complex number satisfying that $i^2 = -1$. Consequently, it is natural to consider the function $\varphi : \mathbb{Z}/n\mathbb{Z} \to G$ defined by $\varphi(k + n\mathbb{Z}) = \operatorname{cis}(2\pi k/n)$. **Caution:** this function is defined on the left cosets of $n\mathbb{Z}$ in $\mathbb{Z}$, so we must check that this rule is well-defined, i.e., that it does not depend on our choice of coset representative. We assume to this end that we have two different coset representative for the same coset, i.e., suppose that $k+n\mathbb{Z} = \ell+n\mathbb{Z}$. By subtracting $\ell+n\mathbb{Z}$ from both sides, we have that $(k-\ell)+n\mathbb{Z} = 0+n\mathbb{Z}$. Observe that this implies that $k - \ell = mn$ for some integer m so that $k = mn + \ell$ and $2\pi k/n = 2\pi m + 2\pi\ell/n$. Considering that $\cos(2\pi m + \theta) = \cos(\theta)$ and $\sin(2\pi m + \theta) = \sin(\theta)$, we conclude that $\operatorname{cis}(2\pi k/n) = \operatorname{cis}(2\pi m + 2\pi\ell/n) = \operatorname{cis}(2\pi\ell/n)$, hence φ is well-defined. Even more, φ is a group homomorphism because we have that

$$\varphi((k + n\mathbb{Z}) + (\ell + n\mathbb{Z})) = \varphi(k + \ell + n\mathbb{Z}) = \operatorname{cis}(2\pi(k + \ell)/n) = \operatorname{cis}(2\pi k/n) \operatorname{cis}(2\pi\ell/n)$$

for any integers k and ℓ by Proposition 2.4.4. Observe that φ respects the ostensibly different binary operations of each group: it takes the sum of two cosets of $\mathbb{Z}$ to a product of complex numbers.

Example 3.3.3. Given any element g of a group G, we claim that the function $\chi_g : G \to G$ defined by $\chi_g(x) = gxg^{-1}$ is a group homomorphism. Observe that for any pair of elements $x, y \in G$, we have that $\chi_g(xy) = gxyg^{-1} = (gxg^{-1})(gyg^{-1}) = \chi_g(x)\chi_g(y)$. Consequently, we have that χ_g is a group homomorphism; it is an endomorphism that sends $x \in G$ to its **conjugate** gxg^{-1} by g.

Example 3.3.4. Consider an abelian group G. We will demonstrate that the inversion function $\varphi : G \to G$ defined by $\varphi(g) = g^{-1}$ is a group endomorphism. By assumption that G is abelian, for any elements $g, h \in G$, we have that $\varphi(gh) = (gh)^{-1} = h^{-1}g^{-1} = g^{-1}h^{-1} = \varphi(g)\varphi(h)$.

We begin our more general discussion with some basic properties of group homomorphisms.

Proposition 3.3.5. *Consider a group homomorphism* $\varphi : (G, *) \to (H, \star)$.

1.) *We have that* $\varphi(e_G) = e_H$.

2.) *We have that* $\varphi(g^{-1}) = \varphi(g)^{-1}$ *for all elements* $g \in G$.

3.) *Given any element* $g \in G$, *we have that* $\varphi(g^n) = \varphi(g)^n$ *for any integer* n.

4.) *Given any element* $g \in G$, *the order of* g *divides the order of* $\varphi(g)$.

5.) *Given any subgroup* K *of* G, *we have that* $\varphi(K)$ *is a subgroup of* H.

Proof. (1.) Observe that $e_G = e_G e_G$, hence we have that $\varphi(e_G) = \varphi(e_G e_G) = \varphi(e_G)\varphi(e_G)$. Cancelling a factor of $\varphi(e_G)$ from both sides yields that $\varphi(e_G) = e_H$.

(2.) Observe that $gg^{-1} = e_G$, hence part (1.) yields that $e_H = \varphi(e_G) = \varphi(gg^{-1}) = \varphi(g)\varphi(g^{-1})$. By multiplying on the left of each side, we find that $\varphi(g^{-1}) = \varphi(g)^{-1}$.

(3.) Observe that if n is a non-negative integer, then $\varphi(g^n) = \varphi(g) \star \cdots \star \varphi(g)$ with n factors of $\varphi(g)$. By definition, this implies that $\varphi(g^n) = \varphi(g)^n$. Conversely, if n is a negative integer, then $\varphi(g^n) = \varphi(g^{-1} * \cdots * g^{-1}) = \varphi(g^{-1}) \star \cdots \star \varphi(g^{-1})$ with $-n$ factors of $\varphi(g^{-1})$.

(4.) If $\operatorname{ord}(g) = r$, then $e_H = \varphi(e_G) = \varphi(g^r) = \varphi(g)^r$, and the result holds by Corollary 2.3.13.

(5.) Consider a subgroup K of G. We claim that $\varphi(K) = \{\varphi(k) \mid k \in K\}$ is a subgroup of H. Considering that $e_G \in K$, we have that $\varphi(e_G) = e_H$ lies in $\varphi(K)$, hence it is nonempty. We proceed by the One-Step Subgroup Test. Explicitly, for any elements $\varphi(k_1), \varphi(k_2) \in \varphi(K)$, we have that $k_1 * k_2^{-1}$ lies in the subgroup K so that $\varphi(k_1) \star \varphi(k_2)^{-1} = \varphi(k_1) \star \varphi(k_2^{-1}) = \varphi(k_1 * k_2^{-1}) \in \varphi(K)$. □

Because it encodes a lot of important data about the underlying group G, we will take much care to determine the **kernel** $\ker \varphi = \{g \in G \mid \varphi(g) = e_H\}$ of a group homomorphism $\varphi : G \to H$. Our first result along these lines is that the kernel of a group homomorphism detects injectivity.

Proposition 3.3.6. *Given a group homomorphism $\varphi : (G, *) \to (H, \star)$, we have that φ is injective if and only the kernel of φ is the trivial subgroup of G, i.e., $\ker \varphi = \{e_G\}$.*

Proof. We will assume first that φ is injective. Given any element $g \in \ker \varphi$, by the first part of Proposition 3.3.5, we have that $\varphi(g) = e_H = \varphi(e_G)$ so that $g = e_G$ by the injectivity of φ.

Conversely, we will assume that $\ker \varphi$ is trivial. Given any elements $g_1, g_2 \in G$ for which $\varphi(g_1) = \varphi(g_2)$, by the second part of Proposition 3.3.5, we have that $e_H = \varphi(g_1)\varphi(g_2)^{-1} = \varphi(g_1)\varphi(g_2^{-1}) = \varphi(g_1 g_2^{-1})$. By hypothesis that $\ker \varphi$ is trivial, it follows that $g_1 g_2^{-1} = e_G$ so that $g_1 = g_2$. □

Example 3.3.7. Consider the group homomorphism $\varphi_n : \mathbb{Z} \to \mathbb{Z}$ defined by $\varphi_n(m) = mn$ for some nonzero integer n. Observe if m is an integer and $mn = 0$, then we must have that $m = 0$. We conclude that $\ker \varphi_n = \{m \in \mathbb{Z} \mid mn = 0\} = \{0\}$, hence φ_n is injective. We could have also proven this directly: indeed, if $mn = \varphi_n(m) = \varphi_n(\ell) = \ell n$, then cancelling n from both sides gives $m = \ell$.

Example 3.3.8. Let n be a positive integer. Let G denote the multiplicative group of nth roots of unity. Consider the group homomorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \to G$ defined by $\varphi(k + n\mathbb{Z}) = \operatorname{cis}(2\pi k/n)$. We have that $\operatorname{cis}(2\pi k/n) = 1$ if and only if $k = mn$ for some integer m if and only if $k + n\mathbb{Z} = 0 + n\mathbb{Z}$. Consequently, we conclude by Proposition 3.3.6 that φ is injective.

Example 3.3.9. Conjugation by a group element is an injective group endomorphism. Explicitly, we have that $gxg^{-1} = e_G$ if and only if $x = e_G$ for every pair of elements $g, x \in G$.

Example 3.3.10. Inversion is an injective group endomorphism of any abelian group because for any element $g \in G$, we have that $g^{-1} = e_G$ if and only if $g = e_G$.

Example 3.3.11. Consider the function $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\pi(k) = k + n\mathbb{Z}$, where n is a positive integer. Observe that $\pi(k + \ell) = (k + \ell) + n\mathbb{Z} = (k + n\mathbb{Z}) + (\ell + n\mathbb{Z}) = \pi(k) + \pi(\ell)$, hence π is a group homomorphism called the **projection** of $\mathbb{Z}$ onto $\mathbb{Z}/n\mathbb{Z}$. Observe that an integer m lies in $\ker \pi$ if and only if $m + n\mathbb{Z} = 0 + n\mathbb{Z}$ if and only if $m = nr$ for some integer r if and only if m lies in $n\mathbb{Z}$. Consequently, the kernel of π is $n\mathbb{Z}$, hence π is not injective.

We refer to a bijective (i.e., injective and surjective) group homomorphism as a **group isomorphism**. Group isomorphisms can be thought of as a means of relabelling elements in the target group with elements in the domain. Explicitly, if $\varphi : (G, *) \to (H, \star)$ is a group isomorphism, then for every element $h \in H$, there exists an element $g \in G$ such that $h = \varphi(g)$. Put another way, every element of H can be labelled with an element of G. Even more, this labelling is unique because φ is injective, hence if $\varphi(g_1) = \varphi(g_2)$, then $g_1 = g_2$. Otherwise stated, if two elements of H have the same label by an element of G, then the two elements of H are equal. Every element of H may therefore be labelled uniquely with an element of G. Even more, this labelling respects the binary operations of G and H because it is a group homomorphism. We say that $(G, *)$ and $(H, \star)$ are **isomorphic** if there exists a group isomorphism between them, and we write $(G, *) \cong (H, \star)$.

Example 3.3.12. Let n be a positive integer. Let G denote the multiplicative group of nth roots of unity. Consider the group homomorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \to G$ defined by $\varphi(k + n\mathbb{Z}) = \operatorname{cis}(2\pi k/n)$. Considering that $\mathbb{Z}/n\mathbb{Z}$ and G are finite sets of the same cardinality and φ is injective by Example 3.3.8, we conclude by Exercise 1.10.5 that φ is surjective, hence it is an isomorphism.

Example 3.3.13. Conjugation by a group element is an injective group endomorphism; it is also surjective because every element $x \in G$ can be written as $x = g(g^{-1}xg)g^{-1} = \chi_g(g^{-1}xg)$. Consequently, conjugation is an isomorphism from a group to itself; it is a **group automorphism**.

Example 3.3.14. Inversion is an injective group endomorphism of any abelian group; even if the group is not abelian, it is both injective and surjective because every element $g \in G$ satisfies that $g = (g^{-1})^{-1}$. Consequently, inversion is a group automorphism of any abelian group.

Example 3.3.15. Observe that if n is an integer other than ±1, then the injective group homomorphism $\varphi_n : \mathbb{Z} \to \mathbb{Z}$ defined by $\varphi_n(m) = mn$ is not surjective because $mn \neq 1$ for any integer m. Consequently, φ_n is not an isomorphism for any integer other than $n = \pm 1$.

Example 3.3.16. Generally, the projection map $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\pi(k) = k + n\mathbb{Z}$ for a positive integer n is not an isomorphism because it is not injective; it is always surjective.

By the paragraph preceding Example 3.3.12, we intuitively suspect that the function inverse of a group isomorphism is a group isomorphism. We could also reasonably expect that two groups are isomorphic only if they have the same properties, e.g., if two groups are isomorphic and one of the groups is cyclic or abelian, then the other group must also be cyclic or abelian.

Proposition 3.3.17. *Consider a group isomorphism $\varphi : (G, *) \to (H, \star)$.*

1.) *We have that $|G| = |H|$.*

2.) *We have that $\varphi^{-1} : H \to G$ is a group isomorphism.*

3.) *We have that G is abelian if and only if H is abelian.*

4.) *We have that G is cyclic if and only if H is cyclic.*

5.) *Every subgroup of G induces a subgroup of H and vice-versa. Particularly, if G and H are isomorphic, then G and H must have the same number of (proper non-trivial) subgroups.*

Proof. (1.) Exercise 1.10.6(a.) demonstrates that $|G| = |H|$ for any pair of sets G and H for which there exists a bijection $\varphi : G \to H$, hence we may move onto the group-theoretic properties.

(2.) Considering that $\varphi : G \to H$ is a bijection, every element of H can be written uniquely as $\varphi(g)$ for some element $g \in G$, hence the function $\varphi^{-1} : H \to G$ defined by $\varphi^{-1}(\varphi(g)) = g$ is well-defined. Certainly, φ^{-1} is surjective; it is injective because if $\varphi^{-1}(\varphi(g)) = g = h = \varphi^{-1}(\varphi(h))$, then $\varphi(g) = \varphi(h)$ by applying φ to each side of the identity $g = h$. Last, we have that

$$\varphi^{-1}(\varphi(g_1) \star \varphi(g_2)) = \varphi^{-1}(\varphi(g_1 * g_2)) = g_1 * g_2 = \varphi^{-1}(\varphi(g_1)) * \varphi^{-1}(\varphi(g_2)).$$

(3.) Given any elements $h_1, h_2 \in H$, we claim that $h_1 \star h_2 = h_2 \star h_1$. By assumption that φ is surjection, there exist elements $g_1, g_2 \in G$ such that $h_1 = \varphi(g_1)$ and $h_2 = \varphi(g_2)$. We conclude that

$$h_1 \star h_2 = \varphi(g_1) \star \varphi(g_2) = \varphi(g_1 * g_2) = \varphi(g_2 * g_1) = \varphi(g_2) \star \varphi(g_1) = h_2 \star h_1$$

by assumption that G is abelian; the same argument applied to $\varphi^{-1} : H \to G$ yields the converse.

(4.) If G is cyclic, then there exists an element $g \in G$ such that every element of G can be written as g^n for some integer n. Considering that φ is surjective, every element of H can be written as $h = \varphi(g^n) = \varphi(g * \cdots * g) = \varphi(g) \star \cdots \star \varphi(g) = \varphi(g)^n$ for some integer n. Consequently, we find that H is cyclic; it is generated by the image of the generator of G under the isomorphism φ.

(5.) By the fifth part of Proposition 3.3.5, every subgroup K of G induces the subgroup $\varphi(K)$ of H, hence H has at least as many subgroups as G. Conversely, every subgroup L of H induces the subgroup $\varphi^{-1}(L)$ of G, hence G has at least as many subgroups as H. We conclude that G and H possess the same number of subgroups. Last, we have that $\varphi(K) = \{e_H\}$ if and only if $K = e_G$ and $\varphi(K) = H$ if and only if $K = G$ because φ is a bijective group homomorphism. □

Using the language of group isomorphisms, we will formally establish that there is "essentially" only one infinite cyclic group, and there is "essentially" only one finite cyclic group.

Theorem 3.3.18. *Every infinite cyclic group is isomorphic to* $(\mathbb{Z}, +)$.

Proof. Consider any infinite cyclic group G. By definition, there exists an element $g \in G$ such that every element of G can be written as g^n for some integer $n \in \mathbb{Z}$. Observe that if $g^m = g^n$ for some integers m and n, then $e_G = g^m(g^n)^{-1} = g^m g^{-n} = g^{m-n}$ by the Group Exponent Laws, hence the order of g (i.e., the order of G) is finite — a contradiction. Consequently, every element of G can be written uniquely as g^n for some integer $n \in \mathbb{Z}$. We may therefore define a bijective function $\varphi : \mathbb{Z} \to G$ by $\varphi(n) = g^n$. Considering that $\varphi(m+n) = g^{m+n} = g^m g^n = \varphi(m)\varphi(n)$ by the Group Exponent Laws, we conclude that φ is an isomorphism, hence G is isomorphic to $(\mathbb{Z}, +)$. □

Lemma 3.3.19. *If $\varphi : G \to H$ is a group homomorphism of finite groups of the same order, then φ is an isomorphism if and only if φ is injective if and only if φ is surjective.*

Proof. By definition, we have that φ is an isomorphism if and only if φ is bijective if and only if φ is injective and surjective. By assumption that G and H are finite groups of the same order, Exercise 1.10.5(d.) implies that φ is bijective if and only if it is injective if and only if it is surjective. □

Theorem 3.3.20. *Every finite cyclic group of order n is isomorphic to* $(\mathbb{Z}/n\mathbb{Z}, +)$.

Proof. Consider a finite cyclic group G of order n. By definition, there exists an element $g \in G$ such that $G = \{g^k \mid 0 \leq k \leq n-1\}$. Consequently, we may define a function $\varphi : \mathbb{Z}/n\mathbb{Z} \to G$ by the assignment $\varphi(k + n\mathbb{Z}) = g^k$. We must demonstrate that φ is well-defined, i.e., $\varphi(k + n\mathbb{Z})$ does not depend upon the coset representative of $k + n\mathbb{Z}$. We will assume to this end that $k + n\mathbb{Z} = \ell + n\mathbb{Z}$. By definition, this means that $k - \ell = mn$ and $k = mn + \ell$ for some integer m so that

$$\varphi(k + n\mathbb{Z}) = g^k = g^{mn+\ell} = g^{mn} g^\ell = (g^n)^m g^\ell = (e_G)^m g^\ell = e_G g^\ell = g^\ell = \varphi(\ell + n\mathbb{Z})$$

by the Group Exponent Laws. We conclude that φ is well-defined; it is surjective by definition of $\mathbb{Z}/n\mathbb{Z}$ and φ, hence we conclude by Lemma 3.3.19 that φ is an isomorphism. □

Example 3.3.21. Consider the multiplicative group of complex numbers $G = \{1, -1, i, -i\}$. Observe that $i^2 = -1$, $i^3 = -i$, and $i^4 = 1$, hence G is a finite cyclic group of order four; it is generated by i. Consequently, by Theorem 3.3.20, we conclude that $(G, \cdot) \cong (\mathbb{Z}/4\mathbb{Z}, +)$. Explicitly, by the proof of the theorem, the isomorphism $\varphi : \mathbb{Z}/4\mathbb{Z} \to G$ is defined by $\varphi(n + 4\mathbb{Z}) = i^n$.

3.4 Cayley's Theorem

Cayley's Theorem is an example of a simple observation with larger implications.

Theorem 3.4.1 (Cayley's Theorem)**.** *Every group is isomorphic to a group of permutations.*

Proof. Given a group G and any element $g \in G$, consider the function $\varphi_g : G \to G$ defined by $\varphi_g(x) = gx$. By hypothesis that G is a group, it follows that g^{-1} is an element of G so that

$$\varphi_g \circ \varphi_{g^{-1}}(x) = \varphi_g(g^{-1}x) = gg^{-1}x = e_G x = x = e_G x = g^{-1}gx = \varphi_{g^{-1}}(gx) = \varphi_{g^{-1}} \circ \varphi_g(x)$$

for every element $x \in G$. Consequently, it follows that $\varphi_{g^{-1}}$ is the function inverse of φ_g so that φ_g is a bijection from G to itself by Exercise 1.10.6(b.). By definition, φ_g is a permutation of G, hence it is an element of the symmetric group $\mathfrak{S}_G$ on the set G. We claim that $\sigma : G \to \mathfrak{S}_G$ defined by $\sigma(g) = \varphi_g$ is a group homomorphism. Observe that for any elements $g_1, g_2, x \in G$, we have that

$$\sigma(g_1 g_2)(x) = \varphi_{g_1 g_2}(x) = g_1 g_2 x = \varphi_{g_1}(g_2 x) = \varphi_{g_1} \circ \varphi_{g_2}(x).$$

Considering that $x \in G$ is arbitrary, it follows that $\sigma(g_1 g_2) = \varphi_{g_1} \circ \varphi_{g_2} = \sigma(g_1) \circ \sigma(g_2)$ so that σ is a group homomorphism. Observe that $g \in \ker \sigma$ if and only if if and only if φ_g is the identity function from G to itself if and only if $gx = \varphi_g(x) = x$ for all elements $x \in G$ if and only if $g = e_G$ by cancellation in G. We conclude that σ is injective; its image $\sigma(G)$ is a subgroup of $\mathfrak{S}_G$ by the fifth part of Proposition 3.3.5. Consequently, we conclude that $G \cong (\sigma(G), \circ)$. □

Corollary 3.4.2. *Every finite group of order n is isomorphic to a subgroup of $\mathfrak{S}_n$.*

Proof. By Cayley's Theorem, every finite group G of order n is isomorphic to a subgroup of $\mathfrak{S}_G$. We claim that $(\mathfrak{S}_G, \circ) \cong (\mathfrak{S}_n, \circ)$. By Exercise 1.10.5(c.), there exists a bijection $f : G \to \{1, 2, \dots, n\}$ because these are finite sets of the same cardinality by assumption. We can extend f to a group isomorphism $\varphi : \mathfrak{S}_G \to \mathfrak{S}_n$ by declaring that for any permutation σ in $\mathfrak{S}_G$, we have that $\varphi(\sigma)$ is the permutation in $\mathfrak{S}_n$ that maps $f(g)$ to $f(h)$ whenever $\sigma(g) = h$. Explicitly, we may define a function $\varphi : \mathfrak{S}_G \to \mathfrak{S}_n$ by $\varphi(\sigma) = f \circ \sigma \circ f^{-1}$. We claim that φ is a well-defined group isomorphism.

1.) $\varphi(\sigma)$ is a permutation of $\{1, 2, \dots, n\}$ because it is a bijection from $\{1, 2, \dots, n\}$ to itself. Consequently, the function φ is well-defined because its image is a subset of $\mathfrak{S}_n$.

2.) φ is a group homomorphism because $\varphi(\sigma \circ \tau) = f \circ (\sigma \circ \tau) \circ f^{-1} = (f \circ \sigma \circ f^{-1}) \circ (f \circ \tau \circ f^{-1})$ shows that $\varphi(\sigma \circ \tau) = \varphi(\sigma) \circ \varphi(\tau)$ by the associativity of function composition.

3.) φ is a bijection with function inverse $\psi : \mathfrak{S}_n \to \mathfrak{S}_G$ defined by $\psi(\rho) = f^{-1} \circ \rho \circ f$. By Exercise 1.10.6, it suffices to note that $\psi \circ \varphi(\sigma) = \psi(f \circ \sigma \circ f^{-1}) = f^{-1} \circ (f \circ \sigma \circ f^{-1}) \circ f = \sigma$. □

Cayley's Theorem expressly motivates the study of symmetric groups because it demonstrates that every group can be "identified with" a group of permutations. Consequently, we will later return to our discussion of symmetric groups to develop even more tools to understand them.

3.5 The Group Isomorphism Theorems

Earlier in this chapter, we mentioned that one of the principal motivations in group theory (and the focus of this chapter) is the classification of groups. Explicitly, we seek to distinguish two groups based on properties such as their order, whether they are cyclic, whether they are abelian, and what kinds of subgroups they admit. Exercise 3.18.20 demonstrates that the existence of a group isomorphism between two groups is an equivalence relation; we say that two groups lie in the same equivalence class modulo this equivalence relation if and only if they are equal **up to isomorphism**. Consequently, we wish to determine all groups with a specified property $\mathcal{P}$ up to isomorphism.

Example 3.5.1. We have already seen in Example 2.2.12 that $\mathbb{Z}/4\mathbb{Z}$ and $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ are two groups of order four that are not isomorphic to one another. Explicitly, the only non-trivial proper subgroup of $\mathbb{Z}/4\mathbb{Z}$ is $2\mathbb{Z}/4\mathbb{Z} = \{0 + 4\mathbb{Z}, 2 + 4\mathbb{Z}\}$; however, there are three non-trivial proper subgroups of $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$, hence these two groups cannot be isomorphic by Proposition 3.3.17. Every cyclic group of order four is isomorphic to $\mathbb{Z}/4\mathbb{Z}$ by Proposition 3.3.20; we will soon see that every non-cyclic abelian group of order four is isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$.

We will now state and prove the **Group Isomorphism Theorems**; these four theorems provide us with a road map by which we may begin to tackle the classification problem of groups.

Theorem 3.5.2 (First Isomorphism Theorem)**.** *Given any groups $(G, *)$ and $(H, \star)$ and a group homomorphism $\varphi : G \to H$, there exists a group isomorphism $\psi : G/\ker\varphi \to \varphi(G)$.*

Proof. We note that $\varphi(G)$ is a subgroup of H by the fifth part of Proposition 3.3.5. Exercise 3.18.19 shows that $\ker\varphi$ is a normal subgroup of G, hence we may view $G/\ker\varphi$ as a group with respect to the operation $*$ of G. Even more, in order to prove the claim, it suffices to find a group isomorphism $\psi : G/\ker\varphi \to \varphi(G)$. Consider the function $\psi : G/\ker\varphi \to \varphi(G)$ defined by $\psi(g * \ker\varphi) = \varphi(g)$. Considering that ψ is defined on the equivalence classes of an equivalence relation, we must establish that ψ is well-defined, i.e., we must show that if $g*\ker\varphi = h*\ker\varphi$, then $\psi(g*\ker\varphi) = \psi(h*\ker\varphi)$. By Proposition 3.1.4, we have that $g * \ker\varphi = h * \ker\varphi$ if and only if $h^{-1}g \in \ker\varphi$ if and only if $\varphi(h^{-1}g) = e_H$ if and only if $\varphi(h^{-1}) \star \varphi(g) = e_H$ if and only if $\varphi(h)^{-1} \star \varphi(g) = e_H$ if and only if $\varphi(g) = \varphi(h)$ if and only if $\psi(g * \ker\varphi) = \psi(h * \ker\varphi)$. We conclude that ψ is well-defined. By hypothesis that φ is a group homomorphism, it follows that ψ is a group homomorphism. Even more, ψ is surjective because its image is $\varphi(G)$, hence it suffices to show that ψ is injective. Observe that $g * \ker\varphi \in \ker\psi$ if and only if $\varphi(g) = \psi(g * \ker\varphi) = e_H$ if and only if $g \in \ker\varphi$ if and only if $g * \ker\varphi = e_G * \ker\varphi$, hence we conclude that $\ker\psi$ is trivial so that ψ is injective, as desired. □

Theorem 3.5.3 (Second Isomorphism Theorem)**.** *Given any group G with a subgroup H and a normal subgroup N, we have that HN/N and $H/(H \cap N)$ are isomorphic groups.*

Proof. We must first demonstrate that HN is a subgroup of G such that N is a normal subgroup of HN; this proves that HN/N is a group. By Exercise 3.18.11, we find that HN is a subgroup of G, so we will prove that N is a normal subgroup of HN. Every element $n \in N$ can be written as $e_G n$ so that $N \subseteq HN$; moreover, N is a subgroup of G, so it is a subgroup of HN. Last, by Proposition 3.2.1, we have that $gN = Ng$ for all elements $g \in G$, so this identity also holds for all elements $g \in HN$. Put another way, if N is normal in G, then it is normal in any subgroup containing it.

By Exercise 3.18.12, it follows that $H \cap N$ is a normal subgroup of H and $H/(H \cap N)$ is a group. We may now appeal to the First Isomorphism Theorem, hence it suffices to find a surjective group homomorphism $\varphi : H \to HN/N$ such that $\ker \varphi = H \cap N$. Consider the function $\varphi : H \to HN/N$ defined by $\varphi(h) = hN$. Every element of HN/N is of the form $(hn)N$ for some elements $h \in H$ and $n \in N$. Considering that N is a subgroup of G, it follows that $nN = N$, hence every element of HN/N is of the form hN for some element $h \in H$. We conclude that φ is well-defined and surjective. Even more, we have that $\varphi(h_1 h_2) = h_1 h_2 N = (h_1 N)(h_2 N)$ because N is a normal subgroup of HN. Consequently, φ is a group homomorphism; its kernel consists of those elements $h \in H$ such that $hN = e_G N$. By Proposition 3.1.4, we have that $hN = e_G N$ if and only if $h \in N$, from which it follows that $\ker \varphi = H \cap N$. Our proof is complete by the First Isomorphism Theorem. □

Theorem 3.5.4 (Third Isomorphism Theorem)**.** *Given any group G with normal subgroups N and H such that $N \subseteq H$, we have that $(G/N)/(H/N)$ and G/N are isomorphic groups.*

Proof. By Proposition 3.2.1, we have that $gN = Ng$ for all elements $g \in G$, hence in particular, this identity also holds for all elements $g \in H$. We conclude that N is a normal subgroup of H because it is a subset of H that is a group with respect to the binary operation on G and N is normal in H. Consequently, it follows that H/N is a group; likewise, it is a subgroup of G/N because it is a subset of G/N that is a group under the binary operation on G/N. Even more, we claim that H/N is a normal subgroup of G/N. Consider an element gN of G/N and an element hN of H/N. By definition of the binary operation of G/N, we have that $(gN)(hN) = ghN$. By assumption that H is a normal subgroup of G, we have that $gH = Hg$ for all elements $g \in G$. Explicitly, there exists an element $k \in H$ such that $gh = kg$, from which it follows that $ghN = kgN = (kN)(gN)$. Considering that this holds for all elements $gN \in G/N$ and $hN \in H/N$, we conclude that $(gN)(H/N) \subseteq (H/N)(gN)$ for all elements $gN \in G/N$ so that H/N is a normal subgroup of G/N by Proposition 3.2.1.

We seek a surjective group homomorphism $\varphi : G/N \to G/H$ such that $\ker \varphi = H/N$. Consider the function $\varphi : G/N \to G/H$ defined by $\varphi(gN) = gH$. We must first establish that φ is well-defined because its domain consists of the left cosets of a group. Observe that if $g_1 N = g_2 N$, then $g_2^{-1} g_1$ is an element of N by Proposition 3.1.4. By assumption that $N \subseteq H$, it follows that $g_2^{-1} g_1$ is an element of H, hence the same proposition demonstrates that $\varphi(g_1 N) = g_1 H = g_2 H = \varphi(g_2 N)$ and φ is well-defined. Every element of G/H can be written as gH for some element $g \in G$. Even more, if g does not lie in H, then it does not lie in N because N is a subset of H, hence every left coset gH is the image of the left coset gN, i.e., φ is surjective. Last, we have that gN lies in $\ker \varphi$ if and only if $gH = \varphi(gN) = e_G H$ if and only if $g \in H$ by Proposition 3.1.4, hence we conclude that $\ker \varphi = H/N$. By the First Isomorphism Theorem, we conclude that $(G/H)/(H/N) \cong G/N$. □

Theorem 3.5.5 (Fourth Isomorphism Theorem)**.** *Given a group G with a normal subgroup N, there exists a one-to-one correspondence between the subgroups of G that contain N and the subgroups of G/N induced by the assignment of a subgroup H of G with $N \subseteq H$ to the subgroup H/N of G/N. Even more, this one-to-one correspondence satisfies the following properties.*

1.) *Given any subgroups H and K of G such that $N \subseteq H$ and $N \subseteq K$, we have that $H \subseteq K$ if and only if $H/N \subseteq K/N$. Put another way, this bijective correspondence is inclusion-preserving.*

2.) *Given any subgroups H and K of G such that $N \subseteq H \subseteq K$, we have that*

$$[K : H] = [K/N : H/N].$$

3.) *Given any subgroups H and K of G such that $N \subseteq H$ and $N \subseteq K$, we have that*

$$(H \cap K)/N = (H/N) \cap (K/N).$$

4.) *Given any subgroup H of G such that $N \subseteq H$, we have that $H \trianglelefteq G$ if and only if $H/N \trianglelefteq G/N$.*

Proof. We must prove first that the assignment of a subgroup H of G with $N \subseteq H$ to the subgroup H/N of G/N is both injective and surjective. Observe that if $H/N = K/N$, then for every element $h \in H$, there exists an element $k \in K$ such that $hN = kN$. Consequently, there exist elements $n_1, n_2 \in N$ such that $hn_1 = kn_2$ so that $h = kn_2n_1^{-1}$. By assumption that $N \subseteq K$, it follows that $h = kn_2n_1^{-1}$ is an element of K. We conclude that $H \subseteq K$. Conversely, an analogous argument demonstrates that $K \subseteq H$, from which it follows that $H = K$, and this assignment is injective. Given a subgroup Q of G/N, in order to prove that this assignment is surjective, we must furnish a subgroup H of G that contains N with the property that $Q = H/N$. Every element of G/N is a left coset of N in G, hence every element of Q is a left coset of N in G. Consider the collection $H = \{g \in G \mid gN \in Q\}$ of elements of G that give rise to elements of Q. By assumption that Q is a subgroup of G/N, the left coset e_GN lies in Q, hence we have that $e_G \in H$. Even more, for any elements $h_1, h_2 \in H$, we have that $h_1h_2N = (h_1N)(h_2N)$ lies in Q implies that $h_1h_2 \in H$ and $h_1^{-1}N = (h_1N)^{-1}$ lies in Q implies that $h_1^{-1} \in H$. We conclude by the Two-Step Subgroup Test that H is a subgroup of G. Given any element $n \in N$, we have that $nN = e_GN$ lies in Q, from which it follows that $N \subseteq H$ and $Q = H/N$. Ultimately, this shows that this assignment is surjective.

We turn our attention to the four asserted properties. We note that the first property holds by the first paragraph. Explicitly, if H and K are subgroups of G that contain N and satisfy that $H/N \subseteq K/N$, then it must be the case that $H \subseteq K$. Conversely, if we assume that $H \subseteq K$, then the inclusion $H/N \subseteq K/N$ holds by definition of left cosets. We note that the second property holds by the Third Isomorphism Theorem: if H and K are subgroups of G such that $N \subseteq H \subseteq K$, then the quotient groups K/H and $(K/N)/(H/N)$ are isomorphic; in particular, there is a bijection between K/H and $(K/N)/(H/N)$, hence the number of left cosets of H in K is equal to the number of left cosets of H/N in K/N. Put another way, we have that $[K : H] = [K/N : H/N]$. Even more, the third property holds by straightforward inspection: every element of $(H \cap K)/N$ is of the form $n(H \cap K)$, hence it is a left coset of N in both H and K. Conversely, every element of $(H/N) \cap (K/N)$ is a left coset of N in both H and K, hence it is a left coset of N in $H \cap K$.

Last, we turn our attention to the fourth property. We will assume to this end that H is a subgroup of G that contains N. We have already demonstrated in the proof of the Third Isomorphism Theorem that if H is a normal subgroup of G, then H/N is a normal subgroup of G/N. Conversely, suppose that H/N is a normal subgroup of G/N. Consequently, the **canonical surjections** $\pi_1 : G \to G/N$ and $\pi_2 : G/N \to (G/N)/(H/N)$ are group homomorphisms by Exercise 3.18.19; the composite function $\pi_2 \circ \pi_1 : G \to (G/N)/(H/N)$ defined by $\pi_2 \circ \pi_1(g) = gH$ is a group homomorphism with kernel H, hence H is a normal subgroup of G by Exercise 3.18.19. □

Example 3.5.6. Consider the general linear group $\operatorname{GL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) \neq 0\}$ under matrix multiplication. We will use the First Isomorphism Theorem to prove that the multiplicative group $(\mathbb{R}^\times, \cdot)$ of nonzero real numbers is isomorphic to a subgroup of $\operatorname{GL}(2, \mathbb{R})$. Consider the function $\varphi : \mathbb{R}^\times \to \operatorname{GL}(2, \mathbb{R})$ defined by $\varphi(c) = cI$, where I is the 2×2 identity matrix. Observe that φ is injective because $cI = dI$ if and only if $c = d$. Even more, it is a group homomorphism because for any real numbers c and d, we have that $\varphi(cd) = (cd)I = (cI)(dI) = \varphi(c)\varphi(d)$. Consequently, $(\mathbb{R}^\times, \cdot)$ is isomorphic to $\varphi(\mathbb{R}^\times) = \{cI \mid c \in \mathbb{R}^\times\}$, i.e., the nonzero real multiples of the identity matrix.

Example 3.5.7. We will prove next that multiplicative group $(\mathbb{R}_{>0}, +)$ of positive real numbers is isomorphic to a proper quotient of the multiplicative group $(\mathbb{R}^\times, \cdot)$ of nonzero real numbers, hence these groups are not isomorphic. Consider the function $\nu : \mathbb{R}^\times \to \mathbb{R}_{>0}$ defined by $\nu(x) = |x|$. Every positive real number can be written as its own absolute value, hence ν is surjective. Even more, ν is a group homomorphism because $\nu(xy) = |xy| = |x| \cdot |y|$. Consequently, we have that $x \in \ker \nu$ if and only if $|x| = 1$ if and only if $x = \pm 1$, hence we have that $\ker \nu = \{-1, 1\}$. By the First Isomorphism Theorem, we conclude that $(\mathbb{R}^\times/\{-1, 1\}, \cdot) \cong (\mathbb{R}_{>0}, \cdot)$ and $(\mathbb{R}_{>0}, \cdot) \ncong (\mathbb{R}^\times, \cdot)$.

Example 3.5.8. Before we conclude this section, we provide an example of the Third Isomorphism Theorem. Consider the additive group $(\mathbb{Z}/n\mathbb{Z}, +)$ of integers modulo a positive integer n. Given any integer m, we may also consider the additive group $(\mathbb{Z}/mn\mathbb{Z}, +)$. Observe that $n\mathbb{Z}/mn\mathbb{Z}$ the cyclic subgroup of $\mathbb{Z}/mn\mathbb{Z}$ generated by the image of n modulo $mn\mathbb{Z}$; in particular, we have that $n\mathbb{Z}/mn\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}/mn\mathbb{Z}$. By the Third Isomorphism Theorem, we have that

$$\frac{\mathbb{Z}/mn\mathbb{Z}}{n\mathbb{Z}/mn\mathbb{Z}} \cong \frac{\mathbb{Z}}{n\mathbb{Z}}.$$

Consequently, it grants no additional information to take subsequent quotients of $\mathbb{Z}$.

3.6 Group Automorphisms

If G is a group, then we refer to a group isomorphism $\varphi : G \to G$ as a **group automorphism** of G. Earlier, we demonstrated in Example 3.3.13 that for every element $g \in G$, the conjugation function $\chi_g : G \to G$ defined by $\chi_g(x) = gxg^{-1}$ is a group automorphism. We will consider the set $\operatorname{Inn}(G) = \{\chi_g : G \to G \mid g \in G\}$ of **inner automorphisms** of G. Even more, if G is an abelian group, then Example 3.3.15 illustrates that the inversion function $\varphi : G \to G$ defined by $\varphi(g) = g^{-1}$ is a group automorphism. We denote by $\operatorname{Aut}(G)$ the set of automorphisms of G, i.e., we have that

$$\operatorname{Aut}(G) = \{\varphi : G \to G \mid \varphi \text{ is a group isomorphism}\}.$$

Proposition 3.6.1. *If G is a group, then $\operatorname{Aut}(G)$ is a group under function composition.*

Proof. Observe that the identity function $\iota : G \to G$ defined by $\iota(g) = g$ is an automorphism of G. Consequently, $\operatorname{Aut}(G)$ is nonempty: ι is the identity element of $\operatorname{Aut}(G)$. Composition of functions is associative, and compositions of bijective homomorphisms are bijective homomorphisms. Last, every group isomorphism has an inverse that is a group isomorphism by Proposition 3.3.17. □

Proposition 3.6.2. *If G is a group, then $\operatorname{Inn}(G)$ is a group under function composition.*

Proof. Observe that the identity function $\iota : G \to G$ defined by $\iota(g) = g$ for every element g in G is an inner automorphism of G. Explicitly, we may identify ι with conjugation by e_G. Consequently, $\operatorname{Inn}(G)$ is a nonempty subset of $\operatorname{Aut}(G)$. By the One-Step Subgroup Test, it suffices to show that if χ_g and χ_h are in $\operatorname{Inn}(G)$, then $\chi_g \circ \chi_{h^{-1}}$ is in $\operatorname{Inn}(G)$. But this composite function is simply $\chi_{gh^{-1}}$. □

Proposition 3.6.3. *Let G be a group G with center $Z(G)$. We have that $G/Z(G) \cong \operatorname{Inn}(G)$.*

Proof. Consider the function $\varphi : G \to \operatorname{Inn}(G)$ defined by $\varphi(g)(x) = gxg^{-1}$. Given any elements $g, h \in G$, we have that $\varphi(gh)(x) = (gh)x(gh)^{-1} = (gh)x(h^{-1}g^{-1}) = g(hxh^{-1})g^{-1} = \varphi(g) \circ \varphi(h)(x)$ for all elements $x \in G$. We conclude that $\varphi(gh) = \varphi(g) \circ \varphi(h)$, hence φ is a group homomorphism. Certainly, φ is surjective because any element of $\operatorname{Inn}(G)$ is conjugation by an element $g \in G$ and $\varphi(g)$ is simply conjugation by g. Last, observe that $g \in \ker \varphi$ if and only if $\varphi(g)(x) = x$ for all elements $x \in G$ if and only if $gxg^{-1} = x$ for all elements $x \in G$ if and only if $gx = xg$ for all elements $x \in G$ if and only if $g \in Z(G)$. We conclude that $G/Z(G) \cong \operatorname{Inn}(G)$ by the First Isomorphism Theorem. □

Proposition 3.6.4. *If G is a group, then $\operatorname{Inn}(G)$ is cyclic if and only if $\operatorname{Inn}(G) = \{\iota\}$.*

Proof. Observe that if $\operatorname{Inn}(G) = \{\iota\}$, then $\operatorname{Inn}(G)$ is cyclic because the identity function $\iota : G \to G$ is the multiplicative identity element of $\operatorname{Inn}(G)$. Conversely, we will assume that $\operatorname{Inn}(G)$ is cyclic. By definition, there exists an element $g \in G$ such that $\operatorname{Inn}(G) = \langle \chi_g \rangle$. One can readily verify that the n-fold composite of χ_g with itself is conjugation by g^n for every integer n. Consequently, for any element $h \in G$, there exists an integer n such that $hxh^{-1} = \chi_h(x) = \chi_{g^n}(x) = g^n x (g^n)^{-1} = g^n x g^{-n}$ for all elements $x \in G$. Particularly, it must hold that $hgh^{-1} = g^n g g^{-n} = g^{n+1} g^{-n} = g$ so that $hg = gh$. But the same argument can be made to show that all elements of G commute with g. Ultimately, we conclude that $\chi_g(x) = gxg^{-1} = xgg^{-1} = x$ for all elements $x \in G$ and $\operatorname{Inn}(G) = \{\iota\}$. □

Corollary 3.6.5. *If G is a group with center $Z(G)$, then the following properties are equivalent.*

(i.) *G is abelian.*

(ii.) *$G/Z(G)$ is cyclic.*

(iii.) *$\operatorname{Inn}(G)$ is cyclic.*

(iv.) *$\operatorname{Inn}(G)$ is the trivial subgroup of $\operatorname{Aut}(G)$.*

Proof. We leave the proof for the reader in Exercise 3.18.30. □

Curiously, a group isomorphism induces an isomorphism of the automorphism groups; however, it is possible to find non-isomorphic groups with the same automorphism group (up to isomorphism).

Proposition 3.6.6. *If two groups $(G, *)$ and $(H, \star)$ are isomorphic, then $\operatorname{Aut}(G)$ and $\operatorname{Aut}(H)$ are isomorphic. Conversely, there exist non-isomorphic groups with the same automorphism group.*

Proof. We will assume first that G and H are isomorphic. Explicitly, consider a group isomorphism $\varphi : (G, *) \to (H, \star)$. By Proposition 3.3.17, there exists a group isomorphism $\varphi^{-1} : (H, \star) \to (G, *)$. Consequently, for any group automorphism $\gamma : (G, *) \to (G, *)$ of G, we may consider the function $\varphi \circ \gamma \circ \varphi^{-1} : (H, \star) \to (H, \star)$. Composition of group homomorphisms gives a group homomorphism. Likewise, composition of bijective functions yields a bijective function. We conclude therefore that the function $\psi : (\operatorname{Aut}(G), \circ) \to (\operatorname{Aut}(H), \circ)$ defined by $\psi(\gamma) = \varphi \circ \gamma \circ \varphi^{-1}$ is well-defined. Even more, it is a group homomorphism: given any automorphisms γ and δ of G, we have that

$$\psi(\gamma \circ \delta) = \varphi \circ (\gamma \circ \delta) \circ \varphi^{-1} = (\varphi \circ \gamma \circ \varphi^{-1}) \circ (\varphi \circ \delta \circ \varphi^{-1}) = \psi(\gamma) \circ \psi(\delta).$$

Every group automorphism $\sigma : (H, \star) \to (H, \star)$ of H can be written as

$$\sigma = \iota \circ \sigma \circ \iota = (\varphi \circ \varphi^{-1}) \circ \sigma \circ (\varphi \circ \varphi^{-1}) = \varphi \circ (\varphi^{-1} \circ \sigma \circ \varphi) \circ \varphi^{-1} = \psi(\varphi^{-1} \circ \sigma \circ \varphi)$$

for the group isomorphism $\varphi^{-1} \circ \sigma \circ \varphi : (G, *) \to (G, *)$, hence ψ is surjective. Even more, we have that $\gamma \in \ker \psi$ if and only if $\varphi \circ \gamma \circ \varphi^{-1} = \iota$ if and only if $\iota \circ \gamma \circ \iota = \varphi^{-1} \circ \iota \circ \varphi$. Observe that the function on the right-hand side is the identity function on G, hence the function on the left-hand side is the identity function on G. Considering that $\gamma(g) = (\iota \circ \gamma \circ \iota)(g) = g$ for all elements $g \in G$, we conclude that $\gamma = \iota$, hence ψ is injective. Consequently, ψ is a group isomorphism.

Conversely, consider the multiplicative group $(\{1\}, \cdot)$ and the cyclic group $(\mathbb{Z}/2\mathbb{Z}, +)$. We claim that these two non-isomorphic groups possess the same automorphism group $(\{\iota\}, \circ)$. Clearly, the only group automorphism of $(\{1\}, \cdot)$ is the identity function $\iota : \{1\} \to \{1\}$ because this is the only function from the set $\{1\}$ to itself. On the other hand, if $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is a group automorphism, then we must have that $\varphi(0 + 2\mathbb{Z}) = 0 + 2\mathbb{Z}$; this implies that $\varphi(1 + 2\mathbb{Z}) = 1 + 2\mathbb{Z}$ because φ must be injective, hence we conclude that φ is the identity function on $\mathbb{Z}/2\mathbb{Z}$. □

One particularly interesting application of automorphism groups is to provide a group-theoretic proof of the so-called **Euler's Theorem** and **Fermat's Little Theorem** from elementary number theory. Explicitly, let us consider the automorphism group $\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})$ for a positive integer n. By Proposition 3.3.20, every cyclic group of order n is isomorphic to $(\mathbb{Z}/n\mathbb{Z}, +)$, hence $\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})$ is the unique automorphism group for cyclic groups of order n (up to isomorphism) by Proposition 3.6.6. Consider the integer-valued **Euler totient function** $\phi : \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 1}$ defined by $\phi(1) = 1$ and $\phi(n) = \#\{k \mid 1 \leq k \leq n - 1 \text{ and } \gcd(k, n) = 1\}$, i.e., $\phi(n)$ is the number of positive integers that are relatively prime to n; we will demonstrate next that $|\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})| = \phi(n)$.

Proposition 3.6.7. *We have that $|\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})| = \phi(n)$. Put another way, there are exactly $\phi(n)$ distinct group automorphisms of $(\mathbb{Z}/n\mathbb{Z}, +)$ for each positive integer n.*

Proof. We claim that every group homomorphism $\psi : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ satisfies the property that $\psi(k + n\mathbb{Z}) = k\psi(1 + n\mathbb{Z})$ for some positive integer $1 \leq k \leq n - 1$. Explicitly, we have that

$$\psi(k + n\mathbb{Z}) = \psi(\underbrace{(1 + n\mathbb{Z}) + \cdots + (1 + n\mathbb{Z})}_{k \text{ summands}}) = \underbrace{\psi(1 + n\mathbb{Z}) + \cdots + \psi(1 + n\mathbb{Z})}_{k \text{ summands}} = k\psi(1 + n\mathbb{Z}).$$

Consequently, every group homomorphism $\psi : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ is uniquely determined by $\psi(1 + n\mathbb{Z})$ in the sense that a group homomorphism $\psi(1 + n\mathbb{Z}) = k\psi(1 + n\mathbb{Z}) + n\mathbb{Z}$. Even more, we have that ψ is a group automorphism if and only if there exists a group homomorphism $\psi^{-1} : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ such that $(\psi^{-1} \circ \psi)(k + n\mathbb{Z}) = k + n\mathbb{Z}$. Explicitly, we must have that

$$\psi(1 + n\mathbb{Z})\psi^{-1}(1 + n\mathbb{Z}) + n\mathbb{Z} = \psi^{-1}(\psi(1 + n\mathbb{Z}) + n\mathbb{Z}) = (\psi^{-1} \circ \psi)(1 + n\mathbb{Z}) = 1 + n\mathbb{Z}.$$

Consider the case that $\psi(1 + n\mathbb{Z}) = a + n\mathbb{Z}$ and $\psi^{-1}(1 + n\mathbb{Z}) = b + n\mathbb{Z}$. By definition, the above displayed identity holds if and only if there exists an integer m such that $ab - 1 = nm$ if and only if there exists an integer m such that $ab - nm = 1$ if and only if $\gcd(a, n) = 1$ by Bézout's Identity. Combined, these observations imply that $\psi : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ is a group automorphism if and only if $\psi(1 + n\mathbb{Z}) = a + n\mathbb{Z}$ satisfies that $\gcd(a, n) = 1$ so that $|\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})| = \phi(n)$. □

Theorem 3.6.8 (Euler's Theorem)**.** *If n is a positive integer, then for any integer a such that* $\gcd(a, n) = 1$, *then we must have that* $a^{\phi(n)} \equiv 1 \pmod{n}$.

Proof. Consider the function $\psi : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ defined by $\psi(1 + n\mathbb{Z}) = a + n\mathbb{Z}$. By Proposition 3.6.7, ψ is a group automorphism. Even more, we have that the composite function $\psi^{\phi(n)}$ of ψ with itself $\phi(n)$ times must be the identity function on $\mathbb{Z}/n\mathbb{Z}$ by Corollary 3.1.16. Consequently, we have that $1 + n\mathbb{Z} = \psi^{\phi(n)}(1 + n\mathbb{Z}) = (a + n\mathbb{Z})^{\phi(n)} = a^{\phi(n)} + n\mathbb{Z}$, i.e., $a^{\phi(n)} \equiv 1 \pmod{n}$. □

Theorem 3.6.9 (Fermat's Little Theorem)**.** *If p is a prime number, then for any integer a such that* $\gcd(a, p) = 1$, *we have that* $a^{p-1} \equiv 1 \pmod{p}$.

Proof. By Euler's Theorem, we have that $a^{\phi(p)} \equiv 1 \pmod{p}$. Consequently, it suffices to note that every integer $1 \leq k \leq p - 1$ satisfies that $\gcd(k, p) = 1$, hence we conclude that $\phi(p) = p - 1$. □

Euler's Theorem can be used to perform fast modular arithmetic as follows.

Example 3.6.10. Observe that if a is any integer, then $a \pmod{10}$ is the digit in the ones place of a. Explicitly, we have that $11 \pmod{10} = 1$. Considering that $10 = 2 \cdot 5$ is the prime factorization of 10, we have that $\phi(10) = 4$ because the positive integers 1, 3, 5, 7, and 9 are all relatively prime to 10. By Euler's Theorem, we have that $999999^4 \equiv 1 \pmod{10}$ because it holds that $\gcd(999999, 10) = 1$.

Even more, the proof of Proposition 3.6.7 directly implies the following two results.

Corollary 3.6.11. *Given any positive integer n, we have that* $\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$ *for the multiplicative group* $(\mathbb{Z}/n\mathbb{Z})^\times = \{a + n\mathbb{Z} \in \mathbb{Z}/n\mathbb{Z} \mid \gcd(a, n) = 1\}$ *of integers modulo n.*

Corollary 3.6.12. *Given any positive integer n, we have that* $\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})$ *is abelian.*

Proof. By Proposition 3.6.7, every group automorphism $\psi : (\mathbb{Z}/n\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ is uniquely determined by $\psi(1 + n\mathbb{Z}) = a + n\mathbb{Z}$ for some integer $1 \leq a \leq n - 1$ such that $\gcd(a, n) = 1$. Consequently, the function $\varphi : \operatorname{Aut}(\mathbb{Z}/n\mathbb{Z}) \to (\mathbb{Z}/n\mathbb{Z})^\times$ defined by $\varphi(\psi) = a + n\mathbb{Z}$ is surjective. By Exercises 1.10.5(d.) and 2.8.28, we conclude that φ is a bijection, so it suffices to prove that φ is a group homomorphism. Given any group automorphism γ of $(\mathbb{Z}/n\mathbb{Z}, +)$, we have that $\gamma(1 + n\mathbb{Z}) = b + n\mathbb{Z}$ for some integer $1 \leq b \leq n - 1$ such that $\gcd(b, n) = 1$. Consequently, $\gamma \circ \psi$ satisfies that

$$(\gamma \circ \psi)(1 + n\mathbb{Z}) = \gamma(a + n\mathbb{Z}) = a\gamma(1 + n\mathbb{Z}) = a(b + n\mathbb{Z}) = ab + n\mathbb{Z} = (a + n\mathbb{Z})(b + n\mathbb{Z}).$$

We conclude that $\varphi(\gamma \circ \psi) = (b + n\mathbb{Z})(a + n\mathbb{Z}) = \varphi(\gamma)\varphi(\psi)$, i.e., φ is a group homomorphism. □

3.7 The Symmetric Group on n Letters, Revisited

Back in Section 2.5, we introduced the group $(\mathfrak{S}_n, \circ)$ of bijections $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ under function composition, and we referred to this as the **symmetric group on n letters**. Quite a bit can be said about this very interesting group, so we return here to explore and discuss its properties further. By Proposition 2.5.13, we have that $\mathfrak{S}_n$ is not abelian for any integer $n \geq 3$. One way to measure the degree to which a group G fails to be abelian is by examining its **center** $Z(G) = \{x \in G \mid gx = xg \text{ for all elements } g \in G\}$. Observe that $Z(G)$ is the collection of all elements of G that commute with all other elements of G. By Exercise 2.8.18, the center of a group is the largest abelian subgroup of the group. Consequently, the center of a group measures how "abelian" the group is — namely, the "larger" $Z(G)$ is, the "more abelian" G must be.

Proposition 3.7.1. *For every integer $n \geq 3$, the center $Z(\mathfrak{S}_n)$ of the symmetric group $\mathfrak{S}_n$ is $\{\iota\}$.*

Proof. Certainly, the identity function $\iota : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ defined by $\iota(i) = i$ for all integers $1 \leq i \leq n$ commutes with every bijection $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$; ι is a bijection, so we have that $Z(\mathfrak{S}_n) \supseteq \{\iota\}$. On the contrary, we will assume that there exists a non-trivial permutation σ of $Z(\mathfrak{S}_n)$. Consequently, there exist distinct integers i and j such that $\sigma(i) = j$. By hypothesis that $n \geq 3$, there exists another integer k distinct from i and j. Consider the transposition $\tau = (i, k)$. We have that $\sigma\tau(i) = \sigma(k) \neq j = \tau(j) = \tau\sigma(i)$. For if it were the case that $\sigma(k) = j$, then we would have that $\sigma(k) = \sigma(i)$ so that $k = i$ by hypothesis that σ is a bijection — a contradiction. But then, σ does not commute with τ, contradicting our assumption that σ is in $Z(\mathfrak{S}_n)$. □

Given any permutation $\sigma \in \mathfrak{S}_n$ with cycle decomposition $\sigma_1 \cdots \sigma_k$ such that σ_i has length r_i, we may rearrange (if necessary) σ_i so that $r_1 \leq \cdots \leq r_k$. We refer to the ordered k-tuple $(r_1, \dots, r_k)$ as the **cycle type** of σ. Considering that an ordered k-tuple $(r_1, \dots, r_k)$ with $r_1 \leq \cdots \leq r_k$ and $r_1 + \cdots + r_k = n$ is an integer partition of n with k parts by definition, we have the following.

Proposition 3.7.2. *Given any positive integer n, the number of distinct cycle types of permutations in $\mathfrak{S}_n$ is equal to the number of distinct integer partitions of n.*

Given any element x of a group G, we say that an element of the form gxg^{-1} for some element $g \in G$ is **conjugate** to x. We will discuss this notation later in our section on the group actions and the Class Equation. Our next proposition states that cycle type is unique up to conjugation. One direction of this statement is established in the following proposition; the other is proved afterward.

Proposition 3.7.3. *Given any k-cycle $\sigma = (a_1, \dots, a_k)$, we have that $\tau\sigma\tau^{-1} = (\tau(a_1), \dots, \tau(a_k))$ for every element τ in $\mathfrak{S}_n$ with $n \geq k$. Put another way, cycle type is preserved under conjugation.*

Proof. Given any integer $1 \leq i \leq n$, we will assume that $\sigma(i) = j$. By hypothesis that τ is a permutation, it follows that τ^{-1} exists and satisfies $\tau^{-1}(\tau(i)) = i$ so that $\sigma\tau^{-1}(\tau(i)) = \sigma(i) = j$. Consequently, we have that $\tau\sigma\tau^{-1}(\tau(i)) = \tau(j)$ so that $\tau\sigma\tau^{-1}$ sends $\tau(i)$ to $\tau(j)$.

Given that σ is the k-cycle $\sigma = (a_1, \dots, a_k)$, it follows that σ fixes all integers in $[n] \setminus \{a_1, \dots, a_k\}$, hence $\tau\sigma\tau^{-1}$ fixes all integers in $[n] \setminus \{\tau(a_1), \dots, \tau(a_k)\}$. Likewise, we have that $\sigma(a_i) = a_{i+1}$ for all integers $1 \leq i \leq k-1$ and $\sigma(a_k) = a_1$, hence $\tau\sigma\tau^{-1}$ maps $\tau(a_i)$ to $\tau(a_{i+1})$ for all integers $1 \leq i \leq k-1$ and $\tau\sigma\tau^{-1}$ maps $\tau(a_k)$ to $\tau(a_1)$. Put another way, we have that $\tau\sigma\tau^{-1} = (\tau(a_1), \dots, \tau(a_k))$. □

Proposition 3.7.4. *For any permutations $\rho, \sigma \in \mathfrak{S}_n$, there exists a permutation $\tau \in \mathfrak{S}_n$ such that $\tau\rho\tau^{-1} = \sigma$ (i.e., ρ and σ are conjugate in $\mathfrak{S}_n$) if and only if ρ and σ have the same cycle type.*

Proof. By definition, if ρ and σ are conjugate in $\mathfrak{S}_n$, then there exists a permutation $\tau \in \mathfrak{S}_n$ such that $\tau\rho\tau^{-1} = \sigma$. We may assume that $\rho = \rho_k \cdots \rho_2\rho_1$ is the cycle decomposition of ρ so that

$$\sigma = \tau\rho\tau^{-1} = (\tau\rho_k\tau^{-1}) \cdots (\tau\rho_2\tau^{-1})(\tau\rho_1\tau^{-1})$$

is the cycle decomposition of σ. By Proposition 3.7.3, it follows that $\tau\rho_i\tau^{-1}$ are cycles of the same length as ρ_i, hence we must have that ρ and σ have the same cycle type.

Conversely, we will assume that ρ and σ have the same cycle type $(r_1, \dots, r_k)$. Consequently, we have that $\rho = \rho_k \cdots \rho_1$ and $\sigma = \sigma_k \cdots \sigma_1$ for some disjoint cycles ρ_i and some disjoint cycles σ_i

with $\text{length}(\rho_i) = r_i = \text{length}(\sigma_i)$. Considering that $[n]$ is a finite set, we may construct a bijection $\tau : [n] \to [n]$ that maps the cycle ρ_i to the cycle σ_i. Even more, we may construct τ in such a way that for any cycle $\rho_i = (a_{i,1}, \dots, a_{i,r_i})$ and the corresponding cycle $\sigma_i = (b_{i,1}, \dots, b_{i,r_i})$, we have that $\tau(a_{i,j}) = b_{i,j}$. We claim that $\tau\rho\tau^{-1} = \sigma$. By Proposition 3.7.3, we have that

$$\tau\rho\tau^{-1}(\tau(a_{i,j})) = \tau\rho(a_{i,j}) = \tau(a_{i+1,j}) = b_{i+1,j} = \sigma(b_{i,j}) = \sigma(\tau(a_{i,j})).$$

By construction, we have that τ is a bijection from $[n]$ to itself, hence every element of $[n]$ can be written as $\tau(a_{i,j})$ for some integer $1 \leq a_{i,j} \leq n$. We conclude therefore that $\tau\rho\tau^{-1} = \sigma$. □

Example 3.7.5. Exhibit a permutation that conjugates (123) and (132).

Solution. By the proof of Proposition 3.7.4, we have that $\tau(1) = 1$, $\tau(2) = 3$, and $\tau(3) = 2$, i.e., $\tau = (1)(23)$. Let us verify that $\tau\rho\tau^{-1} = \sigma$. Considering that $\tau\tau = (1)(23)(1)(23) = (1)(2)(3) = \iota$, it follows that $\tau = \tau^{-1}$ so that $\tau\rho\tau^{-1} = (1)(23)(123)(1)(23) = (132) = \sigma$, as desired. ◇

Example 3.7.6. Exhibit a permutation that conjugates (135)(27)(48)(6) and (1)(258)(34)(67).

Solution. Certainly, we could proceed as outlined in the proof of Proposition 3.7.4; however, this answer from Arturo Magidin gives a simpler approach. Begin by writing down the cycle types of the permutations $\rho = (135)(27)(48)(6)$ and $\sigma = (1)(258)(34)(67)$; then, arrange the cycles of ρ and σ in some (not necessarily unique) manner so that the cycles have non-decreasing length; and last, construct a 2×8 array whose first row is ρ and second row is σ. Observe that the cycle type of ρ and σ is $(1, 2, 2, 3)$, hence we may arrange $\rho = (6)(27)(48)(135)$ and $\sigma = (1)(34)(67)(258)$ to obtain

$$\tau = \begin{pmatrix} 6 & 2 & 7 & 4 & 8 & 1 & 3 & 5 \\ 1 & 3 & 4 & 6 & 7 & 2 & 5 & 8 \end{pmatrix}.$$

By reading off the array, we find that $\tau = (12358746)$. Observe that τ conjugates ρ to σ if and only if $\tau\rho\tau^{-1} = \sigma$ if and only if $\tau\rho = \sigma\tau$. We leave it to the reader to verify that $\tau\rho = \sigma\tau$, as desired. ◇

We turn our attention now to the matter of the combinatorics (or mathematics of counting) in the symmetric group. Our first result follows immediately from Propositions 3.7.2 and 3.7.4.

Proposition 3.7.7. *Given any positive integer n, the number of distinct conjugacy classes of the symmetric group $\mathfrak{S}_n$ is equal to the number of distinct integer partitions of n.*

Proof. By Proposition 3.7.4 above, there exists a bijection

$$\{\text{distinct conjugacy classes of } \mathfrak{S}_n\} \leftrightarrow \{\text{distinct cycle types of permutations in } \mathfrak{S}_n\}$$

that sends the conjugacy class of some permutation ρ with cycle type $(r_1, \dots, r_k)$ to the cycle type $(r_1, \dots, r_k)$. Explicitly, the permutations ρ and σ are conjugate (and hence in the same conjugacy class) if and only if they have the same cycle type, hence this map is injective. Further, this map is surjective because for any cycle type $(r_1, \dots, r_k)$, we can construct a permutation ρ with cycle type $(r_1, \dots, r_k)$. By Proposition 3.7.4, conjugation preserves cycle type. Consequently, we have that

$$\#\{\text{distinct conjugacy classes of } \mathfrak{S}_n\} = \#\{\text{distinct cycle types of permutations in } \mathfrak{S}_n\}.$$

By Proposition 3.7.2 above, the latter is equal to the number of distinct integer partitions of n. □

Often, the best way to count something is to establish a bijection between what we want to count and something for which we already know the cardinality; however, counting can sometimes be successfully accomplished by naïvely underestimating and multiplying by the number of times each element in the set was undercounted. We illustrate this principle in the following proposition.

Proposition 3.7.8. *Given any positive integer n, the number of distinct k-cycles in $\mathfrak{S}_n$ is $\frac{n!}{k(n-k)!}$.*

Proof. Every k-cycle in $\mathfrak{S}_n$ is constructed in the following manner.

1.) Choose k elements from among the n elements of $[n]$. We can do this in $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ways.

2.) Order the k elements in some way. Bear in mind that there is no “first” term in the ordering because $(a_1, \dots, a_k)$ is the same as $(a_k, a_1, \dots, a_{k-1})$, etc. Consequently, the order only matters for $k-1$ of the elements, hence there are $(k-1)!$ ways to order the k elements.

By the Fundamental Counting Principle, there are $\frac{n!}{k!(n-k)!}(k-1)! = \frac{n!}{k(n-k)!}$ k-cycles in $\mathfrak{S}_n$. □

3.8 The Alternating Group on n Letters

We turn our attention to an important subgroup of the symmetric group on n letters. Until now, we have only briefly mentioned the notion of a transposition, i.e., a cycle of length two; however, transpositions are in fact ubiquitous in the study of permutations, as we demonstrate next.

Proposition 3.8.1. *Every permutation can be written as the product of a unique number of transpositions; however, the transpositions need not be disjoint.*

Proof. Considering that every permutation can be written as the product of disjoint cycles, it suffices to show that any cycle $(a_1, \dots, a_k)$ can be written as a product of (not necessarily disjoint) transpositions. We note that $(a_1, \dots, a_k) = (a_1, a_k)(a_1, a_{k-1}) \cdots (a_1, a_2)$ is a product of (not necessarily disjoint) transpositions. By Proposition 3.7.3, we have that cycle type is unique up to conjugation, hence the number of transpositions is uniquely determined by the cycle type of a permutation. □

Considering that every permutation σ in $\mathfrak{S}_n$ can be written as the product of a unique number of (not necessarily disjoint) transpositions, we can define the **parity** of a permutation to be the parity (even or odd) of the number $t(\sigma)$ of transpositions in the transposition decomposition of σ. Further, we refer to the number $\operatorname{sgn}(\sigma) = (-1)^{t(\sigma)}$ as the **sign** of the permutation σ. Observe that σ is even if and only if $\operatorname{sgn}(\sigma) = 1$, and likewise, σ is odd if and only if $\operatorname{sgn}(\sigma) = -1$.

Proposition 3.8.2. *Consider the function* $\operatorname{sgn} : \mathfrak{S}_n \to \{-1, 1\}$ *defined by* $\operatorname{sgn}(\sigma) = (-1)^{t(\sigma)}$. *We have that* $\ker(\operatorname{sgn})$ *is a normal subgroup of* $\mathfrak{S}_n$ *such that* $[\mathfrak{S}_n : \ker(\operatorname{sgn})] = 2$ *and* $|\ker(\operatorname{sgn})| = n!/2$.

Proof. Observe that $\{-1, 1\}$ is a multiplicative group with identity 1. Consequently, the fact that

$$\operatorname{sgn}(\rho\sigma) = (-1)^{t(\rho\sigma)} = (-1)^{t(\rho)+t(\sigma)} = (-1)^{t(\rho)}(-1)^{t(\sigma)} = \operatorname{sgn}(\rho)\operatorname{sgn}(\sigma)$$

illustrates that sgn is a group homomorphism. By Exercise 3.18.19, the kernel of any group homomorphism from G is a normal subgroup of G, hence $\ker(\operatorname{sgn})$ is a normal subgroup of $\mathfrak{S}_n$. By definition of the index, we have that $[\mathfrak{S}_n : \ker(\operatorname{sgn})] = |\mathfrak{S}_n / \ker(\operatorname{sgn})|$. By the First Isomorphism Theorem, we conclude that $\mathfrak{S}_n / \ker(\operatorname{sgn}) \cong \{-1, 1\}$ so that $|\mathfrak{S}_n / \ker(\operatorname{sgn})| = |\{-1, 1\}| = 2$. Last, by Lagrange's Theorem and Proposition 2.5.1, we conclude that $|\ker(\operatorname{sgn})| = |\mathfrak{S}_n| / [\mathfrak{S}_n : \ker(\operatorname{sgn})] = n!/2$. □

We refer to the normal subgroup $\ker(\operatorname{sgn})$ of $\mathfrak{S}_n$ of Proposition 3.8.2 as the **alternating group on n letters**; it is denoted by $\mathfrak{A}_n$ (Fraktur "A") in accordance with the notation $\mathfrak{S}_n$. Observe that a permutation σ lies in $\ker(\operatorname{sgn})$ if and only if $\operatorname{sgn}(\sigma) = 1$ if and only if σ is even, hence the alternating group on n letters is precisely the subgroup of $\mathfrak{S}_n$ consisting of even permutations. Even more, this matches with our intuition: the identity map $\iota : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ is the identity element of $\mathfrak{S}_n$; it can be represented as a the product of 1-cycles $\iota = (1)(2)\cdots(n)$ with 0 transpositions, hence we have that $\operatorname{sgn}(\iota) = (-1)^{t(\iota)} = (-1)^0 = 1$ so that ι is even. Given any two even permutations ρ and σ, we have that $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$ because σ^{-1} has the same cycle type as σ and hence the same number of transpositions. By the One-Step Subgroup Test, we conclude that $\operatorname{sgn}(\rho\sigma^{-1}) = (-1)^{t(\rho\sigma^{-1})} = (-1)^{t(\rho)+t(\sigma^{-1})} = (-1)^{t(\rho)+t(\sigma)} = (-1)^{2r+2s} = 1$ so that $\rho\sigma^{-1}$ is even.

Proposition 3.8.3. *Every permutation of odd order is even; however, the converse is not true. Explicitly, there exist even permutations with even order.*

Proof. Given that σ is a permutation of odd order, it follows that $\operatorname{lcm}(r_1, \dots, r_k)$ is odd, where $(r_1, \dots, r_k)$ is the cycle type of σ. Consequently, we must have that r_i is odd for each integer $1 \leq i \leq k$. By the proof of Proposition 3.8.1, an r_i-cycle is the product of $r_i - 1$ transpositions, hence σ is the product of $(r_1 - 1) + \cdots + (r_k - 1)$ transpositions. Each of the integers $r_i - 1$ is even, so this sum is even, and σ is a product of an even number of transpositions, i.e., σ is even.

Conversely, if σ is the product of an even number of disjoint transpositions, then σ is even by definition of the parity of a permutation, and the order of σ is two by Proposition 2.5.7. □

3.9 External and Internal Direct Products

Earlier in these notes, we encountered the **Klein four-group** $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. We tacitly assumed that the binary operation on this group was inherited from the binary operation on the underlying factors of the group. Explicitly, the group operation on $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ is defined as the componentwise sum of a pair of elements in the group, i.e., we have that $(a, b) + (c, d) = (a + c, b + d)$.

Generally, for any finite collection of groups $G_1, \dots, G_n$, the **Cartesian product** $G_1 \times \cdots \times G_n$ of these groups can be equipped with a binary operation determined componentwise by the binary operations of the underlying groups. Explicitly, if $(G, *)$ and $(H, \star)$ are groups, then the Cartesian product $G \times H$ has an induced binary operation $(g_1, h_1)(g_2, h_2) = (g_1 * g_2, h_1 \star h_2)$. Out of desire for notational convenience, when working with the Cartesian product of an arbitrary number of groups, we will simply adopt the following convention to express a product in $G_1 \times \cdots \times G_n$.

$$(g_1, \dots, g_n)(h_1, \dots, h_n) = (g_1h_1, \dots, g_nh_n)$$

Proposition 3.9.1. *If $G_1, \dots, G_n$ are any groups, then their Cartesian product $G_1 \times \cdots \times G_n$ is a group with respect to the componentwise binary operation $(g_1, \dots, g_n)(h_1, \dots, h_n) = (g_1h_1, \dots, g_nh_n)$.*

Proof. Essentially, the three group axioms are satisfied because the underlying objects in the Cartesian product are groups, but we bear out the details for the sake of completeness. By definition, the binary operation on $G_1 \times \cdots \times G_n$ is associative because the binary operations of each underlying group are associative; the identity element of $G_1 \times \cdots \times G_n$ must be $(e_{G_1}, \dots, e_{G_n})$; and the inverse of any element $(g_1, \dots, g_n)$ of $G_1 \times \cdots \times G_n$ must be the element $(g_1^{-1}, \dots, g_n^{-1})$ of $G_1 \times \cdots \times G_n$. □

We refer to the group $G_1 \times \cdots \times G_n$ as the **external direct product** of $G_1, \dots, G_n$. We note that if each group G_i is finite, then the external direct products of $G_1, \dots, G_n$ must also be finite. Explicitly, we have that $|G_1 \times \cdots \times G_n| = |G_1| \cdots |G_n|$ by the Fundamental Counting Principle.

Example 3.9.2. Considering that the real numbers form an additive group $(\mathbb{R}, +)$, it follows that $(\mathbb{R} \times \mathbb{R}, +)$ is a group under componentwise addition. Often, this group is denoted simply by $\mathbb{R}^2$; its elements are ordered pairs (x, y) of real numbers. Generally, $\mathbb{R}^n$ is defined analogously for $n \geq 3$.

Example 3.9.3. Consider the group $(\mathbb{Z}/2\mathbb{Z})^n$, i.e., the external direct product of $\mathbb{Z}/2\mathbb{Z}$ with itself n times for any positive integer n. Elements of $(\mathbb{Z}/2\mathbb{Z})^n$ are by definition n-tuples consisting of zeros and ones. Explicitly, if $n = 3$, the elements $(0, 1, 0)$ and $(1, 0, 1)$ satisfy that $(0, 1, 0)(1, 0, 1) = (1, 1, 1)$. We note that this operation coincides with addition of integers base two: indeed, there is no significant reason to distinguish between $(0, 1, 0)$ and 010 or $(1, 0, 1)$ and 101.

Example 3.9.4. We may also form the external direct product of $(\mathbb{R}, +)$ and $(\mathbb{R}^\times, \cdot)$ whose elements are ordered pairs (x, y) such that x is a real number and y is a nonzero real number. Unlike the previous example, the operation on each component is different: $(x_1, y_1)(x_2, y_2) = (x_1 + x_2, y_1 \cdot y_2)$.

We will demonstrate that many of the properties of the external direct product of finitely many groups are intimately intertwined with the properties of the underlying groups themselves.

Proposition 3.9.5. *Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. We have that $G_1 \times \cdots \times G_n$ is abelian if and only if $G_1, \dots, G_n$ are abelian.*

Proposition 3.9.6. *Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. For any permutation σ of the indices $1, \dots, n$, we have that $G_1 \times \cdots \times G_n \cong G_{\sigma(1)} \times \cdots \times G_{\sigma(n)}$. Put another way, the arrangement of the factors in the external direct product is unique up to isomorphism.*

We leave the proofs of the previous two propositions as Exercises 3.18.44 and 3.18.46. We demonstrate next that the order of an n-tuple in the external direct product of groups is uniquely determined in an elegant manner by the order of the individual components of the n-tuple.

Proposition 3.9.7. *Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. We have that $\operatorname{ord}(g_1, \dots, g_n) = \operatorname{lcm}(\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n))$.*

Proof. By definition, the least common multiple ℓ of $\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n)$ is divisible by each of the integers $\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n)$, hence we have that $(g_1, \dots, g_n)^\ell = (g_1^\ell, \dots, g_n^\ell) = (e_{G_1}, \dots, e_{G_n})$. Consequently, we must have that $\operatorname{ord}(g_1, \dots, g_n)$ divides $\operatorname{lcm}(\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n))$ by Corollary 2.3.13. Conversely, if we denote by r the order $\operatorname{ord}(g_1, \dots, g_n)$ of $(g_1, \dots, g_n)$, then by definition of the order of an element of a group, we have that $(g_1^r, \dots, g_n^r) = (g_1, \dots, g_n)^r = (e_{G_1}, \dots, e_{G_n})$. Corollary 2.3.13 implies yet again that $\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n)$ divide r, hence r is a common multiple of $\operatorname{ord}(g_1, \dots, \operatorname{ord}(g_n)$ — namely, we conclude that $\operatorname{lcm}(\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n))$ divides r. Each of these integers is positive, and they divide each other, hence they must be equal to one another. □

Proposition 3.9.8. *Let $G_1, \dots, G_n$ be finite cyclic groups with external direct product $G_1 \times \cdots \times G_n$. We have that $G_1 \times \cdots \times G_n$ is cyclic if and only if $\gcd(|G_1|, \dots, |G_n|) = 1$.*

Proof. By Exercise 1.10.34, if $\gcd(|G_1|, \dots, |G_n|) = 1$, then $\operatorname{lcm}(|G_1|, \dots, |G_n|) = |G_1| \cdots |G_n|$ so that the external direct product $G_1 \times \cdots \times G_n$ is the cyclic group generated by $(g_1, \dots, g_n)$ such that $G_i = \langle g_i \rangle$ for each integer $1 \leq i \leq n$. Conversely, we will assume that $G_1 \times \cdots \times G_n$ is cyclic with a generator $(g_1, \dots, g_n)$. We claim that $G_i = \langle g_i \rangle$ for each integer $1 \leq i \leq n$. Every element of $G_1 \times \cdots \times G_n$ can be written as $(g_1, \dots, g_n)^k = (g_1^k, \dots, g_n^k)$ for some integer $1 \leq k \leq |G_1| \cdots |G_n|$, hence every element of G_i is equal to g_i^k for some integer k. By Proposition 3.9.7, we conclude that

$$|G_1| \cdots |G_n| = |G_1 \times \cdots \times G_n| = \operatorname{ord}(g_1, \dots, g_n) = \operatorname{lcm}(\operatorname{ord}(g_1), \dots, \operatorname{ord}(g_n)) = \operatorname{lcm}(|G_1|, \dots, |G_n|).$$

Last, by Exercise 1.10.34 again, we have that $|G_1| \cdots |G_n| = \operatorname{lcm}(|G_1|, \dots, |G_n|) \gcd(|G_1|, \dots, |G_n|)$, and we conclude the desired result from the above displayed equation. □

One of the most important features of the external direct product $G_1 \times \cdots \times G_n$ of groups is that each of the direct factors G_i can be identified with the normal subgroup of $G_1 \times \cdots \times G_n$ obtained by "forgetting about" the other direct factors. Even more, the external direct product $G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$ can be identified with the normal subgroup of $G_1 \times \cdots \times G_n$ obtained by "forgetting about" the direct factor G_i. Conversely, any direct factor and any direct subproduct can be viewed as a quotient of $G_1 \times \cdots \times G_n$. We make these notions rigorous as follows.

Proposition 3.9.9. *Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$.*

1.) *Given any integer $1 \leq i \leq n$, we may view G_i as a subgroup of $G_1 \times \cdots \times G_n$ via the group homomorphism $\gamma_i : G_i \to G_1 \times \cdots \times G_n$ defined by $\gamma_i(g) = (e_{G_1}, \dots, g, \dots, e_{G_n})$. Explicitly, the function γ_i sends an element $g \in G_i$ to the n-tuple in $G_1 \times \cdots \times G_n$ whose ith component is g and jth component is the identity element $e_{G_j} \in G_j$ for each integer $1 \leq j \leq n$ except i.*

$$G_i \cong \{(e_{G_1}, \dots, g, \dots, e_{G_n}) \mid g \in G_i\} \subseteq G_1 \times \cdots \times G_i \times \cdots \times G_n$$

2.) *Given any integer $1 \leq i \leq n$, we may view $G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$ as a subgroup of $G_1 \times \cdots \times G_n$ via the group homomorphism $\varepsilon_i : G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n \to G_1 \times \cdots \times G_n$ defined by $\varepsilon_i(g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n) = (g_1, \dots, g_{i-1}, e_{G_i}, g_{i+1}, \dots, g_n)$.*

$$G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n \cong G_1 \times \cdots \times G_{i-1} \times \{e_{G_i}\} \times G_{i+1} \times \cdots \times G_n$$

3.) *Given any integer $1 \leq i \leq n$, the group G_i is isomorphic to the quotient of $G_1 \times \cdots \times G_n$ by its normal subgroup $G_1 \times \cdots \times G_{i-1} \times \{e_{G_i}\} \times G_{i+1} \times \cdots \times G_n$.*

$$G_i \cong (G_1 \times \cdots \times G_n)/(G_1 \times \cdots \times G_{i-1} \times \{e_{G_i}\} \times G_{i+1} \times \cdots \times G_n)$$

4.) *Given any integer $1 \leq i \leq n$, the external direct product $G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$ is isomorphic to the quotient of $G_1 \times \cdots \times G_n$ by its normal subgroup $\gamma_i(G_i)$.*

$$G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n \cong (G_1 \times \cdots \times G_i \times \cdots \times G_n)/\gamma_i(G_i)$$

Proof. (1.) Certainly, γ_i is a group homomorphism because the identities $e_{G_j} = e_{G_j} e_{G_j}$ yield that

$$\gamma_i(gh) = (e_{G_1}, \dots, gh, \dots, e_{G_n}) = (e_{G_1}, \dots, g, \dots, e_{G_n})(e_{G_1}, \dots, h, \dots, e_{G_n}) = \gamma_i(g)\gamma_i(h).$$

Even more, we have that $g \in \ker \gamma_i$ if and only if $(e_{G_1}, \dots, g, \dots, e_{G_n}) = (e_{G_1}, \dots, e_{G_i}, \dots, e_{G_n})$ if and only if $g = e_{G_i}$ so that $\ker \gamma_i = \{e_{G_i}\}$ and γ_i is injective by Proposition 3.3.6. Consequently, by the First Isomorphism Theorem, there exists a group isomorphism $G_i \cong \gamma_i(G_i)$.

(2.) Considering that $e_{G_i} = e_{G_i} e_{G_i}$, once again, we find that ε_i is a group homomorphism because the n-tuple $(g_1 h_1, \dots, g_{i-1} h_{i-1}, e_{G_i}, g_{i+1} h_{i+1}, \dots, g_n h_n)$ and the n-tuple

$$(g_1, \dots, g_{i-1}, e_{G_i}, g_{i+1}, \dots, g_n)(h_1, \dots, h_{i-1}, e_{G_i}, h_{i+1}, \dots, h_n)$$

are equal. By construction, the group homomorphism ε_i is injective: if $(g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$ lies in $\ker \varepsilon_i$, then we must have that $(g_1, \dots, g_{i-1}, e_{G_i}, g_{i+1}, \dots, g_n) = (e_{G_1}, \dots, e_{G_{i-1}}, e_{G_i}, e_{G_{i+1}}, \dots, e_{G_n})$, hence the component g_j must be the identity element of G_j for each integer $1 \leq j \leq n$ other than i. We conclude the desired result by the First Isomorphism Theorem.

(3.) Consider the function $\pi_i : G_1 \times \cdots \times G_n \to G_i$ defined by $\pi_i(g_1, \dots, g_n) = g_i$. We may view π_i as "forgetting about" all other components of an n-tuple of $G_1 \times \cdots \times G_n$ other than the ith component. Exercise 3.18.47(a.) shows that π_i is a group homomorphism with kernel

$$G_1 \times \cdots \times G_{i-1} \times \{e_{G_i}\} \times G_{i+1} \times \cdots \times G_n = \{(g_1, \dots, g_{i-1}, e_{G_i}, g_{i+1}, \dots, g_n) \mid g_j \in G_j \text{ for } j \neq i\}.$$

By Exercise 3.18.19, it follows that the above displayed subgroup of $G_1 \times \cdots \times G_n$ is normal. Even more, the First Isomorphism Theorem yields the desired isomorphism.

(4.) Consider the function $\delta_i : G_1 \times \cdots \times G_n \to G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$ defined by $\delta_i(g_1, \dots, g_n) = (g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$. We may view δ_i as "forgetting about" the ith component of an n-tuple in $G_1 \times \cdots \times G_n$. Exercise 3.18.47(b.) shows that δ_i is a group homomorphism with kernel $\gamma_i(G_i) = \{(e_{G_1}, \dots, g, \dots, e_{G_n}) \mid g \in G_i\}$, hence $\gamma_i(G_i)$ is a normal subgroup of $G_1 \times \cdots \times G_n$ by Exercise 3.18.19. Last, the desired isomorphism exists by the First Isomorphism Theorem. □

Example 3.9.10. By Proposition 3.9.9, we may view $(\mathbb{Z}, +)$ as a subgroup of $(\mathbb{Z} \times \mathbb{Z}, +)$ by simply "forgetting about" either one of the factors of $\mathbb{Z}$. Explicitly, every integer a can be viewed as the ordered pair $(a, 0)$. Conversely, we may view $(\mathbb{Z}, +)$ as the quotient group $(\mathbb{Z} \times \mathbb{Z}, +)/(\{0\} \times \mathbb{Z}, +)$, in which case an integer a will be identified with the left coset $(a, 0) + \{0\} \times \mathbb{Z}$.

Usually, the external direct product is constructed in order to define a "larger" group from some known "smaller" groups. Conversely, the **internal direct product** can be defined to express a "large" group as a product of two "smaller" groups. Explicitly, we say that a group G can be expressed as an internal direct product of some subgroups H and K of G if and only if

1.) we have that $H \cap K = \{e_G\}$, i.e., the intersection of H and K is trivial;

2.) we have that $G = HK = \{hk \mid h \in H \text{ and } k \in K\}$, i.e., every element of G can be written (not necessarily uniquely) as a product of an element of H and an element of K; and

3.) we have that $hk = kh$ for all elements $h \in H$ and $k \in K$, i.e., H and K commute.

Caution: the third requirement that every element of H commutes with every element of K does not imply that every internal direct product is abelian. Explicitly, if H is not abelian, then HK cannot be abelian because H is a subgroup of HK, and subgroups of abelian groups are abelian.

Example 3.9.11. Consider the dihedral group D_6 of order twelve with subgroups $H = \{1, r^3\}$ and $K = \{1, r^2, r^4, s, r^2s, r^4s\}$. We leave it as Exercise 3.18.52 to demonstrate that $D_6 = HK$.

We provide next a sufficient condition to conclude that a group is an internal direct product. Curiously, we make use of the external direct product of a group on our way to our objective.

Proposition 3.9.12. *Let G be a group with normal subgroups N and M. If $N \cap M = \{e_G\}$, then $NM = MN$. Put another way, every pair normal subgroups that intersect trivially commute.*

Proof. Given any elements $n \in N$ and $m \in M$, consider the element $n^{-1}m^{-1}nm$ of G. We must establish that $nm = mn$ or $n^{-1}m^{-1}nm = e_G$. By hypothesis that N is normal in G, it follows that $m^{-1}nm \in N$ so that $n^{-1}m^{-1}nm \in N$. Likewise, by hypothesis that M is normal in G, we have that $n^{-1}m^{-1}n \in M$ so that $n^{-1}m^{-1}nm \in M$. Consequently, we have that $n^{-1}m^{-1}nm = e_G$. □

Proposition 3.9.13. *Let G be a group with normal subgroups $N_1, \dots, N_k$. If every element of G can be written uniquely as $n_1 \cdots n_k$ for some elements $n_i \in N_i$, then we have that $G \cong N_1 \times \cdots \times N_k$.*

Proof. By assumption that every element of G can be written uniquely as $n_1 \cdots n_k$ for some elements $n_i \in N_i$, it follows that the function $\varphi : N_1 \times \cdots \times N_k \to G$ defined by $\varphi(n_1, \dots, n_k) = n_1 \cdots n_k$ is a bijection. We must establish that φ is a group homomorphism. Explicitly, we claim that

$$n_1n_1' \cdots n_kn_k' = \varphi((n_1, \dots, n_k)(n_1', \dots, n_k')) = \varphi(n_1, \dots, n_k)\varphi(n_1', \dots, n_k') = n_1 \cdots n_kn_1' \cdots n_k'.$$

By Proposition 3.9.12, it suffices to show that $N_i \cap N_j = \{e_G\}$ for all pairs of integers $1 \leq i < j \leq k$. Consider an element $n \in N_i \cap N_j$. Every element of G can be written uniquely as $n_1 \cdots n_k$ for some elements $n_i \in N_i$, hence the unique expression for n according to this is as the product of $i - 1$ copies of e_G followed by n followed by $k - i$ copies of e_G because n is an element of N_i. By the same token, we may also express n as the product of $j - 1$ copies of e_G followed by n followed by $k - j$ copies of e_G because n is an element of N_j. We conclude therefore that $n = e_G$, as desired. □

Corollary 3.9.14. *Let G be a group with normal subgroups $N_1, \dots, N_k$ such that $N_i \cap N_j = \{e_G\}$ for each integer $1 \leq i < j \leq k$. If $G = N_1 \cdots N_k$, then $G \cong N_1 \times \cdots \times N_k$. Put another way, the internal direct product of groups is isomorphic to the external direct product of groups.*

Proof. We proceed by the Principle of Ordinary Induction on k. We will assume first that $k = 2$. By Proposition 3.9.13, it suffices to show that every element of G can be written uniquely as n_1n_2 for some elements $n_1 \in N_1$ and $n_2 \in N_2$. Consider the case that $n_1n_2 = m_1m_2$ for some elements $m_1 \in N_1$ and $m_2 \in N_2$. We note that $m_1^{-1}n_1 = m_2n_2^{-1}$ with $m_1^{-1}n_1 \in N_1$ and $m_2n_2^{-1} \in N_2$. Consequently, the products $m_1^{-1}n_1$ and $m_2n_2^{-1}$ both lie in $N_1 \cap N_2$, hence they must both be equal to e_G, and we conclude that $m_1 = n_1$ and $m_2 = n_2$. We have successfully established the base case, hence we may assume inductively that the result holds for some integer $k \geq 3$. By Exercise 3.18.11, it follows that $N_1 \cdots N_k$ is a normal subgroup of G with the property that $N_1 \cdots N_k \cap N_{k+1} = \{e_G\}$, hence we conclude from the base case that $G \cong (N_1 \cdots N_k) \times N_{k+1} \cong N_1 \times \cdots \times N_k \times N_{k+1}$. □

3.10 Finite Abelian Groups

Earlier in this chapter, we proved that every infinite cyclic group is isomorphic to $(\mathbb{Z}, +)$ and that every finite cyclic group is isomorphic to $(\mathbb{Z}/n\mathbb{Z}, +)$ for some positive integer n. We turn our attention in this section to the less restrictive property of abelianness. Like before, we will provide a characterization of finite abelian groups in terms of isomorphism classes with respect to order.

Consider any positive integer n. By the Fundamental Theorem of Arithmetic, there exist prime numbers $p_1, \dots, p_k$ and unique non-negative integers $e_1, \dots, e_k$ such that $n = p_1^{e_1} \cdots p_k^{e_k}$ and

$$\left|\frac{\mathbb{Z}}{n\mathbb{Z}}\right| = n = p_1^{e_1} \cdots p_k^{e_k} = \left|\frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}}\right| \cdots \left|\frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}\right| = \left|\frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}\right|.$$

Consequently, by Exercise 1.10.5, there exists a bijective function $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \mathbb{Z}/p_k^{e_k}\mathbb{Z}$. One might naturally ask if (and we would certainly hope that) this bijective function carries the additional structure of a group homomorphism. We answer this in the affirmative as follows.

Proposition 3.10.1. *Given any positive integer $n \geq 2$ with prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$ for some prime numbers $p_1, \dots, p_k$ and unique non-negative integers $e_1, \dots, e_k$, we have that*

$$\frac{\mathbb{Z}}{n\mathbb{Z}} \cong \frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}.$$

Proof. Observe that each direct factor of the external direct product $(\mathbb{Z}/p_1^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})$ is a finite cyclic group with $|\mathbb{Z}/p_i^{e_i}\mathbb{Z}| = p_i^{e_i}$ for each integer $1 \leq i \leq k$. Consequently, we have that

$$\gcd\left(\left|\frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}}\right|, \dots, \left|\frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}\right|\right) = \gcd(p_1^{e_1}, \dots, p_k^{e_k}) = 1$$

by assumption that $p_1, \dots, p_k$ are distinct prime numbers. By Proposition 3.9.8, we conclude that $(\mathbb{Z}/p_1^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})$ is cyclic of order $n = p_1^{e_1} \cdots p_k^{e_k}$; the result holds by Proposition 3.3.20. □

Consequently, every finite cyclic group of order n can be written as the external product of cyclic groups whose orders correspond to the unique prime-power factors of the unique prime factorization of n. Conversely, it is only possible to achieve such a representation as an external direct product of finite cyclic groups by factors of n that are pairwise relatively prime to one another.

Proposition 3.10.2. *Given any positive integers $n_1, \dots, n_k$, we have that*

$$\frac{\mathbb{Z}}{n_1 \cdots n_k\mathbb{Z}} \cong \frac{\mathbb{Z}}{n_1\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{n_k\mathbb{Z}}$$

if and only if for all indices $1 \leq i < j \leq k$, we have that $\gcd(n_i, n_j) = 1$.

Proof. Each of the quotient groups $\mathbb{Z}/n_i\mathbb{Z}$ is a cyclic group of order n_i for each integer $1 \leq i \leq k$. By Proposition 3.9.8, we have that $(\mathbb{Z}/n_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})$ is cyclic if and only if $\gcd(n_1, \dots, n_k) = 1$. Consequently, if we assume that $\gcd(n_i, n_j) = 1$ for all indices $1 \leq i < j \leq k$, then we have that $\gcd(n_1, \dots, n_k) = 1$ so that $(\mathbb{Z}/n_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})$ is a cyclic group of order $n_1 \cdots n_k$; the displayed isomorphism is therefore guaranteed by Proposition 3.3.20. Conversely, if we assume to the contrapositive that $\gcd(n_1, n_2) = d \geq 2$ (after possibly relabelling the indices), then the integer $n_1 \cdots n_k/d$ is well-defined because d divides both n_1 and n_2; it is strictly smaller than $n_1 \cdots n_k$. On the contrary, if $(\mathbb{Z}/n_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})$ were cyclic and generated by $(a_1 + n_1\mathbb{Z}, \dots, a_k + n_k\mathbb{Z})$, then the sum of $(a_1 + n_1\mathbb{Z}, \dots, a_k + n_k\mathbb{Z})$ with itself $n_1 \cdots n_k/d$ times would be equal to

$$\left(\frac{n_2 \cdots n_k}{d} a_1 n_1 + n_1\mathbb{Z}, \dots, \frac{n_1 \cdots n_{k-1}}{d} a_k n_k + n_k\mathbb{Z}\right) = (0 + n_1\mathbb{Z}, \dots, 0 + n_k\mathbb{Z});$$

this is impossible because the order of $(a_1 + n_1\mathbb{Z}, \dots, a_k + n_k\mathbb{Z})$ must be $n_1 \cdots n_k$. □

Example 3.10.3. We have that $\mathbb{Z}/12\mathbb{Z} \cong (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$ because the prime factorization of 12 is $12 = 2^2 \cdot 3 = 4 \cdot 3$; however, $\mathbb{Z}/12\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$ are not isomorphic because $\gcd(2, 6) = 2$.

Example 3.10.4. We have that $\mathbb{Z}/30\mathbb{Z} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/15\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/5\mathbb{Z})$ because it holds that $30 = 2 \cdot 15 = 2 \cdot 3 \cdot 5$ and $\gcd(2, 15) = 1$ and $\gcd(3, 5) = 1$.

Even if we do not restrict our attention to finite cyclic groups, every finite abelian group can be written as the direct product of finitely generated cyclic groups. Later, we will develop the tools to prove the following theorem; however, we will simply state it in two versions for now.

Theorem 3.10.5 (Fundamental Theorem of Finite Abelian Groups)**.** *Given any finite abelian group G, there exist (not necessarily distinct) primes $p_1, \dots, p_k$ and integers $e_1, \dots, e_k \geq 0$ such that*

$$G \cong \frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}.$$

We refer to the prime powers $p_i^{e_i}$ as the **elementary divisors** *of G; this representation of G as a direct product of finite cyclic groups is called the* **elementary divisor decomposition** *of G.*

Corollary 3.10.6. *Given any finite abelian group G, there exist (not necessarily distinct) positive integers $n_1, \dots, n_\ell$ such that $n_i \mid n_{i+1}$ for each integer $1 \leq i \leq \ell - 1$ and*

$$G \cong \frac{\mathbb{Z}}{n_1\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{n_\ell\mathbb{Z}}.$$

We refer to the positive integers n_i as the **invariant factors** *of G; this representation of G as a direct product of finite cyclic groups is called the* **invariant factor decomposition** *of G.*

Out of desire for convenience (and because they are isomorphic groups by Proposition 3.3.20), we will throughout this section freely alternate between the notation of $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}_n$.

Example 3.10.7. Given any positive integer n, the elementary divisors of the cyclic group $\mathbb{Z}/n\mathbb{Z}$ are obtained via the prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$. Explicitly, the elementary divisors are $p_1^{e_1}, \dots, p_k^{e_k}$ by Proposition 3.10.1. On the other hand, the only invariant factor of $\mathbb{Z}/n\mathbb{Z}$ is n.

Example 3.10.8. Observe that for any finite abelian group

$$G \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_8 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_{25},$$

the elementary divisors are 2, 2, 2^2, 2^3, 3, 3, 5, and 5^2; however, the invariant factors are unclear.

Algorithm 3.10.9 (Converting Elementary Divisors to Invariant Factors)**.** Let G be a finite abelian group whose elementary divisors are known. Use the following to find the invariant factors of G.

1.) Find the prime number p that appears the most times in the elementary divisor decomposition of G. Choose one arbitrarily if two or more primes appear an equal number of times.

2.) Create an array whose first row consists of all powers of p that appear in the elementary divisor decomposition of G, listing these powers in non-decreasing order from left to right.

3.) Repeat the second step in the second row with the prime q that appears the second most times (or the same number of times as p) in the elementary divisor decomposition of G.

4.) Continue this process until all primes appearing in the elementary divisor decomposition of G have been written in a row. One should end with an upper-triangular array.

5.) By multiplying the elements of each consecutive column, we obtain the invariant factors of G.

Example 3.10.10. Consider the finite abelian group G of Example 3.10.8. By following the method outlined in Algorithm 3.10.9 with the group G at hand, we have the following array.

$$\begin{array}{cccc} 2 & 2 & 2^2 & 2^3 \\ & 3 & 3 & 3 \\ & & 5 & 5^2 \end{array}$$

By multiplying the elements of each consecutive column, we obtain the invariant factors of G: they are 2, 6, 60, and 600. By Corollary 3.10.6, we conclude that $G \cong \mathbb{Z}_2 \times \mathbb{Z}_6 \times \mathbb{Z}_{60} \times \mathbb{Z}_{600}$.

Conversely, one can obtain the elementary divisors from the invariant factors.

Algorithm 3.10.11 (Converting Invariant Factors to Elementary Divisors)**.** Let G be a finite abelian group whose invariant factors are known. Use the following to find the elementary divisors.

1.) Given the invariant factors n_i of G with $n_1 \mid n_2 \mid n_3 \mid \cdots \mid n_\ell$, express each invariant factor n_i as a product of distinct prime powers by the Fundamental Theorem of Arithmetic.

2.) Construct an upper-triangular array whose ith column consists of the distinct prime powers $p_{i1}^{e_{i1}}, \dots, p_{ik}^{e_{ik}}$ such that $n_i = p_{i1}^{e_{i1}} \cdots p_{ik}^{e_{ik}}$.

3.) We obtain the elementary divisors of G as the components of the upper-triangular array.

Example 3.10.12. Consider any finite abelian group $G \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{14} \times \mathbb{Z}_{98} \times \mathbb{Z}_{294}$ with invariant factors 2, 2, $14 = 2 \cdot 7$, $98 = 2 \cdot 7^2$, and $294 = 2 \cdot 3 \cdot 7^2$. By following the procedure outlined in Algorithm 3.10.11 with the group G at hand, we obtain the following array.

$$\begin{array}{ccccc} 2 & 2 & 2 & 2 & 2 \\ & & 7 & 7^2 & 3 \\ & & & & 7^2 \end{array}$$

We find that the elementary divisors of G are 2, 2, 2, 2, 2, 3, 7, 7^2, and 7^2. By the Fundamental Theorem of Finite Abelian Groups, we conclude that $G \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_7 \times \mathbb{Z}_{49} \times \mathbb{Z}_{49}$.

Ultimately, the Fundamental Theorem of Finite Abelian Groups implies that the structure of a finite abelian group G is uniquely determined (up to isomorphism) by its elementary divisors (or equivalently its invariant factors). Even more, we note that the elementary divisors of G are (not necessarily uniquely) determined by the unique prime factorization of $|G|$. By Exercise 2.8.14, every group of order four is abelian, hence it suffices to note that there are two distinct (up to isomorphism) groups of order four: they are the Klein four-group $\mathbb{Z}_2 \times \mathbb{Z}_2$ and the cyclic group $\mathbb{Z}_4$ of order four. Each of these groups corresponds to a distinct **integer partition** of the integer two. Generally, we define a partition of an integer n as a k-tuple $(n_1, \dots, n_k)$ such that $n = n_1 + \cdots + n_k$ and $1 \leq n_1 \leq n_2 \leq \cdots \leq n_k$. Considering that $2 = 1 + 1$ and $2 = 2$, there are two distinct integer partitions of 2. Consequently, there are two distinct abelian groups of order $4 = 2^2$. We will denote by $\rho(n)$ the number of distinct integer partitions of the positive integer n.

Proposition 3.10.13. *If n is a positive integer with unique prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$, then there are $\rho(e_1) \cdots \rho(e_k)$ distinct finite abelian groups of order n up to isomorphism.*

Proof. By the Fundamental Theorem of Finite Abelian Groups, for any finite abelian group G, there exist (not necessarily distinct) prime numbers q_i and non-negative integers f_i such that

$$G \cong \frac{\mathbb{Z}}{q_1^{f_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{q_\ell^{f_\ell}\mathbb{Z}}.$$

Observe that if the order of G is n, then by the uniqueness of the prime factorization of n, we must have that $\{p_1, \dots, p_k\} = \{q_1, \dots, q_\ell\}$ and $e_1 + \cdots + e_k = f_1 + \cdots + f_\ell$. Considering that there are $\rho(e_1) \cdots \rho(e_k)$ solutions $(f_1, \dots, f_k)$ to the positive integer identity $e_1 + \cdots + e_k = f_1 + \cdots + f_\ell$, we conclude that there are at most $\rho(e_1) \cdots \rho(e_k)$ distinct finite abelian groups of order n up to isomorphism. Conversely, each of the finite abelian groups obtained as the external direct product corresponding to some integer partition of e_i is distinct by Proposition 3.10.2 because the cyclic groups $\mathbb{Z}_{p_i^{e_i}}$ and $\mathbb{Z}_{p_i^{e_i - 1}} \times \mathbb{Z}_{p_i}$ are not isomorphic to one another whenever $e_i \geq 2$. □

Example 3.10.14. Consider a finite abelian group G of order $14553000 = 2^3 \cdot 3^3 \cdot 5^3 \cdot 7^2 \cdot 11$. Observe that $3 = 3$, $3 = 1 + 2$, and $3 = 1 + 1 + 1$ are the three distinct integer partitions of three. Even more, there are only two integer partitions $2 = 2$ and $2 = 1 + 1$ of two. Consequently, by Proposition 3.10.13, there are $\rho(3) \cdot \rho(3) \cdot \rho(3) \cdot \rho(2) = \rho(3)^3 \cdot \rho(2) = 3^3 \cdot 2 = 54$ abelian groups of order 14553000.

We have for now stated the Fundamental Theorem of Finite Abelian Groups without proof, but the tools that are developed for the proof are quite useful to understand on their own. Given any abelian group G and any prime number p, consider the set of all elements of G of p-power order

$$G(p) = \{g \in G \mid \operatorname{ord}(g) = p^n \text{ for some integer } n \geq 0\}.$$

Proposition 3.10.15. *If G is an abelian group and p is a prime number, then the set $G(p)$ of all elements $g \in G$ such that $\operatorname{ord}(g) = p^n$ for some integer $n \geq 0$ is subgroup of G.*

Proof. Considering that $\operatorname{ord}(e_G) = 1 = p^0$ for any prime number p, it follows that $G(p)$ is nonempty. By the One-Step Subgroup Test, it suffices to prove that if $\operatorname{ord}(g) = p^m$ and $\operatorname{ord}(h) = p^n$ for some positive integers m and n, then $\operatorname{ord}(gh^{-1}) = p^\ell$ for some integer $\ell \geq 0$; this is Exercise 2.8.32. □

One can view our next proposition as something of a prime factorization property for elements of finite order in an abelian group. Essentially, we prove that every element of finite order in an abelian group can be written as a product of elements of prime-power order.

Proposition 3.10.16. *If G is an abelian group and $g \in G$ has finite order, then for each distinct prime number p_i that divides $\operatorname{ord}(g)$, there exist elements $g_i \in G(p_i)$ such that $g = g_1 \cdots g_k$.*

Proof. We proceed by induction on the number k of distinct prime factors in the unique prime factorization of $\operatorname{ord}(g)$ (cf. Fundamental Theorem of Arithmetic). If $k = 1$, then $\operatorname{ord}(g) = p^n$ for some prime number p and some non-negative integer n, from which it follows that $g \in G(p)$.

We will assume inductively that the proposition holds for all elements of G whose order is divisible by at most $k - 1$ distinct primes. If $\operatorname{ord}(g) = p_1^{e_1} \cdots p_k^{e_k}$ for some distinct prime numbers p_i and positive integers e_i, we may factor $\operatorname{ord}(g) = ab$ for $a = p_1^{e_1}$ and $b = p_2^{e_2} \cdots p_k^{e_k}$. By hypothesis that the p_i are distinct, we have that $\gcd(a, b) = 1$, hence by Bézout's Identity, we find that $ax+by = 1$ for some integers x and y. Consequently, we have that $g = g^{ax+by} = g^{ax}g^{by}$. Observe that

$$(g^{by})^a = g^{aby} = g^{y \operatorname{ord}(g)} = (g^{\operatorname{ord}(g)})^y = e_G^y = e_G,$$

from which it follows that $\operatorname{ord}(g^{by}) \mid p_1^{e_1}$ so that g^{by} lies in $G(p_1)$. Likewise, we have that

$$(g^{ax})^b = g^{abx} = g^{x \operatorname{ord}(g)} = (g^{\operatorname{ord}(g)})^x = e_G^x = e_G,$$

from which it follows that $\operatorname{ord}(g^{ax}) \mid b$ so that $\operatorname{ord}(g^{ax}) = p_2^{f_2} \cdots p_k^{f_k}$ for some non-negative integers f_i. By induction, we may write $g^{ax} = g_2 \cdots g_k$ for some elements $g_i \in G(p_i)$ for each distinct prime number p_i and $g^{by} = g_1$ for some element $g_1 \in G(p_1)$ so that $g = g^{by+ax} = g^{by}g^{ax} = g_1g_2 \cdots g_k$. □

Last, we prove that every finite abelian group can be decomposed as an external direct product of its subgroups $G(p_i)$ for each prime number that divides the order of G.

Theorem 3.10.17. *If G is any finite abelian group, then $G \cong G(p_1) \times \cdots \times G(p_k)$ for some distinct prime numbers $p_1, \dots, p_k$ that divide the order of G.*

Proof. We claim that if p_i and p_j are distinct prime numbers, then $G(p_i) \cap G(p_j) = \{e_G\}$. By definition, if $g \in G(p_i) \cap G(p_j)$, then there are non-negative integers e_i and e_j with $\operatorname{ord}(g) = p_i^{e_i}$ and $\operatorname{ord}(g) = p_j^{e_j}$. On the contrary, if neither e_i is zero nor e_j is zero, then the identity $p_i^{e_i} = p_j^{e_j}$ yields that p_i divides $p_j^{e_j}$. But by assumption that p_i and p_j are distinct prime numbers, this is impossible. We conclude therefore that either $e_i = 0$ or $e_j = 0$ so that $G(p_i) \cap G(p_j) = \{e_G\}$.

By Proposition 3.9.14, it suffices to show that every element of G can be written uniquely as $g_1 \cdots g_k$ for some elements $g_i \in G(p_i)$, where we define $g_i = e_G$ if the prime number p_i does not divide $\operatorname{ord}(g)$. By Proposition 3.10.16, every element of G can be written as $g_1 \cdots g_k$ for some elements $g_i \in G(p_i)$ for each distinct prime number p_i that divides $\operatorname{ord}(g)$. Consider two representations $g_1 \cdots g_k = h_1 \cdots h_k$ of an element $g \in G$ such that g_i and h_i both lie in $G(p_i)$. By hypothesis that G is abelian, we have that $g_1h_1^{-1} = h_2g_2^{-1} \cdots h_kg_k^{-1}$. Considering that $G(p_i)$ is a subgroup of G for each integer $1 \leq i \leq k$ by Proposition 3.10.15, it follows by the One-Step Subgroup Test that $h_ig_i^{-1}$ lies in $G(p_i)$ for each integer $2 \leq i \leq k$. By definition, for each integer $2 \leq i \leq k$, we have that $\operatorname{ord}(h_ig_i^{-1}) = p_i^{e_i}$ for some integer $e_i \geq 0$. Consider the integer $n = p_2^{e_2} \cdots p_k^{e_k}$. We have that

$$(g_1h_1^{-1})^n = (h_2g_2^{-1} \cdots h_kg_k^{-1})^n = (h_2g_2^{-1})^n \cdots (h_kg_k^{-1})^n = e_G$$

by hypothesis that G is abelian and $\operatorname{ord}(h_ig_i^{-1}) = p_i^{e_i}$ divides n. But this implies that $\operatorname{ord}(g_1h_1^{-1}) \mid n$ by Corollary 2.3.13; on the other hand, $\operatorname{ord}(g_1h_1^{-1}) = p_1^{e_1}$ for some integer $e_1 \geq 0$ gives that $e_1 = 0$ so that $g_1h_1^{-1} = e_G$ or $g_1 = h_1$. Repeating this yields that $g_i = h_i$ for all integers $2 \leq i \leq k$. □

Consequently, it is enough to consider $G(p_i)$ for each prime number p_i dividing $|G|$.

Proposition 3.10.18. *If G is an abelian group of order p^n for some prime number p and integer $n \geq 0$, then G admits a subgroup H such that $G \cong (\mathbb{Z}/p^k\mathbb{Z}) \times H$ and $k = \max\{\operatorname{ord}(g) \mid g \in G\}$.*

Proof. We will prove a slightly different statement by appealing to the Principle of Complete Induction on n. Explicitly, we will prove that for any element $g \in G$ with $\operatorname{ord}(g) = \max\{\operatorname{ord}(x) \mid x \in G\}$, there exists a subgroup H of G such that $G \cong \langle g \rangle \times H$. Once this is established, then we may employ Corollary 3.1.13 and Proposition 3.3.20 to conclude that $G \cong \langle g \rangle \times H \cong (\mathbb{Z}/p^k\mathbb{Z}) \times H$.

If $n = 1$, then G is a finite group of order p, hence we have that $G \cong \mathbb{Z}/p\mathbb{Z} \cong (\mathbb{Z}/p\mathbb{Z}) \times \{e_G\}$. We will assume therefore that the proposition holds for all integers $1 \leq k \leq n-1$. Consider an element $g \in G$ such that $\operatorname{ord}(g) = \max\{\operatorname{ord}(x) \mid x \in G\}$. Observe that if $\operatorname{ord}(g) = p^n$, then G is cyclic, hence we conclude as before that $G \cong \mathbb{Z}/p^n\mathbb{Z} \cong (\mathbb{Z}/p^n\mathbb{Z}) \times \{e_G\}$. Consequently, it suffices to consider the case that $\operatorname{ord}(g) = p^k$ for some integer $1 \leq k \leq n-1$. We note that $G \setminus \langle g \rangle$ is nonempty because the order of g is strictly smaller than the order of G, hence we may find an element $h \in G \setminus \langle g \rangle$ such that $\operatorname{ord}(h) = \min\{\operatorname{ord}(x) \mid x \in G \text{ and } x \notin \langle g \rangle\}$ by the Well-Ordering Principle.

Observe that $\operatorname{ord}(h) = p^\ell$ for some integer $\ell \geq 1$ by Lagrange's Theorem. By Proposition 2.3.14, we have that $\operatorname{ord}(h^p) = p^\ell / \gcd(p, p^\ell) = p^\ell/p = p^{\ell-1}$, hence we must have that $h^p \in \langle g \rangle$ by assumption that $\operatorname{ord}(h) = \min\{\operatorname{ord}(x) \mid x \notin \langle g \rangle\}$. By definition of $\langle g \rangle$, there exists a positive integer q such that $h^p = g^q$. By raising each of these identical elements to the power of p^{k-1}, we find that $e_G = h^{p^k} = (h^p)^{p^{k-1}} = (g^q)^{p^{k-1}}$ and $\operatorname{ord}(g^q) \leq p^{k-1}$. Corollary 2.3.15 yields that $\gcd(p^{k-1}, q)$ is nonzero, hence p must divide q, i.e., there exists an integer r such that $q = pr$ and $h^p = g^q = g^{pr}$. Consider the product $x = g^{-r}h$. Observe that $x^p = g^{-pr}h^p = g^{-q}h^p = (g^q)^{-1}h^p = (h^p)^{-1}h^p = e_G$, hence $\operatorname{ord}(x)$ must be either one or p; however, if it were the case that $\operatorname{ord}(x) = 1$, then we would have that $e_G = x = g^{-r}h$ so that $h = g^r$ lies in $\langle g \rangle$ — contradicting the definition of h. We conclude therefore that $\operatorname{ord}(x) = p$. Once again, if x were an element of $\langle g \rangle$, then $h = g^r x$ would be an element of $\langle g \rangle$, hence x must be an element of order p that is not contained in $\langle g \rangle$. By construction of h as an element of $G \setminus \langle g \rangle$ of smallest possible order, we conclude that $\operatorname{ord}(h) = p$.

We claim next that $\langle g \rangle \cap \langle h \rangle = \{e_G\}$. Given any element $x \in \langle g \rangle \cap \langle h \rangle$, we have that $\operatorname{ord}(x) \mid p$ by Lagrange's Theorem so that $\operatorname{ord}(x) = 1$ or $\operatorname{ord}(x) = p$. On the contrary, if it were the case that $\operatorname{ord}(x) = p$, then it would follow that $\langle x \rangle = \langle h \rangle$ because they both generate a subgroup of G of order p and $\langle x \rangle \subseteq \langle h \rangle$ by assumption. But at the same time, because x lies in $\langle g \rangle$, this would imply that $h = x^a = (g^b)^a = g^{ab}$ for some positive integers a and b — contradicting the construction of h as an element of $G \setminus \langle g \rangle$. We conclude therefore that $\operatorname{ord}(x) = 1$ so that $\langle g \rangle \cap \langle h \rangle = \{e_G\}$.

By Lagrange's Theorem, the order of the quotient group $G/\langle h \rangle$ is $|G| / \operatorname{ord}(h) = p^n/p = p^{n-1}$. We claim that $\operatorname{ord}(g\langle h \rangle) = \operatorname{ord}(g) = p^k$, from which it will follow by Exercise 3.18.7 that $g\langle h \rangle$ is an element of $G/\langle h \rangle$ of maximum possible order. Observe that for any integer $1 \leq \ell \leq p^k$ such that $e_G\langle h \rangle = (g\langle h \rangle)^\ell = g^\ell \langle h \rangle$, we must have that $g^\ell \in \langle g \rangle \cap \langle h \rangle = \{e_G\}$. Consequently, the only possibility for ℓ is $\operatorname{ord}(g) = p^k$, and we conclude that $\operatorname{ord}(g\langle h \rangle) = p^k$. By our inductive hypothesis applies to $G/\langle h \rangle$ and its subgroup $g\langle h \rangle$ and the Fourth Isomorphism Theorem, there exists a subgroup H of G such that $G/\langle h \rangle \cong \langle g\langle h \rangle \rangle \times H/\langle h \rangle$. We claim at last that $G \cong \langle g \rangle \times H$. Given any element $x \in \langle g \rangle \cap H$, we have that $x\langle h \rangle \in \langle g\langle h \rangle \rangle \cap H/\langle h \rangle = \{e_G\langle h \rangle\}$ by the Fourth Isomorphism Theorem. We conclude therefore that $x \in \langle g \rangle \cap \langle h \rangle = \{e_G\}$, as desired. Considering that $p^{n-1} = p^k|H|/p$, we conclude that $|H| = p^{n-k}$ so that $G = \langle g \rangle H$ and $G \cong \langle g \rangle \times H$ by Corollary 3.9.14. □

3.11 Finitely Generated Groups

Back in Section 2.3, we demonstrated that cyclic groups are in some sense the "simplest possible" groups to consider. Considering this interpretation, we sought to understand cyclic groups as much as possible by studying the structure of the subgroups of a cyclic group. By Propositions 3.3.18 and 3.3.20, we were easily able to completely solve the classification problem for cyclic groups.

Given any group G and any nonempty subset X of G, consider the collection

$$\langle X \rangle = \{g_1^{n_1} \cdots g_k^{n_k} \mid k \text{ is a positive integer, } n_1, \dots, n_k \text{ are integers, and } g_i \in X\}$$

of all possible integral powers of all possible products of elements of X. We note that if X is the singleton $X = \{g\}$ for any element $g \in G$, then the set $\langle X \rangle$ is simply the cyclic subgroup generated by g. Consequently, this collection of elements of G is a generalization of the cyclic subgroup G.

Proposition 3.11.1. *If X is a nonempty subset of a group G, then $\langle X \rangle$ is a subgroup of G.*

Proof. By assumption that X is nonempty, we have that $e_G = g_i^0$ lies in $\langle X \rangle$ for any element $g_i \in X$. By the One-Step Subgroup Test, it suffices to prove that if g and h are elements of $\langle X \rangle$, then gh^{-1} is in $\langle X \rangle$. By definition of $\langle X \rangle$, for any elements $g, h \in \langle X \rangle$, there exist positive integers k and ℓ, integers $m_1, \dots, m_k, n_1, \dots, n_\ell$, and elements $g_1, \dots, g_k, h_1, \dots, h_\ell \in X$ such that $g = g_1^{m_1} \cdots g_k^{m_k}$ and $h = h_1^{n_1} \cdots h_\ell^{n_\ell}$. Consequently, we have that $gh^{-1} = g_1^{m_1} \cdots g_k^{m_k} h_\ell^{-n_\ell} \cdots h_1^{-n_1} \in \langle X \rangle$. □

We refer to $\langle X \rangle$ of G as the subgroup of G **generated** by X; the elements of X are called the **generators** of $\langle X \rangle$. If $|X|$ is finite, we say that $\langle X \rangle$ is **finitely generated**. Even more, if G admits some subset X such that $G = \langle X \rangle$, then we say that G is a **finitely generated group**.

Example 3.11.2. Every cyclic group is finitely generated. Explicitly, if G is a cyclic group generated by some element $g \in G$, then we have that $G = \langle X \rangle$ for the finite set $X = \{g\}$.

Example 3.11.3. Every finite group G is finitely generated by $X = \{g \mid g \in G\} = G$. On the other hand, it is possible in some cases to find a proper system of generators for a finite group. Explicitly, by Exercise 3.18.64, the symmetric group $\mathfrak{S}_3$ on three letters is finitely generated by (12) and (123).

Example 3.11.4. By Exercise 3.18.26, the group $(\mathbb{Z} \times \mathbb{Z}, +)$ cannot be cyclic; however, it is finitely generated. Explicitly, every element of $\mathbb{Z} \times \mathbb{Z}$ is of the form $(a, b) = (a, 0) + (0, b) = a(1, 0) + b(0, 1)$ for some integers a and b. Consequently, we have that $(\mathbb{Z} \times \mathbb{Z}, +) = \langle (1, 0), (0, 1) \rangle$. Generally, it is true that the direct product $(\mathbb{Z} \times \cdots \times \mathbb{Z}, +)$ of n copies of $\mathbb{Z}$ is finitely generated by the n-tuples $E_1, \dots, E_n$ such that the ith column of E_i is one and all other columns are zero.

Proposition 3.11.5. *Given any nonempty subset X of any group G, consider the collection $\mathscr{C} = \{H \leq G \mid X \subseteq H\}$ of subgroups of G that contain X. We have that $\langle X \rangle = \bigcap_{H \in \mathscr{C}} H$. Put another way, we have that $\langle X \rangle$ is the smallest (with respect to inclusion) subgroup of G that contains X.*

Proof. Consider a subgroup H of G such that $X \subseteq H$. Given any element $g_1^{n_1} \cdots g_k^{n_k}$ of $\langle X \rangle$, it follows that $g_1^{n_1} \cdots g_k^{n_k}$ is in H by hypothesis that H is a subgroup of G that contains X. Certainly, this argument holds for all subgroups H of G with $X \subseteq H$, hence we have that $\langle X \rangle \subseteq \bigcap_{H \in \mathscr{C}} H$.

Conversely, observe that $\langle X \rangle$ is a subgroup of G that contains X. Explicitly, by Proposition 3.11.1, we have that $\langle X \rangle$ is a subgroup of G, and every element $g \in X$ can be written as itself to the power of one. Consequently, we have that $\bigcap_{H \in \mathscr{C}} H \subseteq \langle X \rangle$ because $\langle X \rangle$ lies in $\mathscr{C}$. □

Given a finitely generated group G with set of generators X, we refer to a **relation** among the generators of G as an equation involving the elements of $X \cup \{e_G\}$. We are already familiar with some relations on G. Given any element $g \in G$ of finite order, we obtain the relation $g^{\operatorname{ord}(g)} = e_G$. Given any element $g \in Z(G)$ (i.e., in the center of G) and any element $h \in G$, we obtain the relation $gh = hg$ or $g^{-1}h^{-1}gh = e_G$. Further, if we assume that every relation among the generators of G can be deduced from the finitely many relations $\mathscr{R}_1, \dots, \mathscr{R}_n$ of the elements of $X \cup \{e_G\}$, then we refer to the object $G = \langle X \mid \mathscr{R}_1, \dots, \mathscr{R}_n \rangle$ as a finite **presentation** of the group G.

Example 3.11.6. Consider the group G presented by

$$G = \langle r, s \mid r^3 = 1 \text{ and } s^2 = 1 \text{ and } (sr)^2 = 1 \rangle.$$

By definition of this presentation, every element of G is of the form $r^i s^j$ for some integers i and j because G is generated by $X = \{r, s\}$. By hypothesis, we have that $\operatorname{ord}(r) = 3$, hence every element of G is of the form s^j, rs^j, or $r^2 s^j$ for some integer j. Likewise, we have that $\operatorname{ord}(s) = 2$; it follows that $1, s, r, rs, r^2$, and $r^2 s$ are all possible elements of G. One can verify that $G \cong D_3$.

Example 3.11.7. Certainly, the number of relations in a group can be zero: this is equivalent to the condition that the set of relations is empty, and we omit this part of the presentation. Explicitly, this means that for every element $g \in G$, we have that g^m and g^n are distinct if the integers m and n are distinct. By Proposition 3.3.18, the function $\varphi : G \to \mathbb{Z}$ defined by $\varphi(g^n) = n$ is a group isomorphism, hence up to isomorphism, the unique group with this presentation is $(\mathbb{Z}, +)$.

Example 3.11.8. We will construct a group presentation for $(\mathbb{Z} \times \mathbb{Z}, +)$. By Example 3.11.4, this group is finitely generated by $g = (1, 0)$ and $h = (0, 1)$; neither of the generators has finite order because $g^n = 0$ if and only if $n(1, 0) = (n, 0) = (0, 0)$ if and only if $n = 0$, and the same holds for the element $h = (0, 1)$. Even more, we have that $gh = (1, 0) + (0, 1) = (1, 1) = (0, 1) + (1, 0) = hg$ has infinite order; however, this yields the relation $g^{-1}h^{-1}gh = 1$ that signifies commutativity of the generators. Consequently, we find that $\mathbb{Z} \times \mathbb{Z} \cong \langle g, h \mid g^{-1}h^{-1}gh = 1 \rangle$.

3.12 Finitely Generated Abelian Groups

By the Fundamental Theorem of Finite Abelian Groups, every finite abelian group can be written as a direct product of (not necessarily distinct) cyclic groups of prime-power order. Our aim throughout this section is to produce an analogous statement for **finitely generated abelian groups**.

We will henceforth adopt the terminology that the direct product of r copies of $(\mathbb{Z}, +)$ is the **free abelian group of rank** r. Explicitly, we will denote this by $(\mathbb{Z}^r, +) = (\mathbb{Z} \times \cdots \times \mathbb{Z}, +)$ with exactly r direct factors. Conventionally, we will assume that $\mathbb{Z}^0 = \{0\}$. Every element of $(\mathbb{Z}^r, +)$ can be written uniquely as $(a_1, a_2, \dots, a_r) = a_1(1, 0, \dots, 0) + a_2(0, 1, \dots, 0) + \cdots + a_r(0, 0, \dots, 1)$. Consequently, if we denote by E_i the r-tuple whose ith column is one and whose other columns are zero, then every element of $(\mathbb{Z}^r, +)$ can be written uniquely as $a_1E_1 + a_2E_2 + \cdots + a_rE_r$ for some integers $a_1, a_2, \dots, a_r$. We conclude that $(\mathbb{Z}^r, +)$ is finitely generated by the r-tuples $E_1, E_2, \dots, E_r$.

Theorem 3.12.1 (Fundamental Theorem of Finitely Generated Abelian Groups)**.** *Every finitely generated abelian group G can be written uniquely as $G \cong \mathbb{Z}^r \times \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_\ell}$ for some nonnegative integers r and ℓ and some integers $n_1, \dots, n_\ell \geq 2$ with the property that $n_1 \mid n_2 \mid \cdots \mid n_\ell$.*

Like before, we refer to the positive integers $n_1 \mid n_2 \mid \cdots \mid n_\ell$ that appear in the above decomposition of a finitely generated abelian group G as the **invariant factors** of G; however, for a finitely generated abelian group G, it is possible that there is a nontrivial direct factor of a free abelian group $\mathbb{Z}^r$ of rank r in the invariant factor decomposition of G. We refer to this non-negative integer r as the **free rank** of the finitely generated abelian group G. Ultimately, we will find that the Fundamental Theorem of Finitely Generated Abelian Groups is a consequence of a more general and powerful fact known as the Fundamental Theorem of Finitely Generated Modules over a Principal Ideal Domain, but for now, we will simply continue our discussion without proof.

Example 3.12.2. By the Fundamental Theorem of Finite Abelian Groups, every finite abelian group is a finitely generated abelian group with free rank zero.

Example 3.12.3. We have already seen that $(\mathbb{Z}, +)$ is the unique (up to isomorphism) cyclic group of infinite order: indeed, we have that $(\mathbb{Z}, +) = \langle 1 \rangle$; however, we may also view $(\mathbb{Z}, +)$ as a finitely generated group with any positive integer number of generators. Explicitly, by Bézout's Identity, there exists integers x and y such that $2x + 3y = 1$ because $\gcd(2, 3) = 1$. Consequently, every integer n can be written as $n = n \cdot 1 = n(2x + 3y) = 2nx + 3ny$. Under this identification, it is clear that $\mathbb{Z} = \langle 2, 3 \rangle$. Likewise, we have that $\gcd(6, 10, 15) = \gcd(\gcd(6, 10), 15) = \gcd(2, 15) = 1$, hence by Bézout's Identity, there exist integers x, y, and z such that $6x + 10y + 15z = 1$; as before, we conclude that $\mathbb{Z} = \langle 6, 10, 15 \rangle$. One can readily verify that the same argument extends to any number of distinct prime numbers. Explicitly, if $p_1, \dots, p_n$ are distinct prime numbers, then $\mathbb{Z}$ is finitely generated by the n positive integers $n_i = p_1 \cdots p_n / p_i$ for each integer $1 \leq i \leq n$.

Example 3.12.4. Observe that the group $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_2$ is a finitely generated abelian group of free rank two. Explicitly, we have that $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_2 = \langle (1, 0, 0), (0, 1, 0), (0, 0, 1 + 2\mathbb{Z}) \rangle$.

Consequently, finitely generated abelian groups generalize cyclic groups by allowing for systems of generators with more than one element. Our next proposition extends Proposition 2.3.10.

Proposition 3.12.5. *Every subgroup of a finitely generated abelian group is finitely generated.*

Proof. Consider a finitely generated abelian group G. We will proceed by induction on the number n of generators of G. By Proposition 2.3.10, the claim holds if $n = 1$ because G is cyclic in this case. Consequently, we may assume by induction that the proposition is true for some integer $n \geq 2$. Consider a system of generators $g_1, \dots, g_{n+1}$ for the abelian group G. Observe that the generator g_{n+1} generates a cyclic subgroup $C = \langle g_{n+1} \rangle$ of G. Every element of G can be written as $g_1^{k_1} \cdots g_{n+1}^{k_{n+1}}$, hence every element of G/C can be written as $g_1^{k_1} \cdots g_n^{k_n} C$ for some integers $k_1, \dots, k_n$. Crucially, we conclude that G/C is a finitely generated abelian group with a system of generators $g_1 C, \dots, g_n C$. By induction, every subgroup of G/C is finitely generated. Explicitly, for any subgroup H of G, the induced subgroup H/C of G/C is finitely generated by some elements $h_1 C, \dots, h_\ell C \in H/C$. We will produce an element $h_{\ell+1} \in H$ for which it holds that $H = \langle h_1, \dots, h_\ell, h_{\ell+1} \rangle$.

Consider the surjective function $\pi : G \to G/C$ defined by $\pi(g_1^{k_1} \cdots g_{n+1}^{k_{n+1}}) = g_1^{k_1} \cdots g_n^{k_n} C$. By hypothesis that G is abelian, π is a group homomorphism with $\ker \pi = C$ and $\pi(H) = H/C$. Given any element $h \in H$, there exist integers $k_1, \dots, k_\ell$ such that $\pi(h) = h_1^{k_1} \cdots h_\ell^{k_\ell} C = \pi(h_1^{k_1} \cdots h_\ell^{k_\ell})$ so that $\pi(h h_1^{-k_1} \cdots h_\ell^{-k_\ell}) = \pi(h)\pi(h_1^{-k_1} \cdots h_\ell^{-k_\ell}) = \pi(h)\pi(h_1^{k_1} \cdots h_\ell^{k_\ell})^{-1} = \pi(h)\pi(h)^{-1} = e_G C$ and $h h_1^{-k_1} \cdots h_\ell^{-k_\ell}$ lies in C. By assumption that H is a subgroup of G, it follows that $h h_1^{-k_1} \cdots h_\ell^{-k_\ell}$ is an element of H because the elements $h_1, \dots, h_\ell$ belong to H. We conclude that $h h_1^{-k_1} \cdots h_\ell^{-k_\ell} \in H \cap C$. Considering that H and C are both subgroups of G, it follows that $H \cap C$ is a subgroup of G. Even more, by Proposition 2.3.10, it is cyclic because it is a subgroup of the cyclic group C. Consequently, there exists an element $h_{\ell+1} \in H \cap C$ such that $h h_1^{-k_1} \cdots h_\ell^{-k_\ell} = h_{\ell+1}^{k_{\ell+1}}$ for some integer $k_{\ell+1}$. □

3.13 The Smith Normal Form

We turn our attention throughout this section to an indispensable tool in the theory of finitely generated abelian groups (and in general, in the theory of finitely generated modules over principal ideal domains). Explicitly, we will demonstrate how to determine the free rank and the invariant factor decomposition of a finitely generated abelian group as guaranteed by the Fundamental Theorem of Finitely Generated Abelian Groups. Even more, we will prove this theorem, and in the process, we will take a significant step toward proving the Fundamental Theorem of Finitely Generated Modules over a Principal Ideal Domain — leaving only some terminology to verify.

Given any positive integers m and n, we will henceforth denote by $\mathbb{Z}^{m \times n}$ the collection of $m \times n$ **integer matrices**. Explicitly, an $m \times n$ integer matrix is an **array** A consisting of m rows and n columns such that the data in the ith row and jth column is an integer for every pair of indices $1 \leq i \leq m$ and $1 \leq j \leq n$. Each array is written using rounded parentheses as follows.

Example 3.13.1. Consider the following 3×3 integer matrix.

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix}$$

We note that the data in the ith row and jth column of A is the integer $i + j$. Generally, we will refer to the data in the ith row and jth column of A as the (i, j)th **component** of A, and we will write a_{ij} as shorthand for the (i, j)th component of A. Consequently, we can describe the (i, j)th component of A in this case as $a_{ij} = i + j$. We refer to an $n \times n$ matrix as a **square** matrix.

One can define **matrix addition** and **matrix multiplication** of integer matrices using the usual addition and multiplication of integers. Explicitly, if A and B are $m \times n$ integer matrices, then we may define the matrix sum $A + B$ of A and B by specifying that the (i, j)th component of $A + B$ is the sum of the (i, j)th component of A and the (i, j)th component of B. Crucially, in order to add to matrices, they must both possess the same number of rows and columns. On the other hand, the matrix product is defined on the rows and columns of the two matrices. Explicitly, if A is an $m \times n$ integer matrix and B is an $n \times r$ integer matrix, then the matrix product AB is defined such that the (i, j)th component of AB is the sum of the products of the (i, k)th components of A and the (k, j)th components of B for all integers $1 \leq k \leq n$. Put another way, we have that

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + \cdots + a_{in} b_{nj}.$$

Consequently, the matrix product AB is defined only when the number of columns of A and the number of rows of B are equal. Observe that if A is an $m \times n$ integer matrix and B is an $n \times r$ integer matrix, then AB is an $m \times r$ integer matrix. Conversely, unless $m = r$ (i.e., the number of rows of A is equal to the number of columns of B), the matrix product BA is not defined.

One can readily verify that $\mathbb{Z}^{m \times n}$ is an abelian group with respect to matrix addition. Explicitly, the $m \times n$ zero matrix O is an integer matrix such that $A + O = A = O + A$ for all $m \times n$ integer matrices A. Consequently, the zero matrix is the **additive identity** of $\mathbb{Z}^{m \times n}$. Even more, matrix addition is simply a generalization of integer addition, hence it is associative and commutative.

Last, for any $m \times n$ integer matrix A, the $m \times n$ integer matrix $-A$ whose (i, j)th component is the **additive inverse** of the (i, j)th component of A is the additive inverse of A. Generally, we can define for any integer r the $m \times n$ integer matrix rA whose (i, j)th component is r times the (i, j)th component of A. We refer to this is the **scalar multiple** of A by r. Certainly, scalar multiplication is associative, distributive, and commutative. We summarize the results of this paragraph as follows.

Proposition 3.13.2. *We have that $\mathbb{Z}^{m \times n}$ is an abelian group with respect to matrix addition. Further, integer scalar multiplication of integer matrices is a well-defined function from $\mathbb{Z} \times \mathbb{Z}^{m \times n}$ to $\mathbb{Z}^{m \times n}$ that sends an ordered pair (r, A) to the $m \times n$ integer matrix rA that satisfies the following properties for all integers r and s and all $m \times n$ integer matrices A and B.*

1.) $r(A + B) = rA + rB$

2.) $(r + s)A = rA + sA$

3.) $r(sA) = (rs)A$

4.) $1A = A$

Generally, any algebraic structure for which there exists a well-defined notion of integer scalar multiplication is called a $\mathbb{Z}$-**module**. We note that $\mathbb{Z}$-modules are in fact ubiquitous.

Proposition 3.13.3. *Every abelian group G can be viewed as a $\mathbb{Z}$-module via the action $r \cdot g = g^r$.*

Proof. Given any two elements g and h in G and any integers r and s, we have that

1.) $r \cdot (gh) = (gh)^r = g^r h^r = (r \cdot g)(r \cdot h)$ by hypothesis that G is abelian;

2.) $(r + s) \cdot g = g^{r+s} = g^r g^s = (r \cdot g)(s \cdot g)$ by the Group Exponent Laws;

3.) $r \cdot (s \cdot g) = r \cdot (g^s) = (g^s)^r = g^{rs} = (rs) \cdot g$ by the Group Exponent Laws; and

4.) $1 \cdot g = g^1 = g$, as desired. □

Later, we will define the notion of an R-**module** for any commutative unital ring R. Essentially, R-modules generalize the notion of vector spaces; for now, we are ready for the main theorem.

Theorem 3.13.4 (Smith Normal Form)**.** *Given any nonzero $m \times n$ integer matrix A, there exists an invertible $m \times m$ integer matrix P and an invertible $n \times n$ integer matrix Q such that*

$$PAQ = \begin{pmatrix} n_1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & n_3 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n_\ell & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

and the integers $n_i \geq 1$ are unique (up to sign) and satisfy that $n_1 \mid n_2 \mid n_3 \mid \cdots \mid n_\ell$. Further, one can compute the integers n_i by the recursive formula $n_i = d_i / d_{i-1}$, where d_i is the greatest common divisor of all $i \times i$ minors of the $m \times n$ matrix A and d_0 is defined to be one.

Generally, the Smith Normal Form exists for any $m \times n$ matrix with entries in a **principal ideal domain**, e.g., the integers $\mathbb{Z}$ or any univariate polynomial ring $k[x]$, where k is a field (such as $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$). Computing the Smith Normal Form for an $m \times n$ integer matrix A amounts to carrying out some **elementary row operations** and elementary column operations on A to reduce the given matrix to the desired form; that it is possible to obtain the Smith Normal Form from such a process remains to be seen, but we will prove the theorem before the end of this section. Explicitly, we will find that the invertible $m \times m$ matrix P is obtained from the $m \times m$ **identity matrix** by performing the specified elementary row operations on A; likewise, the invertible $n \times n$ matrix Q is obtained from the $n \times n$ identity matrix by performing the specified elementary column operations on A. Recall that there are three elementary row (or column) operations as follows.

1.) We may multiply any row (or column) of the matrix by -1.

2.) We may add an integer multiple of a row (or column) of the matrix to another row (or column).

3.) We may interchange any pair of rows (or columns) of the matrix.

Going forward, we will use the shorthand $R_i \mapsto -R_i$ to denote the operation of multiplying the ith row of the matrix by -1; we will use the shorthand $R_j + aR_i \mapsto R_j$ to denote the operation of adding an integer multiple a of the ith row of the matrix to the jth row of the matrix (for any distinct indices i and j); and we will use the shorthand $R_i \leftrightarrow R_j$ to denote the operation of interchanging the ith and jth rows of the matrix. Each of these elementary row operations can also be performed with the ith and jth columns C_i and C_j of the matrix for any pair of distinct indices i and j.

Algorithm 3.13.5 (Smith Normal Form Algorithm)**.** Let A be an $m \times n$ integer matrix. Complete the following steps to determine the Smith Normal Form $\operatorname{SNF}(A)$ of A, the invertible $m \times m$ integer matrix P, and the invertible $n \times n$ integer matrix Q such that $\operatorname{SNF}(A) = PAQ$.

1.) Consider the components of each of the rows and columns of A. By Bézout's Identity, the greatest common divisor of the entries of any row (or column) of A can be obtained as an integer linear combination of the entries of that row (or column) of A. Consequently, if the greatest common divisor of the entries of any row (or column) of A is 1, then perform elementary column (or row) operations on A to obtain 1 in the first column (or row) of that row (or column). Otherwise, if the greatest common divisor of each row and column of A is not 1, then arbitrarily choose a row or column, and find its greatest common divisor as before. Perform each elementary row operation successively on the $m \times m$ identity matrix, and perform each elementary column operation successively on the $n \times n$ identity matrix.

2.) Once we have obtained the greatest common divisor of the entries of some row (or column), we can perform elementary column (or row) operations to eliminate every entry of that row (or column) other than the entry with the greatest common divisor. Once again, perform each elementary row operation successively on the $m \times m$ matrix obtained in the previous step from the $m \times m$ identity matrix, and perform each elementary column operation successively on the $n \times n$ matrix obtained in the previous step from the $n \times n$ identity matrix.

3.) We have reduced one row (or column) of A to a row consisting of $n - 1$ zeros and the greatest common divisor of that row (or a column consisting of $m - 1$ zeros and the greatest common divisor of that column). Observe that if the greatest common divisor of some row (or column) is 1, then we have obtained a row consisting of $n-1$ zeros and 1 (or a column consisting of $m-1$ zeros and 1). Otherwise, we may temporarily disregard this greatest common divisor and proceed with the algorithm on the other rows and columns of A.

Basically, the Smith Normal Form of A is obtained by performing elementary row and column operations on the matrix A, keeping track of the elementary row operations in an $m \times m$ matrix obtained from the $m \times m$ identity matrix by successively performing each row operation and keeping track of the elementary column operations in an $n \times n$ matrix obtained from the $n \times n$ identity matrix by successively performing each column operation.

Example 3.13.6. Let us compute the Smith Normal Form of the following 2×3 integer matrix.

$$A = \begin{pmatrix} 4 & -2 & 4 \\ 2 & 4 & 4 \end{pmatrix}$$

Using the Smith Normal Form Algorithm, we must first check whether any rows or columns of A have a greatest common divisor of 1. Unfortunately, this is not the case because all of the entries of A are even. Consequently, we must perform the algorithm somewhat arbitrarily. We will keep track of the elementary row operations by performing each such operation on the 2×2 identity matrix; likewise, we will keep track of the column operations by manipulating the columns of the 3×3 identity matrix according to the column operations on A. We achieve this as follows.

1.) $R_1 \leftrightarrow R_2$ $\qquad A \sim \begin{pmatrix} 2 & 4 & 4 \\ 4 & -2 & 4 \end{pmatrix}$ $\qquad P \sim \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

2.) $R_2 - 2R_1 \mapsto R_2$ $\qquad A \sim \begin{pmatrix} 2 & 4 & 4 \\ 0 & -10 & -4 \end{pmatrix}$ $\qquad P \sim \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}$

3.) $C_2 - 2C_1 \mapsto C_2$ $\qquad A \sim \begin{pmatrix} 2 & 0 & 4 \\ 0 & -10 & -4 \end{pmatrix}$ $\qquad Q \sim \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

4.) $C_3 - 2C_1 \mapsto C_3$ $\qquad A \sim \begin{pmatrix} 2 & 0 & 0 \\ 0 & -10 & -4 \end{pmatrix}$ $\qquad Q \sim \begin{pmatrix} 1 & -2 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

5.) $C_2 - 3C_3 \mapsto C_2$ $\qquad A \sim \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -4 \end{pmatrix}$ $\qquad Q \sim \begin{pmatrix} 1 & 4 & -2 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix}$

6.) $C_3 + 2C_2 \mapsto C_3$ $\qquad A \sim \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$ $\qquad Q \sim \begin{pmatrix} 1 & 4 & 6 \\ 0 & 1 & 2 \\ 0 & -3 & -5 \end{pmatrix}$

Consequently, the Smith Normal Form for A and the invertible matrices P and Q are as follows.

$$\operatorname{SNF}(A) = PAQ = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 4 & -2 & 4 \\ 2 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 & 4 & 6 \\ 0 & 1 & 2 \\ 0 & -3 & -5 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$$

Example 3.13.7. Let us compute the Smith Normal Form for the following 3×3 integer matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Considering that the first and second columns of A satisfy $\gcd(1, 4, 7) = 1$ and $\gcd(2, 5, 8) = 1$, we may begin with either of these columns; however, we will use the first column of A. Our aim is to first reduce the first column of A to the column consisting of 1 in the first row and zeros in the second and third rows; then, we will use this column to eliminate the nonzero entries of the first row of A. We will then restrict our attention to the resulting 2×2 submatrix of A in the bottom right-hand corner. We achieve this as follows, condensing the first four steps into two steps.

1.) $\begin{matrix} R_2 - 4R_1 \mapsto R_2 \\ R_3 - 7R_1 \mapsto R_3 \end{matrix}$ $\quad A \sim \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}$ $\quad P \sim \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -7 & 0 & 1 \end{pmatrix}$

2.) $\begin{matrix} C_2 - 2C_1 \mapsto C_2 \\ C_3 - 3C_1 \mapsto C_3 \end{matrix}$ $\quad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}$ $\quad Q \sim \begin{pmatrix} 1 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

3.) $R_3 - 2R_2 \mapsto R_3$ $\quad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad P \sim \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix}$

4.) $C_3 - 2C_1 \mapsto C_3$ $\quad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad Q \sim \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$

5.) $-C_2 \mapsto C_2$ $\quad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad Q \sim \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$

Consequently, the Smith Normal Form for A and the invertible matrices P and Q are as follows.

$$\operatorname{SNF}(A) = PAQ = \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Our next proposition illustrates that if we perform any elementary row or column operations on an $m \times n$ integer matrix A, then the quotient of the free abelian group $\mathbb{Z}^m$ of rank m by the abelian group generated by the columns of A is isomorphic to the quotient of the $\mathbb{Z}^m$ spanned by the columns of the resulting matrix that is row-and-column equivalent to A.

Proposition 3.13.8. *Given any $m \times n$ integer matrix A, let K_A denote the subgroup of $\mathbb{Z}^m$ generated by the columns of A. Given any invertible $m \times m$ integer matrix P and any invertible $n \times n$ integer matrix Q, the abelian groups $\mathbb{Z}^m/K_A$ and $\mathbb{Z}^m/K_{PAQ}$ are isomorphic.*

Proof. Consider the jth column $A_j = (a_{1j}, \dots, a_{mj})^t$ of the $m \times n$ integer matrix A. By definition, the group K_A generated by the columns of A is $K_A = \{c_1 A_1 + \dots + c_n A_n \mid c_1, \dots, c_n \in \mathbb{Z}\}$. Crucially, for any n-tuple $(c_1, \dots, c_n) \in \mathbb{Z}^n$, we have that $A(c_1, \dots, c_n) = c_1 A_1 + \dots + c_n A_n$. Consequently, the elements of K_A are of the form $A(c_1, \dots, c_n)$ for some element $(c_1, \dots, c_n) \in \mathbb{Z}^n$. Put another way, we have that $K_A = A\mathbb{Z}^n$. By the same argument applied to PAQ, we have that $K_{PAQ} = (PAQ)\mathbb{Z}^n$. By hypothesis that P is invertible, it follows that the group homomorphism $\rho : \mathbb{Z}^m \to \mathbb{Z}^m$ defined by $\rho(v) = Pv$ admits an inverse group homomorphism $\rho^{-1}(v) = P^{-1}v$, hence ρ is a group isomorphism. Consequently, we have that $(PAQ)\mathbb{Z}^n = P(AQ\mathbb{Z}^n) = \rho(AQ\mathbb{Z}^n) = AQ\mathbb{Z}^n$ by the associativity of matrix multiplication and the invertibility of P. Likewise, the group homomorphism $\sigma : \mathbb{Z}^n \to \mathbb{Z}^n$ defined by $\sigma(v) = Qv$ admits an inverse group homomorphism $\sigma^{-1}(v) = Q^{-1}v$, hence it is a group isomorphism. We conclude that $Q\mathbb{Z}^n = \sigma(\mathbb{Z}^n) = \mathbb{Z}^n$ so that $(AQ)\mathbb{Z}^n = A(Q\mathbb{Z}^n) = A\mathbb{Z}^n$. By part (c.) of Exercise 3.18.27, we have that $\mathbb{Z}^m/AQ\mathbb{Z}^n \cong \mathbb{Z}^m/(PAQ)\mathbb{Z}^n$, hence the conclusion follows.

$$\frac{\mathbb{Z}^m}{K_{PAQ}} = \frac{\mathbb{Z}^m}{(PAQ)\mathbb{Z}^n} \cong \frac{\mathbb{Z}^m}{(AQ)\mathbb{Z}^n} = \frac{\mathbb{Z}^m}{A\mathbb{Z}^n} = \frac{\mathbb{Z}^m}{K_A} \qquad \square$$

We provide at last a proof of the Fundamental Theorem of Finitely Generated Abelian Groups.

Proof. Given any finitely generated abelian group G with generators $g_1, \dots, g_k$, every element of G can be written as $g_1^{n_1} \cdots g_k^{n_k}$ for some integers $n_1, \dots, n_k$. Consequently, the function $\varphi : \mathbb{Z}^k \to G$ defined by $\varphi(n_1, \dots, n_k) = g_1^{n_1} \cdots g_k^{n_k}$ is surjective. Even more, it is a group homomorphism by assumption that G is abelian, hence by the First Isomorphism Theorem, we have that $G \cong \mathbb{Z}^k / \ker \varphi$. Considering that $(\mathbb{Z}^k, +)$ is a finitely generated group, it follows by Proposition 3.12.5 that $\ker \varphi$ is finitely generated, i.e., there exists an integer $r \geq 1$ and integers $a_{11}, \dots, a_{k1}, \dots, a_{1r}, \dots, a_{kr}$ such that every element of $\ker \varphi$ can be written as $c_1(a_{11}, \dots, a_{k1})^t + \dots + c_r(a_{1r}, \dots, a_{kr})^t$ for some integers $c_1, \dots, c_r$. Put another way, we have that $\ker \varphi = A\mathbb{Z}^r = \psi(\mathbb{Z}^r)$, where A is the $k \times r$ integer matrix whose jth column is $(a_{1j}, \dots, a_{kj})$ and $\psi : \mathbb{Z}^r \to \mathbb{Z}^k$ is the group homomorphism defined by $\psi(v) = Av$. Observe that if φ is injective, then $\ker \varphi$ consists of the zero k-tuple, hence the $k \times r$ matrix A must be the zero matrix; otherwise, A is a nonzero $k \times r$ integer matrix. By the Smith Normal Form, there exists an invertible $k \times k$ integer matrix P and an invertible $r \times r$ integer matrix Q such that PAQ consists of m nonzero entries $s_1 \mid s_2 \mid s_3 \mid \cdots \mid s_m$ followed by $k - m$ zeros along the diagonal, and all other components of PAQ are zero. By Proposition 3.13.8, we conclude that

$$G \cong \frac{\mathbb{Z}^k}{\ker \varphi} = \frac{\mathbb{Z}^k}{A\mathbb{Z}^r} \cong \frac{\mathbb{Z}^k}{PAQ\mathbb{Z}^r}.$$

Consequently, it remains to be seen that the latter group is simply the external direct product $\mathbb{Z}^{k-m} \times \mathbb{Z}_{s_1} \times \mathbb{Z}_{s_2} \times \mathbb{Z}_{s_3} \times \dots \times \mathbb{Z}_{s_m}$ of the free abelian group of $k - m$ and the finite abelian group with invariants factors $s_1, \dots, s_m$. Observe that for any element $(c_1, \dots, c_r) \in \mathbb{Z}^r$, we have that

$$(PAQ)(c_1, \dots, c_r)^t = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_m & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_m \\ c_{m+1} \\ c_{m+2} \\ \vdots \\ c_r \end{pmatrix} = \begin{pmatrix} c_1 s_1 \\ c_2 s_2 \\ c_3 s_3 \\ \vdots \\ c_m s_m \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

so that $PAQ\mathbb{Z}^r$ is the abelian group spanned by the integer $s_1, s_2, s_3, \dots, s_m$, and $k - m$ copies of 0. Explicitly, we have that $PAQ\mathbb{Z}^r = \langle s_1 \rangle \times \langle s_2 \rangle \times \langle s_3 \rangle \times \cdots \times \langle s_m \rangle \times \langle 0 \rangle \times \cdots \times \langle 0 \rangle$ with $k - m$ direct factors of 0. Going back to our above displayed isomorphism and appealing to the isomorphism of quotients of external direct products of Exercise 3.18.48, we conclude the result as follows.

$$\begin{aligned} \frac{\mathbb{Z}^k}{\langle s_1 \rangle \times \langle s_2 \rangle \times \langle s_3 \rangle \times \cdots \times \langle s_m \rangle \times \underbrace{\langle 0 \rangle \times \cdots \times \langle 0 \rangle}_{k-m \text{ factors}}} &\cong \frac{\mathbb{Z}}{s_1\mathbb{Z}} \times \frac{\mathbb{Z}}{s_2\mathbb{Z}} \times \frac{\mathbb{Z}}{s_3\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{s_m\mathbb{Z}} \times \underbrace{\mathbb{Z} \times \cdots \times \mathbb{Z}}_{k-m \text{ factors}} \\ &\cong \mathbb{Z}^{k-m} \times \mathbb{Z}_{s_1} \times \mathbb{Z}_{s_2} \times \mathbb{Z}_{s_3} \times \cdots \times \mathbb{Z}_{s_m} \qquad \square \end{aligned}$$

Before we conclude this section, let us discuss how to use the Smith Normal Form to compute the invariant factor decomposition guaranteed by the Fundamental Theorem of Finitely Generated Abelian Groups. Like we have seen in the two previous proofs, any $m \times n$ integer matrix A determines a group homomorphism $\alpha : \mathbb{Z}^n \to \mathbb{Z}^m$ defined by $\alpha(v) = Av$. One can readily verify that the image of $\mathbb{Z}^n$ under α is the subgroup of $\mathbb{Z}^m$ generated by the columns of A, i.e., we have that $\alpha(\mathbb{Z}^n) = K_A$ as in the notation of Proposition 3.13.8. We define the **cokernel** of a group homomorphism $\varphi : G \to H$ as the quotient of the codomain H by the image of G under φ, i.e., we have that $\operatorname{coker}(\alpha) = H/\varphi(G)$. Explicitly, it holds that $\operatorname{coker}(\alpha) = \mathbb{Z}^m/\alpha(\mathbb{Z}^n) = \mathbb{Z}^m/K_A$. Consider also the group homomorphism $\sigma : \mathbb{Z}^n \to \mathbb{Z}^m$ defined by $\sigma(v) = \operatorname{SNF}(A)v$. By definition, there exists an invertible $m \times m$ integer matrix P and an invertible $n \times n$ integer matrix Q such that $\operatorname{SNF}(A) = PAQ$, hence by Proposition 3.13.8, we have that $\operatorname{coker}(\alpha) = \mathbb{Z}^m/K_A \cong \mathbb{Z}^m/K_{\operatorname{SNF}(A)} = \operatorname{coker}(\sigma)$. Considering that $\operatorname{SNF}(A)$ is a diagonal matrix by construction, the cokernel of σ is the external direct product of a free abelian group of finite rank and a finite abelian group written as its invariant factor decomposition.

Consider for simplicity a finitely generated abelian group G that is a proper quotient of a free abelian group $\mathbb{Z}^r$ of rank r. By Proposition 3.12.5, every subgroup of the finitely generated abelian group $\mathbb{Z}^r$ is finitely generated, hence there exist integers $a_{11}, \dots, a_{r1}, \dots, a_{1s}, \dots, a_{rs}$ such that

$$G \cong \frac{\mathbb{Z}^r}{\langle (a_{11}, \dots, a_{r1}), \dots, (a_{1s}, \dots, a_{rs}) \rangle}.$$

Once again, the subgroup of $\mathbb{Z}^r$ generated by the elements $(a_{11}, \dots, a_{r1}), \dots, (a_{1s}, \dots, a_{rs})$ is generated by the columns of the $r \times s$ integer matrix A whose jth column is $(a_{1j}, \dots, a_{rj})^t$ for each integer $1 \leq j \leq s$. We will therefore refer to this matrix as the **matrix representation** of G.

By the previous paragraph, in order to express G in terms of its invariant factor decomposition, it suffices to find the Smith Normal Form for A. On the other hand, it is desirable to produce an explicit isomorphism between this proper quotient of $\mathbb{Z}^r$ and its invariant factor decomposition.

Algorithm 3.13.9 (Finding an Explicit Isomorphism of Finitely Generated Abelian Groups)**.** Let G be any quotient of a free abelian group $\mathbb{Z}^r$ represented by the $r \times s$ integer matrix A. Use the following algorithm to find an explicit isomorphism between G and its invariant factor decomposition.

1.) Compute the Smith Normal Form for A by performing a sequence elementary row and column operations on A to obtain a diagonal matrix with positive integers $n_1 \mid n_2 \mid \cdots \mid n_\ell$ followed by zeros along the diagonal. Be sure to keep track of all elementary row and column operations $R_i \leftrightarrow R_j$ and $R_j + aR_i \mapsto R_j$ and $C_i \leftrightarrow C_j$ and $C_j + aC_i \mapsto C_j$ performed in this step.

2.) Perform the elementary row operations from the previous step on the $r \times r$ identity matrix; if done correctly, the resulting matrix is the invertible $r \times r$ integer matrix P.

3.) Perform Gaussian Elimination (if possible) on the invertible matrix P to obtain its inverse P^{-1}. On the other hand, if this is not possible, then one can alternatively begin with the r-tuples $E_1, \dots, E_r$ of $\mathbb{Z}^r$ whose ith column is one and whose other columns are zero. Using the same order as the elementary row operations were performed in the first step, perform the **inverse** elementary column operation on the columns of the $r \times r$ matrix $\left(E_1^t \cdots E_r^t\right)$. Explicitly, if the row operation $R_i \leftrightarrow R_j$ was performed, then perform the column operation $C_i \leftrightarrow C_j$; if the row operation $R_j + aR_i \mapsto R_j$ was performed, then perform the column operations $C_i - aC_j \mapsto C_i$. Ultimately, the resulting matrix is the invertible $r \times r$ matrix P^{-1}.

4.) Construct a surjective group homomorphism φ from $\mathbb{Z}^r$ to the invariant factor decomposition of G by declaring that φ sends the jth column of P^{-1} to the generator of the cyclic group corresponding to the jth row of $\operatorname{SNF}(A)$. By the First Isomorphism Theorem, the induced group homomorphism from $\mathbb{Z}^r / \ker \varphi$ to the invariant factor decomposition of G is an isomorphism.

Proof. We prove that this algorithm achieves the desired result. Considering that Q is an invertible $s \times s$ matrix, it follows that the columns of Q generate $\mathbb{Z}^s$: explicitly, for any s-tuple $v \in \mathbb{Z}^s$, we have that $v = Q(Q^{-1}v)$ is an integer linear combination of the columns of Q so that $\mathbb{Z}^s = Q\mathbb{Z}^s$. Likewise, the columns of P^{-1} generate $\mathbb{Z}^r$ for a similar reason so that $\mathbb{Z}^r = P\mathbb{Z}^r$. Consequently, the group homomorphism $\varphi : \mathbb{Z}^r \to \mathbb{Z}^r / PAQ\mathbb{Z}^r$ defined by $\varphi(c_1 P_1^{-1} + \cdots + c_r P_r^{-1}) = (c_1, \dots, c_r)^t + PAQ\mathbb{Z}^s$ is surjective. Considering that $\varphi(P_j^{-1}) = E_j^t + PAQ\mathbb{Z}^r$, we conclude that φ sends the jth column P_j^{-1} of P^{-1} to the generator of the jth cyclic group $\mathbb{Z}/n_j\mathbb{Z}$ of $\mathbb{Z}^r / PAQ\mathbb{Z}^s$, hence this is in fact the group homomorphism described in the algorithm. Even more, by the opening remarks of the proof, we have that $c_1 P_1^{-1} + \cdots + c_r P_r^{-1}$ lies in $\ker \varphi$ if and only if $(c_1, \dots, c_r)^t + PAQ\mathbb{Z}^s = (0, \dots, 0)^t + PAQ\mathbb{Z}^r$ if and only if $(c_1, \dots, c_r)^t$ lies in $PAQ\mathbb{Z}^s = PA\mathbb{Z}^s = A\mathbb{Z}^s$. We conclude therefore that $\ker \varphi = A\mathbb{Z}^s$. By the First Isomorphism Theorem, the surjective group homomorphism φ induces a group isomorphism $\psi : \mathbb{Z}^r / A\mathbb{Z}^s \to \mathbb{Z}^r / PAQ\mathbb{Z}^s$, the latter of which is precisely the invariant factor decomposition of G by the proof of the Fundamental Theorem of Finitely Generated Abelian Groups. □

Example 3.13.10. Consider the following finitely generated abelian group.

$$G = \frac{\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}}{\langle (0,0,3,1)^t, (0,6,0,0)^t, (0,1,0,1)^t \rangle}.$$

By the preceding discussion, the matrix representation of G is as follows.

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 6 & 1 \\ 3 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

By performing elementary row and column operations, we convert A into its Smith Normal Form, i.e., a diagonal matrix whose positive entries are $n_1 \mid n_2 \mid \cdots \mid n_\ell$. One way to do this is as follows.

$$1.)\ R_1 \leftrightarrow R_4 \qquad A \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 6 & 1 \\ 3 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad P \sim \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$2.)\ R_3 - 3R_1 \mapsto R_3 \qquad A \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 6 & 1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix} \qquad P \sim \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -3 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$3.)\ C_3 - C_1 \mapsto C_3 \qquad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix} \qquad Q \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$4.)\ R_3 + 3R_2 \mapsto R_3 \qquad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 1 \\ 0 & 18 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad P \sim \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & -3 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$5.)\ C_2 - 6C_3 \mapsto C_2 \qquad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 18 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad Q \sim \begin{pmatrix} 1 & 6 & -1 \\ 0 & 1 & 0 \\ 0 & -6 & 1 \end{pmatrix}$$

$$6.)\ C_2 \leftrightarrow C_3 \qquad A \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 18 \\ 0 & 0 & 0 \end{pmatrix} \qquad Q \sim \begin{pmatrix} 1 & -1 & 6 \\ 0 & 0 & 1 \\ 0 & 1 & -6 \end{pmatrix}$$

We conclude from $\operatorname{SNF}(A)$ that $G \cong \mathbb{Z}^4/(\langle 1 \rangle \times \langle 1 \rangle \times \langle 18 \rangle \times \langle 0 \rangle) \cong \{0\} \times \{0\} \times (\mathbb{Z}/18\mathbb{Z}) \times \mathbb{Z}$.

Consider the quadruples $E_1 = (1, 0, 0, 0)$, $E_2 = (0, 1, 0, 0)$, $E_3 = (0, 0, 1, 0)$, and $E_4 = (0, 0, 0, 1)$. Observe that the 4×4 identity matrix is the 4×4 matrix $\begin{pmatrix} E_1^t & E_2^t & E_3^t & E_4^t \end{pmatrix}$ whose ith column is

E_i. Even though it is possible to find P^{-1} via the method of Gaussian Elimination because each column of P possesses an entry with 1, for illustrative purposes, we will find P^{-1} by performing the inverse of the elementary row operations on the columns of the 4×4 identity matrix. Explicitly, if we used the elementary row operation $R_j + aR_i \mapsto R_j$, then we will now use the corresponding elementary column operation $C_i - aC_j \mapsto C_i$, and we will use $C_i \leftrightarrow C_j$ where we used $R_i \leftrightarrow R_j$.

1.) $C_1 \leftrightarrow C_4$ $\quad P^{-1} \sim \begin{pmatrix} E_4^t & E_2^t & E_3^t & E_1^t \end{pmatrix}$

2.) $C_1 + 3C_3 \mapsto C_1$ $\quad P^{-1} \sim \begin{pmatrix} 3E_3^t + E_4^t & E_2^t & E_3^t & E_1^t \end{pmatrix}$

4.) $C_2 - 3C_3 \mapsto C_2$ $\quad P^{-1} \sim \begin{pmatrix} 3E_3^t + E_4^t & E_2^t - 3E_3^t & E_3^t & E_1^t \end{pmatrix}$

Consequently, the first column of P^{-1} is $3E_3 + E_4 = (0, 0, 3, 1)$; the second column is $E_2 - 3E_3 = (0, 1, -3, 0)$; the third column is $E_3 = (0, 0, 1, 0)$; and the fourth column is $E_1 = (1, 0, 0, 0)$.

$$P^{-1} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 3 & -3 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

We conclude by Algorithm 3.13.9 that the function $\varphi : \mathbb{Z}^4 \to \mathbb{Z}/18\mathbb{Z} \times \mathbb{Z}$ defined by

$$\varphi(c_1 P_1^{-1} + c_2 + P_2^{-1} + c_3 P_3^{-1} + c_4 P_4^{-1}) = (0, 0, c_3 + 18\mathbb{Z}, c_4)$$

is a surjective group homomorphism with kernel $A\mathbb{Z}^4 = \langle (0, 0, 3, 1), (0, 6, 0, 0), (0, 1, 0, 1) \rangle$, hence the function $\psi : G \to \mathbb{Z}/18\mathbb{Z} \times \mathbb{Z}$ induced by the group homomorphism φ is a group isomorphism.

3.14 Group Actions and the Class Equation

Consider any group G and any nonempty set X. We say that a function $* : G \times X \to X$ that sends $(g, x) \mapsto g * x$ is a **group action** whenever this function satisfies the properties that

1.) $g * (h * x) = gh * x$ for all elements $g, h \in G$ and all elements $x \in X$ and

2.) $e_G * x = x$ for all elements $x \in X$.

Given any group G and any nonempty set X such that there exists a group action $* : G \times X \to X$, we say that G **acts on** X by $*$. We have already tacitly encountered some examples of group actions.

Example 3.14.1. Consider the general linear group $\mathrm{GL}(2, \mathbb{R})$ of invertible 2×2 real matrices. By identifying the external direct product $\mathbb{R}^2$ with the collection of 2×1 real column vectors, matrix multiplication on the left by an element of $\mathrm{GL}(2, \mathbb{R})$ defines a group action as follows.

$$\mathrm{GL}(2, \mathbb{R}) \times \mathbb{R}^2 \to \mathbb{R}^2 \text{ that sends } \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right) \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

Explicitly, matrix multiplication is associative, hence the first property of a group action is satisfied; the second is satisfied because the 2×2 identity matrix I yields $IX = X$ for all elements $X \in \mathbb{R}^2$.

Example 3.14.2. Consider the dihedral group D_3 consisting of the six symmetry-preserving rotations and reflections of a triangle. By the running example of Section 1.8, we may view the elements of D_3 as the six permutations (1), (12), (13), (23), (123), and (132). Even more, it follows that D_3 acts on the set $X = \{1, 2, 3\}$ of the three vertices of the triangle via $(\sigma, x) \mapsto \sigma(x)$ for each permutation $\sigma \in D_3$ and each integer $x \in X$: indeed, we have that $\tau(\sigma(x)) = (\tau \circ \sigma)(x)$ for every pair of permutations $\sigma, \tau \in D_3$, and the identity permutation is given by $\iota = (1)$.

Example 3.14.3. Given any group G and any subgroup H of G, we may define a group action of H on G by declaring that $h * g = hgh^{-1}$. Explicitly, we have that $e_G * g = e_G g e_G^{-1} = g$ for all elements $g \in G$. Likewise, for all elements $h_1, h_2 \in H$ and any element $g \in G$, it holds that

$$h_1 * (h_2 * g) = h_1 * (h_2 g h_2^{-1}) = h_1(h_2 g h_2^{-1})h_1^{-1} = h_1 h_2 g (h_1 h_2)^{-1} = (h_1 h_2) * g.$$

Conversely, it is not necessarily true that G acts on an arbitrary subgroup H of G by conjugation. Explicitly, this action is well-defined if and only if $g * h = ghg^{-1}$ is an element of H for all elements $g \in G$ and $h \in H$ if and only if H is a normal subgroup of G by Proposition 3.2.1.

Other than in the previous example, one thing to notice about group actions in general is that the set X on which the group G acts need not be related in any obvious manner to the group. Explicitly, for any group G and any set X, we can always define the **trivial action** $G \times X \to X$ by declaring that $(g, x) \mapsto x$ for all elements $g \in G$; however, the most profitable group actions will be non-trivial, as we shall soon see. Given any group action $* : G \times X \to X$ of a group G on a set X, we define the **kernel** of $*$ as the subset $K_* = \{g \in G \mid g * x = x \text{ for all elements } x \in X\}$ of G. Put another way, the kernel of a group action consists of all elements $g \in G$ that act trivially on all elements $x \in X$. By definition of a group action, the identity element $e_G \in G$ acts trivially on all elements $x \in X$, hence K_* is nonempty. We will say that a group action is **faithful** if and only if the identity element e_G of G is the only element of G that acts trivially on all elements $x \in X$ if and only if the kernel K_* of the group action is trivial if and only if $K_* = \{e_G\}$.

Given any element $x \in X$, we may wish to consider the collection of elements $g \in G$ under which x is fixed by the group action; this gives rise to the **stabilizer** $\operatorname{Stab}_G(x) = \{g \in G \mid g * x = x\}$ of the element $x \in X$. Observe that the kernel of a group action is precisely given by the intersection of the stabilizers of every element $x \in X$, i.e., we have that $K_* = \cap_{x \in X} \operatorname{Stab}_G(x)$.

Example 3.14.4. Given any group G, we may define a group action of G on itself by declaring that $g_1 * g_2 = g_1 g_2$; this action is simply left-multiplication by a group element. One can verify that this is a well-defined group action by the very definition of a group: the associativity of the group operation yields that $g_1 * (g_2 * g_3) = g_1 * (g_2 g_3) = g_1(g_2 g_3) = (g_1 g_2) g_3 = (g_1 g_2) * g_3$ for all elements $g_1, g_2, g_3 \in G$, and the identity element by definition satisfies that $e_G * x = e_G x = x$ for all elements $x \in G$. By definition, we have that $K_* = \{g \in G \mid g * x = x \text{ for all elements } x \in G\}$, hence this action is faithful because it satisfies that $gx = g * x = x$ for all elements $x \in G$ if and only if $g = e_G$. Even more, $\operatorname{Stab}_G(x)$ is trivial for any element $x \in G$ because $gx = g * x = x$ implies that $g = e_G$.

Example 3.14.5. Given any group G, consider the group action $g * x = gxg^{-1}$ of conjugation as defined in Example 3.14.3. By definition, we have that $g \in K_*$ if and only if $g * x = x$ for all elements $x \in G$ if and only if $gxg^{-1} = x$ for all elements $x \in G$ if and only if $gx = xg$ for all elements $x \in G$ if and only if g lies in the center $Z(G)$ of G. Consequently, the kernel of the conjugation action is $Z(G)$. Even more, for any element $x \in X$, we have that $g \in \operatorname{Stab}_G(x)$ if and only if $g * x = x$ if and only if $gxg^{-1} = x$ if and only if $gx = xg$ if and only if g lies in the centralizer $Z_G(x)$ of x in G (cf. Exercise 2.8.19). Ultimately, this example illustrates that group actions generalize familiar algebraic structures related to groups. Explicitly, we may realize subgroups of G such as its center $Z(G)$ and the centralizer $Z_G(x)$ of an element $x \in G$ respectively as the kernel K_* and stabilizer $\operatorname{Stab}_G(x)$ of the element $x \in G$ under conjugation; this alone motivates the study of group actions.

Our next proposition demonstrates that the kernel of any group action and the stabilizer of any element under a group action are both normal subgroups of the ambient group.

Proposition 3.14.6. *Consider a group G acting on a nonempty set X via $(g, x) \mapsto g * x$. We have that $K_* = \{g \in G \mid g * x = x \text{ for all elements } x \in X\}$ and $\operatorname{Stab}_G(x) = \{g \in G \mid g * x = x\}$ for some element $x \in X$ are both subgroups of G. Even more, K_* is a normal subgroup of G.*

Proof. We will prove that for any element $x \in X$, the stabilizer $\operatorname{Stab}_G(x)$ of x in G is a subgroup of G; then, we may appeal to Exercise 2.8.22 and the fact that $K_* = \cap_{x \in X} \operatorname{Stab}_G(x)$ to conclude that K_* is a subgroup of G. We may use the Two-Step Subgroup Test to prove that $\operatorname{Stab}_G(x)$ is a subgroup of G because it is true by definition of a group action that $e_G \in \operatorname{Stab}_G(x)$ for any element $x \in X$. Observe that if $g, h \in \operatorname{Stab}_G(x)$, then it holds that $(gh) * x = g * (h * x) = g * x = x$ by definition of a group action and by definition of the stabilizer of x in G. Consequently, it suffices to prove that for every element $g \in \operatorname{Stab}_G(x)$, we have that $g^{-1} \in \operatorname{Stab}_G(x)$. But this is true because $e_G * x = (g^{-1}g) * x = g^{-1} * (g * x) = g^{-1} * x$ by definition of a group action and by assumption that $g \in \operatorname{Stab}_G(x)$. We conclude therefore that $\operatorname{Stab}_G(x)$ is a subgroup of G, hence K_* is a subgroup of G because it is the intersection of subgroups of G. Even more, it holds that K_* is normal in G because for any element $g \in K_*$ and any element $h \in G$, we have that

$$(hgh^{-1}) * x = h * (g * (h^{-1} * x)) = h * (h^{-1} * x) = (hh^{-1}) * x = e_G * x = x.$$

Explicitly, the first, third, and fourth equalities hold because $*$ is a group action, and the second equality holds because every element $g \in K_*$ satisfies that $g * x = x$ for all elements $x \in X$. □

Consequently, an action of a group G on a nonempty set set X gives rise to two subgroups of G, hence we may study the group G by studying these subgroups for each possible action of G on nonempty sets — especially the actions of G on itself. One of the most fruitful ways to exploit this observation is by considering the relation $X_* = \{(x, y) \in X \times X \mid y = g*x \text{ for some element } g \in G\}$.

Proposition 3.14.7. *Given any group G acting on a nonempty set X via $(g, x) \mapsto g*x$, the relation $X_* = \{(x, y) \in X \times X \mid y = g * x \text{ for some element } g \in G\}$ is an equivalence relation on X.*

Proof. We must demonstrate that X_* is (1.) reflexive, (2.) symmetric, and (3.) transitive.

1.) By definition of a group action, we have that $x = e_G * x$ so that (x, x) lies in X_*.

2.) Given any element $(x, y) \in X_*$, there exists an element $g \in G$ such that $y = g * x$. Observe that $g^{-1} * y = g^{-1} * (g * x) = (g^{-1}g) * x = e_G * x = x$, hence (y, x) lies in X_*.

3.) Last, for any pair of elements $(x, y), (y, z) \in X_*$, by definition, we can find elements $g, h \in G$ such that $y = g * x$ and $z = h * y = h * (g * x) = (hg) * x$; this shows that $(x, z) \in X_*$. □

By Proposition 1.4.5, the equivalence relation X_* on X induces a partition of X via its equivalence classes $[x] = \{y \in X \mid (x, y) \in X_*\} = \{y \in X \mid y = g * x \text{ for some element } g \in G\}$. We will adopt the very literal notation $G * x = \{y \in X \mid y = g * x \text{ for some element } g \in G\}$ to denote the equivalence class of x modulo X_*, and we will refer to $G * x$ as the **orbit** of x under the group action $*$. We will say that a group action is **transitive** if there is only one orbit of X modulo X_*. Put another way, a group action is transitive if and only if for every pair of elements $x, y \in X$, there exists an element $g \in G$ such that $y = g * x$; this is because if $*$ is transitive, then there is only one orbit of X modulo X_*, hence for every element $x \in X$, we must have that $X = G * x$.

Our next theorem establishes that the number of elements in any orbit $G * x$ of X modulo X_* is precisely the index of the stabilizer $\operatorname{Stab}_G(x)$ of the element $x \in X$ in G.

Theorem 3.14.8 (Orbit-Stabilizer Theorem)**.** *Given any group G acting on a nonempty set X via $(g, x) \mapsto g * x$, the number of elements in the equivalence class of any element $x \in X$ modulo X_* is the index of the stabilizer of x in G, i.e., we have that $|G * x| = \#\{g * x \mid g \in G\} = [G : \operatorname{Stab}_G(x)]$. Particularly, if G is finite, then for every element $x \in X$, we have that $|G| = |G * x| \cdot |\operatorname{Stab}_G(x)|$.*

Proof. Consider the function $\varphi : G * x \to G/\operatorname{Stab}_G(x)$ defined by $\varphi(g * x) = g \operatorname{Stab}_G(x)$. We claim that φ is a well-defined bijection, hence we have that $\#\{g * x \mid g \in G\} = |G * x| = [G : \operatorname{Stab}_G(x)]$. Given any elements $x, y \in X$ such that $g * x = g * y$, it follows that $g^{-1} * (g * x) = g^{-1} * (g * y)$ so that $x = e_G * x = (g^{-1}g) * x = (g^{-1}g) * y = e_G * y = y$ and $g \operatorname{Stab}_G(x) = g \operatorname{Stab}_G(y)$. We conclude that φ is well-defined. Even more, it is surjective by definition. On the other hand, we have that $g \operatorname{Stab}_G(x) = h \operatorname{Stab}_G(x)$ if and only if $h^{-1}g \operatorname{Stab}_G(x) = e_G \operatorname{Stab}_G(x)$ if and only if $h^{-1}g$ is in $\operatorname{Stab}_G(x)$ if and only if $(h^{-1}g) * x = x$ if and only if $h * (h^{-1}g * x) = h * x$ if and only if $g * x = e_G * (g * x) = (hh^{-1}) * (g * x) = h * x$, from which it follows that φ is injective.

Last, if G is finite, the result follows from Lagrange's Theorem and Proposition 3.14.6: the former ensures that $|G| = [G : \operatorname{Stab}_G(x)]|\operatorname{Stab}_G(x)|$, and the latter shows that $|G * x| = [G : \operatorname{Stab}_G(x)]$. □

By Example 3.14.3, for any group G, we may always consider the group action $* : G \times G \to G$ defined by $(g, x) \mapsto gxg^{-1}$ that we have previously referred to as conjugation. By Proposition 3.14.7, this action induces an equivalence relation G_* on G. We will say that two elements $g, x \in G$ are **conjugate** in G if and only if they lie in the same equivalence class of G modulo G_* if and only if there exists an element $h \in G$ such that $g = hxh^{-1} = h * x$; we will refer to the equivalence classes of G modulo G_* as the **conjugacy classes** of elements of G. Crucially, we note that $G * x$ is a singleton if and only if for every element $g \in G$, we have that $gxg^{-1} = g * x = x$ if and only if $gx = xg$ for all elements $g \in G$ if and only if x lies in the center $Z(G)$ of G. Consequently, the distinct conjugacy classes of elements of G that do not lie in the center of G will have cardinality exceeding one. By Example 3.14.5, we have that $\operatorname{Stab}_G(x)$ is the centralizer $Z_G(x)$ of x in G, hence the Orbit-Stabilizer Theorem guarantees that the size of any orbit of G modulo the action of G on itself by conjugation is $[G : Z_G(x)]$. We obtain the following formula for the order of a finite group.

Theorem 3.14.9 (Class Equation)**.** *Given any finite group G with center $Z(G)$ and representatives $x_1, \dots, x_n$ of the distinct conjugacy classes of elements of G not contained in $Z(G)$, we have that*

$$|G| = |Z(G)| + \sum_{i=1}^{n} [G : Z_G(x_i)].$$

Proof. Considering that G acts on itself by conjugation, it follows by Propositions 3.14.7 and 1.4.5 that the equivalence relation $x \sim y$ if and only if $y = gxg^{-1}$ for some element $g \in G$ partitions G into the disjoint union of the singleton orbits $G * z_i$ for each element $z_i \in Z(G)$ and the larger orbits $G * x_1, \dots, G * x_n$ represented by the elements $x_1, \dots, x_n$ that lie in the distinct conjugacy classes of the elements of G that are not contained in $Z(G)$. Explicitly, we have the following disjoint union.

$$G = \bigcup_{x \in G} (G * x) = (G * z_1) \cup \cdots \cup (G * z_k) \cup (G * x_1) \cup \cdots \cup (G * x_n)$$

Consequently, the formula for $|G|$ follows from this partition of G and the Orbit-Stabilizer Theorem.

$$|G| = \sum_{i=1}^{k} |(G * z_i)| + \sum_{i=1}^{n} |(G * x_i)| = \sum_{i=1}^{k} 1 + \sum_{j=1}^{n} [G : \operatorname{Stab}_G(x_i)] = |Z(G)| + \sum_{i=1}^{n} [G : Z_G(x_i)]. \qquad \square$$

Generally, for any group G acting on a nonempty finite set X via $(g, x) \mapsto g * x$, we can derive a formula for the cardinality of X in terms of some generalization of the "center" of the X under the group action. Explicitly, we wish to consider the set of elements of X whose orbits under the action of G are singletons. Put another way, we want to define a set consisting of all elements $x \in X$ such that $G * x$ has size one. Considering that $e_G * x = x$ by definition of a group action, we have that $|G * x| = 1$ if and only if for every element $g \in G$, we have that $g * x = e_G * x = x$. We define the set of **fixed points** of X under the action of G by $\operatorname{Fix}_G(X) = \{x \in X \mid g * x = x \text{ for all elements } g \in G\}$.

Theorem 3.14.10 (Class Equation of a Group Action). *Given any group G acting on a nonempty finite set X and any representatives $x_1, \dots, x_n$ for the distinct orbits $G * x_i$ of elements of X that are not contained in $\operatorname{Fix}_G(X) = \{x \in X \mid g * x = x \text{ for all elements } g \in G\}$, we have that*

$$|X| = |\operatorname{Fix}_G(X)| + \sum_{i=1}^{n} [G : \operatorname{Stab}_G(x_i)].$$

Proof. Essentially, the proof of this fact is the same as the proof of the Class Equation. We leave it to the reader as Exercise 3.18.78 to verify that the same argument will suffice. □

3.15 Sylow's Theorems

One of the foremost uses of the Class Equation is in the study of elements of prime-power order in finite groups. Given a prime number p and a non-negative integer n, we say that a group of order p^n is a **p-group**. By Lagrange's Theorem, every subgroup of a p-group is also a p-group because the only positive integers that divide p^n are $1, p, p^2, \dots, p^n$. Even more, every group contains at least one p-subgroup — namely, the trivial p-subgroup $\{e_G\}$. Our first significant result of this section guarantees the existence of non-trivial p-subgroups of G for any prime number p dividing $|G|$.

Theorem 3.15.1 (Cauchy's Theorem for Finite Groups). *Given any finite group G and any prime number p that divides $|G|$, there exists an element $g \in G$ such that* $\operatorname{ord}(g) = p$.

Proof. We proceed by the Principle of Complete Induction on $|G|$. Crucially, we note that if p is a prime number that divides $|G|$, then the smallest positive integer that $|G|$ can be is p. Consequently, we may take this as our base case. By Proposition 3.1.13, G must be cyclic, and the result holds.

We will assume inductively that the proposition is true for all integers $2p, 3p, \dots, (k-1)p$ for some integer $k \geq 3$. We must prove the result for $|G| = kp$. By the Class Equation, we have that

$$kp = |G| = |Z(G)| + \sum_{i=1}^{n} [G : Z_G(x_i)]$$

for some representatives $x_1, \dots, x_n$ of the distinct conjugacy classes of the elements of G that do not lie in the center $Z(G)$. By Lagrange's Theorem, it follows that $kp = |G| = [G : Z_G(x_i)]|Z_G(x_i)|$, hence by Exercise 1.10.27, one of the positive integers $[G : Z_G(x_i)]$ or $|Z_G(x_i)|$ must be divisible by p. By the Law of the Excluded Middle, either $[G : Z_G(x_i)]$ is divisible by p for all integers $1 \leq i \leq n$, or $[G : Z_G(x_i)]$ is not divisible by p for some integer $1 \leq i \leq n$. Observe that if the former holds, then by the Class Equation, we must have that $|Z(G)|$ is divisible by p; therefore, by the Fundamental Theorem of Finite Abelian Groups, there exists an integer $e \geq 1$ such that the cyclic group $\mathbb{Z}/p^e\mathbb{Z}$ is a direct factor of $Z(G)$, hence we may identify $\mathbb{Z}/p^e\mathbb{Z}$ with a subgroup of $Z(G)$. Ultimately, the cyclic subgroup $p^{e-1}\mathbb{Z}/p^e\mathbb{Z}$ of $\mathbb{Z}/p^e\mathbb{Z}$ of order p induces a cyclic subgroup of $Z(G)$ of order p; its generator is an element of G of order p. Conversely, if $[G : Z_G(x_i)]$ is not divisible by p for some integer $1 \leq i \leq n$, then $|Z_G(x_i)|$ must be divisible by p. By our inductive hypothesis, there exists an element $g \in Z_G(x_i)$ of order p, hence G admits an element of order p, as desired. □

Corollary 3.15.2. *Given any finite abelian group G and any prime number p that divides the order of G, the subgroup $G(p) = \{g \in G \mid \operatorname{ord}(g) = p^n \text{ for some integer } n \geq 0\}$ of G satisfies that $|G(p)| = p^k$ for some integer $k \geq 0$. Put another way, for every finite abelian group G and every prime factor p of $|G|$, the subgroup of G consisting of p-power order elements of G is a p-group.*

Proof. By Lagrange's Theorem, we have that $|G(p)|$ is divisible by p, hence we must have that $|G| = p^k m$ for some positive integers k and m. On the contrary, suppose that m is divisible by some prime number q that is distinct from p. By Cauchy's Theorem for Finite Groups, it follows that $G(p)$ admits an element of order q — a contradiction to the definition of $G(p)$. □

Even more, we can strengthen the previous corollary for p-groups as follows.

Proposition 3.15.3. *Given any group G of order p^n for some prime number p and some positive integer n, for each integer $0 \leq k \leq n$, there exists a p-subgroup of G of order p^k.*

Proof. We proceed by the Principle of Complete Induction on the positive integer n. Clearly, if $n = 1$, then the statement holds trivially. Consequently, we may assume that the statement of the proposition is true for all positive integers not exceeding $n-1$. By Cauchy's Theorem for Finite Groups, there exists an element $g \in G$ of order p. Consider the cyclic subgroup $H = \langle g \rangle$. By the Law of the Excluded Middle, either H is a normal subgroup of G, or it is not. Observe that if the former holds, then G/H is a group of order p^{n-1} by Lagrange's Theorem and Proposition 3.2.1. Consequently, by our inductive hypothesis, for each integer $0 \leq \ell \leq n-1$, there exists a p-subgroup of G/H of order p^ℓ. By the Fourth Isomorphism Theorem, each of these p-subgroups induces a subgroup K of G such that $p^\ell = |K/H| = |K|/|H| = |K|/p$ and $|K| = p^{\ell+1}$. Conversely, if H is not a normal subgroup of G, then the normalizer $N_G(H)$ of H in G is the "largest" subgroup of G with H as a normal subgroup by Exercise 3.18.15. By Cauchy's Theorem for Finite Groups, there exists a subgroup of $N_G(H)/H$ of order p that induces a subgroup K of G of order p^2. By repeating the above process with K, we obtain p-subgroups of G of order p^k for each integer $3 \leq k \leq n-1$. □

Cauchy's Theorem for Finite Groups provides a powerful tool for the study of finite groups; however, by the end of this section, we will also develop the so-called **Sylow's Theorems** that extend Cauchy's Theorem for Finite Groups to demonstrate the existence of p-subgroups of maximal prime-power order in any finite group. Explicitly, for any finite group G such that $|G| = p^n m$ for some non-negative integers m and n such that $\gcd(p, m) = 1$ (i.e., p^n is the largest power of p that divides the order of G), we refer to any subgroup of G of maximal p-power order p^n as a **Sylow p-subgroup** of G. We will denote by $\operatorname{Syl}_p(G)$ the collection of all Sylow p-subgroups of G; its cardinality $n_p(G)$ is precisely the number of distinct Sylow p-subgroups of G.

Theorem 3.15.4 (Sylow's Theorems)**.** *Consider any group G of order $p^n m$ for some prime number p and some positive integers n and m such that $\gcd(p, m) = 1$.*

1.) *Given any integer $0 \leq k \leq n$, there exists a p-subgroup of G of order p^k. Particularly, there exists at least one Sylow p-subgroup of G. Put another way, $\operatorname{Syl}_p(G)$ is nonempty.*

2.) *If P is a Sylow p-subgroup of G and Q is any p-subgroup of G, then there exists an element g in G such that $Q \subseteq gPg^{-1}$. Particularly, any two Sylow p-subgroups of G are conjugate in G.*

3.) *We have that $n_p(G) \equiv 1 \pmod{p}$ and $n_p(G) \mid m$.*

Often, Sylow's Theorems are presented as the First Sylow Theorem, the Second Sylow Theorem, and the Third Sylow Theorem, respectively. Even though we will not view them separately in this way, considering that the proofs are quite substantial, we will tackle those one at a time instead of all together. We are prepared already to prove the First Sylow Theorem.

Proof. We will prove the first of Sylow's Theorems. Considering that this is merely a strengthening of Cauchy's Theorem for Finite Groups, it is natural to anticipate that a similar strategy as in the proof of the aforementioned fact will work at present. Consequently, we proceed by the Principle of Complete Induction on $|G|$. Our base case is that $|G| = p$, hence the result here follows from the fact that G must be cyclic by Proposition 3.1.13. We will assume by induction that the statement of the proposition holds for all positive integers that are strictly smaller than $p^n m$. By the Class Equation, once again, if we consider any representatives $x_1, \dots, x_r$ of the distinct conjugacy classes of the elements of G that do not lie in the center $Z(G)$, then we have that

$$p^n m = |G| = |Z(G)| + \sum_{i=1}^{r} [G : Z_G(x_i)].$$

By the Law of the Excluded Middle, either $[G : Z_G(x_i)]$ is divisible by p for all integers $1 \leq i \leq r$, or $[G : Z_G(x_i)]$ is not divisible by p for some integer $1 \leq i \leq n$. Observe that if the former holds, then $|Z(G)|$ must be divisible by p, hence Cauchy's Theorem for Finite Groups guarantees that $Z(G)$ possesses an element $g \in Z(G)$ of order p. Consider the cyclic subgroup $H = \langle g \rangle$. By assumption that $g \in Z(G)$, it follows that H is a normal subgroup of G because g commutes with all elements of G. Consequently, G/H is a group by Proposition 3.2.1; its order is $p^{n-1}m$ by Lagrange's Theorem, so our inductive hypothesis yields a subgroup of G/H of order p^ℓ for each integer $0 \leq \ell \leq n-1$. By the Fourth Isomorphism Theorem, each of these subgroups induces a subgroup of G of order $p^{\ell+1}$ for each integer $0 \leq \ell \leq n-1$, as desired. Conversely, if there exists an integer $1 \leq i \leq r$ such that $[G : Z_G(x_i)]$ is not divisible by p, then we must have that $Z_G(x_i)$ is divisible by p^n because Lagrange's Theorem yields that $p^n m = |G| = [G : Z_G(x_i)]|Z_G(x_i)|$. We may now apply our inductive hypothesis to $Z_G(x_i)$ to extract a subgroup of $Z_G(x_i)$ of order p^k for each integer $1 \leq k \leq n$. □

Before we are able to prove the second and third of Sylow's Theorems, we must develop some additional machinery and terminology. Particularly, the second of Sylow's Theorems asserts that any pair of Sylow p-subgroups of a finite group G are conjugate in G. We have until now only discussed conjugacy of elements of G, but conjugacy of subgroups of G works as expected on the level of sets. Explicitly, conjugacy forms an equivalence relation on the set of subgroups of G; the equivalence class of H consists of all subgroups K of G such that $K = gHg^{-1}$ for some element $g \in G$ and $gHg^{-1} = \{ghg^{-1} \mid h \in H\}$. Conjugacy of subgroups is well-defined by Exercise 2.8.20.

Proposition 3.15.5. *Consider any finite group G, any prime number p dividing $|G|$, and any Sylow p-subgroup P of G. Every element $g \in G$ of p-power order satisfying that $gPg^{-1} = P$ lies in P.*

Proof. Given any element $g \in G$ such that $gPg^{-1} = P$, we must have that $g \in N_G(P)$ by definition of the normalizer of P in G. By Exercise 3.18.15, it holds that P is a normal subgroup of $N_G(P)$, hence we find that $N_G(P)/P$ is a group by Proposition 3.2.1. Consider the cyclic subgroup $\langle gP \rangle$ of $N_G(P)/P$. By Exercise 3.18.7, the order of $\langle gP \rangle$ divides the order of g, hence it must be a power of p by assumption. By the Fourth Isomorphism Theorem, there exists a subgroup H of G containing P such that $H/P = \langle gP \rangle$. By Lagrange's Theorem, we conclude that $|H| = [H : P]|P| = |\langle gP \rangle| \cdot |P|$ is a power of p. By definition of a Sylow p-subgroup, the order of H must be at most the order of P, hence we conclude that $H = P$; this yields that $\langle gP \rangle = H/P$ is the trivial subgroup of $N_G(P)/P$. Consequently, we find that $gP = e_G P$ so that g is an element of P, as desired. □

Corollary 3.15.6. *Consider any finite group G, any prime number p dividing $|G|$, and any Sylow p-subgroup P of G. Every p-subgroup of the normalizer $N_G(P)$ of P in G is a p-subgroup of P. Particularly, it follows that P is the unique Sylow p-subgroup of $N_G(P)$.*

Proof. By Corollary 3.1.15 and by definition, every element x of a p-subgroup Q of $N_G(P)$ has p-power order and satisfies that $xPx^{-1} = P$. Consequently, by Proposition 3.15.5, we must have that $Q \subseteq P$. Particularly, every Sylow p-subgroup Q of $N_G(P)$ satisfies that $Q \subseteq P$ and $|Q| = |P|$. □

Lemma 3.15.7. *Given any group G with any pair of subgroups H and K, the number of distinct elements in the orbit of K under the action of conjugation by H is equal to $[H : N_G(K) \cap H]$.*

Proof. We leave it as Exercise 3.18.77 to demonstrate that H acts on the subgroups of G via conjugation. Consequently, it suffices to prove that $|H * K| = [H : N_G(K) \cap H]$ for every subgroup K of G. We achieve this by defining a bijection between the orbit of K under the action of conjugation by H and the collection of left cosets of $N_G(K) \cap H$ in H. Given any pair of elements $h, x \in H$, we have that $hKh^{-1} = xKx^{-1}$ if and only if $x^{-1}hKh^{-1}x = K$ if and only if $x^{-1}hK(x^{-1}h)^{-1} = K$ if and only if $x^{-1}h \in N_G(K) \cap H$ if and only if $h(N_G(K) \cap H) = x(N_G(K) \cap H)$ by Proposition 3.1.4. Consequently, we may construct a well-defined injective function $\varphi : H * K \to H/(N_G(K) \cap H)$ by declaring that $\varphi(hKh^{-1}) = h(N_G(K) \cap H)$; this function is also clearly surjective. □

Proof. We will prove the second of Sylow's Theorems for any group G of order $p^n m$ for some prime number p and some positive integers n and m such that $\gcd(p, m) = 1$. By definition, for any Sylow p-subgroup P of G, the order of P is p^n. Consider any p-subgroup Q of G. By Exercise

3.18.77, it follows that Q acts on the conjugates of P in G via conjugation. By Lemma 3.15.7, the number of distinct elements in the orbit of gPg^{-1} under the action of Q by conjugation is equal to $[Q : N_G(gPg^{-1}) \cap Q]$. By Lagrange's Theorem, it follows that $|Q|$ is divisible by $[Q : N_G(gPg^{-1}) \cap Q]$, hence the latter must be a power of p. Even more, we have that $|G| = p^n m = [G : N_G(P)]|N_G(P)|$. Considering that P is a normal subgroup of $N_G(P)$, we conclude that $|N_G(P)|$ is divisible by p^n. Consequently, we obtain the identity $m = (|N_G(P)|/p^n)[G : N_G(P)]$, and our assumption that $\gcd(p, m) = 1$ together with the aforementioned identity imply that $[G : N_G(P)]$ is not divisible by p. Crucially, observe that $[G : N_G(P)] = [G : N_G(P) \cap G]$, hence this quantity is precisely the number of distinct conjugacy classes of P in G under the action of conjugation by G. By Proposition 3.14.7, the action of Q on the conjugates of P in G partitions these G-conjugates of P, i.e.,

$$[G : N_G(P)] = \sum_{i=1}^{r} [Q : N_G(g_i P g_i^{-1}) \cap Q]$$

for some representatives $g_1, \dots, g_r \in G$ of the distinct conjugacy classes of P in G. We note that $[Q : N_G(g_i P g_i^{-1}) \cap Q]$ is a power of p, hence if each of them were a non-trivial power of p, then it would follow that $[G : N_G(P)]$ were divisible by p. Consequently, there exists an integer $1 \leq i \leq n$ such that $x(g_i P g_i^{-1})x^{-1} = g_i P g_i^{-1}$ for all elements $x \in Q$, from which it follows that $(g_i^{-1} x g_i) P (g_i^{-1} x g_i)^{-1} = P$ for all elements $x \in Q$ and $g_i^{-1} Q g_i \subseteq P$ by Proposition 3.15.5. We conclude that $Q \subseteq g_i P g_i^{-1}$ for some element $g_i \in G$. Particularly, if Q is a Sylow p-subgroup of G, then $Q = g_i P g_i^{-1}$. □

Proof. We will prove the third of Sylow's Theorems for any group G of order $p^n m$ for some prime number p and some positive integers n and m such that $\gcd(p, m) = 1$. By the second of Sylow's Theorems, every pair of Sylow p-subgroups of G are conjugate in G. Consequently, the action of G on the conjugates of P in G is transitive, and we conclude that the number of distinct Sylow p-subgroups of G is equal to the number of distinct conjugacy classes of P in G under the action of conjugation by G; this is precisely $[G : N_G(P)]$ by the proof of the second of Sylow's Theorems. By Lagrange's Theorem, this quantity divides $|G| = p^n m$, hence it suffices to demonstrate that the number $n_p(G)$ of distinct Sylow p-subgroups of G does not divide p. Even more, we will prove that $n_p(G) \equiv 1 \pmod{p}$. Consider to this end the action of P on the Sylow p-subgroups of G via conjugation. Clearly, we have that $xPx^{-1} = P$ for all elements $x \in P$ because P is always contained in $N_G(P)$, hence the orbit of P under the action of conjugation by P has size one. On the other hand, for any other Sylow p-subgroup Q of G, we have that $|P * Q| = \#\{xQx^{-1} \mid x \in P\}$ is equal to $[P : N_G(Q) \cap P]$ by Lemma 3.15.7; this quantity must be divisible by a power of p by Lagrange's Theorem. We conclude that the number $n_p(G)$ of Sylow p-subgroups of G is the sum of the sizes of the orbits of the Sylow p-subgroups of G under the action of conjugation by P, and this is precisely the sum of positive powers of p (from the non-trivial orbits) and one (from the orbit of P). □

Proposition 3.15.8. *Given any group G and any prime number p dividing $|G|$ such that G admits a unique Sylow p-subgroup P, we must have that P is a normal subgroup of G.*

Proof. Given any element $g \in G$, consider the bijection $\chi_g : G \to G$ of Example 3.3.13 defined by $\chi_g(x) = gxg^{-1}$. Considering that $|gPg^{-1}| = |P|$, it follows that gPg^{-1} is a Sylow p-subgroup of G so that $gPg^{-1} = P$ by hypothesis; this holds for all elements $g \in G$, hence P is normal in G. □

Proposition 3.15.9. *Given any group G, any prime number p dividing $|G|$, and any pair of distinct Sylow p-subgroups P and Q of G, we must have that $P \cap Q = \{e_G\}$.*

Proof. Considering that $P \cap Q$ is a subgroup of P by Exercise 2.8.22, it follows by Lagrange's Theorem that $|P \cap Q|$ divides $|P| = p$. By assumption that p is a prime number, we conclude that $|P \cap Q| = 1$ or $|P \cap Q| = p$. On the contrary, if it were the case that $|P \cap Q| = p$, then it would be true that $P = P \cap Q = Q$ — a contradiction. We must have that $|P \cap Q| = 1$ and $P \cap Q = \{e_G\}$. □

Once we have Sylow's Theorems at hand, we can begin to understand how they inform the structure of a finite group based on the unique prime factorization of its order. We call a group G **simple** if the only normal subgroups of G are the trivial subgroup $\{e_G\}$ and the group G itself.

Example 3.15.10. We will use Sylow's Theorems to prove that a group G of order $1365 = 3 \cdot 5 \cdot 7 \cdot 13$ cannot be simple. By the third part of Sylow's Theorems, we may make the following observations.

1.) We must have that $n_3 \equiv 1 \pmod 3$ and $n_3 \mid 5 \cdot 7 \cdot 13$ so that $n_3 \in \{1, 7, 13, 5 \cdot 7 \cdot 13\}$.

2.) We must have that $n_5 \equiv 1 \pmod 5$ and $n_3 \mid 3 \cdot 7 \cdot 13$ so that $n_5 \in \{1, 3 \cdot 7, 7 \cdot 13\}$.

3.) We must have that $n_7 \equiv 1 \pmod 7$ and $n_7 \mid 3 \cdot 5 \cdot 13$ so that $n_7 \in \{1, 3 \cdot 5\}$.

4.) We must have that $n_{13} \equiv 1 \pmod{13}$ and $n_{13} \mid 3 \cdot 5 \cdot 7$ so that $n_{13} \in \{1, 3 \cdot 5 \cdot 7\}$.

Observe that if any of these integers is one, then our proof is complete by Proposition 3.15.8. On the contrary, we will assume that none of these integers is one. Consequently, we have that $n_3 \geq 7$, $n_5 \geq 21$, $n_7 = 15$, and $n_{13} = 105$. By Corollary 3.1.15, a Sylow p-subgroup of order p must have exactly $p - 1$ elements of order p. By Proposition 3.15.9, the distinct Sylow p-subgroups of order p intersect trivially, hence we have that $\#\{\text{elements of order } p \text{ in } G\} = (p-1)n_p$ by the Fundamental Counting Principle. We conclude that G admits at least $2 \cdot 7 + 4 \cdot 21 + 6 \cdot 15 + 12 \cdot 105$ elements of order 3, 5, 7, or 13. But this is impossible because $|G| = 1365 < 1448 = 2 \cdot 7 + 4 \cdot 21 + 6 \cdot 15 + 12 \cdot 105$.

Example 3.15.11. We will use Sylow's Theorems to prove that any group G of order 15 must be cyclic; this generalizes the fact that any abelian group of order $15 = 3 \cdot 5$ is cyclic by the Fundamental Theorem of Finite Abelian Groups and Proposition 3.10.2. By the third part of Sylow's Theorems, we must have that $n_3 \equiv 1 \pmod 3$ and $n_3 \mid 5$, hence it follows that $n_3 = 1$. Likewise, we have that $n_5 \equiv 1 \pmod 5$ and $n_5 \mid 3$ so that $n_5 = 1$. We conclude that there exists a unique Sylow 3-subgroup P and a unique Sylow 5-subgroup Q of G of order 3 and 5, respectively; they are cyclic by Proposition 3.1.13, and they must therefore satisfy that $P \cong \mathbb{Z}/3\mathbb{Z}$ and $Q \cong \mathbb{Z}/5\mathbb{Z}$ by Proposition 3.3.20. Even more, by Proposition 3.15.8, P and Q are normal subgroups of G. By Lagrange's Theorem, we must have that $P \cap Q = \{e_G\}$ because $|P \cap Q|$ divides $|P|$ and $|Q|$; these are distinct prime numbers, hence we must have that $|P \cap Q| = 1$. Consequently, we may apply Corollary 3.9.14 to conclude that $G = PQ \cong P \times Q \cong (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/5\mathbb{Z}) \cong \mathbb{Z}/15\mathbb{Z}$ by Proposition 3.10.2.

3.16 The Commutator Subgroup

Until now, we have primarily studied the structure of abelian groups — especially finitely generated abelian groups. Given a non-abelian group G, one might wish to quantify "how far" G is from being abelian. By Exercise 2.8.14 and Proposition 3.1.14, a non-abelian group must have at least six elements. Particularly, we may find elements $g, h \in G$ such that $gh \neq hg$ so that $g^{-1}h^{-1}gh \neq e_G$. We refer to the element $g^{-1}h^{-1}gh$ as the **commutator** of g and h in G. Given any nonempty sets $X, Y \subseteq G$, we may define the **commutator subgroup** $[X, Y] = \langle x^{-1}y^{-1}xy \mid x \in X \text{ and } y \in Y \rangle$ of X and Y generated by the commutators $[x, y]$ for any element $x \in X$ and any element $y \in Y$. Our primary interests here lie with the **commutator subgroup** $[G, G] = \langle g^{-1}h^{-1}gh \mid g, h \in G \rangle$ of G.

Proposition 3.16.1. *Consider any group G and any subgroup H of G.*

1.) *We have that $gh = hg$ if and only if $g^{-1}h^{-1}gh = e_G$ for any elements $g, h \in G$.*

2.) *We have that H is a normal subgroup of G if and only if $[H, G]$ is a subgroup of H.*

3.) *We have that $[G, G]$ is a normal subgroup of G.*

4.) *We have that $G/[G, G]$ is an abelian group.*

5.) *We have that G/H is an abelian group if and only if $[G, G]$ is a subgroup of the normal subgroup H. Put another way, the quotient group $G/[G, G]$ is the "largest" abelian quotient of G; thus, the "larger" that $[G, G]$ is with respect to inclusion, the "less abelian" the group G is.*

6.) *Every group homomorphism $\varphi : G \to A$ from G to an abelian group A must* **factor through** *the commutator subgroup of G, i.e., $[G, G]$ must be a subgroup of* $\ker \varphi$. *Even more, there exists a group homomorphism $\psi : G/[G, G] \to A$ such that $\varphi = \psi \circ \pi$, where $\pi : G \to G/[G, G]$ is the canonical surjection. Put another way, we obtain the following* **commutative diagram**.

$$\begin{tikzcd}
G \arrow[r, "\pi"] \arrow[dr, "\varphi"'] & G/[G,G] \arrow[d, dashed, "\psi"] \\
& A
\end{tikzcd}$$

Proof. 1.) We have that $gh = hg$ if and only if $h^{-1}gh = g$ if and only if $g^{-1}h^{-1}gh = e_G$.

2.) By Proposition 3.2.1, we have that H is a normal subgroup of G if and only if $g^{-1}hg \in H$ for all elements $g \in G$ and all elements $h \in H$. Consequently, if H is a normal subgroup of G, then for any commutator $h^{-1}g^{-1}hg \in [H, G]$, we have that $g^{-1}hg \in H$ so that $h^{-1}g^{-1}hg$ is in H and $[H, G]$ is a subgroup of H. Conversely, if $[H, G]$ is a subgroup of H, then every element $h^{-1}g^{-1}hg \in [H, G]$ can be written as $h^{-1}g^{-1}hg = k$ for some element $k \in H$. But this implies that $g^{-1}hg = h(h^{-1}g^{-1}hg) = hk \in H$ for all elements $g \in G$ and all elements $h \in H$, i.e., H is normal.

3.) We must establish that $g^{-1}(x^{-1}y^{-1}xy)g \in [G, G]$ for all elements $g, x, y \in G$. Considering that $(g^{-1}xg)^{-1} = g^{-1}x^{-1}g$ holds for all elements $g, x \in G$, we have that

$$g^{-1}(x^{-1}y^{-1}xy)g = (g^{-1}x^{-1}g)(g^{-1}y^{-1}g)(g^{-1}xg)(g^{-1}yg) = (g^{-1}xg)^{-1}(g^{-1}yg)^{-1}(g^{-1}xg)(g^{-1}yg)$$

is an element of $[G, G]$. By Proposition 3.2.1, $[G, G]$ is a normal subgroup of G.

4.) By Proposition 3.2.1 and third part of the present proposition, we have that $G/[G, G]$ is a group with respect to the binary operation $(g[G, G])(h[G, G]) = gh[G, G]$. Given any elements $g[G, G]$ and $h[G, G]$ of $G/[G, G]$, observe that $g^{-1}h^{-1}gh$ lies in $[G, G]$, hence we have that

$$(g[G, G])^{-1}(h[G, G])^{-1}(g[G, G])(h[G, G]) = g^{-1}h^{-1}gh[G, G] = e_G[G, G].$$

We conclude that $(g[G,G])(h[G,G]) = (h[G,G])(g[G,G])$ so that $G/[G,G]$ is abelian.

5.) By Proposition 3.2.1, if G/H is an abelian group, then H is a normal subgroup of G such that $xyH = (xH)(yH) = (yH)(xH) = yxH$ for all elements $x, y \in G$. We conclude that $x^{-1}y^{-1}xyH = e_G H$ for all elements $x, y \in G$, hence we find that $x^{-1}y^{-1}xy \in H$ for all elements $x, y \in G$ by Proposition 3.1.4. By definition of $[G,G]$, we conclude that $[G,G]$ is a subgroup of H. Conversely, if $[G,G]$ is a subgroup of H, then for any elements $g \in G$ and $h \in H$, we have that $h^{-1}g^{-1}hg \in H$; thus, it follows that $g^{-1}hg \in H$ for all elements $g \in G$ and $h \in H$ so that H is normal in G and G/H is a group. Even more, for any elements $x, y \in G$, we have that $x^{-1}y^{-1}xy \in H$ so that $e_G H = x^{-1}y^{-1}xyH = (xH)^{-1}(yH)^{-1}(xH)(yH)$. By rearranging this identity, we find that $(yH)(xH) = (xH)(yH)$ for all elements $x, y \in G$ so that G/H is abelian.

6.) Given any group homomorphism $\varphi : G \to A$ from G to an abelian group A, we have that

$$\varphi(g^{-1}h^{-1}gh) = \varphi(g^{-1})\varphi(h^{-1})\varphi(g)\varphi(h) = \varphi(g)^{-1}\varphi(g)\varphi(h)^{-1}\varphi(h) = e_A.$$

We conclude that $[G,G]$ is a subgroup of $\ker \varphi$. Consider the function $\psi : G/[G,G] \to A$ defined by $\psi(g[G,G]) = \varphi(g)$. Observe that $g[G,G] = h[G,G]$ if and only if $h^{-1}g[G,G] = e_G[G,G]$ if and only if $h^{-1}g \in [G,G]$ if and only if $h^{-1}g = x^{-1}y^{-1}xy$ for some elements $x, y \in G$ so that

$$\varphi(h)^{-1}\varphi(g) = \varphi(h^{-1})\varphi(g) = \varphi(h^{-1}g) = \varphi(x^{-1}y^{-1}xy) = e_A,$$

hence $\varphi(g) = \varphi(h)$ and ψ is well-defined. By hypothesis that φ is a group homomorphism, it follows that ψ is a group homomorphism, and it follows directly that $\varphi = \psi \circ \pi$ by definition of π. □

Generally, it is a difficult (and perhaps tedious) task to determine the commutator subgroup of a non-abelian group; however, there are several known cases that the interested reader can investigate.

Theorem 3.16.2. *Given any positive integer n, the commutator subgroup of the symmetric group on n letters is the alternating group on n letters, i.e., we have that $[\mathfrak{S}_n, \mathfrak{S}_n] = \mathfrak{A}_n$.*

Theorem 3.16.3. *Given any integer $n \geq 3$, the commutator subgroup of the general linear group of size n over $\mathbb{R}$ is the special linear group of size n over $\mathbb{R}$, i.e., $[\mathrm{GL}(n,\mathbb{R}), \mathrm{GL}(n,\mathbb{R})] = \mathrm{SL}(n,\mathbb{R})$.*

We conclude this section by directing the reader to Exercise 3.18.91 for additional practice.

3.17 Semidirect Products

Earlier in this chapter, we studied the external direct product $H \times K$ of two groups H and K. We found that $H \times K$ is a group with operation given componentwise by $(h_1, k_1)(h_2, k_2) = (h_1h_2, k_1k_2)$. By Proposition 3.9.14, if H and K are normal subgroups of some group G such that $H \cap K = \{e_G\}$ and $G = HK$, then we must have that $G \cong H \times K$. Unfortunately, if only one of H or K is normal in G, then we cannot say that $G \cong H \times K$; however, we can still say that HK is a subgroup of G by Exercise 3.18.11. Even worse, if neither H nor K is normal in G, then we cannot say much of anything at all. Consequently, it is natural to wonder if there exist a group G such that H is a normal subgroup of G and K is a (not necessarily normal) subgroup of G and $H \cap K = \{e_G\}$.

Before we answer this question in the affirmative, let us outline what we already know. Given a group G with a normal subgroup H and a subgroup K, we have that HK is a subgroup of G;

thus, for any two elements $h_1k_1, h_2k_2 \in HK$, we have that $(h_1k_1)(h_2k_2) \in HK$. We may write the element $h_1k_1h_2k_2$ in the form h_3k_3 for some elements $h_3 \in H$ and $k_3 \in K$ by observing that

$$h_1k_1h_2k_2 = h_1k_1h_2k_1^{-1}k_1k_2 = h_1(k_1h_2k_1^{-1})(k_1k_2)$$

and using that H is normal in G to find elements $h_3 = h_1(k_1h_2k_1^{-1}) \in H$ and $k_3 = k_1k_2 \in K$.

By Proposition 3.2.1 and Section 3.6, conjugation of H by any element $k_1 \in K$ yields a well-defined group automorphism $\chi_{k_1} : H \to H$ defined by $\chi_{k_1}(h) = k_1hk_1^{-1}$ that in turn induces a group homomorphism from $\varphi : K \to \operatorname{Aut}(H)$ that sends an element $k_1 \in K$ to its conjugation function $\chi_{k_1} \in \operatorname{Aut}(H)$. By rewriting the above displayed equation in terms of χ_{k_1}, we find that

$$(h_1k_1)(h_2k_2) = (h_1\chi_{k_1}(h_2))(k_1k_2).$$

Observe that this defines a binary operation on HK intrinsically in terms of H and K.

Definition 3.17.1. Given any groups H and K with a group homomorphism $\varphi : K \to \operatorname{Aut}(H)$, we define the **semidirect product** of H and K to be the set of ordered pairs in $H \times K$ endowed with the binary operation outlined in the above paragraph. Put another way, we define the semidirect product of H and K to be the following set endowed with the prescribed binary operation.

$$H \rtimes_\varphi K = \{(h, k) \mid h \in H,\, k \in K, \text{ and } (h_1, k_1)(h_2, k_2) = (h_1\varphi(k_1)(h_2), k_1k_2)\}$$

Under this binary operation, the semidirect product of groups is itself a group. Even more, it has the same order as the external direct product of the underlying groups, and it admits isomorphic copies of the underlying groups as subgroups — one of which is a normal subgroup.

Proposition 3.17.2. *Consider any groups H and K and a group homomorphism $\varphi : K \to \operatorname{Aut}(H)$.*

1.) *We have that $H \rtimes_\varphi K$ is a group of order $|H \times K|$.*

2.) *We have that $H_\varphi = \{(h, e_K) \mid h \in H\}$ and $K_\varphi = \{(e_H, k) \mid k \in K\}$ are subgroups of $H \rtimes_\varphi K$ that are isomorphic to the underlying groups H and K, i.e., $H_\varphi \cong H$ and $K_\varphi \cong K$.*

3.) *We have that H_φ is a normal subgroup of $H \rtimes_\varphi K$ such that $H_\varphi \cap K_\varphi = \{e_{H \rtimes_\varphi K}\}$.*

4.) *Given any elements $(h, e_K) \in H_\varphi$ and $(e_H, k) \in K_\varphi$, we have that*

$$(e_H, k)(h, e_K)(e_H, k)^{-1} = (\varphi(k)(h), e_K).$$

Particularly, the centralizer of H_φ in K_φ is given by $\ker \varphi$.

Proof. 1.) Clearly, we may define a bijective function $f : H \times K \to H \rtimes K$ such that $f(h, k) = (h, k)$, hence the order of $H \rtimes K$ is equal to the order of $H \times K$. Considering that $\varphi(K)$ is a subgroup of $\operatorname{Aut}(H)$ by Proposition 3.3.5, it follows that $\varphi(k)(h) \in H$ for all elements $k \in K$. Consequently, we have that $h_1\varphi(k_1)(h_2) \in H$ for all elements $h_1, h_2 \in H$ and $k_1 \in K$ by hypothesis that H is a group. Likewise, we have that $k_1k_2 \in K$ by hypothesis that K is a group. We conclude that $(h_1, k_1)(h_2, k_2) = (h_1\varphi(k_1)(h_2), k_1k_2)$ is a binary operation on $H \rtimes_\varphi K$. Observe that for every element $h \in H$ and $k \in K$, we have that $(e_H, e_K)(h, k) = (e_H\varphi(e_K)(h), e_Kk) = (\operatorname{id}_H(h), k) = (h, k)$

and $(h, k)(e_H, e_K) = (h\varphi(k)(e_H), ke_K) = (he_H, k) = (h, k)$ because any automorphism of H must send e_H to itself. Consequently, the identity element of $H \rtimes_\varphi K$ is simply (e_H, e_K). Given any element (h, k) of $H \rtimes_\varphi K$, its two-sided inverse is given by $(\varphi(k)^{-1}(h^{-1}), k^{-1})$: indeed, we have that $(h, k)(\varphi(k)^{-1}(h^{-1}), k^{-1}) = (h\varphi(k)(\varphi(k)^{-1}(h^{-1})), kk^{-1}) = (hh^{-1}, e_K) = (e_H, e_K)$ and

$$\begin{aligned}(\varphi(k)^{-1}(h^{-1}), k^{-1})(h, k) = (\varphi(k)^{-1}(h^{-1})\varphi(k^{-1})(h), k^{-1}k) &= (\varphi(k)^{-1}(h^{-1})\varphi(k)^{-1}(h), e_K) \\ &= (\varphi(k)^{-1}(h^{-1}h), e_K) = (\varphi(k)^{-1}(e_H), e_K) \\ &= (e_H, e_K)\end{aligned}$$

because φ is a group homomorphism, hence $\varphi(k^{-1}) = \varphi(k)^{-1}$ for all elements $k \in K$. Proving that this binary operation is associative is just a (tedious) matter of writing out the details.

2.) Considering that H_φ and K_φ both contain the identity (e_H, e_K) of $H \rtimes_\varphi K$, they are nonempty. Given any pair of elements (h_1, e_K) and (h_2, e_K) of H_φ, observe that

$$(h_1, e_K)(h_2, e_K)^{-1} = (h_1, e_K)(\varphi(e_K)^{-1}(h_2^{-1}), e_K^{-1}) = (h_1, e_K)(h_2^{-1}, e_K) = (h_1\varphi(e_K)(h_2^{-1}), e_Ke_K),$$

hence $(h_1, e_K)(h_2, e_K)^{-1}$ is in H_φ, as its second component is e_K. By the One-Step Subgroup Test, H_φ is a subgroup of $H \rtimes_\varphi K$. Given any pair of elements (e_H, k_1) and (e_H, k_2) of K_φ, we have that

$$(e_H, k_1)(e_H, k_2)^{-1} = (e_H, k_1)(\varphi(k_1)^{-1}(e_H^{-1}), k_2^{-1}) = (e_H, k_1)(e_H, k_2^{-1}) = (e_H\varphi(k_1)(e_H), k_1k_2^{-1})$$

so that $(e_H, k_1)(e_H, k_2)^{-1}$ is in K_φ, as its first component is e_H. Once again, appealing to the One-Step Subgroup Test, we conclude that K_φ is a subgroup of $H \rtimes_\varphi K$. Consider the surjective function $\eta : H \to H_\varphi$ defined by $\eta(h) = (h, e_K)$. Given any elements h_1, h_2 of H, we have that

$$\eta(h_1h_2) = (h_1h_2, e_K) = (h_1\varphi(e_K)(h_2), e_Ke_K) = (h_1, e_K)(h_2, e_K) = \eta(h_1)\eta(h_2),$$

hence η is a group homomorphism; it is injective because it is easy to see that $\ker \eta = \{e_H\}$, hence η is a bijective group homomorphism. We conclude that $H \cong H_\varphi$. By an analogous argument applied to the surjective function $\kappa : K \to K_\varphi$ defined by $\kappa(k) = (e_H, k)$, we conclude that $K \cong K_\varphi$.

3.) Given any element $(h_1, k_1) \in H \rtimes_\varphi K$ and any element $(h, e_K) \in H_\varphi$, we have that

$$\begin{aligned}(h_1, k_1)(h, e_K)(h_1, k_1)^{-1} &= (h_1\varphi(k_1)(h), k_1e_K)(\varphi(k_1)^{-1}(h_1^{-1}), k_1^{-1}) \\ &= (h_1\varphi(k_1)(h)\varphi(k_1e_K)(\varphi(k_1)^{-1}(h_1^{-1})), k_1e_Kk_1^{-1}) \\ &= (h_1\varphi(k_1)(h)\varphi(k_1)(\varphi(k_1^{-1})(h_1^{-1})), e_K)\end{aligned}$$

is an element of H_φ and H_φ is a normal subgroup of $H \rtimes_\varphi K$ by Proposition 3.2.1. Observe that $(h, k) \in H_\varphi \cap K_\varphi$ if and only if $(h, k) = (h', e_K)$ and $(h, k) = (e_H, k')$ for some elements $h' \in H$ and $k' \in K$ if and only if $h = e_H$ and $k = e_K$, hence we conclude that $H_\varphi \cap K_\varphi = \{(e_H, e_K)\} = \{e_{H \rtimes_\varphi K}\}$.

4.) Given any ordered pair $(h, k) \in H \times K$, we have that

$$\begin{aligned}(e_H, k)(h, e_K)(e_H, k)^{-1} &= (e_H\varphi(k)(h), ke_K)(\varphi(k)^{-1}(e_H^{-1}), k^{-1}) \\ &= (e_H\varphi(k)(h)\varphi(k)(\varphi(k^{-1})(e_H^{-1})), e_K) = (\varphi(k)(h), e_K).\end{aligned}$$ □

Our next proposition illustrates to what extent a semidirect product is not an external direct product. Essentially, a semidirect product is almost never isomorphic to the external direct product.

Proposition 3.17.3. *Given any groups H and K with a group homomorphism $\varphi : K \to \operatorname{Aut}(H)$, the following properties of the semidirect product $H \rtimes_\varphi K$ are equivalent.*

(i.) *The set-theoretic identity function $\iota : H \rtimes_\varphi K \to H \times K$ is a group isomorphism.*

(ii.) *The group homomorphism $\varphi : K \to \operatorname{Aut}(H)$ is trivial, i.e., we have that $\varphi(k)(h) = h$ for all elements $h \in H$ and $k \in K$, i.e., $\varphi(k)$ is the identity automorphism for all elements $k \in K$.*

(iii.) *We have that K_φ is a normal subgroup of $H \rtimes_\varphi K$.*

Proof. Observe that if the set-theoretic identity function $\iota : H \rtimes_\varphi K \to H \times K$ defined by $\iota(h, k) = (h, k)$ is a group isomorphism, then for all elements $(h_1, k_1), (h_2, k_2) \in H \times K$, we have that

$$\underbrace{(h_1h_2, k_1k_2) = (h_1, k_1)(h_2, k_2)}_{\text{the binary operation on } H\times K} = \underbrace{\iota((h_1, k_1)(h_2, k_2)) = \iota(h_1\varphi(k_1)(h_2), k_1k_2)}_{\text{the binary operation on } H\rtimes_\varphi K} = (h_1\varphi(k_1)(h_2), k_1k_2).$$

Comparing the left-hand side to the right-hand side and using the cancellative property of H, we find that $h_2 = \varphi(k_1)(h_2)$ for all elements $h_2 \in H$ and all elements $k_1 \in K$, as desired.

Likewise, if φ is trivial, then for any elements $(h_1, k_1) \in H \rtimes_\varphi K$ and $(e_H, k) \in K_\varphi$, we have that

$$\begin{aligned}(h_1, k_1)(e_H, k)(h_1, k_1)^{-1} &= (h_1\varphi(k_1)(e_H), k_1k)(\varphi(k_1)^{-1}(h_1^{-1}), k_1^{-1})\\ &= (h_1, k_1k)(h_1^{-1}, k_1^{-1})\\ &= (h_1\varphi(k_1k)(h_1^{-1}), k_1kk_1^{-1}) = (h_1h_1^{-1}, k_1kk_1^{-1}) = (e_H, k_1kk_1^{-1}).\end{aligned}$$

Considering that $k_1kk_1^{-1}$ is an element of K, it follows that K_φ is a normal subgroup of $H \rtimes_\varphi K$.

Last, if we assume that K_φ is a normal subgroup of $H \rtimes_\varphi K$, then for all elements $h_1 \in H$ and for any elements $k, k_1 \in K$, there exists an element $k_2 \in K$ such that $(h_1, k_1)(e_H, k)(h_1, k_1)^{-1} = (e_H, k_2)$. Consequently, we have that $(h_1, k_1)(e_H, k) = (e_H, k_2)(h_1, k_1)$ as elements of $H \rtimes_\varphi K$ so that

$$(h_1, kk_1) = (h_1\varphi(k_1)(e_H), kk_1) = (h_1, k_1)(e_H, k) = (e_H, k_2)(h_1, k_1) = (\varphi(k_2)(h_1), k_2k_1).$$

Considering these as elements of the external direct product $H \times K$, we have that $\varphi(k_2)(h_1) = h_1$ and $k = k_2$ by the cancellative property of K. Considering that h_1 and k are arbitrary, it follows that $\varphi(k)$ is the identity automorphism on H for all elements $k \in K$. But this implies that

$$\iota((h_1, k_1)(h_2, k_2)) = \iota(h_1\varphi(k_1)(h_2), k_1k_2) = \iota(h_1h_2, k_1k_2) = (h_1h_2, k_1k_2) = (h_1, k_1)(h_2, k_2)$$

for any pair of elements $(h_1, k_1), (h_2, k_2) \in H \times K$, hence ι is a group isomorphism. □

Example 3.17.4. Consider the semidirect product of $H = \mathbb{Z}_3$ and $K = \mathbb{Z}_4$ with respect to the group homomorphism $\varphi : \mathbb{Z}_4 \to \operatorname{Aut}(\mathbb{Z}_3)$ defined by $\varphi(n + 4\mathbb{Z}) = \nu_n$ and $\nu_n : \mathbb{Z}_3 \to \mathbb{Z}_3$ is the **inversion** automorphism defined by $\nu(k + 3\mathbb{Z}) = (-1)^n k + 3\mathbb{Z}$. We will prove that $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ has a cyclic Sylow 2-subgroup; then, we will deduce that $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ is not isomorphic to the alternating group $\mathfrak{A}_4$ on four letters or the dihedral group D_6 on on 12 elements.

By Proposition 3.17.2, we have that $|\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4| = |\mathbb{Z}_3| \cdot |\mathbb{Z}_4| = 3 \cdot 4 = 2^2 \cdot 3$ so that $(\mathbb{Z}_4)_\varphi \cong \mathbb{Z}_4$ is a cyclic subgroup of order four, i.e., a cyclic Sylow 2-subgroup of order four. Considering that $\mathfrak{A}_4$ does not have any elements of order four by Exercise 3.18.39, it follows that $\mathfrak{A}_4$ does not have a cyclic subgroup of order four so that $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ cannot be isomorphic to $\mathfrak{A}_4$ by Proposition 3.3.17.

On the other hand, we will demonstrate that D_6 has at least three elements of order two and $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ has only one element of order two, so these groups cannot be isomorphic. Observe that the elements of D_6 are of the form $r^i s^j$ for some integers $0 \leq i \leq 5$ and $0 \leq j \leq 1$ with $srs = r^{-1}$. Evidently, we have that s, r^3, and rs are all elements of order two. Each element of $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ looks like $(a + 3\mathbb{Z}, b + 4\mathbb{Z})$ and satisfies that $(a+3\mathbb{Z}, b+4\mathbb{Z})(a+3\mathbb{Z}, b+4\mathbb{Z}) = (a + (-1)^b a + 3\mathbb{Z}, 2b + 4\mathbb{Z})$; thus, an element of $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ has order two if and only if $3 \mid (a + (-1)^b a)$ and $4 \mid 2b$. Considering that $4 \mid 2b$ if and only if $2 \mid b$, it follows that $b + 4\mathbb{Z} = 0 + 4\mathbb{Z}$ or $b + 4\mathbb{Z} = 2 + 4\mathbb{Z}$. Either way, we have that $a + (-1)^b a = 2a$, from which it follows that $3 \mid 2a$ if and only if $3 \mid a$ if and only if $a + 3\mathbb{Z} = 0 + 3\mathbb{Z}$. Consequently, the only element of order two in $\mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_4$ is $(0 + 3\mathbb{Z}, 2 + 4\mathbb{Z})$.

Example 3.17.5. Given any group H, we define the **holomorph** of H to be the semidirect product of $H \rtimes_\iota \operatorname{Aut}(H)$ with respect to the identity isomorphism $\iota : \operatorname{Aut}(H) \to \operatorname{Aut}(H)$, i.e., we have that $\operatorname{Hol}(H) = H \rtimes_\iota \operatorname{Aut}(H)$. We will demonstrate that for the Klein four-group $H = \mathbb{Z}_2 \times \mathbb{Z}_2$, we have that $\operatorname{Hol}(H) \cong \mathfrak{S}_4$. By Exercise 3.18.33, it follows that $\operatorname{Aut}(H) \cong \mathfrak{S}_4$, from which we conclude that $|\operatorname{Hol}(H)| = |H| \times |\operatorname{Aut}(H)| = 4 \cdot 3! = 4!$ by Proposition 3.17.2. Consequently, by Exercise 1.10.5, in order to prove that $\operatorname{Hol}(H) \cong \mathfrak{S}_4$, it suffices to find an injective group homomorphism from $\operatorname{Hol}(H)$ to $\mathfrak{S}_4$. By Exercise 3.18.79, if the action of $\operatorname{Hol}(H)$ on the left cosets of its subgroup $K = \operatorname{Aut}(H)_\iota$ is faithful, then we will obtain an injective group homomorphism $\sigma : \operatorname{Hol}(H) \to \mathfrak{S}_{G/K}$. By Lagrange's Theorem, we will ultimately conclude that $\mathfrak{S}_{G/K} \cong \mathfrak{S}_4$ so that $\operatorname{Hol}(H) \cong \mathfrak{S}_4$.

Explicitly, we may define a group action of $\operatorname{Hol}(H)$ on the left cosets of K in $\operatorname{Hol}(H)$ by declaring that for any element $g \in \operatorname{Hol}(H)$ and any left coset xK, we have that $g * xK = gxK$. Observe that for any element $g \in \operatorname{Hol}(H)$, we have that $gxK = g * xK = xK$ for all elements $xK \in \operatorname{Hol}(H)/K$ if and only if $x^{-1}gx \in K$ for all elements $x \in \operatorname{Hol}(H)$ by Proposition 3.1.4. Particularly, this inclusion holds for the identity element of $\operatorname{Hol}(H)$, hence we find that $g \in K$. Even more, for every element $x \in H_\iota$, we must have that $x^{-1}gx \in K$ so that $x^{-1}gxg^{-1} \in K$ and $x^{-1}gxg^{-1} \in H_\iota$ because H_ι is a normal subgroup of $\operatorname{Hol}(H)$, i.e., $gxg^{-1} \in H_\iota$ for all elements $g \in \operatorname{Hol}(H)$. By Proposition 3.17.2, it follows that $H_\iota \cap K$ is the trivial subgroup of $\operatorname{Hol}(H)$, hence we conclude that $x^{-1}gxg^{-1}$ is trivial; this implies that $gx = xg$ for all elements $x \in H_\iota$ so that every element of the kernel of this action lies in the centralizer of H_ι in K. Considering that the centralizer of H_ι in K is $\ker \iota$ by Proposition 3.17.2, we conclude that g is trivial, hence the group action is faithful, as desired.

We conclude this section with a useful result about the center of a semidirect product.

Proposition 3.17.6. *Consider the semidirect product of $H \rtimes_\varphi K$ of some groups H and K with respect to some group homomorphism $\varphi : K \to \operatorname{Aut}(H)$. We have that*

$$[Z(H) \cap \operatorname{Fix}(\varphi(K))] \times [Z(K) \cap \ker \varphi] \subseteq Z(H \rtimes_\varphi K),$$

where $\operatorname{Fix}(\varphi(K))$ is the set of elements in H that are fixed by all automorphisms of $\varphi(K)$.

Proof. Given any elements $h \in Z(H) \cap \operatorname{Fix}(\varphi(K))$ and $k \in Z(K) \cap \ker \varphi$, we have that

$$(h,k)(h_1,k_1) = (h\varphi(k)(h_1), kk_1) = (hh_1, kk_1) = (h_1 h, k_1 k) = (h_1 \varphi(k_1)(h), k_1 k) = (h_1,k_1)(h,k)$$

for any element $(h_1, k_1) \in H \rtimes_\varphi K$. Explicitly, we have that $\varphi(k)(h_1) = h_1$ because k is in $\ker \varphi$, i.e., $\varphi(k)$ is the identity automorphism on H; $hh_1 = h_1 h$ and $kk_1 = k_1 k$ by assumption that $h \in Z(H)$ and $k \in Z(K)$; and $h = \varphi(k_1)(h)$ because h is in $\operatorname{Fix}(\varphi(K))$, i.e., h is fixed by all automorphisms. □

3.18 Chapter 3 Exercises

3.18.1 Cosets and Lagrange's Theorem

Exercise 3.18.1. Prove that (i.) $\implies$ (ii.) $\implies$ (iii.) $\implies$ (iv.) of Proposition 3.1.4 hold.

Exercise 3.18.2. Use the One-Step Subgroup Test to establish that the rational numbers $\mathbb{Q}$ form a subgroup of the additive group $(\mathbb{R}, +)$ of real numbers; then, prove that $[\mathbb{R} : \mathbb{Q}]$ is infinite.

Exercise 3.18.3. Let G be a group. Prove that if H is a subgroup of G such that $[G : H] = 2$, then it holds that $gH = Hg$ for all elements $g \in G$.

Exercise 3.18.4. Let G be a group with subgroups H and K. Consider the set

$$HK = \{hk \mid h \in H \text{ and } k \in K\}$$

of Exercise 2.8.23. Prove that if H and K are finite, then $|HK| = \dfrac{|H| \cdot |K|}{|H \cap K|}$.

Exercise 3.18.5. Let G be a cyclic group of order n. By Exercise 2.8.34, for each positive integer $d \mid n$, there exists a cyclic subgroup of G of order d. Complete the following steps to demonstrate that n is the sum of Euler's totient function over its positive divisors, i.e., we have that $n = \sum_{d \mid n} \phi(d)$ (cf. the paragraph preceding Exercise 2.8.28 for the definition of Euler's totient function).

(a.) Prove that every element of G lies in one and only one cyclic subgroup of order $d \mid n$. Conclude that the cyclic subgroup of G of order $d \mid n$ is unique, i.e., these subgroups partition G.

(b.) Prove that for each positive integer $d \mid n$, there exist $\phi(d)$ generators for the cyclic subgroup of G generated by d. Conclude that the cyclic subgroup of order d contains exactly $\phi(d)$ elements.

Conclude the desired result by combining parts (a.) and (b.) above.

Exercise 3.18.6. Consider Euler's totient function $\phi : \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 1}$ defined before Exercise 2.8.28. Complete the following steps to prove that $\phi(n)$ is an even integer for all integers $n \geq 2$.

(a.) Conclude by Exercise 2.8.28 that $\phi(n) = |\mathbb{Z}_n^*|$ for every positive integer n.

(b.) Conclude by Exercise 2.8.26 that there exists an element of order two in $\mathbb{Z}_n^*$.

(c.) Conclude the desired result by Lagrange's Theorem.

3.18.2 Quotient Groups and Normal Subgroups

Exercise 3.18.7. Let G be a group with a normal subgroup H. Prove that $\operatorname{ord}(gH) \mid \operatorname{ord}(g)$ for every element $g \in G$. Conclude that $\max\{\operatorname{ord}(gH) \mid g \in G\} \leq \max\{\operatorname{ord}(g) \mid g \in G\}$.

Exercise 3.18.8. Let G be a group with a subgroup H. Prove that if the index of H in G satisfies that $[G : H] = 2$, then H is a normal subgroup of G.

Exercise 3.18.9. Exhibit a non-cyclic group G and a normal subgroup H of G such that H and G/H are cyclic. Conclude that the converse of the first statement of Proposition 3.2.6 is false.

(**Hint:** By Corollary 3.1.13 and Exercise 3.18.8, it suffices to find a non-abelian group G of order $4 = 2 \cdot 2$ and any subgroup H of G of order two. We have already encountered one.)

Exercise 3.18.10. Exhibit a non-abelian group G and a normal subgroup H of G such that H and G/H are abelian. Conclude that the converse of the second statement of Proposition 3.2.6 is false.

(**Hint:** By Corollary 3.1.14 and Exercise 3.18.8, it suffices to find a non-abelian group G of order $6 = 2 \cdot 3$ and any subgroup H of G of order three. We have already encountered one.)

Exercise 3.18.11. Let G be a group with subgroups H and K. Consider the set

$$HK = \{hk \mid h \in H \text{ and } k \in K\}$$

of Exercises 2.8.23 and 3.18.4. Prove that if H is normal in G or K is normal in G, then HK is a subgroup of G. Even more, prove that if H and K are both normal in G, then HK is normal in G.

Exercise 3.18.12. Let G be a group with a subgroup H and a normal subgroup N. Prove that $H \cap N$ is a normal subgroup of H. Conclude that $H/(H \cap N)$ is a group.

Exercise 3.18.13. Let G be a group. Prove that if H is a subgroup of G such that no other subgroup of G has the same order as H, then H must be a normal subgroup of G.

(**Hint:** Consider the function $\chi_g : H \to gHg^{-1}$ defined by $\chi_g(h) = ghg^{-1}$. Use Exercise 1.10.6(a.).)

Exercise 3.18.14. Let G be a group. Let $Z(G) = \{x \in G \mid gx = xg \text{ for all elements } g \in G\}$ denote the center of G. By Exercise 2.8.18, we note that $Z(G)$ is a subgroup of G.

(a.) Prove that $Z(G)$ is a normal subgroup of G.

(b.) Prove that if $G/Z(G)$ is cyclic, then G is abelian.

Exercise 3.18.15. Let G be a group. Let $N_G(H) = \{g \in G \mid gHg^{-1} = H\}$ denote the normalizer of any subgroup $H \subseteq G$. By Exercise 2.8.21, we note that $N_G(H)$ is a subgroup of G.

(a.) Prove that H is a normal subgroup of $N_G(H)$.

(b.) Prove that if K is a subgroup of G and H is a normal subgroup of K, then K is a subgroup of $N_G(H)$. Conclude that $N_G(H)$ is the "largest" subgroup of G with H as a normal subgroup.

Given any element $h \in H$, consider the centralizer $Z_G(h) = \{g \in G \mid gh = hg\}$ of h in G. By Exercise 2.8.19, $Z_G(x)$ is a subgroup of G. We define the **centralizer** of the subgroup H as the union of the centralizers of all elements $h \in H$ in G, i.e., $Z_G(H) = \{g \in G \mid gh = hg \text{ for all elements } h \in H\}$.

(c.) Prove that $Z_G(H)$ is a subgroup of G.

(d.) Prove that H is a normal subgroup of $Z_G(H)$. Conclude that $Z_G(H)$ is a subgroup of $N_G(H)$.

(e.) Prove that $Z_G(H)$ is a normal subgroup of $N_G(H)$.

(**Hint:** By Proposition 3.2.1, it suffices to prove that for every triple of elements $h \in H$, $x \in N_G(H)$, and $g \in Z_G(H)$, it holds that $(xgx^{-1})h = h(xgx^{-1})$. Use the fact that for every element $x \in N_G(H)$, there exists an element $k \in H$ such that $x^{-1}hx = k$ by definition.)

Exercise 3.18.16. Consider the additive group $(\mathbb{Q}, +)$ of rational numbers. Prove that $\mathbb{Z}$ is a normal subgroup of $\mathbb{Q}$; then, prove that every element of the quotient group $\mathbb{Q}/\mathbb{Z}$ has finite order. Conclude that there exists a group of infinite order such that each of its elements has finite order.

(**Hint:** Determine all possible left coset representatives for $0 + \mathbb{Z}$.)

3.18.3 Group Homomorphisms

Exercise 3.18.17. Prove that if G is a group of even order, then there are an odd number of elements of G of order two. Conclude that a group of even order admits a subgroup of order two.

(**Hint:** Prove that the function $\varphi : G \to G$ defined by $\varphi(g) = g^{-1}$ is a bijection. Conclude that for every non-identity element $g \in G$, there exists a non-identity element $h \in G$ such that $g^{-1} = h$. Count the number of non-identity elements of G to deduce the result of the statement.)

Exercise 3.18.18. Prove that if G is an abelian group of odd order, then for each element $g \in G$, there exists a unique element $h \in G$ such that $h^2 = g$. Conclude that in an abelian group of odd order, every element admits a unique square root.

(**Hint:** Prove that the function $\varphi : G \to G$ defined by $\varphi(g) = g^2$ is a homomorphism. Compute its kernel; then, use Lagrange's Theorem to conclude that φ is in fact an isomorphism.)

Exercise 3.18.19. Let $(G, *)$ and $(H, \star)$ be groups. Consider a group homomorphism $\varphi : G \to H$.

(a.) Prove that $\ker \varphi$ is a normal subgroup of G.

(b.) Conversely, suppose that N is a normal subgroup of G. Prove that there exists a group $(K, \cdot)$ and homomorphism $\pi : G \to K$ such that $N = \ker \pi$. Conclude that the normal subgroups of any group G are precisely the kernels of group homomorphisms from G.

(**Hint:** Consider the **canonical surjection** $\pi : G \to G/N$ defined by $\pi(g) = gN$.)

Exercise 3.18.20. Consider the collection $\mathfrak{G}$ of all groups. Prove that $(G, *) \sim (H, \star)$ if and only if there exists an isomorphism $\varphi : (G, *) \to (H, \star)$ is an equivalence relation on $\mathfrak{G}$.

We note that the equivalence classes of the above equivalence relation are called the **isomorphism classes** of groups. If there are n distinct equivalence classes of groups that satisfy a certain property $\mathcal{P}$, then we say that there are n groups that satisfy property $\mathcal{P}$ **up to isomorphism**.

Exercise 3.18.21. Prove that there is only one infinite cyclic group up to isomorphism.

Exercise 3.18.22. Prove that there is only one cyclic group of order n up to isomorphism.

3.18.4 Cayley's Theorem

Exercise 3.18.23. Let G be a group with a subgroup H such that $[G : H]$ is finite. Prove that there exists a normal subgroup N of G such that $N \subseteq H$ and $[G : N] \mid [G : H]!$.

(**Hint:** We note that by Cayley's Theorem and the First Isomorphism Theorem, it suffices to find a group homomorphism from G to the symmetric group on the left cosets G/H of H in G whose kernel is contained in H. Consider the function $\varphi : G \to \mathfrak{S}_{G/H}$ defined by $\varphi(g)(xH) = gxH$.)

Exercise 3.18.24. Let G be a finite group with a subgroup H such that $[G : H]$ is the smallest prime number p that divides $|G|$. Prove that H is a normal subgroup of G.

(**Hint:** We note that by Exercise 3.18.23, there exists a normal subgroup N of G such that $N \subseteq H$ and $[G : N] \mid p!$. Even more, by Lagrange's Theorem, we have that $[G : H]|H| = |G| = [G : N]|N|$, hence we find that $p!|N| = [G : N]|N|q = pq|H|$. Conclude that $|H| \mid |N|$ so that $H = N$.)

3.18.5 The Group Isomorphism Theorems

Exercise 3.18.25. Prove that $(\mathbb{Q}, +)$ and $(\mathbb{Z}, +)$ are not isomorphic.

(**Hint:** Exercise 2.8.24 and Proposition 3.3.18 directly imply this.)

Exercise 3.18.26. Prove that $(\mathbb{Z} \times \mathbb{Z}, +)$ and $(\mathbb{Z}, +)$ are not isomorphic.

(**Hint:** Prove that the function $\varphi : (\mathbb{Z} \times \mathbb{Z}, +) \to (\mathbb{Z}, +)$ defined by $\varphi(a, b) = a - b$ is a surjective group homomorphism with $\ker \varphi = \langle (1, 1) \rangle$. Conclude by the First Isomorphism Theorem that $(\mathbb{Z}, +)$ is isomorphic to a proper quotient of $(\mathbb{Z} \times \mathbb{Z}, +)$, hence the groups cannot be isomorphic.)

Exercise 3.18.27. Let G be a group with a normal subgroup K. Let H be any group such that there exists a group homomorphism $\varphi : G \to H$.

(a.) Prove that if φ is surjective, then $\varphi(K)$ is a normal subgroup of H.

(b.) Prove that if φ is an isomorphism, then we have that $G/K \cong H/\varphi(K)$.

(**Hint:** Consider the function $\psi : G \to H/\varphi(K)$ defined by $\psi(g) = \varphi(g)\varphi(K)$.)

(c.) Conclude that for every group isomorphism $\varphi : G \to G$, we have that $G/K \cong G/\varphi(K)$.

Exercise 3.18.28. Consider the general linear group $\operatorname{GL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) \neq 0\}$ and the special linear group $\operatorname{SL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) = 1\}$. Let $\mathbb{R}^\times$ denote the set of nonzero real numbers. Complete the following steps to prove that $(\operatorname{GL}(2, \mathbb{R})/\operatorname{SL}(2, \mathbb{R}), \cdot) \cong (\mathbb{R}^\times, \cdot)$.

(a.) Prove that the determinant function is a group homomorphism from $\operatorname{GL}(2, \mathbb{R})$ to $(\mathbb{R}^\times, \cdot)$.

(**Hint:** Prove that $\det(AB) = \det(A)\det(B)$ for any real 2×2 matrices A and B.)

(b.) Prove that φ is surjective with $\ker \varphi = \operatorname{SL}(2, \mathbb{R})$.

(c.) Conclude the desired result by the First Isomorphism Theorem.

Exercise 3.18.29. Consider the circle group $\mathbb{T} = \{\cos\theta + i \sin\theta \mid \theta \in \mathbb{R}\}$ under complex multiplication. Complete the following steps to prove that $(\mathbb{R}/\mathbb{Z}, +) \cong (\mathbb{T}, \cdot)$.

(a.) Prove that the function $\varphi : (\mathbb{R}, +) \to (\mathbb{T}, \cdot)$ defined by $\varphi(\theta) = \cos\theta + i\sin\theta$ is a surjective group homomorphism with $\ker\varphi = \{2\pi n \mid n \in \mathbb{Z}\}$.

(b.) Prove that the function $\mu : (\mathbb{R}, +) \to (\mathbb{R}, +)$ defined by $\mu(x) = 2\pi x$ is a group isomorphism.

(c.) Conclude from parts (b.) and (c.) of Exercise 3.18.27 that $(\mathbb{R}/\mathbb{Z}, +) \cong (\mathbb{T}, \cdot)$.

3.18.6 Group Automorphisms

Exercise 3.18.30. Prove that (i.) $\implies$ (ii.) $\implies$ (iii.) $\implies$ (iv.) $\implies$ (i.) in Corollary 3.6.5.

Exercise 3.18.31. Prove that if $\varphi : (\mathbb{Z}, +) \to (\mathbb{Z}, +)$ is a group isomorphism, then we must have that $\varphi(1) = 1$ or $\varphi(1) = -1$. Conclude that $\operatorname{Aut}(\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z}, +)$.

Exercise 3.18.32. Prove that if $\varphi : (\mathbb{Z} \times \mathbb{Z}, +) \to (\mathbb{Z} \times \mathbb{Z}, +)$ is a group isomorphism, then we must have that $\varphi(1, 0) = (1, 0)$ or $\varphi(1, 0) = (-1, 0)$ and $\varphi(0, 1) = (0, 1)$ or $\varphi(0, 1) = (0, -1)$. Conclude that $\operatorname{Aut}(\mathbb{Z} \times \mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z}, +) \times (\mathbb{Z}/2\mathbb{Z}, +)$. Conjecture a formula for $\operatorname{Aut}(\mathbb{Z} \times \cdots \times \mathbb{Z})$ in general.

Exercise 3.18.33. Exhibit an explicit isomorphism $\psi : \mathfrak{S}_3 \to \operatorname{Aut}(\mathbb{Z}_2 \times \mathbb{Z}_2)$.

(**Hint:** Observe that for any group isomorphism $\varphi : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_2$, there are three possibilities for the image of $(1, 0)$ under φ — namely, they are given by $\varphi(1, 0) = (1, 0)$ or $\varphi(1, 0) = (1, 1)$ or $\varphi(1, 0) = (0, 1)$. Each of these gives rise to two possibilities for $\varphi(0, 1)$. Once these two images have been specified, the image of $(1, 1)$ under φ is uniquely determined by $\varphi(1, 1) = \varphi(1, 0) + \varphi(0, 1)$.)

3.18.7 The Symmetric Group on n Letters, Revisited

Exercise 3.18.34. Prove that if τ is a transposition, then $\sigma\tau\sigma^{-1}$ is a transposition.

Exercise 3.18.35. [Hun13, Exercise 7.5.27] Prove that if σ is a $(2n + 1)$-cycle for some positive integer n, then there exists a cycle τ such that $\sigma = \tau^2$.

Exercise 3.18.36. [Hun13, Exercise 7.5.31] Prove that if σ is a product of disjoint cycles of the same length, then σ can be written as a power of a cycle.

Exercise 3.18.37. Compute the number of distinct conjugacy classes in $\mathfrak{S}_5$.

Exercise 3.18.38. Compute the number of distinct 3-cycles in $\mathfrak{S}_5$.

3.18.8 The Alternating Group on n Letters

Exercise 3.18.39. Prove that $\mathfrak{A}_4$ does not contain an element of order four.

Exercise 3.18.40. Prove that $\mathfrak{A}_n$ is generated by all three-cycles on n letters.

Exercise 3.18.41. We say that a group G is **simple** if its only normal subgroups are the trivial subgroup $\{e_G\}$ and the group itself. Prove that $\mathfrak{A}_n$ is simple for $n = 3$ and all integers $n \geq 5$. Conclude that $\mathfrak{A}_5$ is the smallest (in terms of order) non-abelian simple group.

Exercise 3.18.42. Prove that $\mathfrak{A}_4$ has a proper nontrivial normal subgroup, hence $\mathfrak{A}_4$ is not simple.

(**Hint:** Consider the injective function $\varphi : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathfrak{A}_4$ defined by $\varphi(0, 0) = (1)$, $\varphi(1, 0) = (12)(34)$, $\varphi(0, 1) = (13)(24)$, and $\varphi(1, 1) = (14)(23)$. Prove that φ is a group homomorphism; then, conclude by the First Isomorphism Theorem that $\varphi(\mathbb{Z}_2 \times \mathbb{Z}_2)$ is a proper nontrivial normal subgroup of $\mathfrak{A}_4$.)

Exercise 3.18.43. We say that a group G is **solvable** if there exist subgroups

$$G_0 = \{e_G\} \subseteq G_1 \subseteq \cdots \subseteq G_{n-1} \subseteq G_n = G$$

such that G_i is a normal subgroup of G_{i+1} and the quotient groups G_{i+1}/G_i are abelian for each integer $0 \leq i \leq n-1$. Prove that $\mathfrak{A}_5$ is the smallest (in terms of order) non-solvable group.

We will eventually come to understand that Exercises 3.18.42 and 3.18.43 combined imply that quartic polynomials can be solved by radicals (i.e., there exists a quartic formula) because $\mathfrak{A}_4$ is not simple, but there is no quintic formula because $\mathfrak{A}_5$ is simple. Even more, polynomials of degree exceeding five are not solvable by radicals (hence the name "solvable") because every symmetric group on $n \geq 5$ letters contains an isomorphic copy of $\mathfrak{A}_5$. We note that this rationale was originated by the French mathematician Évariste Galois, who perished in a duel in 1832 at the age of 20.

3.18.9 External and Internal Direct Products

Exercise 3.18.44. Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. Prove that $G_1 \times \cdots \times G_n$ is abelian if and only if $G_1, \dots, G_n$ are abelian.

Exercise 3.18.45. Let G, H, and K be groups with external direct product $G \times H \times K$. Prove that the external direct product is associative, i.e., we have that $(G \times H) \times K \cong G \times (H \times K)$.

(**Hint:** Prove that $(G \times H) \times K$ and $G \times (H \times K)$ are both isomorphic to $G \times H \times K$.)

Exercise 3.18.46. Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. Prove that for any permutation σ of the indices $1, \dots, n$, we have that $G_1 \times \cdots \times G_n \cong G_{\sigma(1)} \times \cdots \times G_{\sigma(n)}$.

(**Hint:** By Exercise 3.18.45, it suffices to prove that $G \times H \cong H \times G$ for any groups G and H. One can also prove this directly by exhibiting an isomorphism $\varphi : G_1 \times \cdots \times G_n \to G_{\sigma(1)} \times \cdots \times G_{\sigma(n)}$. Or third, it is also possible to proceed by induction on the number n of factors in the product.)

Exercise 3.18.47. Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$.

(a.) Prove that the function $\pi_i : G_1 \times \cdots \times G_n \to G_i$ defined by $\pi_i(g_1, \dots, g_n) = g_i$ is a group homomorphism with kernel $G_1 \times \cdots \times G_{i-1} \times \{e_{G_i}\} \times G_{i+1} \times \cdots \times G_n$.

(b.) Prove that the function $\delta_i : G_1 \times \cdots \times G_n \to G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times \cdots \times G_n$ defined by $\delta_i(g_1, \dots, g_n) = (g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_n)$ is a group homomorphism with kernel

$$\gamma_i(G_i) = \{(e_{G_1}, \dots, g, \dots, e_{G_n}) \mid g \in G_i\}.$$

Exercise 3.18.48. Let $G_1, \dots, G_n$ be groups with external direct product $G_1 \times \cdots \times G_n$. Let H_i be a normal subgroup of G_i for each integer $1 \leq i \leq n$. Prove that the following isomorphism holds.

$$\frac{G_1 \times \cdots \times G_n}{H_1 \times \cdots \times H_n} \cong \frac{G_1}{H_1} \times \cdots \times \frac{G_n}{H_n}$$

(**Hint:** By the First Isomorphism Theorem, the proof reduces to finding a surjective group homomorphism $\varphi : G_1 \times \cdots \times G_n \to (G_1/H_1) \times \cdots \times (G_n/H_n)$ such that $\ker \varphi = H_1 \times \cdots \times H_n$.)

Given any function $f : X \to Y$ and any subset W of X, we may define a function $f|_W : W \to Y$ by declaring that $f|_W(w) = f(w)$. We refer to this function as the **restriction** of f to W.

Exercise 3.18.49. [Hun13, Exercise 8.4.32] Consider a group homomorphism $\varphi : (G, *) \to (H, \star)$ such that there exists a normal subgroup K of G for which the restriction $\varphi|_K : K \to H$ of the function φ to K is a group isomorphism. Complete the following steps to prove that $G \cong K \times \ker \varphi$.

(a.) Prove that the function $\psi : K \times \ker \varphi \to G$ defined by $\psi(k, x) = kx$ is a group homomorphism.

(b.) Prove that (k, x) lies in $\ker \psi$ if and only if $k = e_G$ and $x = e_G$. Conclude that ψ is injective.

(c.) Prove that for every element $g \in G$, there exists an element $k \in K$ such that $\varphi(k^{-1}g) = e_G$. Conclude that for every element $g \in G$, there exist elements $k \in K$ and $x \in \ker \varphi$ such that $g = kx$, hence ψ is surjective; thus, it follows that ψ is a group isomorphism.

Exercise 3.18.50. [Hun13, Exercise 9.2.21] Prove that if G is any abelian group that admits a surjective group homomorphism $\varphi : G \to (\mathbb{Z}, +)$, then there exists a subgroup H of G such that $H \cong (\mathbb{Z}, +)$. Conclude by Exercise 3.18.49 that $G \cong \mathbb{Z} \times \ker \varphi$ in this case.

(**Hint:** Convince yourself that it is possible to find an element $g \in G$ such that g is not in the kernel of φ; then, look at the cyclic subgroup H generated by g, i.e., suppose that $H = \{g^n \mid n \in \mathbb{Z}\}$.)

Exercise 3.18.51. Prove that the dihedral group $D_4 = \{1, r, r^2, r^3, s, rs, r^2s, r^3s\}$ cannot be realized as an internal direct product of any pair of proper non-trivial subgroups of D_4.

(**Hint:** On the contrary, if $D_4 = HK$ for some proper non-trivial subgroups H and K of D_4, then we must have that $|H| = 4$ and $|K| = 2$. Prove that H and K must be abelian.)

Exercise 3.18.52. Consider the dihedral group $D_6 = \{1, r, r^2, r^3, r^4, r^5, s, rs, r^2s, r^3s, r^4s, r^5s\}$.

(a.) Prove that $K = \{1, r^2, r^4, s, r^2s, r^4s\}$ is a subgroup of D_6.

(**Hint:** Every element of K can be written as $r^{2m}s^n$ for some integers m and n. Use this and the fact that $sr = r^5s$ to conclude that K is closed under the binary operation of D_6.)

(b.) Prove that if $H = \{1, r^3\}$, then D_6 is the internal direct product of H and K.

3.18.10 Finite Abelian Groups

Exercise 3.18.53. Compute the elementary divisors and invariant factors of the following groups.

(a.) $\mathbb{Z}_{169}$

(b.) $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$

(c.) $\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \times \mathbb{Z}_8 \times \mathbb{Z}_{13}$

(d.) $\mathbb{Z}_5 \times \mathbb{Z}_{10} \times \mathbb{Z}_{20} \times \mathbb{Z}_{40}$

(e.) $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_6 \times \mathbb{Z}_8$

(f.) $\mathbb{Z}_6 \times \mathbb{Z}_{15} \times \mathbb{Z}_{25} \times \mathbb{Z}_{30}$

Exercise 3.18.54. Compute the invariant factor decomposition of each of the following groups.

(a.) $\mathbb{Z}_2 \times \mathbb{Z}_9 \times \mathbb{Z}_{16} \times \mathbb{Z}_{25}$

(b.) $\mathbb{Z}_3 \times \mathbb{Z}_6 \times \mathbb{Z}_9 \times \mathbb{Z}_{12}$

Exercise 3.18.55. Compute the elementary divisor decomposition of each of the following groups.

(a.) $\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_{12} \times \mathbb{Z}_{84}$

(b.) $\mathbb{Z}_{10} \times \mathbb{Z}_{20} \times \mathbb{Z}_{60} \times \mathbb{Z}_{120}$

Exercise 3.18.56. Compute the number of distinct finite abelian groups with the following orders.

(a.) $2022 = 2 \cdot 3 \cdot 337$

(b.) $44100 = 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^2$

(c.) $75600 = 2^4 \cdot 3^3 \cdot 5^2 \cdot 7$

(d.) $100000 = 2^5 \cdot 5^5$

Exercise 3.18.57. Prove that if G is a finite abelian group with invariant factors $n_1 \mid n_2 \mid \cdots \mid n_\ell$, then every prime number that divides $|G|$ must divide the largest invariant factor n_ℓ.

Exercise 3.18.58. Determine if the following lists are the invariant factors of a finite abelian group.

(a.) 1, 2, 4

(b.) 2, 4, 6, 8

(c.) 2, 6, 30, 210

(d.) 11, 22, 44, 132

Exercise 3.18.59. Determine all possible invariant factors of a finite abelian group G whose order and largest invariant factor n as follows. Explain if it is not possible for such a group to exist.

(a.) $|G| = 4$ and $n = 4$

(b.) $|G| = 16$ and $n = 4$

(c.) $|G| = 20$ and $n = 5$

(d.) $|G| = 36$ and $n = 6$

(e.) $|G| = 210$ and $n = 10$

(f.) $|G| = 256$ and $n = 8$

Exercise 3.18.60. Consider any positive integer n with unique prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$ for some distinct primes p_i and non-negative integers e_i. Complete the following steps to prove that

$$\frac{\mathbb{Z}}{n\mathbb{Z}} \cong \frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}.$$

(a.) We proceed by induction on the number k of prime factors of n. By definition, if $k = 2$, then there exist integers p and q such that $\gcd(p, q) = 1$ and $n = pq$. Prove that the order of $(a + p\mathbb{Z}, b + q\mathbb{Z})$ is the smallest positive integer k with $ak \equiv 0 \pmod{p}$ and $bk \equiv 0 \pmod{q}$.

(b.) Conclude that $\operatorname{ord}(1 + p\mathbb{Z}, 1 + q\mathbb{Z})$ is the smallest positive integer k such that $p \mid k$ and $q \mid k$.

(c.) Explain why $\operatorname{lcm}(p, q)$ divides $\operatorname{ord}(1 + p\mathbb{Z}, 1 + q\mathbb{Z})$; then, use Exercise 1.10.34 to verify that $\operatorname{lcm}(p, q) = pq$. Conclude that pq divides the order of $(1 + p\mathbb{Z}, 1 + q\mathbb{Z})$.

(d.) Conversely, explain why the order of $(1 + p\mathbb{Z}, 1 + q\mathbb{Z})$ is at most pq.

(e.) Conclude that $(\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/q\mathbb{Z})$ is a cyclic group with generator $(1 + p\mathbb{Z}, 1 + q\mathbb{Z})$; then, use Proposition 3.3.20 to establish that $\mathbb{Z}/pq\mathbb{Z} \cong (\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z} \times q\mathbb{Z})$.

(f.) By induction, we may assume that the claim holds for k. We must establish the isomorphism in the case that $n = p_1^{e_1} \cdots p_{k+1}^{e_{k+1}}$. Explicitly find positive integers r and s for which $\gcd(r, s) = 1$ and $n = rs$; then, use our induction hypothesis to conclude that $\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/r\mathbb{Z}) \times (\mathbb{Z}/s\mathbb{Z})$. Be sure to choose r and s such that $\mathbb{Z}/r\mathbb{Z} \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})$ and $\mathbb{Z}/s\mathbb{Z} \cong \mathbb{Z}/p_{k+1}^{e_{k+1}}$.

Exercise 3.18.61 (Cauchy's Theorem for Abelian Groups)**.** Complete the following steps to prove that if G is a finite abelian group and p is any prime number such that p divides the order of G, then G admits a cyclic subgroup of order p, i.e., G possesses an element of order p.

(a.) By the Fundamental Theorem of Finite Abelian Groups, there exist (not necessarily distinct) prime numbers $p_1, \dots, p_k$ and positive integers $e, e_1, \dots, e_k$ such that

$$G \cong \frac{\mathbb{Z}}{p^e\mathbb{Z}} \times \frac{\mathbb{Z}}{p_1^{e_1}\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{p_k^{e_k}\mathbb{Z}}.$$

Prove that there exists an injective group isomorphism $\varphi : (\mathbb{Z}/p^e\mathbb{Z}, +) \to G$.

(b.) Conclude from the previous step that G admits a cyclic subgroup of order p^e.

(c.) Conclude from Exercise 2.8.34 that there exists a cyclic subgroup of $(\mathbb{Z}/p^e\mathbb{Z}, +)$ of order p. (Even better, if you can, try to write down the generator of this subgroup explicitly.)

(d.) Conclude from the previous step that G admits a cyclic subgroup of order p; then, verify that the generator of this subgroup induces the desired element of G of order p.

Exercise 3.18.62. Prove that if G is an abelian group such that $|G| = p^k m$ for some prime number p, some non-negative integer k, and some positive integer m such that $\gcd(p, m) = 1$, then there are exactly p^k elements of G of order p^n for some integer $0 \leq n \leq k$.

(**Hint:** One should recognize that this is equivalent to showing that the p-subgroup of G given by $G(p) = \{g \in G \mid \operatorname{ord}(g) = p^n \text{ for some integer } 0 \leq n \leq k\}$ has order p^k.)

3.18.11 Finitely Generated Groups

Exercise 3.18.63. Prove that the symmetric group $\mathfrak{S}_n$ is finitely generated by transpositions of the form $(i, i+1)$ for each integer $1 \leq i \leq n-1$).

(**Hint:** By Proposition 3.8.1, every permutation can be written as a product of some transpositions.)

Exercise 3.18.64. Prove that the symmetric group $\mathfrak{S}_3$ is finitely generated by (12) and (123).

Exercise 3.18.65. Prove that the symmetric group $\mathfrak{S}_4$ is finitely generated by (12) and (1234).

Exercise 3.18.66. Prove that the symmetric group $\mathfrak{S}_n$ is finitely generated by two elements.

(**Hint:** Use Exercises 3.18.64 and 3.18.65 to deduce that the generators are some transposition τ and some cycle σ; then, prove that $\sigma^i \tau \sigma^{-i} = (i+1, i+2)$ for each integer $1 \leq i \leq n-2$.)

Consider the **free group on two generators** presented by $F \times F = \langle g, h \rangle$. By definition, every element of $F \times F$ is a **word** in the **letters** g and h, and there are no relations among the generators g and h. Explicitly, the products $gghh$, $ghgh$, $hggh$, $hghg$, and $hhgg$ are all distinct words in $F \times F$.

Exercise 3.18.67. Prove that the function $\varphi : F \times F \to (\mathbb{Z} \times \mathbb{Z}, +)$ that sends a word $w \in F \times F$ consisting of m copies of the letter g and n copies of the letter h to the ordered pair (m, n) is a surjective group homomorphism with $\ker \varphi = \langle g^{-1}h^{-1}gh \rangle$. Conclude that these groups are not isomorphic; then, explain in your own words why it is intuitively true that $F \times F \not\cong (\mathbb{Z} \times \mathbb{Z}, +)$.

Exercise 3.18.68. Complete the following steps to prove that every subgroup of finite index in a finitely generated group is finitely generated.

(a.) Consider finitely generated group G with a finite system of generators $g_1, \dots, g_n$. Prove that for each integer $1 \leq i \leq n$, we have that g_i^{-1} is a generator of G.

(b.) Prove that G admits finitely many distinct right cosets $He_G, Hx_1, \dots, Hx_k$ of H in G. Conclude that every generator of G lies in some unique right coset of H in G.

(c.) Prove that for all pairs of integers $1 \leq i, j \leq k$, there exists an element $h_{ij} \in H$ and a right coset representative $x_{r_{ij}}$ such that $x_i g_j = h_{ij} x_{r_{ij}}$.

(d.) Given any element $h \in H$, we may write $h = g_{i_1} \cdots g_{i_\ell}$ for some generators $g_{i_1}, \dots, g_{i_\ell}$ of G. Conclude by the previous step that there exists an element $h_{i1} \in H$ and a right coset representative $x_{r_{i1}}$ such that such that $g_{i_1} = e_G g_{i_1} = h_{i1} x_{r_{i1}}$ and $h = h_{i1} x_{r_{i1}} g_{i_2} \cdots g_{i_\ell}$.

(e.) Conclude by the previous two steps that there exists an element $h_{i2} \in H$ and a right coset representative $x_{r_{i2}}$ such that $x_{r_{i1}} g_{i_2} = h_{i2} x_{r_{i2}}$ and $h = h_{i1} h_{i2} x_{r_{i2}} g_{i_3} \cdots g_{i_\ell}$.

(f.) Continue this process until we have that $h = h_{i1} h_{i2} \cdots h_{i\ell} x_{r_{i\ell}}$. Conclude that $h_{i1} h_{i2} \cdots h_{i\ell} x_{r_{i\ell}}$ lies in the right coset represented by e_G so that $x_{r_{i\ell}} = e_G$ and $h = h_{i1} h_{i2} \cdots h_{i\ell}$.

(g.) Conclude that H is generated by the elements h_{ij} for all pairs of integers $1 \leq i, j \leq n$.

3.18.12 Finitely Generated Abelian Groups

Exercise 3.18.69. Consider the free abelian group $(\mathbb{Z}^r, +)$ of rank r for some positive integer r. Given any positive integer n, consider the subgroup $n\mathbb{Z}^r = \{(na_1, \dots, na_r) \mid a_1, \dots, a_r \in \mathbb{Z}\}$ of $(\mathbb{Z}^r, +)$. Consider the direct product $\mathbb{Z}_n^r = (\mathbb{Z}/n\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n\mathbb{Z})$ with r direct factors of $\mathbb{Z}/n\mathbb{Z}$. Prove that the function $\varphi : (\mathbb{Z}^r, +) \to (\mathbb{Z}_n^r, +)$ defined by $\varphi(a_1, \dots, a_r) = (a_1 + n\mathbb{Z}, \dots, a_r + n\mathbb{Z})$ is a surjective group homomorphism with $\ker \varphi = n\mathbb{Z}^r$. Conclude that

$$\frac{\mathbb{Z} \times \cdots \times \mathbb{Z}}{n(\mathbb{Z} \times \cdots \times \mathbb{Z})} \cong \frac{\mathbb{Z}}{n\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{n\mathbb{Z}}.$$

3.18.13 Smith Normal Form

Exercise 3.18.70. Compute the Smith Normal Form of the following integer matrix.

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

Use this information to prove that $(\mathbb{Z} \times \mathbb{Z})/\langle (a, 0), (b, 0) \rangle \cong (\mathbb{Z}/\gcd(a, b)\mathbb{Z}) \times (\mathbb{Z}/\operatorname{lcm}(a, b)\mathbb{Z})$.

Exercise 3.18.71. Compute the Smith Normal Form of the following integer matrix.

$$A = \begin{pmatrix} 0 & 5 \\ 30 & 3 \end{pmatrix}$$

Use this information to prove that $(\mathbb{Z} \times \mathbb{Z}_{30})/\langle (5, 3 + 30\mathbb{Z}) \rangle \cong \{0\} \times \mathbb{Z}_{150}$.

Exercise 3.18.72. Compute the Smith Normal Form of the following integer matrix

$$\begin{pmatrix} 4 & 6 \\ 1 & 3 \end{pmatrix}$$

Use this information to exhibit an explicit isomorphism between the finitely generated abelian group $(\mathbb{Z} \times \mathbb{Z})/\langle (4,1),(6,3)\rangle$ and its invariant factor decomposition.

One of the foremost uses of the Smith Normal Form is to compute the Rational Canonical Form of a matrix over a field k or the Jordan Canonical Form of a matrix over an algebraically closed field — often taken to be the complex numbers $\mathbb{C}$. One can prove that the **minimal polynomial** of an $n \times n$ matrix A is the largest invariant factor of the **characteristic matrix** $xI - A$; likewise, the **characteristic polynomial** of A is the product of the invariant factors of $xI - A$.

Exercise 3.18.73. Compute the Smith Normal Form of the matrix $xI - A$ given that

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

to find the invariant factors, elementary divisors, and minimal and characteristic polynomials of A.

Exercise 3.18.74. Compute the Smith Normal Form of the matrix $xI - A$ given that

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

to find the invariant factors, elementary divisors, and minimal and characteristic polynomials of A.

Exercise 3.18.75. Compute the Smith Normal Form of the matrix $xI - A$ given that

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

to find the invariant factors, elementary divisors, and minimal and characteristic polynomials of A.

3.18.14 Group Actions and the Class Equation

Exercise 3.18.76. Consider a group G. We say that a pair of elements $x, y \in G$ are **conjugate** in G if and only if there exists an element $g \in G$ such that $y = gxg^{-1}$.

(a.) Prove that conjugacy forms an equivalence relation on G with equivalence classes

$$[x] = \{gxg^{-1} \mid g \in G\}.$$

(b.) Prove that $[x]$ is a singleton if and only if x lies in the center $Z(G)$ of G.

(c.) Prove that $\#[x] = [G : Z_G(x)]$, where $Z_G(x) = \{g \in G \mid gx = xg\}$ is the centralizer of x in G and $[G : Z_G(x)]$ is the index of the centralizer of x in G.

(d.) Prove that if $|G|$ is odd, then x and x^{-1} are conjugate if and only if $x = e_G$.

(e.) Prove that if $|G|$ is odd, then every normal subgroup N of G of order three lies in $Z(G)$.

Exercise 3.18.77. Prove that every subgroup H of a group G induces a well-defined group action $h * K = hKh^{-1}$ on the collection of subgroups of G. We refer to this action as H-**conjugation**.

Exercise 3.18.78. Prove that the Class Equation of a Group Action holds.

Exercise 3.18.79. Complete the following steps to prove the following generalization of Cayley's Theorem: if G is any group that acts faithfully on any nonempty set X via $(g, x) \mapsto g * x$, then G is isomorphic to a subgroup of the group $\mathfrak{S}_X$ of permutations of X.

(a.) Prove that this action induces a bijection $\varphi_g : X \to X$ defined by $\varphi_g(x) = g * x$.

(b.) Prove that the function $\sigma : G \to \mathfrak{S}_X$ defined by $\sigma(g) = \varphi_g$ is a group homomorphism.

(c.) Prove that σ is injective if and only if G is faithful.

(d.) Conclude that $\sigma(G)$ is isomorphic to a subgroup of $\mathfrak{S}_X$. Explain why this implies that if X is a finite group, then $|G|$ divides $|X|!$. Conclude that G must also be finite.

Exercise 3.18.80. Prove that if G is any group of prime power order p^2, then G is abelian.

Exercise 3.18.81. Consider a group G of order p^n for some prime number p and integer $n \geq 1$.

(a.) Prove that the center $Z(G)$ of G is non-trivial.

(b.) Prove that if N is a normal subgroup of G of order p, then N lies in $Z(G)$.

Exercise 3.18.82. Consider any prime number p. We will use the Orbit-Stabilizer Theorem in this exercise to count the number of integer matrices that are their own matrix inverse modulo p.

(a.) Prove that $\mathbb{Z}/p\mathbb{Z}$ admits a well-defined multiplication $(a + p\mathbb{Z})(b + p\mathbb{Z}) = ab + p\mathbb{Z}$.

(b.) Prove that the following set forms a group with respect to matrix multiplication.

$$\mathrm{GL}(2, \mathbb{Z}/p\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{Z}/p\mathbb{Z} \right\}$$

(We have suppressed the coset notation for simplicity; the products are taken modulo p.)

(c.) Prove that the order of $\mathrm{GL}(2, \mathbb{Z}/p\mathbb{Z})$ is exactly $(p^2 - 1)(p^2 - p)$.

(d.) Prove that two matrices lie in the same orbit under conjugation if and only if the two matrices are similar if and only if the two matrices have the same characteristic polynomial.

(e.) Prove that a matrix B lies in the stabilizer of a matrix A if and only if B commutes with A.

(f.) Conclude from the previous steps that if $A^2 = I$, then the characteristic polynomial of A is $x^2 - 1$, hence it suffices to consider matrices annihilated by the divisors of $x^2 - 1$.

(g.) Prove that if $p = 2$, then the only divisors of $x^2 - 1$ are $x - 1$ and $(x - 1)^2$.

(h.) Conclude by the previous step that if $p = 2$, then $A = I$ or A is similar to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

(i.) Prove that the only matrices that commute with $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ are of the form $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with $a, b, d \in \mathbb{Z}/p\mathbb{Z}$. Conclude that there are two such matrices — namely, the 2×2 identity matrix and $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

(j.) Conclude that if $p = 2$, then there are $\frac{(2^2-1)(2^2-2)}{2} + 1 = 3 + 1 = 4$ matrices with $A^2 = I$.

(k.) Prove that if $p > 2$, then the only divisors of $x^2 - 1$ are $x - 1$, $x + 1$, and $x^2 - 1$.

(l.) Conclude by the previous step that if $p > 2$, then $A = \pm I$ or A is similar to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

(m.) Prove that the only matrices that commute with $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ are of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right)$ with $a, d \in \mathbb{Z}/p\mathbb{Z}$. Conclude that there are $(p-1)^2$ such matrices.

(n.) Conclude that if $p > 2$, then there are $\frac{(p^2-1)(p^2-p)}{(p-1)^2} + 1 + 1 = p(p+1) + 2$ matrices with $A^2 = I$.

Exercise 3.18.83. Given any group G, denote by $Z(G)$ the center of G. By Exercise 3.18.14, the center of G is a normal subgroup of G. Construct subgroups $Z_i(G)$ inductively as follows.

1.) Begin with the trivial subgroup $Z_0(G) = \{e_G\}$.

2.) For each integer $i \geq 0$, let $Z_{i+1}(G)$ be the subgroup of G that is the pre-image of the center of the group $G/Z_i(G)$ so that $Z_{i+1}(G)/Z_i(G)$ is the center of $G/Z_i(G)$.

We refer to the series of subgroups $Z_i(G)$ as an **upper central series** of G, and we say that the group G is **nilpotent** if there exists an integer $n \geq 1$ such that $Z_n(G) = G$.

(a.) Prove that every abelian group is nilpotent.

(b.) Prove that $Z_i(G)$ is a normal subgroup of G for each integer $i \geq 0$.

(c.) Prove that if $|G| = p^r$ for some prime number p and some integer $r \geq 0$, then G is nilpotent.

3.18.15 Sylow's Theorems

Exercise 3.18.84. Consider a finite group G and a prime number p that divides $|G|$. Complete the following steps to provide an alternative proof for Cauchy's Theorem for Finite Groups.

(a.) Consider the set X of p-tuples $(x_1, \dots, x_p)$ of elements $x_1, \dots, x_p \in G$ such that $x_1 \cdots x_p = e_G$. Prove that X is a nonempty finite set.

(b.) Use the Fundamental Counting Principle to prove that $|X| = |G|^{p-1}$.

(**Hint:** Observe that if $x_1 \cdots x_p = e_G$, then $x_p = (x_1 \cdots x_{p-1})^{-1}$.)

(c.) Prove that we may define a group action of $\mathbb{Z}/p\mathbb{Z}$ on X by declaring that

$$(1 + p\mathbb{Z}) * (x_1, x_2, \dots, x_{p-1}, x_p) = (x_p, x_1, x_2, \dots, x_{p-1}).$$

(d.) Conclude by the Orbit-Stabilizer Theorem that the orbit of $(x_1, \dots, x_p)$ has size 1 or p.

(e.) Conclude by the Class Equation of a Group Action that $|G|^{p-1}$ is the sum of the number of orbits of X size 1 and the number of orbits of X of size p.

(f.) Conclude that the number of orbits of X of size 1 is divisible by p.

(g.) Prove that the trivial p-tuple $(e_G, \dots, e_G)$ induces an orbit of X of size 1. Conclude that from the previous step that there are at least p orbits of X of size 1, hence there exists a p-tuple $(x_1, \dots, x_p)$ such that $(x_p, x_1, x_2, \dots, x_{p-1}) = (x_1, x_2, x_3, \dots, x_{p-1})$.

(h.) Prove that all components of the p-tuple from the previous step are equal — say to some element $x \in G$ — hence G contains at least p elements whose order divides p.

(i.) Conclude that $p-1$ of the elements of G whose order divides p are non-trivial, hence G contains at least $p-1$ elements of order p so that G contains a subgroup of order p.

Exercise 3.18.85. Consider a finite abelian group G. Given any prime number p that divides $|G|$, prove that G admits a unique Sylow p-subgroup.

Exercise 3.18.86. Give an example of a group G with a normal subgroup H such that both H and G/H are nilpotent but G is not nilpotent.

Exercise 3.18.87. Prove that any group of order 30 admits a cyclic subgroup of order 15.

Exercise 3.18.88. Prove that the center of any non-abelian group of order 21 is trivial.

Exercise 3.18.89. Prove that any group of order $435 = 3 \cdot 5 \cdot 29$ must be abelian.

Exercise 3.18.90. Consider a prime number $p > 5$ such that $p \not\equiv 1 \pmod 5$. Prove that any group of order $15p$ must contain a subgroup of order $5p$.

3.18.16 The Commutator Subgroup

Exercise 3.18.91. Consider a group G with commutator subgroup $[G, G] = \langle g^{-1}h^{-1}gh \mid g, h \in G \rangle$.

(a.) Prove that $x(g^{-1}h^{-1}gh)x^{-1} = (xg^{-1}x^{-1})(xh^{-1}x^{-1})(xgx^{-1})(xhx^{-1})$ for all elements $g, h, x \in G$. Conclude that conjugation defines a group action of G on its commutator subgroup $[G, G]$.

(b.) Prove that for any group homomorphism $\varphi : G \to H$, we have that $\varphi([G, G]) \subseteq [H, H]$.

(**Hint:** Considering that G is generated by elements of the form $g^{-1}h^{-1}gh$, it suffices to prove that $\varphi(g^{-1}h^{-1}gh) \in [H, H]$ for some element $g, h \in G$.)

(c.) Prove that if G is abelian, then $[G, G] = \{e_G\}$.

(d.) We say that G is **perfect** if it holds that $[G, G] = G$. Prove that every simple group is perfect.

(**Hint:** By the Law of the Excluded Middle, either G is abelian, or it is not.)

3.18.17 Semidirect Products

Exercise 3.18.92. Consider the semidirect product $H \rtimes_\varphi K$ of some groups H and K with respect to some group homomorphism $\varphi : K \to \operatorname{Aut}(K)$. Prove that the centralizer of K_φ in H_φ is equal to the normalizer of K_φ in H_φ.

Exercise 3.18.93. Consider a prime number p. Give an example of a non-abelian group of order p^n for some integer $n \geq 1$ whose center contains more than one normal subgroup of order p.

Chapter 4

Ring Theory I

Ring theory is the study of objects for which there exists a notion of addition and multiplication. Common mathematical structures such as the real numbers, real polynomials, and real square matrices are all examples of rings with respect to the appropriate notion of addition and multiplication. Often, the assumption is made that the multiplication defined in a ring is commutative, i.e., the order of two elements in a product does not matter. Broadly, this area of ring theory is referred to as commutative algebra, and it involves more general algebraic structures associated to rings. Commutative algebra hosts many interesting and challenging unresolved questions; however, the techniques inherent to the field can also be used to study objects arising in combinatorics, geometry, number theory, and topology. Elsewhere, there exists a rich theory of non-commutative rings; these sorts of rings arise naturally in relation to operator theory and topological ring theory.

4.1 Rings and Ring Homomorphisms

Consider an additive abelian group $(R, +)$ equipped with a binary operation $\cdot : R \times R \to R$ that sends $(r, s) \mapsto r \cdot s$. Crucially, observe that this operation is written multiplicatively. We say that R forms a (unital) **ring** with respect to $+$ and $\cdot$ if the triple $(R, +, \cdot)$ satisfies the following properties.

1.) We have that $r \cdot (s \cdot t) = (r \cdot s) \cdot t$ for any elements $r, s, t \in R$, i.e., $\cdot$ is associative.

2.) We have that $r \cdot (s + t) = r \cdot s + r \cdot t$ and $(r + s) \cdot t = r \cdot t + s \cdot t$ for any elements $r, s, t \in R$, i.e., $\cdot$ is distributive on both the left- and the right-hand side.

3.) R admits an element $1_R \in R$ such that $1_R \cdot r = r = r \cdot 1_R$ for all elements $r \in R$.

Caution: even though this situation is growing increasingly uncommon over time, it is possible at this point to come across an author who defines a ring as an additive abelian group with multiplication that satisfies properties (1.) and (2.) but *not necessarily* property (3.). We will refer to such an object as a **rng** because it has no multiplicative "i"dentity; however, these authors refer to our element $1_R \in R$ as the **unity** of R, and they refer to our rings as **unital rings** or rings with unity.

Example 4.1.1. Consider the abelian group $(\mathbb{Z}, +)$. Certainly, multiplication of integers is associative and distributive, and the multiplicative identity of the integers is the integer 1. Consequently, we conclude that $\mathbb{Z}$ forms a commutative unital ring because integer multiplication is commutative.

Example 4.1.2. Consider the abelian group $(n\mathbb{Z}, +)$ for any nonzero integer n. Once again, multiplication of integers is associative, distributive, and commutative. Even more, if we take any pair of integers $na, nb \in n\mathbb{Z}$, then their product $(na)(nb) = n^2ab = n(nab)$ lies in $n\mathbb{Z}$, hence $n\mathbb{Z}$ is closed under integer multiplication; however, unless we impose the condition that $n = \pm 1$, there does not exist an integer of the form na such that $(na)(nb) = nb$ for all integers b. Consequently, we conclude that $n\mathbb{Z}$ forms a commutative rng; in particular, $n\mathbb{Z}$ is not unital except when $n = \pm 1$.

Example 4.1.3. Consider the additive abelian group $(\mathbb{Z}/n\mathbb{Z}, +)$ for some positive integer n. We define multiplication on $\mathbb{Z}/n\mathbb{Z}$ by declaring that $(a+n\mathbb{Z})(b+n\mathbb{Z}) = ab+n\mathbb{Z}$. Once we establish that this operation is well-defined, it will follow shortly thereafter that $\mathbb{Z}/n\mathbb{Z}$ is a commutative unital ring with respect to this multiplication because integer multiplication is associative, commutative, distributive, and the left coset $1+n\mathbb{Z}$ is the multiplicative identity of $\mathbb{Z}/n\mathbb{Z}$. Consider any left coset representatives $a + n\mathbb{Z} = c + n\mathbb{Z}$ and $b + n\mathbb{Z} = d + n\mathbb{Z}$. Observe that $a = a + n \cdot 0$ is an element of $a+n\mathbb{Z}$, hence there exists an integer r such that $a = c+nr$. Likewise, there exists an integer s such that $b = d + ns$. Consequently, we have that $ab = (c + nr)(d + ns) = cd + n(cs) + n(dr) + n(nrs)$. We conclude therefore that $ab + n\mathbb{Z} = cd + n(cs) + n(dr) + n(nrs) + n\mathbb{Z} = cd + n\mathbb{Z}$ by Proposition 3.1.4 because the left cosets $n(cs) + n(dr) + n(nrs) + n\mathbb{Z}$ and $0 + n\mathbb{Z}$ are equal; this shows that the multiplication on $\mathbb{Z}/n\mathbb{Z}$ is well-defined, hence $\mathbb{Z}/n\mathbb{Z}$ is a commutative unital ring.

Example 4.1.4. Consider the collection $\mathbb{R}[x]$ of real polynomials in indeterminate x. One might recall from linear algebra that $\mathbb{R}[x]$ is a real vector space (of infinite dimension), hence it is in particular an abelian group under polynomial addition. Even more, polynomial multiplication is associative, commutative, and distributive, and the multiplicative identity of $\mathbb{R}[x]$ is the constant polynomial 1. Consequently, it follows that $\mathbb{R}[x]$ is a commutative unital ring. Generally, if R is any rng, then we may define the **polynomial rng** $R[x]$ in indeterminate x by generalizing the usual polynomial addition such that $r_ix^i + s_ix^i = (r_i + s_i)x^i$ and by declaring that $(r_ix^i)(s_jx^j) = r_is_jx^{i+j}$. Consequently, it follows that $R[x]$ is a rng that is commutative if and only if R is commutative and unital if and only if R is unital. We will study polynomial rngs in greater depth in Section 4.6.

Example 4.1.5. Consider the abelian group $(\mathbb{R}^{n\times n}, +)$ of real $n\times n$ matrices under matrix addition. Back in linear algebra, we learn that matrix multiplication is associative and distributive, and the product of two real $n \times n$ matrices is once again a real $n \times n$ matrix; the $n \times n$ identity matrix I satisfies that $IA = A = AI$ for all real $n \times n$ matrices A, hence it is the multiplicative identity of $\mathbb{R}^{n\times n}$. Consequently, it follows that $\mathbb{R}^{n\times n}$ is a unital ring; however, it is not commutative except when $n = 1$. Explicitly, the following real 2×2 matrices do not commute with one another.

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 7 & 1 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 4 & 6 \end{bmatrix}$$

Considering that we can realize these two non-commuting real 2×2 matrices A and B as 2×2 submatrices of any real $n \times n$ matrices with $n \geq 2$, it follows that $\mathbb{R}^{n\times n}$ is not commutative.

Example 4.1.6. Consider the collection $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$. One can readily verify that $F(\mathbb{R}, \mathbb{R})$ forms an abelian group with respect to function addition; in fact, one might recall from linear algebra $F(\mathbb{R}, \mathbb{R})$ is a real vector space. Even more, function composition is associative and distributive, and the identity function defined by $f(x) = x$ is the multiplicative identity of $F(\mathbb{R}, \mathbb{R})$. Consequently, it follows that $F(\mathbb{R}, \mathbb{R})$ forms a unital ring with respect to function composition.

On the other hand, it is also entirely valid to define multiplication of functions according to the rule $(fg)(x) = f(x)g(x)$. Explicitly, under this assignment, the product of functions corresponds to **pointwise multiplication** of their images. Observe that with respect to this product, the constant function $f(x) = 1$ satisfies that $(fg)(x) = g(x) = (gf)(x)$, hence it is the multiplicative identity of $F(\mathbb{R}, \mathbb{R})$ in this case. Considering that multiplication of real numbers is associative, distributive, and commutative, it follows that $F(\mathbb{R}, \mathbb{R})$ is a commutative unital ring with respect to this product.

Example 4.1.7. Consider any finite collections of rngs $R_1, \dots, R_n$. By Proposition 3.9.1, it follows that $R_1 \times \cdots \times R_n$ is an additive abelian group with respect to componentwise addition. Likewise, one can show along the same lines as the proof of the aforementioned proposition that componentwise multiplication of the elements of $R_1 \times \cdots \times R_n$ constitutes a binary operation on $R_1 \times \cdots \times R_n$ so that $R_1 \times \cdots \times R_n$ is a rng with respect to componentwise addition and multiplication. We refer to this rng as the **direct product** of $R_1, \dots, R_n$. Even more, as before, the properties of $R_1 \times \cdots \times R_n$ are intimately connected with the properties of $R_1, \dots, R_n$. Explicitly, we have that

1.) $R_1 \times \cdots \times R_n$ is a unital ring if and only if $R_1, \dots, R_n$ are unital rings and

2.) $R_1 \times \cdots \times R_n$ is commutative if and only if $R_1, \dots, R_n$ are commutative.

Explicitly, for the first property, the multiplicative identity must be the n-tuple $(1_{R_1}, \dots, 1_{R_n})$.

Going forward, we will omit the multiplicative notation $\cdot$ of a rng R and simply resort to the usual concatenation, e.g., $r \cdot s = rs$ as we had done in our study of group theory.

Each of the properties inherent to a rng R will either be discovered anew or inherited from the additive abelian group structure of R. We remind the reader that for any element $r \in R$ and any integer n, we have that $n \cdot r = r + r + \cdots + r$ with n summands if n is non-negative and $n \cdot r = (-r) + (-r) + \cdots + (-r)$ with n summands if n is negative, where $-r$ is the **additive inverse** of r satisfying that $r + (-r) = 0_R = (-r) + r$ for the **additive identity** element 0_R of R as guaranteed by the additive group structure of R. Our next proposition demonstrates that the additive and multiplicative operations of a rng interact with each other in a civilized manner.

Proposition 4.1.8. *Consider any rng R with additive identity element 0_R.*

1.) *We have that $0_R r = 0_R = r 0_R$ for all elements $r \in R$.*

2.) *We have that $r(-s) = -(rs) = (-r)s$ for all elements $r, s \in R$.*

3.) *We have that $(-r)(-s) = rs$ for all elements $r, s \in R$.*

4.) *If R is unital, then its multiplicative identity 1_R is unique.*

Proof. 1.) By definition of the additive identity element of R, for every element $r \in R$, we have that $0_R r = (0_R + 0_R)r = 0_R r + 0_R r$ by the distributive property. Cancelling one summand of $0_R r$ from both sides of this identity illustrates that $0_R r = 0_R$; the fact that $r 0_R = 0_R$ follows similarly.

2.) Observe that the additive inverse of an element of an additive abelian group is unique by Proposition 2.2.2, hence it suffices to prove that $rs + r(-s) = 0_R$ for all elements $r, s \in R$. But this holds by the distributive property of R: we have that $rs + r(-s) = r(s + (-s)) = r 0_R = 0_R$.

3.) Like before, we have that $(-r)(-s) - (rs) = (-r)(-s) + r(-s) = ((-r) + r)(-s) = 0_R$.

4.) Consider the multiplicative identity 1_R of R. Given any element $1 \in R$ such that $1r = r = r1$ for all elements $r \in R$, it follows that $1 = 1_R 1 = 1_R$: on the left-hand side, we use the property of 1_R as the multiplicative identity of R, and on the right-hand side, we use the property of 1. □

Proposition 4.1.9 (Ring Exponent Laws)**.** *Let R be any rng. Let m and n be positive integers.*

1.) *We have that $r^m r^n = r^{m+n}$ for any element $r \in R$.*

2.) *We have that $(r^m)^n = r^{mn}$ for any element $r \in R$.*

3.) *If R is commutative, then $(r_1 r_2)^n = r_1^n r_2^n$ for all elements $r_1, r_2 \in R$.*

Given any nonzero element r of a rng R, it stands in contrast to the situation with multiplicative groups that r must possess a multiplicative inverse in R. Explicitly, a generic rng R only carries the structure of a **semigroup** under multiplication; therefore, a unital ring can be viewed as a **monoid** under multiplication. Even still, there is no guarantee (or requirement) that r possesses a multiplicative inverse. Consequently, if there exists a nonzero element $s \in R$ such that $rs = 1_R = sr$, then we refer to r as a **unit**. Occasionally, a unit r is referred to as an **invertible** element of R. Exercise 4.8.4 yields that such an element s is unique to r; it is called the **multiplicative inverse** of r, and it is denoted by $s = r^{-1}$. Once again, we make no assumption that every nonzero element of R has a multiplicative inverse; in fact, a ring with this property is called a **skew field**. We will henceforth adopt the notation $U(R)$ to denote the collection of units of a unital ring R.

Example 4.1.10. We have seen in Example 2.1.6 that the only integers with multiplicative inverses in $\mathbb{Z}$ are 1 and -1. Consequently, the units of the ring $\mathbb{Z}$ are 1 and -1, i.e., $U(\mathbb{Z}) = \{1, -1\}$.

Example 4.1.11. Consider the commutative unital ring $\mathbb{Z}/n\mathbb{Z}$ for any positive integer n. Observe that a left coset $a + n\mathbb{Z}$ is a unit of $\mathbb{Z}/n\mathbb{Z}$ if and only if there exists a left coset $b + n\mathbb{Z}$ such that $ab + n\mathbb{Z} = (a + n\mathbb{Z})(b + n\mathbb{Z}) = 1 + n\mathbb{Z}$ if and only if $ab - 1 = nq$ for some integer q by Proposition 3.1.4 if and only if $ab + n(-q) = 1$. By Bézout's Identity, the units of $\mathbb{Z}/n\mathbb{Z}$ are the cosets $a + n\mathbb{Z}$ such that $\gcd(n, a) = 1$ so that $|U(\mathbb{Z}/n\mathbb{Z})| = \phi(n)$ by the paragraph preceding Exercise 2.8.28.

Example 4.1.12. Given any real $n \times n$ matrix A, we have that A is a unit of $\mathbb{R}^{n \times n}$ if and only if A is invertible if and only if $\det(A)$ is nonzero. Consequently, the units of the unital ring of real $n \times n$ matrices are precisely the real invertible $n \times n$ matrices, i.e., we have that $U(\mathbb{R}^{n \times n}) = \mathrm{GL}(n, \mathbb{R})$.

Example 4.1.13. Observe that a real number is a unit of the ring of real numbers $\mathbb{R}$ if and only if it is nonzero. Consequently, we have that $U(\mathbb{R}) = \mathbb{R}^\times$ (the nonzero real numbers), and $\mathbb{R}$ is a **field**.

Like with groups, we are concerned with structure-preserving functions of rngs R and S. We say that a function $\varphi : R \to S$ is a **rng homomorphism** if and only if for all elements $r_1, r_2 \in R$,

1.) $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$, i.e., φ is a group homomorphism and

2.) $\varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2)$, i.e., φ is compatible with multiplication.

Even more, if both R and S are unital rings, then we impose a third condition that

3.) $\varphi(1_R) = 1_S$, i.e., the multiplicative identity of R maps to the multiplicative identity of S.

We say in this case that φ is a **unital ring homomorphism**. Once again, this last condition serves to underline the fundamental differences between groups and rings. Explicitly, it is required in the definition of a unital ring homomorphism that $\varphi(1_R) = 1_S$; however, for a group homomorphism, it is possible to prove from the definition and the group axioms that the identity element of one group maps to the identity element of the other group under any group homomorphism. Essentially, it is not possible to derive such a conclusion here because we cannot a priori guarantee that $\varphi(1_R)$ is a unit of S. Let us do a few examples of rng homomorphisms to illustrate this idea.

Example 4.1.14. Consider the rng $n\mathbb{Z}$ and the unital ring $\mathbb{Z}$ for some positive integer n. We may define a rng homomorphism $\varphi : n\mathbb{Z} \to \mathbb{Z}$ by declaring that $\varphi(na) = na$: indeed, for any pair of elements $na, nb \in n\mathbb{Z}$, we have that $\varphi(na+nb) = \varphi(n(a+b)) = n(a+b) = na+nb = \varphi(na)+\varphi(nb)$ and $\varphi((na)(nb)) = \varphi(n(nab)) = n(nab) = (na)(nb) = \varphi(na)\varphi(nb)$. On the other hand, unless we assume that $n = 1$, then $n\mathbb{Z}$ does not possess a multiplicative identity, so we are done.

Example 4.1.15. Consider the function $\varphi : \mathbb{Z} \to \mathbb{Z}$ defined by $\varphi(n) = 2n$. Even though it holds that $\varphi(m+n) = 2(m+n) = 2m+2n = \varphi(m)+\varphi(n)$, we have that $\varphi(mn) = 2mn$ is not equal to $\varphi(m)\varphi(n) = (2m)(2n) = 4mn$ unless one of the integers m or n is zero. Consequently, this is not a unital ring homomorphism. Explicitly, a function $\psi : \mathbb{Z} \to \mathbb{Z}$ is a unital ring homomorphism if and only if $\psi(n) = \psi(1+1+\cdots+1) = \psi(1)+\psi(1)+\cdots+\psi(1) = n\psi(1)$ for all integers n; moreover, a unital ring homomorphism must satisfy that $\psi(1) = 1$, hence it follows that $\psi(n) = n$ for all integers n, i.e., the only unital ring homomorphism from $\mathbb{Z}$ to itself is the identity homomorphism.

Example 4.1.16. Given any unital ring R, let us classify all unital ring homomorphisms $\varphi : \mathbb{Z} \to R$. Once again, by the first and third conditions above, φ is a unital ring homomorphism only if for all integers n, it holds that $\varphi(n) = \varphi(1+1+\cdots+1) = \varphi(1)+\varphi(1)+\cdots+\varphi(1) = n\varphi(1) = n1_R$. Consequently, the only ring homomorphisms from the integers to a unital ring are multiplication by the multiplicative identity 1_R. On the other hand, if we assume that S is any rng, then a rng homomorphism $\psi : \mathbb{Z} \to S$ must be uniquely determined by $\psi(1)$ because we have that $\psi(n) = n\psi(1)$ by the previous computation. Because we are not imposing any additional structure on S, it follows that $\psi(1)$ could be anything in this case, and ψ can be viewed simply as multiplication by $\psi(1)$.

Like with group homomorphisms, we refer to a bijective rng homomorphism as a **rng isomorphism**. We say that the rngs R and S are **isomorphic** if there exists a rng isomorphism $\varphi : R \to S$, and we write $R \cong S$. We will come to find that it is more difficult to find rng homomorphisms (and hence rng isomorphisms) than it was to find group homomorphisms (isomorphisms) because a rng homomorphism must satisfy additional properties. Explicitly, Exercises 4.8.9 and 4.8.35 underscore the differences between the group structure and the rng structure of certain familiar sets.

Given any rng homomorphism $\varphi : R \to S$, as with group homomorphisms, we are interested in the kernel of φ, i.e., the set $\ker\varphi = \{r \in R \mid \varphi(r) = 0_S\}$. Considering that kernel membership is a property of the addition in R, the following can be deduced immediately from Proposition 3.3.6.

Proposition 4.1.17. *Given any rng homomorphism $\varphi : R \to S$, we have that φ is injective if and only if the kernel of φ is the trivial subgroup of R, i.e., $\ker\varphi = \{0_R\}$.*

Example 4.1.18. Consider the function $\varphi : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ defined by $\varphi(n) = (n, n)$. Given any elements $m, n \in \mathbb{Z}$, we have that $\varphi(m+n) = (m+n, m+n) = (m, m)+(n, n) = \varphi(m)+\varphi(n)$ and $\varphi(mn) = (mn, mn) = (m, m)(n, n) = \varphi(m)\varphi(n)$. Even more, we have that $\varphi(1) = (1, 1)$ is the multiplicative identity of $\mathbb{Z} \times \mathbb{Z}$, hence φ is a unital ring homomorphism. We have that $n \in \ker \varphi$ if and only if $\varphi(n) = (0,0)$ if and only if $(n,n) = (0,0)$ if and only if $n = 0$, hence φ is injective.

Example 4.1.19. Consider the function $\varphi : \mathbb{C} \to \mathbb{C}$ defined by $\varphi(a+bi) = a - bi$. Explicitly, one may recognize φ as complex conjugation. Given any real numbers a and b, we have that

$$\varphi((a+c)+(b+d)i) = (a+c)-(b+d)i = (a-bi)+(c-di) = \varphi(a+bi)+\varphi(c+di)$$

and $\varphi((ac-bd)+(ad+bc)i) = (ac-bd)-(ad+bc)i = (a-bi)(c-di) = \varphi(a+bi)\varphi(c+di)$. Even more, we have that $\varphi(1+0i) = 1-0i$, hence φ sends the multiplicative identity of $\mathbb{C}$ to itself. We conclude that φ is a unital ring homomorphism. Last, we note that $a+bi \in \ker\varphi$ if and only if $a - bi = 0 + 0i$ if and only if $a = 0$ and $b = 0$, and we conclude as usual that φ is injective.

Given any unital ring R, we demonstrated in Example 4.1.16 that every unital ring homomorphism $\varphi : \mathbb{Z} \to R$ is defined by $\varphi(n) = n \cdot 1_R$. By definition, $\ker\varphi$ consists of all integers n such that $n \cdot 1_R = 0_R$. We refer to the **characteristic** $\operatorname{char}(R)$ of R as the smallest (with respect to divisibility) positive integer n for which $n \cdot 1_R = 0_R$. Conventionally, if $n \cdot 1_R$ is nonzero for all positive integers n, then the characteristic of R is zero; otherwise, the characteristic of R is a positive integer.

Example 4.1.20. Certainly, the characteristic of the commutative unital rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are zero: by definition, these rings do not admit any multiples of 1 that result in 0 other than $0 \cdot 1$.

Example 4.1.21. Consider the commutative unital ring $\mathbb{Z}/n\mathbb{Z}$ for some positive integer n. Each of the left cosets $k(1+n\mathbb{Z}) = k + n\mathbb{Z}$ is nonzero for each integer $1 \leq k \leq n-1$. On the other hand, we have that $n(1+n\mathbb{Z}) = n + n\mathbb{Z} = 0 + n\mathbb{Z}$, hence we conclude that $\operatorname{char}(\mathbb{Z}/n\mathbb{Z}) = n$.

4.2 Ideals and Quotient Rings

Given any nonempty set $S \subseteq R$, we say that S is a **subrng** of R whenever S is a rng with respect to the prescribed binary operations of R. Often, in order to determine if $S \subseteq R$ is a subrng of R, it is most practical and convenient to use the following generalization of the Subgroup Test.

Proposition 4.2.1 (Subrng Test)**.** *Consider any rng R and any set $S \subseteq R$. We have that S is a subrng of R if and only if the following three properties hold.*

1.) *We have that S is nonempty.*

2.) *We have that $r - s \in S$ for all elements $r, s \in S$, i.e., S is closed under subtraction.*

3.) *We have that $rs \in S$ for all elements $r, s \in S$, i.e., S is closed under multiplication.*

Even more, if R is commutative, then S is commutative. Likewise, if R is a unital ring such that S contains the multiplicative identity 1_R of R, then S is a unital ring with multiplicative identity 1_R.

Proof. We note that if S is any subset of R that satisfies the first and second properties above, then $(S, +)$ is a subgroup of $(R, +)$ by the One-Step Subgroup Test. Even more, if S satisfies the third property above, then the multiplication of R is a binary operation on S; it is automatically associative and distributive because the elements of S can all be viewed as elements of R.

Conversely, if S is a subrng of R, then $(S, +)$ is a subgroup of $(R, +)$, hence S cannot be empty because it must contain the additive identity 0_R by the Subgroup Test. Even more, we must have that $r - s \in S$ for all elements $r, s \in R$ by the One-Step Subgroup Test. Last, we must have that $rs \in S$ for all elements $r, s \in R$ because the multiplication of R must be a binary operation on S.

We turn our attention now to the inheritance of properties of R. Certainly, if R is commutative, then any subrng S of R is commutative because the elements of S can be viewed as elements of R. Further, if R is a unital ring with multiplicative identity 1_R and S is a subrng of R that contains 1_R, then by Proposition 4.1.8, we conclude that S is a unital ring with multiplicative identity 1_R. □

Caution: the Subrng Test does not say that a unital ring has the same multiplicative identity as any overring; in fact, it is possible to find a unital subring S of a unital ring R whose multiplicative identity 1_S is distinct from the multiplicative identity 1_R of R (cf. Exercises 4.8.17 and 4.8.29).

Example 4.2.2. Each of the subset containments $\mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}$ induce subring containments.

Example 4.2.3. Consider the commutative unital ring $\mathbb{Z}/n\mathbb{Z}$ for some positive integer n. By the Fourth Isomorphism Theorem, the subgroups of $\mathbb{Z}/n\mathbb{Z}$ are precisely the groups $k\mathbb{Z}/n\mathbb{Z}$ such that $n\mathbb{Z} \subseteq k\mathbb{Z}$; the latter happens if and only if $n \in k\mathbb{Z}$ if and only if k divides n. Consequently, the only possible subrngs of $\mathbb{Z}/n\mathbb{Z}$ are $k\mathbb{Z}/n\mathbb{Z}$ for some positive integer $k \mid n$; these are always subrngs because $k\mathbb{Z}$ is closed under multiplication and subtraction, hence the cosets of $n\mathbb{Z}$ in $k\mathbb{Z}$ are, as well.

Below, we provide several useful properties that relate rng homomorphisms and subrngs.

Proposition 4.2.4. *Given any rng homomorphism $\varphi : R \to S$, the following hold.*

1.) *We have that $\varphi(0_R) = 0_S$.*

2.) *We have that $\varphi(r - s) = \varphi(r) - \varphi(s)$ for all elements $r, s \in R$.*

3.) *We have that $\varphi(T) = \{\varphi(t) \mid t \in T\}$ is a subrng of S for every subrng $T \subseteq R$.*

4.) *If φ is surjective and R is a unital ring, then S is a unital ring. Explicitly, if the multiplicative identity of R is 1_R, then the multiplicative identity of S must be $\varphi(1_R)$.*

5.) *If φ is surjective and R is a unital ring, then for any unit $u \in R$ with multiplicative inverse u^{-1}, we have that $\varphi(u)$ is a unit of S with multiplicative inverse $\varphi(u)^{-1} = \varphi(u^{-1})$.*

Proof. We prove (1.) and (2.) as follows. Considering that φ is a group homomorphism, we have that $\varphi(0_R) = 0_S$ by Proposition 3.3.5. Even more, by Proposition 4.1.8, it follows that

$$\varphi(r - s) = \varphi(r + (-s)) = \varphi(r) + \varphi(-s) = \varphi(r) - \varphi(s)$$

for all elements $r, s \in R$ because the additive inverse of $\varphi(-s)$ is none other than $-\varphi(s)$.

(3.) By the Subrng Test, it suffices to prove that $\varphi(T)$ is nonempty, closed under subtraction, and closed under multiplication. By (1.), it follows that $\varphi(T)$ is nonempty: indeed, as T is a subrng of R, it contains the additive identity 0_R of R, hence $\varphi(T)$ contains 0_S. Likewise, for any elements $s, t \in T$, we have that $\varphi(s) - \varphi(t) = \varphi(s - t)$ is an element of $\varphi(T)$ because the difference $s - t$ is an element of T by the Subrng Test. Last, for any elements $s, t \in T$, we have that $\varphi(s)\varphi(t) = \varphi(st)$ is an element of $\varphi(T)$ because the product st is an element of T by the Subrng Test.

We will assume toward a proof of properties (4.) and (5.) that φ is surjective and R is a unital ring with multiplicative identity 1_R. Given any element $s \in S$, there exists an element $r \in R$ such that $s = \varphi(r)$. Consequently, we have that $\varphi(1_R)s = \varphi(1_R)\varphi(r) = \varphi(1_R r) = \varphi(r) = s$, and the analogous argument shows that $s\varphi(1_R) = \varphi(r)\varphi(1_R) = \varphi(r1_R) = \varphi(r) = s$. We conclude therefore that S is a unital ring with multiplicative identity $\varphi(1_R)$. Given any unit $u \in R$, by Exercise 4.8.4, there exists a unique element $u^{-1} \in R$ such that $uu^{-1} = 1_R$. By applying our rng homomorphism φ, we find that $\varphi(1_R) = \varphi(uu^{-1}) = \varphi(u)\varphi(u^{-1})$ so that $\varphi(u)$ is a unit of S and $\varphi(u)^{-1} = \varphi(u^{-1})$. □

We state next and prove the following crucial property the kernel of a rng homomorphism.

Proposition 4.2.5. *Given any rng homomorphism $\varphi : R \to S$, we have that $\ker \varphi$ is a subrng of R that is closed under multiplication on the left and on the right by elements of R.*

Proof. By Proposition 4.2.4, it follows that $\varphi(0_R) = 0_S$ and $\varphi(r-s) = \varphi(r)-\varphi(s) = 0_S-0_S = 0_S$ for any elements $r, s \in \ker \varphi$, hence $\ker \varphi$ is a nonempty subset of R that is closed under subtraction. By the Subrng Test, it suffices to show that $\ker \varphi$ is closed under multiplication. We will prove moreover that $\ker \varphi$ is closed under multiplication on the left and on the right by elements of R. Given any elements $r \in \ker \varphi$ and $s \in R$, we have that $\varphi(rs) = \varphi(r)\varphi(s) = 0_S\varphi(s) = 0_S$ by Proposition 4.1.8; a similar argument illustrates that $rs \in \ker \varphi$ in the case that $r \in R$ and $s \in \ker \varphi$. □

We refer to a subrng I of R that is closed under left multiplication by elements of R as a left **ideal** of R; the analogous statement can be made to define right ideals of R; and ideals that are closed under multiplication on the left and right by elements of R are called two-sided ideals. Proposition 4.2.5 shows that the kernel of any rng homomorphism is a two-sided ideal of the rng on which the function is defined. Often, we will deal with commutative rngs, hence we will not qualify ideals as two-sided because any left ideal is automatically right ideal by commutativity (and vice-versa); however, in the case that R is non-commutative, we must distinguish between left ideals and right ideals. We say that an ideal I of a rng R is **proper** if it holds that $I \subsetneq R$. Observe that a proper ideal I of a unital ring R cannot contain the multiplicative identity 1_R of R: indeed, if 1_R lies in I, then by definition, we must have that $r = r1_R$ lies in I for all elements $r \in R$ so that $I = R$.

Example 4.2.6. Observe that $n\mathbb{Z}$ is an ideal of the commutative unital ring $\mathbb{Z}$ for any non-negative integer n: it is a nonempty subrng of $\mathbb{Z}$ satisfying that $s(nr) = n(rs) \in n\mathbb{Z}$ for any integers r and s.

Example 4.2.7. Consider the non-commutative unital ring $\mathbb{R}^{n\times n}$ consisting of real $n\times n$ matrices for some positive integer $n \geq 2$. Consider the set $I \subseteq \mathbb{R}^{n\times n}$ of all real $n \times n$ matrices whose first column consists entirely of zeros. Certainly, the zero matrix O lies in I, hence I is nonempty. Given any elements $A, B \in I$, we have that $A-B$ lies in I because matrix addition is performed componentwise and the first columns of A and $-B$ consist entirely of zeros. Last, matrix multiplication is carried out row-by-column, hence the first column of AB must consist entirely of zeros: explicitly, the first column of AB is determined by the product of the rows of A with the first column of B, so it is zero by assumption that the first column of B is zero. Consequently, we conclude that I is a subrng of $\mathbb{R}^{n\times n}$. We claim that it is a left ideal but not a right ideal. By the same rationale as before, for any $n \times n$ matrix A, the first column of AB must be zero, hence I is closed under multiplication on the left. On the other hand, the first column of BA is determined by the product of the rows of B with the first column of A, so if the first row of B is nonzero and the first column of A is nonzero, then it is possible that the first column of BA is nonzero, hence I is not closed under right multiplication.

Like with subrngs, there is a simple test to determine if a nonempty subset of a rng is an ideal.

Proposition 4.2.8 (Three-Step Ideal Test)**.** *Consider any rng R and any nonempty set $I \subseteq R$. We have that I is a two-sided ideal of R if and only if the following three properties hold.*

1.) *We have that $i - j \in I$ for all elements $i, j \in I$.*

2.) *We have that $ri \in I$ for all elements $r \in R$ and $i \in I$.*

3.) *We have that $ir \in I$ for all elements $r \in R$ and $i \in I$.*

Generally, if I satisfies the first and second conditions, then I is a left ideal of R. Likewise, if I satisfies the first and third conditions, then I is a right ideal of R.

Proof. By definition, a two-sided ideal I of R is a subrng of R that is closed under multiplication by elements of R. Consequently, if I is a two-sided ideal of R, then I must satisfy the three conditions above. Conversely, if I satisfies the first condition above, then by the One-Step Subgroup Test, we conclude that $(I, +)$ is a subgroup of $(R, +)$. Even more, if I satisfies the second and third conditions, then I is closed under multiplication by elements of R, hence I is a subrng of R (by the Subrng Test) that is closed under multiplication by elements of R, i.e., a two-sided ideal of R. □

Like with groups, we may consider ideals generated by a subset of elements of R.

Proposition 4.2.9. *Given any elements $x_1, \dots, x_n$ of any commutative rng R, we have that*

$$(x_1, \dots, x_n) = \{r_1 x_1 + \cdots + r_n x_n \mid r_1, \dots, r_n \in R\}$$

is an ideal of R called that is said to be **finitely generated** *by $x_1, \dots, x_n$.*

Proof. We note that $(x_1, \dots, x_n)$ contains $0_R = 0_R + \cdots + 0_R = 0_R x_1 + \cdots + 0_R x_n$ by Proposition 4.1.8, hence it is nonempty. By the Three-Step Ideal Test, in order to prove that $(x_1, \dots, x_n)$ is an ideal of R, it suffices to show that it is closed under subtraction and multiplication by elements of R. Both of these properties are straightforward to verify by the distributive property and commutativity: indeed, we have that $(r_1 x_1 + \cdots + r_n x_n) - (s_1 x_1 + \cdots + s_n x_n) = (r_1 - s_1)x_1 + \cdots + (r_n - s_n)x_n$ and $r(r_1 x_1 + \cdots + r_n x_n) = r(r_1 x_1) + \cdots + r(r_n x_n) = (rr_1)x_1 + \cdots + (rr_n)x_n$ for any elements $r \in R$ and $r_1 x_1 + \cdots + r_n x_n \in (x_1, \dots, x_n)$ by the associativity of multiplication. We conclude that $(x_1, \dots, x_n)$ is a left ideal of R, hence it must also be a right ideal of R by assumption that R is commutative. □

Caution: if R is not a unital ring, then it might not be the case that the ideal $(x_1, \dots, x_n)$ of R generated by $x_1, \dots, x_n$ contains the elements $x_1, \dots, x_n$ themselves. Consequently, we will restrict our attention to commutative unital rings when investigating these types of ideals.

Proposition 4.2.10. *Given any elements $x_1, \dots, x_n$ of any commutative unital ring R and any ideal I of R, we have that $I \supseteq (x_1, \dots, x_n)$ if and only if $x_1, \dots, x_n \in I$.*

Proof. Consider any ideal $I \subseteq R$ such that $x_1, \dots, x_n \in I$. We have that $r_1 x_1, \dots, r_n x_n \in I$ for all possible elements $r_1, \dots, r_n \in R$ because I is an ideal of R and must therefore be closed under multiplication by elements of R. Even more, we have that $r_1 x_1 + \cdots + r_n x_n \in I$ because I is an ideal of R and must therefore be closed under addition because it is a subrng of R. We conclude therefore that $I \supseteq (x_1, \dots, x_n)$. Conversely, every ideal $I \supseteq (x_1, \dots, x_n)$ must contain each of the generators $x_1, \dots, x_n$ since it holds that $x_i = 1_R x_i = 0_R x_1 + \cdots + 0_R x_{i-1} + 1_R x_i + 0_R x_{i+1} + \cdots + 0_R x_n$. □

Corollary 4.2.11. *Given any elements $x_1, \dots, x_n$ of any commutative unital ring R, $(x_1, \dots, x_n)$ is the smallest (with respect to inclusion) ideal of R that contains $x_1, \dots, x_n$.*

Remark 4.2.12. One can define finitely generated left ideals and finitely generated right ideals by mimicking the definition of Proposition 4.2.9. Explicitly, for any elements $x_1, \dots, x_n$ of a rng R,

$$R(x_1, \dots, x_n) = \{r_1 x_1 + \cdots + r_n x_n \mid r_1, \dots, r_n \in R\}$$

is a left ideal of R by the proof of the above proposition, and by analogy, one can demonstrate that

$$(x_1, \dots, x_n)R = \{x_1 r_1 + \cdots + x_n r_n \mid r_1, \dots, r_n \in R\}$$

is a right ideal of R. Of course, if R is commutative, then these ideals are equal. Each of Proposition 4.2.10 and Corollary 4.2.11 can likewise be generalized to the case of a non-commutative unital ring R. Explicitly, a left ideal I of R contains $R(x_1, \dots, x_n)$ if and only if it contains $x_1, \dots, x_n$, hence $R(x_1, \dots, x_n)$ is the smallest (with respect to inclusion) left ideal of R that contains $x_1, \dots, x_n$.

Example 4.2.13. Consider the polynomial ring $\mathbb{Z}[x]$. We may form the ideal $(2, x)$ of $\mathbb{Z}[x]$ generated by 2 and x. By definition, the elements of $(2, x)$ are of the form $ax + 2b$ for some integers a and b. Explicitly, every element of $(2, x)$ is either a constant polynomial that is divisible by 2 or a linear polynomial whose constant term is divisible by 2. Examples include 2, x, $-x + 2$, and $3x + 8$.

Example 4.2.14. Consider the commutative unital ring $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ under pointwise multiplication $(fg)(x) = f(x)g(x)$. We may form the ideal $(x, \sin x, \cos x)$ of all functions of the form $xf(x) + (\sin x)g(x) + (\cos x)h(x)$ for some real functions $f(x), g(x)$, and $h(x)$. Of course, each of the functions x, $\sin x$, and $\cos x$ lies in $F(\mathbb{R}, \mathbb{R})$, but it is also the case that the function $\sin^2(x) + \cos^2(x) = 1$ lies in this ideal. Consequently, $(x, \sin x, \cos x)$ is in fact $F(\mathbb{R}, \mathbb{R})$ in disguise.

Given any element x of a commutative rng R, we refer to the ideal $(x) = Rx = \{rx \mid r \in R\}$ as the **principal ideal** generated by x; this is a special case of Proposition 4.2.9. Even more, we say that a system of generators $x_1, \dots, x_n$ of an ideal $(x_1, \dots, x_n)$ is a **minimal system of generators** whenever $\{x_1, \dots, x_n\} \setminus \{x_i\}$ does not generate I for any integer $1 \leq i \leq n$. Put another way, if we delete one generator, then we obtain an ideal that is strictly contained in I. If an ideal I admits a finite system of generators, we say that I is **finitely generated**. Consequently, we may define

$$\mu(I) = \inf\{n \geq 0 \mid x_1, \dots, x_n \text{ form a minimal system of generators of } I\}.$$

Later, we will concern ourselves with the **minimal number of generators** $\mu(I)$ of an ideal I, but for now, we use the next example as an interesting motivational example for the reader.

Example 4.2.15. By Example 3.12.3, it follows that the abelian group $(\mathbb{Z}, +)$ is finitely generated. Even more, $\mathbb{Z}$ admits a minimal system of generators consisting of n integers for each integer $n \geq 1$. Explicitly, for any integer $n \geq 1$ and any collection of n distinct prime numbers $p_1, \dots, p_n$, the positive integers $x_i = p_1 \cdots p_n / p_i$ satisfy that $\gcd(x_1, \dots, x_n) = 1$, hence Bézout's Identity yields that $a_1 x_1 + \cdots + a_n x_n = 1$ for some integers $a_1, \dots, a_n$. Consequently, we may view $\mathbb{Z}$ as an ideal of itself generated by $(x_1, \dots, x_n)$. Even more, this system of generators is minimal since the greatest common divisor of the integers in the set $\{x_1, \dots, x_n\} \setminus \{x_i\}$ is in fact the prime number p_i. Put another way, the ideal generated by $\{x_1, \dots, x_n\} \setminus \{x_i\}$ is the principal ideal $p_i\mathbb{Z}$ and not $\mathbb{Z}$.

Our next proposition establishes that the generators of an ideal are not unique; rather, they can be chosen strategically so that the presentation of the ideal is as simple as possible.

Proposition 4.2.16. *Let R be a commutative unital ring with an ideal $I = (x_1, \dots, x_n)$. Consider the ideal $J = (x_1, \dots, x_{i-1}, u_1 x_1 + \cdots + u_n x_n, x_{i+1}, \dots, x_n)$ for some units $u_1, \dots, u_n \in R$, i.e., the ideal of R generated by the elements of $\{x_1, \dots, x_n, u_1 x_1 + \cdots + u_n x_n\} \setminus \{x_i\}$. We have that $I = J$.*

Proof. We can immediately verify that $J \subseteq I$ by Proposition 4.2.10 because each of the generators of J is an element of I. Conversely, each of the generators x_j of I for $j \neq i$ is an element of J, hence it suffices to prove that x_i is in J. Observe that $u_i x_i = u_1 x_1 + \cdots + u_n x_n + \sum_{j \neq i} (-u_j) x_j$ is an element of J so that $x_i = 1_R x_i = (u_i^{-1} u_i) x_i = u_i^{-1}(u_i x_i)$ is in J. We conclude that $I \subseteq J$. □

Example 4.2.17. Let us find the simplest system of generators for the ideal $I = (4, 6)$ in $\mathbb{Z}$. Every element of I can be written as $4m + 6n = 2(2m + 3n)$, from which it follows that $(4, 6) \subseteq (2)$. Conversely, we have that $2 = 4(-1) + 6(1)$ is an element of I, hence we must have that $I = (2)$.

Example 4.2.18. Let us find the simplest system of generators for the ideal $I = (2, 4, 6, 9)$ in $\mathbb{Z}$. Observe that $-4 \cdot 2 + 0 \cdot 4 + 0 \cdot 6 + 1 \cdot 9 = 1$ is an element of I, hence we conclude that $I = \mathbb{Z} = (1)$. Exercise 4.8.28 demonstrates that the previous examples are indicative of a general phenomenon.

Example 4.2.19. Consider the ideal $I = (x^2 - 1, x^3 - x, x^4 - x^2)$ of $\mathbb{R}[x]$. By Proposition 4.2.16, we can replace any of the generators of I by a linear combination of the generators so long as the coefficients of this linear combination are units of $\mathbb{R}[x]$. Particularly, we may replace $x^4 - x^2$ by $x^4 - 1 = (x^4 - x^2) + (x^2 - 1)$. On the other hand, we have that $x^4 - 1 = (x^2 - 1)(x^2 + 1)$, hence every polynomial of the form $p(x)(x^4 - 1)$ can now be realized as a polynomial $p(x)(x^2 + 1)(x^2 - 1)$, and we can dispose of the generator $x^4 - x^2$ of I. Likewise, we have that $x^3 - x = x(x^2 - 1)$, hence every polynomial of the form $q(x)(x^3 - x)$ can be realized as a polynomial $q(x)x(x^2 - 1)$, and we can dispose of the generator $x^3 - x$. Consequently, we find that $I = (x^2 - 1)$.

We will now discuss how to construct important new ideals by performing set and rng operations on existing ideals. Given any left ideals I and J of a rng R, it is natural to consider the behavior of I and J with respect to set operations such as intersection and union. By Exercise 4.8.22, it turns out that the intersection $I \cap J = \{k \in R \mid k \in I \text{ and } k \in J\}$ of left ideals yields a left ideal of R; however, it is rarely the case that the union $I \cup J = \{k \in R \mid k \in I \text{ or } k \in J\}$ of ideals is an ideal of R. Even more, considering that I and J are normal subgroups of the abelian group $(R, +)$, it is possible by Exercise 2.8.23 to form the normal subgroup $I + J = \{i + j \mid i \in I \text{ and } j \in J\}$ of $(R, +)$; it is not difficult to check that $I + J$ is a left ideal of R. Last, if I is a left ideal and J is a right ideal, we may also define the **product ideal** $IJ = \{i_1 j_1 + \cdots + i_n j_n \mid n \geq 1, i_1, \dots, i_n \in I, j_1, \dots, j_n \in J\}$ of I and J. Crucially, notice the definition of this two-sided ideal as the set of all possible sums of products of an element of I and an element of J. Even though it is most natural to hope that $I * J = \{ij \mid i \in I \text{ and } j \in J\}$ is an ideal of R, Exercise 4.8.23 shows that this is not true in general. Our next proposition illuminates the relationship between the ideals IJ, $I \cap J$, I, J, and $I + J$.

Proposition 4.2.20. *Given any left ideals I and J of any rng R, we have the left ideal containments $I \cap J \subseteq I \subseteq I + J$ and $I \cap J \subseteq J \subseteq I + J$. If I and J are two-sided ideals, then $IJ \subseteq I \cap J$.*

Proof. We leave it as Exercise 4.8.22 to prove that $I \cap J$ and $I + J$ are left ideals of R and that IJ is a two-sided ideal of R if I is a left ideal and J is a right ideal of R. Given any element $k \in I \cap J$, we have that $k \in I$ and $k \in J$ so that $I \cap J \subseteq I$ and $I \cap J \subseteq J$. Even more, for any elements $i \in I$ and $j \in J$, we have that $i = i + 0_R \in I + J$ and $j = 0_R + j \in I + J$, from which it follows that $I \subseteq I + J$ and $J \subseteq I + J$. Last, if I is a right ideal and J is a left ideal of R, then for every element $i \in I$ and $j \in J$, we have that $ij \in I \cap J$ because I is a right ideal of R and J is a left ideal of R. Consequently, for every integer $n \geq 1$ and any elements $i_1, \dots, i_n \in I$ and $j_1, \dots, j_n \in I$, the closure of I and J under addition yields that $i_1 j_1 + \cdots + i_n j_n \in I \cap J$ so that $IJ \subseteq I \cap J$. □

Products of finitely generated ideals of a commutative rng are especially simple to describe.

Proposition 4.2.21. *For any elements $x_1, \dots, x_m, y_1, \dots, y_n$ of a commutative rng R, we have that*

$$(x_1, \dots, x_m)(y_1, \dots, y_n) = (x_i y_j \mid 1 \leq i \leq m \text{ and } 1 \leq j \leq n).$$

Put another way, the product of any finitely generated ideals of a commutative rng is a finitely generated ideal that can be generated by the products of the generators of the underlying ideals.

Proof. By Exercise 4.8.22, it follows that $(x_1, \dots, x_m)(y_1, \dots, y_n)$ is a two-sided ideal of R. Each of the products $x_i y_j$ with $1 \leq i \leq m$ and $1 \leq j \leq n$ lies in this ideal, hence it follows by Proposition 4.2.10 that the ideal generated by $x_i y_j$ for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$ lies in $(x_1, \dots, x_m)(y_1, \dots, y_n)$. Conversely, every element of the product ideal is of the form $i_1 j_1 + \cdots + i_\ell j_\ell$ for some elements $i_1, \dots, i_\ell \in (x_1, \dots, x_m)$ and $j_1, \dots, j_\ell \in (y_1, \dots, y_n)$. Consequently, it suffices to prove that every product ij of an element $i \in (x_1, \dots, x_m)$ and an element $j \in (y_1, \dots, y_n)$ is an element of the ideal $(x_i y_j \mid 1 \leq i \leq m$ and $1 \leq j \leq n)$. But this is not difficult: observe that if $i = r_1 x_1 + \cdots + r_m x_m$ and $j = s_1 y_1 + \cdots + s_n y_n$ for some elements $r_1, \dots, r_m, s_1, \dots, s_n \in R$, then

$$ij = (r_1 x_1 + \cdots + r_m x_m)(s_1 y_1 + \cdots + s_n y_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} r_i s_j x_i y_j$$

is an element of the finitely generated ideal $(x_i y_j \mid 1 \leq i \leq m$ and $1 \leq j \leq n)$ by distributivity. □

By the One-Step Subgroup Test and by the Subrng Test, it follows that an ideal I of a rng R is a normal subgroup of the abelian group $(R, +)$, hence we have that $(R/I, +)$ is an abelian group with respect to the usual left coset addition defined by $(r + I) + (s + I) = (r + s) + I$. Even more, if I is a two-sided ideal, consider the multiplication of left cosets prescribed by $(r + I)(s + I) = rs + I$. We must check that this is well-defined. Given that any pair of left coset representatives $r + I = x + I$ and $s + I = y + I$, it follows that $r = r + 0_R = x + i$ and $s = s + 0_R = y + j$ for some elements $i, j \in I$. Consequently, we have that $rs = (x + i)(y + j) = xy + xj + iy + ij$. By hypothesis that I is a two-sided ideal of R, it follows that xj, iy, and ij are elements of I so that $xj + iy + ij$ lies in I and $xj + iy + ij + I = 0_R + I$. We conclude that $(r + I)(s + I) = rs + I = xy + I = (x + I)(y + I)$, as desired. Ultimately, this demonstrates that R/I is a rng with respect to the prescribed addition and multiplication defined on the left cosets of I in R: it is called the **quotient rng** of R modulo I. One can readily verify that if R is commutative, then R/I is commutative, and if R is a unital ring with multiplicative identity 1_R, then R/I is a unital ring with multiplicative identity $1_R + I$.

Example 4.2.22. Given any positive integer n, we have seen that $n\mathbb{Z}$ is a two-sided ideal of the ring $\mathbb{Z}$, hence we can form the quotient ring $\mathbb{Z}/n\mathbb{Z}$; this is the ring defined in Example 4.1.3.

Example 4.2.23. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Every ideal of $\mathbb{R}[x]$ is two-sided, hence we can form the quotient ring $\mathbb{R}[x]/(x)$ of $\mathbb{R}[x]$ modulo the principal ideal (x) generated by the monomial x. By definition, every element of $\mathbb{R}[x]/(x)$ is of the form $p(x) + (x)$ for some polynomial $p(x) \in \mathbb{R}[x]$. Considering that every element of (x) is of the form $q(x)x$ for some polynomial $q(x) \in \mathbb{R}[x]$, it follows that (x) consists precisely of the univariate real polynomials that are divisible by x; thus, if we write $p(x) = a_n x^n + \cdots + a_1 x + a_0$ for some real numbers $a_n, \dots, a_1, a_0$, then $a_n x^n + \cdots + a_1 x = (a_n x^{n-1} + \cdots + a_1)x \in (x)$ so that $p(x) + (x) = a_n x^n + \cdots + a_1 x + a_0 + (x) = (a_n x^{n-1} + \cdots + a_1)x + a_0 + (x) = a_0 + (x)$ because (x) absorbs any polynomial that is divisible by x. We conclude that $\mathbb{R}[x]/(x) = \{a + (x) \mid a \in \mathbb{R}\}$.

Example 4.2.24. Consider the commutative unital ring $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ under pointwise multiplication $(fg)(x) = f(x)g(x)$. Consider the collection I of real functions that pass through the origin, i.e., the set $I = \{f : \mathbb{R} \to \mathbb{R} \mid f(0) = 0\}$. Observe that the constant function zero lies in I, hence it is nonempty; the Three-Step Ideal Test yields that I is an ideal of $F(\mathbb{R}, \mathbb{R})$ because it holds that $(f - g)(0) = f(0) - g(0) = 0$ and $(fg)(0) = f(0)g(0) = 0$ for all functions $f, g \in I$. Consequently, we may form the quotient ring $F(\mathbb{R}, \mathbb{R})/I$ of $F(\mathbb{R}, \mathbb{R})$ modulo I whose elements are by definition left cosets of the form $f(x) + I$. Every real function $f : \mathbb{R} \to \mathbb{R}$ that passes through the origin is identified with the zero function modulo I, hence the nonzero elements of $F(\mathbb{R}, \mathbb{R})/I$ are precisely those functions $f : \mathbb{R} \to \mathbb{R}$ that do not pass through the origin. Explicitly, every polynomial function $p(x) = a_n x^n + \cdots + a_1 x + a_0$ is identified with its constant term modulo I, and the functions $\sin x$ and x^2 satisfy that $\sin x + I = 0 + I = x^2 + I$. Even more bizarrely, we have that $e^0 = 1$ so that $e^x - 1$ is identically zero modulo I and $e^x + I = 1 + I = \cos x + I$, i.e., the images of $\cos x$ and the exponential function e^x are identified with the constant function 1 modulo I.

We conclude with an indispensable result indicating how to construct two-sided ideals of a rng.

Proposition 4.2.25. *Every two-sided ideal of a rng R is the kernel of a rng homomorphism from R. Consequently, the two-sided ideals of R are precisely the kernels of rng homomorphisms from R.*

Proof. Given any two-sided ideal I of R, we have that R/I is a rng with respect to the multiplication $(r + I)(s + I) = rs + I$. Consequently, the **canonical projection** function $\pi : R \to R/I$ defined by $\pi(r) = r + I$ is a rng homomorphism. Observe that r lies in $\ker \pi$ if and only if $r + I = 0_R + I$ if and only if $r \in I$ by Proposition 3.1.4, from which it follows that $\ker \pi = I$, as desired. Conversely, Proposition 4.2.5 shows that $\ker \varphi$ is a two-sided ideal for any rng homomorphism $\varphi : R \to S$. □

4.3 The Ring Isomorphism Theorems

We provide in this section analogs of the Group Isomorphism Theorems of Section 3.5 for rngs.

Theorem 4.3.1 (First Isomorphism Theorem for Rngs)**.** *Given any rngs R and S and any rng homomorphism $\varphi : R \to S$, there exists a rng isomorphism $\psi : R/\ker \varphi \to \varphi(R)$.*

Proof. By Proposition 4.2.4, we have that $\varphi(R)$ is a subrng of S. Considering that $\ker \varphi$ is a two-sided ideal of R by Proposition 4.2.5, we may view $R/\ker \varphi$ as a rng with multiplication defined by $(r + \ker \varphi)(s + \ker \varphi) = rs + \ker \varphi$. We claim that the function $\psi : R/\ker \varphi \to \varphi(R)$ defined by $\psi(r + \ker \varphi) = \varphi(r)$ is a well-defined rng isomorphism. We must show that if $r + \ker \varphi = s + \ker \varphi$, then $\psi(r+\ker\varphi) = \psi(s+\ker\varphi)$. By Propositions 3.1.4 and 4.2.4, we have that $r+\ker\varphi = s+\ker\varphi$ if and only if $(r-s)+\ker\varphi = 0_R+\ker\varphi$ if and only if $r-s \in \ker\varphi$ if and only if $\varphi(r-s) = 0_S$ if and only if $\varphi(r)-\varphi(s) = 0_S$ if and only if $\varphi(r) = \varphi(s)$ if and only if $\psi(r+\ker\varphi) = \psi(s+\ker\varphi)$. We conclude that ψ is well-defined. By hypothesis that φ is a rng homomorphism, it follows that ψ is a rng homomorphism, and ψ is clearly surjective, hence it suffices to show that ψ is injective. Observe that $r+\ker\varphi$ is in $\ker\psi$ if and only if $\varphi(r) = \psi(r+\ker\varphi) = 0_S$ if and only if r is in $\ker\varphi$ if and only if $r+\ker\varphi = 0_R+\ker\varphi$ so that $\ker\psi$ is trivial and ψ is injective, as desired. □

Theorem 4.3.2 (Second Isomorphism Theorem for Rngs)**.** *Given any rng R with a subrng S and a two-sided ideal I, we have that $(S+I)/I$ and $S/(I\cap S)$ are isomorphic rngs.*

Theorem 4.3.3 (Third Isomorphism Theorem for Rngs)**.** *Given a rng R with two-sided ideals I and J such that $J\subseteq I$, we have that $(R/J)/(I/J)$ and R/I are isomorphic rngs.*

We leave these as exercises; they are proved analogously to the Group Isomorphism Theorems.

Theorem 4.3.4 (Fourth Isomorphism Theorem for Rngs)**.** *Given a rng R with a two-sided ideal I, there exists a one-to-one correspondence between the subrngs of R that contain I and the subrngs of R/I induced by the assignment of a subrng S of R with $I\subseteq S$ to the subrng S/I of R/I. Even more, this one-to-one correspondence satisfies the following properties.*

1.) *Given any subrngs S and T of R such that $I\subseteq S$ and $I\subseteq T$, we have that $S\subseteq T$ if and only if $S/I\subseteq T/I$. Put another way, this bijective correspondence is inclusion-preserving.*

2.) *Given any subrng S of R that contains the two-sided ideal I, we have that S is an ideal of R if and only if the set S/I of left cosets of I in S is an ideal of R/I.*

Proof. We must first establish that the assignment of a subrng S of R with $I\subseteq S$ to a subrng S/I of R/I is well-defined, injective, and surjective. Considering that I is a two-sided ideal of R that is contained in S, it is a two-sided ideal of S, hence the quotient rng S/I is well-defined; it is a subrng of R/I because S is a subrng of R. Consider any pair of subrngs S and T of R such that $S/I = T/I$. We must prove that $S=T$. Given any element $s\in S$, there exist elements $t\in T$ and $i\in I$ such that $s = s+0_R = t+i$, from which it follows that s lies in T and $S\subseteq T$ because T is a subrng of R that contains I. By the same argument applied to the elements of T, we conclude that $T\subseteq S$, as desired. We will demonstrate next that every subrng Q of R/I is of the form S/I for some subrng S of R such that $I\subseteq S$. Consider the collection $S = \{r\in R \mid r+I\in Q\}$ of elements of R whose images modulo I lie in the subrng Q of R/I. We claim that S is a subrng of R that contains I and satisfies that $Q = S/I$. By assumption that Q is a subrng of R/I, it follows by the Subrng Test that $0_R+I\in Q$ so that $0_R\in S$. Likewise, for any elements $r,s\in S$, we have that

$$(r-s)+I = (r+(-s))+I = (r+I)+(-s+I) = (r+I)-(s+I)$$

by Proposition 4.1.8. Considering that $r+I$ and $s+I$ lie in the subrng Q of R/I, their difference lies in Q, hence we find that $r-s\in S$. By the same rationale, the product rs lies in S because it satisfies that $rs+I = (r+I)(s+I)$ and the left cosets $r+I$ and $s+I$ both lie in Q. We conclude that S is a subrng of R; it contains I because for every element $i\in I$, we have that $i+I = 0_R+I$ by Proposition 3.1.4; and it is straightforward to verify that $Q = \{r+I \mid r+I\in Q\} = S/I$.

By the previous paragraph, the only assertion that remains to be seen is the second property. One need not think too hard to prove that if S is an ideal of R, then S/I is an ideal of R/I: indeed, this follows because $(r+I)(s+I) = rs + I$ lies in S/I for every element $r + I \in R/I$ by assumption that S is an ideal of R. Conversely, if S/I is an ideal of R/I, then for every element $r \in R$ and every element $s \in S$, we have that $rs + I = (r+I)(s+I)$ is an element of S/I so that $rs \in S$. □

Example 4.3.5. Consider the commutative unital ring $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ under pointwise multiplication $(fg)(x) = f(x)g(x)$. We may define a function $\varphi : F(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ by declaring that $\varphi(f(x)) = f(0)$; explicitly, φ evaluates the function $f(x)$ at 0. Observe that φ is a group homomorphism because $\varphi(f(x) + g(x)) = \varphi((f+g)(x)) = (f+g)(0) = f(0) + g(0) = \varphi(f(x)) + \varphi(g(x))$ and $\varphi(f(x)g(x)) = \varphi((fg)(x)) = (fg)(0) = f(0)g(0)$, hence φ is a unital ring homomorphism; often, it is referred to simply as the **evaluation homomorphism** at 0. Considering that for every real number C, the constant function $f_C(x) = C$ satisfies that $C = f_C(0) = \varphi(f_C(x))$, it follows that φ is surjective. Even more, we have that $f(x) \in \ker \varphi$ if and only if $f(0) = \varphi(f(x)) = 0$, hence the kernel of φ consists of all functions $f : \mathbb{R} \to \mathbb{R}$ that pass through the origin, i.e., it is the ideal from Example 4.2.24. By the First Isomorphism Theorem for Rngs, we have that $R/\ker \varphi \cong \mathbb{R}$.

Example 4.3.6. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x as a unital subring of $F(\mathbb{R}, \mathbb{R})$. We have seen previously that evaluation at 0 is a unital ring homomorphism, i.e., the function $\varphi : \mathbb{R}[x] \to \mathbb{R}$ defined by $\varphi_0(p(x)) = p(0)$ is a unital ring homomorphism; its kernel consists of all polynomials that pass through the origin. Observe that a polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$ passes through the origin if and only if $0 = p(0) = a_0$ if and only if $p(x) = (a_n x^{n-1} + \cdots + a_1)x$, hence the kernel consists of all polynomials that can be written as $p(x) = q(x)x$ for some polynomial $q(x)$. Put another way, we have that $\ker \varphi_0 = (x)$. Once again, the First Isomorphism Theorem for Rngs guarantees that $\mathbb{R}[x]/(x) \cong \mathbb{R}$ (cf. Example 4.2.23).

Example 4.3.7. Consider the commutative unital rings $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}/mn\mathbb{Z}$ of integers modulo some positive integer n and mn, respectively. Observe that $n\mathbb{Z}/mn\mathbb{Z}$ is a two-sided ideal $\mathbb{Z}/mn\mathbb{Z}$: indeed, we have already seen that $n\mathbb{Z}/mn\mathbb{Z}$ is an additive abelian group, and moreover, for any integer a and any left coset representative $nk + mn\mathbb{Z}$ of $n\mathbb{Z}$ in $mn\mathbb{Z}$, we have that $(a + mn\mathbb{Z})(nk + mn\mathbb{Z}) = n(ak) + mn\mathbb{Z}$ lies in $n\mathbb{Z}/mn\mathbb{Z}$. By the Third Isomorphism Theorem for Rngs, we conclude that

$$\frac{\mathbb{Z}/mn\mathbb{Z}}{n\mathbb{Z}/mn\mathbb{Z}} \cong \frac{\mathbb{Z}}{n\mathbb{Z}}.$$

Consequently, it grants no additional information to take subsequent quotients of $\mathbb{Z}$.

4.4 Integral Domains and Fields

We have seen so far that a rng R is an abelian group $(R, +)$ with an associative and distributive multiplication. We have reserved the terminology of unital ring for any rng R that admits a unique multiplicative identity element 1_R satisfying that $r1_R = r = 1_R r$ for every element r of R. We make no assumption that the order of multiplication in a rng is irrelevant; rather, we distinguish a rng R as commutative if it holds that $rs = sr$ for all elements $r, s \in R$. Generally, the order of multiplication matters in non-commutative rngs such as the unital ring $\mathbb{R}^{n \times n}$ of real $n \times n$ matrices.

We say that an element r of a rng R is left **regular** if $rs = 0_R$ implies that $s = 0_R$. We will soon alternatively refer to these elements as left **cancellable** (cf. Proposition 4.4.10). Conversely, a left **zero divisor** is any element $r \in R$ for which $rs = 0_R$ for some nonzero element $s \in R$.

Example 4.4.1. Consider the commutative unital ring $\mathbb{Z}/n\mathbb{Z}$ for any positive integer n. Observe that if k is any positive non-trivial divisor of n, then $k + n\mathbb{Z}$ is a zero divisor of $n\mathbb{Z}$. Explicitly, in this case, there exists an integer $q > 1$ such that $n = kq$, hence the left cosets $k + n\mathbb{Z}$ and $q + n\mathbb{Z}$ are nonzero and satisfy that $(k + n\mathbb{Z})(q + n\mathbb{Z}) = kq + n\mathbb{Z} = n + n\mathbb{Z} = 0 + n\mathbb{Z}$ by Proposition 3.1.4. Concretely, if $n = 30$, then the zero divisors of $\mathbb{Z}/30\mathbb{Z}$ are $2 + 30\mathbb{Z}$, $3 + 30\mathbb{Z}$, $5 + 30\mathbb{Z}$, $6 + 30\mathbb{Z}$, $10 + 30\mathbb{Z}$, and $15 + 30\mathbb{Z}$ because the non-trivial divisors of 30 are 2, 3, 5, 6, 10, and 15.

Conversely, by definition, the regular elements of $\mathbb{Z}/n\mathbb{Z}$ are those left cosets $a + n\mathbb{Z}$ satisfying that $ab + n\mathbb{Z} = (a + n\mathbb{Z})(b + n\mathbb{Z}) = 0 + n\mathbb{Z}$ implies that $b + n\mathbb{Z} = 0 + n\mathbb{Z}$. Put another way, the left coset $a + n\mathbb{Z}$ is a regular element of $\mathbb{Z}/n\mathbb{Z}$ if and only if $n \mid ab$ implies that $n \mid b$ if and only if $\gcd(n, a) = 1$ by Exercise 1.10.28 if and only if $a + n\mathbb{Z}$ is a unit by Example 4.1.11. Explicitly, if $\gcd(a, n) = d > 1$, then by Exercise 1.10.34, it follows that $a + n\mathbb{Z}$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$.

Example 4.4.2. Consider the following real 2×2 matrices.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$$

Observe that AB is the zero matrix, hence A is a left zero divisor and B is a right zero divisor. Conversely, we have that A^2 is the zero matrix, hence A is a right zero divisor. Likewise, we have that BC is the zero matrix, hence B is a right zero divisor and C is a left zero divisor.

Example 4.4.3. External direct products of rngs always admit non-trivial left zero divisors. Explicitly, if R and S are any rngs, then for any nonzero elements $r \in R$ and $s \in S$, we have that $(r, 0_S)$ and $(0_R, s)$ are nonzero elements of $R \times S$ such that $(r, 0_S)(0_R, s) = (0_R, 0_S) = 0_{R \times S}$.

Example 4.4.4. Observe that if n is any nonzero integer, then $mn = 0$ if and only if $m = 0$ because we can divide both sides of the equation $mn = 0$ by n. Consequently, there are no non-trivial zero divisors of $\mathbb{Z}$. Put another way, every nonzero element of $\mathbb{Z}$ is regular. By the same argument, the nonzero elements of the commutative unital rings $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all regular.

Definition 4.4.5 (Zero Product Property)**.** Given any rng R, we say that a nonzero element $r \in R$ satisfies the **Zero Product Property** if it holds that $rs = 0_R$ only if $s = 0_R$.

Consequently, every regular element of a rng satisfies the Zero Product Property. We refer to a unital ring R in which every nonzero element satisfies the Zero Product Property as a **domain**; commutative domains are called **integral domains**. Recall that an element u of a unital ring R is a unit if there exists a unique element $u^{-1} \in R$ such that $uu^{-1} = 1_R = u^{-1}u$. Unital rings in which every nonzero element is a unit are called **skew fields**; commutative skew fields are simply **fields**.

Example 4.4.6. Observe that the integers $\mathbb{Z}$ form an integral domain that is not a field: for any integer $n \geq 2$, the multiplicative inverse of n is a non-integral rational number. Put another way, by Example 4.1.10, we have that $U(\mathbb{Z}) = \{1, -1\}$, but there are infinitely many nonzero integers.

Example 4.4.7. Consider the commutative unital ring $\mathbb{Z}/p\mathbb{Z}$ for any prime number p. By Example 4.1.11, every nonzero element of $\mathbb{Z}/p\mathbb{Z}$ is a unit. Consequently, the only non-unit in $\mathbb{Z}/p\mathbb{Z}$ is the zero coset $0 + p\mathbb{Z}$, hence $\mathbb{Z}/p\mathbb{Z}$ is a field with p elements, i.e., it is a **finite field**.

Example 4.4.8. Observe that the rational numbers $\mathbb{Q}$ form a field because every nonzero element of $\mathbb{Q}$ can be written as $\frac{r}{s}$ for some nonzero integers r and s with $\gcd(r, s) = 1$ so that $\frac{r}{s} \cdot \frac{s}{r} = 1$.

Our aim throughout the rest of this section and in the next chapter is to understand to what extent an (integral) domain fails to be a (skew) field. Before this, we record several immediate propositions regarding the especially nice properties of (integral) domains and (skew) fields.

Proposition 4.4.9. *Every skew field is a domain. Consequently, every field is an integral domain.*

Proof. We must prove that every nonzero element of a skew field k satisfies the Zero Product Property, i.e., if u is a nonzero element of k and $uv = 0_k$, then $v = 0_k$. Every nonzero element $u \in k$ admits a unique multiplicative inverse u^{-1} such that $u^{-1}u = 1_k$, hence for any element $v \in k$ such that $uv = 0_k$, it follows that $0_k = u^{-1}0_k = u^{-1}(uv) = (u^{-1}u)v = 1_k v = v$ by Proposition 4.1.8. □

Proposition 4.4.10. *Cancellation of nonzero factors in products is a valid operation in a domain.*

Proof. Consider any elements $r, s,$ and t of any domain R such that $rs = rt$ and r is nonzero. We claim that $s = t$. We have that $rs - rt = 0_R$ so that $r(s - t) = 0_R$. By assumption that R is a domain and r is a nonzero element of R, we conclude that $s - t = 0_R$ so that $s = t$. □

Proposition 4.4.11. *Every nonzero unital subring of a skew field is a domain.*

Proof. Consider a nonzero element r of a nonzero unital subring R of a skew field k. Observe that if $rs = 0_k$ for some element $s \in R$, then we must have that $s = 0_k$ by Proposition 4.4.9. Explicitly, if we view the equation $rs = 0_k$ as an equation in the elements of k, then we may multiply on the left-hand side by the unique multiplicative inverse r^{-1} of R to obtain the desired result. □

Corollary 4.4.12. *Every nonzero unital subring of a field is an integral domain.*

Example 4.4.13. Consider the nonempty subset $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ of the complex numbers $\mathbb{C}$. By the Subrng Test, it is straightforward to verify that $\mathbb{Z}[i]$ is a commutative unital subring of $\mathbb{C}$ called the **Gaussian integers**: complex subtraction obeys $(a+bi)-(c+di) = (a-c)+(b-d)i$, and complex multiplication satisfies that $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$. Both of these complex numbers will have integral components so long as $a, b, c,$ and d are integers in the first place. Corollary 4.4.12 ensures that the Gaussian integers form an integral domain.

Proposition 4.4.14. *The characteristic of an integral domain is either zero or a prime number.*

Proof. We will assume that the characteristic of R is nonzero. By definition, there exists a smallest positive integer $n \geq 2$ such that $n \cdot 1_R = 0_R$. Consider the smallest prime number p that divides n. Observe that if $p \cdot 1_R = 0_R$, then the characteristic of R is by definition the prime number p; otherwise, we have that $p \cdot 1_R$ is cancellable in R by assumption that R is an integral domain. Consequently, the element $p^k \cdot 1_R$ of R corresponding to the largest power p^k of p that divides n must be cancellable in R; otherwise, $p \cdot 1_R$ would be a zero divisor of R. We conclude that

$$(p^k \cdot 1_R)\left(\frac{n}{p^k} \cdot 1_R\right) = n \cdot 1_R = 0_R \text{ only if } \frac{n}{p^k} \cdot 1_R = 0_R.$$

Continuing in this manner for the smallest prime number that divides each subsequent quotient of n must eventually produce a smallest prime number q satisfying that $q \cdot 1_R = 0_R$. □

Example 4.4.15. By Example 4.1.20, the characteristic of $\mathbb{Z}/p\mathbb{Z}$ is p whenever p is prime.

Considering our examples so far, we have the following hierarchy of commutative unital rings.

$$\text{finite fields} \subsetneq \text{(skew) fields} \subsetneq \text{(integral) domains} \subsetneq \text{(commutative) unital rings}$$

Later, in Chapter 5, we will specialize this hierarchy to discuss different types of integral domains; for now, we continue to explore the relationship between (integral) domains and (skew) fields.

Proposition 4.4.16. *Every nonzero finite integral domain is a field.*

Proof. We must demonstrate that for any nonzero element x of an integral domain R, there exists a nonzero element $y \in R$ such that $xy = 1_R$. Consider the function $\varphi_x : R \to R$ defined by $\varphi_x(r) = rx$. By hypothesis that R is a domain and x is a nonzero element of R, it follows that x is cancellable. Consequently, φ is injective: indeed, we have that $\varphi_x(r) = \varphi_x(s)$ if and only if $rx = sx$ if and only if $r = s$ by Proposition 4.4.10. Considering that R is finite, it follows that φ is surjective by Exercise 1.10.5, hence there exists a nonzero element $y \in R$ such that $1_R = xy = \varphi_x(y)$, as desired. □

Consequently, if we wish to study (integral) domains that are not (skew) fields, then we must focus our attention on those (integral) domains with infinitely many elements. Our next proposition yields a very restrictive and useful conditions on the two-sided ideals of a (skew) field.

Proposition 4.4.17. *Given any unital ring homomorphism $\varphi : k \to R$ from any skew field k to any unital ring R, we must have that either φ is injective or φ is the zero function.*

Proof. If φ is not injective, then there exists a nonzero element $x \in \ker \varphi$. By hypothesis that k is a skew field, there exists a unique element $x^{-1} \in k$ such that $x^{-1}x = 1_k$. Considering that $\ker \varphi$ is a two-sided ideal by Proposition 4.2.5, it follows that $1_R = x^{-1}x$ is an element of $\ker \varphi$. But this implies that for every element $y \in k$, we have that $\varphi(y) = \varphi(1_k y) = \varphi(1_k)\varphi(y) = 0_R \varphi(y) = 0_R$. □

Corollary 4.4.18. *Every surjective unital ring homomorphism $\varphi : k \to R$ from any skew field k to any nonzero unital ring R is a unital ring isomorphism.*

Proof. Considering that φ is surjective and R is a nonzero unital ring, it follows that φ is not the zero function. Consequently, by Proposition 4.4.17, we conclude that φ is injective. □

Corollary 4.4.19. *If k is a skew field, then its only two-sided ideals are the zero ideal and itself.*

Proof. By Proposition 4.2.25, every two-sided ideal of k is the kernel of some unital ring homomorphism from k. By Proposition 4.4.17, the kernel of a unital ring homomorphism from k is either the zero ideal (in the case that the unital ring homomorphism is injective) or the entire skew field k itself (in the case that the unital ring homomorphism is the zero function). □

4.5 Prime and Maximal Ideals

We have seen thus far in this chapter that the underlying structure of a rng as an additive abelian group endows a rng with many of the same properties as an abelian group. Explicitly, rng homomorphisms are group homomorphisms that preserve multiplication; two-sided ideals of rngs are analogous to normal subgroups; quotient rngs are analogous to quotient groups; and there four isomorphism theorems for rngs that extend the four isomorphism theorems for groups.

Our immediate aim throughout this section is to impress that the multiplicative structure of a rng makes it much more interesting; it is to this end that (at last) we restrict our attention to commutative unital rings. We will therefore not make any distinction between ideals and two-sided ideals here because they are the same notion. Even more, all of the work that we have done so far is valid in this setting because it holds in the much broader context of arbitrary rngs.

We begin by saying that a proper ideal P of a commutative unital ring R is **prime** if it has the property that for all elements $r, s \in R$ such that $rs \in P$, we must have that either $r \in P$ or $s \in P$.

Example 4.5.1. Consider the principal ideal $5\mathbb{Z} = \{5k : k \in \mathbb{Z}\}$ of the commutative unital ring $\mathbb{Z}$ of integers. Given any integers m and n such that $mn \in 5\mathbb{Z}$, by definition, we must have that $mn = 5k$ for some integer k, from which it follows that $5 \mid mn$. Considering that 5 is a prime number, we must have that $5 \mid m$ or $5 \mid n$ so that $m \in 5\mathbb{Z}$ or $n \in 5\mathbb{Z}$. Put another way, $5\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$. Ultimately, this example serves to show that prime ideals are a generalization of prime numbers. We conclude by noting that $\mathbb{Z}/5\mathbb{Z}$ is an integral domain by Example 4.4.7.

Example 4.5.2. Consider the principal ideal $(x) = \{r(x)x : r(x) \in \mathbb{R}[x]\}$ of the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Given any real polynomials $p(x)$ and $q(x)$ such that $p(x)q(x) \in (x)$, we must have that $p(x)q(x) = r(x)x$ for some real polynomial $r(x)$. Observe that if neither $p(x)$ nor $q(x)$ were divisible by x, then their product would not be divisible by x: indeed, we have that $p(x)$ is not divisible by x if and only if the constant term of $p(x)$ is nonzero. Consequently, if neither $p(x)$ nor $q(x)$ has constant term zero, then the constant term of $p(x)q(x)$ cannot possibly be zero because the real numbers form a field. On the other hand, the constant term of the polynomial $r(x)x$ is zero, so it follows that either the constant term of $p(x)$ is zero or the constant term of $q(x)$ is zero, i.e., we must have that $p(x) \in (x)$ or $q(x) \in (x)$. We conclude that (x) is a prime ideal of $\mathbb{R}[x]$. By Example 4.3.6, we have that $\mathbb{R}[x]/(x) \cong \mathbb{R}$ is an integral domain.

Our next proposition illustrates that the conclusions of the previous examples hold in general.

Proposition 4.5.3. *Given any commutative unital ring R and any proper ideal P of R, we have that P is a prime ideal of R if and only if the quotient ring R/P is an integral domain.*

Proof. We will assume first that P is a prime ideal of R. We claim that R/P is an integral domain. Considering that R is a commutative unital ring, it follows that R/P is a commutative unital ring, hence it suffices to demonstrate that for any left cosets $r + P$ and $s + P$ of P in R such that $(r + P)(s + P) = 0_R + P$, we have that $r + P = 0_R + P$ or $s + P = 0_R + P$. Coset multiplication is defined such that $0_R + P = (r + P)(s + P) = rs + P$, hence we have that $rs \in P$ by Proposition 3.1.4. By assumption that P is a prime ideal, either $r \in P$ or $s \in P$ so that either $r + P = 0_R + P$ or $s + P = 0_R + P$. Conversely, suppose that R/P is an integral domain. Given any elements $r, s \in R$ such that $rs \in P$, we have that $(r + P)(s + P) = rs + P = 0_R + P$. By hypothesis that R/P is an integral domain, it follows that $r + P = 0_R + P$ or $s + P = 0_R + P$ so that $r \in P$ or $s \in P$. □

Corollary 4.5.4. *There exists a commutative unital ring that admits an ideal that is not prime.*

Proof. Consider the principal ideal $4\mathbb{Z}$ of the commutative unital ring $\mathbb{Z}$ of integers. Observe that the left coset $2+4\mathbb{Z}$ of $4\mathbb{Z}$ in $\mathbb{Z}$ is nonzero and satisfies that $(2+4\mathbb{Z})(2+4\mathbb{Z}) = 4+4\mathbb{Z} = 0+4\mathbb{Z}$, hence $\mathbb{Z}/4\mathbb{Z}$ is not an integral domain. By Proposition 4.5.3, it follows that $4\mathbb{Z}$ is not a prime ideal. □

We say that a proper ideal M of a commutative unital ring R is **maximal** if it has the property that $M \subseteq I$ for some ideal I of R implies that $I = M$ or $I = R$. Put another way, a maximal ideal M is maximal (with respect to inclusion) among the proper ideals of R that contains M. Crucially, this definition implies that if M is a maximal ideal of a commutative unital ring R, then the only ideal of R that properly contains M is the entire ring R. Explicitly, if $M \subsetneq I$, then $I = R$.

Example 4.5.5. Consider the principal ideal $5\mathbb{Z} = \{5k : k \in \mathbb{Z}\}$ of the commutative unital ring $\mathbb{Z}$ of integers. Observe that $5\mathbb{Z} \subseteq n\mathbb{Z}$ for some positive integer n if and only if $5 = 5(1) \in n\mathbb{Z}$ if and only if $5 = nq$ for some integer q if and only if $n = 1$ or $n = 5$. Consequently, the only ideals of $\mathbb{Z}$ containing $5\mathbb{Z}$ are the entire ring $\mathbb{Z}$ and the ideal $5\mathbb{Z}$ itself, hence $5\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$. By the Fourth Isomorphism Theorem for Rngs, the ideals of $\mathbb{Z}/5\mathbb{Z}$ are in one-to-one correspondence with the ideals of $\mathbb{Z}$ containing $5\mathbb{Z}$. Consequently, the only ideals of $\mathbb{Z}/5\mathbb{Z}$ are $\mathbb{Z}/5\mathbb{Z}$ (with pre-image $\mathbb{Z}$) and the zero ideal (with pre-image $5\mathbb{Z}$); this agrees with Example 4.4.7 and Corollary 4.4.19.

Example 4.5.6. Consider the principal ideal (x) of the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. By Proposition 4.2.10, (x) is contained in an ideal I of $\mathbb{R}[x]$ if and only if $x \in I$. We claim that if I is a proper ideal of $\mathbb{R}[x]$, then $I = (x)$. By the Polynomial Division Algorithm, for any polynomial $p(x) \in I$, there exist unique polynomials $q(x)$ and $r(x)$ such that $p(x) = q(x)x + r(x)$ and $r(x)$ is a constant polynomial. Observe that if $r(x)$ were a nonzero constant polynomial, then I would contain a nonzero constant $r(x) = p(x) - q(x)x$ by assumption that I is an ideal that contains $p(x)$ and x. Every nonzero constant polynomial in $\mathbb{R}[x]$ is a nonzero real number, hence we may multiply $r(x)$ by its multiplicative inverse to find that 1 lies in I — a contradiction to our assumption that I is a proper ideal of $\mathbb{R}[x]$. We conclude that every element of I is divisible by x so that $I \subseteq (x)$. Consequently, the only ideals of $\mathbb{R}[x]$ that contain (x) are the entire ring $\mathbb{R}[x]$ and the ideal (x) itself, and (x) is maximal. By Example 4.3.6, $\mathbb{R}[x]/(x) \cong \mathbb{R}$ is a field.

Proposition 4.5.7. *Given any commutative unital ring R and any proper ideal M of R, we have that M is a maximal ideal of R if and only if the quotient ring R/M is a field.*

Proof. We will assume first that M is a maximal ideal of R. We claim that R/M is a field. Considering that R is a commutative unital ring, it follows that R/M is a commutative unital ring, hence it suffices to demonstrate that for any nonzero left coset $x + M$ of M in R, there exists a nonzero left coset $r + M$ of M in R such that $(r + M)(x + M) = 1_R + M$. Coset multiplication is defined such that $(r + M)(x + M) = rx + M$, hence we have that $(r + M)(x + M) = 1_R + M$ if and only if $rx + M = 1_R + M$ if and only if $rx = rx + 0_R = 1_R - m$ for some element $m \in M$ if and only if $1_R = m + rx$ for some elements $m \in M$ and $r \in R \setminus M$. By Exercise 4.8.25, the set $M + Rx = \{m + rx \mid m \in M \text{ and } r \in R\}$ is an ideal of R that properly contains M, hence by the maximality of M, it follows that $R = M + Rx$. Consequently, there exist elements $m \in M$ and $r \in R$ such that $1_R = m + rx$. Crucially, we must have that $r \in R \setminus M$ because M is a proper ideal of R: for if it were true that $r \in M$, then $1_R = m + rx$ would be an element of M.

Conversely, we will assume that R/M is a field. By the Fourth Isomorphism Theorem for Rngs, every ideal of R/M is of the form I/M for some ideal I of R such that $M \subseteq I$. By Corollary 4.4.19, the only ideals of R/M are the zero ideal and R/M itself, hence we have that $I/M = \{0_R + M\}$ or $I/M = R/M$. But this implies that $I = M$ or $I = R$, hence M is maximal, as desired. □

Corollary 4.5.8. *Every maximal ideal of a commutative unital ring is a prime ideal. Conversely, there exists a commutative unital ring with a prime ideal that is not maximal.*

Proof. By Proposition 4.4.9, every field is an integral domain. Consequently, if M is a maximal ideal of a commutative unital ring R, then R/M is a field by Proposition 4.5.7 so that it is an integral domain. We conclude by Proposition 4.5.3 that M is a prime ideal of R. We will need to develop a bit more machinery for the proof of the converse, hence we reserve this task for later. □

Consequently, we have the following hierarchy of ideals of commutative unital rings.

$$\text{maximal ideals} \subsetneq \text{prime ideals} \subsetneq \text{ideals}$$

By the exposition preceding proposition 4.2.20, every pair of ideals I and J of a commutative unital ring R induce a product ideal $IJ = \{i_1 j_1 + \cdots + i_n j_n \mid n \geq 1, i_1, \dots, i_n \in I, j_1, \dots, j_n \in J\}$. Prime numbers have the property that if p is a prime number and a and b are integers such that $p \mid ab$, then it must be the case that $p \mid a$ or $p \mid b$. Our next proposition reasserts that prime ideals behave analogously to prime numbers with respect to an abstraction of this divisibility property.

Proposition 4.5.9. *Given any commutative unital ring R, any prime ideal P of R, and any ideals I and J of R such that $IJ \subseteq P$, we have that $I \subseteq P$ or $J \subseteq P$.*

Proof. By Exercise 1.10.10, we may assume that $J \not\subseteq P$ and subsequently establish that $I \subseteq P$. Given any element $i \in I$, we have that $ij \in P$ for every element $j \in J$ by hypothesis that $IJ \subseteq P$. Considering that $J \not\subseteq P$, there exists an element $j_0 \in J$ such that $j_0 \notin P$. By the primality of P and the fact that $ij_0 \in P$, we must have that $i \in P$. We conclude therefore that $I \subseteq P$, as desired. □

Until now, we have tacitly assumed that every commutative unital ring admits prime and maximal ideals. By Proposition 4.5.8, in order to prove that this is indeed the case, it suffices to prove that every commutative unital ring admits a maximal ideal. We achieve this using Zorn's Lemma. Explicitly, if R is a nonzero commutative unital ring, then the zero ideal $\{0_R\}$ is a proper ideal of R. By Exercise 1.10.19, set inclusion constitutes a partial order on the nonempty set of proper ideals of R, hence if we can demonstrate that every ascending chain of proper ideals in R has an upper bound that is a proper ideal of R, then we will conclude that R admits a maximal ideal.

Proposition 4.5.10. *Given any ascending chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of proper ideals of a commutative unital ring R, we have that $\cup_{n=1}^\infty I_n$ is a proper ideal of R.*

Proof. Considering that 0_R lies in I_1, it follows that 0_R lies in $\cup_{n=1}^\infty I_n$ so that this set is nonempty. By the Three-Step Ideal Test, it suffices to prove that $\cup_{n=1}^\infty I_n$ is closed under subtraction and multiplication by elements of R. Given any elements $r, s \in \cup_{n=1}^\infty I_n$, there exist indices $m \geq \ell$ such that $r \in I_\ell$ and $s \in I_m$. By assumption that $I_\ell \subseteq I_m$, it follows that $r, s \in I_m$ so that $r - s \in I_m$ because it is an ideal of R. We conclude that $r - s \in \cup_{n=1}^\infty I_n$. Even more, for any element $x \in R$, we have that $xr \in I_\ell$ so that $xr \in \cup_{n=1}^\infty I_n$, hence it is an ideal of R. On the contrary, suppose that $\cup_{n=1}^\infty I_n$ is not a proper ideal of R. Consequently, there exists an integer $m \geq 1$ such that $1_R \in I_m$ so that I_m is not a proper ideal of R — contradicting the assumptions of the proposition statement. □

Leveraging the previous proposition, we carry out the strategy outlined in the previous paragraph to illustrate that every commutative unital ring possesses at least one maximal ideal, and moreover, that maximal ideals are actually ubiquitous in commutative unital rings. We note that the ideas contained in the following proofs are quite common in commutative algebra, so read them carefully.

Theorem 4.5.11. *Every nonzero commutative unital ring possesses a maximal ideal.*

Proof. Consider the collection $\mathcal{P}$ of proper ideals of a commutative unital ring R. Observe that $\mathcal{P}$ is partially ordered by set inclusion, and it is nonempty because it contains the zero ideal $\{0_R\}$. Consequently, we seek to employ Zorn's Lemma. Consider any ascending chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of ideals in $\mathcal{P}$. By Proposition 4.5.10, it follows that $\cup_{n=1}^\infty I_n$ is a proper ideal of R; this demonstrates that every chain of elements of $\mathcal{P}$ has an upper bound in $\mathcal{P}$, hence $\mathcal{P}$ admits a maximal element with respect to set inclusion. By definition, this maximal element is a maximal ideal of R. □

Theorem 4.5.12. *Every proper ideal of a nonzero commutative unital ring lies in a maximal ideal.*

Proof. Given any proper ideal I of a commutative unital ring R, consider the collection $\mathcal{P}$ of proper ideals of R that contain I. Certainly, the set $\mathcal{P}$ is nonempty because I is a proper ideal of R that contains I. Even more, every ascending chain $J_1 \subseteq J_2 \subseteq J_3 \subseteq \cdots$ of ideals of $\mathcal{P}$ induces an upper bound $\cup_{n=1}^\infty J_n$ that is a proper ideal of R that contains I by the proof of Proposition 4.5.10. Consequently, by Zorn's Lemma, the set $\mathcal{P}$ admits a maximal element M with respect to set inclusion. We claim that M is a maximal ideal of R. By construction, M is the largest (with respect to set inclusion) proper ideal of R that contains I. Consequently, for any ideal J of R that contains M, we have that J contains I, hence it must be the case that $J = M$ or $J = R$. □

4.6 Polynomial Rings and Polynomial Long Division

Given any rng R, define the collection of **univariate polynomials** in **indeterminate** x over R by

$$R[x] = \{r_n x^n + \cdots + r_1 x + r_0 \mid n \geq 0 \text{ is an integer and } r_0, r_1, \dots, r_n \in R\}.$$

Each rng element r_i is called the **coefficient** of the **monomial** x^i; the element r_0 is the **constant term**; the largest non-negative integer n for which the coefficient r_n of the monomial x^n is nonzero is the **degree** of the polynomial; and the coefficient r_n of the monomial x^n in this case is called the **leading coefficient**. Conventionally, the degree of the **zero polynomial** 0_R is $-\infty$.

Polynomials over arbitrary rngs can be equipped with an addition and multiplication extending that of real polynomials. Explicitly, for any rng R, any integers $n \geq m \geq 0$, and any polynomials $p(x) = r_m x^m + \cdots + r_1 x + r_0$ and $q(x) = s_n x^n + \cdots + s_1 x + s_0$ in $R[x]$, we define the following.

$$p(x) + q(x) = s_n x^n + \cdots + s_{m+1} x^{m+1} + (r_m + s_m) x^m + \cdots + (r_1 + s_1) x + (r_0 + s_0)$$

$$p(x)q(x) = \sum_{j=0}^{m+n} \left(\sum_{i=0}^{j} r_i s_{j-i} \right) x^j = r_m s_n x^{m+n} + \cdots + (r_0 s_1 + r_1 s_0) x + r_0 s_0$$

Each of the sums $r_i + s_i$ for each integer $0 \leq i \leq m$ is an element of R because $(R, +)$ is an abelian group. Likewise, for each integer $0 \leq j \leq m + n$, the product $r_i s_{j-i}$ is an element of R for each integer $0 \leq i \leq j$ because R is closed under multiplication, hence the sum of these products

$$\sum_{i=0}^{j} r_i s_{j-i} = r_0 s_j + r_1 s_{j-1} + \cdots + r_{j-1} s_1 + r_j s_0$$

yields an element of R once again because $(R, +)$ is an abelian group. We conclude therefore that this addition and multiplication both constitute binary operations on $R[x]$. Even more, this addition is associative and commutative because R is an abelian group; the zero polynomial 0_R satisfies the property that $p(x) + 0_R = p(x) = 0_R$ for all polynomials $p(x) \in R[x]$; and the additive inverse of a polynomial $p(x) = r_m x^m + \cdots + r_1 x + r_0$ must be the polynomial of $R[x]$ whose coefficients are the additive inverses of the coefficients of $p(x)$, i.e., $-p(x) = (-r_m)x^m + \cdots + (-r_1)x + (-r_0)$. Combined, these observations all yield that $R[x]$ is an abelian group under polynomial addition.

Proposition 4.6.1. *Given any rng R, the collection $R[x]$ of univariate polynomials in indeterminate x with coefficients in R forms a rng with respect to polynomial addition and polynomial multiplication of which R is a subrng. Even more, if R is a unital ring with multiplicative identity 1_R, then $R[x]$ is a unital ring with multiplicative identity 1_R. Likewise, if R is commutative, then $R[x]$ is commutative.*

Proof. By the exposition preceding the statement of the proposition, it suffices to prove that polynomial multiplication is associative and distributive. Consider any polynomials $p(x) = \sum_{i=0}^{\ell} r_i x^i$, $q(x) = \sum_{i=0}^{m} s_i x^i$, and $r(x) = \sum_{i=0}^{n} t_i x^i$. We demonstrate that $p(x)(q(x)r(x)) = (p(x)q(x))r(x)$.

$$\begin{aligned}
p(x)(q(x)r(x)) &= p(x)\left(\sum_{j=0}^{m+n}\left(\sum_{i=0}^{j} s_i t_{j-i}\right)x^j\right) \\
&= \sum_{k=0}^{\ell+m+n}\left(\sum_{j=0}^{k} r_j\left(\sum_{i=0}^{k-j} s_i t_{k-j-i}\right)\right)x^k \\
&= \sum_{k=0}^{\ell+m+n}\left(\sum_{j=0}^{k}\sum_{i=0}^{k-j} r_j(s_i t_{k-j-i})\right)x^k \\
&= \sum_{k=0}^{\ell+m+n}\left(\sum_{j=0}^{k}\sum_{i=0}^{j-i}(r_i s_{j-i})t_{k-j}\right)x^k \\
&= \sum_{k=0}^{\ell+m+n}\left(\sum_{j=0}^{k}\left(\sum_{i=0}^{j} r_i s_{j-i}\right)t_{k-j}\right)x^k \\
&= \left(\sum_{j=0}^{\ell+m}\left(\sum_{i=0}^{j} r_i s_{j-i}\right)x^j\right)r(x) \\
&= (p(x)q(x))r(x)
\end{aligned}$$

Explicitly, the sums involved in the previous displayed equations are all finite, hence we may combine and reindex as desired; the associativity of R guarantees that the third and fifth equalities holds. Likewise, the distributive property of polynomial multiplication can be established by performing a similar argument as before using the definitions of polynomial addition and polynomial multiplication and appealing to the distributive property of R; we omit the details. Last, we note that R is a subrng of $R[x]$ because polynomial addition and polynomial multiplication are binary operations on R: indeed, these are simply the addition and multiplication already defined on R.

We will assume now that R is a unital ring with multiplicative identity 1_R. By definition of polynomial multiplication, it follows that $p(x)1_R = p(x) = 1_R p(x)$ for all polynomials $p(x) \in R[x]$. Consequently, by the fourth part of Proposition 4.1.8, we conclude that $R[x]$ is a unital ring with multiplicative identity 1_R. Even more, if R is commutative, then $R[x]$ must be commutative because $r_i s_{j-i} = s_{j-i} r_i$ for all integers $0 \leq i \leq j$ and $0 \leq j \leq \ell + m$ so that $p(x)q(x) = q(x)p(x)$. □

Example 4.6.2. We are quite familiar with real polynomials already; however, we can also restrict our attention to polynomials with integer coefficients. By Proposition 4.6.1, the unital subring $\mathbb{Z}$ of $\mathbb{R}$ induces a unital subring $\mathbb{Z}[x]$ of $\mathbb{R}[x]$. Polynomials with integer coefficients behave in some ways quite differently than polynomials with real coefficients. Explicitly, the polynomial $x^2 - 2$ of $\mathbb{Z}[x]$ cannot be factored in $\mathbb{Z}[x]$ because it has no roots in $\mathbb{Z}$. Consequently, $x^2 - 2$ is **irreducible** in $\mathbb{Z}[x]$; however, in $\mathbb{R}[x]$, we know that it factors non-trivially as $(x - \sqrt{2})(x + \sqrt{2})$.

Example 4.6.3. Consider the commutative unital polynomial ring $(\mathbb{Z}/4\mathbb{Z})[x]$. Conventionally, the coefficients of the polynomials in this ring are not written as left cosets; rather, they are simply written as integers with the tacit understanding that addition and multiplication of polynomials occurs modulo 4. Occasionally, it is beneficial to write a polynomial of $(\mathbb{Z}/4\mathbb{Z})[x]$ as $p(x) \pmod 4$ to underscore the fact that the coefficients are taken modulo 4. Explicitly, the polynomial $2x + 3$ of $(\mathbb{Z}/4\mathbb{Z})[x]$ satisfies that $(2x + 3)(2x + 3) = 4x^2 + 12x + 9 \equiv 1 \pmod 4$, hence $2x + 3$ is a unit of $(\mathbb{Z}/4\mathbb{Z})[x]$. Even more, the polynomial $2x + 2$ of $(\mathbb{Z}/4\mathbb{Z})[x]$ satisfies that $(2x + 2)(2x + 2) = 4x^2 + 8x + 4 \equiv 0 \pmod 4$. Consequently, it is possible to find non-constant polynomials in $(\mathbb{Z}/4\mathbb{Z})[x]$ that are units, and the degree of a product of polynomials in $(\mathbb{Z}/4\mathbb{Z})[x]$ is not necessarily the sum of the degrees of the polynomials; this stands in stark contrast to the situation with real polynomials.

Generally, polynomials over arbitrary rngs exhibit very strange and unpredictable behavior, and they can be difficult to understand beyond the details we have provided (cf. Exercises 4.8.64 and 4.8.65). Our next propositions illustrate that polynomial rings over domains are more civilized. Particularly, they do not admit any of the wonky behavior of polynomial rngs over general rngs.

Proposition 4.6.4. *Given any rng R, we have that R is a domain if and only if $R[x]$ is a domain. Even more, if R is a domain, then* $\deg(pq) = \deg(p) + \deg(q)$ *for all polynomials $p(x), q(x) \in R[x]$.*

Proof. If $R[x]$ is a domain, then R is a domain: indeed, R is a subring of $R[x]$ by Proposition 4.6.1, and any subring of a domain is a domain. Conversely, we will assume that R is a domain. Consider any nonzero polynomials $p(x) = r_m x^m + \cdots + r_1 x + r_0$ and $q(x) = s_n x^n + \cdots + s_1 x + s_0$ of $R[x]$ with respective degrees m and n. Observe that the leading coefficient of $p(x)q(x)$ is by definition $r_m s_n$. By hypothesis that R is a domain and r_m and s_n are nonzero elements of R, it follows that $r_m s_n$ is nonzero, hence $p(x)q(x)$ is a nonzero polynomial such that $\deg(pq) = m + n = \deg(p) + \deg(q)$. Consequently, every nonzero element of $R[x]$ is regular, hence $R[x]$ is a domain. □

Proposition 4.6.5. *Given any domain R, we have that u is a unit of $R[x]$ if and only if u is a unit of R. Put another way, the units of $R[x]$ and the units of R coincide.*

Proof. Certainly, if u is a unit of R, then the constant polynomial u is a unit of $R[x]$. Conversely, we will assume that $u = r_n x^n + \cdots + r_1 x + r_0$ is a unit of $R[x]$. Consequently, there exist elements $s_0, s_1, \dots, s_m \in R$ not all of which are zero such that $u^{-1} = s_m x^m + \cdots + s_1 x + s_0$ and $uu^{-1} = 1_R$. By hypothesis that R is a domain, it follows by Proposition 4.6.4 that $R[x]$ is a domain, and we must have that $0 = \deg(1_R) = \deg(uu^{-1}) = \deg(u) + \deg(u^{-1})$. Considering that u and u^{-1} are nonzero polynomials, their degrees must be non-negative; they sum to 0 if and only if u and u^{-1} are constant. We conclude that $u = r_0$ and $u^{-1} = s_0$ with $r_0 s_0 = 1_R$, i.e., u is a unit of R. □

Even in the case of polynomials with coefficients lying in a domain, there exist subtle obstructions. Explicitly, it is not possible to obtain the integer polynomial $2x + 3$ as a polynomial of the form $2x+3 = q(x)(3x-4)+r(x)$ for some integer polynomials $q(x)$ and $r(x)$ such that $r(x)$ is either the zero polynomial or a constant polynomial: indeed, the leading coefficient of $q(x)(3x-4)+r(x)$ must be divisible by 3, so it cannot be $2x + 3$. Consequently, polynomials with coefficients lying in an arbitrary domain do not necessarily admit some analogy of the Division Algorithm.

Conversely, if we restrict our attention to **monic** polynomials (i.e., polynomials with leading coefficient 1_R) with coefficients in an arbitrary rng R, then it is possible to uniquely divide any polynomial of $R[x]$ by a monic polynomial of $R[x]$ (possibly with remainder). Explicitly, for the integer polynomial $2x+3$ and the monic polynomial $x+1$, we may write $2x+3 = 2(x+1)+1$ such that the polynomials 2 and 1 are uniquely determined. We reserve the general case of this fact as Exercise 4.8.70; however, we will prove this for polynomials with coefficients in a domain.

Theorem 4.6.6 (Polynomial Division Algorithm)**.** *Given any domain R, consider any monic polynomial $p(x)$ and any polynomial $f(x)$ in $R[x]$. There exist unique polynomials $q(x)$ and $r(x)$ in $R[x]$ such that $f(x) = p(x)q(x) + r(x)$ and either $r(x) = 0_R$ or $0 \leq \deg(r) \leq \deg(p) - 1$.*

Proof. By Proposition 4.6.4 and our assumption that R is a domain, for all polynomials $q(x) \in R[x]$, we have that $\deg(pq) = \deg(p) + \deg(q)$. Consequently, the unique expression of 0_R in the desired form of the theorem statement is $0_R = 0_R + 0_R = p(x)0_R + 0_R$. We may assume therefore that $f(x)$ is nonzero so that $\deg(f) = n$ is a non-negative integer. Observe that if $\deg(p) - 1 \geq n \geq 0$, then $f(x) = 0_R + f(x) = p(x)0_R + f(x)$ is the unique expression of $f(x)$ in the desired form of the theorem statement. Consequently, we may assume that $\deg(f) = n \geq m = \deg(p)$, in which case we may proceed by the Principle of Complete Induction on n. Consider the leading coefficient r_n of $f(x)$. By assumption that $p(x)$ is a monic polynomial of degree $m \leq n$, the polynomial $r_n x^{n-m} p(x)$ has degree n with leading coefficient r_n so that $f(x) - r_n x^{n-m} p(x)$ is a polynomial of strictly lesser degree than $f(x)$. By our inductive hypothesis, there exist polynomials $q(x)$ and $r(x)$ in $R[x]$ such that $f(x) - r_n x^{n-m} p(x) = p(x)q(x) + r(x)$ and either $r(x) = 0_R$ or $0 \leq \deg(r) \leq \deg(p) - 1$. By rearranging this expression, we find that $f(x) = p(x)(r_n x^{n-m} + q(x)) + r(x)$, hence the existence of the desired polynomial of the theorem statement is established. We prove that they are unique.

Consider any polynomials $q_1(x), q_2(x), r_1(x)$, and $r_2(x)$ such that $f(x) = p(x)q_1(x) + r_1(x)$ and $f(x) = p(x)q_2(x) + r_2(x)$ and either both $r_1(x)$ and $r_2(x)$ are the zero polynomial or one of the inequalities $0 \leq \deg(r_1) \leq \deg(p) - 1$ or $0 \leq \deg(r_2) \leq \deg(p) - 1$ holds. Crucially, observe that either way, we must have that $\deg(r_2 - r_1) \leq \deg(p) - 1$. By rearranging the two aforementioned identities of $f(x)$, we obtain an identity $p(x)(q_1(x) - q_2(x)) = r_2(x) - r_1(x)$. On the contrary, if it were the case that $q_1(x)$ and $q_2(x)$ were not equal, then their difference $q_1(x) - q_2(x)$ would be a nonzero polynomial so that $\deg(q_1 - q_2) \geq 0$. By the first paragraph of the proof, we would have that $\deg(p) - 1 \geq \deg(r_2 - r_1) = \deg(p(q_1 - q_2)) = \deg(p) + \deg(q_1 - q_2) \geq \deg(p)$ — a contradiction. We conclude therefore that $q_1(x) = q_2(x)$, from which it follows that $r_1(x) = r_2(x)$. □

Essentially, the Polynomial Division Algorithm allows us to perform the usual polynomial long division from high school algebra in the more general setting of polynomial rngs over arbitrary rngs. Given any polynomial identity of the form $f(x) = p(x)q(x) + r(x)$, we refer to the polynomial $f(x)$ as the **dividend**; $p(x)$ is the **divisor**; $q(x)$ is the **quotient**; and $r(x)$ is the **remainder**.

Example 4.6.7. Let us perform polynomial long division to find the quotient $q(x)$ and the remainder $r(x)$ of the polynomial $f(x) = 3x^3 + 2x^2 - x + 1$ divided by the monic polynomial $p(x) = x - 1$.

$$\begin{array}{r|rrrr}
\multicolumn{1}{r}{} & & 3x^2 & +5x & +4 \\ \cline{2-5}
x-1 & 3x^3 & +2x^2 & -x & +1 \\
 & -3x^3 & +3x^2 & & \\ \cline{2-3}
\multicolumn{1}{r}{} & & 5x^2 & -x & \\
\multicolumn{1}{r}{} & & -5x^2 & +5x & \\ \cline{3-4}
\multicolumn{1}{r}{} & & & 4x & +1 \\
\multicolumn{1}{r}{} & & & -4x & +4 \\ \cline{4-5}
\multicolumn{1}{r}{} & & & & 5
\end{array}$$

Explicitly, we begin by eliminating the leading term of $3x^3$ by multiplying $x-1$ by $3x^2$ and subtracting the resulting polynomial $3x^3 - 3x^2$ from the dividend; the resulting polynomial is $5x^2 - x + 1$, hence we multiply $x-1$ by $5x$ and subtract the resulting polynomial $5x^2 - 5x$ from $5x^2 - x + 1$ to obtain $4x+1$; and last, we multiply $x-1$ by 4 and subtract the resulting polynomial $4x-4$ from $4x+1$ to obtain a remainder of 5. Ultimately, we have that $3x^3 + 2x^2 - x + 1 = (x-1)(3x^2 + 5x + 4) + 5$.

Example 4.6.8. Let us perform polynomial long division to find the quotient $q(x)$ and the remainder $r(x)$ of the polynomial $f(x) = x^3 + x + 1$ divided by the monic polynomial $p(x) = x + 2$.

$$\begin{array}{r|rrrr}
\multicolumn{1}{r}{} & & x^2 & -2x & +5 \\ \cline{2-5}
x+2 & x^3 & & +x & +1 \\
 & -x^3 & -2x^2 & & \\ \cline{2-3}
\multicolumn{1}{r}{} & & -2x^2 & +x & \\
\multicolumn{1}{r}{} & & 2x^2 & +4x & \\ \cline{3-4}
\multicolumn{1}{r}{} & & & 5x & +1 \\
\multicolumn{1}{r}{} & & & -5x & -10 \\ \cline{4-5}
\multicolumn{1}{r}{} & & & & -9
\end{array}$$

Consequently, we have that $x^3 + x + 1 = (x+2)(x^2 - 2x + 5) - 9$. Observe that if we view the coefficients of these polynomials as elements of $\mathbb{Z}/3\mathbb{Z}$, then $x^3 + x + 1$ is divisible by $x + 2$ modulo 3 because we have that $x^3 + x + 1 = (x+2)(x^2 - 2x + 5) - 9 \equiv (x+2)(x^2 + x + 2) \pmod 3$.

Often, polynomials are viewed throughout mathematics as functions for which the indeterminate x is viewed as a variable that can be substituted with values from the rng of coefficients; however, we have not and will continue not to adopt this viewpoint. Explicitly, in our case, polynomials offer a construction that allow us to understand the properties of the rng of coefficients. On the other hand, for any polynomial rng $R[x]$ with coefficients in a rng R and for each element $\alpha \in R$, we are afforded a rng homomorphism $\varphi_\alpha : R[x] \to R$ defined by $\varphi_\alpha(p(x)) = p(\alpha)$ that is called the **evaluation homomorphism** at α: indeed, for any polynomials $p(x) = r_m x^m + \cdots + r_1 x + r_0$ and $q(x) = s_n x^n + \cdots + s_1 x + s_0$ such that $n \geq m \geq 0$, the following properties hold.

$$p(\alpha) + q(\alpha) = s_n \alpha^n + \cdots + s_{m+1}\alpha^{m+1} + (r_m + s_m)\alpha^m + \cdots + (r_1 + s_1)\alpha + (r_0 + s_0)$$

$$p(\alpha)q(\alpha) = (r_m \alpha^m + \cdots + r_1 \alpha + r_0)(s_n \alpha^n + \cdots + s_1 \alpha + s_0) = \sum_{j=0}^{m+n} \left(\sum_{i=0}^{j} r_i s_{i-j} \right) \alpha^j$$

Consequently, the first equation above demonstrates that $\varphi_\alpha(p(x) + q(x)) = \varphi_\alpha(p(x)) + \varphi_\alpha(q(x))$, and the second equation above shows that $\varphi_\alpha(p(x)q(x)) = \varphi_\alpha(p(x))\varphi)\alpha(q(x))$. We have already demonstrated in Example 4.3.6 that evaluation homomorphisms can be used to determine explicit isomorphisms of quotients of polynomial rngs, and we will return to this notion again later.

Consider any element α of any rng R. We will say that α is a **root** of a polynomial $p(x)$ in $R[x]$ if and only if $p(x)$ lies in the kernel of the evaluation homomorphism at α if and only if $p(\alpha) = 0_R$. Our next two propositions relate the roots of a polynomial to its linear factors.

Theorem 4.6.9 (Remainder Theorem)**.** *Given any rng R, any polynomial $p(x)$ in $R[x]$, and any element $\alpha \in R$, the remainder of $p(x)$ modulo the monic linear polynomial $x - \alpha$ is $p(\alpha)$.*

Proof. Considering that $x - \alpha$ is a monic polynomial, by Exercise 4.8.70, there exist unique polynomials $q(x)$ and $r(x)$ in $R[x]$ such that $p(x) = (x - \alpha)q(x) + r(x)$ and $r(x)$ is a constant polynomial. By applying the evaluation homomorphism at α, we find that

$$p(\alpha) = (\alpha - \alpha)q(\alpha) + r(\alpha) = 0_R q(\alpha) + r(\alpha) = r(\alpha).$$

Considering that $r(\alpha)$ is a constant polynomial that takes the value of $p(\alpha)$ under the evaluation homomorphism at α, we conclude that $r(x) = p(\alpha)$, as desired. □

Theorem 4.6.10 (Factor Theorem)**.** *Given any rng R, any polynomial $p(x)$ in $R[x]$, and any element $\alpha \in R$, we have that $x - \alpha$ is a factor of $p(x)$ if and only if α is a root of $p(x)$.*

Proof. By the Remainder Theorem, if α is a root of $p(x)$, then the remainder of $p(x)$ modulo $x - \alpha$ is $p(\alpha) = 0_R$, hence there exists a unique polynomial $q(x)$ such that $p(x) = (x - \alpha)q(x)$ and $x - \alpha$ is a factor of $p(x)$. Conversely, we will assume that $x - \alpha$ is a factor of $p(x)$. By definition, there exists a unique polynomial $q(x)$ in $R[x]$ with $p(x) = (x - \alpha)q(x)$. By applying the evaluation homomorphism at α, we find that $p(\alpha) = (\alpha - \alpha)q(\alpha) = 0_R q(\alpha) = 0_R$. We conclude that α is a root of $p(x)$. □

Example 4.6.11. By Example 4.6.7 and the Remainder Theorem, we note that the polynomial $f(x) = 3x^3 + 2x^2 - x + 1$ satisfies that $f(1) = 5$ because the remainder of $f(x)$ modulo $x - 1$ is 5.

Example 4.6.12. By Example 4.6.8 and the Remainder Theorem, we note that the polynomial $f(x) = x^3 + x + 1$ satisfies that $f(-2) = -9$ because the remainder of $f(x)$ modulo $x + 2$ is -9.

Combined, the Remainder Theorem and the Factor Theorem provide powerful tools that significantly reduce the amount of work required to compute the roots of a polynomial in a large number of cases. Even more, the Rational Roots Theorem narrows down the search for roots of polynomials with integer coefficients to a finite number of possibilities! We will explore more properties about the existence of roots of polynomials with integer coefficients in the next section.

4.7 Polynomial Irreducibility

We turn our attention throughout this section to the question of factoring polynomials. By the Factor Theorem, the linear factors of a polynomial are in one-to-one correspondence with the roots of the polynomial (up to multiplicity), hence the factorization of a polynomial is intimately connected with the possible roots of the polynomial; however, as we have seen, the degree of a product of polynomials over an arbitrary rng need not be the sum of the degrees of the polynomial unless the rng is in fact a domain. Consequently, we will henceforth assume throughout this section that R is an integral domain (i.e., a commutative unital ring in which any nonzero element is cancellable) so that the degree of a polynomial is the sum of the degrees of its proper factors. Given any element $r \in R$, we say that an element $d \in R$ **divides** R if and only if there exists an element $s \in R$ such that $r = ds$. By analogy to the greatest common divisor of integers, we say that for any pair of elements $r, s \in R$, a **greatest common divisor** of r and s is any element $d \in R$ such that

(a.) $d \mid r$ and $d \mid s$, i.e., d divides both r and s and

(b.) if d' is a nonzero element of R such that $d' \mid r$ and $d' \mid s$, then $d' \mid d$.

Later, we will prove that if the greatest common divisor of an element of an arbitrary integral domain exists, then it is unique up to multiplication by a unit of R. We will not concern ourselves with either the existence or the uniqueness of the greatest common divisor at the moment; rather, we will assume that R is an integral domain in which any pair of elements r and s admits a greatest common divisor $\gcd(r, s)$. Per usual, the greatest common divisor $\gcd(r_0, r_1, \dots, r_n)$ of any collection of elements $r_0, r_1, \dots, r_n \in R$ can be computed recursively via $\gcd(r_0, r_1, \dots, r_n) = \gcd(\gcd(r_0, r_1), r_2, \dots, r_n)$. Often, it will behoove us to restrict our attention to the integral domains $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ in which case our assumptions hold. By Proposition 4.6.4, we note that under our present hypotheses, the polynomial ring $R[x]$ is an integral domain. Given any polynomial $p(x) = r_n x^n + \cdots + r_1 x + r_0$ in $R[x]$, the **content** of $p(x)$ is the element $\operatorname{content}(p) = \gcd(r_0, r_1, \dots, r_n)$ of R. We say that a polynomial $p(x)$ is **primitive** if $\operatorname{content}(p)$ is a unit of R; however, this terminology will eventually become ambiguous, so it is important to bear in mind the context. Our next observation is natural.

Proposition 4.7.1. *Let R be an integral domain in which any pair of elements admits a greatest common divisor. Let x be an indeterminate over R. Every polynomial $p(x)$ in $R[x]$ can be factored as the product $p(x) = \operatorname{content}(p)q(x)$ for some primitive polynomial $q(x)$ in $R[x]$.*

Proof. We will assume that $p(x) = r_n x^n + \cdots + r_1 x + r_0$. By definition of greatest common divisor, each of the coefficients of $p(x)$ is divisible by $\operatorname{content}(p) = \gcd(r_0, r_1, \dots, r_n)$ — namely, there exist elements $s_0, s_1, \dots, s_n$ such that $r_i = \operatorname{content}(p) s_i$ for each integer $0 \leq n$. Consider the polynomial $q(x) = s_n x^n + \cdots + s_1 x + s_0$. Observe that $p(x) = \operatorname{content}(p)q(x)$, hence it suffices to prove that

$q(x)$ is primitive, i.e., $\text{content}(q) = \gcd(s_0, s_1, \dots, s_n)$ is a unit of R. Eventually, we will prove a generalization of Bézout's Identity that implies the existence of elements $t_0, t_1, \dots, t_n \in R$ such that $\text{content}(p) = \gcd(r_0, r_1, \dots, r_n) = r_0 t_0 + r_1 t_1 + \cdots + r_n t_n$. Considering that $r_i = \text{content}(p) s_i$, we find that $\text{content}(p)(s_0 t_0 + s_1 t_1 + \cdots + s_n t_n) = \text{content}(p)$. Cancellation holds in R by assumption that it is a domain, hence we conclude that $s_0 t_0 + s_1 t_1 + \cdots + s_n t_n = 1_R$. Last, any greatest common divisor of $s_0, s_1, \dots, s_n$ divides $s_0 t_0 + s_1 t_1 + \cdots + s_n t_n = 1_R$, hence it must be a unit. Put another way, we conclude that $\text{content}(q) = \gcd(s_0, s_1, \dots, s_n)$ is a unit so that $q(x)$ is primitive. □

Even more, we will refer to a polynomial $p(x) \in R[x]$ as **reducible** in $R[x]$ (or over R) if either

(a.) $\text{content}(p)$ is not a unit of R or

(b.) there exist non-constant polynomials $q(x), r(x) \in R[x]$ such that $p(x) = q(x)r(x)$.

Conversely, we say that $p(x)$ is **irreducible** over R if and only if it is non-constant and not reducible if and only if $\text{content}(p)$ is a unit of R and $p(x)$ does not factor as a product of two non-constant polynomials. Consequently, by definition, a primitive polynomial in $R[x]$ is irreducible if and only if it does not factor as a product of non-constant polynomials in $R[x]$. Let us look at some examples.

Example 4.7.2. Every monic polynomial in $R[x]$ is primitive because its leading coefficient is 1_R.

Example 4.7.3. Consider the polynomial $p(x) = 3x^2 + 7x + 1$ in $\mathbb{Z}[x]$. By definition, we have that $\text{content}(p) = \gcd(3, 7, 1) = 1$, hence $p(x)$ is primitive in $\mathbb{Z}[x]$. By the Quadratic Formula and the Factor Theorem, $p(x)$ is irreducible over $\mathbb{Z}$ because its roots are non-rational real numbers.

Example 4.7.4. Consider the polynomial $p(x) = 4x^2 + 6x + 2$ in $\mathbb{Z}[x]$. By definition, we have that $\text{content}(p) = \gcd(4, 6, 2) = 2$, hence $p(x)$ is not primitive in $\mathbb{Z}[x]$; in fact, the primitive polynomial $q(x) = 2x^2 + 3x + 1$ satisfies that $p(x) = 2q(x) = 2(2x^2 + 3x + 1)$. Even still, observe that $q(x)$ is not irreducible in $\mathbb{Z}[x]$ because it holds that $q(x) = 2x^2 + 3x + 1 = (2x + 1)(x + 1)$.

Example 4.7.5. Consider the polynomial $p(x) = x^2 + 2x + 3$ in $\mathbb{Q}[x]$. By definition, we have that $\text{content}(p) = \gcd(1, 2, 3) = 1$, hence $p(x)$ is primitive in $\mathbb{Q}[x]$. Every nonzero polynomial in $\mathbb{Q}[x]$ is primitive in $\mathbb{Q}[x]$ because $\mathbb{Q}$ is a field, hence any nonzero element of $\mathbb{Q}$ is a unit. By the Quadratic Formula and the Factor Theorem, $p(x)$ is irreducible over $\mathbb{Q}$ because its roots are the complex numbers $-1 - i\sqrt{2}$ and $-1 + i\sqrt{2}$. By the same rationale, $p(x)$ is primitive and irreducible in $\mathbb{R}[x]$.

Example 4.7.6. Let us find all irreducible quadratic polynomials in $(\mathbb{Z}/2\mathbb{Z})[x]$. Considering that the only elements of $\mathbb{Z}/2\mathbb{Z}$ are 0 and 1 (modulo 2), it follows by the Fundamental Counting Principle that there are only $2 \cdot 2 = 4$ quadratic polynomials in $(\mathbb{Z}/2\mathbb{Z})[x]$. Explicitly, the only quadratic polynomials in $(\mathbb{Z}/2\mathbb{Z})[x]$ are x^2, $x^2 + 1$, $x^2 + x$, and $x^2 + x + 1$. Clearly, the polynomials x^2 and $x^2 + x$ are reducible because they each admit a linear factor of x. On the other hand, it follows by the Factor Theorem that $x^2 + 1$ is reducible: indeed, we have that $1^2 + 1 = 2 \equiv 0 \pmod{2}$, hence 1 is a root of $x^2 + 1$ (modulo 2) and $x - 1 \equiv x + 1 \pmod{2}$ is a factor of $x^2 + 1$. One can verify by polynomial long division that $x^2 + 1 \equiv (x + 1)(x + 1) \pmod{2}$, but it is also possible to see this by noticing that $x^2 + 1 \equiv x^2 + 2x + 1 = (x + 1)^2 \pmod{2}$. We refer to the phenomenon $x^2 + 1 \equiv (x + 1)^2 \pmod{2}$ as the **Freshman's Dream**. Generally, a factorization of this form holds for any pair of elements over any ring of prime characteristic (cf. Exercise 4.8.74 for more on the Freshman's Dream). Last, we note that $x^2 + x + 1$ is irreducible in $(\mathbb{Z}/2\mathbb{Z})[x]$ because $0^2 + 0 + 1 = 1$ and $1^2 + 1 + 1 = 3 \equiv 1 \pmod{2}$ are both nonzero in $\mathbb{Z}/2\mathbb{Z}$, hence $x^2 + x + 1$ has no linear factors.

Quadratic and cubic polynomials are generally dealt with by appealing to the Factor Theorem: indeed, the Factor Theorem immediately implies that a primitive quadratic or cubic polynomial is irreducible in $R[x]$ if and only if it does not admit a root in R. Explicitly, a quadratic polynomial that is the product of non-constant polynomials must be the product of two linear polynomials, and a cubic polynomial that is the product of non-constant polynomials must be the product of a linear polynomial and a quadratic polynomial. Often, however, it requires a bit more machinery to deduce the irreducibility of polynomials of larger degree. Explicitly, in order to deduce whether a polynomial of degree exceeding three is reducible, one must check that the polynomial admits no linear factors or quadratic factors or cubic factors and so on. Continuing in this manner eventually exhausts all possibilities; however, this process can be tedious, and it is not clear how to discern the irreducibility of polynomials of large degree. Consequently, we set out to develop some criteria that will simplify this process. We will restrict our present attention to polynomials with integer, rational, and real coefficients; however, we note that these tools can be extended to certain "nice" integral domains. Our first results relate factorizations of primitive polynomials in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

Lemma 4.7.7. *Given any factorization $p(x) = q(x)r(x)$ of polynomials $p(x), q(x)$, and $r(x)$ in $\mathbb{Z}[x]$, if n is a prime number that divides every coefficient of $p(x)$, then n divides every coefficient of $q(x)$ or n divides every coefficient of $r(x)$. Put another way, if $n \mid q(x)r(x)$, then $n \mid q(x)$ or $n \mid r(x)$.*

Proof. We will assume that $q(x) = a_\ell x^\ell + \cdots + a_1 x + a_0$ and $r(x) = b_m x^m + \cdots + b_1 x + b_0$ are polynomials in $\mathbb{Z}[x]$. On the contrary, suppose that the prime number n that divides $p(x) = q(x)r(x)$ does not divide $q(x)$ or $r(x)$. Consequently, there exists a least integer i such that n does not divide the ith coefficient a_i of $q(x)$, and there exists a least integer j such that n does not divide the jth coefficient b_j of $r(x)$ by the Well-Ordering Principle. Considering that n is a prime number, n cannot divide $a_i b_j$; however, by assumption, n divides the $(i+j)$th coefficient of $q(x)r(x)$

$$\sum_{k=0}^{i+j} a_k b_{i+j-k} = a_0 b_{i+j} + \cdots + a_{i-1} b_{j+1} + a_i b_j + a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0$$

and each of the integers $a_0, \dots, a_{i-1}$ and $b_0, \dots, b_{j-1}$ is divisible by n by construction of a_i and b_j, hence n divides $a_i b_j$ — a contradiction. We conclude that n divides $q(x)$ or n divides $r(x)$. □

Proposition 4.7.8. *Consider a primitive polynomial $q(x)$ in $\mathbb{Z}[x]$. Given any polynomial $p(x)$ in $\mathbb{Z}[x]$, if $q(x)$ divides $p(x)$ as polynomials in $\mathbb{Q}[x]$, then $q(x)$ divides $p(x)$ as polynomials in $\mathbb{Z}[x]$.*

Proof. We will assume that $q(x)$ divides $p(x)$ as polynomials in $\mathbb{Q}[x]$. By definition, there exists a polynomial $r(x)$ in $\mathbb{Q}[x]$ such that $p(x) = q(x)r(x)$. Explicitly, we may write

$$r(x) = \frac{a_i}{b_i} x^i + \cdots + \frac{a_1}{b_1} x + \frac{a_0}{b_0}$$

for some integers $a_i, \dots, a_1, a_0$ and some nonzero integers $b_i, \dots, b_1, b_0$. Consider the nonzero integer $b = b_i \cdots b_1 b_0$. Clearing the denominators of the terms of $r(x)$ by multiplying $r(x)$ by b yields a polynomial $br(x)$ in $\mathbb{Z}[x]$ such that $bp(x) = q(x)(br(x))$, i.e., $q(x)$ divides $bp(x)$ as polynomials in $\mathbb{Z}[x]$. We conclude that $C = \{c \in \mathbb{Z} \mid q(x) \text{ divides } cp(x) \text{ as polynomials in } \mathbb{Z}[x]\}$ is a nonempty set. Consider the least positive integer m such that $q(x)$ divides $mp(x)$ as polynomials in $\mathbb{Z}[x]$. By the

Well-Ordering Principle, such an integer exists. We claim that $m = 1$ so that $q(x)$ divides $p(x)$ as polynomials in $\mathbb{Z}[x]$. On the contrary, we will assume that $m \geq 2$. By the Fundamental Theorem of Arithmetic, there exists a prime number n that divides m, i.e., $\frac{m}{n}$ is a positive integer smaller than m. Given any polynomial $s(x)$ such that $mp(x) = q(x)s(x)$ as polynomials in $\mathbb{Z}[x]$, n must divide $q(x)$ or n must divide $s(x)$ by Lemma 4.7.7. By assumption that $q(x)$ is primitive, it cannot be divisible by a prime number n, hence we conclude that $s(x)$ is divisible by n. But this implies that

$$q(x)\frac{s(x)}{n} = \frac{q(x)s(x)}{n} = \frac{mp(x)}{n} = \frac{m}{n}p(x)$$

as polynomials in $\mathbb{Z}[x]$ — contradicting the minimal property defining m. □

Put another way, Proposition 4.7.8 states that if a polynomial with integer coefficients has a primitive factor when viewed as a polynomial in $\mathbb{Q}[x]$, then that primitive factor remains when we consider the factorization as polynomials in $\mathbb{Z}[x]$. Our next two theorems generalize this to any factorizations of integer polynomials in $\mathbb{Q}[x]$; they are similarly named after Carl Friedrich Gauss.

Theorem 4.7.9 (Gauss's Lemma)**.** *Given any polynomial $p(x)$ in $\mathbb{Z}[x]$, if there exist polynomials $Q(x)$ and $R(x)$ in $\mathbb{Q}[x]$ with $p(x) = Q(x)R(x)$, then there exist polynomials $q(x)$ and $r(x)$ in $\mathbb{Z}[x]$ with $p(x) = q(x)r(x)$. Put another way, if an integer polynomial admits a factorization by polynomials with rational coefficients, then it admits a factorization by polynomials with integer coefficients.*

Proof. We will assume that $p(x) = Q(x)R(x)$ for some polynomials $Q(x)$ and $R(x)$ in $\mathbb{Q}[x]$. By the proof of Proposition 4.7.8, we may clear the denominators of the coefficients of $Q(x)$ to obtain a polynomial $Q_0(x) = \alpha Q(x)$ of $\mathbb{Z}[x]$. By Proposition 4.7.1, we may factor the polynomial $Q_0(x)$ of $\mathbb{Z}[x]$ as $Q_0(x) = \operatorname{content}(Q_0)q(x)$ for some primitive polynomial $q(x)$ in $\mathbb{Z}[x]$. We have therefore established that $q(x)$ divides $p(x)$ as polynomials in $\mathbb{Q}[x]$, as we have the polynomial identity

$$p(x) = Q(x)R(x) = \frac{1}{\alpha}Q_0(x)H(x) = \frac{\operatorname{content}(Q_0)}{\alpha}q(x)H(x) = q(x)\left(\frac{\operatorname{content}(Q_0)}{\alpha}R(x)\right).$$

By Proposition 4.7.8, there exists a polynomial $r(x)$ in $\mathbb{Z}[x]$ such that $p(x) = q(x)r(x)$ in $\mathbb{Z}[x]$. □

Theorem 4.7.10 (Gauss's Little Lemma)**.** *Consider the univariate polynomial ring $\mathbb{Z}[x]$.*

1.) *We have that $p(x)$ and $q(x)$ are primitive in $\mathbb{Z}[x]$ if and only if $p(x)q(x)$ is primitive in $\mathbb{Z}[x]$.*

2.) *Given any non-constant primitive polynomial $p(x)$ in $\mathbb{Z}[x]$, we have that $p(x)$ is irreducible in $\mathbb{Z}[x]$ if and only if $p(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. (1.) We will assume first that $p(x)$ and $q(x)$ are primitive polynomials in $\mathbb{Z}[x]$. By Proposition 4.7.1, there exists a primitive polynomial $r(x)$ in $\mathbb{Z}[x]$ such that $p(x)q(x) = \operatorname{content}(pq)r(x)$. Observe that as polynomials in $\mathbb{Q}[x]$, we have the following factorization of $r(x)$ by $p(x)$.

$$r(x) = p(x)\left(\frac{1}{\operatorname{content}(pq)}q(x)\right)$$

By assumption that $p(x)$ is a primitive polynomial and $r(x)$ is an integer polynomial, Proposition 4.7.8 yields a polynomial $s(x)$ in $\mathbb{Z}[x]$ such that $r(x) = p(x)s(x)$. By multiplying both sides of this equation by $q(x)$ and using the fact that $p(x)q(x) = \text{content}(pq)r(x)$, it follows that $q(x)r(x) = \text{content}(pq)r(x)s(x)$. Considering that $\mathbb{Z}[x]$ is a domain by Proposition 4.6.4, we may cancel the nonzero polynomial $r(x)$ on both sides of this equation to find that $q(x) = \text{content}(pq)s(x)$. Comparing the content of each polynomial, we find that $\text{content}(q) = \text{content}(pq)\,\text{content}(s)$. By assumption that $q(x)$ is primitive, we have that $\text{content}(q) = \pm 1$ so that $\text{content}(pq) = \pm 1$ and $p(x)q(x)$ is primitive. Conversely, suppose that either $p(x)$ or $q(x)$ is not primitive in $\mathbb{Z}[x]$. Consequently, there exists a prime number n such that n divides $p(x)$ or n divides $q(x)$; either way, n divides $p(x)q(x)$ so that $p(x)q(x)$ is not primitive because its content is divisible by n.

(2.) By Gauss's Lemma, if $p(x)$ admits a factorization in $\mathbb{Q}[x]$, then $p(x)$ admits a factorization in $\mathbb{Z}[x]$, hence if $p(x)$ is irreducible in $\mathbb{Z}[x]$, then it is irreducible in $\mathbb{Q}[x]$. Conversely, if $p(x)$ is irreducible in $\mathbb{Q}[x]$, then it is irreducible in $\mathbb{Z}[x]$ because any $\mathbb{Z}[x]$-factorization is a $\mathbb{Q}[x]$-factorization. □

Corollary 4.7.11. *Every polynomial with integer coefficients that admits a rational number as a root also admits an integer as a root; this integer divides the constant term of the polynomial.*

Proof. We may assume that $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ is a polynomial with integer coefficients such that a_0 is nonzero: indeed, the leading coefficient of a polynomial does not affect the roots of the polynomial, and every polynomial that is divisible by x admits 0 as a root. By assumption, there exists a rational number α such that $p(\alpha) = 0$. Consequently, by the Factor Theorem, it follows that $p(x) = (x - \alpha)Q(x)$ for some polynomial $Q(x)$ in $\mathbb{Q}[x]$. By Gauss's Lemma and its proof, there exists a linear polynomial $x - a$ and a monic polynomial $q(x)$ in $\mathbb{Z}[x]$ of degree $n-1$ such that $p(x) = (x-a)q(x)$. Observe that the constant term of $p(x)$ is $a_0 = p(0) = -q(0)a$. □

Example 4.7.12. We claim that the quartic polynomial $p(x) = x^4 + x + 1$ is irreducible in $\mathbb{Q}[x]$. Consequently, it suffices to prove that $p(x)$ has no linear factors and no quadratic factors: indeed, if $p(x)$ has no linear factors, then it has no cubic factors, either. By the Factor Theorem, the linear factors of $p(x)$ correspond to the roots of the polynomial $p(x)$. Corollary 4.7.11 guarantees that the existence of a rational root induces an integer root dividing the constant term of $p(x)$, so it suffices to check that 1 and -1 are not roots of $p(x)$; this is clear because $p(1) = 3$ and $p(-1) = 1$. By Gauss's Lemma, we may restrict our attention to monic quadratic factors of $p(x)$ in $\mathbb{Z}[x]$, hence we may assume that $p(x) = (x^2 + ax + b)(x^2 + cx + d)$. Expand the right-hand side and compare the coefficients of $x^4 + 0x^3 + 0x^2 + 1x + 1 = x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd.$

$$\begin{aligned} a + c &= 0 & \qquad ad + bc &= 1 \\ ac + b + d &= 0 & bd &= 1 \end{aligned}$$

Considering that $bd = 1$ and b and d are integers, it follows that $b = d = \pm 1$. Either way, the identities $ad + bc = 1$ and $a + c = 0$ yield that $0 = b \cdot 0 = b(a + c) = ab + bc = ad + bc = 1$ — a contradiction. We conclude that $p(x)$ admits no quadratic factors, so it is irreducible in $\mathbb{Q}[x]$.

Example 4.7.13. We claim that the quintic polynomial $p(x) = x^5 - 4x^2 + 2$ is irreducible in $\mathbb{Q}[x]$. Like before, we eliminate the possibility of linear or quartic factors by checking the roots of $p(x)$; then, we dismiss the possibility of quadratic or cubic factors by inspection. Corollary 4.7.11 reduces our search for rational roots of $p(x)$ to integer roots that divide 2; therefore, the only possibilities for a linear factor are the linear polynomials corresponding to the integers ± 1 and ± 2. One can verify that $p(\pm 1)$ and $p(\pm 2)$ are all nonzero, hence $p(x)$ does not admit any linear or quartic factors.

Once again, by Gauss's Lemma, we may restrict our search for quadratic factors to monic quadratic factors in $\mathbb{Z}[x]$. Consider the case that $p(x) = (x^2 + ax + b)(x^3 + cx^2 + dx + e)$. Expanding these polynomials and comparing the coefficients yields the following integer equations.

$$\begin{array}{ccc} a + c = 0 & ad + bc + e = -4 & be = 2 \\ ac + b + d = 0 & ae + bd = 0 & \end{array}$$

Consequently, we must consider the following four cases arising from the integer equation $be = 2$.

(1.) Observe that if $b = 2$ and $e = 1$, then $ae + bd = 0$ implies that $a + 2d = 0$ and $a = -2d$; then, $c = -a$ and $ac + b + d = 0$ yield that $-4d^2 + d = -2$ or $(4d - 1)d = 2$ — a contradiction.

(2.) Observe that if $b = -2$ and $e = -1$, then we arrive at the same contradiction as above.

(3.) Observe that if $b = 1$ and $e = 2$, then $ae + bd = 0$ implies that $2a + d = 0$ and $d = -2a$; then, $c = -a$ and $ac + b + d = 0$ yield that $-a^2 - 2a = -1$ or $(a + 2)a = 1$ — a contradiction.

(4.) Observe that if $b = -1$ and $e = -2$, then we arrive at the same contradiction as above.

We conclude therefore that $p(x)$ admits no quadratic factors, so it is irreducible in $\mathbb{Q}[x]$.

Generally, the method of proof outlined in the previous two examples is in theory possible to carry out for polynomials of arbitrarily large degree; however, as we have seen, this process has its limitations, as it requires us to solve non-linear systems of integer equations. Carrying this out by hand on a case-by-case basis can be extremely tedious and ad hoc — even in the case of quartic and quintic polynomials — when either the constant term of the integer polynomial has a large number of prime factors (with multiplicity) or when the polynomial has many nonzero terms.

We turn our attention to a criterion for the irreducibility of a polynomial whose constant term shares a (multiplicity one) prime factor with each non-leading coefficient. Given any prime number p, we say that a polynomial $q(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ in $\mathbb{Z}[x]$ is p-**Eisenstein** if

(1.) p divides each of the coefficients $a_0, a_1, \ldots, a_{n-1}$ and

(2.) p does not divide the leading coefficient a_n and

(3.) p^2 does not divide the constant term a_0.

Theorem 4.7.14 (Eisenstein's Criterion). *If $q(x)$ is a polynomial in $\mathbb{Z}[x]$ that is p-Eisenstein for some prime number p, then $q(x)$ cannot be written as the product of two non-constant polynomials. Consequently, if $q(x)$ is primitive, then $q(x)$ is irreducible in $\mathbb{Z}[x]$, hence $q(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. On the contrary, we will assume that $q(x) = r(x)s(x)$ for some non-constant polynomials $r(x)$ and $s(x)$ in $\mathbb{Z}[x]$. We may write $r(x) = b_i x^i + \cdots + b_1 x + b_0$ and $s(x) = c_j x^j + \cdots + c_1 x + c_0$ for some integers $b_0, b_1, \ldots, b_i, c_0, c_1, \ldots, c_j$. Consequently, we have that $a_0 = b_0 c_0$, $a_1 = b_1 c_0 + b_0 c_1$, etc. By hypothesis that p divides a_0 and p^2 does not divide a_0, one of b_0 and c_0 must be divisible by p but not both. We may assume that p divides b_0 so that p does not divide c_0, in which case the identity $a_1 - b_0 c_1 = b_1 c_0$ yields that p divides b_1. Continuing in this manner, we find that $b_0, b_1, \ldots, b_i$ are divisible by p so that $a_n = b_i c_j$ is divisible by p — a contradiction. We conclude that one of $r(x)$ or

$s(x)$ is constant, hence $q(x)$ cannot be written as a product of two non-constant polynomials. If $q(x)$ is primitive, then the constant factor of $q(x)$ must be a unit of $\mathbb{Z}$, hence $q(x)$ must be irreducible over $\mathbb{Z}[x]$. By Gauss's Little Lemma, we conclude that $q(x)$ is irreducible in $\mathbb{Q}[x]$. □

Example 4.7.15. Observe that $p(x) = x^3 - 2$ is 2-Eisenstein and hence irreducible in $\mathbb{Q}[x]$.

Example 4.7.16. Observe that $p(x) = x^3 - 9x + 3$ is 3-Eisenstein and hence irreducible in $\mathbb{Q}[x]$.

One other technique for demonstrating the irreducibility of a polynomial with integer coefficients is the following so-called **reduction modulo** p for a prime number p.

Proposition 4.7.17 (Reduction Modulo p)**.** *Consider any polynomial $q(x)$ in $\mathbb{Z}[x]$. If there exists a prime number p that does not divide the leading coefficient of $q(x)$ such that the image of $q(x)$ modulo p cannot be written as the product of two non-constant polynomials in $(\mathbb{Z}/p\mathbb{Z})[x]$, then $q(x)$ cannot be written as the product of two non-constant polynomials in $\mathbb{Z}[x]$. Consequently, if $q(x)$ is a primitive polynomial in $\mathbb{Z}[x]$, then $q(x)$ is irreducible in $\mathbb{Z}[x]$, hence $q(x)$ is irreducible in $\mathbb{Q}[x]$.*

Proof. Consider any prime number p that does not divide the leading coefficient of $q(x)$ such that the image of $q(x)$ modulo p cannot be written as the product of two non-constant polynomials in $(\mathbb{Z}/p\mathbb{Z})[x]$. On the contrary, we will assume that $q(x) = r(x)s(x)$ for some non-constant polynomials $r(x)$ and $s(x)$ in $\mathbb{Z}[x]$. By assumption that p does not divide the leading coefficient of $q(x)$, it follows that neither the leading coefficient of $r(x)$ nor the leading coefficient of $s(x)$ is divisible by p: indeed, because $\mathbb{Z}[x]$ is a domain by Proposition 4.6.4, the leading coefficient of $q(x)$ is the product of the leading coefficient of $r(x)$ and the leading coefficient of $s(x)$. Consequently, the degree of $r(x) \pmod p$ is the degree of $r(x)$, and the degree of $s(x) \pmod p$ is the degree of $s(x)$. Particularly, the polynomials $r(x) \pmod p$ and $s(x) \pmod p$ are non-constant, and their product is $q(x) \pmod p$ — a contradiction. We conclude that such a factorization of $q(x)$ is not possible. □

Example 4.7.18. Consider the polynomial $q(x) = 7x^3 + 6x^2 + 4x + 4$. We note that $\gcd(7, 6, 4) = 1$, hence $q(x)$ is a primitive polynomial in $\mathbb{Z}[x]$. Employing the technique of reduction modulo $p = 5$, by Proposition 4.7.17, in order to prove that $q(x)$ is irreducible in $\mathbb{Q}[x]$, it suffices to note that $q(x)$ admits no roots modulo 5: indeed, $q(0) = 4$, $q(1) = 21 \equiv 1 \pmod 5$, $q(2) = 92 \equiv 2 \pmod 5$, $q(3) = 259 \equiv 4 \pmod 5$, and $q(4) = 564 \equiv 4 \pmod 5$ are nonzero modulo 5.

Last, we turn our attention to the irreducibility of polynomials in $\mathbb{R}[x]$. Our first result toward this end uses calculus to vastly reduce the types of possible irreducible polynomials in $\mathbb{R}[x]$.

Proposition 4.7.19. *Every real polynomial of odd degree admits a real root.*

Proof. Consider any real polynomial $p(x)$ of odd degree. We may view $p(x)$ as a continuous real function via the unital ring homomorphism $\mathbb{R}[x] \to F(\mathbb{R}, \mathbb{R})$ that sends $p(x)$ to the polynomial function $p(x)$. Considering that $\lim_{x \to -\infty} p(x)$ and $\lim_{x \to \infty} p(x)$ are infinite of opposite sign, by the Intermediate Value Theorem, there exists a real number α such that $p(\alpha) = 0$, as desired. □

Consequently, every real polynomial of odd degree can be written as a product of a real polynomial of even degree and a real linear polynomial, so it is natural to seek to understand real polynomials of even degree. Quadratic polynomials that admit one real root must admit two real roots by the Factor Theorem, hence it suffices to note by the Quadratic Formula that a quadratic polynomial $ax^2 + bx + c$ is reducible if and only if its **discriminant** $b^2 - 4ac$ is non-negative. Put another way, the following holds for real quadratic polynomials. (We assume that $a > 0$.)

Proposition 4.7.20. *Every real polynomial $ax^2 + bx + c$ is irreducible if and only if $b^2 - 4ac < 0$.*

We conclude this section by demonstrating that every irreducible real polynomial is either a real linear polynomial or a real quadratic polynomial whose discriminant is negative. Even more, we show that every real polynomial is the product of real linear and irreducible quadratic polynomials.

Theorem 4.7.21. *Consider the commutative unital ring $\mathbb{R}[x]$ of univariate real polynomials in x.*

1.) *If $p(x)$ is an irreducible real polynomial, then either $p(x)$ is a real linear polynomial or $p(x)$ is a real quadratic polynomial whose discriminant is negative.*

2.) *Every real polynomial is a product of real linear and irreducible quadratic polynomials.*

Proof. Every nonzero real polynomial is primitive because every nonzero element of $\mathbb{R}$ is a unit, hence in order to deduce the irreducibility of a real polynomial, it suffices to prove that the real polynomial does not factor as a product of two non-constant polynomials. Constant polynomials are never irreducible by definition, hence we may restrict our attention to polynomials of positive degree. Linear polynomials are always irreducible because a linear polynomial cannot be written as a product of two non-constant polynomials. Continuing our role call of real polynomials, by Proposition 4.7.20, real quadratic polynomials with negative discriminant are irreducible. Conversely, by Proposition 4.7.19, real polynomials of odd degree exceeding one are not irreducible. We are therefore left to deal only with real polynomials of even degree exceeding two. By the Fundamental Theorem of Algebra, every real polynomial $p(x)$ of degree $2k$ admits exactly $2k$ complex roots. Consider a complex root $z = a + bi$ of $p(x)$. By Example 4.1.19, complex conjugation distributes across complex addition and complex multiplication, hence $p(a - bi)$ is the complex conjugate of $p(a + bi)$; the latter is zero by assumption, hence $\overline{z} = a - bi$ is a root of $p(x)$. By the Factor Theorem, we conclude that

$$p(x) = (x - z)(x - \overline{z})q(x) = (x^2 - \overline{z}x - zx + z\overline{z})q(x) = (x^2 - 2ax + a^2 + b^2)q(x)$$

for some polynomial $q(x)$ in $\mathbb{C}[x]$ of degree $2k - 2$. We claim that $q(x)$ has real coefficients; if this holds, then by the Principle of Ordinary Induction applied to the real polynomial $q(x)$ of even degree, we may conclude the desired result that $p(x)$ is a product of real quadratic polynomials.

Considering $p(x)$ and $x^2 - 2ax + a^2 + b^2$ as real polynomials, the Polynomial Division Algorithm yields unique real polynomials $q_0(x)$ and $r(x)$ such that $p(x) = (x^2 - 2ax + a^2 + b^2)q_0(x) + r(x)$ and either $r(x)$ is the zero polynomial or $0 \leq \deg(r) \leq 1$. Considering $p(x)$ and $x^2 - 2ax + a^2 + b^2$ as complex polynomials, the uniqueness of the Polynomial Division Algorithm applied to the identity $p(x) = (x^2 - 2ax + a^2 + b^2)q(x)$ implies that $r(x) = 0$ and $q(x) = q_0(x)$ is a real polynomial. □

4.8 Chapter 4 Exercises

4.8.1 Rings and Ring Homomorphisms

Exercise 4.8.1. Prove the Ring Exponent Laws.

Exercise 4.8.2. We say that a rng R is **Boolean** if it holds that $r^2 = r$ for all elements $r \in R$.

(a.) Prove that $r = -r$ holds for every element r of a Boolean rng R.

(b.) Prove that every Boolean rng is commutative.

(**Hint:** Observe that if R is Boolean, then $(r+s)^2 = r+s$ for all elements $r, s \in R$.)

Exercise 4.8.3. Consider a unital ring R with multiplicative identity 1_R. Given an element $r \in R$ such that there exists an element $s \in R$ for which $rs = 1_R$, must it be true that $sr = 1_R$? Explain.

Exercise 4.8.4. Consider a unital ring R with multiplicative identity 1_R. Prove that for any element $r \in R$ such that there exists an element $s \in R$ for which $rs = 1_R = sr$, the element s is unique to r.

Exercise 4.8.5. Given a unital ring R, consider the set of units of R

$$U(R) = \{u \in R \mid uv = 1_R \text{ for some element } v \in R\}.$$

(a.) Prove that if u is a unit of R, then u^{-1} is a unit of R.

(b.) Prove that if u and v are units of R, then uv is a unit of R.

(c.) Prove that $U(R)$ forms a group with respect to multiplication; it is called the **multiplicative group of units** of R. Conclude that if R is commutative, then $U(R)$ is abelian.

(d.) Prove that if u is a unit of R, then $-u$ is a unit of R.

(e.) Prove that $U(R)$ is not closed under addition. Conclude that $U(R)$ is not a rng.

Exercise 4.8.6. Given a unital ring R with multiplicative identity 1_R, prove that for any elements $u, v \in R$ such that uv is a unit of R, we must have that u and v are units of R.

Exercise 4.8.7. Given any unital rings R and S, prove that $U(R \times S) = U(R) \times U(S)$.

Exercise 4.8.8. Determine if each of the following functions are unital ring homomorphisms.

(a.) $\varphi : \mathbb{Z} \to \mathbb{Z}$ defined by $\varphi(x) = -x$

(b.) $\varphi : \mathbb{Q} \to \mathbb{Q}$ defined by $\varphi(x) = \dfrac{2x}{x+1}$

(c.) $\varphi : \mathbb{R} \to \mathbb{C}$ defined by $\varphi(x) = \sqrt{x}$

(d.) $\varphi : \mathbb{R} \to \mathbb{R}^{2\times 2}$ defined by $\varphi(x) = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$

Exercise 4.8.9. Determine if the following pairs of commutative rngs are isomorphic.

(a.) $\dfrac{\mathbb{Z}}{6\mathbb{Z}}$ and $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{3\mathbb{Z}}$

(b.) $\dfrac{\mathbb{Z}}{16\mathbb{Z}}$ and $\dfrac{\mathbb{Z}}{4\mathbb{Z}} \times \dfrac{\mathbb{Z}}{4\mathbb{Z}}$

(c.) $n\mathbb{Z}$ and $\mathbb{Z}$ for any integer $n \geq 2$

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

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

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

(**Hint:** Consider the Fundamental Theorem of Finite Abelian Groups for the rings in parts (a.) and (b.). Consider the possible square roots of elements of the rings in parts (e.) and (f.).)

Exercise 4.8.10. Prove that $m\mathbb{Z}$ and $n\mathbb{Z}$ are not isomorphic as rngs for any integers $m > n \geq 2$.

(**Hint:** Consider the possible rng homomorphisms $\varphi : m\mathbb{Z} \to n\mathbb{Z}$ as in Example 4.1.16.)

Exercise 4.8.11. Consider the set $\operatorname{End}(R) = \{\varphi : R \to R \mid \varphi \text{ is a rng homomorphism}\}$ of rng endomorphisms of a rng R. Prove that $\operatorname{End}(R)$ is a non-commutative unital ring under composition.

Exercise 4.8.12. Consider any unital ring R with multiplicative identity 1_R.

(a.) Prove that for any unit u of R, the function $\chi_u : R \to R$ defined by $\chi_u(r) = uru^{-1}$ is a unital ring automorphism. By analogy to group theory, we refer to χ_u as an **inner automorphism** of R. We denote by $\operatorname{Inn}(R) = \{\chi_u : R \to R \mid u \in U(R)\}$ the set of inner automorphisms of R.

(b.) Prove that $\operatorname{Inn}(R)$ forms a non-abelian group under composition.

(c.) Prove that $\operatorname{Inn}(R)$ is not an additive group. Conclude that $\operatorname{Inn}(R)$ is not a rng.

(d.) Prove that the function $\psi : U(R) \to \operatorname{Inn}(R)$ defined by $\psi(u) = \chi_u$ is a group homomorphism.

(e.) Compute the kernel of the group homomorphism $\psi : U(R) \to \operatorname{Inn}(R)$.

Exercise 4.8.13. Compute the characteristic of each of the following commutative unital rings.

(a.) $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{2\mathbb{Z}}$ (b.) $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{3\mathbb{Z}}$ (c.) $\dfrac{\mathbb{Z}}{4\mathbb{Z}} \times \dfrac{\mathbb{Z}}{6\mathbb{Z}}$

(d.) $\mathbb{Q}$ (e.) $\mathbb{R}$ (f.) $\mathbb{C}$

Exercise 4.8.14. Conjecture a formula for the characteristic of $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ for any pair of positive integers m and n; then, prove that your formula holds.

Exercise 4.8.15. Prove that if R is a finite unital ring, then the characteristic of R is positive.

Exercise 4.8.16. Consider a rng R such that $r^3 = r$ for all elements $r \in R$.

(a.) Prove that $(r+s)^3 = r^3 + rsr + sr^2 + s^2r + r^2s + rs^2 + srs + s^3$ for all elements $r, s \in R$. Conclude that $rsr + sr^2 + s^2r + r^2s + rs^2 + srs = 0_R$ for all elements $r, s \in R$.

(b.) Conclude from the previous step that $(r+r)^3 = 8r^3$ for all elements $r \in R$.

(c.) Conclude from the previous step that $6r = 0_R$ for all elements $r \in R$.

(d.) Prove that $(r-s)^3 = r^3 - rsr - sr^2 + s^2r - r^2s + rs^2 + srs - s^3$ for all elements $r, s \in R$. Conclude that $-rsr - sr^2 + s^2r - r^2s + rs^2 + srs = 2s$ for all elements $r, s \in R$.

(e.) Conclude from the previous steps that $2(s^2r + rs^2 + srs) = 2s$ for all elements $r, s \in R$.

(f.) Conclude from the previous steps that $2r = 0_R$ for all elements $r \in R$.

(g.) Conclude from the previous step and Exercise 4.8.2 that R is commutative.

Conclude that commutativity of a rng is a stronger condition than commutativity of a group.

4.8.2 Ideals and Quotient Rings

Exercise 4.8.17. Prove that the following is a subrng of the non-commutative unital ring $\mathbb{R}^{2\times 2}$.

$$S = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \,\middle|\, a, b \in \mathbb{R} \right\}$$

Prove that S admits an element A such that $AB = B = BA$ for all elements $B \in S$. Explain why this does not violate the conclusion of the Subrng Test regarding unital rings.

Exercise 4.8.18. Prove that $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a commutative unital subring of $\mathbb{R}$.

Exercise 4.8.19. Consider any rng R with any pair of subrngs S and T.

(a.) Prove that $S \cap T$ is a subrng of R.

(b.) Provide an example of a non-commutative rng R and a subrng S of R such that S is commutative. Conclude that non-commutativity is not inherited under intersection of rngs.

(c.) Provide an example of a unital ring R and a subrng S of R such that S does not possess a multiplicative identity. Conclude that unity is not inherited under intersection of rngs.

(d.) Prove that $S \cup T$ is a subrng of R if and only if $R = S$ or $R = T$.

Exercise 4.8.20. Let R be a rng. Prove that $Z(R) = \{x \in R \mid rx = xr \text{ for all elements } r \in R\}$ is a subrng of R called the **center** of R. Conclude that if R is unital, then $Z(R)$ is unital.

Exercise 4.8.21. Consider the commutative unital ring $\mathbb{R}[x]$ of univariate real polynomials.

(a.) Prove that $C = \{p(x) \in \mathbb{R}[x] : p(x) \text{ is constant}\}$ is a commutative unital subring of $\mathbb{R}[x]$.

(b.) Prove that $C = \{p(x) \in \mathbb{R}[x] : p(x) \text{ is constant}\}$ is not an ideal of $\mathbb{R}[x]$.

(c.) Prove that $I = \{p(x) \in \mathbb{R}[x] : p(0) = 2\alpha \text{ for some real number } \alpha\}$ is an ideal of $\mathbb{R}[x]$.

(d.) Prove that $J = \{p(x) \in \mathbb{R}[x] : p(0) = 2\}$ is not an ideal of $\mathbb{R}[x]$.

Exercise 4.8.22. Consider any rng R with left ideals I and J.

(a.) Prove that $I + J = \{i + j \mid i \in I \text{ and } j \in J\}$ is a left ideal of R.

(b.) Prove that $I \cap J = \{x \in R \mid x \in I \text{ and } x \in J\}$ is a left ideal of R.

(c.) Prove that $I \cup J = \{x \in R \mid x \in I \text{ or } x \in J\}$ is not an ideal if $I \setminus J$ and $J \setminus I$ are nonempty.

(**Hint:** Consider an element $i \in I \setminus J$ and an element $j \in J \setminus I$. On the contrary, if $I \cup J$ were an ideal, then it would be closed under subtraction. Conclude that either $j \in I$ or $i \in J$.)

Consider the case that I is a left ideal and J is a right ideal of R.

(d.) Prove that $IJ = \{i_1 j_1 + \cdots + i_n j_n \mid n \geq 1, i_1, \dots i_n \in I, j_1, \dots, j_n \in J\}$ is a two-sided ideal.

(e.) Conclude that for any integer $n \geq 1$, the set I^n consisting of all finite sums of n-fold products $i_1 \cdots i_n$ of elements of I is a left ideal of R called the nth **power** of I.

Exercise 4.8.23. Complete the following exercise to prove that for any two-sided ideals I and J of a rng R, it is not in general true that $I * J = \{ij \mid i \in I \text{ and } j \in J\}$ is an ideal of R.

(a.) Prove that for $R = \mathbb{Z}[x]$, the ideals $I = (2, x)$ and $J = (3, x)$ satisfy that the monomial x can be written as $f(x)g(x) + h(x)k(x)$ for some polynomials $f(x), h(x) \in I$ and $g(x), k(x) \in J$.

(b.) Prove that x cannot be written as $p(x)q(x)$ for any polynomials $p(x) \in I$ and $q(x) \in J$.

(c.) Conclude from the previous two steps that $I * J$ is not closed under addition.

Exercise 4.8.24. Prove that if $R \supseteq S$ are rngs, then $I \cap S$ is an ideal of S for any ideal I of R.

Exercise 4.8.25. Consider a commutative unital ring R. Prove that if M is a proper ideal of R and x is an element of $R \setminus M$, then the set $M + Rx = \{m + rx \mid m \in M \text{ and } r \in R\}$ is an ideal of R such that $M + Rx$ properly contains M, i.e., we have that $M + Rx \supsetneq M$.

Exercise 4.8.26. Consider a commutative unital ring R. Prove that if I is an ideal of R, then the set $\sqrt{I} = \{r \in R \mid r^n \in I \text{ for some integer } n \geq 1\}$ is an ideal of R called the **radical** of I.

(**Hint:** Consider the Binomial Theorem as it applies to the sums of elements of I.)

Exercise 4.8.27. Consider a commutative unital ring R. Prove that if I and J are any ideals of R, then the set $(I : J) = \{r \in R \mid rJ \subseteq I\}$ is an ideal of R called the **ideal quotient** of I by J.

Exercise 4.8.28. Complete the following steps to prove that every ideal of $\mathbb{Z}$ is principal.

(a.) Prove that if I is a nonzero ideal of $\mathbb{Z}$, then I admits a smallest positive element a.

(b.) Conclude from the previous step that I contains the principal ideal $a\mathbb{Z}$, i.e., $I \supseteq a\mathbb{Z}$.

(c.) Conversely, use the Division Algorithm to prove that $I \subseteq a\mathbb{Z}$. Conclude that I is principal.

We refer to a rng R for which all (one-sided) ideals are principal as a **principal ideal rng**.

Exercise 4.8.29. Consider the commutative unital ring $\mathbb{Z} \times \mathbb{Z}$.

(a.) Prove that the diagonal $\Delta_{\mathbb{Z}} = \{(n, n) \mid n \in \mathbb{Z}\}$ of $\mathbb{Z}$ is a commutative unital subring of $\mathbb{Z} \times \mathbb{Z}$.

(b.) Prove that $\Delta_{\mathbb{Z}}$ is not an ideal of $\mathbb{Z} \times \mathbb{Z}$.

(c.) Prove that $\mathbb{Z} \times \{0\} = \{(n, 0) \mid n \in \mathbb{Z}\}$ is an ideal of $\mathbb{Z} \times \mathbb{Z}$.

(d.) Prove that $\mathbb{Z}$ and $\mathbb{Z} \times \{0\}$ are isomorphic as commutative unital rings.

Exercise 4.8.30. Prove that if I is a two-sided ideal of a rng R and J is a two-sided ideal of a rng S, then $I \times J$ is a two-sided ideal of the direct product $R \times S$.

Exercise 4.8.31. Complete the following steps to prove that if R is a unital ring and S is any rng, then a two-sided ideal of $R \times S$ has the form $I \times J$ for some two-sided ideals I of R and J of S.

(a.) Given any ideal K of $R \times S$, consider the sets $I = \{r \in R \mid (r, s) \in K \text{ for some element } s \in S\}$ and $J = \{s \in S \mid (r, s) \in K \text{ for some element } r \in R\}$. Prove that $K \subseteq I \times J$.

(b.) Prove that I is a two-sided ideal of R.

(c.) Prove that J is a two-sided ideal of S.

(d.) By definition of I and J, for every element $(r, s) \in I \times J$, there exist elements $x \in R$ and $y \in S$ such that $(r, y), (x, s) \in K$. Prove that (r, ys) and (x, ys) are elements of K.

(e.) Conclude that $(r - x, 0_S)$ is an element of K so that (r, s) is an element of S and $I \times J \subseteq K$.

(f.) Conclude by the previous steps and Exercise 4.8.30 that every two-sided ideal of the direct product of unital rings has the form $I \times J$ for some two-sided ideals I of R and J of S.

Exercise 4.8.32. Determine all ideals of the following direct products of commutative unital rings.

(a.) $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{3\mathbb{Z}}$

(b.) $\dfrac{\mathbb{Z}}{6\mathbb{Z}} \times \dfrac{\mathbb{Z}}{15\mathbb{Z}}$

(c.) $m\mathbb{Z} \times n\mathbb{Z}$ for any integers $m \geq n \geq 0$

(d.) $\mathbb{Z} \times \mathbb{Z}$

(e.) $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$

(f.) $\mathbb{Q} \times \mathbb{Q}$

(**Hint:** Use Exercise 4.8.31. Consider the Fourth Isomorphism Theorem for Rngs for parts (a.) and (b.). Consider Exercise 4.8.28 for parts (c.), (d.), and (e.). Use Corollary 4.4.19 for part (f.).)

Exercise 4.8.33. Consider the non-commutative unital ring $\mathbb{R}^{n \times n}$ consisting of real $n \times n$ matrices for some positive integer $n \geq 2$. Prove that the set $I \subseteq \mathbb{R}^{n \times n}$ of all real $n \times n$ matrices whose first row consists entirely of zeros is a right ideal of $\mathbb{R}^{n \times n}$ that is not a left ideal of $\mathbb{R}^{n \times n}$.

Exercise 4.8.34. Consider the non-commutative unital ring $\mathbb{R}^{n \times n}$ consisting of real $n \times n$ matrices for some positive integer $n \geq 2$. Complete the following steps to prove that there are no non-trivial two sided ideals of $\mathbb{R}^{n \times n}$, i.e., the only nonzero ideal of $\mathbb{R}^{n \times n}$ is the entire ring $\mathbb{R}^{n \times n}$ itself.

(a.) If I is a nonzero ideal of $\mathbb{R}^{n \times n}$, then there exists a nonzero real $n \times n$ matrix $A \in I$. Prove that for any nonzero component a_{ij} of A, the matrix consisting of zeros in every component other than the (i, j)th component and whose (i, j)th component is a_{ij} lies in I.

(b.) Conclude from the previous step that the matrix E_{ij} consisting of zeros in every component other than the (i, j)th component and whose (i, j)th component is 1 lies in I.

(c.) Prove that the matrices E_{ij} consisting of zeros in every component other than the (i, j)th component and whose (i, j)th component is 1 lie in I for all integers $1 \leq i \leq n$ and $1 \leq j \leq n$.

(**Hint:** Once we have one of them, can we take products to find all of them?)

(d.) Conclude from the previous step that every real $n \times n$ matrix lies in I so that $I = \mathbb{R}^{n \times n}$.

Exercise 4.8.35. Determine if the following pairs of commutative unital rings are isomorphic.

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

(b.) $\mathbb{C}$ and $\mathbb{R}^{2 \times 2}$

(c.) $\mathbb{R}^{n \times n}$ and $\mathbb{R}^{n^2}$

(d.) $\mathbb{C}^n$ and $\mathbb{R}^{2n}$

4.8.3 The Ring Isomorphism Theorems

Exercise 4.8.36. Consider the commutative unital rings $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/12\mathbb{Z}$.

(a.) Prove that the function $\varphi : \mathbb{Z}/12\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z}$ defined by $\varphi(n + 12\mathbb{Z}) = n + 4\mathbb{Z}$ is a well-defined surjective unital ring homomorphism.

(b.) Compute the kernel of φ; then, use the First Isomorphism Theorem for Rngs to express $\mathbb{Z}/4\mathbb{Z}$ as a proper quotient of $\mathbb{Z}/12\mathbb{Z}$ by the ideal $\ker \varphi$ of $\mathbb{Z}/12\mathbb{Z}$.

Exercise 4.8.37. Consider the commutative unital rings $\mathbb{Z}$ and $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$.

(a.) Prove that the function $\varphi : \mathbb{Z} \to (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ defined by $\varphi(n) = (n + 2\mathbb{Z}, n + 3\mathbb{Z})$ is a unital ring homomorphism.

(b.) Prove that φ is surjective.

(**Hint:** Explain why every element of $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ can be written as $n(1 + 2\mathbb{Z}, 1 + 3\mathbb{Z})$ for some integer $1 \leq n \leq 6$.)

(c.) Compute the kernel of φ; then, employ the First Isomorphism Theorem for Rngs to express $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ as a proper quotient of $\mathbb{Z}$ by the ideal $\ker \varphi$ of $\mathbb{Z}$.

Exercise 4.8.38. Consider any commutative unital rings R and S.

(a.) Prove that $\varphi : R \times S \to R$ defined by $\varphi(r, s) = r$ is a surjective ring unital ring homomorphism.

(b.) Compute the kernel of φ; then use the First Isomorphism Theorem for Rngs to express R as a proper quotient of $R \times S$ by the ideal $\ker \varphi$ of $R \times S$.

Exercise 4.8.39. Complete the following proof of the Second Isomorphism Theorem for Rngs.

(a.) Prove that $S + I = \{s + i \mid s \in S \text{ and } i \in I\}$ is a subrng of R.

(b.) Prove that I is a two-sided ideal of $S + I$. Conclude that $(S + I)/I$ is a rng.

(c.) Conclude by Exercise 4.8.24 that $I \cap S$ is a two-sided ideal of S.

(d.) Prove that the function $\varphi : S \to (S + I)/I$ defined by $\varphi(s) = s + I$ is a well-defined surjective rng homomorphism such that $\ker \varphi = I \cap S$.

(e.) Conclude by the First Isomorphism Theorem for Rngs that $S/(I \cap S) \cong (S + I)/I$.

Exercise 4.8.40. Complete the following proof of the Third Isomorphism Theorem for Rngs.

(a.) Prove that J is a two-sided ideal of the subrng I of R.

(b.) Prove that that I/J is a two-sided ideal of R/J.

(c.) Prove that the function $\varphi : R/J \to R/I$ defined by $\varphi(r + J) = r + I$ is a well-defined surjective rng homomorphism such that $\ker \varphi = I/J$.

(d.) Conclude by the First Isomorphism Theorem for Rngs that $(R/J)/(I/J) \cong R/I$.

4.8.4 Integral Domains and Fields

Exercise 4.8.41. We say that an element r of a rng R is **idempotent** if it holds that $r^2 = r$.

(a.) Prove that if R is a domain, then the only idempotent elements of R are 0_R and 1_R.

(b.) Exhibit a commutative unital ring R with a nonzero idempotent element.

Exercise 4.8.42. We say that an element r of a rng R is **nilpotent** if there exists an integer $n \geq 1$ such that $r^n = 0_R$. We refer to $\operatorname{ind}(r) = \min\{k \geq 1 \mid r^k = 0_R\}$ as the **index of nilpotency** of r.

(a.) Prove that if R is a domain, then the only nilpotent element of R is 0_R.

(b.) Exhibit a commutative unital ring R with a nonzero nilpotent element.

Exercise 4.8.43. We say that a subset S of a commutative unital ring R with multiplicative identity 1_R is **multiplicatively closed** if $1_R \in S$ and $st \in S$ for any elements $s, t \in S$. Prove that $S = \{r \in R \mid r \text{ is not a zero divisor}\}$ is a multiplicatively closed subset of R.

Exercise 4.8.44. Prove or disprove that if R and S are domains, then $R \times S$ is a domain.

Exercise 4.8.45. Prove that $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a field that contains $\mathbb{Q}$.

Exercise 4.8.46. Prove that any domain R whose only ideals are $\{0_R\}$ and R is a skew field.

Exercise 4.8.47. Prove that if R is a nonzero finite commutative rng with no zero divisors, then R admits a multiplicative identity element. Conclude that R must be a field.

(**Hint:** Every injective rng homomorphism from R to itself must be surjective by Exercise 1.10.5.)

Exercise 4.8.48. Let R be an integral domain that contains a field k. Prove that if R is a finite-dimensional vector space over k, then R must be a field.

(**Hint:** Prove that for every nonzero element $x \in R$, the function $\varphi_x : R \to R$ defined by $\varphi_x(r) = xr$ is a k-linear transformation. Use the Rank-Nullity Theorem to prove that φ is surjective.)

Exercise 4.8.49. Complete the following steps to prove that every element of a finite unital ring must be either a zero divisor or a unit of the ring.

(a.) Given any nonzero element x of a finite unital ring R, prove that the function $\varphi_x : R \to R$ defined by $\varphi_x(r) = xr$ is injective if and only if x is not a left zero divisor of R.

(b.) Given any nonzero element x of a finite unital ring R, prove that the function $\psi_x : R \to R$ defined by $\psi_x(r) = rx$ is injective if and only if x is not a right zero divisor of R.

(c.) Prove that a nonzero element x of a finite unital ring R is a left zero divisor of R if and only if it is a right zero divisor of R. Conclude that x is either a zero divisor of R or not.

(d.) Conclude that if a nonzero element x of a finite unital ring R is not a zero divisor of R, then there exist nonzero elements $y, z \in R$ such that $xy = 1_R$ and $zx = 1_R$. Prove that $y = z$; then, conclude that if x is not a zero divisor of R, then x must be a unit of R.

Exercise 4.8.50. Prove that there are no nonzero unital ring homomorphisms $\varphi : \mathbb{Q} \to \mathbb{Z}$.

Exercise 4.8.51. Complete the following steps to prove that (up to isomorphism), the only finite field is $\mathbb{Z}/p\mathbb{Z}$, i.e., every finite field is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for some prime number p.

(a.) Consider any finite field k. Prove that the function $\varphi : \mathbb{Z} \to k$ defined by $\varphi(n) = n \cdot 1_k$ is a surjective unital ring homomorphism.

(b.) Prove that the kernel of φ is equal to $p\mathbb{Z}$ for some prime number p.

(c.) Conclude by the First Isomorphism Theorem for Rngs that k is isomorphic to $\mathbb{Z}/p\mathbb{Z}$.

Exercise 4.8.52. Consider the non-commutative unital ring $k^{n \times n}$ consisting of all $n \times n$ matrices over the field k for some positive integer $n \geq 2$. Generalize the proof of Exercise 4.8.34 to prove that there are no non-trivial two sided ideals of $k^{n \times n}$, i.e., the only nonzero ideal of $k^{n \times n}$ is $k^{n \times n}$.

4.8.5 Prime and Maximal Ideals

Exercise 4.8.53. Determine all prime ideals of the following commutative unital rings.

(a.) $\dfrac{\mathbb{Z}}{7\mathbb{Z}}$ (b.) $\dfrac{\mathbb{Z}}{30\mathbb{Z}}$ (c.) $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{3\mathbb{Z}}$

(d.) $\mathbb{Z}$ (e.) $\mathbb{Q}$ (f.) $\mathbb{R}$

Exercise 4.8.54. Determine all maximal ideals of the following commutative unital rings.

(a.) $\dfrac{\mathbb{Z}}{7\mathbb{Z}}$ (b.) $\dfrac{\mathbb{Z}}{30\mathbb{Z}}$ (c.) $\dfrac{\mathbb{Z}}{2\mathbb{Z}} \times \dfrac{\mathbb{Z}}{3\mathbb{Z}}$

(d.) $\mathbb{Z}$ (e.) $\mathbb{Q}$ (f.) $\mathbb{R}$

Exercise 4.8.55. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Prove that for any elements $a, b \in \mathbb{R}$ such that a is nonzero, the ideal $(ax+b)$ of $\mathbb{R}[x]$ is maximal.

(**Hint:** Prove that the quotient ring $\mathbb{R}[x]/(ax+b)$ is isomorphic to the field $\mathbb{R}$.)

Exercise 4.8.56. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Complete the following steps to prove that (x^2+1) is a maximal ideal of $\mathbb{R}[x]$.

(a.) Prove that $\mathbb{R}[x]/(x^2+1) = \{ax + b + (x^2+1) \mid a, b \in \mathbb{R}\}$.

(b.) Prove that the function $\varphi : \mathbb{R}[x]/(x^2+1) \to \mathbb{C}$ defined by $\varphi(a + bx + (x^2+1)) = a + bi$ is a well-defined bijective unital ring homomorphism.

Exercise 4.8.57. Consider the commutative unital ring $\mathcal{C}^0(\mathbb{R})$ consisting of continuous real functions $f : \mathbb{R} \to \mathbb{R}$ under pointwise multiplication $(fg)(x) = f(x)g(x)$.

(a.) Prove that for every real number α, the ideal $I_\alpha = \{f : \mathbb{R} \to \mathbb{R} \mid f(\alpha) = 0\}$ is maximal.

(b.) Prove that for the ideals I_e and I_π defined in part (a.), the ideal $I_e \cap I_\pi$ is not prime.

(c.) Prove that the ideal of $\mathcal{C}^0(\mathbb{R})$ generated by the zero function is not prime.

Exercise 4.8.58. Consider the commutative unital ring $\mathbb{C}[x, y]$ of complex polynomials in indeterminates x and y. (One can view this as the ring of polynomials in y with coefficients in $\mathbb{C}[x]$.)

(a.) Prove that the quotient ring $\mathbb{C}[x, y]/(xy)$ is not an integral domain.

(b.) Prove that (x) is an ideal of $\mathbb{C}[x, y]$ that contains (xy).

(c.) Prove that the quotient ring $\dfrac{\mathbb{C}[x, y]/(xy)}{(x)/(xy)}$ is isomorphic to $\mathbb{C}[x, y]/(x)$.

(d.) Prove that the function $\varphi : \mathbb{C}[x, y] \to \mathbb{C}[x]$ defined by $\varphi(p(x, y)) = p(x, 0)$ is a surjective unital ring homomorphism such that $\ker \varphi = (x)$.

(**Hint:** One can readily verify that $(x) \subseteq \ker \varphi$. Conversely, we may write any polynomial $p(x, y)$ as $p(x, y) = q(x, y)x + r(y)$ for some polynomial $r(y)$ in indeterminate y alone.)

(e.) Conclude by the First Isomorphism Theorem for Rngs that $\mathbb{C}[x, y]/(x) \cong \mathbb{C}[x]$

(f.) Conclude that $(x)/(xy)$ is a prime ideal of $\mathbb{C}[x, y]/(xy)$.

Exercise 4.8.59. Consider the commutative unital ring $\mathbb{R}[x, y, z]$ of real polynomials in the indeterminates x, y, and z. Prove that $I = (x, y)$ is a prime ideal $\mathbb{R}[x, y, z]$.

Exercise 4.8.60. Consider any prime ideal P of a commutative unital ring R. Prove that if I and J are any ideals of R such that $IJ \subseteq P$, then we must have that either $I \subseteq P$ or $J \subseteq P$.

Exercise 4.8.61. Consider distinct maximal ideals M_1 and M_2 of a commutative unital ring R.

(a.) Prove that $M_1 \cap M_2$ is not prime.

(b.) Prove that $M_1 + M_2 = R$. We say in this case that the ideals M_1 and M_2 are **comaximal**.

(c.) Prove that for any integer $n \geq 1$, we have that $M_1^n + M_2^n = R$.

(**Hint:** Prove that if M is any maximal ideal of R such that $M_1^n + M_2^n \subseteq M$, we have that $M_1 + M_2 \subseteq M$. Conclude that $M_1^n + M_2^n$ must contain every maximal ideal of R.)

Exercise 4.8.62. Consider the commutative unital ring $\mathbb{R}[x, y]$ of bivariate real polynomials. Given any positive integer $n \geq 1$ and any distinct points $P_1 = (a_1, b_1)$ and $P_2 = (a_2, b_2)$ in $\mathbb{R} \times \mathbb{R}$, prove that for each polynomial $f(x, y) \in \mathbb{R}[x, y]$, there exist polynomials $g(x, y), h(x, y) \in \mathbb{R}[x, y]$ with

(1.) $f(x, y) = g(x, y) + h(x, y)$;

(2.) $g(x, y)$ and all of its partial derivatives of order less than n vanish at P_1; and

(3.) $h(x, y)$ and all of its partial derivatives of order less than n vanish at P_2.

(**Hint:** Consider the result of Exercise 4.8.61.)

Exercise 4.8.63. Consider an integral domain R and a collection $\{P_n\}_{n=1}^{\infty}$ of prime ideals.

(a.) Prove that if $P_1 \supseteq P_2 \supseteq P_3 \supseteq \cdots$ is a descending chain, then $\cap_{n=1}^{\infty} P_n$ is a prime ideal.

(b.) Give an explicit counterexample to part (a.) when the primes do not form a descending chain.

(**Hint:** Consider any pair of distinct prime ideals of the ring of integers $\mathbb{Z}$.)

4.8.6 Polynomials Rings and Polynomial Long Division

Exercise 4.8.64. Compute each of the following polynomials in the indicated polynomial ring.

(a.) $(x+1)^3$ in $(\mathbb{Z}/3\mathbb{Z})[x]$

(b.) $(3x+2)^2$ in $(\mathbb{Z}/4\mathbb{Z})[x]$

(c.) $(2x+1)(3x+1)$ in $(\mathbb{Z}/6\mathbb{Z})[x]$

(d.) $(2x^2+x+7)(4x^2+6x+7)$ in $(\mathbb{Z}/8\mathbb{Z})[x]$

(e.) $(x+3)^3$ in $(\mathbb{Z}/9\mathbb{Z})[x]$

(f.) $(5x^2+5)(6x^3+2x)$ in $(\mathbb{Z}/10\mathbb{Z})[x]$

Exercise 4.8.65. Compute the roots of each polynomial in the indicated polynomial ring.

(a.) x^3-x+1 in $(\mathbb{Z}/2\mathbb{Z})[x]$

(b.) x^5-x in $(\mathbb{Z}/5\mathbb{Z})[x]$

(c.) $x^5+6x^4+3x^2+1$ in $(\mathbb{Z}/7\mathbb{Z})[x]$

(d.) $3x-7$ in $\mathbb{Z}[x]$

(e.) x^6-16x^3+64 in $\mathbb{Q}[x]$

(f.) x^4-4 in $\mathbb{R}[x]$

Exercise 4.8.66. Use the Rational Roots Theorem to find the rational roots of the following.

(a.) x^3+x+1

(b.) $2x^3-x^2+2x-1$

(c.) $x^3-6x^2+11x-6$

(d.) $4x^4-13x^2+9$

Exercise 4.8.67. Complete the polynomial long division in the indicated polynomial ring.

(a.) $\dfrac{x^3-6x^2+11x-6}{x-1}$ in $\mathbb{Z}[x]$

(b.) $\dfrac{x^4+x^2+1}{x^2-x+1}$ in $\mathbb{Z}[x]$

(c.) $\dfrac{x^5-x^2+1}{x^2+1}$ in $\mathbb{Z}[x]$

(d.) $\dfrac{x^5-x^3+x^2+1}{x^2+1}$ in $(\mathbb{Z}/2\mathbb{Z})[x]$

(e.) $\dfrac{x^4+x^3+x^2+x+1}{x-1}$ in $(\mathbb{Z}/5\mathbb{Z})[x]$

Exercise 4.8.68. Consider the univariate polynomial rng $R[x]$ over an arbitrary rng R. Prove that if $p(x)$ is any polynomial of $R[x]$ whose leading coefficient is a regular element of R, we have that $\deg(pq) = \deg(p) + \deg(q)$ for any polynomial $q(x) \in R[x]$.

Exercise 4.8.69. Consider the univariate polynomial rng $R[x]$ over an arbitrary rng R. Prove that if $p(x)$ is any polynomial of $R[x]$ whose leading coefficient is a regular element of R, then ever polynomial of the form $p(x)q(x) + r(x)$ such that $q(x)$ and $r(x)$ are polynomials of $R[x]$ and either $r(x)$ is the zero polynomial or $0 \leq \deg(r) \leq \deg(p) - 1$ is uniquely determined by $q(x)$ and $r(x)$.

Exercise 4.8.70. Complete the following steps to prove that any polynomial $f(x)$ of a polynomial rng $R[x]$ over an arbitrary rng R can be uniquely divided by any monic polynomial $p(x)$. Explicitly, prove that for any polynomial $f(x)$ in $R[x]$, there exist unique polynomials $q(x)$ and $r(x)$ such that $f(x) = p(x)q(x) + r(x)$ and either $r(x)$ is the zero polynomial or $0 \leq \deg(r) \leq \deg(p) - 1$.

(a.) We proceed by the Principle of Complete Induction on the degree of the polynomial $f(x)$ that we wish to divide by the monic polynomial $p(x)$. Prove the statement in the following cases.

(i.) $f(x)$ is the zero polynomial.

(ii.) $p(x)$ is the constant polynomial 1_R.

(b.) Conclude that we may assume that neither $f(x)$ is the zero polynomial nor $p(x)$ is the constant polynomial 1_R. Particularly, we may assume that $\deg(p) - 1 \geq 0$. Prove that the statement holds in the case that $f(x)$ is a nonzero constant polynomial (so that $\deg(f) = 0$).

(c.) Based on the previous part of the exercise, we may assume inductively that the statement holds for all polynomials of degree at most $n - 1$. Consider the case that $f(x)$ has degree n. Prove the existence part of the statement in the case that $\deg(p) - 1 \geq n$.

(d.) We may assume next that the degree m of $p(x)$ is at most the degree of $f(x)$. Consider the leading coefficient r_n of $f(x)$. Prove that $f(x) - r_n x^{n-m} p(x)$ is a polynomial of degree strictly smaller than n; then, appeal to complete induction to find polynomials $q(x)$ and $r(x)$ such that $f(x) - r_n x^{n-m} p(x) = p(x)q(x) + r(x)$ and either $r(x)$ is the zero polynomial or $0 \leq \deg(r) \leq \deg(p) - 1$. Conclude the existence by the Principle of Complete Induction.

(e.) Last, we will prove the uniqueness of the polynomials $q(x)$ and $r(x)$. Consider any polynomials $q_1(x)$, $q_2(x)$, $r_1(x)$, and $r_2(x)$ such that $f(x) = p(x)q_1(x) + r_1(x)$ and $f(x) = p(x)q_2(x) + r_2(x)$. Compare the identities to conclude that $p(x)(q_1(x) - q_2(x)) = r_2(x) - r_1(x)$. On the contrary, suppose that $q_1(x) - q_2(x)$ and $r_2(x) - r_1(x)$ are nonzero polynomials. Compare the degrees of the polynomials $p(x)(q_1(x) - q_2(x))$ and $r_2(x) - r_1(x)$ to derive a contradiction.

Explain how the above proof can be generalized to demonstrate that any polynomial $f(x)$ of $R[x]$ can be uniquely divided by a polynomial $p(x)$ whose leading coefficient is a unit.

Exercise 4.8.71. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Convert Exercise 4.8.28 to use the Polynomial Division Algorithm to prove the following.

(a.) $(ax + b)$ is a maximal ideal of $\mathbb{R}[x]$ for any real numbers a and b such that a is nonzero.

(b.) $(x^2 + 1)$ is a maximal ideal of $\mathbb{R}[x]$.

Exercise 4.8.72 (Rational Roots Theorem)**.** Complete the following steps to prove that for any polynomial $p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$ of degree n with integer coefficients, if a and b are relatively prime integers, then the rational number $\frac{a}{b}$ (written in lowest terms) is a root of $p(x)$ only if a divides the constant term c_0 and b divides the leading coefficient c_n of $p(x)$.

(a.) Consider any rational number $\frac{a}{b}$ such that $\gcd(a, b) = 1$. Prove that if

$$p\left(\frac{a}{b}\right) = c_n \left(\frac{a}{b}\right)^n + c_{n-1} \left(\frac{a}{b}\right)^{n-1} + \cdots + c_1 \left(\frac{a}{b}\right) + c_0 = 0,$$

then it follows that $c_n a^n + c_{n-1} a^{n-1} b + \cdots + c_1 a b^{n-1} + c_0 b^n = 0$.

(b.) Conclude from the previous step that a divides $c_0 b^n$.

(c.) Conclude from Exercise 1.10.28 that a divides c_0.

(d.) Conclude from the first step that b divides $c_n a^n$.

(e.) Conclude from Exercise 1.10.28 that b divides c_n.

4.8.7 Polynomial Irreducibility

Exercise 4.8.73. Determine if each polynomial is irreducible in the indicated polynomial ring.

(a.) $2x + 3$ in $\mathbb{Z}[x]$

(b.) $2x + 3$ in $\mathbb{Q}[x]$

(c.) $x^2 - 4$ in $\mathbb{Z}[x]$

(d.) $x^2 + x + 1$ in $\mathbb{Z}[x]$

(e.) $x^2 + x + 1$ in $\mathbb{C}[x]$

(f.) $x^3 - 8$ in $\mathbb{Z}[x]$

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

(h.) $x^3 + x + 1$ in $\mathbb{R}[x]$

(i.) $2x^4 + 9x - 6$ in $\mathbb{Z}[x]$

(j.) $x^4 + x^2 + 1$ in $\mathbb{Z}[x]$

(k.) $x^5 - 32$ in $\mathbb{Z}[x]$

(l.) $5x^5 - 11x^4 + 22x^2 - 33$ in $\mathbb{Z}[x]$

(m.) $7x^3 + 6x^2 + 4x + 6$ in $\mathbb{Z}[x]$

(n.) $9x^4 + 4x^3 - 3x + 7$ in $\mathbb{Z}[x]$

Exercise 4.8.74 (Freshman's Dream)**.** Consider any commutative unital ring R of prime characteristic p. Prove that the identity $(r+s)^p = r^p + s^p$ holds for any elements $r, s \in R$.

(**Hint:** Use the Binomial Theorem to write $(r+s)^p$ as a sum of products of the form $r^i s^{p-i}$ for each integer $0 \leq i \leq p$; then, express the binomial coefficients $\binom{p}{i}$ as integers in fraction form. Conclude that for each integer $1 \leq i \leq n-1$, the binomial coefficient $\binom{p}{i}$ is divisible by p.)

Exercise 4.8.75. Consider any prime number p. Prove that the polynomial $x^p - x$ has p distinct roots in $\mathbb{Z}/p\mathbb{Z}$. Conclude that $x^p - x = x(x-1)(x-2)\cdots(x-(p-1))$ in $(\mathbb{Z}/p\mathbb{Z})[x]$.

(**Hint:** Combine Fermat's Little Theorem and the evaluation homomorphisms from $(\mathbb{Z}/p\mathbb{Z})[x]$.)

Exercise 4.8.76. Prove that there are infinitely many irreducible polynomials in $\mathbb{Q}[x]$.

Exercise 4.8.77. Prove that there are irreducible polynomials of arbitrary positive degree in $\mathbb{Q}[x]$.

Exercise 4.8.78. Given any positive integer n, the nth **cyclotomic polynomial** is given by

$$\Phi_n(x) = \frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + x + 1.$$

Complete the following steps to prove that if p is a prime number, then $\Phi_p(x)$ is irreducible in $\mathbb{Q}[x]$.

(a.) Prove that $\Phi_p(x)$ is reducible if and only if $\Phi_p(x+1)$ is reducible.

(b.) Prove that every non-leading coefficient of $\Phi_p(x+1)$ is divisible by p.

(c.) Prove that the constant term of $\Phi_p(x)$ is not divisible by p^2.

(d.) Conclude by Eisenstein's Criterion the $\Phi_p(x+1)$ is p-Eisenstein. Conclude that $\Phi_p(x+1)$ is irreducible so that $\Phi_p(x)$ is irreducible by the first part above.

Remarkably, it is true that if n is composite, then $\Phi_n(x)$ is reducible! Even though the proof when n is odd is far from trivial, prove that if $n = 2k$ for some integer $k \geq 2$, then $\Phi_n(-1) = 0$. Conclude that if n is any even integer exceeding two, then $\Phi_n(x)$ is divisible by $x + 1$, hence it is reducible.

Given any prime number p, call a polynomial $q(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ with integer coefficients p-**Steisenein** if it holds that

(1.) p divides each of the coefficients $a_1, a_2, \ldots, a_n$ and

(2.) p does not divide the constant term a_0 and

(3.) p^2 does not divide the leading coefficient a_n.

Exercise 4.8.79 (Eisenstein's Criterion, Revisited)**.** Prove that if $q(x)$ is a primitive p-Steisenein polynomial in $\mathbb{Z}[x]$ for some prime number p, then $q(x)$ is irreducible in $\mathbb{Q}[x]$.

(**Hint:** Prove that $x^n q(1/x)$ is a primitive p-Eisenstein polynomial in $\mathbb{Z}[x]$.)

Exercise 4.8.80. Let k be any field. Complete the following steps to prove that every non-constant polynomial in $k[x]$ can be written as a product of irreducible polynomials in $k[x]$.

(a.) Consider the collection N of non-constant polynomials in $k[x]$ that *cannot* be written as a product of irreducible polynomials in $k[x]$. We seek to demonstrate that N is empty. On the contrary, suppose that it is nonempty. Explain why no polynomial in N is irreducible.

(b.) Prove that N admits a polynomial $p(x)$ such that every polynomial of $k[x]$ of degree strictly smaller than $\deg(p)$ admits a factorization as a product of irreducible polynomials.

(c.) Conclude from the previous two steps that $p(x)$ can be written as a product of irreducible polynomials; then, conclude from this contradiction that N is empty.

Exercise 4.8.81. Prove that if $p(x)$ is a polynomial of odd degree in $\mathbb{R}[x]$ that does not admit a root with multiplicity exceeding one, then $p(x)$ has an odd number of real roots.

(**Hint:** By Theorem 4.7.21, write $p(x) = p_1(x) \cdots p_n(x)$ for some real polynomials $p_1(x), \ldots, p_n(x)$ of degree one and two. Express $\deg(p)$ in terms of $\deg(p_1), \ldots, \deg(p_n)$ and rearrange.)

Chapter 5

Ring Theory II

We have demonstrated thus far that a rng is an algebraic structure in which there exist well-defined notions of addition and multiplication. Generally, the order of the rng elements in a product cannot be interchanged; however, in a commutative rng, this is not the case. Unital rings admit unique multiplicative identity elements; the elements of a unital ring that admit multiplicative inverses are called units. We say that a unital ring is a skew field if all of its nonzero elements are units; a unital ring is called a domain if all of nonzero its elements satisfy the Zero Product Property. Commutative skew fields are simply called fields; commutative domains are called integral domains. Our main objective throughout this chapter is to examine the rich hierarchy of integral domains.

5.1 Euclidean Domains

Given any integer a and any nonzero integer b, recall that the Division Algorithm guarantees that there exist unique integers q and r such that $a = bq + r$ and $0 \leq r < |b|$. Even more, for any pair of nonzero integers a and b, we have that $|a| \leq |a| \cdot |b| = |ab|$. Consequently, we refer to the absolute value function $|\cdot| : \mathbb{Z} \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ as a **Euclidean function**. Generally, a Euclidean function (or **valuation**) on an integral domain R is a function $v : R \setminus \{0_R\} \to \mathbb{Z}_{\geq 0}$ satisfying that

(1.) for any element $a \in R$ and any nonzero element $b \in R$, there exist elements $q, r \in R$ such that $a = bq + r$ and either $r = 0_R$ or $0 \leq v(r) < v(b)$ and

(2.) for all nonzero elements $a, b \in R$, we have that $v(a) \leq v(ab)$.

If R is an integral domain that admits a valuation $v : R \setminus \{0_R\} \to \mathbb{Z}_{\geq 0}$, then R is called a **Euclidean domain**. We say that a valuation v is **multiplicative** if $v(a) \geq 1$ and $v(ab) = v(a)v(b)$.

Example 5.1.1. We have already verified that the absolute value function $|\cdot| : \mathbb{Z} \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ constitutes a multiplicative valuation of the integers. Consequently, $\mathbb{Z}$ is a Euclidean domain.

Example 5.1.2. Given any field k, consider the degree function $v : k[x] \setminus \{0_k\} \to \mathbb{Z}_{\geq 0}$ defined by $v(p(x)) = \deg(p)$. Every nonzero element of k is a unit, hence every nonzero polynomial in $k[x]$ gives rise to a monic polynomial by multiplying by the inverse of the leading coefficient. Consequently, by the Polynomial Division Algorithm, for any polynomial $f(x)$ and any nonzero polynomial $p(x)$ in $k[x]$, there exist polynomials $q(x), r(x) \in k[x]$ such that $f(x) = p(x)q(x) + r(x)$ and either $r(x) = 0_k$

or $0 \leq \deg(r) < \deg(p)$. Even more, we have that $\deg(p) \leq \deg(p) + \deg(q) = \deg(pq)$ for any pair of nonzero polynomials $p(x)$ and $q(x)$ in $k[x]$. We conclude that $k[x]$ is a Euclidean domain.

Example 5.1.3. Given any field k, consider the function $v : k \setminus \{0_k\} \to \mathbb{Z}_{\geq 0}$ defined by $v(x) = 1$. Every nonzero element of k is a unit, hence for any element $x \in k$ and any nonzero element $y \in k$, there exists an element $y^{-1} \in k$ such that $x = 1_R x = (yy^{-1})x = y(y^{-1}x)$. Clearly, for all nonzero elements $xy \in k$, we have that xy is nonzero so that $v(x) = v(xy) = 1 = v(x)v(y)$. We conclude that k is a Euclidean domain. Even more, observe that v is a multiplicative valuation on k.

Example 5.1.4. Consider the commutative unital subring $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ of the complex numbers. Back in Example 4.4.13, we referred to this as the ring of Gaussian integers. By Exercise 5.2.1, $v(a + bi) = a^2 + b^2$ is a (multiplicative) valuation on $\mathbb{Z}[i]$, hence $\mathbb{Z}[i]$ is a Euclidean domain.

We demonstrate next that the units of a Euclidean domain are precisely those elements whose valuation agrees with the valuation of the multiplicative identity.

Proposition 5.1.5. *Let R be a Euclidean domain with a valuation $v : R \setminus \{0_R\} \to \mathbb{Z}_{\geq 0}$. We have that $u \in R$ is a unit if and only if $v(u) = v(1_R)$. Even more, if v is multiplicative, then $v(1_R) = 1$.*

Proof. We will assume first that $v(u) = v(1_R)$. By hypothesis that v is a valuation on R, there exist elements $q, r \in R$ such that $1_R = qu + r$ and either $r = 0_R$ or $v(r) < v(u)$. On the contrary, suppose that $v(r) < v(u)$. By rearranging the identity $1_R = qu + r$, we find that $r = 1_R - qu$ so that $v(r) = v(1_R - qu)$ and $v(u) = v(1_R) \leq v(1_R - qu) = v(r) < v(u)$ — a contradiction. We conclude that $1_R = qu$ so that u is a unit. Conversely, if u is a unit, then there exists a unique element $u^{-1} \in R$ such that $1_R = uu^{-1}$ and $v(u) \leq v(uu^{-1}) = v(1_R) \leq v(1_R u) = v(u)$.

Even more, if v is multiplicative, then $v(1_R) = v(1_R 1_R) = v(1_R)v(1_R)$. Cancelling the nonzero integer $v(1_R)$ from both sides of this identity, we conclude that $v(1_R) = 1$. □

One of the fundamental properties of a Euclidean domain is that it is a principal ideal ring.

Proposition 5.1.6. *Given any nonzero ideal I of a Euclidean domain R, there exists a nonzero element $x \in I$ such that $I = xR$. Consequently, every ideal of a Euclidean domain is principal.*

Proof. Consider the collection $V = \{v(i) \mid i \in I \text{ is nonzero}\} \subseteq \mathbb{Z}_{\geq 0}$, where v is a valuation of R. By the Well-Ordering Principle, there exists a least element $v(x)$ of V. We claim that $I = xR$. By hypothesis that I is an ideal containing x, we have that $xr \in I$ for all elements $r \in R$ so that $xR \subseteq I$. Conversely, for any element $y \in I$, there exist elements $q, R \in R$ such that $y = qx + r$ and either $r = 0_R$ or $v(r) < v(x)$. By assumption that I is an ideal, we must have that $qx \in I$ so that $r = y - qx \in I$. But as such, the minimality of x with respect to the valuation v of R precludes the possibility that $v(r) < v(x)$. We conclude that $y = qx$ is in xR so that $I \subseteq xR$, as desired. □

Contrapositively, Proposition 5.1.6 states that if R is an integral domain in which there exists an ideal that *is not* principal, then R is not a Euclidean domain. We put this to use immediately.

Example 5.1.7. We will prove that the ideal $I = (2, x)$ of the polynomial ring $\mathbb{Z}[x]$ is not principal. Consequently, it will follow that $\mathbb{Z}[x]$ is neither a Euclidean domain nor a principal ideal domain. On the contrary, suppose that $I = p(x)\mathbb{Z}[x]$ for some polynomial $p(x) \in \mathbb{Z}[x]$. Considering that $2 = 2 \cdot 1 + x \cdot 0 \in I = p(x)\mathbb{Z}[x]$, there must exist a polynomial $q(x) \in \mathbb{Z}[x]$ such that $2 = p(x)q(x)$.

By Proposition 4.6.4, we have that $0 = \deg(2) = \deg(pq) = \deg(p) + \deg(q)$, from which it follows that $\deg(p) = \deg(q) = 0$ so that $p(x)$ and $q(x)$ are both integers. Considering that 2 is prime and $2 = p(x)q(x)$, we must have that either $p(x) = \pm 2$ or $p(x) = \pm 1$. But in any case, we have a contradiction: indeed, if $p(x) = \pm 2$, then x does not lie in $p(x)\mathbb{Z}[x]$ because every element of $p(x)\mathbb{Z}[x]$ must be divisible by 2; if $p(x) = \pm 1$, then $I = p(x)\mathbb{Z}[x] = \mathbb{Z}[x]$ despite the fact that 1 is not in I. By the contrapositive of Proposition 5.1.6, it follows that $\mathbb{Z}[x]$ is not a Euclidean domain.

Example 5.1.7 illustrates that the contrapositive of Proposition 5.1.6 affords a tool to decipher when an integral domain is not a Euclidean domain. One can show that for any prime number p, the ideal (p, x) of $\mathbb{Z}[x]$ is not principal; however, we shall soon see that there exist non-Euclidean integral domains in which all ideals are principal, hence the task of finding non-principal ideals of an arbitrary integral domain becomes more complicated (and sometimes impossible) in general.

5.2 Chapter 5 Exercises

5.2.1 Euclidean Domains

Exercise 5.2.1. Complete the following steps to prove that the Gaussian integers $\mathbb{Z}[i]$ admit a multiplicative valuation $v : \mathbb{Z}[i] \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ defined by $v(a + bi) = a^2 + b^2$.

(a.) Prove that $v(a + bi) \geq 1$ for all nonzero integers a and b.

(b.) Prove that if $w = a + bi$ and $z = c + di$, then $v(wz) = v(w)v(z)$.

(c.) Prove that if z is nonzero, then $v(w) \leq v(wz)$.

(d.) Prove that if c and d are nonzero, then $\dfrac{a + bi}{c + di} = \alpha + \beta i$ lies in $\mathbb{Q}[i]$.

(e.) Prove that there exist integers m and n such that $|m - \alpha| \leq \frac{1}{2}$ and $|n - \beta| \leq \frac{1}{2}$.

(f.) Prove that $a = b(m + ni) + b[(\alpha - m) + (\beta - n)i]$.

(g.) Prove that $r = b[(\alpha - m) + (\beta - n)i$ lies in $\mathbb{Z}[i]$.

(h.) Prove that $v(r) < v(b)$.

(i.) Conclude that $\mathbb{Z}[i]$ is a Euclidean domain.

Exercise 5.2.2. Consider the Gaussian integers $\mathbb{Z}[i]$ with the usual multiplicative valuation $v(a + bi) = a^2 + b^2$. We say that $a + bi$ is **prime** if the valuation of any factor of $a + bi$ in $\mathbb{Z}[i]$ is either one or $a^2 + b^2$. Prove that if $a^2 + b^2$ is a prime number, then $a + bi$ is a prime element of $\mathbb{Z}[i]$.

Exercise 5.2.3. Verify that any unit multiple of the following Gaussian integers is prime.

(a.) $1 + i$

(b.) $2 - 3i$

(c.) $p + 0i$ such that p is a prime number and $p \equiv 3 \pmod 4$

By [Con22, Theorem 9.9], the only prime elements of $\mathbb{Z}[i]$ are of the $a+bi$ such that a^2+b^2 is a prime number or $p+0i$ such that p is a prime number and $p \equiv 3 \pmod 4$.

Exercise 5.2.4. Compute the prime factorizations of the following Gaussian integers.

(a.) $1-i$

(b.) $3+0i$

(c.) $6+0i$

(d.) $5+3i$

Exercise 5.2.5. Compute the number of distinct ideals of the quotient ring $\mathbb{Z}[i]/6\mathbb{Z}[i]$.

Exercise 5.2.6. Consider the commutative unital ring $\mathbb{C}[t,t^{-1}]$ of complex **Laurent polynomials** in indeterminate t. Explicitly, every element of $\mathbb{C}[t,t^{-1}]$ can be written as

$$p(t) = \sum_{i=-m}^{n} a_i t^i = a_n t^n + \cdots + a_1 t + a_0 + \frac{a_{-1}}{t} + \cdots + \frac{a_{-m}}{t^m}$$

for some non-negative integers m and n and some complex numbers $a_n, \dots, a_1, a_0, a_{-1}, \dots, a_{-m}$.

(a.) Prove that every element of $\mathbb{C}[t,t^{-1}]$ can be written as $p(t) = t^{-m}P(t)$ with $P(t) \in \mathbb{C}[t]$.

(b.) Prove that the function $v : \mathbb{C}[t,t^{-1}] \to \mathbb{Z}_{\geq 0}$ defined by $v(p(t)) = \deg(P)$ is a valuation.

(c.) Conclude from the previous step that every ideal of $\mathbb{C}[t,t^{-1}]$ is principal.

Chapter 6

Field Theory I

Classically, the development of field theory began as early as the sixteenth century with the development of the Quadratic Formula, the Cubic Formula, and the Quartic Formula. Culminating in one of the landmark results of the field, the eponymous works of the precocious French mathematician Évariste Galois in the early 1800s inspired the development of Galois Theory that is still used extensively in contemporary mathematics. Particularly, it is a consequence of the theory of Galois that there is not (in general) a formula to produce the roots of real polynomials of degree greater than or equal to five. We begin our studies in field theory with a view toward Galois Theory.

6.1 Roots of Polynomials and Field Extensions

We will concern ourselves throughout this chapter with the univariate polynomial ring $k[x]$ for some field k. Like in Section 4.7, we will typically deal with the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$, the field of complex numbers $\mathbb{C}$, or the finite fields $\mathbb{Z}/p\mathbb{Z}$ for some prime number p. We remind the reader that the elements of $k[x]$ are polynomials $p(x) = a_n x^n + \cdots + a_1 x + a_0$ with coefficients $a_n, \dots, a_1, a_0$ that are elements of the field k. Every nonzero element of a field is a unit, hence for any polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$ such that a_n is nonzero, we have that $q(x) = a_n^{-1} p(x) = x^n + a_n^{-1} a_{n-1} x^{n-1} + \cdots + a_n^{-1} a_1 x + a_n^{-1} a_0$ is a monic polynomial; therefore, we may restrict our attention to monic polynomials in $k[x]$. We say that an element $\alpha \in k$ is a **root** of $p(x) = x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ if and only if $p(\alpha) = \alpha^n + a_{n-1} \alpha^{n-1} + \cdots + a_1 \alpha + a_0 = 0_k$. Unfortunately, as we have seen, there exist fields that admit polynomials with no roots in the field. Explicitly, the polynomial $x^2 + 1$ in $\mathbb{R}[x]$ admits no root in $\mathbb{R}$: indeed, we have that α is a root of $x^2 + 1$ if and only if $\alpha^2 + 1 = 0$ if and only if $\alpha^2 = -1$ if and only if $\alpha = \pm\sqrt{-1}$, and this is not a real number. Consequently, it is in this sense that the field $\mathbb{R}$ of real numbers is deficient, and we set out to look for the smallest field k that contains $\mathbb{R}$ and all roots of polynomials in $\mathbb{R}[x]$. We know already from Theorem 4.7.21 and the Quadratic Formula that the only polynomials in $\mathbb{R}[x]$ that do not admit roots in $\mathbb{R}$ are the quadratic polynomials $ax^2 + bx + c$ for which the discriminant $b^2 - 4ac < 0$, hence it seems that $\mathbb{C}$ is the smallest field containing $\mathbb{R}$ and all roots of polynomials in $\mathbb{R}[x]$. Our aim throughout this section is to verify this intuition and use it to investigate similar situations over the field of rational numbers $\mathbb{Q}$ and the finite field $\mathbb{Z}/p\mathbb{Z}$ for a prime number p.

Given any monic polynomial $p(x) = x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ in $k[x]$, recall that the principal ideal $(p(x)) = \{p(x)q(x) : q(x) \in k[x]\}$ generated by $p(x)$ consists of all polynomials in $k[x]$ that are divisible by $p(x)$. Even more, the quotient ring $k[x]/(p(x))$ is a commutative unital ring for which the left coset $\bar{x} = x + (p(x))$ of the indeterminate x in $(p(x))$ satisfies that

$$\begin{aligned}
p(\bar{x}) + (p(x)) &= \bar{x}^n + a_{n-1}\bar{x}^{n-1} + \cdots + a_1\bar{x} + a_0 + (p(x)) \\
&= [x + (p(x))]^n + a_{n-1}[x + (p(x))]^{n-1} + \cdots + a_1[x + (p(x))] + a_0 + (p(x)) \\
&= [x^n + (p(x))] + a_{n-1}[x^{n-1} + (p(x))] + \cdots + a_1[x + (p(x))] + a_0 + (p(x)) \\
&= x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 + (p(x)) \\
&= p(x) + (p(x)) \\
&= 0_k + (p(x)),
\end{aligned}$$

hence $\bar{x}$ is a root of $p(x)$ in $k[x]/(p(x))$. Considering that the inclusion $k \to k[x]/(p(x))$ that sends an element α of k to the left coset $\alpha + (p(x))$ of $k[x]/(p(x))$ is a unital ring homomorphism, we may identify the field k with a unital subring of $k[x]/(p(x))$ by the First Isomorphism Theorem for Rngs. We have found a commutative unital ring that contains (an isomorphic copy of) k and a root of $p(x)$. Our next proposition gives a sufficient condition under which $k[x]/p(x)$ is a field.

Proposition 6.1.1. *Consider the univariate polynomial ring $k[x]$ over any field k. If $p(x)$ is any monic irreducible polynomial in $k[x]$, then $k[x]/(p(x))$ is a field that contains k and a root of $p(x)$.*

Proof. By Proposition 4.5.7, it suffices to prove that $(p(x))$ is a maximal ideal of $k[x]$. Given any proper ideal I of $k[x]$ such that $I \supseteq (p(x))$, we must demonstrate that $I \subseteq (p(x))$. By assumption that I is a proper ideal of $k[x]$, the monic constant polynomial 1_R does not lie in I. Consequently, the degrees of the nonzero monic polynomials of I form a nonempty subset of positive integers, hence by the Well-Ordering Principle, there exists a nonzero monic polynomial $f(x) \in I$ of least positive degree. Even more, by the Polynomial Division Algorithm, there exist unique polynomials $q(x)$ and $r(x)$ in $k[x]$ such that $p(x) = f(x)q(x) + r(x)$ and either $r(x) = 0_k$ or $0 \leq \deg(r) \leq \deg(f) - 1$. Considering that $p(x)$ and $-f(x)q(x)$ both lie in I, their sum $r(x) = p(x) - f(x)q(x)$ lies in I. We note that if the leading coefficient a_n of $r(x)$ were nonzero, then we could find a monic polynomial $a_n^{-1}r(x)$ of strictly lesser degree than $f(x)$ — a contradiction. We conclude therefore that $r(x) = 0_k$ so that $p(x) = f(x)q(x)$. By assumption that $p(x)$ is irreducible, it must be the case that $q(x)$ is a nonzero constant, hence the degree of $p(x)$ and $f(x)$ are the same, i.e., we find that $p(x)$ is a monic polynomial of least positive degree in I. Given any polynomial $g(x) \in I$, once again, by the Polynomial Division Algorithm, there exist unique polynomials $Q(x)$ and $R(x)$ in $k[x]$ such that $g(x) = p(x)Q(x) + R(x)$ and $R(x) = 0_k$ or $0 \leq \deg(R) \leq \deg(p) - 1$. Like before, the polynomial $-p(x)Q(x)$ lies in I, hence the sum $R(x) = g(x) - p(x)Q(x)$ lies in I. But this forces $R(x) = 0_k$ by the same rationale as before. We conclude that $g(x) = p(x)Q(x) \in (p(x))$ so that $I \subseteq (p(x))$.

By the paragraph preceding this proposition, $k[x]/(p(x))$ contains k and a root of $p(x)$. □

Given any field k, we refer to a field F for which there exists an injective unital ring homomorphism $k \to F$ as an **extension field** of k. Consequently, Proposition 6.1.1 states that if $p(x)$ is an irreducible polynomial in $k[x]$, then $k[x]/(p(x))$ is an extension field of k that contains a root of $p(x)$. Considering its importance, we bear out the details of the following theorem of Kronecker.

Theorem 6.1.2 (Fundamental Theorem of Field Theory)**.** *Every non-constant univariate polynomial $p(x)$ over a field k induces an extension field F of k and an element $\alpha \in F$ such that $p(\alpha) = 0$.*

Proof. Every monic irreducible factor $q(x)$ of $p(x)$ induces a field $k[x]/(q(x))$ in which $q(x)$ admits the root $\alpha = x + (q(x))$ by the paragraph preceding Proposition 6.1.1 and the proposition itself. Considering that every root of $q(x)$ is a root of $p(x)$, the existence of the field F and the element $\alpha \in F$ such that $p(\alpha) = 0_F$ are established; in order to demonstrate that F is an extension field of k, it suffices to find an injective unital ring homomorphism $\varphi : k \to F$. Like we mentioned previously, the inclusion $\varphi(a) = a + (q(x))$ is clearly a unital ring homomorphism; it is injective because $a + (q(x)) = 0_k + (q(x))$ if and only if $a = q(x)f(x)$ if and only if $a = 0_k$ and $f(x) = 0_k$. □

Example 6.1.3. Consider the monic polynomial $x^2 - 2$ in $\mathbb{Q}[x]$. By the Quadratic Formula, the only roots of $x^2 - 2$ are $\pm\sqrt{2}$. Considering that $\sqrt{2}$ is not rational, it follows that $x^2 - 2$ is an irreducible monic polynomial in $\mathbb{Q}[x]$, hence $\mathbb{Q}[x]/(x^2 - 2)$ is an extension field of $\mathbb{Q}$ that contains a root of $x^2 - 2$. We will prove that the commutative unital ring $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ defined in Exercise 4.8.18 and the field $\mathbb{Q}[x]/(x^2 - 2)$ are isomorphic. Consider the function $\varphi : \mathbb{Q}[x]/(x^2 - 2) \to \mathbb{Q}(\sqrt{2})$ defined by $\varphi(a + bx + (x^2 - 2)) = a + b\sqrt{2}$. Clearly, it follows that φ is surjective. Given any pair of elements $a + b\sqrt{2}$ and $c + d\sqrt{2}$ of $\mathbb{Q}(\sqrt{2})$ such that $a + b\sqrt{2} = c + d\sqrt{2}$, we must have that $a - c = (d - b)\sqrt{2}$. Consequently, if $d - b$ were nonzero, then we would find that $\sqrt{2}$ is rational — a contradiction. We conclude that $b = d$ so that $a = c$ and $a + bx + (x^2 - 2) = c + dx + (x^2 - 2)$, i.e., φ is injective. Last, it is straightforward to verify that φ is a unital ring homomorphism: indeed, it is a group homomorphism because $(a + c) + (b + d)x + (x^2 - 2) = (a + bx) + (c + dx) + (x^2 - 2)$ and $\varphi((a + c) + (b + d)x + (x^2 - 2)) = (a + c) + (b + d)\sqrt{2} = (a + b\sqrt{2}) + (c + d\sqrt{2})$, and we have that

$$\begin{aligned}(a + bx + (x^2 - 2))(c + dx + (x^2 - 2)) &= ac + (ad + bc)x + bdx^2 + (x^2 - 2) \\ &= (ac + 2bd) + (ad + bc)x + bd(x^2 - 2) + (x^2 - 2) \\ &= (ac + 2bd) + (ad + bc)x + (x^2 - 2)\end{aligned}$$

gives $\varphi(a + bx + (x^2 - 2))(c + dx + (x^2 - 2)) = (ac + 2bd) + (ad + bc)\sqrt{2} = (a + b\sqrt{2})(c + d\sqrt{2})$. Explicitly, we have that $\varphi(x + (x^2 - 2)) = \sqrt{2}$, hence we obtain an algebraic description of $\sqrt{2}$. We note that in $\mathbb{Q}(\sqrt{2})[x]$, we have a complete factorization $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$.

Example 6.1.4. Consider the monic polynomial $x^2 + 1$ in $\mathbb{R}[x]$. We are well aware by now that the only roots of $x^2 + 1$ are the non-real complex numbers $\pm\sqrt{-1}$. Consequently, $x^2 + 1$ is an irreducible monic polynomial in $\mathbb{R}[x]$, hence $\mathbb{R}[x]/(x^2 + 1)$ is an extension field of $\mathbb{R}$ that contains both roots of $x^2 + 1$. By Exercise 4.8.56, $\mathbb{R}[x]/(x^2 + 1)$ and $\mathbb{C}$ are isomorphic via the unital ring homomorphism $\varphi : \mathbb{R}[x]/(x^2 + 1) \to \mathbb{C}$ defined by $\varphi(a + bx + (x^2 + 1)) = a + bi$. We may therefore identify the left coset $x + (x^2 + 1)$ of $\mathbb{R}[x]/(x^2 + 1)$ with the complex number $i = \sqrt{-1}$ to obtain a purely algebraic description of i. Even more, as a polynomial in $\mathbb{C}[x]$, we have that $x^2 + 1 = (x - i)(x + i)$.

Example 6.1.5. Observe that $0^2 + 0 + 1 = 1$ and $1^2 + 1 + 1 = 3 \equiv 1 \pmod 2$, hence the monic quadratic polynomial $x^2 + x + 1$ does not admit a root in $(\mathbb{Z}/2\mathbb{Z})[x]$. Consequently, $x^2 + x + 1$ is an irreducible monic polynomial in $(\mathbb{Z}/2\mathbb{Z})[x]$ so that $(\mathbb{Z}/2\mathbb{Z})[x]/(x^2+x+1)$ is an extension field of $\mathbb{Z}/2\mathbb{Z}$ that contains a root of $x^2 + x + 1$. By the Polynomial Division Algorithm, every polynomial $p(x)$ in $(\mathbb{Z}/2\mathbb{Z})[x]$ can be written uniquely as $p(x) = (x^2 + x + 1)q(x)$ for some unique polynomials $q(x)$ and $r(x) = ax + b$ in $(\mathbb{Z}/2\mathbb{Z})[x]$. Considering that $\mathbb{Z}/2\mathbb{Z}$ has two elements, there are simultaneously two choices for each of the elements $a, b \in \mathbb{Z}/2\mathbb{Z}$. We conclude that $(\mathbb{Z}/2\mathbb{Z})[x]/(x^2 + x + 1)$ is a field with 2^2 elements $0 + (x^2 + x + 1), 1 + (x^2 + x + 1), x + (x^2 + x + 1)$, and $x + 1 + (x^2 + x + 1)$. Like in the previous examples, there exists an isomorphism $\varphi : (\mathbb{Z}/2\mathbb{Z})[x]/(x^2 + x + 1) \to (\mathbb{Z}/2\mathbb{Z})(\alpha)$ defined by $\varphi(a + bx + (x^2 + x + 1)) = a + b\alpha$ for any root α of $x^2 + x + 1$, hence $(\mathbb{Z}/2\mathbb{Z})(\alpha)$ is a field with four elements $0, 1, \alpha$, and $\alpha + 1$. We note that $\alpha^2 = \alpha^2 + \alpha + 1 + \alpha + 1 = \alpha + 1$ because 2 and $\alpha^2 + \alpha + 1$ are both zero in $(\mathbb{Z}/2\mathbb{Z})(\alpha)$. Even more, we have that $\alpha(\alpha + 1) = \alpha^2 + \alpha = \alpha^2 + \alpha + 1 + 1 = 1$.

6.2 Simple Extensions

We will continue to assume that k is a field. Given any field F such that there exists an injective unital ring homomorphism $k \to F$, we say that F is an extension field of k; the injective unital ring homomorphism $k \to F$ is itself called the field extension of F over k. Often, in the literature, the two concepts are conflated; however, we will try to keep them separate for the sake of clarity.

By the Fundamental Theorem of Field Theory, every non-constant polynomial in $k[x]$ induces an extension field F of k in which there lies a root α of $p(x)$, i.e., we can always find a field F and an element $\alpha \in F$ such that $p(\alpha) = 0_k$. Conversely, given any extension field F over k, an element $\alpha \in F$ is **algebraic** over k if there exists a nonzero polynomial $p(x)$ in $k[x]$ such that $p(\alpha) = 0_F$.

Example 6.2.1. Considering that $\sqrt{2}$ is a root of the nonzero polynomial $x^2 - 2$ in $\mathbb{Q}[x]$, it follows that the real number $\sqrt{2}$ is algebraic over $\mathbb{Q}$. Likewise, the real number $-\sqrt{2}$ is algebraic over $\mathbb{Q}$.

Example 6.2.2. Observe that the complex number $i = \sqrt{-1}$ is a root of the polynomial $x^2 + 1$ in $\mathbb{Q}[x]$, hence i is algebraic over $\mathbb{Q}$. Likewise, the complex number $-i$ is algebraic over $\mathbb{Q}$.

Example 6.2.3. We will demonstrate that $\alpha = \sqrt{2 + \sqrt{3}}$ is algebraic over $\mathbb{Q}$.

$$\begin{aligned} \alpha^2 &= 2 + \sqrt{3} \\ \alpha^2 - 2 &= \sqrt{3} \\ (\alpha^2 - 2)^2 &= 3 \\ \alpha^4 - 4\alpha^2 + 4 &= 3 \\ \alpha^4 - 4\alpha^2 + 1 &= 0 \end{aligned}$$

Consequently, we find that α is a root of the rational polynomial $x^4 - 4x^2 + 1$.

Example 6.2.4. Elements of an extension field F of k need not be algebraic: indeed, it is a nontrivial fact that the real numbers π and e are not algebraic over $\mathbb{Q}$. Put another way, there is no nonzero polynomial $p(x)$ in $\mathbb{Q}[x]$ for which $p(\pi) = 0$ or $p(e) = 0$. We refer to the real numbers π and e as **transcendental**. Generally, an element $\alpha \in F$ is transcendental over k if α is not the root of any nonzero polynomial in $k[x]$, i.e., the evaluation homomorphism $\varphi_\alpha : k[x] \to F$ is injective.

We will say that a field extension $k \to F$ is an **algebraic extension** of k if every element of F is algebraic over k. Given any algebraic elements $\alpha_1, \dots, \alpha_n$ of F over k, we write $k(\alpha_1, \dots, \alpha_n)$ to denote the smallest extension field of k lying in F that contains k and the elements $\alpha_1, \dots, \alpha_n$. Explicitly, if α is any algebraic element of F over k, then $k(\alpha)$ is called a **simple extension** of k. Generally, an extension of the form $k(\alpha_1, \dots, \alpha_n)$ is called a **finitely generated extension** of k.

Example 6.2.5. By Example 6.2.1, we have that $\mathbb{Q}(\sqrt{2})$ is a simple extension of $\mathbb{Q}$.

Example 6.2.6. By Example 6.2.2, we have that $\mathbb{Q}(i)$ is a simple extension of $\mathbb{Q}$.

Example 6.2.7. We will demonstrate that the field $(\mathbb{Z}/2\mathbb{Z})(\alpha)$ of Example 6.1.5 is an algebraic extension of $\mathbb{Z}/2\mathbb{Z}$. Explicitly, we have that $(\mathbb{Z}/2\mathbb{Z})(\alpha) = \{0, 1, \alpha, \alpha + 1\}$ such that $\alpha^2 + \alpha + 1 = 0$. Certainly, the elements 0 and 1 are algebraic over $\mathbb{Z}/2\mathbb{Z}$ because they are the respective roots the polynomials x and $x - 1$ in $(\mathbb{Z}/2\mathbb{Z})[x]$. Even more, by construction, we have that α is the root of $x^2 + x + 1$ in $(\mathbb{Z}/2\mathbb{Z})[x]$, so it suffices to prove that $\alpha + 1$ is the root of a nonzero polynomial in $(\mathbb{Z}/2\mathbb{Z})[x]$. We note that $(\alpha+1)^2 = \alpha^2+2\alpha+1 = \alpha^2+1$ so that $(\alpha+1)^2+(\alpha+1) = (\alpha^2+\alpha+1)+1$ and $(\alpha+1)^2+(\alpha+1)+1 = (\alpha^2+\alpha+1)+1+1 = 0$. Consequently, $\alpha+1$ is a root of x^2+x+1. We note that the complete factorization of x^2+x+1 in $(\mathbb{Z}/2\mathbb{Z})(\alpha)[x]$ is $x^2+x+1 = (x+\alpha)(x+\alpha+1)$.

Generally, we have not yet discussed a description of the elements of a simple extension $k(\alpha)$, so we turn our attention to this matter next. We seek to leverage the Fundamental Theorem of Field Theory and the Proposition 6.1.1 that implies it. Before this, we need the following lemma.

Lemma 6.2.8. *Given any algebraic element α of any extension field F of any field k, there exists a unique monic irreducible polynomial $\mu_\alpha(x)$ in $k[x]$ of least positive degree that has α as a root. We refer to the unique monic irreducible polynomial $\mu_\alpha(x)$ as the* **minimal polynomial** *of α over k. Particularly, for any polynomial $p(x)$ in $k[x]$ such that $p(\alpha) = 0_k$, we have that $\mu_\alpha(x)$ divides $p(x)$.*

Proof. Consider the evaluation homomorphism $\varphi_\alpha : k[x] \to F$ at α defined by $\varphi_\alpha(p(x)) = p(\alpha)$. By hypothesis that α is algebraic over k, there exists a nonzero polynomial $p(x)$ in $k[x]$ such that $p(\alpha) = 0_k$, i.e., the kernel of φ_α is a nonzero proper ideal of $k[x]$. By the Well-Ordering Principle applied to the degrees of the nonzero polynomials in $\ker \varphi_\alpha$, there exists a polynomial $p(x)$ in $\ker \varphi_\alpha$ of least positive degree. We claim that $p(x)$ is divides every polynomial in the kernel of φ_α, i.e., we claim that $\ker \varphi_\alpha = (p(x))$. By the Polynomial Division Algorithm, for any polynomial $f(x)$ in $\ker \varphi_\alpha$, there exist unique polynomials $q(x)$ and $r(x)$ in $k[x]$ such that $f(x) = p(x)q(x) + r(x)$ and $r(x)$ is either zero or the degree of $r(x)$ is a non-negative integer that is strictly less than the degree of $p(x)$. Considering that $r(x) = f(x) - p(x)q(x)$ lies in $\ker \varphi_\alpha$, we must have that $r(x)$ is the zero polynomial; otherwise, we would have found a nonzero polynomial in $\ker \varphi_\alpha$ of strictly lesser degree than $p(x)$ — a contradiction. We conclude that $p(x)$ divides every polynomial in $\ker \varphi_\alpha$. Even more, we claim that $p(x)$ is irreducible: indeed, if we write $p(x) = q(x)r(x)$ for some polynomials $q(x)$ and $r(x)$ in $k[x]$, it follows that $0 = p(\alpha) = q(\alpha)r(\alpha)$ in the field F, hence the Zero Product Property yields that either $q(\alpha) = 0_k$ or $r(\alpha) = 0_k$. Certainly, either $q(x)$ or $r(x)$ must have the same degree as $p(x)$; otherwise, we would have found a nonzero polynomial in $\ker \varphi_\alpha$ of strictly lesser degree than $p(x)$ — a contradiction. We conclude that $p(x)$ is irreducible. Even more, if a is the leading coefficient of $p(x)$, then $\mu_\alpha(x) = a^{-1}p(x)$ is a monic irreducible polynomial of least positive degree that has α as a root; it is unique because it divides any polynomial with α as a root. □

Corollary 6.2.9. *Given any algebraic element α of any extension field F of any field k, if $p(x)$ is a monic irreducible polynomial in $k[x]$ and $p(\alpha) = 0_k$, we must have that $p(x) = \mu_\alpha(x)$.*

Proof. By Lemma 6.2.8, we must have that $p(x) = \mu_\alpha(x)q(x)$ for some polynomial $q(x)$ in $k[x]$. Considering that $p(x)$ is irreducible, the degree of $\mu_\alpha(x)$ must coincide with the degree of $p(x)$. Even more, $p(x)$ and $\mu_\alpha(x)$ are monic, hence we must have that $q(x) = 1_k$ so that $p(x) = \mu_\alpha(x)$. □

Example 6.2.10. We note that $x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ because it is the unique monic irreducible polynomial of least positive degree that has $\sqrt{2}$ as a root: indeed, there are no linear polynomials in $\mathbb{Q}[x]$ with $\sqrt{2}$ as a root because $\sqrt{2}$ is not a rational number.

Example 6.2.11. We note that $x^2 + 1$ is the minimal polynomial of $i = \sqrt{-1}$ over $\mathbb{Q}$ because it is the unique monic irreducible polynomial of least positive degree that has i as a root: indeed, there are no linear polynomials in $\mathbb{Q}[x]$ with i as a root because i is not a rational number.

Example 6.2.12. We have seen already in Example 6.2.3 that $x^4 - 4x^2 + 1$ is a monic polynomial of $\mathbb{Q}[x]$ that has $\sqrt{2 + \sqrt{3}}$ as a root; we will prove that it is the minimal polynomial of $\sqrt{2 + \sqrt{3}}$ over $\mathbb{Q}$. By the Rational Roots Theorem, the only possible rational roots of $x^4 - 4x^2 + 1$ are 1 and -1; it is not difficult to check that neither of them is actually a root. Consequently, we may assume that $x^4 - 4x^2 + 1 = (x^2 + ax + b)(x^2 + cx + d)$ for some integers a, b, c, and d. Expanding the product and comparing the coefficients of the monomials x^3, x^2, x, and 1 yields the following.

$$\begin{aligned} a + c &= 0 & ad + bc &= 0 \\ ac + b + d &= -4 & bd &= 1 \end{aligned}$$

We must have that $b = d = \pm 1$. Given that $b = d = 1$, the equations $a + c = 0$ and $ac + b + d = -4$ yield that $-a^2 = -6$ so that $a^2 = 6$ — a contradiction. Likewise, if $b = d = -1$, then we have that $-a^2 = -2$ so that $a^2 = 2$ — a contradiction. We conclude that $x^4 - 4x^2 + 1$ is irreducible in $\mathbb{Q}[x]$.

Given any algebraic element α of any extension field F of any field k, we refer to the degree of the minimal polynomial $\mu_\alpha(x)$ of α over k as the **degree** of the simple extension $k(\alpha)$ over k (or as the **degree** of α over k) and we write $[k(\alpha) : k]$ to denote this common degree.

Example 6.2.13. Example 6.2.10 illustrates that $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.

Example 6.2.14. Example 6.2.11 illustrates that $[\mathbb{Q}(i) : \mathbb{Q}] = 2$.

Example 6.2.15. Example 6.2.12 illustrates that $[\mathbb{Q}(\sqrt{2 + \sqrt{3}}) : \mathbb{Q}] = 4$.

We are now in a position to explicitly describe the elements of the simple extension $k(\alpha)$.

Proposition 6.2.16. *Consider any algebraic element α of any extension field F of any field k.*

1.) *We have that $k(\alpha) \cong k[x]/(\mu_\alpha(x))$ for the minimal polynomial $\mu_\alpha(x)$ of α.*

2.) *We have that $k(\alpha)$ is a k-vector space.*

3.) *We have that $\{1_k, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ is a k-vector space basis for $k(\alpha)$ if the degree of $\mu_\alpha(x)$ is n. Put another way, the degree of the simple extension $k(\alpha)$ over k is n.*

Proof. (1.) By Lemma 6.2.8 and the First Isomorphism Theorem for Rngs, the evaluation homomorphism $\varphi_\alpha : k[x] \to F$ at α induces a unital ring isomorphism $k[x]/(\mu_\alpha(x)) \cong \varphi(k[x])$. Considering that $k(\alpha)$ is a field that contains α, it must hold that $k(\alpha)$ contains the powers $1_k, \alpha, \alpha^2$, etc. Even more, $k(\alpha)$ contains k and must be closed under addition and multiplication, hence $k(\alpha)$ contains every polynomial of the form $a_n\alpha^n + \cdots + a_1\alpha + a_0$. Consequently, it holds that $k(\alpha)$ contains the field $\varphi(k[x])$. By definition, $k(\alpha)$ is the smallest field lying in F that contains k and α, hence $k(\alpha)$ must be equal to the field $\varphi(k[x])$ because $\varphi(k[x])$ is a field lying in F that contains k and α.

(2.) Every element of $k(\alpha)$ is of the form $a_n\alpha^n + \cdots + a_1\alpha + a_0$ for some elements $a_n, \dots, a_1, a_0$ of k. Consequently, we may realize $k(\alpha)$ as the collection of polynomials in α with coefficients in k. Considering that these polynomials form a k-vector space, so must the field $k(\alpha)$.

(3.) We have already seen that every element of $k(\alpha)$ is a polynomial in α with coefficients in k, hence it suffices to prove that $\{1_k, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ span $k(\alpha)$ as a k-vector space and are linearly independent over k. By the Polynomial Division Algorithm, every polynomial $p(x)$ in $k[x]$ can be written as $p(x) = \mu_\alpha(x)q(x) + r(x)$ for some polynomials $q(x)$ and $r(x) = a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. Consequently, we find that $p(\alpha) = \mu_\alpha(\alpha)q(\alpha) + r(\alpha) = r(\alpha) = a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha_1 + a_0$ because $\mu_\alpha(\alpha) = 0_k$ by definition, hence every element of $k(\alpha)$ can be written as a k-linear combination of the elements $1_k, \alpha, \alpha^2, \dots, \alpha^{n-1}$. Even more, these elements of $k(\alpha)$ are linearly independent over k: indeed, any expression of linear dependence $a_{n-1}\alpha^{n-1} + \cdots + a_2\alpha^2 + a_1\alpha + a_0 = 0_k$ induces a polynomial $p(x) = a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0$ of $k[x]$ that has α as a root. By Lemma 6.2.8, we must have that $\mu_\alpha(x)$ divides $p(x)$. Considering that the degree of $p(x)$ is strictly lesser than the degree of $\mu_\alpha(x)$, we must have that $p(x)$ is the zero polynomial so that $a_0 = a_1 = a_2 = \cdots = a_{n-1} = 0_k$. □

Example 6.2.17. Considering that $x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$, it follows by the previous proposition that the simple extension $\mathbb{Q}(\sqrt{2})$ is a $\mathbb{Q}$-vector space of dimension two with a basis of $\{1, \sqrt{2}\}$. Consequently, every element of $\mathbb{Q}(\sqrt{2})$ can be written as $a + b\sqrt{2}$ for some rational numbers a and b. We note that this justifies the description of $\mathbb{Q}(\sqrt{2})$ in Exercise 4.8.18.

Example 6.2.18. We have that $\mathbb{Q}(i)$ is a $\mathbb{Q}$-vector space of dimension two with a basis of $\{1, i\}$ because the minimal polynomial of i over $\mathbb{Q}$ is $x^2 + 1$. We conclude that $\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$.

Example 6.2.19. We have seen that $x^4 - 4x^2 + 1$ is the minimal polynomial of $\sqrt{2 + \sqrt{3}}$ over $\mathbb{Q}$, hence $\mathbb{Q}(\sqrt{2 + \sqrt{3}})$ is a $\mathbb{Q}$-vector space with a basis $\{1, \sqrt{2 + \sqrt{3}}, (\sqrt{2 + \sqrt{3}})^2, (\sqrt{2 + \sqrt{3}})^3\}$.

6.3 Finite Extensions

Consider any extension field F of any field k. We may view F as a k-vector space by virtue of the fact that F is an additive abelian group by definition with the additional property that for any element $\alpha \in F$ and any element $a \in k$, we have that $a\alpha$ lies in F because k can be identified (by the First Isomorphism Theorem for Rngs) with a subfield of F. Even more, we say that F is a **finite extension** of k if F is a finite-dimensional k-vector space, i.e., F admits a finite basis over k.

Every extension field we have encountered thus far in this chapter has been a finite extension. Even more, these extensions have all been algebraic, and every finite extension is algebraic.

Proposition 6.3.1. *Every finite extension of fields is algebraic, i.e., if $k \to F$ is a field extension such that F is a finite-dimensional k-vector space, then every element of F is algebraic over k.*

Proof. Considering that F is a finite-dimensional k-vector space, there exists an integer $n \geq 0$ such that for any element $\alpha \in F$, the powers $1, \alpha, \alpha^2, \dots, \alpha^n$ are linearly dependent over k. Consequently, there exist elements $a_0, a_1, a_2, \dots, a_n \in k$ not all of which are zero such that we obtain a relation of linear dependence $a_n\alpha^n + \cdots + a_2\alpha^2 + a_1\alpha + a_0 = 0_F$ over k. We conclude that α is a root of the nonzero polynomial $a_nx^n + \cdots + a_2x^2 + a_1x + a_0$ in $k[x]$, hence α is algebraic over k. □

Corollary 6.3.2. *Given any algebraic element α of any extension field F of any field k, we have that $k(\alpha)$ is an algebraic extension of k. Explicitly, every element of $k(\alpha)$ is algebraic over k.*

Proof. By Proposition 6.2.16, it follows that $k(\alpha)$ is a finite-dimensional k-vector space. □

Caution: the converse of Proposition 6.3.1 does not hold: indeed, we will see in the next section that the collection of real numbers that are algebraic over $\mathbb{Q}$ forms an algebraic extension of $\mathbb{Q}$ that is an infinite-dimensional $\mathbb{Q}$-vector space (for reasons that are beyond the scope of these notes).

Every simple extension $k(\alpha)$ over k is a finite-dimensional k-vector space of dimension $[k(\alpha) : k]$ equal to the degree of the minimal polynomial of α over k by Proposition 6.2.16, hence every simple extension is itself a finite extension. Conventionally, we adopt the notation $[F : k]$ to denote the k-vector space dimension of any finite extension F over k. We demonstrate next the crucial fact that finiteness of a field extension is transitive and the dimension of a finite extension is multiplicative.

Proposition 6.3.3. *Given any finite extension of a field F over a field k and any finite extension of a field E over F, we have that E is a finite extension over k such that $[E : k] = [E : F][F : k]$.*

Proof. Each of the claims will be achieved simultaneously by demonstrating that if $\alpha_1, \dots, \alpha_m$ form a k-vector space basis of F and $\beta_1, \dots, \beta_n$ form an F-vector space basis of E over F, then their products $\alpha_i\beta_j$ for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$ form a k-vector space basis of E over k. Every element of E can be written as $a_1\beta_1 + \cdots + a_n\beta_n$ for some unique elements $a_1, \dots, a_n \in F$ by assumption that E is a finite-dimensional F-vector space. Considering that F is a finite-dimensional k-vector space, each of the elements a_i of F can be written as $a_i = b_{1i}\alpha_1 + \cdots + b_{mi}\alpha_m$ for some unique elements $b_{1i}, \dots, b_{mi} \in k$. Combined, these observations demonstrate that every element of E is of the form $(b_{11}\alpha_1 + \cdots + b_{m1}\alpha_m)\beta_1 + \cdots + (b_{1n}\alpha_1 + \cdots + b_{mn}\alpha_n)\beta_n$. Expanding the products and rearranging the summands gives a k-linear combination of the products $\alpha_i\beta_j$, hence $\alpha_i\beta_j$ span E as a k-vector space. Even more, they are linearly independent over k: any relation of k-linear dependence $\sum_{j=1}^n (\sum_{i=1}^m a_{ij}\alpha_i)\beta_j = \sum_{j=1}^n \sum_{i=1}^m a_{ij}\alpha_i\beta_j = 0_E$ gives rise to a relation of k-linear dependence $\sum_{i=1}^m a_{ij}\alpha_i$ for each integer $1 \leq j \leq n$. Considering that $\alpha_1, \dots, \alpha_m$ form a basis of F over k, we must have that $a_{ij} = 0_k$ for all integers $1 \leq i \leq m$ and $1 \leq j \leq n$, as desired. □

Corollary 6.3.4. *Given any finite extensions $F_n \supseteq F_{n-1} \supseteq \cdots \supseteq F_2 \supseteq F_1 \supseteq k$, we have that*

$$[F_n : k] = [F_n : F_{n-1}] \cdots [F_2 : F_1][F_1 : k].$$

Proof. We obtain this as a corollary to Proposition 6.3.3 by the Principle of Ordinary Induction: we have that $[F_n : k] = [F_n : F_{n-1}][F_{n-1} : k]$, and the formula holds for $[F_{n-1} : k]$ by induction. □

Corollary 6.3.5. *Given any algebraic elements $\alpha_1, \dots, \alpha_n$ of any extension field F of any field k, we have that $k(\alpha_1, \dots, \alpha_i)$ is an algebraic extension of $k(\alpha_1, \dots, \alpha_{i-1})$ for each integer $1 \leq i \leq n$. Consequently, every finitely generated extension by algebraic elements is an algebraic extension.*

Proof. Given any pair of algebraic elements α and β in any extension field F of k, we must first check that $k(\alpha, \beta)$ is an extension field over $k(\alpha)$. By definition, we have that $k(\alpha, \beta)$ is the smallest extension field of k lying in F that contains k and the elements α and β. Consequently, it follows that $k(\alpha, \beta)$ contains every polynomial in α with coefficients in k, hence $k(\alpha, \beta)$ contains $k(\alpha)$. Even more, because $k(\alpha, \beta)$ contains β, it must contain the smallest extension field of $k(\alpha)$ lying in F that contains $k(\alpha)$ and β, i.e., $k(\alpha, \beta)$ contains $k(\alpha)(\beta)$. Conversely, we note that $k(\alpha)(\beta)$ contains $k(\alpha)$ and β, hence it must contain k, α, and β. Considering that $k(\alpha, \beta)$ is the smallest extension field of k lying in F that contains k and the elements α and β, we conclude that $k(\alpha)(\beta)$ contains $k(\alpha, \beta)$. By the same rationale, it follows that $k(\alpha_1, \dots, \alpha_i) = k(\alpha_1, \dots, \alpha_{i-1})(\alpha_i)$ for each integer $1 \leq i \leq n$, hence every finitely generated extension by algebraic elements induces a **tower** of simple extensions by algebraic elements. Each of these simple extensions is finite by Proposition 6.2.16, hence we find that $k(\alpha_1, \dots, \alpha_n)$ is a finite extension of k; it must be algebraic by Proposition 6.3.1. □

We have thus far in this chapter only explicitly dealt with simple extensions, so it is natural to seek to determine the structure of any algebraic extension $k(\alpha_1, \dots, \alpha_n)$ over k. One immediate idea is to view $k(\alpha_1, \dots, \alpha_n)$ as a simple extension $k(\alpha_1, \dots, \alpha_{n-1})(\alpha_n)$; then, it suffices to determine the structure of $k(\alpha_1)$ over k, the structure of $k(\alpha_1)(\alpha_2)$ over $k(\alpha_1)$, etc. Combined with the following proposition, this strategy can be used to great effect to simplify our study of finite field extensions.

Proposition 6.3.6. *Consider any algebraic element α of any extension field F of any field k. Given any element β of the simple extension $k(\alpha)$, the minimal polynomial $\mu_\beta(x)$ of β in $k[x]$, the degree of $\mu_\beta(x)$ in $k[x]$ divides the degree of the minimal polynomial $\mu_\alpha(x)$ in $k[x]$.*

Proof. Given any element β of the simple extension $k(\alpha)$, we must have that β is a polynomial in α by Proposition 6.2.16. Even more, it follows that β is algebraic over k by Corollary 6.3.2, hence the simple extension $k(\beta)$ over k is finite by Proposition 6.2.16. We may therefore consider the minimal polynomial $\mu_\beta(x)$ of β over k. Every element of $k(\beta)$ is a polynomial in β, and β is a polynomial in α, hence every element of $k(\beta)$ is a polynomial in α, and it follows that $k(\alpha)$ is an extension field of $k(\beta)$. Considering that $k(\alpha)$ is a finite extension of k, it must be the case that $k(\alpha)$ is a finite extension of $k(\beta)$ because the minimal polynomial $\mu_\alpha(x)$ of α over k is divisible by the minimal polynomial of α over $k(\beta)$. We conclude that $k \to k(\beta) \subseteq k(\alpha)$ is a tower of finite extensions, hence Proposition 6.3.3 yields that $\deg(\mu_\alpha) = [k(\alpha) : k] = [k(\alpha) : k(\beta)][k(\beta) : k] = [k(\alpha) : k(\beta)] \deg(\mu_\beta)$. □

Example 6.3.7. Consider the finitely generated extension $\mathbb{Q}(\sqrt{2}, i)$ of $\mathbb{Q}$. By Example 6.2.13, we have that $\mathbb{Q}(\sqrt{2})$ is a simple extension of degree two over $\mathbb{Q}$. Considering that $x^2 + 1$ is a monic polynomial that does not admit a root over $\mathbb{Q}(\sqrt{2})$ because i is not a real number, it follows that $x^2 + 1$ is the minimal polynomial of i over $\mathbb{Q}(\sqrt{2})$. We conclude by Proposition 6.3.3 that $\mathbb{Q}(\sqrt{2}, i)$ is a finite algebraic extension of $\mathbb{Q}$ of degree $[\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = (2)(2) = 4$.

Example 6.3.8. By Example 6.2.15, the simple extension $\mathbb{Q}(\sqrt{2 + \sqrt{3}})$ of $\mathbb{Q}$ has degree four over $\mathbb{Q}$. We will establish this fact by providing an alternative to the previous proof. Considering that $\sqrt{3} = (\sqrt{2 + \sqrt{3}})^2 - 2$, it follows that $\mathbb{Q}(\sqrt{2 + \sqrt{3}})$ is an extension field of $\mathbb{Q}(\sqrt{3})$; it is a finite extension of $\mathbb{Q}(\sqrt{3})$ because $\mathbb{Q}(\sqrt{2 + \sqrt{3}})$ and $\mathbb{Q}(\sqrt{3})$ are both finite extensions of $\mathbb{Q}$. We claim that $x^2 - (2 + \sqrt{3})$ is the minimal polynomial of $\sqrt{2 + \sqrt{3}}$ over $\mathbb{Q}(\sqrt{3})$. By the Factor Theorem, it suffices to prove that $x^2 - (2 + \sqrt{3})$ admits no roots in $\mathbb{Q}(\sqrt{3})$. On the contrary, suppose that

$a + b\sqrt{3}$ satisfies that $(a^2 + 3b^2) + 2ab\sqrt{3} = (a + b\sqrt{3})^2 = 2 + \sqrt{3}$ for some rational numbers a and b. By rearranging this expression, we could write $\sqrt{3}$ as a rational number — a contradiction.

$$\sqrt{3} = \frac{a^2 + 3b^2 - 2}{1 - 2ab}$$

We conclude therefore that $[\mathbb{Q}(\sqrt{2+\sqrt{3}}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2+\sqrt{3}}) : \mathbb{Q}(\sqrt{3})][\mathbb{Q}(\sqrt{3}) : \mathbb{Q}] = (2)(2) = 4.$

Example 6.3.9. We note that the method of the previous example can be applied more generally to determine the degree of finitely generated extensions by algebraic elements. Consider the finite algebraic extension $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ of $\mathbb{Q}$. By Exercise 6.4.7, we have that $\mathbb{Q}(\sqrt{3})$ has degree two over $\mathbb{Q}$. Consequently, we may turn our attention to the extension field $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ of $\mathbb{Q}(\sqrt{3})$. We claim that $x^2 - 5$ is the minimal polynomial of $\sqrt{5}$ over $\mathbb{Q}(\sqrt{3})$. Like before, we may assume on the contrary that there exist rational numbers a and b such that $(a^2 + 3b^2) + 2ab\sqrt{3} = (a + b\sqrt{3})^2 = 5$, and in the same way as the previous example, we arrive at a contradiction that $\sqrt{3}$ is a rational number.

$$\sqrt{3} = \frac{5 - a^2 - 3b^2}{2ab}$$

We conclude by the Factor Theorem that $x^2 - 5$ is irreducible over $\mathbb{Q}(\sqrt{3})$, hence we have that $[\mathbb{Q}(\sqrt{3}, \sqrt{5}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{3}, \sqrt{5}) : \mathbb{Q}(\sqrt{3})][\mathbb{Q}(\sqrt{3}) : \mathbb{Q}] = (2)(2) = 4$ by Proposition 6.3.3.

Consider the element $\sqrt{3}+\sqrt{5}$ of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$. We note that $\mathbb{Q}(\sqrt{3}+\sqrt{5})$ lies in $\mathbb{Q}(\sqrt{3}, \sqrt{5})$, hence the degree of $\mathbb{Q}(\sqrt{3} + \sqrt{5})$ divides the degree of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ by Proposition 6.3.6. Considering that $\mathbb{Q}(\sqrt{3}, \sqrt{5})$ is a finite extension of $\mathbb{Q}(\sqrt{3} + \sqrt{5})$ by the proof of the aforementioned proposition, it follows from general considerations in linear algebra that $\mathbb{Q}(\sqrt{3} + \sqrt{5}) = \mathbb{Q}(\sqrt{3}, \sqrt{5})$ if and only if $[\mathbb{Q}(\sqrt{3} + \sqrt{5}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{3}, \sqrt{5}) : \mathbb{Q}] = 4$. We know by Corollary 6.3.1 that $\sqrt{3} + \sqrt{5}$ is algebraic over $\mathbb{Q}$, hence we may find a candidate for the minimal polynomial of $\sqrt{3} + \sqrt{5}$.

$$\begin{aligned} \alpha &= \sqrt{3} + \sqrt{5} \\ \alpha^2 &= 8 + 2\sqrt{15} \\ \alpha^2 - 8 &= 2\sqrt{15} \\ (\alpha^2 - 8)^2 &= 60 \\ \alpha^4 - 16\alpha^2 + 4 &= 0 \end{aligned}$$

Consequently, we have found a monic polynomial $x^4 - 16x^2 + 4$ in $\mathbb{Q}[x]$ for which $\sqrt{3}+\sqrt{5}$ is a root. By the Rational Roots Theorem, the only possible rational roots of $x^4 - 16x^2 + 4$ are ±1, ±2, and ±4. Check that none of these is a root, hence $x^4 - 16x^2 + 4$ does not admit any linear factors by the Factor Theorem. Even more, by Gauss's Lemma, it suffices to prove that $x^4 - 16x^2 + 4$ does not factor as a product of quadratics $x^4 - 16x^2 + 4 = (x^2 + ax + b)(x^2 + cx + d)$.

$$\begin{aligned} a + c &= 0 & ad + bc &= 0 \\ ac + b + d &= -16 & bd &= 4 \end{aligned}$$

Considering that $bd = 4$, it follows that $b = d = \pm 2$ so that $-16 - 2b = -16 - b - d = ac = -a^2$ or $a^2 = 16 + 2b$ by the first and second equations in the left-hand column. Given that $b = d = 2$, it follows that $a^2 = 20$ — a contradiction to the result of Exercise 6.4.7. Conversely, if $b = d = -2$, then $a^2 = 12$ — a contradiction. We conclude therefore that $x^4 - 16x^2 + 4$ is irreducible over $\mathbb{Q}$, hence we have that $[\mathbb{Q}(\sqrt{3} + \sqrt{5}) : \mathbb{Q}] = 4$ so that $\mathbb{Q}(\sqrt{3}, \sqrt{5}) = \mathbb{Q}(\sqrt{3} + \sqrt{5})$ is simple.

We conclude this section with an important result that completely characterizes finite extensions.

Theorem 6.3.10. *Given any extension field F of any field k, the following conditions are equivalent.*

(i.) *We have that F is a finite extension of k, i.e., F is finite-dimensional as a k-vector space.*

(ii.) *We have that F is a finitely generated algebraic extension of k, i.e., there exist $\alpha_1, \dots, \alpha_n \in F$ such that $F = k(\alpha_1, \dots, \alpha_n)$ and α_i is algebraic over k for each integer $1 \leq i \leq n$.*

(iii.) *We have that F is obtained from a finite sequence of simple algebraic extensions over k, i.e., there exist elements $\alpha_1, \dots, \alpha_n$ such that $F = k(\alpha_1, \dots, \alpha_n)$ and for each integer $1 \leq i \leq n$, we have that $k(\alpha_1, \dots, \alpha_i) = k(\alpha_1, \dots, \alpha_{i-1})(\alpha_i)$ is an algebraic extension of $k(\alpha_1, \dots, \alpha_{i-1})$.*

Proof. Essentially, the proof of this fact follows from a careful recollection of the observations of this section. We note that if F is a finite extension of k, then we may find a basis $\alpha_1, \dots, \alpha_n \in F$ of F as a k-vector space. Every element of F can be written as a k-linear combination of the elements $\alpha_1, \dots, \alpha_n$, hence F is contained in $k(\alpha_1, \dots, \alpha_n)$. Considering that $k(\alpha_1, \dots, \alpha_n)$ is by definition the smallest extension field of k lying in F that contains k and the elements $\alpha_1, \dots, \alpha_n$, we conclude that $F = k(\alpha_1, \dots, \alpha_n)$. By Proposition 6.3.1, we must have that F is algebraic over k.

We will assume next that F admits elements $\alpha_1, \dots, \alpha_n$ such that $F = k(\alpha_1, \dots, \alpha_n)$ and α_i is algebraic over k for each integer $1 \leq i \leq n$. Corollary 6.3.5 ensures the desired result.

Last, we will assume that there exist elements $\alpha_1, \dots, \alpha_n$ such that $F = k(\alpha_1, \dots, \alpha_n)$ and for each integer $1 \leq i \leq n$, we have that $k(\alpha_1, \dots, \alpha_i)$ is an algebraic extension of $k(\alpha_1, \dots, \alpha_{i-1})$. Each of the simple extensions $k(\alpha_1, \dots, \alpha_i)$ over $k(\alpha_1, \dots, \alpha_{i-1})$ is algebraic by assumption, hence each one is finite by Proposition 6.2.16; we conclude that F is finite over k by Corollary 6.3.4. □

Corollary 6.3.11. *Given any algebraic extension of any field F over any field k and any algebraic extension of any field E over F, we have that E is algebraic extension over k.*

Proof. Given any element $\alpha \in E$, we must demonstrate that α is algebraic over k. By hypothesis that E is algebraic over F, there exist elements $a_0, a_1, \dots, a_n \in F$ not all of which are zero such that $a_n \alpha^n + \cdots + a_1 \alpha + a_0 = 0_F$. Considering that coefficients $a_0, a_1, \dots, a_n$ induce a finitely generated extension $k(a_0, a_1, \dots, a_n)$ over k, we find that α is algebraic over $k(a_0, a_1, \dots, a_n)$ so that $k(a_0, a_1, \dots, a_n, \alpha)$ is a simple algebraic extension over $k(a_0, a_1, \dots, a_n)$ by Corollary 6.3.2. Even more, by assumption, each of the elements $a_0, a_1, \dots, a_n$ is algebraic over k, hence the extension field $k(a_0, a_1, \dots, a_n, \alpha)$ is obtained from a finite sequence of simple algebraic extensions over k. We conclude by Theorem 6.3.10 that $k(a_0, a_1, \dots, a_n, \alpha)$ is a finite extension of k, hence there exists an integer $m \geq 0$ such that $1, \alpha, \alpha^2, \dots, \alpha^m$ are linearly dependent over k. We obtain from here a nonzero polynomial $p(x) = a_m x^m + \cdots + a_2 x^2 + a_1 x + a_0$ in $k[x]$ such that $p(\alpha) = 0_k$. □

6.4 Chapter 6 Exercises

6.4.1 Roots of Polynomials and Field Extensions

Exercise 6.4.1. Prove or disprove that each of the following commutative unital rings is a field.

(a.) $\mathbb{Q}[x]/(x^2-4)$

(b.) $\mathbb{Q}[x]/(x^2+3)$

(c.) $\mathbb{Q}[x]/(x^3+x+1)$

(d.) $\mathbb{Q}[x]/(x^3-2x^2+2x-1)$

(e.) $(\mathbb{Z}/2\mathbb{Z})[x]/(x^3+1)$

(f.) $(\mathbb{Z}/2\mathbb{Z})[x]/(x^3+x+1)$

(g.) $(\mathbb{Z}/3\mathbb{Z})[x]/(x^3+2x+1)$

(h.) $(\mathbb{Z}/5\mathbb{Z})[x]/(x^5+1)$

Exercise 6.4.2. Consider the monic polynomial x^2-3 in $\mathbb{Q}[x]$.

(a.) Use the Polynomial Division Algorithm to find polynomials $p(x)$ and $q(x)$ such that

$$(x^2-3)p(x)+(2x+3)q(x)=1.$$

(b.) Prove that $F=\mathbb{Q}[x]/(x^2-3)$ is a field that contains $\mathbb{Q}$ and a root α of x^2-3.

(c.) Express the element $\alpha^4-3\alpha^3+\alpha^2-\alpha$ as a polynomial in α of degree at most one.

(d.) Express the element $(2\alpha+3)^{-1}$ as a polynomial in α of degree at most one.

Exercise 6.4.3. Consider the monic polynomial x^3+x+1 in $(\mathbb{Z}/2\mathbb{Z})[x]$.

(a.) Use the Polynomial Division Algorithm to find polynomials $p(x)$ and $q(x)$ such that

$$(x^3+x+1)p(x)+(x^2+1)q(x)\equiv 1 \pmod 2.$$

(b.) Prove that $F=(\mathbb{Z}/2\mathbb{Z})[x]/(x^3+x+1)$ is a field that contains k and a root α of x^3+x+1.

(c.) Express each of the eight elements of F as a polynomial in α of degree at most two.

(d.) Express the element $\alpha^5+\alpha^4+\alpha^3+\alpha^2+\alpha+1$ as a polynomial in α of degree at most two.

(e.) Express the element $(\alpha^2+1)^{-1}$ as a polynomial in α of degree at most two.

6.4.2 Simple Extensions

Exercise 6.4.4. Consider any field F. Prove that if $\{F_i\}_{i\in I}$ is any nonempty collection of fields indexed by I such that $F_i\subseteq F$, then $\cap_{i\in I}F_i$ is a field contained in F.

Exercise 6.4.5. Prove that each of the following complex numbers is algebraic over $\mathbb{Q}$.

(a.) $\sqrt[4]{4}$

(b.) $1+\sqrt[3]{3}$

(c.) $-1+i$

(d.) $\sqrt{i-\sqrt{2}}$

(e.) $\cos\left(\frac{\pi}{4}\right)+i\sin\left(\frac{\pi}{4}\right)$

(f.) $\cos\left(\frac{2\pi}{5}\right)$

Exercise 6.4.6. Compute the minimal polynomial of the following complex numbers over $\mathbb{Q}$.

(a.) $\sqrt{3}$

(b.) $1 + \sqrt[3]{3}$

(c.) $\sqrt{3} + \sqrt{5}$

(d.) $\cos\left(\frac{\pi}{4}\right) + i \sin\left(\frac{\pi}{4}\right)$

(e.) $i\sqrt[6]{2}$

(f.) $\cos\left(\frac{2\pi}{5}\right)$

Exercise 6.4.7. Consider any positive integer a. We say that a is a **perfect square** if there exists a positive integer b such that $a = b^2$.

(a.) Prove that if a is a perfect square, then the positive integer b such that $a = b^2$ is uniquely determined by a. Conclude that we may write in this case that $b = \sqrt{a}$.

(b.) Prove that if a is a perfect square, then $x^2 - a = (x - \sqrt{a})(x + \sqrt{a})$ is a $\mathbb{Q}$-factorization.

(c.) Prove that if a is not a perfect square, then $\sqrt{a}$ is not a rational number.

(**Hint:** On the contrary, if $\sqrt{a}$ were a rational number, then it must be an integer; otherwise, we could find relatively prime positive integers p and q such that $\sqrt{a} = \frac{p}{q}$ and $p^2 = aq^2$.)

(d.) Prove that if a is not a perfect square, then the minimal polynomial of $\sqrt{a}$ over $\mathbb{Q}$ is $x^2 - a$.

(e.) Prove that if a is not a perfect square, then $\mathbb{Q}(\sqrt{a})$ is a finite algebraic extension of $\mathbb{Q}$.

Exercise 6.4.8. Prove that $\sqrt{\pi}$ is algebraic over $\mathbb{Q}(\pi))$; then, find the degree of $\mathbb{Q}(\sqrt{\pi})$ over $\mathbb{Q}(\pi)$.

Exercise 6.4.9. Prove that π is algebraic over $\mathbb{Q}(\pi^4)$; then, find the degree of $\mathbb{Q}(\pi)$ over $\mathbb{Q}(\pi^4)$.

Exercise 6.4.10. Prove that $\mathbb{C}$ is an algebraic extension of $\mathbb{R}$.

6.4.3 Finite Extensions

Exercise 6.4.11. Consider a finite field k of prime characteristic p.

(a.) Prove that if F is any extension field of k, then F is a field of prime characteristic p.

(b.) Prove that if F is any extension field of k such that $|F|$ is finite, then F is algebraic over k.

References

[Bag19] J. Bagaria. *Zermelo-Fraenkel Set Theory*. 2019. URL: https://plato.stanford.edu/entries/set-theory/ZF.html.

[CKK22] P. Corn, S. Kallasa, and J. Khim. *Axiom of Choice*. 2022. URL: https://brilliant.org/wiki/axiom-of-choice/.

[Con22] K. Conrad. *The Gaussian Integers*. 2022. URL: https://kconrad.math.uconn.edu/blurbs/ugradnumthy/Zinotes.pdf.

[Cor+22] P. Corn et al. *Binomial Theorem*. 2022. URL: https://brilliant.org/wiki/binomial-theorem-n-choose-k/.

[DF04] D.S. Dummit and R.M. Foote. *Abstract Algebra*. 3rd ed. John Wiley & Sons, Inc., 2004.

[DW00] J.P. D'Angelo and D.B. West. *Mathematical Thinking: Problem Solving and Proofs*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[Gon22] A. Gonzalez. *Fundamental Counting Principle*. 2022. URL: https://brilliant.org/wiki/fundamental-counting-principle/.

[Hen19] J.N. Henry. *Groups Satisfying the Converse to Lagrange's Theorem*. 2019. URL: https://bearworks.missouristate.edu/cgi/viewcontent.cgi?article=4484&context=theses.

[Hun13] T.W. Hungerford. *Abstract Algebra: an Introduction*. 3rd ed. Brooks / Cole, Cengage Learning, 2013.

[JB21] T.W. Judson and R.A. Beezer. *Abstract Algebra: Theory and Applications*. 2021.

[Mag11] A. Magidin. *Why are two permutations conjugate iff they have the same cycle structure?* 2011. URL: https://math.stackexchange.com/a/48137/390180.

[Wil15] R.A. Wilson. *An example of a PID which is not a Euclidean domain*. 2015. URL: http://www.maths.qmul.ac.uk/~raw/MTH5100/PIDnotED.pdf.