Introduction to Abstract Algebra

Dylan C. Beck

Acknowledgements

Primarily, the first two and last chapters of this document were written in the Fall 2022 and Spring 2023 semesters at Baker University with mild revisions and reorganization efforts unfolding during the Spring and Summer 2024. Bearing this in mind, I express my gratitude toward Vance Gaffar for his helpful comments and interesting solutions to some of the exercises and my former students at Baker University — especially those whose thoughtful work and insightful suggestions enhanced this document, including Sabrena Hayles, Kyler Kosanke, Allie Renfro, and April Thomas.

Elsewhere, the bulk of the material on ring theory was written independently from August 2018 to January 2022 at the request of the Algebra Qualifying Exam Study Group at the University of Kansas. Exercises in these chapters may have been adapted from the qualifying exams in algebra at the University of Kansas. I sincerely appreciate these former graduate students for their motivation and participation in those many 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 those sessions and their peripheral contributions to these notes.

Contents

Chapter 0

Essential Topics in Modern Mathematics

0.1 Sets, Relations, and Functions

Contemporary mathematics is communicated rigorously using sets, symbols, functions, relations, certain computational tools, and proofs; 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 language of set theory and the calculus of logic. Each of these branches of mathematics enjoys contemporary ubiquity and significance that make them active areas 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 "naïve" or Zermelo-Fraenkel set theory with the Axiom of Choice.

0.1.1 Sets and Set Operations

We define a **set** X as a collection of "similar" objects, e.g., the names of the 2023-2024 Golden State Warriors, the menu items at the cafeteria this evening, or any collection of real numbers. We refer to an arbitrary object x of a set X as an **element** (or **member**) of X. Concretely, 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 y does not lie in X, then we write $y \notin X$ to signify this fact symbolically.

Order and repetition are irrelevant notions when considering the elements of a set. Explicitly, the set W consisting only of the real numbers 1 and -1 can be realized as $W = \{-1, 1\}$ or $W = \{1, -1\}$ or $W = \{-1, 1, -1, 1\}$. Out of desire for simplicity, we will list only the distinct elements of a set. Consequently, if there are "few enough" distinct elements of a set X, we can explicitly write down X using braces. Observe that $X = \{1, 2, 3, 4, 5, 6\}$ is the unique 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 list the elements of infinite sets. We will henceforth fix the following notation for the sets of **natural numbers** $\mathbb{Z}_{\geq 0} = \{n \mid n \text{ is a non-negative integer}\}$,

integers $\mathbb{Z} = \{n \mid n \text{ is an integer}\}$, and **rational numbers** $\mathbb{Q} = \left\{\frac{a}{b} \mid a \text{ and } b \text{ are integers, } b \neq 0\right\}$. Using the rational numbers, one can construct the **real numbers** $\mathbb{R} = \{x \mid x \text{ is a real number}\}$.

Example 0.1.1. Crucially, we must be able to convert between set-builder notation and explicit ("curly braces") notation. Given the set $S = \{n \mid n \text{ is an integer and } |n| \leq 3\}$, we find that n is an integer such that $-3 \leq n \leq 3$, hence we conclude that $S = \{-3, -2, -1, 0, 1, 2, 3\}$.

Example 0.1.2. Consider the finite set $T = \{-7, -5, -3, \ldots, 11, 13\}$. We use an ellipsis in this case to signify that the pattern repeats up to the integer 11. Each of the elements -7, -5, -3, 11, and 13 of T is an odd integer, hence the set T consists of all odd integers t such that $-7 \leq t \leq 13$. We may likewise use set-builder notation to express that $T = \{t \mid t \text{ is an odd integer and } -7 \leq t \leq 13\}$. We could have perhaps more easily described this set as $T = \{t \in \mathbb{Z} \mid t \text{ is odd and } -7 \leq t \leq 13\}$.

Example 0.1.3. Consider the infinite set $U = \{x^2 \mid x \in \mathbb{Z}_{\geq 0}\}$. Every element of U is the square of some non-negative integers, hence we have that $U = \{0, 1, 4, 9, 16, \ldots\}$. Once again, we use an ellipsis to signify that the pattern continues; however, in this case, it does so indefinitely.

One important consideration in the arithmetic of sets is the number of elements that belong to the set. One can readily verify that the set $X = \{1, 2, 3, 4, 5, 6\}$ consists of six elements, but the set $Y = \{1, 2, 3, 4, 5\}$ 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 $|Y| = 5$. Cardinality can be defined even for infinite sets, but additional care must be taken in this case, so we will not bother.

Example 0.1.4. Consider the following four sets written in set-builder notation.

$$A = \{n \in \mathbb{Z}_{\geq 0} \mid n \leq 9\} \qquad C = \{x \in \mathbb{R} \mid x^2 - 2 = 0\}$$
$$B = \{q \in \mathbb{Q}_{\geq 0} \mid q \leq 9\} \qquad D = \{q \in \mathbb{Q} \mid q^2 - 2 = 0\}$$

We will illustrate some of the concepts of this section by answering the following.

(a.) (Explicit Notation) By definition, we have the set membership $n \in A$ if and only if n is a non-negative integer such that $n \geq 9$. Consequently, we conclude that $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

(b.) (Set Membership) By definition, we have that $q \in A$ if and only if q is a non-negative rational number such that $q \leq 9$. We note that there are infinitely many elements of B that do not lie in A. Concretely, any rational number $\frac{1}{2^n}$ for some integer $n \geq 1$ lies in B but not in A.

(c.) (Explicit Notation) By the Square Root Property, we have that $x^2 - 2 = 0$ if and only if $x^2 = 2$ if and only if $x = \pm\sqrt{2}$. Consequently, the elements of C are given by $C = \{-\sqrt{2}, \sqrt{2}\}$.

(d.) (Set Membership) By part (c.), there are no elements in D because neither $\pm\sqrt{2}$ is rational.

(e.) (Cardinality) By parts (a.), (c.), and (d.), we have that $|A| = 10$, $|C| = 2$, and $|D| = 0$.

Commonly in mathematics, in order to understand an object, it is beneficial to study its subobjects. Consequently, for a given set, we may seek to determine all sets that can be constructed with the elements of the specified set. Concretely, it is straightforward to verify that every element of the set $Y = \{1, 2, 3, 4, 5\}$ is also an element of the set $X = \{0, 1, 2, 3, 4, 5, 6\}$, but there are elements

of X that do not lie in Y: namely, we have that $0, 6 \in X$ and yet $0, 6 \notin Y$. We express this by saying that Y is a **proper subset** of X: the modifier "proper" indicates that X and Y are not the same set (since they do not have the same members). Put into symbols, we write $Y \subset X$ if and only if

(a.) every element of Y is an element of X and

(b.) there exists an element of X that is not contained in Y.

We read $Y \subset X$ as "Y is contained in but does not equal X." We may also say that Y is "included in" X or that Y "lies in" X. One other way to indicate that Y is a (proper) subset of X is to say that X is a (proper) **superset** of Y, in which case we write $X \supseteq Y$ (or $X \supset Y$ if the containment is proper). Observe that if we could step through the paper and look at the superset containment $X \supseteq Y$ from the other side, we would simply see that $Y \subseteq X$; however, it is sometimes preferable to use this notation to emphasize that X is the object of our concern rather than Y.

Containment of subsets is **transitive** in the sense that if $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$: indeed, every element $x \in X$ is an element of Y so that $x \in Y$; moreover, every element of Y is an element of Z so that $x \in Z$ ultimately holds. Compare this with inequalities of real numbers.

Proposition 0.1.5 (Set Containment Is Transitive)**.** *Given any sets X, Y, and Z such that $X \subseteq Y$ and $Y \subseteq Z$, we have that $X \subseteq Z$. Put another way, set containment is transitive.*

Example 0.1.6. Consider the sets $A = \{-1, 1\}$, $B = \{-1, 0, 1\}$, and $C = \{-2, -1, 1, 2\}$. Observe that the strict inclusions $A \subset B$ and $A \subset C$ hold, but neither $B \subseteq C$ or $C \subseteq B$ holds.

Example 0.1.7. Every non-negative integer is an integer; every integer is a rational number; and every rational number is a real number. Consequently, we have the subset containments

$$\mathbb{Z}_{\geq 0} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Each of these containments is strict because -1 is an integer that is not non-negative; $\frac{1}{2}$ is a rational number that is not an integer; and $\sqrt{2}$ is a real number that is not a rational number. We will from now on refer to the collection of real numbers that are not rational as **irrational numbers**.

Equality of sets is determined by simultaneous subset and superset containments. Explicitly, a pair of sets X and Y are **equal** if and only if it holds that $X \subseteq Y$ and $X \supseteq Y$. Put another way, the sets X and Y are equal if and only if X and Y possess exactly the same elements: indeed, for any element $x \in X$, we have that $x \in Y$ because $X \subseteq Y$, and for any element $y \in Y$, we have that $y \in X$ because $X \supseteq Y$. Crucially, one can demonstrate that two finite sets are equal if and only if they have the same cardinality and one of the sets is a subset of the other (cf. Proposition 0.1.86).

Often, we will view a set X as a subset of a specified **universal set** (or **ambient set**). Explicitly, in each of the examples from the previous two sections, we typically dealt with integers, hence we could have taken the ambient set as any of $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$. Context will usually make this clear.

Like with the usual arithmetic of real numbers, we may define mathematical operations on sets. We will explore in this section typical set operations that allow us to combine, compare, and take differences of sets. Consider the sets $X = \{0, 1, 2, 3, 4, 5, 6\}$ and $Y = \{1, 2, 3, 4, 5\}$ of the previous section. We introduce the **relative complement** of Y with respect to X to formalize our previous observation that 0 and 6 belong to X but do not belong to Y. By definition, the relative complement

of Y with respect to X is the set consisting of all elements of X that are not elements of Y. We use the symbolic notation $X \setminus Y$ to denote the relative complement of Y with respect to X so that

$$X \setminus Y = \{w \mid w \in X \text{ and } w \notin Y\}.$$

We note that $X \setminus Y = \{0, 6\}$ 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 of all elements that belong to both X and Y. Going back to our running example of $X = \{0, 1, 2, 3, 4, 5, 6\}$ and $Y = \{1, 2, 3, 4, 5\}$, we have that $X \cap Y = \{1, 2, 3, 4, 5\}$. Order of the sets does not matter with respect to the set intersection. Explicitly, for any sets X and Y, we have that $X \cap Y = Y \cap X$ because every element that lies in both X and Y lies in both Y and X. Consequently, set intersection is a **commutative** (or **order-invariant**) operation.

Exercise 0.1.8. Construct a **Venn diagram** to visualize the sets X, Y, $X \setminus Y$, and $X \cap Y$.

Example 0.1.9. Consider the sets $A = \{1, 2, 3, \ldots, 10\}$, $B = \{1, 4, 9\}$, and $C = \{1, 3, 5, 7, 9\}$. We have that $A \setminus B = \{2, 3, 5, 6, 7, 8, 10\}$, $A \setminus C = \{2, 4, 6, 8, 10\}$, $B \setminus C = \{4\}$, and $C \setminus B = \{3, 4, 7\}$. Each of the sets A and B is a proper subset of A, and we have that $A \cap B = B$ and $A \cap C = C$.

Crucially, if $B \subseteq A$, then $A \cap B = B$: indeed, every element of B is an element of A, hence we have that $A \cap B \supseteq B$. Conversely, every element of $A \cap B$ is an element of B so that $A \cap B \subseteq B$.

Proposition 0.1.10 (Going-Down Property of Set Intersection)**.** *Given any sets X and Y such that $X \subseteq Y$, we have that $X \cap Y = X$. Conversely, if $X \cap Y = X$, then $X \subseteq Y$.*

Proof. By the paragraph preceding the statement of the proposition, the first assertion holds. Conversely, if $X \cap Y = X$, then for every element $x \in X$, we have that $x \in X \cap Y$ so that $x \in Y$. □

Example 0.1.11. Consider the sets $D = \{1, 3, 5, 7\}$, $E = \{1, 4, 7, 10\}$, and $F = \{2, 5, 8, 11\}$. We have that $D \setminus E = \{3, 5\}$, $D \setminus F = \{1, 3, 7\}$, $E \setminus D = \{4, 10\}$, and $F \setminus D = \{2, 8, 11\}$. Even more, we have that $D \cap E = \{1, 7\}$, $D \cap F = \{5\}$, and E and F have no elements in common.

Consider the finite sets $V = \{1, 2, 3\}$ and $W = \{4, 5, 6\}$. Considering that none of the elements of V belongs to W and none of the elements of W belongs to V, the intersection of V and W does not possess any elements; it is empty! Conventionally, this is called the **empty set**; it is denoted by $\varnothing$. Put another way, our observations thus far in this paragraph can be stated as $V \cap W = \varnothing$. 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 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 (or both) — that is, we have that

$$X \cup Y = \{w \mid w \in X \text{ or } w \in Y\}.$$

Like the set intersection, the set union is also a commutative (or order-invariant) operation.

Example 0.1.12. Consider the sets A, B, and C of Example 0.1.9. Each of the elements of B and C are elements of A, hence we have that $A \cup B = A$, $A \cup C = A$, and $B \cup C = \{1, 3, 4, 5, 7, 9\}$.

Crucially, if $B \subseteq A$, then $A \cup B = A$: indeed, every element of A is an element of $A \cup B$, hence we have that $A \cup B \supseteq A$. Conversely, every element of $A \cup B$ is an element of A and $A \cup B \subseteq A$.

Proposition 0.1.13 (Going-Up Property of Set Union)**.** *Given any sets X and Y such that $X \subseteq Y$, we have that $X \cup Y = Y$. Conversely, if $X \cup Y = Y$, then $X \subseteq Y$.*

Proof. By the paragraph preceding the statement of the proposition, the first assertion holds. Conversely, if $X \cup Y = Y$, then for every element $x \in X$, we have that $x \in X \cup Y$ so that $x \in Y$. □

Example 0.1.14. Consider the sets D, E, and F of Example 0.1.11. Excluding any overlap, we have that $D \cup E = \{1, 3, 4, 5, 7, 10\}$, $D \cup F = \{1, 2, 3, 5, 7, 8, 11\}$, and $E \cup F = \{1, 2, 4, 5, 7, 8, 10, 11\}$.

Every set X gives rise to a unique set consisting of all possible subsets of X. Explicitly, for any set X, the **power set** $P(X)$ is the set of all subsets of X — including the empty set.

Example 0.1.15. Consider the set $U = \{-1, 0, 1\}$. Counting the empty set by convention, there are exactly $2^3 = 8$ subsets of U. Each subset is determined by including or excluding each element of U. Label the elements of U in order; then, construct an ordered triple consisting of checks ✓ and crosses × corresponding respectively to whether an element of U is included or excluded as follows.

× × ×: $\varnothing$	✓ ✓ ×: $\{-1, 0\}$
✓ × ×: $\{-1\}$	✓ × ✓: $\{-1, 1\}$
× ✓ ×: $\{0\}$	× ✓ ✓: $\{0, 1\}$
× × ✓: $\{1\}$	✓ ✓ ✓: $\{-1, 0, 1\}$

Consequently, we have that $P(U) = \{\varnothing, \{-1\}, \{0\}, \{1\}, \{-1, 0\}, \{-1, 1\}, \{0, 1\}, \{-1, 0, 1\}\}$.

Crucially, if U is a finite set, then $|P(U)| = 2^{|U|}$: indeed, every subset of U is uniquely determined by its elements, and each element of U can either be included or excluded from a given subset.

Proposition 0.1.16 (Cardinality of the Power Set of a Finite Set)**.** *Given any finite set X, the power set of X has cardinality $2^{|X|}$. Put another way, we have that $|P(X)| = 2^{|X|}$ if $|X|$ is finite.*

Example 0.1.17. Consider the finite sets $\varnothing$, $X = \{\varnothing\}$, and $Y = \{\varnothing, \{\varnothing\}\} = \{\varnothing, X\}$. By the previous proposition, it follows that $|P(\varnothing)| = 2^0 = 1$, $|P(X)| = 2^1 = 2$, and $|P(Y)| = 2^2 = 4$. Explicitly, we have that $P(\varnothing) = \{\varnothing\} = X$, $P(X) = \{\varnothing, \{\varnothing\}\} = Y$, and $P(Y) = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\}$.

Often, we wish to deal with objects from a collection of more than two sets. Considering that the union and intersection of a pair of sets is itself a set, we can apply recursion. We achieve this by first creating an **index set** I that contains all of the labels for the sets in question. Explicitly, if we are working with three distinct sets X_1, X_2, and X_3, then our index set can be taken as $I = \{1, 2, 3\}$ to indicate the first, second, and third set. Bearing in mind that the order of the sets in a set union or intersection does not matter, we do not need to worry about the order of the labels of our sets. Even more, we are ooften at liberty to label our sets in an order-appropriate manner. We have that

$$X_1 \cap X_2 \cap X_3 = \{x \mid x \in X_1 \text{ and } x \in X_2 \text{ and } x \in X_3\} \text{ and}$$
$$X_1 \cup X_2 \cup X_3 = \{x \mid x \in X_1 \text{ or } x \in X_2 \text{ or } x \in X_3\}.$$

Consequently, in order for an element to lie in the intersection $X_1 \cap X_2 \cap X_3$ of three sets, it must lie in each of the three sets; on the other hand, an element belongs to the union $X_1 \cup X_2 \cup X_3$ if and only if it belongs to at least one of the three sets. Generally, we may define the set union and intersection of a finite number $n \geq 2$ of sets $X_1, X_2, \ldots, X_n$ using the index set $[n] = \{1, 2, \ldots, n\}$.

$$\bigcap_{i \in [n]} X_i = \bigcap_{i=1}^{n} X_i = X_1 \cap X_2 \cap \cdots \cap X_n = \{x \mid x \in X_i \text{ for each integer } 1 \leq i \leq n\}$$
$$\bigcup_{i \in [n]} X_i = \bigcup_{i=1}^{n} X_i = X_1 \cup X_2 \cup \cdots \cup X_n = \{x \mid x \in X_i \text{ for some integer } 1 \leq i \leq n\}$$

Example 0.1.18. Consider the sets $A_1 = \{1, 2\}, A_2 = \{2, 3\}, \ldots, A_{10} = \{10, 11\}$. Crucially, we note that $A_i = \{i, i+1\}$ for each integer $1 \leq i \leq 10$. Using the index set $[10] = \{1, 2, \ldots, 10\}$ yields

$$\bigcap_{i=1}^{10} A_i = \{a \mid a \in A_i \text{ for each integer } 1 \leq i \leq 10\} = \varnothing,$$

$$\bigcap_{i=j}^{j+1} A_i = \{a \mid a \in A_j \text{ and } a \in A_{j+1}\} = \{j+1\}, \text{ and}$$

$$\bigcap_{i=j}^{k} A_i = \{a \mid a \in A_i \text{ for each integer } 1 \leq j \leq k \leq 10\} = \begin{cases} \{j, j+1\} & \text{if } k = j, \\ \{j+1\} & \text{if } k = j+1, \text{ and} \\ \varnothing & \text{if } k \geq j+2. \end{cases}$$

Consequently, the intersection of these sets is typically empty; however, the union satisfies that

$$\bigcup_{i=1}^{10} A_i = \{a \mid a \in A_i \text{ for some integer } 1 \leq i \leq 10\} = \{1, 2, \ldots, 11\},$$
$$\bigcup_{i=3}^{7} A_i = \{a \mid a \in A_i \text{ for some integer } 3 \leq i \leq 7\} = \{3, 4, \ldots, 8\}, \text{ and}$$
$$\bigcup_{i=j}^{k} A_i = \{a \mid a \in A_i \text{ for some integer } 1 \leq j \leq k \leq 10\} = \{j, j+1, \ldots, k+1\}.$$

Example 0.1.19. Consider the index set $L = \{a, b, c, \ldots, z\}$ consisting of all 26 letters of the English alphabet. We may define for each letter $\ell \in L$ the set W_ℓ consisting of all English words that contain the letter ℓ; this induces an indexed collection of sets $\{W_\ell\}_{\ell \in L}$. Certainly, we have that

$$\bigcap_{\ell \in L} W_\ell = \varnothing \text{ and } \bigcup_{\ell \in L} W_\ell = \{\omega \mid \omega \text{ is a word in the English language}\}$$

because there is no word in the English language that consists of all letters of the alphabet. Even more, consider the set $V = \{a, e, i, o, u\}$ of all vowels in the English language. We note that $\cap_{\ell \in V} W_\ell$ consists of many words, including satisfying words like "facetious" and "sequoia." Conversely, the word "why" does not belong to $\cup_{\ell \in V} W_\ell$ because it does not contain any of the letters a, e, i, o, or u.

We need not confine ourselves to the case that our index set is finite. Explicitly, we may consider any collection of sets $\{X_i\}_{i\in I}$ indexed by any nonempty (possibly infinite) set I. We have that

$$\bigcap_{i\in I} X_i = \{x \mid x \in X_i \text{ for each element } i \in I\} \text{ and}$$
$$\bigcup_{i\in I} X_i = \{x \mid x \in X_i \text{ for some element } i \in I\}.$$

We may also refer to the elements $i \in I$ as **indices**; the set $\{X_i\}_{i\in I}$ is an indexed collection of sets.

Example 0.1.20. Consider the infinite index set $I = \mathbb{Z}_{\geq 0}$ consisting of all non-negative integers. We may construct an indexed collection of sets $\{X_i\}_{i\in I}$ by declaring that $X_i = \{i, i+1\}$ for each element $i \in I$. Conventionally, the intersection and union over this infinite index set are written as

$$\bigcap_{i\in I} X_i = \bigcap_{i=0}^{\infty} X_i \text{ and } \bigcup_{i\in I} X_i = \bigcup_{i=0}^{\infty} X_i.$$

Computing the former gives the empty set, but the latter yields the index set $I = \mathbb{Z}_{\geq 0}$.

Example 0.1.21. Consider the infinite index set $\mathbb{Z}_{\geq 1}$ consisting of all integers $n \geq 1$, i.e., all positive integers. Each positive integer n gives rise to a closed interval of real numbers

$$C_n = \left[-\frac{1}{n}, \frac{1}{n}\right] = \left\{x \in \mathbb{R} : -\frac{1}{n} \leq x \leq \frac{1}{n}\right\}.$$

Each of these intervals is **nested** within the preceding interval: explicitly, for each integer $n \geq 1$, we have that $C_n \supseteq C_{n+1}$ because for any real number $x \in C_{n+1}$, we have that $x \in C_n$ since

$$-\frac{1}{n} < -\frac{1}{n+1} \leq x \leq \frac{1}{n+1} < \frac{1}{n}.$$

Consequently, it follows that $C_1 \supseteq C_2 \supseteq \cdots$ so that the indexed collection of sets $\{C_n\}_{n=1}^{\infty}$ forms a **descending chain** of sets. Generally, it is true for descending chains of sets that the union of sets in the chain is the largest set in the chain (see Proposition 0.1.13). Put another way, we have that

$$\bigcup_{n=1}^{\infty} C_n = \left\{x \in \mathbb{R} : -\frac{1}{n} \leq x \leq \frac{1}{n} \text{ for some integer } n \geq 1\right\} = [-1, 1].$$

On the other hand, the only real number x satisfying that $|x| \leq -\frac{1}{n}$ for all integers $n \geq 1$ is $x = 0$: indeed, if $|x| > 0$, we can find an integer $n \geq 1$ such that $|x| > -\frac{1}{n}$. We conclude therefore that

$$\bigcap_{n=1}^{\infty} C_n = \left\{x \in \mathbb{R} : -\frac{1}{n} \leq x \leq \frac{1}{n} \text{ for each integer } n \geq 1\right\} = \{0\}.$$

0.1.2 Partitions of Sets

We say that two sets X_i and X_j are **disjoint** if $X_i \cap X_j = \varnothing$. Even more, if the indexed collection of sets $\{X_i\}_{i\in I}$ satisfies the condition that the sets X_i and X_j are disjoint for each pair of distinct indices $i, j \in I$, then we say that $\{X_i\}_{i\in I}$ is **pairwise disjoint** (or **mutually exclusive**). Often, we will abuse terminology by saying that the sets X_i are pairwise disjoint for each element $i \in I$.

Example 0.1.22. Consider the sets $A = \{1, 4, 7\}$, $B = \{2, 5, 8\}$, and $C = \{3, 6, 9\}$. One can readily verify that $A \cap B = A \cap C = B \cap C = \varnothing$, hence the set $\{A, B, C\}$ is pairwise disjoint.

Example 0.1.23. Consider the sets $D = \{1, 3, 5, 7\}$, $E = \{2, 4, 6, 8\}$, and $F = \{3, 5, 7, 9\}$. We have that $D \cap E = E \cap F = \varnothing$ but $D \cap F = \{3, 5, 7\}$, hence the set $\{D, E, F\}$ is not pairwise disjoint.

Observe that if $X_i = \varnothing$ for any index i, then $X_i \cap X_j = \varnothing$ for all indices j by the Going-Down Property of Set Intersection, hence any indexed collection of sets $\{X_i\}_{i \in I}$ containing the empty set is pairwise disjoint. Consequently, we may restrict our attention to collections of nonempty pairwise disjoint sets. We say that an indexed collection of sets $\mathcal{P} = \{X_i\}_{i \in I}$ forms a **partition** of a set X if

(a.) the sets X_i are nonempty, i.e., $X_i \neq \varnothing$ for each element $i \in I$;

(b.) the sets X_i cover the set X, i.e., $X = \cup_{i \in I} X_i$; and

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

Example 0.1.24. Every set X admits a canonical partition $\mathcal{X} = \{\{x\}\}_{x \in X}$ indexed by the **singleton** sets $\{x\}$ for each element $x \in X$; however, many sets admit more interesting partitions.

Example 0.1.25. Consider the sets $A = \{1, 4, 7\}$, $B = \{2, 5, 8\}$, and $C = \{3, 6, 9\}$ of Example 0.1.22. Considering that the sets A, B, and C are pairwise disjoint and $A \cup B \cup C = \{1, 2, \ldots, 9\} = [9]$, it follows that the set $\mathcal{P} = \{A, B, C\}$ constitutes a partition of the finite set $[9] = \{1, 2, \ldots, 9\}$.

Conversely, even though the nonempty sets $D = \{1, 3, 5, 7\}$, $E = \{2, 4, 6, 8\}$, and $F = \{3, 5, 7, 9\}$ of Example 0.1.23 satisfy $[9] = D \cup E \cup F$, they are not pairwise disjoint and do not partition $[9]$.

Example 0.1.26. Consider the set $\mathbb{Z}$ of integers. Given any integer n, divide n by 3 to obtain unique integers q and r (the quotient and remainder of this division) such that $0 \leq r \leq 2$. Consequently, every integer n can be written as $n = 3q + r$ for some unique integers q and $0 \leq r \leq 2$. We conclude that $\mathbb{Z} = X_0 \cup X_1 \cup X_2$ is a partition of $\mathbb{Z}$ with $X_r = \{3q + r \mid q \in \mathbb{Z}\}$ for each integer $0 \leq r \leq 2$.

Example 0.1.27. Every nonzero rational number can be written uniquely as a **reduced fraction** $\frac{p}{q}$ for some nonzero integers p and q that have no common divisors other than 1. Consider the indexed collection of sets $\{D_q\}_{q=1}^{\infty}$ of nonzero reduced fractions with denominator q, i.e.,

$$D_q = \left\{ \frac{p}{q} : p, q \in \mathbb{Z} \setminus \{0\} \text{ and } p \text{ and } q \text{ have no common divisors other than } 1 \right\}.$$

Explicitly, we have that

$$D_1 = \{\ldots, -2, -1, 1, 2, \ldots\},\ D_2 = \left\{\ldots, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\right\},\ \text{and}\ D_3 = \left\{\ldots, -\frac{2}{3}, -\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \ldots\right\}.$$

Later in the semester, we will be able to prove that D_q and D_r are disjoint for any pair of distinct positive integers q and r. Considering that every nonzero rational number can be written as a reduced fraction, it follows that the collection of nonzero rational numbers is partitioned by $\{D_q\}_{q=1}^{\infty}$.

0.1.3 Cartesian Products of Sets

Given any nonempty set X, for any elements $x_1, x_2 \in X$, the **ordered pair** (x_1, x_2) is simply an ordered list with first **coordinate** x_1 and second coordinate x_2. Crucially, the ordered pairs (x_1, x_2) and (x_2, x_3) are equal if and only if $x_1 = x_2 = x_3$ for any elements $x_1, x_2, x_3 \in X$. We are already familiar with ordered pairs of real numbers: indeed, the concept arises naturally in our high school mathematics courses from intermediate algebra to calculus. Concretely, we refer to the set $X \times Y$ of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$ as the **Cartesian product** of X and Y.

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$$

Example 0.1.28. Consider the sets $X = \{-1, 1\}$ and $Y = \{1, 2, 3\}$. We have that

$$X \times Y = \{(-1, 1), (-1, 2), (-1, 3), (1, 1), (1, 2), (1, 3)\} \text{ and}$$
$$Y \times X = \{(1, -1), (1, 1), (2, -1), (2, 1), (3, -1), (3, 1)\}.$$

Consequently, the Cartesian product of sets is in general not commutative: indeed, the sets $X \times Y$ and $Y \times X$ from above are not equal because we have that $(-1, 1) \in X \times Y$ and $(-1, 1) \notin Y \times X$.

Even more, we may also consider the Cartesian product of a set with itself. We have that

$$X \times X = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\} \text{ and}$$
$$Y \times Y = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}.$$

Example 0.1.29. Observe that the Cartesian product $\mathbb{Z} \times \mathbb{Z} = \{(a, b) \mid a \text{ and } b \text{ are integers}\}$ is the collection of all integer points in the **Cartesian plane** $\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \text{ and } y \text{ are real numbers}\}$.

Example 0.1.30. Given any real univariate function $f : \mathbb{R} \to \mathbb{R}$, the **graph** of f consists of all ordered pairs $(x, f(x))$ such that x is in the **domain** of f. Explicitly, if we assume that D_f is the domain of f and R_f is the **range** of f, then the graph of f is given by the Cartesian product

$$G_f = D_f \times R_f = \{(x, f(x)) \mid x \in D_f \text{ and } f(x) \in R_f\}.$$

Concretely, if $f(x) = 2x + 3$, then the graph of f is given by $G_f = \{(x, 2x + 3) \mid x \in \mathbb{R}\}$.

Crucially, if X and Y are finite sets with cardinalities $|X|$ and $|Y|$, then the Cartesian product $X \times Y$ has cardinality $|X||Y|$ because an element of $X \times Y$ is uniquely determined by the ordered pair (x, y). Consequently, we have that $\varnothing \times Y = \varnothing = X \times \varnothing$ for any finite sets X and Y. Even if X and Y are infinite, the Cartesian product with the empty set results in the empty set.

Proposition 0.1.31 (Cartesian Product of Finite Sets)**.** *Consider any finite sets X and Y.*

1.) *We have that $|X \times Y| = |X||Y|$. Consequently, the cardinality of the Cartesian product of any pair of finite sets is the product of the cardinalities of the underlying sets.*

2.) *We have that $\varnothing \times Y = \varnothing = X \times \varnothing$. Consequently, the Cartesian product of any finite set with the empty set is the empty set. Even more, this equality holds whenever X and Y are infinite.*

Proof. We will prove only the last statement of the proposition since the proof of the first statement is provided above. Certainly, if X and Y are finite, then $|\varnothing \times Y| = |\varnothing||Y| = 0 = |X||\varnothing| = |X \times \varnothing|$ so that $\varnothing \times Y = \varnothing = X \times \varnothing$. We may assume therefore that X and Y are infinite. By definition of the Cartesian product, we have that $\varnothing \times Y$ consists of all ordered pairs (x, y) such that $x \in \varnothing$ and $y \in Y$. Considering that there are no such elements $x \in \varnothing$, there are no such ordered pairs. □

0.1.4 Relations, Equivalence Relations, and Partial Orders

Given any sets X and Y, a **relation from** X to Y is any subset R of the Cartesian product $X \times Y$. Explicitly, a relation R from X to Y consists of ordered pairs (x, y) such that $x \in X$ and $y \in Y$. We say that an element $x \in X$ is **related to** an element $y \in Y$ by R if $(x, y) \in R$, and we write that $x \, R \, y$ in this case; otherwise, if $(x, y) \notin R$, then x is not related to y by R, and we write $x \not R \, y$.

Example 0.1.32. Consider the sets $X = \{-1, 1\}$ and $Y = \{1, 2, 3\}$ of Example 0.1.28. Observe that $|X \times Y| = |X||Y| = 6$, hence there are $|P(X \times Y)| = 2^6$ possible relations from X to Y. We may define one such relation $R = \{(1, 1), (1, 2), (1, 3)\}$ from X to Y. Under this relation, it holds that $1 \, R \, 1$, $1 \, R \, 2$, and $1 \, R \, 3$ so that 1 is related to each of the elements of Y. Conversely, we have that $-1 \not R \, 1$, $-1 \not R \, 2$, and $-1 \not R \, 3$ so that -1 is not related to any of the elements of Y.

Every relation R from a set X to a set Y induces two important sets: namely, the collection

$$\text{dom}(R) = \{x \in X \mid (x, y) \in R \text{ for some element } y \in Y\}$$

consists of all elements in X are related to some element of Y by R; it is called the **domain** of the relation R from X to Y. Likewise, the **range** of the relation R from X to Y is given by

$$\text{range}(R) = \{y \in Y \mid (x, y) \in R \text{ for some element } x \in X\}$$

and consists of all elements $y \in Y$ for which there exists an element of $x \in X$ that is related to y by R. Crucially, the domain of a relation R from X to Y only concerns the first coordinate of an element of R, and the range of R only takes into account the second coordinate of an element of R.

Example 0.1.33. Consider the relation $R = \{(1, 1), (1, 2), (1, 3)\}$ from $X = \{-1, 1\}$ to $Y = \{1, 2, 3\}$ of Example 0.1.32. We have that $\text{dom}(R) = \{1\}$ and $\text{range}(R) = \{1, 2, 3\} = Y$.

Given any relation R from a set X to a set Y, we may define the **inverse relation**

$$R^{-1} = \{(y, x) \mid (x, y) \in R\}.$$

Crucially, if R is a relation from X to Y, then R^{-1} is a relation from Y to X, i.e., $R^{-1} \subseteq Y \times X$.

Example 0.1.34. Consider the relation $R = \{(1, 1), (1, 2), (1, 3)\}$ from $\{-1, 1\}$ to $\{1, 2, 3\}$ of Example 0.1.32. We have that $R^{-1} = \{(1, 1), (2, 1), (3, 1)\}$, $\text{dom}(R^{-1}) = \{1, 2, 3\}$, and $\text{range}(R^{-1}) = \{1\}$.

We refer to a subset R of the Cartesian product $X \times X$ as a **relation on** X. Every set X admits a relation Δ_X called the **diagonal** 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 $\Delta_X = \{(x, x) \mid x \in X\}$. Observe that if X is a finite set with cardinality $|X|$, then the cardinality of $X \times X$ is $|X|^2$, hence there are a total of $2^{|X|^2}$ possible relations on a set X simply because there are as many subsets of $X \times X$.

Example 0.1.35. Consider the set $X = \{-1, 1\}$. We may define relations

$$\begin{aligned}
\Delta_X &= \{(-1, -1), (1, 1)\} \text{ with } \text{dom}(\Delta_X) = \{-1, 1\} = \text{range}(\Delta_X), \\
R_1 &= \{(-1, 1), (1, -1)\} \text{ with } \text{dom}(R_1) = \{-1, 1\} = \text{range}(R_1), \text{ and} \\
R_2 &= \{(-1, -1), (-1, 1)\} \text{ with } \text{dom}(R_2) = \{-1\} \text{ and } \text{range}(R_2) = \{-1, 1\}.
\end{aligned}$$

Observe that $\Delta_X^{-1} = \Delta_X$ and $R_1^{-1} = R_1$ but $R_2^{-1} = \{(-1, -1,), (1, -1)\}$ is not its own inverse.

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 will 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 if and only if $R \supseteq \Delta_X$. Even more, if it holds that $(y, x) \in R$ whenever $(x, y) \in R$, then R is **symmetric**. Last, if $(x, y) \in R$ and $(y, z) \in R$ together imply that $(x, z) \in R$, then we refer to the relation R as **transitive**.

Example 0.1.36. Consider the following relations on the set $X = \{x, y, z\}$.

$$
\begin{aligned}
R_1 &= \{(x, y), (y, z)\} \\
R_2 &= \{(x, x), (x, y), (y, y), (y, z), (z, z)\} \\
R_3 &= \{(x, y), (y, x)\} \\
R_4 &= \{(x, y), (y, z), (x, z)\} \\
R_5 &= \{(x, x), (x, y), (y, x), (y, y), (y, z), (z, y), (z, z)\} \\
R_6 &= \{(x, x), (x, y), (x, z), (y, y), (y, z), (z, z)\} \\
R_7 &= \{(x, x), (x, y), (y, x), (y, y)\} \\
R_8 &= \{(x, x), (x, y), (x, z), (y, x), (y, y), (y, z), (z, x), (z, y), (z, z)\}
\end{aligned}
$$

Observe that R_1 is not reflexive because (x, x) does not lie in R_1; it is not symmetric because (x, y) lies in R_1 and yet (y, x) does not lie in R_1; and it is not transitive because (x, y) and (y, z) both lie in R_1 and yet (x, z) does not lie in R_1. We note that R_2 is reflexive, but it is not symmetric because it contains (x, y) but not (y, x), and it is not transitive because it contains (x, y) and (y, z) but not (x, z). Continuing in this manner, the reader should verify the properties of the following table.

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
reflexive		✓			✓	✓		✓
symmetric			✓		✓		✓	✓
transitive				✓		✓	✓	✓

Example 0.1.37. Consider the relation R defined on the set $\mathbb{Z}$ of integers such that for any pair of integers $x, y \in \mathbb{Z}$, we have that $x \; R \; y$ if and only if $x \leq y$. Certainly, every integer x is equal to itself, hence we have that $x \leq x$ so that R is reflexive; however, we note that R is not symmetric since the strict inequality $0 < 1$ implies that $0 \; R \; 1$ and yet $1 \not R \; 0$. Last, it is straightforward to verify that R is transitive because if $x \; R \; y$ and $y \; R \; z$, then $x \leq y \leq z$ so that $x \leq z$ and $x \; R \; z$.

Example 0.1.38. Consider the relation S defined on the set $\mathbb{Z}$ of integers such that for any integers $x, y \in \mathbb{Z}$, we have that $x \; S \; y$ if and only if $x \neq y$. Contrary to Example 0.1.37, this relation is symmetric but neither reflexive nor transitive: indeed, one can readily check that $x \; S \; y$ if and only if $y \; S \; x$, hence S is symmetric; however, we have that $0 = 0$ so that $0 \not S \; 0$ and S is not reflexive. Likewise, we have that $0 \neq 1$ and $1 \neq 0$ so that $0 \; S \; 1$ and $1 \; S \; 0$ but $0 \not S \; 0$, hence S is not transitive.

Example 0.1.39. Consider the relation D defined on the set $\mathbb{R}$ of real numbers such that $x \; D \; y$ if and only if $|x - y| \leq 1$. We can immediately verify that D is reflexive and symmetric: indeed, we have that $|x - x| = 0$ so that $x \; D \; x$ and $|y - x| = |x - y|$ so that $y \; D \; x$ if and only if $x \; D \; y$; however, $0 \; D \; 1$ and $1 \; D \; 2$ do not together imply that $0 \; D \; 2$ because $|2 - 0| > 1$, so D is not transitive.

Relations that are reflexive, symmetric, and transitive are defined as **equivalence relations**.

Example 0.1.40. Consider any set X. We may define a relation R on X by declaring that $x \; R \; y$ if and only if $x = y$. Equality is reflexive because $x = x$ holds for all elements $x \in X$; it is symmetric because $x = y$ implies that $y = x$ for any elements $x, y \in X$; and it is transitive because if $x = y$ and $y = z$, then $x = y = z$ implies that $x = z$ for all elements $x, y, z \in X$. Consequently, equality is an equivalence relation. We synthesize the result of this example in the following proposition.

Proposition 0.1.41. *Given any set X, the diagonal $\Delta_X = \{(x, x) \mid x \in X\}$ of X is an equivalence relation on X. Explicitly, every set admits at least one equivalence relation on itself.*

Proof. Observe that as a relation on X, the diagonal of X captures equality of the elements of X: if $(x, y) \in \Delta_X$, then we must have that $x = y$. Conversely, if $x = y$, then $(x, y) \in \Delta_X$. Put another way, the relation Δ_X can be identified with the equality equivalence relation of Example 0.1.40. □

Example 0.1.42. Consider the collection $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that the first derivative $f'(x)$ of $f(x)$ is continuous for all real numbers x. We may define a relation R on $\mathcal{C}^1(\mathbb{R})$ such that $(f, g) \in R$ if and only if $f'(x) = g'(x)$ for all real numbers x. Because R is defined by equality and equality is reflexive, symmetric, and transitive, it follows that R is an equivalence relation on $\mathcal{C}^1(\mathbb{R})$.

Example 0.1.43. Consider the relation R defined on the set $\mathbb{Z}$ of integers such that $x \; R \; y$ if and only if $y - x = 2k$ for some integer k. Considering that $x - x = 0 = 2 \cdot 0$, it follows that R is reflexive. Even more, if $y - x = 2k$ for some integer k, then $x - y = -(y - x) = 2(-k)$ for the integer $-k$, hence R is symmetric. Last, if $y - x = 2k$ and $z - y = 2\ell$ for some pair of integers k and ℓ, then $z - x = (z - y) + (y - x) = 2\ell + 2k = 2(\ell + k)$ for the integer $\ell + k$. Consequently, the relations $x \; R \; y$ and $y \; R \; z$ together yield that $x \; R \; z$. We conclude that R is an equivalence relation on $\mathbb{Z}$.

Example 0.1.44. Often, it is useful to determine if a relation is an equivalence relation by examining its elements explicitly. Consider the following relation defined on the set $[5] = \{1, 2, 3, 4, 5\}$.

$$R = \{(1,1), (1,3), (1,5), (2,2), (2,4), (3,1), (3,3), (3,5), (4,2), (4,4), (5,1), (5,3), (5,5)\}$$

Considering that R contains the diagonal of $[5]$, it follows that R is reflexive. Put another way, we have that $(x, x) \in R$ for all elements $x \in [5]$. Even more, for each element $(x, y) \in R$, we have that $(y, x) \in R$ so that R is symmetric. Last, one can readily verify that if (x, y) and (y, z) both lie in R, then (x, z) lies in R, hence R is transitive. We conclude that R is an equivalence relation on $[5]$.

Given an equivalence relation E defined on a set X, we say that x and y are **equivalent modulo** E provided that x is related to y by E. We note that this convention is due to Carl Friedrich Gauss to express that x and y are "the same up to differences accounted for by E." We may define the **equivalence class** $[x]$ of an element $x \in X$ modulo the equivalence relation E as the set of elements $y \in X$ that are equivalent to x modulo E. Consequently, the equivalence class of x modulo E is

$$[x] = \{y \in X \mid y \; E \; x\} = \{y \in X \mid (y, x) \in E\}.$$

Example 0.1.45. Every element of a set X lies in its own equivalence class modulo the equivalence relation $\Delta_X = \{(x, x) \mid x \in X\}$ because the elements of Δ_X are precisely the ordered pairs (x, x). Consequently, the equivalence class of any element $x \in X$ modulo Δ_X is the singleton $[x] = \{x\}$.

Example 0.1.46. Consider the equivalence relation R defined on the set $\mathcal{C}^1(\mathbb{R})$ of Example 0.1.42. Given any functions $f, g \in \mathcal{C}^1(\mathbb{R})$, because $f'(x)$ and $g'(x)$ are continuous for all real numbers x, it follows that $f(x) - g(x)$ is continuous and differentiable on every open interval of the form $(0, x)$. Consequently, the Mean Value Theorem ensures the existence of a real number $0 < c < x$ such that

$$f(x) - g(x) = [f'(c) - g'(c)]x + [f(0) - g(0)].$$

Observe that if $f'(x) = g'(x)$ for all real numbers x, then $f'(c) - g'(c) = 0$, and there exists a real number C such that $g(x) = f(x) + C$. Conversely, if $g(x) = f(x) + C$ for some real number C, then $f'(x) = g'(x)$. We conclude that the equivalence classes of $\mathcal{C}^1(\mathbb{R})$ modulo R are given precisely by the sets $[f] = \{g \in \mathcal{C}^1(\mathbb{R}) \mid (g, f) \in R\} = \{g \in \mathcal{C}^1(\mathbb{R}) \mid g(x) = f(x) + C \text{ for some real number } C\}$.

Example 0.1.47. Consider the equivalence relation R of Example 0.1.43. By definition, if $x = 2k$ for some integer k, then $x - 0 = 2k$, hence $(x, 0)$ lies in R. Conversely, if $(x, 0)$ lies in R, then $x = 2k$ for some integer k. We conclude that the equivalence class of 0 modulo R is given by

$$[0] = \{x \in R \mid (x, 0) \in R\} = \{x \in R \mid x = 2k \text{ for some integer } k\}.$$

Likewise, if $x = 2k + 1$ for some integer k, then $x - 1 = 2k$ for some integer k so that $(x, 1)$ lies in R. Even more, if $(x, 1)$ lies in R, then $x - 1 = 2k$ and $x = 2k + 1$ for some integer k. Considering this in terms of R, we conclude that the equivalence class of 1 modulo R is given by

$$[1] = \{x \in R \mid (x, 1) \in R\} = \{x \in R \mid x = 2k + 1 \text{ for some integer } k\}.$$

Every integer is of the form $2k$ or $2k + 1$, hence these are the equivalence classes of $\mathbb{Z}$ modulo R.

Example 0.1.48. Consider the equivalence relation R of Example 0.1.44. Each of the integers 1, 3, and 5 are equivalent modulo R because $(1, 3)$ and $(3, 5)$ lie in the equivalence relation R. On the other hand, the integers 2 and 4 are equivalent modulo R because $(2, 4)$ lies in R; thus, there are two distinct equivalence classes modulo R — namely, $[1] = \{1, 3, 5\} = [3] = [5]$ and $[2] = \{2, 4\} = [4]$.

Given any nonempty relation E defined on a nonempty set X, we recall that E is an equivalence relation provided that E is reflexive, symmetric, and transitive. Each equivalence relation E defined on X induces a collection of sets defined on X called the equivalence classes of the elements of X. Explicitly, the equivalence class $[x]$ of an element $x \in X$ is defined by $[x] = \{y \in X \mid (y, x) \in E\}$. We demonstrate next that a pair of equivalence classes of elements of X modulo E are either equal or disjoint; as a corollary, we obtain a relationship between equivalence relations and partitions.

Proposition 0.1.49 (Equality of Equivalence Classes)**.** *Consider any equivalence relation E defined on a nonempty set X. Given any elements $x, y \in X$, we have that $[x] = [y]$ if and only if $(x, y) \in E$.*

Proof. By definition of $[x]$, for any element $z \in [x]$, we have that $(z, x) \in E$, hence the symmetry of the equivalence relation E yields that $(x, z) \in E$. Given that $[x] = [y]$, we have that $z \in [y]$ so that $(z, y) \in E$. Last, the transitivity of E ensures that $(x, y) \in E$ because (x, z) and (z, y) lie in E.

Conversely, we will assume that $(x, y) \in E$. We must demonstrate that $[x] \subseteq [y]$ and $[y] \subseteq [x]$. Given any element $z \in [x]$, we have that $(z, x) \in E$. By assumption that $(x, y) \in E$, the transitivity of the equivalence relation E yields that $(z, y) \in E$ so that $z \in [y]$. Likewise, for any element $w \in [y]$, we have that $(w, y) \in E$. By the symmetry of the equivalence relation E, we have that $(y, x) \in E$ by assumption that $(x, y) \in E$, hence the transitivity of E yields that $(w, x) \in E$ so that $w \in [x]$. □

Proposition 0.1.50 (Equivalence Classes Are Either the Same or Disjoint)**.** *Consider any equivalence relation E defined on a nonempty set X. Given any elements $x, y \in X$, the classes $[x]$ and $[y]$ of x and y modulo E are the same or disjoint. Explicitly, we must have that $[x] = [y]$ or $[x] \cap [y] = \varnothing$.*

Proof. Consider any pair $[x]$ and $[y]$ of equivalence classes of a set X modulo an equivalence relation E. We have nothing to prove if $[x] \cap [y] = \varnothing$, hence we may assume that this is not the case and prove that $[x] = [y]$. Concretely, we will assume that there exists an element $w \in [x] \cap [y]$. Crucially, by definition of the equivalence classes of X modulo E, we have that $(w, x) \in E$ and $(w, y) \in E$. By assumption that E is an equivalence relation, it follows that $(x, w) \in E$ by symmetry, hence the transitivity of E together with the inclusions $(x, w), (w, y) \in E$ yield that $(x, y) \in E$. By Proposition 0.1.49, we conclude that $[x] = [y]$, hence the proposed result is in fact established. □

Corollary 0.1.51 (Equivalence Relations and Partitions)**.** *Each equivalence relation on a nonempty set X induces a partition of X. Each partition of X induces an equivalence relation on X.*

Proof. By Proposition 0.1.50, for any equivalence relation E on a nonempty set X, the collection $\mathcal{P}$ of distinct equivalence classes of X modulo E is pairwise disjoint. Considering that every element of X lies in its own equivalence class, we conclude that $X = \cup_{C \in \mathcal{P}} C$ is a partition of X.

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

Example 0.1.52. Consider the equivalence relation R of Example 0.1.44. By Corollary 0.1.51, the collection of distinct equivalence classes of $[5]$ modulo R provides a partition of $[5]$. By Example 0.1.48, the distinct equivalence classes of $[5]$ modulo R are $[1] = \{1, 3, 5\}$ and $[2] = \{2, 4\}$, hence the underlying partition of $[5]$ induced by the equivalence relation R is $\mathcal{P} = \{[1], [2]\} = \{\{1, 3, 5\}, \{2, 4\}\}$.

Example 0.1.53. Consider the following partition $\mathcal{P} = \{R_0, R_1, R_2, R_3\}$ of the set $\mathbb{Z}$ of integers.

$$R_0 = \{\dots, -8, -4, 0, 4, \dots\} \qquad R_2 = \{\dots, -6, -2, 2, 6, \dots\}$$
$$R_1 = \{\dots, -7, -3, 1, 5, \dots\} \qquad R_3 = \{\dots, -5, -1, 3, 7, \dots\}$$

By Corollary 0.1.51, the distinct sets in the partition $\mathcal{P}$ constitute the distinct equivalence classes of an equivalence relation $E_{\mathcal{P}}$ on $\mathbb{Z}$. Explicitly, we have that $(x, y) \in E_{\mathcal{P}}$ if and only if $x, y \in R_i$ for some integer $1 \leq i \leq 4$. Consequently, the distinct equivalence classes of $\mathbb{Z}$ modulo the equivalence relation $E_{\mathcal{P}}$ are R_0, R_1, R_2, and R_3. Observe that $(0, 4) \in E_{\mathcal{P}}$ holds because $0, 4 \in R_0$ and $(1, 5) \in E_{\mathcal{P}}$ holds because $1, 5 \in R_1$, but neither $(0, 2)$ nor $(1, 3)$ lie in $E_{\mathcal{P}}$. By Proposition 0.1.49, a pair of equivalence classes are distinct if and only if their **representatives** are related, hence the distinct equivalence classes of $\mathbb{Z}$ modulo $E_{\mathcal{P}}$ are $[0]$, $[1]$, $[2]$, and $[3]$ or similarly $[4]$, $[5]$, $[6]$, and $[7]$ and so on.

Last, we say that a relation R defined on a set X is **antisymmetric** if for every pair of elements $x, y \in X$, the inclusions $(x, y) \in R$ and $(y, x) \in R$ together imply that $x = y$. Equivalence relations

are reflexive, symmetric, and transitive relations on a set; however, if we replace the requirement of the symmetry condition with the property of antisymmetry, then we obtain a **partial order** on the set. 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 since the diagonal is a partial order.

Proposition 0.1.54. *Given any set X, the diagonal Δ_X of X is a partial order on X.*

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

Example 0.1.55. Observe that the real numbers $\mathbb{R}$ are partially ordered via the usual less-than-or-equal-to $\leq$. Put another way, the relation $P = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \leq y\}$ is a partial order on $\mathbb{R}$. Explicitly, we have that $x = x$ so that $x \leq x$ and $(x, x) \in P$ for all real numbers x. Likewise, if we have that $(x, y), (y, x) \in P$, then $x \leq y$ and $y \leq x$ together imply that $x = y$. Last, if we assume that $(x, y), (y, z) \in P$, then $x \leq y$ and $y \leq z$ together imply that $x \leq z$ so that $(x, z) \in P$.

Example 0.1.56. Divisibility constitutes a partial order on the set $\mathbb{Z}_{>0}$ of positive integers. Consider the relation $D = \{(a, b) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \mid a \text{ divides } b\}$. Observe that D is reflexive since a divides a. 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 $a = bn = amn$. 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 positive. Ultimately, this proves that $a = b$, hence D is antisymmetric. Last, if a divides b and b divides c, then a divides c: indeed, we have that $b = am$ and $c = bn$ together yield that $c = bn = (am)n = a(mn)$.

Every set admits a partial order via the diagonal, 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 distinct prime numbers 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 $(y, x_0) \in P$ 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 these ingredients to comprise one of the most ubiquitous results in mathematics (and especially in the ideal theory of commutative algebra).

Theorem 0.1.57 (Zorn's Lemma)**.** *Consider any partial order P defined on any set X. With respect to P, if every chain Y in X has an upper bound in Y, then Y admits a maximal element $y_0 \in Y$.*

0.1.5 Congruence Modulo n

We say that a nonzero integer a **divides** an integer b if there exists an integer c such that $b = ac$. We will write $a \mid b$ in this case, and we will typically say that b is **divisible by** a. Given any nonzero integer n, we say that a pair of integers a and b are **congruent modulo** n if it holds that n divides $b - a$ or $n \mid (b - a)$. Conventionally, if a and b are congruent modulo n, we write $b \equiv a \pmod{n}$.

Example 0.1.58. We have that $7 \equiv 3 \pmod{4}$ because $7 - 3 = 4$ is divisible by 4.

Example 0.1.59. We have that $5 \equiv 21 \pmod 4$ because $5 - 21 = -16 = 4(-4)$ is divisible by 4.

Example 0.1.60. We have that $11 \not\equiv 8 \pmod 4$ because $11 - 8 = 3$ is not divisible by 4.

Given any nonzero integer n, we note that congruence modulo n induces a relation R_n on the set $\mathbb{Z}$ of integers: indeed, for any integers a and b, we have that $(a, b) \in R_n$ if and only if $a \; R_n \; b$ if and only if $b \equiv a \pmod n$ if and only if n divides $b - a$. Even more, the following proposition and Proposition 0.1.62 guarantee that the relation of congruence modulo n admits "nice" properties.

Proposition 0.1.61 (Properties of Congruence Modulo n)**.** *Consider any nonzero integer n and any integers a, b, and c. Each of the following properties of congruence modulo n holds.*

1.) (**Identity Property**) *We have that $a \equiv 0 \pmod n$ if and only if n divides a.*

2.) (**Well-Defined Property**) *We have that $b \equiv a \pmod n$ if and only if $b - a \equiv 0 \pmod n$.*

3.) (**Reflexive Property**) *We have that $a \equiv a \pmod n$ for any integer a.*

4.) (**Symmetric Property**) *We have that $b \equiv a \pmod n$ if and only if $a \equiv b \pmod n$.*

5.) (**Transitive Property**) *If $b \equiv a \pmod n$ and $c \equiv b \pmod n$, then $c \equiv a \pmod n$.*

6.) (**Additive Property**) *We have that $b \equiv a \pmod n$ if and only if $b + c \equiv a + c \pmod n$.*

7.) (**Multiplicative Property**) *If $b \equiv a \pmod n$, then $cb \equiv ca \pmod n$.*

8.) (**Exponentiation Property**) *If $b \equiv a \pmod n$, then $b^k \equiv a^k \pmod n$ for any integer $k \geq 0$.*

Proof. (1.) We have that $a \equiv 0 \pmod n$ if and only if n divides $a - 0$ if and only if n divides a.

(2.) By the definition and Identity Property of congruence modulo n, we have that $b \equiv a \pmod n$ if and only if n divides $b - a$ if and only if $b - a \equiv 0 \pmod n$.

(3.) Considering that $a - a = 0 = n \cdot 0$, it follows that n divides $a - a$ so that $a \equiv a \pmod n$.

(4.) We have that $b \equiv a \pmod n$ if and only if n divides $b - a$ if and only if n divides $-(a - b)$ if and only if n divides $a - b$ if and only if $a \equiv b \pmod n$.

(5.) Given that $b \equiv a \pmod n$ and $c \equiv b \pmod n$, by definition, there exist integers k and ℓ such that $b - a = nk$ and $c - b = n\ell$. Observe that $c - a = (c - b) + (b - a) = nk + n\ell = n(k + \ell)$, hence n divides $c - a$ so that $c \equiv a \pmod n$ by definition of congruence modulo n.

(6.) We have that $b \equiv a \pmod n$ if and only if n divides $b - a$ if and only if $b - a = nk$ for some integer k if and only if $-b + a = (-b) - (-a) = n(-k)$ for some integer k if and only if n divides $-b - (-a)$ if and only if $-b \equiv -a \pmod n$ by definition of congruence modulo n.

(7.) By definition of congruence modulo n, we have that $b \equiv a \pmod n$ if and only if n divides $b - a$ if and only if n divides $(b + c) - (a + c)$ if and only if $b + c \equiv a + c \pmod n$.

(8.) By definition of congruence modulo n, if $b \equiv a \pmod n$, then n divides $b - a$ so that n divides $c(b - a)$. Considering that $c(b - a) = cb - ca$, it follows that $cb \equiv ca \pmod n$.

(9.) By the Multiplicative Property, if $b \equiv a \pmod n$, we have that $b^2 = b \cdot b \equiv b \cdot a \pmod n$ and $a^2 = a \cdot a \equiv a \cdot b \pmod n$. Considering that $b \cdot a = a \cdot b$, the Transitive Property of congruence modulo n yields that $b^2 = b \cdot b \equiv b \cdot a = a \cdot b \equiv a \cdot a = a^2 \pmod n$. By the same rationale, we have that $b^3 = b \cdot b^2 \equiv b \cdot a^2 = a \cdot a^2 = a^3 \pmod n$. Continuing in this manner establishes the result. □

Given any nonzero integer n, the relation R_n defined on the set $\mathbb{Z}$ of integers such that $a \; R_n \; b$ if and only if $b \equiv a \pmod{n}$ is commonly referred to as **congruence modulo** n. By the third, fourth, and fifth properties of Proposition 0.1.61, congruence modulo n is an equivalence relation on $\mathbb{Z}$.

Proposition 0.1.62. *Congruence modulo any nonzero integer n is an equivalence relation on $\mathbb{Z}$.*

Consider the equivalence class $[a]$ of any integer a modulo the equivalence relation of congruence modulo n. Conventionally, we refer to $[a]$ as the class of a **modulo** n. By definition, we have that

$$[a] = \{b \in \mathbb{Z} \mid b \equiv a \pmod{n}\} = \{b \in \mathbb{Z} \mid b - a = nq \text{ for some integer } q\} = \{nq + a \mid q \in \mathbb{Z}\}.$$

Consequently, the equivalence class of a modulo n consists of sums of integer multiples of n and a.

Example 0.1.63. Congruence modulo 1 is an equivalence relation on $\mathbb{Z}$, hence we may seek to determine the equivalence classes of the integers modulo 1. Considering that every integer is divisible by 1, it follows that every pair of integers are related by congruence modulo 1: indeed, for any pair of integers a and b, we have that $b - a = 1 \cdot (b - a)$, hence a and b are congruent modulo 1. But this implies that every integer is congruent to 0 modulo 1, hence there is only one equivalence class of integers modulo 1. Explicitly, we have that $[0] = \{1q + 0 \mid q \in \mathbb{Z}\} = \{q \mid q \in \mathbb{Z}\} = \mathbb{Z}$.

Example 0.1.64. Congruence modulo 2 is an equivalence relation on $\mathbb{Z}$ with equivalence classes

$$[0] = \{2q + 0 \mid q \in \mathbb{Z}\} = \{\dots, -4, -2, 0, 2, 4, \dots\} \text{ and}$$
$$[1] = \{2q + 1 \mid q \in \mathbb{Z}\} = \{\dots, -3, -1, 1, 3, 5, \dots\}.$$

By Proposition 0.1.50, these are all of the distinct equivalence classes of $\mathbb{Z}$ modulo 2. Even more, by Proposition 0.1.51, we obtain a partition of $\mathbb{Z}$ into distinct equivalence classes modulo 2

$$\mathbb{Z} = [0] \cup [1] = \{\dots, -4, -2, 0, 2, 4, \dots\} \cup \{\dots, -3, -1, 1, 3, 5, \dots\}.$$

Example 0.1.65. Congruence modulo 3 is an equivalence relation on $\mathbb{Z}$ with equivalence classes

$$[0] = \{3q + 0 \mid q \in \mathbb{Z}\} = \{\dots, -6, -3, 0, 3, 6, \dots\},$$
$$[1] = \{3q + 1 \mid q \in \mathbb{Z}\} = \{\dots, -5, -2, 1, 4, 7, \dots\}, \text{ and}$$
$$[2] = \{3q + 2 \mid q \in \mathbb{Z}\} = \{\dots, -4, -1, 2, 5, 8, \dots\}.$$

By Proposition 0.1.50, these are all of the distinct equivalence classes of $\mathbb{Z}$ modulo 3. Even more, by Proposition 0.1.51, we obtain a partition of $\mathbb{Z}$ into distinct equivalence classes modulo 3

$$\mathbb{Z} = [0] \cup [1] \cup [2] = \{\dots, -6, -3, 0, 3, 6, \dots\} \cup \{\dots, -5, -2, 1, 4, 7, \dots\} \cup \{\dots, -4, -1, 2, 5, 8, \dots\}.$$

Each of the preceding examples is illustrative of the general structure of the equivalence classes of the integers modulo a nonzero integer n. Concretely, for any nonzero integer n, there are n distinct equivalence classes of the integers modulo n, and each class consists of sums of integer multiples of n and a non-negative integer that is strictly smaller than n. We remark that the proof of this fact follows by the Division Algorithm, hence we will not endeavor to provide such justification at the moment; however, the reader should consider how the result makes sense intuitively according to the process of integer division, quotients, remainders, and the definition of congruence modulo n.

Proposition 0.1.66. *Given any nonzero integer n, there are exactly n distinct equivalence classes of $\mathbb{Z}$ modulo n defined by $[r] = \{nq + r \mid q \in \mathbb{Z}\}$ for each integer $0 \leq r \leq n-1$. Consequently, every nonzero integer n induces a partition of the integers into distinct equivalence classes modulo n*

$$\mathbb{Z} = \bigcup_{r=0}^{n-1} \{nq + r \mid q \in \mathbb{Z}\}.$$

Congruence modulo a nonzero integer also gives rise to other interesting equivalence relations.

Example 0.1.67. Consider the relation R defined on the set $\mathbb{Z}$ of integers such that $a \; R \; b$ if and only if $5b \equiv 2a \pmod 3$ for any integers a and b. We claim that R is an equivalence relation.

1.) We must first establish that $a \; R \; a$ for all integers a. By definition of R, we must prove that $5a \equiv 2a \pmod 3$. But this is true because $5a - 2a = 3a$ is divisible by 3 for all integers a.

2.) We establish next that if $a \; R \; b$, then $b \; R \; a$. By definition of R, if $a \; R \; b$, then $5b \equiv 2a \pmod 3$ so that $5b - 2a = 3k$ for some integer k. Consequently, we have that $2a - 5b = 3(-k)$. By adding $3a$ and $3b$ to both sides of this equation, we obtain $5a - 2b = 3(-k) + 3a + 3b = 3(-k + a + b)$. We conclude that $5a - 2b$ is divisible by 3 so that $5a \equiv 2b \pmod 3$ and $b \; R \; a$.

3.) Last, if $a \; R \; b$ and $b \; R \; c$, then $5b \equiv 2a \pmod 3$ and $5c \equiv 2b \pmod 3$. By definition, there exist integers k and ℓ such that $5b - 2a = 3k$ and $5c - 2b = 3\ell$. By taking their sum, we find that

$$5c - 3b - 2a = (5c - 2b) + (5b - 2a) = 3\ell + 3k = 3(\ell + k)$$

so that $5c - 2a = 3(\ell + k + b)$; therefore, 3 divides $5c - 2a$ so that $5c \equiv 2a \pmod 3$ and $a \; R \; c$.

By definition of an equivalence class of $\mathbb{Z}$ modulo R, the equivalence class of a modulo R is simply

$$[a] = \{b \in \mathbb{Z} \mid a \; R \; b\} = \{b \in \mathbb{Z} \mid 5b \equiv 2a \pmod 3\} = \{b \in \mathbb{Z} \mid 5b - 2a = 3k \text{ for some integer } k\}.$$

Consequently, the class of a modulo R is $[a] = \{b \in \mathbb{Z} \mid 5b = 3k + 2a \text{ for some integer } k\}$. Checking some small values of b yields that $[0] = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$. Likewise, by definition and a brute-force check, we have that $[1] = \{b \in \mathbb{Z} \mid 5b = 3k + 2 \text{ for some integer } k\} = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$ and $[2] = \{b \in \mathbb{Z} \mid 5b = 3k + 4 \text{ for some integer } k\} = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$. Every integer belongs to one of these three distinct equivalence classes modulo R, hence this is an exhaustive list.

0.1.6 The Definition of a Function

Consider any sets X and Y. We have seen previously that a relation from X to Y is any subset of the Cartesian product $X \times Y$. We will distinguish a relation f from X to Y as a **function** if and only if every element of X is the first coordinate of one and only one ordered pair in f. Explicitly, a function $f : X \to Y$ is merely an assignment of each element $x \in X$ to a unique (but not necessarily distinct) element $f(x) \in Y$ called the **direct image** of x under f. We refer to the set X as the **domain** of $f : X \to Y$; the **codomain** of f is Y; and the **range** of f is the set $\text{range}(f) = \{f(x) \mid x \in X\}$ of second coordinates of elements in f. Out of desire for notational convenience, we may sometimes omit the letter $f : X \to Y$ when defining a function and simply use an arrow $X \to Y$ to indicate the sets involved and an arrow $x \mapsto y$ to declare the direct image $y \in Y$ of the element $x \in X$.

Example 0.1.68. Consider the relation $f = \{(-1, 1), (1, -1)\}$ defined on the set $X = \{-1, 1\}$. Each of the elements of X is the first coordinate of one and only one ordered pair in f, hence $f : X \to X$ is a function; its domain and range are both X. Conventionally, we might recognize this function as $f(x) = -x$ because it has the effect of swapping the signs of each element $x \in X$.

Example 0.1.69. Consider the relation $g = \{(x, x-1) \mid x \in \mathbb{R}\}$ on the collection $\mathbb{R}$ of real numbers. Every real number is the first coordinate of one and only one ordered pair in g, hence $g : \mathbb{R} \to \mathbb{R}$ is a function; its domain and range are both $\mathbb{R}$. Conventionally, we might recognize this function as $g(x) = x - 1$ because the ordered pairs $(x, y) \in g$ satisfy that $y = x - 1$ for each real number x.

Example 0.1.70. Often in calculus, a function is defined simply by declaring a rule, e.g., $h(x) = x^2$. Conventionally, the domain of such a function is assumed to be the **natural domain**, i.e., the largest subset of the real numbers for which $h(x)$ can be defined. Considering that the square of any real number is itself a real number, it follows that the domain of $h(x)$ is all real numbers; the range of $h(x)$ is the collection of all non-negative real numbers because if $x \in \mathbb{R}$, then $x^2 \geq 0$.

But strictly speaking, in general, a function depends intimately on its domain and its codomain. We will soon see that the functions $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by $h(x) = x^2$ and $k : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined by $k(x) = x^2$ are quite different from one another, all though the underlying **rule** of both functions is the same. Even more, both of these functions are different from $\ell : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $\ell(x) = x^2$.

Example 0.1.71. Consider the equivalence relation R defined on the set $\{1, 2, 3, 4, 5\}$ as in Example 0.1.44. Crucially, we note that R is not a function since the ordered pairs $(1, 1)$ and $(1, 3)$ lie in R. Generally, an equivalence relation R will never be a function because if (x, y) and (y, x) both lie in R, then by definition, we must have that $(x, x) \in R$ so that R is not a function.

Every set X admits an **identity function** $\mathrm{id}_X : X \to X$ defined by $\mathrm{id}_X(x) = x$. 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 appearing to the left of the arrow $\mapsto$ is viewed as an element of X while the symbol x appearing to the right of the arrow $\mapsto$ is viewed as an element of Y; in the usual notation, the inclusion may be thought of as the function $i : X \to Y$ defined by $i(x) = x$. 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 $\delta_X(x) = (x, x)$. Later in the course, we will prove that the diagonal Δ_X of X is the direct image of the diagonal function δ_X of X, hence there should be no confusion in terminologies.

Even if we have never thought of it as such, algebraic operations such as addition, subtraction, multiplication, and division can be viewed as functions. Explicitly, addition of real numbers is the function $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $(x, y) \mapsto x + y$. Crucially, the sum of two real numbers is a real number, hence this function is **well-defined**, i.e., the image of every element lies in the codomain of the function. Generally, if X is any set, the function $* : X \times X \to X$ that sends $(x, y) \mapsto x * y$ is a **binary operation** if and only if $x * y$ is an element of X for every pair of elements $x, y \in X$. Like we mentioned, addition and multiplication are binary operations on the real numbers $\mathbb{R}$.

Consider any pair of functions $f : X \to Y$ and $g : X \to Y$. Given any element $x \in X$, there exist unique elements $f(x), g(x) \in Y$ such that $(x, f(x)) \in f$ and $(x, g(x)) \in g$. Consequently, if f and g are equal as sets so that $f = g$, then $(x, f(x))$ lies in g; the uniqueness of $g(x)$ yields in turn that $f(x) = g(x)$. Conversely, if $f(x) = g(x)$ for every element $x \in X$, then we have that

$$f = \{(x, f(x)) \mid x \in X\} = \{(x, g(x)) \mid x \in X\} = g$$

so that f and g are equal as sets; this establishes the following important fact about functions.

Proposition 0.1.72 (Equality of Functions)**.** *Given any sets X and Y, the functions $f : X \to Y$ and $g : X \to Y$ are equal as sets if and only if $f(x) = g(x)$ for all elements $x \in X$.*

Every time we define a function $f : X \to Y$, for every subset $V \subseteq X$, we implicitly distinguish the collection of elements $y \in Y$ such that $y = f(v)$ for some element $v \in V$; this is denoted by

$$f(V) = \{f(v) \mid v \in V\}$$

and called the **direct image** of V (in Y) under f. Conversely, if $W \subseteq Y$, the collection of elements $x \in X$ such that $f(x) \in W$ is the **inverse image** of W (in X) under f. Explicitly, we have that

$$f^{-1}(W) = \{x \in X \mid f(x) \in W\}.$$

Example 0.1.73. Consider the function $f = \{(u, 1), (v, 2), (w, 3), (x, 3), (y, 2), (z, 1)\}$ from the set $X = \{u, v, w, x, y, z\}$ to $Y = [6] = \{1, 2, 3, 4, 5, 6\}$. We have that $\text{range}(f) = \{1, 2, 3\}$, but it is just as true that $\text{range}(f) = f(\{u, v, w\}) = f(\{u, x, y\}) = f(\{x, y, z\})$. Even more, we have that

$$f^{-1}(\{2, 3\}) = \{v, w, x, y\} \text{ and } f^{-1}(\{4, 5, 6\}) = \varnothing$$

because the elements $4, 5, 6 \in Y$ do not belong to the second component of any ordered pair in f.

Example 0.1.74. Consider the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$. Observe that for any real number x such that $-1 \leq x \leq 1$, we have that $0 \leq x^2 \leq 1$, hence it follows that $g([-1, 1]) = [0, 1]$. Likewise, if $x^2 = g(x) \geq 4$, then $x \geq 2$ or $x \leq -2$ so that $g^{-1}([4, \infty)) = (-\infty, -2] \cup [2, \infty)$.

Even if the sets X and Y are finite with small cardinalities $|X|$ and $|Y|$, the number of functions $f : X \to Y$ grows astonishingly quickly. Explicitly, a function $f : X \to Y$ is uniquely determined by choosing for each element $x \in X$ one and only one element $y \in Y$ such that $f(x) = y$. Consequently, for each element $x \in X$, there are $|Y|$ possible choices for $f(x)$. By denoting the set of functions $f : X \to Y$ as $Y^X = \{f \subseteq X \times Y \mid f : X \to Y \text{ is a function}\}$, we have that $|Y^X| = |Y|^{|X|}$.

Example 0.1.75. Consider the sets $X = \{u, v, w, x, y, z\}$ and $Y = [6] = \{1, 2, 3, 4, 5, 6\}$ of Example 0.1.73. We have that $|X| = 6 = |Y|$, hence there are $|Y|^{|X|} = 6^6$ possible functions $f : X \to Y$.

0.1.7 One-to-One and Onto Functions

We introduce in this section two indispensable properties of a function $f : X \to Y$ from a set X to a set Y. We say that f is **one-to-one** (or **injective**) if every pair of distinct elements $x_1, x_2 \in X$ induces distinct elements $f(x_1), f(x_2) \in Y$. Equivalently, we say that f is one-to-one if every equality $f(x_1) = f(x_2)$ of elements of Y yields the corresponding equality $x_1 = x_2$ of elements of X.

Example 0.1.76. Consider the function $f = \{(-1, 1), (1, -1)\}$ from the set $X = \{-1, 1\}$ to itself. Each of the elements $x \in X$ corresponds to a distinct element $f(x) \in X$, hence f is one-to-one.

Example 0.1.77. Consider the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 3x + 4$. Observe that if $f(x_1) = f(x_2)$, then $3x_1 + 4 = 3x_2 + 4$ so that $3x_1 = 3x_2$ and $x_1 = x_2$; thus, f is one-to-one.

Example 0.1.78. Consider the real function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined by $f(x) = x^2$. Observe that if $f(x_1) = f(x_2)$, then $x_1^2 = x_2^2$. By taking the square root of both sides and using the fact that the domain of f consists of non-negative real numbers, it follows that $x_1 = x_2$ so that f is one-to-one.

Example 0.1.79. Consider the real function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$. Considering that $g(-1) = 1 = g(1)$ but $-1 \neq 1$, it follows that g is not one-to-one. Compare with Example 0.1.78.

Example 0.1.80. Consider the function $f = \{(u, 1), (v, 2), (w, 3), (x, 3), (y, 2), (z, 1)\}$ from the set $X = \{u, v, w, x, y, z\}$ to $Y = [6] = \{1, 2, 3, 4, 5, 6\}$. Considering that $f(u) = 1 = f(z)$ but $u \neq z$, it follows that f is not one-to-one; the same holds for $f(v) = 2 = f(y)$ and $f(w) = 3 = f(x)$.

Even more, we say that $f : X \to Y$ is **onto** (or **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 the codomain Y is "mapped onto" or "covered" by an element of X. Even more simply, a function $f : X \to Y$ is surjective if and only if $Y = \text{range}(f) = \{f(x) \mid x \in X\}$.

Example 0.1.81. Consider the function $f = \{(-1, 1), (1, -1)\}$ from the set $X = \{-1, 1\}$ to itself. Each of the elements $y \in X$ can be written as $y = f(x)$ for some element $x \in X$, hence f is onto.

Example 0.1.82. Consider the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 3x + 4$ of Example 0.1.77. We claim that f is onto, hence for any real number y, we require a real number x such that $y = f(x) = 3x + 4$. By solving for x in $y = 3x + 4$, we find that $x = \frac{1}{3}(y - 4)$. Computing $f(x)$ yields

$$f(x) = 3x + 4 = 3 \cdot \frac{1}{3}(y - 4) + 4 = (y - 4) + 4 = y$$

because $x = \frac{1}{3}(y - 4)$ by construction, as desired. Consequently, it follows that f is onto.

Example 0.1.83. Consider the real function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ defined by $f(x) = x^2$. Given any real number $y \geq 0$, we claim that there exists a real number x such that $y = x^2$. By taking $x = \sqrt{y}$ (this is well-defined because $y \geq 0$), it follows that $f(x) = x^2 = (\sqrt{y})^2 = y$ so that f is onto.

Example 0.1.84. Consider the function $f = \{(u, 1), (v, 2), (w, 3), (x, 3), (y, 2), (z, 1)\}$ from the set $X = \{u, v, w, x, y, z\}$ to $Y = [6] = \{1, 2, 3, 4, 5, 6\}$ as in Example 0.1.80. Considering that 4, 5, and 6 are not the image of any element of X under f, it follows that f is not onto.

Example 0.1.85. Consider the sets $X = \{a, b, c\}$ and $Y = \{0, 1, 2, 3\}$. We cannot possibly find a function $f : X \to Y$ that is onto because the cardinality of X is strictly smaller than the cardinality of Y; therefore, it is impossible to assign to each element $y \in Y$ a unique element $x \in X$.

We say that a function $f : X \to Y$ is **bijective** if f is both injective and surjective. We may think of a bijection $f : X \to Y$ simply as a relabelling of the elements of Y by the names of elements of X; in this way, two sets X and Y are "essentially the same" if there exists a bijection $f : X \to Y$. Often, this property of a bijective function is emphasized in the literature by using the terminology of "one-to-one correspondence" between X and Y rather than a "bijection" from X to Y.

Proposition 0.1.86. *Consider any pair of arbitrary finite sets X and Y.*

1.) *If there exists an injective function $f : X \to Y$, then $|X| \leq |Y|$.*

2.) *If $|X| \leq |Y|$, then there exists an injective function $f : X \to Y$.*

3.) *If there exists a surjective function* $f : X \to Y$, *then* $|X| \geq |Y|$.

4.) *If* $|X| \geq |Y|$, *then there exists a surjective function* $f : X \to Y$.

5.) *If there exists a bijective function* $f : X \to Y$, *then* $|X| = |Y|$.

6.) *If* $|X| = |Y|$, *then there exists a bijective function* $f : X \to Y$.

7.) *If* $|X| = |Y|$, *then a function* $f : X \to Y$ *is injective if and only if it is surjective.*

Proof. We will assume throughout the proof that $|X| = m$ and $|Y| = n$ are non-negative integers. Certainly, if either m or n is zero, then the empty function satisfies the desired properties. Consequently, we may assume that neither m nor n is zero. We will assume also for notational convenience that $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$. We turn our attention to each claim in turn.

(1.) We will assume that there exists an injective function $f : X \to Y$. Consequently, every element $y \in Y$ is obtained from at most one element $x \in X$ via $y = f(x)$. Considering that every element $x \in X$ corresponds to a unique element $f(x) \in Y$, we conclude that $|X| \leq |Y|$.

(2.) Observe that if $m \leq n$, then we may define an injective function $f : X \to Y$ by declaring that $f(x_i) = y_i$ for each integer $1 \leq i \leq m$. Explicitly, f is a function because every element $x_i \in X$ corresponds to exactly one element $y_i = f(x_i) \in Y$. Even more, f is injective since for each element $y_i \in Y$, there is at most one element $x_i \in X$ such that $y_i = f(x_i)$ by assumption that $n \geq m$.

(3.) We will assume that there exists a surjective function $f : X \to Y$. Consequently, for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. Considering that every element $x \in X$ corresponds to a unique element $f(x) \in Y$, we conclude that $|X| \geq |Y|$.

(4.) Conversely, if $m \geq n$, then we may define a surjective function $f : X \to Y$ by declaring that $f(x_i) = y_i$ for each integer $1 \leq i \leq m$. We have already seen in the previous paragraph that such a relation is a function; however, by assumption that $m \geq n$, it follows that f is surjective because for every element $y_i \in Y$, there exists an element $x_i \in X$ such that $y_i = f(x_i)$.

(5.) Combined, parts (a.) and (c.) imply that $|X| \leq |Y|$ and $|X| \geq |Y|$ so that $|X| = |Y|$.

(6.) Combined, parts (b.) and (d.) yield a bijective function $f : X \to Y$ defined by $f(x_i) = y_i$.

(7.) Last, we will assume that $m = n$. Consider any function $f : X \to Y$. Observe that if f is injective, then every element of X maps to a distinct element of Y under f, hence range(f) is a subset of Y with the same cardinality as Y. We conclude that range$(f) = Y$ so that f is surjective. Conversely, if f is surjective, then for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. By assumption that $m = n$, the element $x \in X$ such that $y = f(x)$ must be uniquely determined by y, hence the image of $x \in X$ under f is unique so that f is injective. □

Caution: if X and Y are infinite sets, then there need not exist a bijective function $f : X \to Y$. Explicitly, there is no bijection $f : \mathbb{Q} \to \mathbb{R}$ between the rational numbers and the real numbers.

Caution: if X and Y are infinite sets, then a function $f : X \to Y$ can be injective without being surjective (and vice-versa). Explicitly, the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = 2x$ is injective but not surjective, and the function $g : \mathbb{Q} \to \mathbb{Z}$ defined by $g(p/q) = p$ is surjective but not injective.

Example 0.1.87. Consider the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = -x$. Cancelling a minus sign, we conclude that if $f(x) = f(y)$, then $-x = -y$ yields that $x = y$ so that f is one-to-one. Likewise, every integer n is the image of $-n$ under f because $n = -(-n) = f(-n)$, hence f is onto.

Example 0.1.88. Consider the rational function $f : \mathbb{R} \setminus \{3\} \to \mathbb{R} \setminus \{1\}$ defined by

$$f(x) = \frac{x-2}{x-3}.$$

Cross-multiplying denominators, we note that $f(x) = f(y)$ if and only if $(x-2)(y-3) = (x-3)(y-2)$ if and only if $xy - 3x - 2y + 6 = xy - 2x - 3y + 6$ if and only if $x = y$, hence f is one-to-one. Conversely, we will prove that f is onto. Behind the scenes, we solve the following equation for x.

$$y = \frac{x-2}{x-3}$$

Observe that this identity holds if and only if $(x-3)y = x-2$ if and only if $xy - 3y = x - 2$ if and only if $xy - x = 3y - 2$ if and only if $x(y-1) = 3y - 2$ if and only if

$$x = \frac{3y-2}{y-1}.$$

Consequently, for every real number $y \in \mathbb{R} \setminus \{1\}$, we have that $y = f(x)$ so that f is onto.

Example 0.1.89. Consider the equivalence relation R_6 of congruence modulo 6 defined on the set $\mathbb{Z}$ of integers. Conventionally, the collection of equivalence classes of $\mathbb{Z}$ modulo 6 is denoted $\mathbb{Z}/6\mathbb{Z}$. Every element of $\mathbb{Z}/6\mathbb{Z}$ is the equivalence class $[a] = \{b \in \mathbb{Z} \mid b \equiv a \pmod 6\}$ of an integer a modulo 6, hence there are six distinct elements of $\mathbb{Z}/6\mathbb{Z}$ by Proposition 0.1.62. We will demonstrate in this example how to define a function from $\mathbb{Z}/6\mathbb{Z}$ to itself. We may define a relation $f : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$ by declaring that $f([x]) = [5x+3]$. By definition, in order to establish that f is a function, we must verify that if $[5x+3]$ and $[5y+3]$ are distinct, then $[x]$ and $[y]$ are distinct. Concretely, this ensures that the function f passes the Vertical Line Test. Consequently, we may assume that $[x] = [y]$ and derive $[5x+3] = [5y+3]$. (Why?) By Equality of Equivalence Classes, it follows that $y \equiv x \pmod 6$ so that 6 divides $y - x$ by the Properties of Congruence Modulo n. By definition of divides, there exists an integer k such that $y - x = 6k$; in turn, this yields the divisibility relation

$$(5y+3) - (5x+3) = 5(y-x) = 6(5k).$$

Considering that $5k$ is an integer, we conclude that 6 divides the integer $(5y+3) - (5x+3)$ so that $5y+3 \equiv 5x+3 \pmod 6$. Once again, by Proposition 0.1.49, we conclude that $[5x+3] = [5y+3]$. We say in this case that the relation f is a **well-defined** function. Quite to our delight, it happens that f is a bijection: indeed, we have that $[0] = f([3])$, $[1] = f([2])$, $[2] = f([1])$, $[3] = f([0])$, $[4] = f([5])$, and $[5] = f([4])$ so that f is both injective and surjective. (One can prove this algebraically.)

0.1.8 Composition of Functions

Every pair of functions $f : X \to Y$ and $g : Y \to Z$ from any three sets X, Y, and Z give rise to a third function $g \circ f : X \to Z$ called the **composite function** defined by $(g \circ f)(x) = g(f(x))$. We may also refer to the function $g \circ f$ as g **composed with** f or the **composition** of f under g.

Example 0.1.90. Consider the sets $X = \{-1, 1\}$, $Y = \{x, y, z\}$, and $Z = \{1, 2, 3\}$. We may define a pair of functions $f : X \to Y$ and $g : Y \to Z$ by $f = \{(-1, x), (1, z)\}$ and $g = \{(x, 2), (y, 3), (z, 1)\}$. Observe that the composite function $g \circ f : X \to Z$ satisfies $(g \circ f)(-1) = g(f(-1)) = g(x) = 2$ and $(g \circ f)(1) = g(f(1)) = g(z) = 1$. Consequently, we find that $g \circ f = \{(-1, 2), (1, 1)\}$.

Example 0.1.91. Consider the sets $A = \{a, b, c, d\}$, $B = \{b, c, d, e\}$, and $C = \{c, d, e, f\}$. We may define a pair of functions $f : A \to B$ and $g : B \to C$ such that $f = \{(a, b), (b, c), (c, d), (d, e)\}$ and $g = \{(b, c), (c, d), (d, e), (e, f)\}$. Observe that the composite function $g \circ f : A \to C$ satisfies that

$$(g \circ f)(a) = g(f(a)) = g(b) = c, \qquad (g \circ f)(c) = g(f(c)) = g(d) = e, \text{ and}$$
$$(g \circ f)(b) = g(f(b)) = g(c) = d, \qquad (g \circ f)(d) = g(f(d)) = g(e) = f.$$

Consequently, we find that $g \circ f : A \to C$ satisfies that $g \circ f = \{(a, c), (b, d), (c, e), (d, f)\}$.

Example 0.1.92. Composition of functions ought to be a familiar concept from calculus: indeed, the Chain Rule for Derivatives gives a formula for the derivative of a composite function. Consider the functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = e^x$ and $g(x) = |x|$. We have that

$$f \circ g : \mathbb{R} \to \mathbb{R} \text{ is defined by } (f \circ g)(x) = f(g(x)) = e^{g(x)} = e^{|x|} \text{ and}$$
$$g \circ f : \mathbb{R} \to \mathbb{R} \text{ is defined by } (g \circ f)(x) = g(f(x)) = |f(x)| = |e^x| = e^x.$$

Crucially, the latter holds because $e^x > 0$ for all real numbers x, hence it follows that $g \circ f = f$ as real functions. On the other hand, for the real identity function $\mathrm{id}_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ defined by $\mathrm{id}_{\mathbb{R}}(x) = x$, we note that $\mathrm{id}_{\mathbb{R}} \circ f = f$ since $(\mathrm{id}_{\mathbb{R}} \circ f)(x) = \mathrm{id}_{\mathbb{R}}(f(x)) = f(x)$ for all real numbers x. Comparing the two identities derived in this example yields that $g \circ f = \mathrm{id}_{\mathbb{R}} \circ f$; however, it is not the case that $g = \mathrm{id}_{\mathbb{R}}$ because $g(-1) = 1 \neq -1 = \mathrm{id}_{\mathbb{R}}(-1)$. Consequently, we obtain the following important fact.

Proposition 0.1.93 (Function Composition Is Not Cancellative)**.** *Given any quadruple of functions $f : X \to Y$, $g : X \to Y$, $h : Y \to Z$, and $j : Y \to Z$ such that $h \circ f = j \circ f$ and $h \circ f = h \circ g$, we cannot conclude that either $h = j$ or $f = g$. Put another way, function composition is not cancellative.*

Proof. We leave it to the reader to adapt the approach of Example 0.1.92 to determine sets X, Y, and Z and distinct functions $f : X \to Y$, $g : X \to Y$, and $h : Y \to Z$ such that $h \circ f = h \circ g$. □

Even though function composition is not cancellative, we will soon come to find that it satisfies several important properties that make it an indispensable operation in the theory of functions.

Proposition 0.1.94 (Composition of Functions Preserves Injectivity and Surjectivity)**.** *Consider any pair of functions $f : X \to Y$ and $g : Y \to Z$.*

1.) *If f and g are injective, then $g \circ f$ is injective.*

2.) *If f and g are surjective, then $g \circ f$ is surjective.*

Put another way, composition of functions preserves injectivity and surjectivity.

Proof. (1.) We must prove that if $(g \circ f)(x_1) = (g \circ f)(x_2)$, then $x_1 = x_2$. By assumption that g is injective, if $g(f(x_1)) = (g \circ f)(x_1) = (g \circ f)(x_2) = g(f(x_2))$, then $f(x_1) = f(x_2)$. But by the same rationale applied to the injective function f, we conclude that $x_1 = x_2$, as desired.

(2.) We must prove that for every element $z \in Z$, we have that $z = (g \circ f)(x)$ for some element $x \in X$. By assumption that g is surjective, for every element $z \in Z$, there exists an element $y \in Y$ such that $z = g(y)$. Even more, by hypothesis that f is surjective, there exists an element $x \in X$ such that $y = f(x)$. Combined, these observations yield that $z = g(y) = g(f(x)) = (g \circ f)(x)$. □

Corollary 0.1.95 (Composition of Bijective Functions Is Bijective)**.** *Given any bijective functions $f : X \to Y$ and $g : Y \to Z$, the composite function $g \circ f : X \to Z$ is bijective.*

Proof. Both f and g are injective and surjective, so $g \circ f$ is injective and surjective. □

Proposition 0.1.96 (Composition of Functions Is Associative)**.** *Consider any triple of functions $f : W \to X$, $g : X \to Y$, and $h : Y \to Z$. We have that $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. We must prove that $[h \circ (g \circ f)](w) = [(h \circ g) \circ f](w)$ for all elements $w \in W$ by Proposition 0.1.72. We will assume that $f(w) = x$, $g(x) = y$, and $h(y) = z$. By definition of the composite function, we have that $(g \circ f)(w) = g(f(w)) = g(x) = y$ and $(h \circ g)(x) = h(g(x)) = h(y) = z$ so that $[h \circ (g \circ f)](w) = h((g \circ f)(w)) = h(y) = z$ and $[(h \circ g) \circ f](w) = (h \circ g)(f(w)) = (h \circ g)(x) = z$. □

Remark 0.1.97. We note that in order to define the composition $g \circ f$ of any function $f : X \to Y$ under any other function $g : Y \to Z$, it is sufficient but not strictly necessary to assume that the domain of g contains the codomain of f. Generally, the composite function $g \circ f$ is well-defined for any function $g : W \to Z$ so long as $W \supseteq \text{range}(f)$. Consider to this end the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$, we have that $\text{range}(f) = \{f(x) \mid x \in \mathbb{R}\} = \{x^2 \mid x \in \mathbb{R}\} = \mathbb{R}_{\geq 0}$, hence for any function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$, the composite function $g \circ f$ is well-defined. Explicitly, if we assume that $g(x) = \sqrt{x}$, then $(g \circ f)(x) = g(f(x)) = g(x^2) = \sqrt{x^2} = |x|$; however, if $g(x) = \ln(x)$ on its natural domain, then the composite function $g \circ f$ is not well-defined because $\ln(0)$ is not well-defined.

Proposition 0.1.98 (Composition of Functions Is Not Commutative)**.** *Given any pair of functions $f : X \to Y$ and $g : Y \to Z$, we cannot conclude that $g \circ f = f \circ g$.*

Proof. By the preceding remark, we must have that $Y \supseteq \text{range}(f)$, so if this is not the case, then $g \circ f$ is not well-defined. We may assume therefore that $Y \supseteq \text{range}(f)$ and $X \supseteq \text{range}(g)$ so that $g \circ f$ and $f \circ g$ are both well-defined. Consider the real functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x + 1$ and $g(x) = 2x - 1$. Certainly, the reader can verify that f and g are both bijective functions (indeed, the graphs of f and g are lines of slope 2), hence the condition that the domain of g contains the codomain of f and vice-versa hold. Likewise, it is simple to check that

$$(f \circ g)(x) = f(g(x)) = f(2x - 1) = 2(2x - 1) + 1 = 4x - 1 \text{ and}$$
$$(g \circ f)(x) = g(f(x)) = g(2x + 1) = 2(2x + 1) - 1 = 4x + 1.$$

Considering that $(f \circ g)(0) = -1$ and $(g \circ f)(0) = 1$ are not equal, $f \circ g$ and $g \circ f$ are not equal. □

0.1.9 Inverse Functions

Considering that any function $f : X \to Y$ between two sets X and Y is by definition a relation, there exists an inverse relation f^{-1} from Y to X defined by $f^{-1} = \{(y, x) \mid (x, y) \in f\}$. One natural curiosity regarding the nature of the inverse relation f^{-1} of a function f is to ask whether the inverse relation f^{-1} of a function f must be a function. Generally, the answer is no.

Proposition 0.1.99 (The Inverse Relation of a Function Is Not Necessarily a Function)**.** *Given any function $f : X \to Y$, the inverse relation $f^{-1} : Y \to X$ is not necessarily a function.*

Proof. Consider the relation $f = \{(-1,1),(1,1)\}$ on the set $X = \{-1,1\}$. We leave it to the reader to verify that f is a function with inverse relation $f^{-1} = \{(1,-1),(1,1)\}$. We note that f^{-1} is not a function because $f^{-1}(1)$ is not well-defined since $(1,-1)$ and $(1,1)$ both lie in f^{-1}. □

Consequently, it would appear that in order for the inverse relation f^{-1} of a function $f : X \to Y$ to be a function, we require that every element $f(x) \in \text{range}(f)$ corresponds uniquely to an element $x \in X$. Put another way, we must have that f is injective. Conversely, by definition, if $f^{-1} : Y \to X$ is a function, then for every element $y \in Y$, we require that $f^{-1}(x)$ is an element of X. Explicitly, it must be the case that for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. Put another way, we must have that f is surjective. We are therefore lead to the following result.

Theorem 0.1.100 (Existence of an Inverse Function)**.** *Given any function $f : X \to Y$, the inverse relation f^{-1} is a function if and only if f is bijective. Even more, if f^{-1} is a function, it is bijective.*

Proof. Observe that if f is a bijective function, then for every element $y \in Y$, there exists a unique element $x \in X$ such that $y = f(x)$. Consequently, the inverse relation $f^{-1} : Y \to X$ of f defined by $(y,x) \in f^{-1}$ if and only if $y = f(x)$ is a function because for every element $y \in Y$, there exists one and only one element $x \in X$ such that $y = f(x)$ by hypothesis that f is a bijection. Concretely, for any pair of elements $(y,x_1),(y,x_2) \in f^{-1}$, we have that $f(x_1) = y = f(x_2)$ so that $x_1 = x_2$ since f is injective, and every element of Y is mapped onto an element of X by f^{-1} since f is surjective.

Conversely, suppose that the inverse relation $f^{-1} = \{(y,x) \mid (x,y) \in f\}$ of f is a function. By definition of a function, for every element $y \in Y$, there exists an element $x \in X$ such that $(y,x) \in f^{-1}$ so that $y = f(x)$. But this implies that f is surjective since every element of Y is the image of some element of X. Even more, for every element $y \in Y$, there exists a unique element $x \in X$ such that $(y,x) \in f^{-1}$ or $y = f(x)$; thus, if $(y,x_1),(y,x_2) \in f^{-1}$, then $x_1 = x_2$. By definition of the inverse relation, we find that if $f(x_1) = f(x_2)$, then $x_1 = x_2$, hence f is injective.

Last, we will demonstrate that if f^{-1} is a function, then f^{-1} is bijective. We will assume first that $f^{-1}(y_1) = f^{-1}(y_2)$. By the previous paragraph, if f^{-1} is a function, then f is surjective, hence there exist elements $x_1, x_2 \in X$ such that $y_1 = f(x_1)$ and $y_2 = f(x_2)$. By definition of the inverse relation, if $f^{-1}(y_1) = f^{-1}(y_2)$, then $x_1 = x_2$ so that $y_1 = f(x_1) = f(x_2) = y_2$. Even more, for every element $x \in X$, there exists one and only one element $y \in Y$ such that $y = f(x)$ since f is bijective. Consequently, for every element $x \in X$, there exists an element $y \in Y$ such that $x = f^{-1}(y)$. □

Remark 0.1.101 (Construction of Inverse Functions)**.** By the previous theorem, the inverse relation f^{-1} of a function $f : X \to Y$ is a function if and only if f is bijective. We demonstrate next that if f is injective but not surjective, it is possible to construct an inverse function related to f. Crucially, every function $f : X \to Y$ **restricts** to a surjective function $F : X \to f(X)$ defined by $F(x) = f(x)$ with the same domain as f but whose codomain is the range of f. Consequently, if $f : X \to Y$ is injective, then $F : X \to f(X)$ is bijective, hence the inverse relation $F^{-1} : f(X) \to X$ is a function. Conversely, even if $f : X \to Y$ is not injective, we may modify the domain of f to obtain a bijection. Consider the set X_f of $x \in X$ such that for every pair of elements $x_1, x_2 \in X_f$, we have that $x_1 = x_2$ if $f(x_1) = f(x_2)$. One can verify that $\tilde{f} : X_f \to f(X_f)$ defined by $\tilde{f}(x) = f(x)$ is bijective.

We illustrate these concepts in the following examples. Consider the real function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$. Observe that $\sqrt{x} \geq 0$ for all real numbers $x \geq 0$, hence f is not surjective: indeed, we have that $f(\mathbb{R}_{\geq 0}) = \mathbb{R}_{\geq 0}$. Consequently, the induced function $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined

by $F(x) = \sqrt{x}$ is a bijection. Likewise, the real function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ is neither injective nor surjective since $g(-1) = 1 = g(1)$ and $x^2 \geq 0$ implies that $g(\mathbb{R}) = \mathbb{R}_{\geq 0}$. On the other hand, the induced function $\tilde{g} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $\tilde{g}(x) = x^2$ is a bijection: indeed, for any real numbers $x_1^2 = x_2^2$ such that $x_1, x_2 \geq 0$, we must have that $x_1 = x_2$ since $-x_1, -x_2 \leq 0$.

Once we have identified that a function $f : X \to Y$ admits an inverse function $f^{-1} : Y \to X$, we seek an explicit definition of that inverse function. We achieve this via the following proposition.

Proposition 0.1.102 (Construction and Uniqueness of Inverse Functions)**.** *Given any bijective function $f : X \to Y$, the inverse function $f^{-1} : Y \to X$ satisfies that $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$. Conversely, if $g : Y \to X$ is any function such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, then we must have that $g = f^{-1}$. Put another way, the inverse function $f^{-1} : Y \to X$ corresponding to any bijective function $f : X \to Y$ is the unique function $g : Y \to X$ satisfying that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$.*

Proof. Given any bijective function $f : X \to Y$, the inverse relation $f^{-1} : Y \to X$ is a function by Theorem 0.1.100. By definition of f^{-1}, we have that $f^{-1}(f(x)) = x = \mathrm{id}_X(x)$ for every element $x \in X$ so that $f^{-1} \circ f = \mathrm{id}_X$ by Proposition 0.1.72. Likewise, we have that $f(f^{-1}(y)) = y = \mathrm{id}_Y(y)$ for every element $y \in Y$ so that $f \circ f^{-1} = \mathrm{id}_Y$. We will assume next that $g : Y \to X$ is any function such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. By Proposition 0.1.96, we have that

$$g(y) = (g \circ \mathrm{id}_Y)(y) = [g \circ (f \circ f^{-1})](y) = [(g \circ f) \circ f^{-1}](y) = (\mathrm{id}_X \circ f^{-1})(y) = f^{-1}(y)$$

for every element $y \in Y$. We leave it to the reader to verify that if $f \circ g = \mathrm{id}_Y$, then $g = f^{-1}$. □

Remark 0.1.103. Generally, Proposition 0.1.102 provides an algorithm for determining the inverse function $f^{-1} : Y \to X$ of any function $f : X \to Y$ that can be defined by an explicit rule $y = f(x)$. Explicitly, we may solve the equation $y = f(x)$ in terms of x to find that $x = f^{-1}(y)$.

Example 0.1.104. We proved in Example 0.1.87 that the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = -x$ is bijective; its inverse function $f^{-1} : \mathbb{Z} \to \mathbb{Z}$ is defined by $f^{-1}(x) = -x$.

Example 0.1.105. We proved in Example 0.1.88 that the function $f : \mathbb{R} \setminus \{3\} \to \mathbb{R} \setminus \{1\}$ with

$$f(x) = \frac{x-2}{x-3}$$

is bijective. Observe that its inverse function is $f^{-1} : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{3\}$ defined by

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

Concretely, this formula is obtained by solving the equation $y = f(x)$ in terms of $x = f^{-1}(y)$ and subsequently replacing each instance of the symbol y with the symbol x. We encourage the reader to revisit the aforementioned example or proceed to the next example for the details.

Example 0.1.106. Consider the rational function $f : \mathbb{R} \setminus \{2\} \to \mathbb{R} \setminus \{1\}$ defined by

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

We may solve the equation $y = f(x)$ to find a function $x = f^{-1}(y)$ that is the inverse of f.

$$y = f(x) = \frac{2x+3}{2x-4}$$

$$2xy - 4y = 2x + 3$$

$$2xy - 2x = 4y + 3$$

$$x(2y-2) = 4y + 3$$

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

Consequently, we obtain a rational function $f^{-1} : \mathbb{R} \setminus \{1\} \to \mathbb{R} \setminus \{2\}$ defined by

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

We will verify that $(f^{-1} \circ f)(x) = x$ for all real numbers $x \neq 2$ and $(f \circ f^{-1})(x) = x$ for all real numbers $x \neq 1$. By Proposition 0.1.107, we will conclude that f^{-1} is the inverse of f.

$$(f^{-1} \circ f)(x) = f^{-1}(f(x)) = \frac{4f(x)+3}{2f(x)-2} = \frac{4 \cdot \frac{2x+3}{2x-4} + 3}{2 \cdot \frac{2x+3}{2x-4} - 2} = \frac{4(2x+3)+3(2x-4)}{2(2x+3)-2(2x-4)} = \frac{14x}{14} = x$$

$$(f \circ f^{-1})(x) = f(f^{-1}(x)) = \frac{2f^{-1}(x)+3}{2f^{-1}(x)-4} = \frac{2 \cdot \frac{4x+3}{2x-2} + 3}{2 \cdot \frac{4x+3}{2x-2} - 4} = \frac{2(4x+3)+3(2x-2)}{2(4x+3)-4(2x-2)} = \frac{14x}{14} = x$$

Currently, our strategy for computing the inverse function of a bijective function is somewhat backwards: in order to determine that the inverse relation of a function is a function, we must prove that the function is a bijection. But this requires us to establish that the function is onto, and this necessitates the computation of the inverse function. We make the process more efficient as follows.

Proposition 0.1.107. *Given any function $f : X \to Y$ such that there exists a function $g : Y \to X$ for which $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, it follows that f and g are bijections satisfying that $g = f^{-1}$.*

Proof. We will prove only that f is bijective. By Propositions 0.1.100 and 0.1.102, the result follows. Consider any elements $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$. By hypothesis, we have that

$$x_1 = \mathrm{id}_X(x_1) = (g \circ f)(x_1) = g(f(x_1)) = g(f(x_2)) = (g \circ f)(x_2) = \mathrm{id}_X(x_2) = x_2,$$

hence we conclude that f is injective. Conversely, for every element $y \in Y$, we have that

$$y = \mathrm{id}_Y(y) = (f \circ g)(y) = f(g(y)).$$

Considering that $g(y) = x$ is an element of X, we conclude that $y = f(x)$ so that f is surjective. $\square$

0.2 Logic and Truth Tables

Generally, the purpose of mathematics is to describe the universe quantitatively in a manner that is consistent, replicable, and unambiguous. Combined with the language of set theory, the calculus of logic provides the basis for mathematical communication: if sets, relations, and functions constitute the skeleton of some structure or organism that can be modelled mathematically, then the connective tissue is represented by (mathematical) statements, logical quantifiers, and truth tables. We introduce in this section several axioms and symbols that are commonplace in modern logic.

0.2.1 Statements

We have thus far garnered a working knowledge of set theory — including the theory of relations and functions — and we have seen examples of mathematical proofs. We turn our attention next to fleshing out some details regarding the calculus of logic that will soon assist us with writing original proofs. We will assume throughout this section that the symbols P and Q are **statements**, i.e., P and Q are complete sentences that assert a property that is unambiguously true (T) or false (F).

Example 0.2.1. "Every positive whole number is an integer" is an example of a true statement.

Example 0.2.2. "The integer 10 is divisible by 3" is an example of a false statement.

Example 0.2.3. "The weather in Kansas City is lovely this time of year" is not a statement because some might think so, but others might not: its truth value is ambiguous. Generally, any sentence that is exclamatory (e.g., any observation), imperative (e.g., any command), or interrogative (e.g., any question) is not a statement because these types of sentences have no inherent truth value.

Exclamatory: "What a story, Mark!"
Imperative: "Don't forget to mow the lawn."
Interrogative: "How about those Chiefs?"

We will henceforth refer to the verity of a statement as its **truth value**. Our ability to determine the truth value of a sentence does not preclude the possibility that the sentence is a valid statement; indeed, there are many unsolved statements throughout mathematics. Generally, a statement whose truth value is undetermined is called a **conjecture**. Common examples of mathematical statements with undetermined veracity include those that involve a potentially unknown or variable quantity x. We have encountered statements of these kinds throughout many of our mathematics courses.

Example 0.2.4. "The real number x is irrational" is an example of a valid statement; it is neither true nor false, but rather, its truth value depends explicitly on the value of the real number x.

Conventionally, any declarative statement of the form $P(x)$ for some variable quantity x is called an **open sentence**; the set of all possible values that x can assume is called the **domain** of x; and the truth value of $P(x)$ depends explicitly upon the determination of the variable x.

Example 0.2.5. Observe that the statement $P(x)$ that "the real number x is irrational" is an open sentence; the domain of x is the set of real numbers; and $P(x)$ is true if and only if $x \in \mathbb{R} \setminus \mathbb{Q}$.

We will typically represent an open sentence in the variable x by the symbol $P(x)$, and we will separate $P(x)$ from the open sentence it represents with a colon, as in the following example.

Example 0.2.6. Consider the following pair of open sentences.

$$P(x)\colon \text{ We have that } x^2 - 1 = 0.$$
$$Q(x)\colon \text{ We have that } x^2 + 1 = 0.$$

By solving for the unknown quantity x, we find that $P(x)$ is a true statement if and only if $x = \pm 1$, hence the natural domain for the statement $P(x)$ is the set $\mathbb{Z}$ of integers. Likewise, it follows that $Q(x)$ is a true statement if and only if $x = \pm\sqrt{-1}$, hence $Q(x)$ is false for any subset of real numbers because $\sqrt{-1}$ is not a real number. We conclude that the natural domain for $Q(x)$ is $\mathbb{C}$.

Example 0.2.7. Consider the following open sentence.

$$P(x, y)\colon \text{ We have that } x^2 + y^2 \geq 0$$

Considering that $x^2 + y^2 \geq 0 + 0 = 0$ for any pair of real numbers x and y, it follows that $P(x, y)$ is a true statement if the domain of x and y is any subset of the set $\mathbb{R}$ of real numbers; however, if the domains of x and y are both the set $\mathbb{C}$ of complex numbers, then we can determine values of x and y such that $P(x, y)$ is false. Concretely, we note that $P(i, i)$ is false since $i^2 + i^2 = -1 - 1 = -2 < 0$.

Example 0.2.8. Consider the following open sentence.

$$P(x, y)\colon \text{ We have that } x + y \text{ is a positive prime number.}$$

Let us assume throughout this example that the domain of x is $X = \{1, 2, 3, 4\}$ and the domain of y is $Y = \{-1, -2, -3, -4\}$. Calculating the sum $x + y$ for each of the sixteen elements of $X \times Y$, we find that $P(x, y)$ is true if and only if $(x, y) \in \{(3, -1), (4, -1), (4, -2)\}$; otherwise, $P(x, y)$ is false.

Often, it is convenient to collect the truth values of some finitely many statements $P_1, P_2, \ldots, P_n$ in a **truth table**. Each column of a truth table contains one statement followed by all of its possible truth values relative to the other statements. Concretely, the first row of a truth table contains the symbols that represent the statements, and the subsequent rows contain the possible truth values of each statement relative to the other. Considering that any statement attains one and only truth value, a truth table for the n statements $P_1, P_2, \ldots, P_n$ admits n columns and $2^n + 1$ rows as follows.

P
T
F

P	Q
T	T
T	F
F	T
F	F

P	Q	R
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Table 1: the truth tables for one, two, and three statements

0.2.2 Conjunction, Disjunction, and Negation

We examine next the myriad ways to construct new statements from any finite number of existing statements. We concern ourselves immediately with a statement P. We refer to the statement "not P" (precisely, "It is not the case that P") as the **negation** of P; symbolically, the negation of any statement P is denoted by $\neg P$. Often, it is possible to represent the negation $\neg P$ of a statement P in a less clunky way than simply by, "It is not the case that P," as the following examples illustrate.

Example 0.2.9. Consider the following statement.

$$P\text{: The integer 2 is even.}$$

By definition, the negation $\neg P$ of the given statement P is the following statement.

$$\neg P\text{: It is not the case that the integer 2 is even.}$$

Considering that any integer is either even or odd, we can rephrase $\neg P$ as follows.

$$\neg P\text{: The integer 2 is odd.}$$

Crucially, we note that P is a true statement, and its negation $\neg P$ is a false statement.

Example 0.2.10. Consider the following statement.

$$P\text{: The integer 111 is prime.}$$

We may express the negation $\neg P$ of the given statement P as follows.

$$\neg P\text{: The integer 111 is not prime.}$$

Better yet, since every integer is either prime or composite, we can rephrase $\neg P$ as follows.

$$\neg P\text{: The integer 111 is composite.}$$

Observe that in this case, the statement P is false, and its negation $\neg P$ is a true statement.

Generally, it ought to be clear to the reader that the statements P and $\neg P$ have opposite truth values: if P is true, then $\neg P$ must be false; however, if P is false, then $\neg P$ must be true.

P	$\neg P$
T	F
F	T

Table 2: the truth table for the negation $\neg P$

Even more, we will soon see for any statement P, it must be the case that either P is true or $\neg P$ is true. Before we arrive at this conclusion, we must discuss other ways to create new statements from a pair of statements P and Q. One way to do this is to consider the case that either P is true or Q is true. Put into symbols, the **disjunction** $P \vee Q$ of the statements P and Q is the statement, "Either it is the case that P or it is the case that Q," for which the upside-down wedge $\vee$ denotes the connective "or." Compare the similarities between the disjunction $\vee$ and the set union $\cup$.

Example 0.2.11. Consider the following pair of statements.

P: Topeka is the capital of Kansas.

Q: The real number $\sqrt{2}$ is a root of $x^2 - 2$.

We may construct the disjunction $P \vee Q$ by placing the connective "or" between the statements.

$P \vee Q$: Either Topeka is the capital of Kansas or the real number $\sqrt{2}$ is a root of $x^2 - 2$.

Both of the statements P and Q are in fact true, hence the disjunction $P \vee Q$ is true.

Example 0.2.12. Consider the following pair of statements.

P: Kansas City is the capital of Missouri.

Q: The real number π is transcendental.

We may construct the disjunction $P \vee Q$ by placing the connective "or" between the statements.

$P \vee Q$: Either Kansas City is the capital of Missouri or the real number π is transcendental.

Even though the statement P is false (since the capital of Missouri is Jefferson City), the disjunction $P \vee Q$ is true because π is a transcendental number (this fact is non-trivial but well-known).

Example 0.2.13. Consider the following pair of statements.

P: The square root of -1 is a real number

Q: The integer 11 is composite.

We may construct the disjunction $P \vee Q$ by placing the connective "or" between the statements.

$P \vee Q$: Either the square root of -1 is a real number or the integer 11 is composite.

Both of these statements P and Q are false: $\sqrt{-1}$ is a non-real complex number, and 11 is prime. Consequently, the disjunction $P \vee Q$ is a false statement because neither P nor Q is a true statement.

Crucially, we note that if either of the statements P or Q is true, then the disjunction $P \vee Q$ must also be true; however, if neither of the statements P or Q is true, then $P \vee Q$ must be false.

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

Table 3: the truth table for the disjunction $P \vee Q$

We may also consider the case that both of the statements P and Q are true simultaneously. Put another way, we may form the statement, "It is the case that P and it is the case that Q." We refer to this statement as the **conjunction** $P \wedge Q$ of P and Q, and we use the wedge $\wedge$ as the connective "and." Compare the similarities between the conjunction $\wedge$ and the set intersection $\cap$.

Example 0.2.14. Consider the following pair of statements.

P: Paris is the capital of France.

Q: The real number 1 is less than the real number $\sqrt{2}$.

We may construct the conjunction $P \wedge Q$ by placing the connective "and" between the statements.

$P \wedge Q$: Paris is the capital of France, and the real number 1 is less than the real number $\sqrt{2}$.

Both of the statements P and Q are in fact true, hence the conjunction $P \wedge Q$ is true.

Example 0.2.15. Consider the following pair of statements.

P: Leticia is the capital of France.

Q: The identity function on a set is injective.

We may construct the conjunction $P \wedge Q$ by placing the connective "and" between the statements.

$P \wedge Q$: Leticia is the capital of France, and the identity function on a set is injective.

Considering that the statement P is false (since we know that Paris is the capital of France), the conjunction $P \wedge Q$ is false. Explicitly, it is not true that both P and Q are true, so $P \wedge Q$ is false.

Example 0.2.16. Consider the following pair of statements.

P: We have that $\cos(k\pi) = 0$ for all integers k.

Q: The integer 8 is a perfect square.

We may construct the conjunction $P \wedge Q$ by placing the connective "and" between the statements.

$P \wedge Q$: We have that $\cos(k\pi) = 0$ for all integers k, and integer 8 is a perfect square.

Both of these statements are false: indeed, $\cos(k\pi) = (-1)^k$ for all integers k, and $\sqrt{8} = 2\sqrt{2}$ is not an integer. Consequently, the conjunction $P \wedge Q$ is false because neither P nor Q is true.

We note that the conjunction $P \wedge Q$ of statements P and Q is true if and only if both P and Q are true. Consequently, if either of the statements P or Q is false, then $P \wedge Q$ is false. Be careful not to confuse the upside-down wedge $\vee$ (meaning "or") with the wedge $\wedge$ (meaning "and").

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

Table 4: the truth table for conjunction $P \wedge Q$

We are now in a position to state and prove two fundamental principles in the calculus of logic.

Theorem 0.2.17 (Law of Excluded Middle)**.** *We have that $P \vee \neg P$ is true for any statement P.*

Proof. Given any statement P, consider the disjunction $P \vee \neg P$ of P and $\neg P$. Observe that if P is true, then $P \vee \neg P$ is true. Conversely, if P is false, then $\neg P$ is true, hence $P \vee \neg P$ is true. □

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

Table 5: the Law of Excluded Middle

Theorem 0.2.18 (Law of Non-Contradiction)**.** *We have that $P \wedge \neg P$ is false for any statement P.*

Proof. Given any statement P, consider the conjunction $P \vee \neg P$ of P and $\neg P$. Observe that if P is true, then $\neg P$ is false, hence $P \wedge \neg P$ is false. Conversely, if P is false, then $P \wedge \neg P$ is false. □

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

Table 6: the Law of Non-Contradiction

0.2.3 Conditional and Biconditional Statements

Going forward, we will be interested primarily in statements of the form $P \Rightarrow Q$, read aloud as "P implies Q" or "If P, then Q." Unsurprisingly, a statement of this form is called an **implication** or a **conditional statement**. We refer to the statement P in this construction as the **antecedent**; the statement Q is called the **consequent**. Observe that the statement $P \Rightarrow Q$ is false if and only if Q is false and P is true (since this is a **lie**); otherwise, the implication $P \Rightarrow Q$ is true.

Example 0.2.19. Consider the following pairs of statements.

P: Madrid is the capital of Spain.

Q: The integer 3 is odd.

We may construct the implication $P \Rightarrow Q$ as follows.

$P \Rightarrow Q$: If Madrid is the capital of Spain, then the integer 3 is odd.

Considering that both P and Q are true statements, it follows that $P \Rightarrow Q$ is true.

Example 0.2.20. Consider the following pairs of statements.

P: The integer 3 divides the integer 243.

Q: The integer 3 is even.

We may construct the implication $P \Rightarrow Q$ as follows.

$P \Rightarrow Q$: If the integer 3 divides the integer 243, then the integer 3 is even.

Observe that $243 = 81 \cdot 3$, hence 3 divides 243; however, we know well that 3 is not an even integer. Consequently, the conditional statement $P \Rightarrow Q$ is false: indeed, we are lying here.

Below is the truth table for the conditional statement $P \Rightarrow Q$, as indicated above.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

Table 7: the truth table for the implication $P \Rightarrow Q$

Crucially, if the statement P is false, then according to Table 7 above, the implication $P \Rightarrow Q$ is true regardless of the truth value of Q; in this case, the conditional statement $P \Rightarrow Q$ is called a **vacuous truth**, or equivalently, we say that $P \Rightarrow Q$ is **vacuously** true. Concretely, the idea is that the antecedent P cannot be satisfied because it is false, so the implication must be true.

Example 0.2.21. Consider the following pairs of statements.

P: The integer 17 is negative.

Q: Dr. Beck is a multi-instrumentalist.

We may construct the implication $P \Rightarrow Q$ as follows.

$P \Rightarrow Q$: If the integer 17 is negative, then Dr. Beck is a multi-instrumentalist.

Considering that the antecedent P is false (since its negation "$\neg P$: The integer 17 is positive." is in fact the true statement), it follows that the conditional statement $P \Rightarrow Q$ is vacuously true.

One way to justify this result (as promised by Table 7) is that no lies were told: Dr. Beck is a multi-instrumentalist, so there was no harm in (falsely) assuming that 17 is a negative integer.

Example 0.2.22. Consider the following pairs of statements.

P: The integer 17 is negative.

Q: Dr. Beck is a multi-millionaire.

We may construct the implication $P \Rightarrow Q$ as follows.

$P \Rightarrow Q$: If the integer 17 is negative, then Dr. Beck is a multi-millionaire.

Considering that the antecedent P is false (since its negation "$\neg P$: The integer 17 is positive." is in fact the true statement), it follows that the conditional statement $P \Rightarrow Q$ is vacuously true. (Unfortunately for Dr. Beck, this makes no difference for his situation: the integer 17 is positive.)

One way to verify this result is that no lies were told: Dr. Beck is in fact not a multi-millionaire, but on the other hand, there was nothing guaranteed unless 17 were in fact a negative integer.

We will typically say that "P implies Q" or "If P, then Q" if the conditional statement $P \Rightarrow Q$ is true. Conventionally, if P implies Q, then we will say that P is **sufficient** for Q. One can rephrase this by saying that P is sufficient for Q if Q is true provided the statement P. Crucially, as Table 7 illustrates, the statement P may be either true or false; it does not actually matter. Equivalently, we may say that "P only if Q" if the conditional statement $P \Rightarrow Q$ is true. We declare in this case that Q is **necessary** for P. Consequently, each of the following statements is equivalent.

(a.) $P \Rightarrow Q$ (c.) Q if P. (e.) P is sufficient for Q.

(b.) If P, then Q. (d.) P only if Q. (f.) Q is necessary for P.

We will fix our attention throughout the rest of the course primarily on conditional statements in which P and Q are open sentences. Consider the following examples along these lines.

Example 0.2.23. Consider the following pairs of statements about a positive integer n.

$$P(n)\colon \text{ The integer } n^4+1 \text{ is prime.}$$
$$Q(n)\colon \text{ The integer } n^2+1 \text{ is prime.}$$

By plugging in different values of the integer $n \geq 1$, we obtain explicit statements $P(n)$ and $Q(n)$.

$P(1)$: The integer 2 is prime. $Q(1)$: The integer 2 is prime.
$P(2)$: The integer 17 is prime. $Q(2)$: The integer 5 is prime.
$P(3)$: The integer 82 is prime. $Q(3)$: The integer 10 is prime.
$P(4)$: The integer 257 is prime. $Q(4)$: The integer 17 is prime.
$P(5)$: The integer 626 is prime. $Q(5)$: The integer 26 is prime.

Consider the conditional statement $P(n) \Rightarrow Q(n)$ defined as follows.

$$P(n) \Rightarrow Q(n)\colon \text{ If the integer } n^4+1 \text{ is prime, then the integer } n^2+1 \text{ is prime.}$$

By Table 7, we know that $P(n) \Rightarrow Q(n)$ is false if and only if $P(n)$ is true and $Q(n)$ is false. Consequently, the statement $P(n) \Rightarrow Q(n)$ is true for all integers $1 \leq n \leq 5$. Quite astonishingly, this statement is in fact true for all integers $1 \leq n \leq 27$; however, we have that $28^4+1 = 614657$ is prime and $28^2+1 = 785$ is not prime, hence the statement $P(28) \Rightarrow Q(28)$ is false.

Example 0.2.24. Consider the following pairs of statements about a positive integer n.

$$P(n)\colon \text{ The integer } n^2+1 \text{ is prime.}$$
$$Q(n)\colon \text{ The integer } n^4-1 \text{ is prime.}$$

By definition, the conditional statement $P(n) \Rightarrow Q(n)$ is given as follows.

$$P(n) \Rightarrow Q(n)\colon \text{ If the integer } n^2+1 \text{ is prime, then the integer } n^4-1 \text{ is prime.}$$

By Table 7, we know that $P(n) \Rightarrow Q(n)$ is false if and only if $P(n)$ is true and $Q(n)$ is false. Considering that $n^4-1 = (n^2-1)(n^2+1)$ is divisible by n^2+1 for all integers n, it follows that n^4-1 is composite for all integers n, hence the open sentence $Q(n)$ is false for every integer n. We conclude that the conditional statement $P(n) \Rightarrow Q(n)$ is false for all integers $n \geq 1$.

Example 0.2.25. Consider the following pairs of statements about a pair of real numbers x and y.

$$P(x,y)\colon \text{ We have that } x+y=1.$$
$$Q(x,y)\colon \text{ We have that } x^2+y^2=1.$$

By definition, the conditional statement $P(x,y) \Rightarrow Q(x,y)$ is given as follows.

$$P(x,y) \Rightarrow Q(x,y)\colon \text{ If } x+y=1, \text{ then } x^2+y^2=1.$$

By Table 7, we know that $P(x,y) \Rightarrow Q(x,y)$ is false if and only if $P(x,y)$ is true and $Q(x,y)$ is false. Observe that if $x+y=1$, then $y=1-x$ so that $y^2=x^2-2x+1$. Consequently, it follows that $x^2+y^2=2x^2-2x+1$; thus, the open sentence $Q(x,y)$ is true if and only if $2x^2-2x+1=1$ if and only if $2x^2-2x=0$ if and only if $2x(x-1)=0$ if and only if $x=0$ or $x=1$. We conclude that the statement $P(x,y) \Rightarrow Q(x,y)$ is true if and only if $x=0$ and $y=1$ or $x=1$ and $y=0$.

Given any pair of statements P and Q, the conditional statement $Q \Rightarrow P$ formed by swapping the antecedent and the consequent is called the **converse** of the conditional statement $P \Rightarrow Q$. Like the implication $P \Rightarrow Q$, its converse $Q \Rightarrow P$ can be understood in different ways.

(a.) $Q \Rightarrow P$

(b.) If Q, then P.

(c.) P if Q.

(d.) Q only if P.

(e.) Q is sufficient for P.

(f.) P is necessary for Q.

P	Q	$Q \Rightarrow P$
T	T	T
T	F	T
F	T	F
F	F	T

Table 8: the truth table for the converse $Q \Rightarrow P$ of the implication $P \Rightarrow Q$

Example 0.2.26. Consider the following pairs of statements.

P: The integer 17 is negative.

Q: Dr. Beck is a multi-instrumentalist.

We may construct the converse $Q \Rightarrow P$ of the implication $P \Rightarrow Q$ as follows.

$Q \Rightarrow P$: If Dr. Beck is a multi-instrumentalist, then the integer 17 is negative.

Unlike the implication $P \Rightarrow Q$ (which is vacuously true since P is false), the converse $Q \Rightarrow P$ is false: Dr. Beck is a multi-instrumentalist, but the integer 17 is not negative.

Example 0.2.27. Consider the following pairs of statements.

P: The integer 17 is negative.

Q: Dr. Beck is a multi-millionaire.

We may construct the converse $Q \Rightarrow P$ of the implication $P \Rightarrow Q$ as follows.

$Q \Rightarrow P$: If Dr. Beck is a multi-millionaire, then the integer 17 is negative.

Considering that the antecedent Q is false, the conditional statement $Q \Rightarrow P$ is vacuously true. One way to verify this result is that no lies were told: Dr. Beck is not a multi-millionaire.

Bearing in mind Examples 0.2.21 and 0.2.26, even if P is sufficient for Q, it may not be true that P is necessary for Q; however, if P is both necessary and sufficient for Q, then both of the conditional statements $P \Rightarrow Q$ and $Q \Rightarrow P$ are true. We say in this case that P holds if and only if Q holds, and we represent this relationship symbolically by $P \Leftrightarrow Q$. We will typically say that the statements P and Q are **(materially) equivalent** if P is true if and only if Q is true. Put another way, the material equivalence $P \Leftrightarrow Q$ is simply the conjunction $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$. Each of the following statements concerning the **biconditional statement** $P \Leftrightarrow Q$ is equivalent.

(a.) $P \Leftrightarrow Q$

(b.) P if and only if Q.

(c.) P is (materially) equivalent to Q.

(d.) P is necessary and sufficient for Q.

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$(P \Rightarrow Q) \wedge (Q \Rightarrow P)$	$P \Leftrightarrow Q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

Table 9: the truth table for the biconditional $P \Leftrightarrow Q$

Put another way, we have that $P \Leftrightarrow Q$ is true if and only if P and Q have the same truth value.

Example 0.2.28. Consider the following statements.

P: The integer 3 divides the integer 243.
Q: The integer 3 is even.
R: The integer 17 is negative.
S: The integer 2027 is prime.

Considering that P and S are true statements, but Q and R are false statements, it follows that $P \Leftrightarrow S$ and $Q \Leftrightarrow R$ are both true statements and $P \Leftrightarrow Q$, $P \Leftrightarrow R$, $Q \Leftrightarrow S$, and $R \Leftrightarrow S$ are all false statements. Examples of these statements in words are provided below.

$P \Leftrightarrow Q$: The integer 3 divides the integer 243 if and only if 3 is even.
$P \Leftrightarrow S$: The integer 3 divides the integer 243 if and only if the integer 2027 is prime.
$Q \Leftrightarrow R$: The integer 3 is even if and only if the integer 17 is negative.

Example 0.2.29. Consider the following statements about an integer n.

$P(n)$: The integer n is even.
$Q(n)$: The integer n^2 is even.

We construct the biconditional statement $P(n) \Leftrightarrow Q(n)$ as follows.

$P(n) \Leftrightarrow Q(n)$: The integer n is even if and only if the integer n^2 is even.

By definition, an integer n is even if and only if $n = 2k$ for some integer k. Consequently, if n is even, then $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$ is even. Conversely, if n^2 is even, then there exists an integer k such that $n^2 = 2k$. Considering that 2 is prime, we must have that 2 divides n, hence n is even. We conclude therefore that the statement $P(n) \Leftrightarrow Q(n)$ is true for all integers n.

Example 0.2.30. Consider the following statements about an integer n.

$$P(n)\text{: The integer } n \text{ is odd.}$$
$$Q(n)\text{: The integer } n^2 \text{ is odd.}$$

We construct the biconditional statement $P(n) \Leftrightarrow Q(n)$ as follows.

$$P(n) \Leftrightarrow Q(n)\text{: For the integer } n \text{ to be odd, it is necessary and sufficient that } n^2 \text{ is odd.}$$

Often, this construction is more awkward than the more natural "if and only if" statement.

$$P(n) \Leftrightarrow Q(n)\text{: The integer } n \text{ is odd if and only if the integer } n^2 \text{ is odd.}$$

We leave it as an exercise for the reader to prove that $P(n) \Leftrightarrow Q(n)$ is true for all integers n.

Example 0.2.31. Consider the following pairs of statements about a pair of real numbers x and y.

$$P(x,y)\text{: We have that } x^2 + y^2 = 1.$$
$$Q(x,y)\text{: We have that } (x,y) \text{ lies on a circle of radius 1 centered at } (0,0).$$

By definition, the statements $P(x,y) \Rightarrow Q(x,y)$ and $Q(x,y) \Rightarrow P(x,y)$ are as follows.

$$P(x,y) \Rightarrow Q(x,y)\text{: If } x^2 + y^2 = 1, \text{ then } (x,y) \text{ lies on a circle of radius 1 centered at } (0,0).$$
$$Q(x,y) \Rightarrow P(x,y)\text{: If } (x,y) \text{ lies on a circle of radius 1 centered at } (0,0), \text{ then } x^2 + y^2 = 1.$$

Recall that the equation of a circle of radius r centered at (h,k) is given by

$$(x-h)^2 + (y-k)^2 = r^2.$$

Consequently, the conditional statement $Q(x,y) \Rightarrow P(x,y)$ is true by definition. Conversely, if $x^2 + y^2 = 1$, then $(x-0)^2 + (y-0)^2 = 1^2$ implies that the point $(x,y) \in \mathbb{R} \times \mathbb{R}$ lies on a circle of radius 1 centered at $(0,0)$. Put another way, we have that $P(x,y) \Rightarrow Q(x,y)$ is true. Ultimately, these observations together yield that the biconditional statement $P(x,y) \Leftrightarrow Q(x,y)$ is true.

0.2.4 Tautologies and Contradictions

By the Law of Excluded Middle, for any statement P, the statement $P \vee \neg P$ ("P or not P") is true; it is a **tautology**. Generally, a tautology is any statement that is true for all possible truth inputs.

Example 0.2.32. Given any statements P and Q, the disjunction $(\neg Q) \vee (P \Rightarrow Q)$ is a tautology. We can convince ourselves of this by realizing that $P \Rightarrow Q$ is true if either P is false or P and Q are both true. Consequently, the statement $(\neg Q) \vee (P \Rightarrow Q)$ is true in the case that P is false or P and Q are both true. But if Q is false, then $\neg Q$ is true, hence $(\neg Q) \vee (P \Rightarrow Q)$ is true.

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

Table 10: the truth table for $(\neg Q) \vee (P \Rightarrow Q)$

Example 0.2.33. Given statements P and Q, the implication $[(P \vee Q) \wedge (\neg Q)] \Rightarrow P$ is a tautology: indeed, it suffices to check that all of its values in the following truth table are T.

P	Q	$\neg Q$	$P \vee Q$	$(P \vee Q) \wedge (\neg Q)]$	$[(P \vee Q) \wedge (\neg Q)] \Rightarrow P$
T	T	F	T	F	T
T	F	T	T	T	T
F	T	F	T	F	T
F	F	T	F	F	T

Table 11: the truth table for $[(P \vee Q) \wedge (\neg Q)] \Rightarrow P$

Beyoncé says, "I break the internet: top two and I ain't number two," so she must be number one.

By the Law of Non-Contradiction, the statement $P \wedge \neg P$ ("P and not P") is always false; it is a **contradiction**. Generally, a contradiction is a statement that is false for all possible truth inputs.

Example 0.2.34. Given statements P and Q, the conjunction $P \wedge [P \Rightarrow (Q \wedge \neg Q)]$ is a contradiction: by the Law of Non-Contradiction, $Q \wedge \neg Q$ is false; thus, the conditional statement $P \Rightarrow (Q \wedge \neg Q)$ is false if P is true. Conversely, the implication $P \Rightarrow (Q \wedge \neg Q)$ is true if P is false. Consequently, the statements P and $P \Rightarrow (Q \wedge \neg Q)$ take opposite truth values, so their conjunction is false.

P	Q	$\neg Q$	$Q \wedge \neg Q$	$P \Rightarrow (Q \wedge \neg Q)$	$P \wedge [P \Rightarrow (Q \wedge \neg Q)]$
T	T	F	F	F	F
T	F	T	F	F	F
F	T	F	F	T	F
F	F	T	F	T	F

Table 12: the truth table for $P \wedge [P \Rightarrow (Q \wedge \neg Q)]$

Example 0.2.35. Given any statements P and Q, the conjunction $(P \wedge Q) \wedge [Q \Rightarrow \neg P]$ is a contradiction. We can verify this by constructing the corresponding truth table as follows.

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

Table 13: the truth table for $(P \wedge Q) \wedge (Q \Rightarrow \neg P)$

0.2.5 Logical Equivalence

Given any statements P and Q, recall from Table 7 that the conditional statement $P \Rightarrow Q$ is vacuously true if P is false; therefore, in order to determine the truth value of $P \Rightarrow Q$, it suffices to consider the case that P is true. Unfortunately, in some situations, it is difficult to establish the verity of Q even if P is known to be true. Under these circumstances, it is not possible to determine if the statement $P \Rightarrow Q$ 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 $S(P, Q)$ that depends on both P and Q that is **logically equivalent** to the implication $P \Rightarrow Q$. We say that two statements S_1 and S_2 are logically equivalent if and only if their values in a truth table are equal; if this is the case, then we write $S_1 \equiv S_2$ to assert symbolically that S_1 and S_2 are logically equivalent. Consequently, if we demonstrate that the statement $S(P, Q)$ is true, then $P \Rightarrow Q$ must be true, as well.

We will concern ourselves primarily with the interplay between the conjunction, disjunction, implication, and negation. We seek to construct a glossary of statements that are logically equivalent to the implication $P \Rightarrow Q$. Conventionally, if the statement P is false, then the implication $P \Rightarrow Q$ is vacuously true. Even more, if the statement Q is true, then the implication $P \Rightarrow Q$ is trivially true regardless of the truth value of P. Consequently, we may deduce that the statements $P \Rightarrow Q$ and $\neg P \vee Q$ are logically equivalent, as the following truth table illustrates.

P	Q	$\neg P$	$P \Rightarrow Q$	$\neg P \vee Q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Table 14: the truth table for the implication $P \Rightarrow Q$ and the disjunction $\neg P \vee Q$

Proposition 0.2.36. *Given any statements P and Q, we have that $(P \Rightarrow Q) \equiv (\neg P \vee Q)$.*

Consider the statement $\neg Q \Rightarrow \neg P$ called the **contrapositive** of the implication $P \Rightarrow Q$. Observe that if Q is true, then $\neg Q$ is false, hence the statement $\neg Q \Rightarrow \neg P$ is vacuously true. Likewise, if Q is true, then the statement $P \Rightarrow Q$ is true regardless of the verity of P. Conversely, if Q is false, then $\neg Q$ is true, hence $\neg Q \Rightarrow \neg P$ is true if and only if $\neg P$ is true if and only if P is false. Consequently, we are lead to the following truth table and the subsequent proposition.

P	Q	$\neg P$	$\neg Q$	$P \Rightarrow Q$	$\neg Q \Rightarrow \neg P$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Table 15: the truth table for the contrapositive $\neg Q \Rightarrow \neg P$ of the implication $P \Rightarrow Q$

Proposition 0.2.37. *Given any statements P and Q, we have that $(P \Rightarrow Q) \equiv (\neg Q \Rightarrow \neg P)$.*

Example 0.2.38. Consider the following statements.

P: Bernard earns an A on his final exam in MA291.

Q: Bernard earns an A as his final grade in MA291.

Let us assume that if Bernard earns an A on his final exam in MA291, then Bernard earns an A as his final grade in MA291. Consider the following statements regarding Bernard's grade in MA291.

R: Either Bernard does not earn an A on his final exam or Bernard earns an A in MA291.

S: If Bernard does not earn an A in MA291, then Bernard did not earn an A on his final exam.

Observe that the statement R is true: indeed, if Bernard does not earn an A on his final exam, then there is no promise as to what his final grade in MA291 will be, so no lies have been told regardless of the outcome. On the other hand, if Bernard earns an A as his final grade, then it does not matter what he earned on his final exam in MA291 because he will surely be happy with his grade. Likewise, the statement S is true: indeed, if Bernard does not earn an A as his final grade, then he must not have earned an A on his final exam because that would have guaranteed him an A in the course. We have corroborated the logical equivalence of the statements $P \Rightarrow Q$, $\neg P \vee Q$, and $\neg Q \Rightarrow \neg P$ for the example at hand, as guaranteed by Propositions 0.2.36 and 0.2.37.

Example 0.2.39. Consider the following statements.

P: It is overcast in Kansas City.

Q: Bernard brings an umbrella to work.

Let us assume as before that if it is overcast in Kansas City, then Bernard brings an umbrella to work. Observe that if Bernard does not bring an umbrella to work, then it must not be overcast in Kansas City; otherwise, if it were overcast in Kansas City, then Bernard would have brought an umbrella to work. Even more, it is either sunny in Kansas City or Bernard brings an umbrella to work: indeed, if Bernard does not bring an umbrella to work, then it must be sunny in Kansas City. Our exposition here bears out the logical equivalence of $P \Rightarrow Q$, $\neg Q \Rightarrow \neg P$, and $\neg P \vee Q$.

Often, it is useful to determine when the conditional statement $P \Rightarrow Q$ is false (i.e., P does not provide sufficient information from which to deduce Q). By Table 7, we have that $P \Rightarrow Q$ is false if and only if P is true and Q is false if and only if $P \wedge \neg Q$ is true, hence the statements $\neg(P \Rightarrow Q)$ and $P \wedge \neg Q$ are logically equivalent, as the following truth table illustrates.

P	Q	$\neg Q$	$P \Rightarrow Q$	$\neg(P \Rightarrow Q)$	$P \wedge \neg Q$
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 16: the truth table for the negated implication $\neg(P \Rightarrow Q)$ and the disjunction $P \wedge \neg Q$

Proposition 0.2.40. *Given any statements P and Q, we have that $\neg(P \Rightarrow Q) \equiv (P \wedge \neg Q)$.*

Example 0.2.41. Consider the following statements.

P: Bernard earns an A on his final exam in MA291.

Q: Bernard earns an A as his final grade in MA291.

Observe that if Bernard earns an A on his final exam in MA291 but Bernard does not earn an A as his final grade, then it is a lie to say that if Bernard earns an A on his final exam in MA291, then Bernard earns an A as his final grade in MA291; this illustrates the result of Proposition 0.2.40.

By Table 3, if $P \vee Q$ is false, then neither P nor Q is true. Likewise, by Table 4, if $P \wedge Q$ is false, then either P is false or Q is false. Combined, these observations form **De Morgan's Laws**.

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

Theorem 0.2.42 (De Morgan's Laws)**.** *Consider any statements P and Q.*

(a.) *We have that $\neg(P \vee Q) \equiv \neg P \wedge \neg Q$, i.e., $\neg(P \vee Q)$ is logically equivalent to $\neg P \wedge \neg Q$.*

(b.) *We have that $\neg(P \wedge Q) \equiv \neg P \vee \neg Q$, i.e., $\neg(P \wedge Q)$ is logically equivalent to $\neg P \vee \neg Q$.*

Example 0.2.43. Consider the following statements.

P: It is overcast in Kansas City.

Q: Bernard brings an umbrella to work.

Observe that if it is not the case that either it is overcast in Kansas City or Bernard brings an umbrella to work, then it must be the case that neither it is overcast in Kansas City nor Bernard brings an umbrella to work. Likewise, if it is not the case that it is overcast in Kansas City and Bernard brings an umbrella to work, then either it is not overcast in Kansas City or Bernard does not bring an umbrella to work. We have thus verified De Morgan's Laws for the given statements.

0.2.6 Quantified Statements

Often, we seek to determine the verity of an open sentence for all possible values in its domain. Explicitly, if $P(x)$ is any open sentence that depends on a variable x with domain S, then for each element $s \in S$, the truth value of the statement $P(s)$ is well-defined and can be determined.

Example 0.2.44. Consider the following statement about an integer n.

$P(n)$: The integer n is even.

We can plainly see that the verity of $P(n)$ depends entirely on the value of n. Each of the following statements in the left-hand column is true, but each statement in the right-hand column is false.

$P(0)$: The integer 0 is even.	$P(1)$: The integer 1 is even.
$P(2)$: The integer 2 is even.	$P(3)$: The integer 3 is even.
$P(4)$: The integer 4 is even.	$P(5)$: The integer 5 is even.

Quantification is another process of converting an open sentence $P(x)$ in the variable x into a statement whose truth value can be determined. **Quantified statements** are expressed using **logical quantifiers**. Primarily, we will study three logical quantifies throughout this course.

We use the **universal quantifier** $\forall$ to symbolically represent the phrases "for all," "for every," or "for each." Consequently, the statement $\forall x \in S$, $P(x)$ can be understood in words as, "For all elements $x \in S$, we have that $P(x)$." Observe that the quantified statement $\forall x \in S$, $P(x)$ is true if $P(x)$ is true for all elements $x \in S$; otherwise, this statement is false. Put another way, if the statement $P(x_0)$ is false for some element $x_0 \in S$, then the statement $\forall x \in S$, $P(x)$ is false.

Summary 0.2.45. Given any open sentence $P(x)$ with domain S, the quantified statement

$$\forall x \in S,\ P(x)\text{: For every element } x \in S, \text{ we have that } P(x).$$

is true if and only if $P(x)$ is true for all elements $x \in S$. Conversely, this quantified statement is false if and only if there exists an element $x_0 \in S$ such that $P(x_0)$ is false.

Example 0.2.46. Consider the following statement about an integer n.

$$P(n)\text{: The integer } n \text{ is even.}$$

By using the universal quantifier $\forall$ ("for all"), we obtain the following quantified statement.

$$\forall n \in \mathbb{Z},\ P(n)\text{: For every integer } n, \text{ we have that } n \text{ is even.}$$

By Example 0.2.44 and Summary 0.2.45, the above quantified statement is false since $P(1)$ is false.

Example 0.2.47. Consider the following statements about an integer n.

$$P(n)\text{: The integer } n \text{ is even.}$$
$$Q(n)\text{: The integer } n^2 \text{ is even.}$$

By using the universal quantifier $\forall$ ("for all"), we obtain the following quantified statement.

$$\forall n \in \mathbb{Z},\ [P(n) \Leftrightarrow Q(n)]\text{: For every integer } n, \text{ we have that } n \text{ is even if and only if } n^2 \text{ is even.}$$

By Example 0.2.29, this statement is true because $P(n) \Leftrightarrow Q(n)$ is true for all integers n.

Example 0.2.48. Consider the following statement about a pair of real numbers x and y.

$$P(x, y)\text{: The real number } x^2 + y^2 \text{ is non-negative.}$$

By using the universal quantifier $\forall$ ("for all"), we obtain the following quantified statement.

$$\forall x, y \in \mathbb{R},\ P(x, y)\text{: For every pair of real numbers } x \text{ and } y, \text{ we have that } x^2 + y^2 \geq 0.$$

Considering that $x^2 \geq 0$ for every real number x, it follows that $x^2 + y^2 \geq 0$ for every pair of real numbers x and y. Consequently, the quantified statement $\forall x, y \in \mathbb{R}, P(x, y)$ is true.

Example 0.2.49. Consider the following statements about a real number x.

$$P(x)\colon \text{ The real number } x^2+4 \text{ satisfies that } x^2+4 \geq 4.$$
$$Q(x)\colon \text{ The real number } x^2+4 \text{ satisfies that } x^2+4 \leq 4.$$

By using the universal quantifier $\forall$ ("for all"), we obtain the following quantified statements.

$$\forall x \in \mathbb{R},\, P(x)\colon \text{ For all real numbers } x, \text{ we have that } x^2+4 \geq 4.$$
$$\forall x \in \mathbb{R},\, Q(x)\colon \text{ For all real numbers } x, \text{ we have that } x^2+4 \leq 4.$$

Considering that $x^2 \geq 0$ for every real number x, it follows that $x^2+4 \geq 4$ for every real number x. Consequently, the quantified statement $\forall x \in \mathbb{R},\, P(x)$ is true; however, $Q(1)$ is false because the real number $5 = 1^2+4$ does not satisfy that $5 \leq 4$. We conclude that $\forall x \in \mathbb{R},\, Q(x)$ is false.

One other indispensable way to view the universally quantified statement $\forall x \in S,\, P(x)$ in words is, "If x is an element of S, then we have that $P(x)$" or "If $x \in S$, then $P(x)$." Observe that in this manner, any statement involving the universal quantifier is simply a conditional statement. Consequently, Proposition 0.2.36 entails the logical equivalence of the universally quantified statement $\forall x \in S,\, P(x)$ and the disjunction $(x \notin S) \vee P(x)$. By De Morgan's Laws, the negation of $\forall x \in S,\, P(x)$ is logically equivalent to the negation of $(x \notin S) \vee P(x)$ — namely, $(x \in S) \wedge \neg P(x)$.

Summary 0.2.50. Given any open sentence $P(x)$ with domain S, the following are equivalent.

(a.) $\forall x \in S,\, P(x)\colon$ For every element $x \in S,$ we have that $P(x)$.

(b.) $(x \notin S) \vee P(x)\colon$ Either x is not an element of S or we have that $P(x)$.

Better yet, the negation of a quantified statement is itself a quantified statement. Explicitly, we use the **existential quantifier** $\exists$ to express the phrases "there exists," "for at least one," or "for some." Consequently, the quantified statement $\exists x \in S,\, P(x)$ can be understood in words as, "There exists an element $x \in S$ such that $P(x)$." Observe that the quantified statement $\exists x \in S,\, P(x)$ is true if $P(x_0)$ is true for some element $x_0 \in S$; otherwise, this statement is false. Put another way, if $P(x)$ is false for every element $x \in S$, then the quantified statement $\exists x \in S,\, P(x)$ is false.

Summary 0.2.51. Given any open sentence $P(x)$ with domain S, the quantified statement

$$\exists x \in S,\, P(x)\colon \text{ There exists an element } x \in S \text{ such that } P(x).$$

is true if and only if $P(x_0)$ is true for some element $x_0 \in S$. Conversely, this quantified statement is false if and only if the statement $P(x)$ is false for all elements $x \in S$.

Example 0.2.52. Consider the following statement about an integer n.

$$P(n)\colon \text{ The integer } n \text{ is even.}$$

By using the existential quantifier $\exists$ ("there exists"), we obtain the following quantified statement.

$$\exists n \in \mathbb{Z},\, P(n)\colon \text{ There exists an integer } n \text{ such that } n \text{ is even.}$$

Certainly, the above quantified statement is true because $P(2)$ is true.

Example 0.2.53. Consider the following statement about an integer n.

$$P(n)\colon \text{ The integer } n^4+1 \text{ is prime.}$$

By using the existential quantifier $\exists$ ("there exists"), we obtain the following quantified statement.

$$\exists n \in \mathbb{Z},\ P(n)\colon \text{ There exists an integer } n \text{ such that } n^4+1 \text{ is prime.}$$

Considering that $2 = 1^4 + 1$ is prime, $P(1)$ is true, hence the above quantified statement is true.

Example 0.2.54. Consider the following statement about a pair of real numbers x and y.

$$P(x,y)\colon \text{ The real numbers } x \text{ and } y \text{ satisfy that } x^2+y^2=4.$$

By using the existential quantifier $\exists$ ("there exists"), we obtain the following quantified statement.

$$\exists x, y \in \mathbb{R},\ P(x,y)\colon \text{ There exist real numbers } x \text{ and } y \text{ such that } x^2+y^2=4.$$

Considering that the set of ordered pairs (x, y) of real numbers satisfying that $x^2 + y^2 = 4$ is the graph of a circle of radius 2 centered at the origin in the Cartesian plane, it follows that the above quantified statement is true: indeed, both of the statements $P(2, 0)$ and $P(0, 2)$ are true.

Example 0.2.55. Consider the following statements about a real number x.

$$P(x)\colon \text{ The real number } x \text{ satisfies that } x^2-2x-3=0.$$
$$Q(x)\colon \text{ The real number } x^3 \text{ satisfies that } x^3 \geq 8.$$

By using the existential quantifier $\exists$ ("there exists"), we obtain the following quantified statements.

$$\exists x \in \mathbb{R},\ [P(x) \Rightarrow Q(x)]\colon \text{ There exists a real number } x \text{ such that } x^3 \geq 8 \text{ if } x^2-2x-3=0.$$
$$\exists x \in \mathbb{R},\ [P(x) \wedge \neg Q(x)]\colon \text{ There exists a real number } x \text{ such that } x^2-2x-3=0 \text{ and } x^3<8.$$

Observe that if $P(x)$ is false, then the conditional statement $P(x) \Rightarrow Q(x)$ is vacuously true. Consequently, the first quantified statement above is true for any real number x such that x^2-2x-3 is nonzero (e.g., suppose that $x = 0$ or $x = 1$). On the other hand, we can determine explicitly the values of x such that $P(x)$ is true since $(x-3)(x+1) = x^2-2x-3 = 0$ if and only if $x = 3$ or $x = -1$. Consequently, we have that $P(3) \Rightarrow Q(3)$ is true. Likewise, the second quantified statement above is true because the real number $x = -1$ satisfies that $(-1)^2 - 2(-1) - 3 = 0$ and $(-1)^3 = 1 < 8$. Put another way, we have that $P(-1)$ is true and $Q(-1)$ is false.

We provide next the crucial theorem that relates the universal and existential quantifiers.

Theorem 0.2.56 (Negation of Quantified Statements)**.** *Consider any open sentence $P(x)$ over S.*

1.) *We have that* $\neg[\forall x \in S,\ P(x)] \equiv [\exists x \in S,\ \neg P(x)]$.

2.) *We have that* $\neg[\exists x \in S,\ P(x)] \equiv [\forall x \in S,\ \neg P(x)]$.

Last, if $P(x)$ is any open sentence whose domain is any nonempty set S, then we say that an element $x_0 \in S$ is the **unique** element of S **satisfying the statement** $P(x_0)$ if and only if

(a.) the statement $P(x_0)$ is true and

(b.) for every element $x \in S$, if $P(x)$ is true, then we must have that $x = x_0$.

We use the **uniqueness quantifier** ! to represent the phrase "unique." Explicitly, we will write $\exists! x \in S,\ P(x)$ to signify that "there exists a unique element $x \in S$ such that $P(x)$."

Example 0.2.57. Consider the following statement about an integer n.

$$P(n)\colon \text{ The integer } n \text{ satisfies that } 3n - 4 = 5.$$

Observe that $P(n)$ is true if and only if $3n - 4 = 5$ if and only if $n = 3$, hence the statement $P(n)$ admits a unique element $n_0 \in \mathbb{Z}$ satisfying that $P(n_0)$ is true: namely, it is the integer $n_0 = 3$. Put another way, the following quantified statement involving the uniqueness quantifier is true.

$$\exists! n \in \mathbb{Z},\ P(n)\colon \text{ There exists a unique integer } n \text{ such that } 3n - 4 = 9.$$

Example 0.2.58. Consider the following statement about a real number x.

$$P(x)\colon \text{ The real number } x \text{ satisfies that } x - 5 + \frac{25}{x+5} = \frac{4x+5}{x+5}.$$

By solving the rational equation that defines $P(x)$, we find that $P(x)$ is true if and only if $x = 1$.

$$\frac{(x-5)(x+5)+25}{x+5} = \frac{4x+5}{x+5}$$

$$(x-5)(x+5)+25 = 4x+5$$

$$x^2 - 25 + 25 = 4x + 5$$

$$x^2 - 4x - 5 = 0$$

$$(x-1)(x+5) = 0$$

Considering that $x + 5$ cannot equal 0, the Zero Product Property yields that $x - 1 = 0$ so that $x = 1$. Put another way, the following statement involving the uniqueness quantifier is true.

$$\exists! x \in \mathbb{R},\ P(x)\colon \text{ There exists a unique real number } x \text{ such that } x - 5 + \frac{25}{x+5} = \frac{4x+5}{x+5}.$$

0.3 Basic Proof Techniques

Generally, mathematical research and problem solving are carried out in two steps: first, one must conduct extensive experimentation to determine some underlying pattern; then, the most significant effort is exerted to establish the veracity of the observed phenomenon in general. Concretely, this is achieved using set theory and the calculus of logic to construct a mathematical proof. Put simply, a mathematical proof is nothing more than a convincing argument that is replicable and unambiguous. We demonstrate in this section how to employ the basic axioms and general principles of certain mathematical structures to write mathematical proofs. We devote our attention to the three most common types of proofs: direct proof, proof by contrapositive, and proof by contradiction.

0.3.1 Direct Proof

Our primary focus throughout this section is to use the foundations of the calculus of logic presented in Section 0.2 to inform and develop the writing of mathematical proofs: indeed, the overwhelming impetus of contemporary mathematics lies in proving statements of the form "if P, then Q" for some statements (or open sentences) P and Q. Consequently, our attention will be by-and-large fixed on conditional statements of the form $P \Rightarrow Q$. Considering the truth table 7 for the implication, if either the statement Q is true or the statement P is false, then the conditional statement $P \Rightarrow Q$ is true. Proofs that are carried out by showing that Q is true are called **trivial proofs**. Conversely, any proof that demonstrates that P is false is called a **vacuous proof**. We begin our discussion of direct proofs with this low-hanging fruit, as illustrated in the following typical examples.

Example 0.3.1. Prove that if n is an even integer, then $n^2 + 4 \geq 3$.

Solution. Consider the following statements involving an integer n.

$$\begin{aligned} P(n)&: \text{ The integer } n \text{ is even.} \\ Q(n)&: \text{ The integer } n \text{ satisfies that } n^2 + 4 \geq 3. \end{aligned}$$

We seek to prove that $\forall n \in \mathbb{Z}$, $[P(n) \Rightarrow Q(n)]$ is a true statement. Considering that $n^2 \geq 0$ for any real number (and hence any integer) n, it follows that $n^2 + 4 \geq 4$. Consequently, the statement $Q(n)$ is true for all integers n, hence the statement $\forall n \in \mathbb{Z}$, $[P(n) \Rightarrow Q(n)]$ is trivially true. ◇

Our above work is merely a suggestion of a proof of the statement in Example 0.3.1. Below, we provide an example of how a proof of this statement might look "in the wild." Crucially, observe that in the following proof, there is no need to provide any symbols for the statements.

Proof. (Example 0.3.1) Considering that $n^2 \geq 0$ for any real number n, it follows that $n^2 + 4 \geq 4$. Consequently, we have that $n^2 + 4 \geq 3$ for every integer n, so the claim holds trivially. □

We point out at this juncture two important features of a mathematical proof. First, it is vitally important for the writer to indicate the beginning of a proof with an italicized "Proof" and a period. Equally as important is the ending of the proof. We will use in this course an empty box □ to signal the conclusion of a proof; however, the reader may alternatively use the acronym "QED" (Latin for "quod erat demonstrandum" or "what was to be shown") depending upon their preference.

Example 0.3.2. Prove that if a real number x satisfies that $x^2 - 2 = 0$, then 7 is an odd integer.

Solution. Like the previous example, the hypothesis that x is a real number satisfying that $x^2 - 2 = 0$ has no bearing on the truth value of the conclusion that 7 is an odd integer: indeed, 7 is an odd integer, so regardless of what hypotheses we make, the if-then statement remains true. ◇

Proof. (Example 0.3.2) Considering that 7 is an odd integer, the statement is trivially true. □

Example 0.3.3. Prove that if the Riemann Hypothesis holds, then $\frac{d}{dx}e^x = e^x$.

Proof. By elementary calculus, it holds that $\frac{d}{dx}e^x = e^x$, hence the statement is trivially true. □

Example 0.3.4. Prove that if -1 is an even integer, then the Riemann Hypothesis holds.

Solution. We are now in the opposite case of a trivial proof: indeed, the hypotheses of the statement are false because -1 is not an even integer, hence the statement is true vacuously. $\diamond$

Proof. (Example 0.3.4) Considering that -1 is an odd integer, the statement is vacuously true. $\square$

Example 0.3.5. Prove that if there exist a pair of real numbers x and y such that $x^2 + y^2 = -4$, then only finitely many positive integers are prime.

Proof. Given any real number x, we have that $x^2 \geq 0$. Consequently, we find that $x^2 + y^2 \geq 0$ for all real numbers x and y. Bearing this in mind, it follows that the statement is vacuously true. $\square$

Example 0.3.6. Prove that if $\{(x, y) \mid x, y \in \mathbb{R} \text{ and } y^2 = x\}$ is a function, then $\frac{1}{0} = 1$.

Proof. Observe that if $x = 1$, then the real numbers $y = 1$ and $y = -1$ both satisfy that $y^2 = x$. Consequently, the ordered pairs $(1, 1)$ and $(1, -1)$ both belong to $\{(x, y) \mid x, y \in \mathbb{R} \text{ and } y^2 = x\}$, hence this relation is not a function. We conclude that the statement is vacuously true. $\square$

Often, it will not be the case that we will encounter a statement that can be proved by a trivial or vacuous proof; rather, we will typically assume that the hypotheses of the statement are true in the first place, and we will subsequently perform some algebraic analysis or arithmetic manipulation in order to rigorously justify that the conclusion of the statement holds. We refer to this process as a **direct proof**. Explicitly, a direct proof of a conditional statement $P \Rightarrow Q$ usually begins with the phrase, "Suppose that P is true" and ends with the phrase, "We conclude that Q is true." Between these two points, the writer is left to fill in the details — how ever complicated they are.

Crucially, the validity of a direct proof relies on the law of inference called **modus ponens** that asserts that the conditional statement $[(P \Rightarrow Q) \wedge P] \Rightarrow Q$ is a tautology. Eliminating the trivial or vacuous cases, in order to establish the verity of a conditional statement $P \Rightarrow Q$, we need only assume that P is true and deduce from this that $P \Rightarrow Q$ is true (because if P is false, then $P \Rightarrow Q$ is true vacuously). Let us construct a truth table to verify the law of modus ponens.

P	Q	$P \Rightarrow Q$	$(P \Rightarrow Q) \wedge P$	$[(P \Rightarrow Q) \wedge P] \Rightarrow Q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

Table 18: the truth table for modus ponens $[(P \Rightarrow Q) \wedge P] \Rightarrow Q$

We conclude this section with several examples of direct proofs that require a bit more work than trivial or vacuous proofs. Before this, we recall the definitions of an even integer versus an odd integer. Explicitly, an integer n is **even** if and only if there exists an integer k such that $n = 2k$. Conversely, an integer n is **odd** if and only if there exists an integer ℓ such that $n = 2\ell + 1$.

Example 0.3.7. Prove that if n is an even integer, then $4n + 7$ is an odd integer.

Proof. By definition, if n is an even integer, then there exists an integer k satisfying that $n = 2k$. Consequently, we have that $4n + 7 = 4(2k) + 7 = 8k + 7 = (8k + 6) + 1 = 2(4k + 3) + 1$. Considering that $4k + 3$ is also an integer, it follows that $4n + 7$ is an odd integer, as desired. $\square$

Example 0.3.8. Prove that if n is an odd integer, then $3n - 1$ is an even integer.

Proof. By definition, if n is an odd integer, then there exists an integer k satisfying that $n = 2k+1$. Consequently, we have that $3n - 1 = 3(2k + 1) - 1 = 6k + 2 = 2(3k + 1)$. Considering that $3k + 1$ is also an integer, it follows that $3n - 1$ is an even integer, as desired. □

Example 0.3.9. Prove that if n is an even integer, then $3n^2 + 5n - 3$ is an odd integer.

Proof. By definition, if n is an even integer, then $n = 2k$ for some integer k. Consequently, we have

$$3n^2 + 5n - 3 = 3(2k)^2 + 5(2k) - 3 = 12k^2 + 10k - 3 = (12k^2 + 10k - 4) + 1 = 2(6k^2 + 5k - 2) + 1.$$

Considering that $6k^2 + 5k - 2$ is an integer, it follows that $3n^2 + 5n - 3$ is an odd integer. □

Example 0.3.10. Prove that if a, b, c are integers, then $ab + ac + bc$ is even if a and b are even.

Proof. We will assume that a, b, and c are integers such that a and b are even. By definition, there exist integers k and ℓ such that $a = 2k$ and $b = 2\ell$. Consequently, we have that

$$ab + ac + bc = (2k)(2\ell) + (2k)c + (2\ell)c = 4k\ell + 2(ck) + 2(c\ell) = 2(ck + c\ell + 2k\ell).$$

Considering that $ck + c\ell + 2k\ell$ is an integer, it follows that $ab + ac + bc$ is an even integer. □

0.3.2 Proof by Contrapositive

Consider any pair of statements P and Q. Recall from Section 0.2.5 that the **contrapositive** of the conditional statement $P \Rightarrow Q$ is the conditional statement $\neg Q \Rightarrow \neg P$. By the result of Table 15 and Proposition 0.2.37, any conditional statement is logically equivalent to its contrapositive. Consequently, the **proof by contrapositive** is a proof technique that exploits this logical equivalence. Explicitly, a proof by contrapositive is used to establish the verity of a conditional statement $P \Rightarrow Q$ by instead demonstrating the truth of its contrapositive statement $\neg Q \Rightarrow \neg P$ and using the logical equivalence of the two statements to conclude the truth of the original implication $P \Rightarrow Q$. Bearing this in mind, a typical proof by contrapositive ought to begin with the phrase, "Suppose that $\neg Q$ is true" and end with the phrase, "We conclude that $\neg P$ is true."

Before we proceed with an illustration of the technique of proof by contrapositive, we turn our attention to the law of inference called **modus tollens** that is closely related to the law of modus ponens and asserts that the conditional statement $[\neg Q \wedge (P \Rightarrow Q)] \Rightarrow \neg P$ is a tautology.

P	Q	$\neg P$	$\neg Q$	$P \Rightarrow Q$	$\neg Q \wedge (P \Rightarrow Q)$	$[\neg Q \wedge (P \Rightarrow Q)] \Rightarrow \neg P$
T	T	F	F	T	F	T
T	F	F	T	F	F	T
F	T	T	F	T	F	T
F	F	T	T	T	T	T

Table 19: the truth table for modus tollens $[\neg Q \wedge (P \Rightarrow Q)] \Rightarrow \neg P$

Proof by contrapositive is a powerful technique that is most useful when either the verity of Q is difficult to deduce from the verity of P or $\neg Q$ is a stronger hypothesis than P itself. We illustrate the importance and usefulness of the proof by contrapositive in the following examples. Be sure to make note of where a direct proof might falter or what difficulties arise from weak assumptions.

Example 0.3.11. Prove that if n is an integer, then n is even if and only if n^2 is even.

Solution. Consider the following statements involving an integer n.

$$P(n)\colon \text{ The integer } n \text{ is even.}$$
$$Q(n)\colon \text{ The integer } n^2 \text{ is even.}$$

We seek to establish the veracity of the biconditional statement $P(n) \Leftrightarrow Q(n)$ for each integer n. Consequently, we must establish that both the implication $P(n) \Rightarrow Q(n)$ and its converse $Q(n) \Rightarrow P(n)$ are true for each integer n. One direction is fairly straightforward: if the integer n is even, then there exists an integer k such that $n = 2k$. By squaring both sides of this equation, we conclude that $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$ is even because $2k^2$ is an integer. Conversely, if we assume that n is an integer such that n^2 is even, then there exists an integer k such that $n^2 = 2k$. Unfortunately, this assumption does not afford us much deductive power: it is unclear to the author (and likely to the reader) at this point why the equation $n^2 = 2k$ entails that n must be even. (Later, we will learn about division by prime numbers, but for now, we make no assumption that the reader is familiar with this technique.) Consequently, the hypothesis of $Q(n)$ is relatively "weak."

We may therefore seek to prove the conditional statement $Q(n) \Rightarrow P(n)$ by contrapositive: indeed, we fare immediately better using this proof technique because the assumption $\neg P(n)$ that n is an odd integer is "stronger" than the assumption that n^2 is an even integer. Concretely, if n is an odd integer, then $n = 2k + 1$ for some integer k. By squaring both sides of this equation, we find that $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$. Considering that $2k^2 + 2k$ is an integer, we conclude that n^2 is an odd integer; thus, our proof by contrapositive is complete. ◇

Proof. (Example 0.3.11) We will assume first that n is an even integer. By definition, there exists an integer k satisfying that $n = 2k$. Consequently, by squaring both sides of this equation, we find that $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$. Considering that $2k^2$ is an integer, it follows that n^2 is even.

Conversely, we will prove the converse by contrapositive. We must assume to this end that n is an odd integer. By definition of an odd integer, there exists an integer k such that $n = 2k + 1$. By squaring both sides of this equation, we find that $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$. Considering that $2k^2 + 2k$ is an integer, it follows that n^2 is an odd integer, as desired. □

Example 0.3.12. Prove that if n is an integer such that $7n + 6$ is even, then n is even.

Solution. We might first attempt a direct proof. Observe that if $7n+6$ is even, then $7n+6 = 2k$ for some integer k. By subtracting 6 from both sides, we find that $7n = 2k-6 = 2(k-3)$; however, it is here that things become unclear without a solid understanding of how prime numbers behave with respect to divisibility. Consequently, a direct proof is unsatisfactory; on the other hand, we might fare better with a proof by contrapositive. Observe that if n is odd, then there exists an integer k such that $n = 2k + 1$ and $7n + 6 = 7(2k + 1) + 6 = 14k + 13 = (14k + 12) + 1 = 2(7k + 6) + 1$. We conclude therefore that if n is odd, then $7n + 6$ is odd, hence the contrapositive is true. ◇

Proof. (Example 0.3.12) We will prove the contrapositive of the statement. We must assume to this end that n is an odd integer. By definition of an odd integer, there exists an integer k such that $n = 2k + 1$. Observe that $7n + 6 = 7(2k + 1) + 6 = 14k + 13 = (14k + 12) + 1 = 2(7k + 6) + 1$. Considering that $7k + 6$ is an integer, it follows that $7n + 6$ is an odd integer, as desired. □

Example 0.3.13. Prove that if n is an integer such that $7n - 3$ is odd, then $11n + 6$ is even.

Solution. We will first attempt a direct proof. We will assume along these lines that $7n-3 = 2k+1$ for some integer k. By adding $4n + 9$ to both sides of this equation, we find that

$$11n + 6 = (7n - 3) + (4n + 9) = (2k + 1) + (4n + 9) = 2k + 4n + 10 = 2(k + 2n + 5)$$

so that $11n+6$ is an even integer because $k+2n+5$ is an integer. But perhaps it seems miraculous to the reader that we were able to add $4n+9$ to both sides of the equation to obtain a direct proof. Bearing this in mind, we might seek a proof by contrapositive; this would entail that $11n+6 = 2k+1$ for some integer k so that $11n = 2k - 5$. We are at this point stuck because it is not clear how to extract any meaning from this equation. Our intuition might suggest that if $7n - 3$ is odd, then n must be even: indeed, an odd integer times an odd integer is an odd integer, and the difference of two odd integers is an odd integer, so n cannot (ostensibly) be odd. We are therefore brought to the potential midpoint in the present problem to prove that if $7n - 3$ is odd, then n is even. ◇

Often, the proof of an assertion could benefit from (or potentially even requires) some more powerful observation. Conventionally, such a helping proposition is referred to as a **lemma**. Let us state and prove a lemma that will make the proof of the previous example follow more efficiently.

Lemma 0.3.14. *If n is an integer such that $7n - 3$ is an odd integer, then n is even.*

Solution. We might first attempt a direct proof: indeed, suppose that $7n - 3 = 2k + 1$ for some integer k. We have that $7n = 2k + 4 = 2(k + 2)$. But again, without knowledge of divisibility of prime numbers, this equation is rather useless; we will therefore attempt a proof by contrapositive for this lemma. Observe that if n is odd, then there exists an integer k satisfying that $n = 2k + 1$. Consequently, we have that $7n - 3 = 7(2k + 1) - 3 = 14k + 4 = 2(7k + 2)$ is even, as desired. ◇

Proof. (Lemma 0.3.14) We will prove the contrapositive of the statement of the lemma. We must assume to this end that n is an odd integer. By definition of an odd integer, there exists an integer k such that $n = 2k + 1$. Observe that $7n - 3 = 7(2k + 1) - 3 = 14k + 4 = 2(7k + 2)$. Considering that $7k + 2$ is an integer, it follows that $7n + 6$ is an even integer, as desired. □

Proof. (Example 0.3.13) By Lemma 0.3.14, if n is an integer such that $7n-3$ is odd, then n is even. Consequently, there exists an integer k such that $n = 2k$. Even more, we have that

$$11n + 6 = 11(2k) + 6 = 22k + 6 = 2(11k + 3).$$

Considering that $11k + 3$ is an integer, we conclude that $11n + 6$ is an even integer. □

Example 0.3.15. Prove that if n is any integer, then $2n^2 + n$ is odd if and only if $\cos\left(\frac{n\pi}{2}\right) = 0$.

Solution. Glancing at this proposition, it might seem quite unwieldy — after all, we are comparing the parity of an integer $2n^2+n$ with the roots of the cosine function — but if one takes a moment to recognize the values this cosine sequences takes, the proof strategy becomes clear: indeed, computing $\cos\left(\frac{n\pi}{2}\right) = 0$ for some integers n, the reader will have a much better handle of the situation.

$$\cos(0) = 1 \qquad \cos\left(\frac{\pi}{2}\right) = 0 \qquad \cos(\pi) = -1 \qquad \cos\left(\frac{3\pi}{2}\right) = 0$$

Consequently, we deduce that $\cos\left(\frac{n\pi}{2}\right) = 0$ if and only if n is odd. We are lead to the following. $\diamond$

Lemma 0.3.16. *If n is an integer, then $\cos\left(\frac{n\pi}{2}\right) = 0$ if and only if n is odd.*

Proof. By elementary trigonometry, we have that $\cos\left(\frac{n\pi}{2}\right) = 0$ if and only if $\frac{n\pi}{2} = \frac{(2k+1)\pi}{2}$ for some integer k if and only if $n = 2k+1$ for some integer k if and only if n is an odd integer. $\square$

Proof. (Example 0.3.15) By Lemma 0.3.16, it suffices to prove that $2n^2 + n$ is odd if and only if n is odd. We will assume first that n is an odd integer. By definition of an odd integer, there exists an integer k such that $n = 2k+1$. Consequently, we have that

$$2n^2 + n = 2(2k+1)^2 + (2k+1) = 2(4k^2 + 4k + 1) + (2k+1) = 2(4k^2 + 5k + 1) + 1.$$

Considering that $4k^2 + 5k + 1$ is an integer, it follows that $2n^2 + n$ is odd.

Conversely, we will prove the contrapositive of the converse. We must assume to this end that n is an even integer. By definition of an even integer, we have that $n = 2k$ for some integer k. Consequently, we find that $2n^2 + n = 2(2k)^2 + (2k) = 8k^2 + 2k = 2(4k^2 + k)$. Considering that $4k^2 + k$ is an integer, it follows that $2n^2 + n$ is an even integer, as desired. $\square$

Example 0.3.17. Prove that if x and y are real numbers such that $x^3 + xy^2 \leq y^3 + x^2y$, then $x \leq y$.

Proof. We will prove the contrapositive statement. We must assume to this end that x and y are real numbers such that $x > y$. By multiplying this inequality by the non-negative real number y^2, we find that $xy^2 \geq y^3$. Likewise, by multiplying this inequality by the non-negative real number x^2, we find that $x^3 \geq x^2y$. By adding these two inequalities, we conclude that $x^3 + xy^2 \geq y^3 + x^2y$. Considering that $x > y$, one of the real numbers x or y must be nonzero, hence one of the inequalities $xy^2 \geq y^3$ or $x^3 \geq x^2y$ must be strict. Consequently, we conclude that $x^3 + xy^2 > y^3 + x^2y$, as desired. $\square$

0.3.3 Proof by Cases

Consider any open sentence $P(x_1, \ldots, x_n)$ involving the n variables $x_1, \ldots, x_n$ with domain S. **Proof by cases** is an exhaustive proof technique that exploits some "finiteness property" of the set S. Often, this "finiteness property" of S can be realized as one of the following situations.

(a.) We have that S is finite and it is possible to prove the statement $P(x)$ for each element $x \in S$.

(b.) We have that S admits a finite partition $S = S_1 \cup S_2 \cup \cdots \cup S_n$ and it is possible to prove the statement $P(x)$ for each element $x \in S_i$ for each integer $1 \leq i \leq n$.

Concretely, we will illustrate the proof by cases by completing the following typical examples.

Example 0.3.18. Consider the finite subset $S = \{1, \sqrt{2}, 2\sqrt{2}\}$ of $\mathbb{R}$. Prove that for every element $x \in S$, there exists an element $y \in S$ such that $x - y \leq 0$ and $x^2 + y^2$ is a perfect square.

Proof. We may consider the following three cases.

1.) If $x = 1$, then observe that for $y = 2\sqrt{2}$, we have that $x < y$ so that $x - y \leq 0$ and

$$x^2 + y^2 = (1)^2 + \left(2\sqrt{2}\right)^2 = 1 + 8 = 9 = 3^2.$$

2.) If $x = \sqrt{2}$, then observe that for $y = \sqrt{2}$, we have that $x = y$ so that $x - y \leq 0$ and

$$x^2 + y^2 = \left(\sqrt{2}\right)^2 + \left(2\sqrt{2}\right)^2 = 2 + 2 = 4 = 2^2.$$

3.) If $x = 2\sqrt{2}$, then observe that for $y = 2\sqrt{2}$, we have that $x = y$ so that $x - y \leq 0$ and

$$x^2 + y^2 = \left(2\sqrt{2}\right)^2 + \left(2\sqrt{2}\right)^2 = 8 + 8 = 16 = 4^2.$$

We have exhausted all possibilities for an element $x \in S$, hence our proof is complete. □

Example 0.3.19. Consider the finite subset $S = \{2, 3, 4\}$ of $\mathbb{N}$. Prove that for every element $x \in S$ such that $x^2(x-1)^2/4$ is even, we have that $x^2(x+1)^2/4$ is even.

Proof. We may consider the following three cases.

1.) If $x = 2$, then $x^2(x-1)^2/4 = 2^2(2-1)^2/4 = 1$ is not even, so we proceed to the next case.

2.) If $x = 3$, then $x^2(x-1)^2/4 = 3^2(3-1)^2/4 = 9$ is not even, so we proceed to the next case.

3.) If $x = 4$, then each of $x^2(x-1)^2/4$ and $x^2(x+1)^2/4$ have a factor of 4, so they are even.

We have exhausted all possibilities for an element $x \in S$, hence our proof is complete. □

Essentially, a proof by cases for an open sentence $P(x)$ with finite domain S amounts to verifying $P(x)$ for each element $x \in S$. Consequently, there are at most $|S|$ cases in this situation.

We turn our attention next to open sentences that involve integers or elements of other infinite sets. Recall that an integer is either even or odd but not both; the quality that an integer is even or odd is called the **parity** of the integer. Consequently, if we encounter a statement involving an integer, then it is possible to construct a proof by cases by inspecting the situation when n is even and when n is odd separately. We illustrate this idea concretely in the following three examples.

Example 0.3.20. Prove that for every integer n, we have that $n^2 + 3n - 4$ is even.

Proof. We may consider the following two cases.

1.) By definition, if n is even, then there exists an integer k such that $n = 2k$. Consequently, we have that $n^2 + 3n - 4 = (2k)^2 + 3(2k) - 4 = 4k^2 + 6k - 4 = 2(2k^2 + 3k - 2)$. Considering that $2k^2 + 3k - 2$ is an integer, we conclude that $n^2 + 3n - 4$ is an even integer.

2.) By definition, if n is odd, then there exists an integer k such that $n = 2k+1$. Consequently, we have that $n^2+3n-4 = (2k+1)^2+3(2k+1)-4 = (4k^2+4k+1)+(6k+3)-4 = 2(2k^2+5k)$. Considering that $2k^2+10k$ is an integer, we conclude that n^2+3n-4 is an even integer.

We have exhausted all possibilities for the parity of the integer n, hence our proof is complete. □

Example 0.3.21. Prove that any integers x and y have the same parity if and only if $x+y$ is even.

Proof. We will first prove the statement that if x and y are any integers of the same parity, then $x+y$ is even. Consider toward this end the following two cases.

1.) By definition, if the integers x and y are both even, then there exist integers k and ℓ satisfying that $x = 2k$ and $y = 2\ell$. Consequently, we have that $x+y = 2k+2\ell = 2(k+\ell)$. Considering that $k+\ell$ is an integer, we conclude that $x+y$ is even, as desired.

2.) By definition, if the integers x and y are both odd, then there exist integers k and ℓ such that $x = 2k+1$ and $y = 2\ell+1$. Consequently, we have that $x+y = (2k+1)+(2\ell+1) = 2(k+\ell+1)$. Considering that $k+\ell+1$ is an integer, we conclude that $x+y$ is even, as desired.

We have exhausted all possibilities for the parity of the integers x and y, hence the statement holds.

Conversely, we will prove the contrapositive of the statement that if $x+y$ is even, then the integers x and y have the same parity. Explicitly, we will demonstrate that if x and y have opposite parity, then the integer $x+y$ is odd. We may assume **without loss of generality** that x is even and y is odd. Consequently, there exist integers k and ℓ such that $x = 2k$ and $y = 2\ell+1$. Observe that $x+y = 2k+(2\ell+1) = 2(k+\ell)+1$. Because $k+\ell$ is an integer, the integer $x+y$ is odd. □

Remark 0.3.22. We reflect here on two important features of the proof of Example 0.3.21.

1.) First, it is important to note that the biconditional ("if and only if") statement was proved by using a proof by cases for one direction of the biconditional (the "only if" direction) and using a proof by contrapositive for the other direction (the "if" direction). Often, we will be required to use multiple proof techniques in tandem to write a satisfactory proof of a proposition.

2.) We have introduced in the body of the proof of Example 0.3.21 an important phrase in the trade of mathematical writing: "without loss of generality." Essentially, what this means is that the author is asserting to the reader that there is no need to distinguish between the two variables x and y in the above proof: indeed, it does not matter if x is even and y is odd or vice-versa; the result would work the same if the names (or roles) of x and y were swapped. One way to think about the phrase "without loss of generality" is that it can be useful to save the author and the reader precious time if the same (or at least a similar) proof could be used for the other cases that would be necessary to consider in the proof by cases; therefore, one might instead use the phrase, "A similar proof can be used to establish the result."

Example 0.3.23. Prove that $3x+5y+7z$ is odd if exactly two of the integers x, y, z are even.

Proof. Observe that $3x+5y+7z = 2(x+2y+3z)+x+y+z$. Consequently, it suffices to prove that $x+y+z$ is odd by Example 0.3.21: indeed, if $x+y+z$ were even, then $3x+5y+7z$ would be even. Consequently, we may assume without loss of generality that x and y are even and z is odd.

Explicitly, suppose that there exist integers k, ℓ, and m such that $x = 2k$, $y = 2\ell$, and $z = 2m + 1$. We have that $x + y + z = 2k + 2\ell + (2m + 1) = 2(k + \ell + m) + 1$, hence $x + y + z$ is odd. □

Remark 0.3.24. Observe that the proof of Example 0.3.23 is quite clever and drastically reduces the amount of work required to prove the statement. We immediately used the result of Example 0.3.21 to reduce the problem at hand to simply demonstrating that $x+y+z$ is odd whenever exactly two of the integers x, y, and z are even; then, because each of the integers x, y, and z appeared as terms of the sum, there was no need to distinguish between them, so we could appeal to the phrase "without loss of generality" to reduce a proof potentially involving three cases to just one case. Compare this with the amount required to write a proof for Example 0.3.21 with three cases.

Last, a proof by cases can sometimes be used to handle statements involving the union of sets.

Example 0.3.25. Prove that if A, B, and C are sets with $x \in A \cup B$, then $x \in A \cup C$ or $x \in B \cup C$.

Proof. Observe that $x \in A \cup B$ if and only if $x \in A$ or $x \in B$. Consequently, there are two cases.

1.) If $x \in A$, then $x \in A \cup C$ by definition of the set union.

2.) If $x \in B$, then $x \in B \cup C$ by definition of the set union.

Either way, we conclude that $x \in A \cup C$ or $x \in B \cup C$, as desired. □

Remark 0.3.26. We note that in the previous proof, the sets A and B are analogous: indeed, our ultimate objective is to verify a disjunctive statement in the sets $A \cup C$ and $B \cup C$. Consequently, it is possible to use the phrase "without loss of generality" rather than appeal to a proof by cases.

Proof. (Example 0.3.25) Considering that $x \in A \cup B$ if and only if $x \in A$ or $x \in B$, we may assume without loss of generality that $x \in A$. We conclude that $x \in A \cup C$, as desired. □

0.3.4 Counterexamples

Before we are able to prove a statement, we must first deduce that it is true. Often, this amounts to computing several examples to convince ourselves that the statement is valid. Best case scenario, either this practice reveals the nature of a potential proof or a **counterexample** is revealed to us. By counterexample, we mean an explicit instance for which the statement in question is false.

Example 0.3.27. Consider the following conditional statement.

$$P(n)\colon \text{ If } n \text{ is an integer, then } 5n + 4 \text{ is even.}$$

Considering that $5(1) + 4 = 9$ is odd, the conditional statement $P(1)$ is false. Consequently, the integer $n_0 = 1$ provides a counterexample and illustrates that $P(n)$ is not a true statement.

Example 0.3.28. Consider the following conditional statement.

$$P(x)\colon \text{ If } x \text{ is a real number, then } x - e^x > 0.$$

Considering that $0 - e^0 = 0 - 1 = -1 \leq 0$, the conditional statement $P(0)$ is false. Consequently, the real number $x_0 = 0$ provides a counterexample and illustrates that $P(x)$ is not a true statement.

Example 0.3.29. Consider the following conditional statement.

$$P(x)\text{: If } x \text{ is a real number, then } \cot^2(x) + 1 = \csc^2(x).$$

Considering that $\cot(0)$ and $\csc(0)$ are undefined, the conditional statement $P(0)$ is false. Consequently, the real number $x_0 = 0$ provides a counterexample and illustrates that $P(x)$ is not true.

Counterexamples can be astonishingly difficult to determine in many cases: in fact, it is a highly active area of mathematical research to find counterexamples to certain statements of particular interest. Explicitly, the desire for a counterexample is illuminated by the following observation. Consider an open sentence $P(x)$ in a variable x with domain S. Recall that the quantified statement "$\forall x \in S$, $P(x)$" is true if and only if $P(x)$ is true for all elements $x \in S$. By Theorem 0.2.56, if we wish to **disprove** the statement "$\forall x \in S$, $P(x)$" (or show that this statement is false), it suffices to exhibit an element $x_0 \in S$ such that $P(x_0)$ is false. By name, this element x_0 is a counterexample.

Example 0.3.30. Disprove the following conditional statement.

$$P(x)\text{: If } x \text{ is a real number, then } \frac{x^3+1}{x^3-1} = \frac{x^2-x+1}{x^2+x+1}.$$

Solution. Observe that if $x = 1$, then $x^3 - 1 = 0$, hence the left-hand fraction in the statement of $P(x)$ is undefined. Consequently, $x = 1$ is a counterexample to the conditional statement $P(x)$. ⋄

Example 0.3.31. Disprove the following conditional statement.

$$P(x, y)\text{: If } x \text{ and } y \text{ are real numbers, then } x^2 - 4xy + y^2 > 0.$$

Solution. Observe that if $x = 1$ and $y = 1$, then $x^2 - 4xy + y^2 = 1 - 4 + 1 = -2 \leq 0$. Consequently, the ordered pair $(x, y) = (1, 1)$ is a counterexample to the conditional statement $P(x, y)$. ⋄

Example 0.3.32. Disprove the following conditional statement.

$$P(x, y, z)\text{: If } x, y, \text{ and } z \text{ are positive real numbers, then } (x^y)(x^z) = x^{yz}.$$

Solution. Observe that if $x = 2$, $y = 1$, and $z = 3$, then $(x^y)(x^z) = (2^1)(2^3) = 16$ and $x^{yz} = 8$, hence the ordered triple $(x, y, z) = (2, 1, 3)$ is a counterexample to the conditional statement $P(x, y, z)$. ⋄

Be sure to make note of the form we use when solving a problem that asks us to disprove something: we begin with an italicized "Solution" and a period; we exhibit an explicit counterexample to the statement; and we conclude with an empty diamond ⋄ to signify the conclusion of our solution.

0.3.5 Proof by Contradiction

Last but certainly not least, 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 proof technique by a combination of the Law of Excluded Middle, the Law of Non-Contradiction, and Table 14. We bear out the details in two cases of particular interest. We will first assume toward this end that P is a statement that we wish to prove is true.

1.) By the Law of the Excluded Middle, either P is true or P is false.

2.) By the Law of Non-Contradiction, if P is not false, then P is true.

3.) Consequently, in order to demonstrate that P is true, it suffices to prove that P is not false. We assume toward this end that P is in fact false, i.e., we assume that $\neg P$ is true.

4.) By some properties of $\neg P$, it might be possible to derive a contradiction C, i.e., a statement C that is false with respect to all possible truth inputs. Crucially, the contradiction C could reveal itself as a direct consequence of the assumption $\neg P$ or it might be possible to derive a contradiction C from some other known facts (e.g., definitions, propositions, and theorems).

5.) We conclude that the conditional statement $\neg P \Rightarrow C$ is true. But C is false, so $\neg P$ must be false; therefore, our initial assumption that P is false is untenable, so P must be true.

Often, a proof by contradiction is desirable to prove a conditional statement $P \Rightarrow Q$. We outline next how a proof by contradiction for such a statement could be carried out and why it is valid.

1.) By the Law of the Excluded Middle, either $P \Rightarrow Q$ is true or $P \Rightarrow Q$ is false.

2.) By the Law of Non-Contradiction, if $P \Rightarrow Q$ is not false, then $P \Rightarrow Q$ is true.

3.) Consequently, it suffices to prove that $P \Rightarrow Q$ is not false. By Table 7, we must show that if Q is false, then P is false. We assume toward this end that Q is false and P is true.

4.) Like in the case of the proof by contradiction discussed above, it might be possible to derive a contradiction C from some properties of $\neg Q$; this would entail that $\neg Q \Rightarrow C$ is true.

5.) Observe that if C is false and $\neg Q \Rightarrow C$ is true, then $\neg Q$ is false; therefore, our assumption that Q is false is untenable, hence Q must be true so that $P \Rightarrow Q$ is true.

We point out at this time that the writer should always mention in the first line of the proof the proof technique that will be used. Best practice dictates (in the case of a proof by contradiction) that this is achieved using the phrase, "Suppose on the contrary that P is true and Q is false" or, "We will assume toward a contradiction that P is true and Q is false." Even more, the writer should take care to point out exactly what contradiction is derived in a proof by contradiction.

Example 0.3.33. Prove that there is no smallest integer.

Proof. Suppose on the contrary that n is the smallest integer. Observe that $n-1$ is an integer. By adding n to both sides of the inequality $-1 < 0$, we find that $n-1 < n$. But this is a contradiction: if n is the smallest integer, there can be no integer less than n. Our assumption that there exists a smallest integer is therefore untenable, hence we conclude that there is no smallest integer. □

Example 0.3.34. Prove that no integer is both even and odd.

Proof. Suppose on the contrary that n is an even integer that is also odd. By definition of an even integer, there exists an integer k such that $n = 2k$. By definition of an odd integer, there exists an integer ℓ such that $n = 2\ell + 1$. Considering that $n = n$ is a tautology, it follows that $2k = 2\ell + 1$ so

that $1 = 2k - 2\ell = 2(k - \ell)$. By dividing both sides of this equation by 2, we find that $k - \ell = \frac{1}{2}$. But this is a contradiction: the difference of two integers is an integer, but the rational number $\frac{1}{2}$ is not an integer. Our assumption that there exists an integer that is both even and odd is therefore untenable, hence we conclude that no integer is both even and odd. □

Example 0.3.35. Prove that no even integer is the sum of three odd integers.

Proof. Suppose on the contrary that n is an even integer that is the sum of three odd integers a, b, and c. By definition of an odd integer, we have that $a = 2k + 1$, $b = 2\ell + 1$, and $c = 2m + 1$ for some integers k, ℓ, and m. Considering that $n = a + b + c$, it follows that

$$n = (2k + 1) + (2\ell + 1) + (2m + 1) = 2(k + \ell + m + 1) + 1,$$

hence n is odd. But this is a contradiction: by Example 0.3.34, we have that no integer is both even and odd. Our assumption that there exists an even integer that is the sum of three odd integers is untenable, hence we conclude that no even integer is the sum of three odd integers. □

Example 0.3.36. Prove that if a, b, and c are integers such that $a^2 + b^2 = c^2$, then a or b is even.

Proof. Suppose on the contrary that a and b are both odd integers. By definition of an odd integer, we have that $a = 2k + 1$ and $b = 2\ell + 1$ for some integers k and ℓ. Consequently, we find that

$$c^2 = a^2 + b^2 = (2k + 1)^2 + (2\ell + 1)^2 = (4k^2 + 4k + 1) + (4\ell^2 + 4\ell + 1) = 2(2k^2 + 2k + 2\ell^2 + 2\ell + 1).$$

Considering that $2k^2 + 2k + 2\ell^2 + 2\ell + 1$ is an integer, we conclude that c^2 is even so that c is even. By definition of an even integer, we have that $c = 2m$ for some integer m so that

$$4m^2 = (2m)^2 = c^2 = 2(2k^2 + 2k + 2\ell^2 + 2\ell + 1).$$

Cancelling one factor of 2 from both sides of this equation yields that

$$2m^2 = 2k^2 + 2k + 2\ell^2 + 2\ell + 1 = 2(k^2 + k + \ell^2 + \ell) + 1.$$

But this is a contradiction: the left-hand side shows an even integer, but the right-hand side shows an odd integer. Our assumption that a and b are odd is untenable, hence a or b is even. □

Example 0.3.37. Prove that if x is even and y is odd, then $x^2 + 2y^2$ is not divisible by 4.

Proof. Suppose on the contrary that x is an even integer and y is an odd integer such that $x^2 + 2y^2$ is divisible by 4. By definition of the parity of an integer, we have that $x = 2k$ and $y = 2\ell + 1$ for some integers k and ℓ. Consequently, we may simplify the expression $x^2 + 2y^2$ to find that

$$x^2 + 2y^2 = (2k)^2 + (2\ell + 1)^2 = 4k^2 + 2(4\ell^2 + 4\ell + 1) = 4(k^2 + 2\ell^2 + 2\ell) + 2.$$

By assumption that $x^2 + 2y^2$ is divisible by 4, there exists an integer m such that $x^2 + 2y^2 = 4m$. Combined with our previous displayed equation, this yields that

$$4m = 4(k^2 + 2\ell^2 + 2\ell) + 2,$$

from which we deduce that $2 = 4m - 4(k^2 + 2\ell^2 + 2\ell) = 4(m - k^2 - 2\ell^2 - 2\ell)$. By cancelling a factor of 2 from both sides, we find that $1 = 2(m - k^2 - 2\ell^2 - 2\ell)$. But this is a contradiction: the integer 1 is odd, so it cannot be divisible by 2 by Example 0.3.34. Our assumption that x is an even integer and y is an odd integer such that $x^2 + 2y^2$ is divisible by 4 is therefore untenable, hence we conclude that if x is an even integer and y is an odd integer, then $x^2 + 2y^2$ is not divisible by 4. □

Example 0.3.38. Prove that $\sqrt{2}$ is irrational.

Proof. Suppose on the contrary that $\sqrt{2}$ is rational. By definition of a rational number, there exist integers a and b such that b is nonzero; a and b possess no common factors other than ± 1; and

$$\sqrt{2} = \frac{a}{b}.$$

By squaring both sides of this equation and clearing the denominator, we find that

$$a^2 = 2b^2.$$

Consequently, the integer a^2 is even. Considering that the square of an integer is even if and only if that integer is even, it follows that a is even so that $a = 2k$ for some integer k. By substituting this identity back into our above displayed equation, we find that

$$4k^2 = (2k)^2 = a^2 = 2b^2.$$

Cancelling a factor of 2 from both sides yields that b^2 is an even integer since

$$b^2 = 2k^2.$$

By the same rationale as before, we conclude that b is even so that $b = 2\ell$ for some integer ℓ. But this is a contradiction: we had originally assumed that a and b possess no common factors other than ± 1, but if a and b are both even, then they have a common factor of 2. Our assumption that $\sqrt{2}$ is rational is therefore untenable, hence we conclude that $\sqrt{2}$ is irrational. □

0.3.6 Existence Proofs

Complementary to counterexamples, proving the existence of certain mathematical objects or structures with desirable properties is also a foremost concern throughout mathematics. We remind the reader at this point that an existence statement is a quantified statement of the form

$$\exists x \in S,\, P(x)\colon \text{ There exists an element } x \in S \text{ such that } P(x).$$

for some open sentence $P(x)$ in a variable x with domain S. Consequently, in order to determine the verity of an existence statement, it suffices to provide an explicit example of an element $x_0 \in S$ such that $P(x_0)$ is true; if this is possible, then the attendant proof of the existence statement is called a **constructive proof** because the element $x_0 \in S$ is often "constructed" or produced by explicitly performing some algebraic manipulation or computation. We provide some examples below.

Example 0.3.39. Prove that there exists an integer whose cube is equal to its square.

Solution. Before we prove this existence statement, we may find it beneficial to write the statement in symbols. Observe that if n is an integer, then n^3 is its cube and n^2 is its square. Consequently,

$$P(n)\colon \text{ The integer } n \text{ satisfies that } n^3 = n^2.$$

is the open sentence that n is an integer whose cube is equal to its square. Ultimately, we are trying to prove the following existentially quantified statement in the variable n over the domain $\mathbb{Z}$.

$$\exists n \in \mathbb{Z},\ P(n)\colon \text{ There exists an integer } n \text{ such that } n^3 = n^2.$$

Observe that if $n^3 = n^2$, then $n^3 - n^2 = 0$ so that $n^2(n-1) = 0$. By the Zero Product Property, it follows that $n = 0$ or $n = 1$. Either one of these integers provides an explicit solution to the integer equation $n^3 = n^2$, hence we have the ingredients to write a constructive proof for the statement. $\diamond$

Proof. Observe that the integer $n = 1$ satisfies that $n^3 = 1^3 = 1 = 1^2 = n^2$, and the claim holds. $\square$

Example 0.3.40. Prove that there exist real numbers x and y such that $(x+y)^2 = x^2 + y^2$.

Solution. Before we determine a proof of the statement, we note that we seek to establish the verity of the following existential statement in the variables x and y over the domain $\mathbb{R}$.

$$\exists x, y \in \mathbb{R},\ P(x,y)\colon \text{ There exist real numbers } x \text{ and } y \text{ such that } (x+y)^2 = x^2 + y^2.$$

Observe that if $(x+y)^2 = x^2 + y^2$, then $x^2 + 2xy + y^2 = x^2 + y^2$ so that $2xy = 0$. By the Zero Product Property, it follows that $x = 0$ or $y = 0$. Either way, the statement that $(x+y)^2 = x^2 + y^2$ will be true for any value of the variables x and y so long as one of them is zero: indeed, if $y = 0$, then $(x+y)^2 = (x+0)^2 = x^2 = x^2 + 0^2 = x^2 + y^2$. We have the makings of a constructive proof. $\diamond$

Proof. Observe that the real numbers $x = 1$ and $y = 0$ satisfy that

$$(x+y)^2 = (1+0)^2 = 1^2 = 1^2 + 0^2 = x^2 + y^2.$$

Consequently, the statement in question holds, and our proof is complete. $\square$

Example 0.3.41. Prove that $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$ has at least one real root.

Solution. By definition of a root of a function, the statement we are tasked to prove is as follows.

$$\exists x \in \mathbb{R},\ P(x)\colon \text{ There exists a real number } x \text{ such that } f(x) = 0.$$

Observe that $f(-1) = (-1)^5 + (-1)^4 + (-1)^3 + (-1)^2 + (-1) + 1 = -3 + 3 = 0$, hence a direct proof is possible because we have found an explicit example of a real root of $f(x)$. $\diamond$

Proof. Observe that the real number $x = -1$ is a root of $f(x)$ since we have that

$$f(-1) = (-1)^5 + (-1)^4 + (-1)^3 + (-1)^2 + (-1) + 1 = -3 + 3 = 0. \qquad \square$$

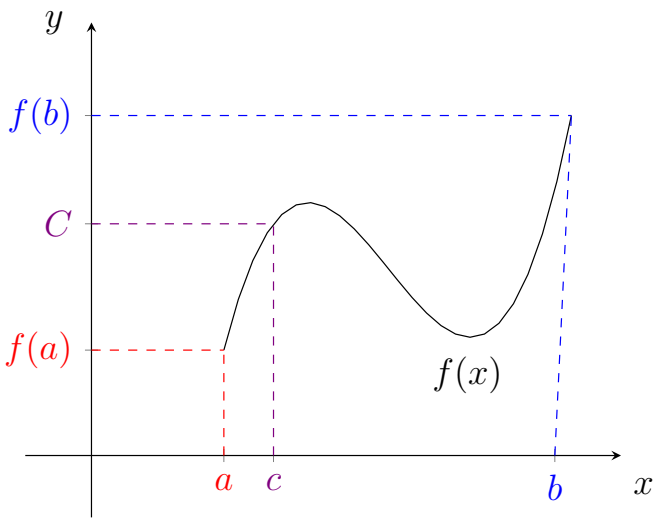
Remark 0.3.42. We make an important and necessary observation about the serendipitous nature of the existence proof provided in Example 0.3.41. Exactly how did we stumble upon the real number $x = -1$, and why did we suspect that it is a root of the polynomial $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$? Unfortunately, this was simply a lucky coincidence — the product of years of schema and knowing where to look. One natural starting point in the aforementioned example is to begin by plugging in integer values of x near zero. Plugging in $x = 0$ yields that $f(0) = 0$, and plugging in $x = 1$ yields that $f(1) = 6$ — both failures. We were exceedingly lucky that our next guess $x = -1$ worked.

Generally, the roots of a real function $f(x)$ are seriously difficult to compute. By the Quadratic Formula, the roots of any real function of the form $f(x) = ax^2 + bx + c$ with a nonzero are known; there are also the Cubic Formula and the Quartic Formula, but these are typically not taught, and students are not expected to know them (the author freely admits to not knowing them, either). Beyond that, it is a landmark result of Galois Theory that there is no closed form expression for the roots of a real polynomial of degree at least five. Consequently, there is little hope for deducing the roots of a polynomial of degree five or larger — let alone trying to find the roots of a real function that is not a polynomial (other than certain trigonometric, inverse trigonometric, or logarithmic functions) — for students in this course without specialized knowledge (such as the Newton-Raphson Method or other recursive numerical methods for finding roots of differentiable functions).

Even still, using elementary calculus, there is a way to determine existence of roots of continuous functions without ever knowing exactly what those roots are! Before we provide a proof along these lines, we must first recall the following important fact about continuous functions from Calculus I.

Theorem 0.3.43 (Intermediate Value Theorem)**.** *Every real univariate function $f : D_f \to \mathbb{R}$ with domain $D_f \subseteq R$ that is continuous on a closed and bounded interval $[a, b]$ satisfies that for every real number C between $f(a)$ and $f(b)$, there exists a real number c such that $a \leq c \leq b$ and $f(c) = C$.*

Concretely, the Intermediate Value Theorem states any every real function $f(x)$ that is continuous on a closed and bounded interval $[a, b]$ achieves every possible y-value between $f(a)$ and $f(b)$ for some x-value between a and b. Graphically, the intuition is that a continuous function can be represented visually by drawing without lifting one's pencil, hence as the curve $y = f(x)$ is traced out from $x = a$ to $x = b$ along the x-axis, every real number along the y-axis between $f(a)$ and $f(b)$ must correspond to some point on the x-axis. Consider the picture below for an illustration.



Consequently, the upshot is that in order to prove the existence of roots of a continuous function $f(x)$, we may find real numbers a and b such that $f(a)$ and $f(b)$ have opposite sign, i.e., $f(a) < 0$ and $f(b) > 0$ (or vice-versa); then, because $f(x)$ is a continuous function such that $f(a)$ and $f(b)$ have opposite sign, there must exist a real number c such that $a \leq c \leq b$ and $f(c) = 0$. We refer to such a proof of the existence of the roots of a continuous function as a **non-constructive proof**: in fact, we are not explicitly exhibiting the roots of the function. We are instead simply relying on the Intermediate Value Theorem to conclude that some root must exist. Generally, a non-constructive proof relies on some well-known fact, theorem, or definition. Consequently, a non-constructive proof may not be direct. We conclude this section with several examples of non-constructive proofs.

Example 0.3.44. Prove that $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$ has at least one real root.

Proof. Observe that the polynomial $f(x)$ is a continuous function. Considering that $f(-2) = -21$ and $f(0) = 1$, by the Intermediate Value Theorem, there exists a real number c such that $-2 < c < 0$ and $f(c) = 0$. By definition, the real number c is a root of the polynomial $f(x)$, as desired. □

Example 0.3.45. Prove that $f(x) = e^x - 3x$ has at least one real root.

Proof. Observe that $f(x)$ is a continuous function since it is the difference of the continuous functions e^x and $3x$. Considering that $f(1) = e - 3 < 0$ and $f(0) = 1 > 0$, by the Intermediate Value Theorem, there exists a real number c such that $0 < c < 1$ and $f(c) = 0$, as desired. □

Example 0.3.46. Prove that $\cos(x) - \sin(x) = \frac{1}{2}$ for some real number x such that $0 \leq x \leq \frac{\pi}{4}$.

Proof. Consider the function $f(x) = \cos(x) - \sin(x)$. Observe that $f(x)$ is continuous because it is the difference of the continuous functions $\cos(x)$ and $\sin(x)$. Considering that

$$f(0) = 1 - 0 = 1 \geq 0 = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} = f\left(\frac{\pi}{4}\right),$$

by the Intermediate Value Theorem, there exists a real number $0 \leq c \leq \frac{\pi}{4}$ and $f(c) = \frac{1}{2}$. □

Example 0.3.47. Prove that some digit appears infinitely often in the decimal expansion of π.

Solution. Before we outline a proof strategy, we note the novelty of the example. We cannot use the Intermediate Value Theorem because the statement does not involve a continuous function. We cannot verify directly that some digit appears infinitely often in the decimal expansion of π because we cannot check the infinitely many digits in the decimal expansion of π. Consequently, there is no hope for a constructive proof. Our statement is not conditional, so there is no contrapositive. We require a proof by contradiction! Observe that the negation of the above statement is, "No digit appears infinitely often in the decimal expansion of π," or, "Every digit in the decimal expansion of π appears finitely many times." Considering that π is irrational, this statement cannot be true, so the statement we seek to prove must be true by the Law of Excluded Middle and the Law of Non-Contradiction. We did not check any digits of π, hence this is a non-constructive proof. ◇

Proof. On the contrary, suppose that every digit in the decimal expansion of π appears finitely many times. Considering that the digits in the decimal expansion of any real number are $0, 1, 2, \ldots, 9$, the decimal expansion of π contains at most ten digits. But this is a contradiction: if the decimal

expansion of π is finite, then π is a rational number; however, π is irrational! Consequently, our assumption that every digit in the decimal expansion of π appears finitely many times is untenable, hence we conclude that some digit in the decimal expansion of π appears infinitely many times. $\square$

0.4 Proofs in the Wild

Once we are satisfactorily acquainted with the basic proof strategies outlined in the previous section, we may consider examples and write proofs in a variety of familiar mathematical contexts. We aim throughout this section to employ the techniques of the previous three sections as they pertain to the study of combinatorics, elementary number theory, modern algebra, and naïve set theory.

0.4.1 Principle of Mathematical Induction

Consider any open sentence $P(n)$ defined for a variable n with domain $S \subseteq \mathbb{Z}$. Observe that if S admits a smallest element n_0, then we may denote $n_0 = \min\{n \mid n \in S\}$ since it is the minimum element of S. We have seen in Section 0.3 that it may be possible to prove the quantified statement $\forall n \in S,\ P(n)$ by cases; however, this may be tedious if S is finite and $|S|$ is large, and it may not be clear why the statement $P(n)$ is true even if we assume that n is either even or odd. Consequently, we may require another technique all together to demonstrate that $P(n)$ is true for all $n \in S$.

We turn our attention thus to one of the most useful proof techniques for establishing the verity of universally quantified statements defined forintegers: a **proof by induction** appeals to one of the three forms of the **Principle of Mathematical Induction**. Before we proceed to the definition, let us explore some examples of properties of integers for which a proof by induction is appropriate.

Example 0.4.1. Consider the sum of the first n consecutive odd positive integers.

$$o(n) = 1 + 3 + 5 + \cdots + (2n-1) = \sum_{k=1}^{n} (2k-1)$$

Computing the values of $o(n)$ for the first four positive integers $1 \leq n \leq 4$ yields that $o(1) = 1$, $o(2) = 1 + 3 = 4$, $o(3) = 1 + 3 + 5 = 9$, $o(4) = 1 + 3 + 5 + 7 = 16$, and so on.

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

Table 20: the sum of first five 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 0.4.2. We have that $o(n) = n^2$ for all integers $n \geq 1$ for $o(n)$ as in Example 0.4.1.

Observe that $o(1) = 1^2$ and $o(n+1) = o(n) + (2n+1)$, hence if we were to assume that $o(n) = n^2$ for some integer $n \geq 1$, then we would conclude that $o(n+1) = n^2 + 2n + 1 = (n+1)^2$. We will soon return to validate this idea as one of the tenants of the Principle of Mathematical Induction!

Example 0.4.3. Consider the sum of the first n consecutive positive integers.

$$c(n) = 1 + 2 + 3 + \cdots + n = \sum_{k=1}^{n} k$$

Computing the values of $c(n)$ for the first four positive integers $1 \leq n \leq 4$ yields that $c(1) = 1$, $c(2) = 1 + 2 = 3$, $c(3) = 1 + 2 + 3 = 6$, and $c(4) = 1 + 2 + 3 + 4 = 10$, and so on.

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

Table 21: the sum of the first five consecutive positive integers

Unfortunately, the pattern here is not obvious; however, due to a young Gauss, the following strategy can be employed. Briefly put, the idea is to write down the sum $1 + 2 + 3 + \cdots + n$ both forwards and backwards, adding each column of the sum to determine the value of $2(1 + 2 + 3 + \cdots + n)$.

$$\begin{array}{ccccccccccc} & 1 & + & 2 & + & 3 & + & \cdots & + & n \\ + & n & + & (n-1) & + & (n-2) & + & \cdots & + & 1 \\ \hline & (n+1) & + & (n+1) & + & (n+1) & + & \cdots & + & (n+1) \end{array}$$

Considering that there are n columns in this table and the sum of each column is $n+1$, we conclude that $2(1 + 2 + 3 + \cdots + n) = n(n + 1)$. Consequently, we have the following conjecture.

Conjecture 0.4.4. We have that $c(n) = \frac{n(n+1)}{2}$ for all integers $n \geq 1$ for $c(n)$ as in Example 0.4.3.

Like before, we can readily verify the facts that $c(1) = 1 = \frac{1 \cdot 2}{2}$ and $c(n + 1) = c(n) + (n + 1)$, hence if we were to assume that $c(n) = \frac{n(n+1)}{2}$ for some integer $n \geq 1$, 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 0.4.5 (Principle of Ordinary Induction)**.** Given any integer n_0, consider any open sentence $P(n)$ defined for all integers $n \geq n_0$. We may define the following criteria.

(a.) We have that $P(n_0)$ is a true statement.

(b.) If $P(n)$ is a true statement for some integer $n \geq n_0$, then $P(n + 1)$ is a true statement.

Provided that both of these statements are true, it follows that $P(n)$ is true for all integers $n \geq n_0$.

By the Principle of Ordinary Induction, we can return to prove Conjectures 0.4.2 and 0.4.4.

Proof. (Conjecture 0.4.2) Consider the following open sentence defined for all integers $n \geq 1$.

$$P(n)\text{: We have that } 1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

We will prove that $P(n)$ is true for all integers $n \geq 1$, i.e., we will prove that "$\forall n \in \mathbb{Z}_{\geq 1}$, $P(n)$" is true. We proceed by the Principle of Ordinary Induction. We must verify the following conditions.

(a.) Observe that $P(1)$ is a true statement because it holds that $1 = 1^2$.

(b.) We will assume that $P(n)$ is true for some integer $n \geq 1$. Consequently, we have that

$$1+3+5+\cdots+(2n+1) = 1+3+5+\cdots+(2n-1)+(2n+1) = n^2+2n+1 = (n+1)^2.$$

Considering that (a.) $P(1)$ is a true statement and (b.) $P(n+1)$ is true whenever $P(n)$ is true for some integer $n \geq 1$, our proof is complete by the Principle of Ordinary Induction. □

Proof. (Conjecture 0.4.4) Consider the following open sentence defined for all integers $n \geq 1$.

$$P(n)\text{: We have that } 1+2+3+\cdots+n = \frac{n(n+1)}{2}.$$

We will prove that $P(n)$ is true for all integers $n \geq 1$, i.e., we will prove that "$\forall n \in \mathbb{Z}_{\geq 1}$, $P(n)$" is true. We proceed by the Principle of Ordinary Induction. We must verify the following conditions.

(a.) Observe that $P(1)$ is a true statement because it holds that $1 = \frac{1\cdot 2}{2}$.

(b.) We will assume that $P(n)$ is true for some integer $n \geq 1$. Consequently, we have that

$$1+2+3+\cdots+n+(n+1) = \frac{n(n+1)}{2}+(n+1) = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$

Considering that (a.) $P(1)$ is a true statement and (b.) $P(n+1)$ is true whenever $P(n)$ is true for some integer $n \geq 1$, our proof is complete by the Principle of Ordinary Induction. □

Going forward, we will begin any inductive proof by simply stating our intention to use a proof by induction; however, we will not typically make any explicit reference to the open sentence $P(n)$ that we intend to prove, and we will abbreviate the steps in an inductive proof under the assumption that our intended audience is familiar with induction. We illustrate a typical proof by induction.

Example 0.4.6. Prove that $2^n > n^2$ for all integers $n \geq 5$.

Proof. We proceed by induction. Observe that $2^5 = 32 > 25 = 5^2$, hence the claim holds for $n = 5$. We will assume inductively that $2^n > n^2$ for some integer $n \geq 5$. By hypothesis, we have that

$$2^{n+1} = 2 \cdot 2^n > 2 \cdot n^2,$$

so it suffices to prove that $2n^2 \geq (n+1)^2$. Considering that $n \geq 5$ by our inductive hypothesis, we have that $n^2 \geq 5n$ and $5n = 4n+n \geq 4n+5 \geq 2n+1$ so that

$$2n^2 = n^2+n^2 \geq n^2+5n \geq n^2+2n+1 = (n+1)^2.$$

We conclude by induction that $2^n > n^2$ for all integers $n \geq 5$. □

Occasionally, it is desirable to strengthen the hypotheses of the Principle of Ordinary Induction in order to simplify proofs defined for induction. Currently, we may view induction as a property of falling dominoes: (a.) if the n_0th domino falls and (b.) the nth domino falling causes the $(n+1)$th domino to fall, then as the n_0th domino falls, all consecutive dominoes after it will fall. But suppose that we could knock down all dominoes from the n_0th 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 0.4.7 (Principle of Complete Induction)**.** Given any integer n_0, consider any open sentence $P(n)$ defined for all integers $n \geq n_0$. We may define the following criteria.

(a.) We have that $P(n_0)$ is a true statement.

(b.) If $P(k)$ is a true statement for each integer $n_0 \leq k \leq n$, then $P(n+1)$ is a true statement.

Provided that both of these statements are true, it follows that $P(n)$ is true for all integers $n \geq n_0$.

Even though the criteria of the Principle of Complete Induction ostensibly appear to be much stronger than the Principle of Ordinary Induction, the two principles are in fact materially equivalent (see Exercise 0.6.23). Last, we obtain another crucial tool that is ubiquitous in mathematics.

Theorem 0.4.8 (Well-Ordering Principle)**.** *Every nonempty set of non-negative integers admits a smallest element with respect to the total order $\leq$ on the real numbers. 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 $P(k)$ is true for each integer $1 \leq k \leq n$. By definition of $P(k)$, 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 this theorem is materially equivalent to both ordinary induction and complete induction (see Exercise 0.6.24). Combined, the Principle of Ordinary Induction, the Principle of Complete Induction, and the Well-Ordering Principle constitute the triumvirate of the Principle of Mathematical Induction.

Before we conclude this section, we provide an example using the Well-Ordering Principle.

Example 0.4.9. Prove that every integer is of the form $2^a b$ for some integers $a \geq 0$ and b odd.

Proof. Certainly, every odd integer n is of the form $n = 2^a b$ since we may take $a = 0$ and $b = n$ in this case. Conversely, if n is even, then there exists an integer k such that $n = 2k$. Observe that if k is odd, then our proof is complete since we may take $a = 1$ and $b = k$ in this case. Otherwise, we must have that k is even, hence we may repeat the same argument for k. Continuing in this manner yields a strictly decreasing sequence $|n| > |k| > \cdots > |b|$ of a positive integers that must eventually terminate in some odd integer b by the Well-Ordering Principle. We conclude that $n = 2^a b$. □

Remark 0.4.10. Canonically, any proof with non-negative integers that invokes the Well-Ordering Principle to ensure the termination of a repeating process is considered a proof by infinite descent (or Fermat's Method of Descent). Crucially, the Well-Ordering Principle ensures there is no infinite strictly decreasing sequence of non-negative integers, hence every process involving non-negative integers that ostensibly results in "infinite descent" must eventually terminate. Classically, such a proof was structured as a proof by contradiction, assuming that an infinite process were possible.

0.4.2 Divisibility Properties of Integers

We say that a nonzero integer a **divides** an integer b if there exists an integer q such that $b = aq$. We will write $a \mid b$ in this case, and we will typically say that b is **divisible by** a. Conversely, the **divisors** of b are the nonzero integers a that divide b. We are already familiar with this notion from Section 0.1.5, but for illustrative purposes, we note that the integers 1, 2, 3, 4, 6, and 12 divide 12 (i.e., the divisors of 12 are the integers 1, 2, 3, 4, 6, and 12) because $12 = 1 \cdot 12 = 2 \cdot 6 = 3 \cdot 4$. We say that an integer $p \geq 2$ is **prime** if its only positive divisors are 1 and p. Conversely, an integer $n \geq 2$ that admits positive divisors other than 1 and n is **composite**. Quite useful is the following property of the divisors of composite integers that provides a characterization of compositeness.

Theorem 0.4.11 (Factorization Criterion for Composite Integers)**.** *Given any integer $n \geq 2$, we have that n is composite if and only if $n = ab$ for some integers a and b such that $2 \leq a, b \leq n - 1$.*

Proof. Observe that if $n \geq 2$ is composite, then by definition, we may write $n = ab$ for some positive integers a and b such that a is neither 1 nor n. Considering that $b \geq 1$, it follows that $n = ab \geq a$ so that $2 \leq a \leq n - 1$ by hypothesis that a is neither 1 nor n. Conversely, if $n \geq 2$ admits positive integers a and b such that $n = ab$ and $2 \leq a \leq n - 1$, then n must be composite by definition. □

We will soon see that primes form the "building blocks" for all integers. Explicitly, every integer $n \geq 2$ can be written as a product of primes. We refer to such an expression of an integer as a product of its prime factors as the **prime factorization** of the integer. Observe that $12 = 4 \cdot 3 = 2^2 \cdot 3$ is the prime factorization of 12 and $30 = 2 \cdot 15 = 2 \cdot 3 \cdot 5$ is the prime factorization of 30. Before we are able to prove that every integer $n \geq 2$ admits a unique prime factorization, we set out to develop some basic tools for understanding divisibility of integers. Our first task is to verify the following.

Proposition 0.4.12 (Properties of Divisibility of Integers)**.** *Consider any nonzero integers a and b and any integers c and d. Each of the following properties of divisibility of integers holds.*

1.) (**Product Property**) *If $a \mid c$ or $a \mid d$, then $a \mid cd$.*

2.) (**Transitive Property**) *If $a \mid b$ and $b \mid c$, then $a \mid c$.*

3.) (**Homogeneity Property**) *If $a \mid c$ and $b \mid d$, then $ab \mid cd$.*

4.) (**Linearity Property**) *If $a \mid c$ and $a \mid d$, then $a \mid (cx + dy)$ for any integers x and y.*

Proof. We may prove each statement in turn directly by appealing to the definition of divisibility.

1.) We may assume without loss of generality that $a \mid c$ since $cd = dc$. By definition, if a divides c, then there exists an integer q such that $c = aq$. Consequently, we have that $cd = (aq)d = a(dq)$. Considering that dq is an integer because d and q are integers, we conclude that a divides cd.

2.) By definition, if $a \mid b$ and $b \mid c$, then $b = aq$ and $c = br$ for some integers q and r. We conclude by substitution that $c = br = (aq)r = a(qr)$ is divisible by a because qr is an integer.

3.) By definition, if $a \mid c$ and $b \mid d$, then $c = aq$ and $d = br$ for some integers q and r. We conclude by substitution that $cd = (aq)(br) = ab(qr)$ is divisible by ab because qr is an integer.

4.) By definition, if $a \mid c$ and $a \mid d$, then $c = aq$ and $d = ar$ for some integers q and r so that

$$cx + dy = (aq)x + (ar)y = a(qx) + a(ry) = a(qx + ry)$$

for any integers x and y. Considering that $qx + ry$ is an integer, we find that $a \mid (cx + dy)$. □

Proposition 0.4.13 (Divisibility and Absolute Value)**.** *Consider any nonzero integers a and b. Each of the following properties relating divisibility of integers and the absolute value function holds.*

1.) (**Divisibility Decreases Absolute Value**) *If $a \mid b$, then $|a| \leq |b|$.*

2.) (**Divisibility Detects Absolute Value**) *If $a \mid b$ and $b \mid a$, then $|a| = |b|$.*

Proof. We will prove the first statement; the second statement then follows from the first statement by noting that if $a \mid b$ and $b \mid a$, then $|a| \leq |b|$ and $|b| \leq |a|$ so that equality holds. By definition, if $a \mid b$, then there exists an integer q such that $b = aq$. Even more, by assumption that b is nonzero, we must have that $|q| \geq 1$. Consequently, we conclude that $|b| = |aq| = |a||q| \geq |a|$, as desired. □

Remark 0.4.14. Each of the properties of divisibility of integers we have discussed thus far can be phrased in terms of congruence modulo some nonzero integers a and b. We remind the reader that a pair of integers c and d are congruent modulo a nonzero integer n if and only if n divides $d - c$ if and only if $n \mid (d - c)$. Conventionally, if c and d are congruent modulo n, we write $d \equiv c \pmod{n}$.

1.) (**Product Property**) If $c \equiv 0 \pmod{a}$ or $d \equiv 0 \pmod{a}$, then $cd \equiv 0 \pmod{a}$.

2.) (**Transitive Property**) If $b \equiv 0 \pmod{a}$ and $c \equiv 0 \pmod{b}$, then $c \equiv 0 \pmod{a}$.

3.) (**Homogeneity Property**) If $c \equiv 0 \pmod{a}$ and $d \equiv 0 \pmod{b}$, then $cd \equiv 0 \pmod{ab}$.

4.) (**Linearity Property**) If $c \equiv 0 \pmod{a}$ and $d \equiv 0 \pmod{a}$, then $cx + dy \equiv 0 \pmod{a}$.

5.) (**Divisibility Decreases Absolute Value**) If $b \equiv 0 \pmod{a}$, then $|a| \leq |b|$.

6.) (**Divisibility Detects Absolute Value**) If $b \equiv 0 \pmod{a}$ and $a \equiv 0 \pmod{b}$, then $|a| = |b|$.

Even with this very basic notion of divisibility, there are many interesting examples to consider.

Example 0.4.15. Prove that if a, b, c are integers, a and b nonzero, $a^2 \mid b$, and $b^3 \mid c$, then $a^6 \mid c$.

Proof. By definition, if $a^2 \mid b$, then there exists an integer q such that $b = a^2q$. Likewise, if $b^3 \mid c$, then there exists an integer r such that $c = b^3r$. Considering that $b = a^2q$, we find that $b^3 = (a^2q)^3 = a^6q^3$ so that $c = b^3r = (a^6q^3)r = a^6(q^3r)$. We conclude that $a^6 \mid c$ because q^3r is an integer. □

Example 0.4.16. Prove that for any integers a and b, if $2 \mid ab$, then $2 \mid a$ or $2 \mid b$.

Proof. We will prove the contrapositive. We may assume to this end that 2 does not divide either a or b. Consequently, the integers a and b must be odd, hence there exist integers k and ℓ such that $a = 2k + 1$ and $b = 2\ell + 1$ so that $ab = (2k + 1)(2\ell + 1) = 4k\ell + 2k + 2\ell + 1 = 2(2k\ell + k + \ell) + 1$. Considering that $2k\ell + k + \ell$ is an integer, we conclude that 2 does not divide ab. □

Remark 0.4.17. Even without the notion of divisibility of integers, we could have established the result of the preceding example using the notion of parity of integers: indeed, we claim that if ab is even for some integers a and b, then a or b must be even. Compare with the third line of the proof.

Example 0.4.18. Prove that if n is an integer such that $7 \mid 4n$, then $7 \mid n$.

Proof. By definition, if $7 \mid 4n$, then there exists an integer q such that $4n = 7q$. Considering that $4n$ is even, we must have that q is even; otherwise, if q were odd, then $q = 2k + 1$ for some integer k so that $4n = 7q = 7(2k + 1) = 2(7k + 3) + 1$ is odd — a contradiction. Consequently, there exists an integer k such that $q = 2k$ and $4n = 7q = 14k$. Cancelling one factor of 2 from each side of this identity yields that $2n = 7k$ so that $2 \mid 7k$. By Example 0.4.16, we conclude that $2 \mid k$. Consequently, there exists an integer ℓ such that $k = 2\ell$ and $q = 2k = 4\ell$. Considering that $4n = 7q$, we find that $4n = 7(4\ell)$. Cancelling one factor of 4 from each side yields $n = 7\ell$ so that $7 \mid n$. □

Remark 0.4.19. Examples 0.4.16 and 0.4.18 can be generalized to demonstrate that for any integers a and b and any prime p, we have that $p \mid ab$ if and only if $p \mid a$ or $p \mid b$ (see Exercise 0.6.26).

Example 0.4.20. Prove that if n if an integer such that $2 \mid (n^2 + 3)$, then $4 \mid (n^2 + 3)$.

Proof. By definition, if $2 \mid (n^2 + 3)$, then there exists an integer k such that $n^2 + 3 = 2k$, hence we have that $n^2 = 2k - 3 = 2(k - 2) + 1$ is odd so that n is odd. Consequently, we have that

$$n^2 + 3 = (2\ell + 1)^2 + 3 = (4\ell^2 + 4\ell + 1) + 3 = 4(\ell^2 + \ell + 1)$$

for some integer ℓ. Considering that $\ell^2 + \ell + 1$ is an integer, we conclude that $4 \mid (n^2 + 3)$. □

Given any nonzero integers a and b, we say that a nonzero integer c is a **common divisor** of a and b if and only if $c \mid a$ and $c \mid b$, i.e., c divides a and c divides b. We distinguish among all common divisors of a and b the unique **greatest common divisor** $d = \gcd(a, b)$ of a and b satisfying that

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

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

Consequently, $\gcd(a, b)$ is the "largest" common divisor of a and b with respect to divisibility.

Example 0.4.21. Consider the integers $a = 12$ and $b = 30$. By writing down the prime factorizations of a and b, their greatest common divisor can be easily determined. Observe that $12 = 2^2 \cdot 3$ and $30 = 2 \cdot 3 \cdot 5$. Consequently, the greatest common divisor of 12 and 30 is $2 \cdot 3$, i.e., $\gcd(12, 30) = 6$.

Example 0.4.22. 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 as in Examples 0.4.21 and 0.4.22 (see Exercise 0.6.32). Certainly, it is possible that $\gcd(a, b) = 1$. One immediate instance of this is that both a and b are prime. Generalizing this notion, we say that positive integers a and b are **relatively prime** if and only if $\gcd(a, b) = 1$.

Example 0.4.23. Observe that 2 and 3 are relatively prime because they are distinct primes, hence they have no prime factors in common. Consequently, we have that $\gcd(2, 3) = 1$.

Example 0.4.24. We claim that 30 and 77 are relatively prime. Observe that the prime factorization of 30 is $30 = 2 \cdot 3 \cdot 5$, and the prime factorization of 77 is $77 = 7 \cdot 11$. Because they have no prime factors in common, we conclude that $\gcd(30, 77) = 1$, hence 30 and 77 are relatively prime.

0.4.3 Division Algorithm

Even as early as grade school, we learn the process of dividing an integer by a nonzero integer. Each time we divide an integer b by a nonzero integer a, we obtain an integer q and a non-negative integer r that is strictly smaller than $|a|$ such that $b = qa + r$. Explicitly, we say that b is the **dividend**; a 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 b by a nonzero integer a unambiguously results in integers q and r such that $b = qa + r$ and $0 \leq r < |a|$. We will also establish an algorithm that will allow us to efficiently find the integers q and r.

Example 0.4.25. Consider the case that $b = 11$ and $a = 2$. One can easily see that $11 = 5 \cdot 2 + 1$, hence the integers $q = 5$ and $r = 1$ satisfy the requirements that $b = qa + r$ and $0 \leq r < |a|$.

Example 0.4.26. Consider the case that $b = -17$ and $a = 6$. We find that $-17 = -3 \cdot 6 + 1$, hence the integers $q = -3$ and $r = 1$ satisfy the requirements that $b = qa + r$ and $0 \leq r < |a|$.

Example 0.4.27. Consider the case that $b = -8$ and $a = -9$. We find that $-8 = 1(-9) + 1$, hence the integers $q = 1$ and $r = 1$ satisfy the requirements that $b = qa + r$ and $0 \leq r < |a|$.

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 $R(a, b) = \{b - qa \mid q \text{ is an integer and } b \geq qa\}$. Our idea is to find the largest (in absolute value) integer q such that $b \geq qa$; then, the difference $b - qa$ 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 $R(a, b)$ in our previous examples.

Example 0.4.28. By definition, we have that $R(2, 11) = \{11 - 2q \mid q \text{ is an integer and } 11 \geq 2q\}$. Observe that $11 \geq 2q$ if and only if $q \leq 11/2$, hence the only valid values of q in $R(11, 2)$ 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 $R(2, 11) = \{1, 3, 5, 7, \dots\}$ consists of all odd positive integers.

Example 0.4.29. We have that $R(6, -17) = \{-17 - 6q \mid q \text{ is an integer and } -17 \geq 6q\}$. Observe that $-17 \geq 6q$ if and only if $q \leq -17/6$, hence the only valid values of q in $R(-17, 6)$ are $q \leq -3$. Consequently, we conclude that $R(6, -17) = \{-17 - 6q \mid q \leq -3 \text{ is an integer}\} = \{1, 7, 13, 19, \dots\}$.

Example 0.4.30. We have that $R(-9, -8) = \{-8 + 9q \mid q \text{ is an integer and } -8 \geq -9q\}$. Observe that $-8 \geq -9q$ if and only if $q \geq 8/9$, hence the only valid values of q in $R(-8, -9)$ are $q \geq 1$. Consequently, we conclude that $R(-9, -8) = \{-8 + 9q \mid q \geq 1 \text{ is an integer}\} = \{1, 10, 19, 28, \dots\}$.

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

Theorem 0.4.31 (Division Algorithm)**.** *Given any integer b and any nonzero integer a, there exist unique integers q and r such that $b = qa + r$ and $0 \leq r < |a|$.*

Proof. Consider the collection $R(a, b) = \{b - qa \mid q \text{ is an integer and } b \geq qa\}$. By definition, $R(a, b)$ consists of non-negative integers. Observe that if $b \geq 0$, then $R(a, b)$ is nonempty because we may take $q = 0$ to demonstrate that $R(a, b)$ contains b. On the other hand, if $b \leq -1$, then if $a \geq 1$, then $R(a, b)$ is yet again nonempty because we may take $q = b - 1$ to demonstrate that $R(a, b)$ contains $b - qa$ since $b \geq b - 1 \geq (b - 1)a = qa$. Last, if $b \leq -1$ and $a \leq -1$, then $R(a, b)$ is once more nonempty because we may take $q = -(b - 1)$ to demonstrate that $R(a, b)$ contains $b - qa$

since $b \geq b-1 \geq -(b-1)a = qa$. Ultimately, this shows that $R(a,b)$ is a nonempty subset of non-negative integers, hence the Well-Ordering Principle implies that there exists a smallest element $r(a,b) = b - qa$ with respect to the total order $\leq$ on the real numbers. Rearranging this identity with $r = r(a,b)$ yields that $b = qa + r$. Considering that $r \geq 0$, it suffices to see that $r < |a|$. On the contrary, suppose that $b - qa = r \geq |a|$. Observe that if $a \geq 1$, then $|a| = a$ yields that $b - qa \geq a$ and $b - (q+1)a \geq 0$. Considering that $b-(q+1)a$ is less than the least element $r(a,b) = b - qa$ of $R(a,b)$, we obtain a contradiction. Likewise, if $a \leq -1$, then $|a| = -a$ yields that $b - qa \geq -a$ and $b - (q-1)a \geq 0$. Considering that $a \leq -1$, we find that $b-(q-1)a = b - qa + a < b - qa = r(a,b)$. Once again, this contradicts the fact that $r(a,b)$ is the smallest element of $R(a,b)$. Consequently, we conclude that there exist integers q and r such that $b = qa + r$ and $0 \leq r < |a|$.

We must prove next that these integers are unique. We accomplish this by assuming that there exist integers q' and r' such that $b = q'a + r'$ and $0 \leq r' < |a|$. Considering that $b = qa + r$ by the previous paragraph, we conclude that $qa + r = q'a + r'$ so that $a(q - q') = r' - r$. 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 $|a| \leq |r' - r|$. Observe that if $r' > r$, then $|r' - r| = r' - r$ implies that $r' \geq |a| + r \geq |a|$ — a contradiction. Likewise, if $r' < r$, then $|r' - r| = r - r'$ implies that $r \geq |a| + r' \geq |a|$ — a contradiction. Either way, we conclude that $r' = r$ so that $a(q - q') = 0$. By hypothesis that a is nonzero, we conclude that $q - q' = 0$ or $q' = q$. □

We have therefore rigorously verified the non-trivial method of division that we have taken for granted since grade school! We remind the reader at this point that if $b = qa + r$ for the unique integers q and r such that $0 \leq r < |a|$, then we refer to the integer b as the **dividend**; the integer a as the **divisor**; the integer q as the **quotient** of b modulo a; and the integer r as the **remainder** of b modulo a. Crucially, the remainder of b modulo a is non-negative and strictly smaller than the absolute value of the divisor. We note that all though the Division Algorithm does not have explicit steps to compute the quotient or remainder of an integer b modulo a nonzero integer b, the proof is constructive in the sense that the unique integers q and $0 \leq r < |a|$ can be deduced from the collection $R(a,b) = \{b - qa \mid q \text{ is an integer and } b \geq qa\}$, as we have done in previous examples.

One of the most fruitful applications of the Division Algorithm is the generalization of the proof by cases technique for divisibility proofs involving any positive integer $n \geq 2$. Explicitly, if we wish to prove that a positive integer $a \geq 2$ divides an integer b, then by the Division Algorithm, we may write $b = qa + r$ for some integers q and r such that $0 \leq r < |a|$. Consequently, it suffices to check each of the $|a|$ cases that $0 \leq r \leq |a| - 1$. We have already tacitly used this kind of proof by cases: in fact, every integer is either even or odd because the remainder an integer modulo 2 is either 0 or 1. Concretely, we illustrate this more general idea for divisibility proofs involving the integer 3.

Example 0.4.32. Prove that if n is an integer, then $3 \mid (2n^2 + 1)$ if and only if $3 \nmid n$.

Proof. We will assume first that $3 \nmid n$. By the Division Algorithm, there are two cases.

1.) Observe that if $n = 3q + 1$ for some integer q, then

$$2n^2 + 1 = 2(3q+1)^2 + 1 = 2(9q^2 + 6q + 1) + 1. = 3(6q^2 + 4q + 1).$$

Considering that $6q^2 + 4q + 1$ is an integer, we conclude that $3 \mid (2n^2 + 1)$.

2.) Observe that if $n = 3q + 2$ for some integer q, then

$$2n^2 + 1 = 2(3q + 2)^2 + 1 = 2(9q^2 + 12q + 4) + 1. = 3(6q^2 + 8q + 3).$$

Considering that $6q^2 + 8q + 3$ is an integer, we conclude that $3 \mid (2n^2 + 1)$.

Conversely, we will prove the contrapositive. We may assume to this end that $3 \mid n$. By definition of divides, there exists an integer q such that $n = 3q$. Consequently, we have that

$$2n^2 + 1 = 2(3q)^2 + 1 = 18q^2 + 1 = 3(6q^2) + 1.$$

Certainly, this is not divisible by 3 because 1 is not divisible by 3, hence $3 \nmid (2n^2 + 1)$. □

Example 0.4.33. Prove that if n is an odd integer such that $3 \nmid n$, then $24 \mid (n^2 - 1)$.

Proof. We will assume that n is an odd integer. By definition of an odd integer, there exists an integer k such that $n = 2k + 1$. By the Division Algorithm, if $3 \nmid n$, then there are two cases.

1.) Observe that if $n = 3q + 1$ for some integer q, then $2k + 1 = 3q + 1$ yields that $2k = 3q$. By Example 0.4.16, we must have that $2 \mid q$ so that $q = 2\ell$ for some integer ℓ, $n = 6\ell + 1$, and

$$n^2 - 1 = (6\ell + 1)^2 - 1 = (36\ell^2 + 12\ell + 1) - 1 = 12(3\ell^2 + \ell).$$

We claim that ℓ is even. On the contrary, if ℓ were odd, then we would have that $\ell = 2m + 1$ for some integer m. Combining this identity with our previous identity that $n = 6\ell + 1$ yields that $n = 6(2m+1)+1 = 12m+2 = 2(6m+1)$ — a contradiction. Consequently, there exists an integer m such that $\ell = 2m$ and $n^2 - 1 = 12[3(2m)^2 + 2m] = 24(6m^2 + m)$.

2.) Observe that if $n = 3q + 2$ for some integer q, then $2k + 1 = 3q + 2$ yields that $2k = 3q + 1$. Consequently, we must have that q is odd; otherwise, if it were the case that $q = 2\ell$ for some integer ℓ, then $2k = 3(2\ell) + 1 = 2(3\ell) + 1$ is odd — a contradiction. We conclude that there exists an integer ℓ such that $q = 2\ell + 1$ and $n = 3q + 2 = 3(2\ell + 1) + 2 = 6\ell + 5$. Observe that

$$n^2 - 1 = (6\ell + 5)^2 - 1 = (36\ell^2 + 60\ell + 25) - 1 = 12(3\ell^2 + 5\ell + 2).$$

We claim that $3\ell^2 + 5\ell + 2$ is even. Certainly, this holds if ℓ is even because the sum of three even integers is even; on the other hand, if $\ell = 2m + 1$ for some integer m, then

$$3\ell^2 + 5\ell + 2 = 3(2m + 1)^2 + 5(2m + 1) + 2 = 3(4m^2 + 4m + 1) + 10m + 7 = 2(6m^2 + 11m + 5).$$

Either way, we conclude that $2 \mid (3\ell^2 + 5\ell + 2)$ so that $24 \mid (n^2 - 1)$, as desired. □

Before we state our next theorem, we remind the reader that if a and b are any nonzero integers and 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 c is any common divisor of a and b, then $c \mid d$.

We say that a pair of nonzero integers a and b are relatively prime if and only if $\gcd(a,b) = 1$. Our next theorem states that $\gcd(a,b)$ can be realized as an integer-linear combination of a and b.

Theorem 0.4.34 (Bézout's Identity)**.** *Given any nonzero integers a and b, 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 set $L(a,b) = \{ax + by \mid x, y \text{ are integers and } ax + by \geq 1\}$ of positive $\mathbb{Z}$-linear combinations of some nonzero integers a and b. One of the integers $a + b$, $a - b$, $-a + b$, or $-a - b$ lies in $L(a,b)$, hence $L(a,b)$ is nonempty. By the Well-Ordering Principle, there exists a smallest element $\ell(a,b) = ax + by$ with respect to the total order $\leq$. We will show that $\gcd(a,b) = \ell(a,b)$.

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

$$r_a = a - q_a\ell(a,b) = a - q_a(ax + by) = (1 - q_ax)a - (q_ay)b.$$

Observe that if r_a were nonzero, then it would lie in $L(a,b)$ and satisfy $1 \leq r_a < \ell(a,b)$, but this is impossible because $\ell(a,b)$ 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 $\ell(a,b)$ divides b. Ultimately, this proves that $\ell(a,b) \mid a$ and $\ell(a,b) \mid b$, hence $\ell(a,b)$ is a common divisor of both a and b.

Consider any other common divisor c of a and b. We must prove that $c \mid \ell(a,b)$. By assumption, there exist integers q_a and q_b such that $a = q_ac$ and $b = q_bc$, from which it follows that

$$\ell(a,b) = ax + by = (q_ac)x + (q_bc)y = (q_ax + q_by)c.$$

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

Last, 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 $av + bw$ for any integers v and w by Proposition 0.4.12. □

Corollary 0.4.35 (Uniqueness of GCD)**.** *Greatest common divisors of nonzero integers are unique.*

Proof. By the proof of Bézout's Identity, $\gcd(a,b)$ is unique for any nonzero integers a and b since it is by construction the smallest (w.r.t. the total order $\leq$) positive integer satisfying a property. □

Corollary 0.4.36 (Characterization of Relatively Prime Integers)**.** *Given any nonzero integers a and b, we have that a and b are relatively prime if and only if $ax + by = 1$ for some integers x, y.*

Even though Bézout's Identity guarantees that the existence of integers x and y such that $\gcd(a,b) = ax + by$ for any pair of nonzero integers a and b, neither the statement of this fact nor its proof provides any tools for explicitly finding these integers x and y. We conclude this section by constructing a step-by-step process for producing the integers x and y for which $\gcd(a,b) = ax + by$. Contrary to the Division Algorithm (that is not in fact an algorithm after all), we will obtain a programmable, reproducible algorithm for this procedure that can be readily coded for computing.

Example 0.4.37. 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 fact 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 0.4.38. Consider the case that $a = 110$ and $b = 24$. Observe that the unique prime factorizations of 110 and 24 are $110 = 10 \cdot 11 = 2 \cdot 5 \cdot 11$ and $24 = 2^3 \cdot 3$, respectively. By Exercise 0.6.32, it follows that $\gcd(110, 24) = 2$. 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 repeat 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. Bearing in mind 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 determine the integers x and y such that $110x + 24y = 2$ by unravelling this process in reverse. Explicitly, our last identity gives that $10 - 2 \cdot 4 = 2$; the identity before that gives 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 $14 = 1 \cdot 10 + 4$ gives 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 $24 = 1 \cdot 14 + 10$ gives 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 0.4.39 (Euclidean Algorithm)**.** Consider any nonzero integers a and b such that $a \geq b$. We may produce integers x and y such that $\gcd(a, b) = ax + by$ according to the following.

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 the remainder 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 and Proposition 0.4.13, we must have that $\gcd(a, b) \leq r_{n+1}$. Conversely, because r_{n+1} divides r_n by (4.), it must divide r_k for all integers $1 \leq k \leq n$ by steps (5.) through (7.) above. Consequently, by step (2.) above, we conclude that r_{n+1} must divide b, and by step (1.) 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 $r_{n+1} = \gcd(a, b)$.

0.4.4 Congruence Modulo n, Revisited

We will assume until further notice 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 0.4.40. *We have that $R_n = \{(a, r) \mid a = qn + r$ for some integer $q\}$ is an equivalence relation on the set $\mathbb{Z}$ of integers 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 (1.) reflexive, (2.) symmetric, and (3.) transitive.

(1.) Observe that $(a, a) \in R_n$ for any integer $a \in \mathbb{Z}$ since we have that $a = 0 \cdot n + a$.

(2.) We must demonstrate 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 $(r, a) \in R_n$.

(3.) 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 $[a] = \{r \in \mathbb{Z} \mid a = qn + r$ 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 0.1.50, we conclude that $[a] = [r_a] = \{r \in \mathbb{Z} \mid r = -qn + r_a$ for some integer $q \in \mathbb{Z}\} = \{qn + r_a \mid q \in \mathbb{Z}\}$, as desired. □

Example 0.4.41. 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 $\mathbb{O} = \{2q + 1 \mid q \in \mathbb{Z}\}$.

Example 0.4.42. Observe that R_3 is an equivalence relation on $\mathbb{Z}$ whose distinct equivalence classes consist of the integer multiples of 3, $[0] = \{3q \mid q \in \mathbb{Z}\}$, the integers that are congruent to 1 modulo 3, $[1] = \{3q + 1 \mid q \in \mathbb{Z}\}$, and the integers that are congruent to 2 modulo 3, $[2] = \{3q + 2 \mid q \in \mathbb{Z}\}$.

We will henceforth refer to the set $\mathbb{Z}_n$ of equivalence classes of $\mathbb{Z}$ modulo R_n as the equivalence classes of $\mathbb{Z}$ **modulo** n. By Proposition 0.4.40, we find that $\mathbb{Z}_n$ 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 need not possess any 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.

Theorem 0.4.43 (Arithmetic Modulo n)**.** *Consider the set $\mathbb{Z}_n$ of the integers modulo n.*

1.) *Given any integers a and b, the addition $[a]+[b]=[a+b]$ is a well-defined binary operation. Even more, this addition is associative, commutative, and $[a]+[0]=[a]=[0]+[a]$.*

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

3.) *Given any integers a and b, the multiplication $[a][b]=[ab]$ is a well-defined binary operation. Even more, this product is associative, commutative, distributive, and $[a][1]=[a]=[1][a]$.*

4.) *Given any integer a, the set $[a]$ 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 exposition preceding the proposition, if we assume that $[a_1]=[a_2]$ and $[b_1]=[b_2]$, then there exist integers q_a and q_b such that $a_1-a_2=q_an$ and $b_1-b_2=q_bn$. Consequently, we have that

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

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 $[a_1b_1]=[a_2b_2]$. By the exposition 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 $a_1-a_2=q_an$ and $b_1-b_2=q_bn$. Consequently, we have that

$$a_1b_1-a_2b_2=a_1b_1-a_1b_2+a_1b_2-a_2b_2=a_1(b_1-b_2)+b_2(a_1-a_2)=q_ba_1n+q_ab_2n=(q_ba_1+q_ab_2)n,$$

from which we conclude that $[a_1b_1]=[a_2b_2]$. 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 this product, 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, we find that $[a]$ 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 **modular arithmetic**.

Proposition 0.4.44 (Properties of Arithmetic Modulo n)**.** *Consider any nonzero integer n and any integers a, b, c, and d. Each of the following properties of congruence modulo n holds.*

1.) (**Modular Addition**) *If $b\equiv a \pmod n$ and $d\equiv c \pmod n$, then $b+d\equiv a+c \pmod n$.*

2.) (**Modular Multiplication**) *If $b\equiv a \pmod n$ and $d\equiv c \pmod n$, then $bd\equiv ac \pmod n$.*

Proof. We will assume that $b\equiv a \pmod n$ and $d\equiv c \pmod n$, i.e., $b-a=nk$ and $d-c=n\ell$ for some integers k and ℓ. Observe that $(b+d)-(a+c)=(b-a)+(d-c)=nk-n\ell=n(k-\ell)$, hence the fact that $k-\ell$ is an integer implies that n divides $(b+d)-(a+c)$ and $b+d\equiv a+c \pmod n$. Likewise, we have that $bd-ac=bd-bc+bc-ac=b(d-c)+c(b-a)=bn\ell+cnk=n(b\ell+ck)$. Considering that $b\ell+ck$ is an integer, it follows by definition that $bd\equiv ac \pmod n$, as desired. □

0.4.5 Proofs Involving Sets, Set Operations, and Functions

Combined, the calculus of logic of Section 0.2 and the basic proof techniques of Section 0.3 allow us to deduce further properties of sets and set operations. On their own, naïve set theory and formal logic are two rich and interesting areas of mathematics, but their utility in the broader patchwork of pure and applied mathematics and computer science makes them indelible tools in our toolkit.

Before we proceed to any new material, we provide first a reinterpretation of Section 0.1 in the language of Section 0.2. We will assume to this end that X and Y are some (possibly empty) sets.

- We may view the **set membership** $x \in X$ as the following statement.

  $$M(x, X)\colon \text{ We have that } x \text{ is an element of the set } X.$$

  Consequently, the negation $x \notin X$ of the set membership $x \in X$ is the following statement.

  $$\neg M(x, X)\colon \text{ We have that } x \text{ is not an element of the set } X.$$

- We may view the **subset containment** $X \subseteq Y$ as the following statement.

  $$C(X, Y)\colon \text{ For every element } x \in X, \text{ we have that } x \in Y.$$

  Considering that any universally quantified statement can be viewed as a conditional statement, we may view the subset containment $X \subseteq Y$ as the following conditional statement.

  $$C(X, Y)\colon \text{ If } x \in X, \text{ then } x \in Y.$$

  Consequently, the empty set $\varnothing$ is a subset of every set X: indeed, $C(\varnothing, X)$ is vacuously true! Observe that the negation $X \not\subseteq Y$ of the subset containment is an existence statement.

  $$\neg C(X, Y)\colon \text{ There exists an element } x \in X \text{ such that } x \notin Y.$$

- We may view the **proper subset containment** $X \subset Y$ as the following statement.

  $$C^*(X, Y)\colon \text{ We have that } X \text{ is a subset of } Y \text{ and there exists an element } y \in Y \setminus X.$$

  Consequently, the proper subset containment is a conjunctive statement.

- We may view the **set equality** $X = Y$ as the following conjunctive statement.

  $$E(X, Y)\colon \text{ We have that } X \subseteq Y \text{ and } Y \subseteq X.$$

- Elements of either the set X or the Y define the **set union** $X \cup Y$ of X and Y.

  $$X \cup Y = \{w \mid (w \in X) \vee (w \in Y) \text{ is true}\}$$

- Elements of both the set X and the set Y define the **set intersection** $X \cap Y$ of X and Y.

  $$X \cap Y = \{w \mid (w \in X) \wedge (w \in Y) \text{ is true}\}$$

- Elements of the set Y but not the set X define the **relative complement** $Y \setminus X$ of X in Y.

$$Y \setminus X = \{w \mid (w \in Y) \wedge (w \notin X) \text{ is true}\}$$

- We may view the **Cartesian product** $X \times Y$ of the sets X and Y as the collection of all ordered pairs (x, y) for which x is an element of X and y is an element of Y.

$$X \times Y = \{(x, y) \mid (x \in X) \wedge (y \in Y) \text{ is true}\}$$

By using the above dictionary between set theory and logic, we can prove many facts about sets.

Example 0.4.45. Prove that for any sets X, Y, W such that $X \subseteq W$ and $Y \subseteq W$, we have that

$$X \setminus Y = X \cap (W \setminus Y).$$

Proof. By the above definition of set equality, we must demonstrate that $X \setminus Y \subseteq X \cap (W \setminus Y)$ and $X \cap (W \setminus Y) \subseteq X \setminus Y$. By definition of $X \setminus Y$, if $x \in X \setminus Y$, then $x \in X$ and $x \notin Y$. By assumption that $X \subseteq W$, we find that $x \in W$ and $x \notin Y$ so that $x \in X$ and $x \in W \setminus Y$. We conclude that $x \in X \cap (W \setminus Y)$, from which it follows that $X \setminus Y \subseteq X \cap (W \setminus Y)$. Conversely, if $x \in X \cap (W \setminus Y)$, then $x \in X$ and $x \in W \setminus Y$. By definition of $W \setminus Y$, we have that $x \in W$ and $x \notin Y$. We conclude that $x \in X \setminus Y$ since $x \in X$ and $x \notin Y$, from which it follows that $X \cap (W \setminus Y) \subseteq X \setminus Y$. □

Example 0.4.46. Prove that for any sets X and Y, we have that $X = (X \cap Y) \cup (X \setminus Y)$.

Proof. By the above definition of set equality, we must demonstrate that $X \subseteq (X \cap Y) \cup (X \setminus Y)$ and $(X \cap Y) \cup (X \setminus Y) \subseteq X$. Given any element $x \in X$, either $x \in Y$ or $x \notin Y$ by the Law of Excluded Middle: if the former holds, then $x \in X \cap Y$; if the latter holds, then $x \in X \setminus Y$. Either way, it follows that $x \in (X \cap Y) \cup (X \setminus Y)$. Conversely, suppose that $x \in (X \cap Y) \cup (X \setminus Y)$. Each of the sets $X \cap Y$ and $X \setminus Y$ is by definition a subset of X, hence we have that $x \in X$. □

Example 0.4.47. Prove that for any sets X and Y, we have that $X \cup Y = X$ if and only if $Y \subseteq X$.

Proof. By the above definition of set equality, we must demonstrate that if $Y \subseteq X$, then $X \cup Y \subseteq X$ and $X \subseteq X \cup Y$. Observe that the latter inclusion is true by definition of the union, hence it suffices to prove that if $Y \subseteq X$, we have that $X \cup Y \subseteq X$. We will assume to this end that $Y \subseteq X$. Observe that if $w \in X \cup Y$, then by definition of the set union, we have that $w \in X$ or $w \in Y$. Either way, by hypothesis that $Y \subseteq X$, it follows that $w \in X$, hence we conclude that $X \cup Y \subseteq X$.

Conversely, we will assume that $X \cup Y = X$. Given any element $y \in Y$, we have that $y \in X \cup Y$ so that $y \in X$ by assumption that $X \cup Y = X$. We conclude that $Y \subseteq X$, as desired. □

We will assume throughout the rest of this section that X, Y, and W are some (possibly empty) sets for which the inclusions $X \subseteq W$ and $Y \subseteq W$ hold. We remind the reader that in this case, we refer to W as our **universe** (or as the **universal set**), 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 0.4.48 (Law of Excluded Middle for Sets)**.** *Consider any (possibly empty) sets $X \subseteq W$. Given any element $w \in W$, we must have that either $w \in X$ or $w \notin X$.*

Theorem 0.4.49 (Law of Non-Contradiction for Sets)**.** *Consider any (possibly empty) sets $X \subseteq W$. Given any element $w \in W$, we cannot have that both $w \in X$ and $w \notin X$.*

We omit the proofs of the aforementioned facts because they follow immediately from the Law of Excluded Middle and the Law of Non-Contradiction for the set membership statement $M(w, X)$. Even more, we have De Morgan's Laws for the relative complements of $X \cup Y$ and $X \cap Y$ in W.

Theorem 0.4.50 (De Morgan's Laws for Sets)**.** *Consider any (possibly empty) sets $X, Y \subseteq W$.*

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

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

Proof. (1.) We will first establish the inclusion $W \setminus (X \cup Y) \subseteq (W \setminus X) \cap (W \setminus Y)$. Given any element $w \in W \setminus (X \cup Y)$, we have that $w \in W$ and $w \notin X \cup Y$ by definition of the set relative complement. Consequently, we must have that $w \notin X$ and $w \notin Y$. But this implies that $w \in W \setminus X$ and $w \in W \setminus Y$ so that $w \in (W \setminus X) \cap (W \setminus Y)$. Conversely, suppose that $w \in (W \setminus X) \cap (W \setminus Y)$. By definition of the set intersection, we have that $w \in W \setminus X$ and $w \in W \setminus Y$. By definition of the relative complement, we have that $w \in W$ and $w \notin X$ and $w \notin Y$ so that $w \in W$ and $w \notin X \cup Y$.

(2.) We will first establish the inclusion $W \setminus (X \cap Y) \subseteq (W \setminus X) \cup (W \setminus Y)$. Given any element $w \in W \setminus (X \cap Y)$, we have that $w \in W$ and $w \notin X \cap Y$ by definition of the relative complement. Consequently, we must have that either $w \notin X$ or $w \notin Y$. But this implies that $w \in W \setminus X$ or $w \in W \setminus Y$ so that $w \in (W \setminus X) \cup (W \setminus Y)$. Conversely, suppose that $w \in (W \setminus X) \cup (W \setminus Y)$. By definition of the union, we have that $w \in W \setminus X$ or $w \in W \setminus Y$. Consequently, we have that $w \in W$ and either $w \notin X$ or $w \notin Y$. Either way, it follows that $w \notin X \cap Y$ so that $w \in W \setminus (X \cap Y)$. □

Often, we will simultaneously deal with $n \geq 2$ (possibly empty) sets $X_1, X_2, \ldots, X_n$ such that $X_i \subseteq W$ for each integer $1 \leq i \leq n$; in this case, it is easiest to adopt the notation of Section 0.1. We will say that each set X_i is **indexed** by an integer subscript $1 \leq i \leq n$. Consider the set union

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

Observe that $w \in \cup_{i=1}^{n} X_i$ if and only if the existence statement "$\exists i \in \{1, 2, \ldots, n\}$, $w \in X_i$" is true. Consider the set intersection

$$\bigcap_{i=1}^{n} X_i = X_1 \cap X_2 \cap \cdots \cap X_n = \{w \mid w \in X_i \text{ for every integer } 1 \leq i \leq n\}.$$

Observe that $w \in \cap_{i=1}^{n} X_i$ if and only if the universal statement "$\forall i \in \{1, 2, \ldots, n\}$, $w \in X_i$" is true. Generally, De Morgan's Laws for Sets hold for finite unions and intersections of sets as follows.

Proposition 0.4.51 (Generalized De Morgan's Laws)**.** *Consider any sets $X_1, X_2, \ldots, X_n \subseteq W$.*

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

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

Likewise, we have the following distributive laws for finite unions and intersections of sets.

Proposition 0.4.52 (Distributive Laws for Sets)**.** *Consider any sets $X_1, X_2, \ldots, X_n$ and Y.*

1.) *We have that* $Y \cap (X_1 \cup X_2 \cup \cdots \cup X_n) = (Y \cap X_1) \cup (Y \cap X_2) \cup \cdots \cup (Y \cap X_n)$.

2.) *We have that* $Y \cup (X_1 \cap X_2 \cap \cdots \cap X_n) = (Y \cup X_1) \cap (Y \cup X_2) \cap \cdots \cap (Y \cup X_n)$.

Proof. (1.) By definition of set equality, we must establish both of the set containments

$$Y \cap (X_1 \cup X_2 \cup \cdots \cup X_n) \subseteq (Y \cap X_1) \cup (Y \cap X_2) \cup \cdots \cup (Y \cap X_n) \text{ and}$$
$$Y \cap (X_1 \cup X_2 \cup \cdots \cup X_n) \supseteq (Y \cap X_1) \cup (Y \cap X_2) \cup \cdots \cup (Y \cap X_n).$$

Consider any element $x \in Y \cap (X_1 \cup X_2 \cup \cdots \cup X_n)$. By definition of set intersection, we have that $x \in Y$ and $x \in X_1 \cup X_2 \cup \cdots \cup X_n$. Likewise, by definition of set union, we have that $x \in X_i$ for some integer $1 \leq i \leq n$. Consequently, it follows that $x \in Y \cap X_i$ for some integer $1 \leq i \leq n$ so that $x \in (Y \cap X_1) \cup (Y \cap X_2) \cup \cdots \cup (Y \cap X_n)$, and the subset containment $\subseteq$ is established. Conversely, suppose that $x \in (Y \cap X_1) \cup (Y \cap X_2) \cup \cdots \cup (Y \cap X_n)$. By definition of set union, we have that $x \in Y \cap X_i$ for some integer $1 \leq i \leq n$. Consequently, it follows that $x \in Y$ and $x \in X_i$ for some integer $1 \leq i \leq n$. But this implies that $x \in X$ and $x \in X_1 \cup X_2 \cup \cdots \cup X_n$, hence $\supseteq$ holds.

(2.) We reserve the proof of the second distributive law for sets as Exercise 0.6.36. □

By appealing to our dictionary between logic and set theory, we may also prove many important properties of functions. We remind the reader that a function $f : X \to Y$ is simply a subset of the Cartesian product $X \times Y$ with the additional property that for each element $x \in X$, there exists one and only one element $f(x) = y \in Y$ such that $(x, f(x)) \in f$. Each function $f : X \to Y$ gives rise to a set $\text{range}(f) = \{f(x) \mid x \in X\}$ of all **images** of elements of X under f. Conversely, for each subset $W \subseteq Y$, we may consider the set $f^{-1}(W) = \{x \in X \mid f(x) \in W\}$ of **inverse images** of elements of W under f. We refer to the function $f : X \to Y$ as **injective** provided that $f(x) = f(y)$ implies that $x = y$ for all elements $f(x) \in \text{range}(f)$. Likewise, we refer to the function $f : X \to Y$ as **surjective** provided that $Y = \text{range}(f)$, i.e., for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. We say that a function is **bijective** if it is injective and surjective.

Proposition 0.4.53. *Consider any function $f : X \to Y$ between any two sets X and Y.*

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

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

Proof. (a.) By Exercise 0.6.38, it suffices to prove that $f^{-1}(f(V)) \subseteq V$. Consider any element $x \in f^{-1}(f(V))$. By definition of the inverse image $f^{-1}(f(V))$ of $f(V)$, we have that $f(x) \in f(V)$. By definition of the image $f(V)$ of V, it follows 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.

(b.) By Exercise 0.6.38, it suffices to prove that $W \subseteq f(f^{-1}(W))$. Consider any element $w \in W$. 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 inverse 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, and likewise, if $f(f^{-1}(W)) = W$ for any set $W \subseteq Y$, then f must be surjective (see Exercise 0.6.39).

0.5 Chapter 0 Overview

We recall that a **set** X is a collection of distinct objects called **elements** (or **members**) that often possess common properties. Each element of a set X is written as a lowercase x. If X possesses only finitely many elements $x_1, x_2, \ldots, x_n$, then we may describe the set X using the **explicit notation** $X = \{x_1, x_2, \ldots, x_n\}$. Often, it is most convenient to express a set X using **set-builder notation** $X = \{x \mid P(x)\}$ for some property $P(x)$ common to all elements $x \in X$. We assume the existence of a set $\varnothing$ that does not possess any elements; it is called the **empty set**. Every collection of sets admits certain operations that allow us to combine, compare, and take differences. Explicitly,

- 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 $X \cap Y = \{w \mid w \in X \text{ and } w \in Y\}$; and
- the **relative complement** of X with respect to Y is the set $Y \setminus X = \{w \mid w \in Y \text{ and } w \notin X\}$.

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 $Y \subset X$. By the Going-Down Property of Set Intersection or the Going-Up Property of Set Union, we have that Y is a subset of X if and only if $X \cap Y = Y$ if and only if $X \cup Y = X$. If $Y \subseteq X$ and $X \subseteq Y$, then $X = Y$; otherwise, the sets X and Y are not equal. One other way to distinguish a (finite) set X is by the number of elements X possesses, called the **cardinality** of X and denoted by $|X|$ (or $\#X$ if the bars are ambiguous).

Conveniently, we may work with large collections of sets by introducing an **index set** I. Concretely, we may denote by $\{X_i \mid i \in I\}$ the family of sets **indexed** by I. If each set X_i is a subset of some set U, we refer to U as a **universal set**. By definition, the union of the sets X_i is the set

$$\bigcup_{i \in I} X_i = \{u \mid u \in X_i \text{ for some element } i \in I\}$$

so that membership of an element $u \in U$ in this arbitrary union is characterized by $u \in \cup_{i \in I} X_i$ if and only if $u \in X_i$ for some index $i \in I$. Likewise, the arbitrary intersection of these sets is

$$\bigcap_{i \in I} X_i = \{u \mid u \in X_i \text{ for all elements } i \in I\}$$

with membership of an element $u \in U$ in the intersection characterized by $u \in \cap_{i \in I} X_i$ if and only if $u \in X_i$ for all indices $i \in I$. We say that two sets X_i and X_j are **disjoint** if $X_i \cap X_j = \varnothing$; if $X_i \cap X_j = \varnothing$ for all distinct indices $i, j \in I$, then the sets in $\{X_i \mid i \in I\}$ are **pairwise disjoint** or **mutually exclusive**. We say that $\mathcal{P} = \{X_i \mid i \in I\}$ forms a **partition** of the set U if and only if

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

(b.) $U = \cup_{i \in I} X_i$; and

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

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., $X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$. Cardinality of finite sets X and Y is multiplicative in the sense that $|X \times Y| = |X| \cdot |Y|$. We refer to any subset R of the Cartesian product $X \times Y$ as a **relation** from the set X to the set Y. We say that an element $x \in X$ is **related to** an element $y \in Y$ under R if $(x, y) \in R$, and we write that $x \, R \, y$ in this case. Every relation $R \subseteq X \times Y$ induces a relation $R^{-1} \subseteq Y \times X$ called the **inverse relation** defined by

$$R^{-1} = \{(y, x) \mid (x, y) \in R\}.$$

If X is any set, a **relation on** X is a subset R of the Cartesian product $X \times X$. Every set X admits a relation $\Delta_X = \{(x, x) \mid x \in X\}$ called the **diagonal**. 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 $(x, y) \in R$ implies that $(y, x) \in R$;
- **antisymmetric** if and only if $(x, y) \in R$ and $(y, x) \in R$ implies that $x = y$; and
- **transitive** if and only if $(x, y) \in R$ and $(y, z) \in R$ together imply that $(x, z) \in R$.

Equivalence relations are precisely those relations that are reflexive, symmetric, and transitive; **partial orders** are precisely those relations that are reflexive, antisymmetric, and transitive. 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 $x, y \in X$ are **equivalent modulo** E if and only if $(x, y) \in E$, in which case we write that $x \, E \, y$; thus, the equivalence class of an element $x \in X$ is the collection of elements $y \in X$ that are equivalent to x modulo E, i.e., the equivalence class of x is simply

$$[x] = \{y \in X \mid y \, E \, x\} = \{y \in X \mid (y, x) \in E\}.$$

Every element of a nonempty set X belongs to one and only one equivalence class of X modulo an equivalence relation E, hence the distinct equivalence classes of X modulo E partition X.

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 may 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 $f(V) = \{f(v) \mid v \in V\}$ of Y for every subset $V \subseteq X$ called the **direct image** of

V (in Y) under f. Given any subset $W \subseteq Y$, the **inverse image** of W (in X) with respect to f is $f^{-1}(W) = \{x \in X \mid f(x) \in W\}$. 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**. We say that a function $f : X \to Y$ is **bijective** provided that it is both injective and surjective.

Given any functions $f : X \to Y$ and $g : Y \to Z$, we may define a function $g \circ f : X \to Z$ called the **composite function** of f under g by declaring that $(g \circ f)(x) = g(f(x))$ for all $x \in X$; the construction of composite functions is an operation known as **function composition**. Composition of functions is **associative** so that $h \circ (g \circ f) = (h \circ g) \circ f$ whenever each of these composite functions is **well-defined**; however, function composition is neither cancellative nor commutative. Concretely, we cannot conclude that $f = g$ simply because $h \circ f = h \circ g$; we cannot conclude that $h = j$ simply because $h \circ g = j \circ g$; and it is not necessarily the case that $f \circ g = g \circ f$. Composition of functions preserves the injectivity and surjectivity of functions, so it preserves bijections, as well. Every function $f : X \to Y$ is a relation from X to Y, hence there exists an inverse relation f^{-1} from Y to X; this inverse relation f^{-1} is a function if and only if f is bijective. Crucially, the **inverse function** $f^{-1} : Y \to X$ of a bijective function $f : X \to Y$ is the unique function satisfying that $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$ for the **identity function** $\mathrm{id}_X : X \to X$ defined by $\mathrm{id}_X(x) = x$.

Generally, if $f : X \to Y$ is an injective function, then the induced function $F : X \to f(X)$ defined by $F(x) = f(x)$ is bijective. Consequently, there exists a function $F^{-1} : f(X) \to X$ defined by $F^{-1}(y) = x$ for every element $y = f(x)$. Computing the inverse function F^{-1} corresponding to the induced function F amounts to solving the equation $y = F(x)$ in terms of x; the solution has the form $F^{-1}(y) = x$, and it is precisely the function F^{-1} that is the inverse function of F.

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, "The integer 3 is an odd" and "The integer 17 is negative." We note that 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** $P \Rightarrow Q$ is the statement, "P implies Q" (or equivalently, "If P, then Q"); the implication is false if and only if P is true and Q is false. Regardless of the verity of the statement Q, if the statement P is false, then the implication $P \Rightarrow Q$ must be **vacuously** true. 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 $\neg P$ is true if and only if P is false. Given any statement P, the disjunction $P \vee \neg P$ is true by the Law of Excluded Middle, and the conjunction $P \wedge \neg P$ is false by the Law of Non-Contradiction.

We use **truth tables** to deduce the verity of a statement $S(P, Q)$ that depends on 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 $S(P, Q)$ in a third column. Considering that the statements P and Q could themselves depend upon other statements $P_1, \ldots, P_n$, a truth table grows arbitrarily large as the number of statements increases. Generally, we need $2^n + 1$ rows and $n + 1$ columns to construct the truth table of a statement $S(P_1, \ldots, P_n)$ defined for n distinct statements $P_1, \ldots, P_n$.

We say that two statements P and W are **logically equivalent** if and only if they induce the same truth table; in particular, if P and Q are equivalent statements, the truth values of P are the

same as the truth values of Q for all possible truth inputs, hence the verity of the statement P is exactly the same as the verity of the statement Q. Even more, if the truth values of P are all true, then P is a **tautology**; if the truth values for P are all false, then P is a **contradiction**.

De Morgan's Laws are two rules of inference that relate the conjunction, disjunction, and negation. Concretely, De Morgan's Laws assert the logical equivalence of the following statements.

(a.) $\neg(P \vee Q)$: It is not the case that either P or Q.

(b.) $\neg P \wedge \neg Q$: It is neither the case that P nor the case that Q.

Likewise, De Morgan's Laws for the negation of a conjunction assert the equivalence of the following.

(c.) $\neg(P \wedge Q)$: It is not the case that both P and Q.

(d.) $\neg P \vee \neg Q$: It is either not the case that P or not the case that Q.

Logical quantifiers allow us to symbolically handle statements involving quantities. We use the **universal quantifier** $\forall$ to express that an open sentence $P(x)$ is true "for all" possible values of x in its domain, and we use the **existential quantifier** $\exists$ to express that "there exists" a value of x in the domain of $P(x)$ such that $P(x)$ is true. We say that an element x_0 in the domain of the open sentence $P(x)$ is **unique** if it is the only value in the domain of $P(x)$ such that $P(x_0)$ is true. We use the **uniqueness quantifier** $\exists!$ to express the existence ($\exists$) and uniqueness ($!$) of x_0.

Often, we seek to prove conditional statements of the form $P \Rightarrow Q$. Generally, a **vacuous proof** amounts to showing that P is false. Conversely, a **trivial proof** follows by showing that Q is true. If neither P is false nor Q is true, then a **direct proof** follows by assuming that P is true and deducing that Q is true. We refer to this rule of inference as **modus ponens**. Conditional statement $P \Rightarrow Q$ is logically equivalent to its **contrapositive** $\neg Q \Rightarrow \neg P$ (see Table 15 and the subsequent Proposition 0.2.37). Consequently, a **proof by contrapositive** follows by assuming that $\neg Q$ is true and deducing that $\neg P$ is true. We refer to this rule of inference as **modus tollens**. Concretely, if $\neg P$ can be deduced from $\neg Q$, then we may construct a proof by contrapositive; on the other hand, if Q can be deduced from P, then we may construct a direct proof. Otherwise, we seek a **proof by contradiction** by assuming that P is true and Q is false and deriving a contradiction using any assumption made in the context of the proof or any definition or well-known fact. Proof by contradiction can be deduced from the Law of Excluded Middle, the Law of Non-Contradiction, and the logical equivalence of the statements $\neg(P \Rightarrow Q)$ and $P \wedge \neg Q$ (see Table 14).

Collectively, **Principle of Mathematical Induction** comprises three equivalent statements: the Principle of Ordinary Induction, the Principle of Complete Induction, and the Well-Ordering Principle. We will assume that n_0 is an integer. Consider any open sentence $P(n)$ defined for all integers $n \geq n_0$. Concretely, the Principle of Ordinary Induction asserts that the statement $P(n)$ is true for all integers $n \geq n_0$ provided that the following pair of statements is true.

(a.) We have that $P(n_0)$ is a true statement.

(b.) If $P(n)$ is a true statement for some integer $n \geq n_0$, then $P(n+1)$ is a true statement.

Likewise, the Principle of Complete Induction asserts that the statement $P(n)$ is true for all integers $n \geq n_0$ provided that the following pair of statements is true.

(c.) We have that $P(n_0)$ is a true statement.

(d.) If $P(k)$ is a true statement for each integer $n_0 \leq k \leq n$, then $P(n+1)$ is a true statement.

One crucial benefit of using complete induction as opposed to ordinary induction is that the stronger hypotheses of complete induction provide more information with which to conveniently write proofs that might otherwise prove difficult with ordinary induction (see Exercise 0.6.22). Even more, the Principle of Mathematical Induction appears in the guise of the Well-Ordering Principle of the non-negative integers — a powerful tool that 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 $b = qa + r$ and $0 \leq r \leq |b| - 1$; 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** of b modulo a. 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 c is any common divisor of a and b, then $c \mid d$.

We say that the nonzero integers a and b are **relatively prime** if and only if $\gcd(a, b) = 1$. Bézout's Identity asserts that $\gcd(a, b) = ax + by$ for some integers x and y. We may employ the Euclidean Algorithm to determine the integers x and y that are guaranteed by Bézout's Identity.

Using logical quantifiers allows us to conveniently state many properties of sets, e.g., the Law of Excluded Middle for Sets, Law of Non-Contradiction for Sets, and De Morgan's Laws for Sets.

0.6 Chapter 0 Exercises

Exercise 0.6.1. Express each of the following sets in set-builder notation.

(a.) $S = \{1, 4, 7, 10\}$

(b.) $T = \{-5, -4, -3, 3, 4, 5\}$

(c.) $U = \{-19, -18, \ldots, -4, 4, 5, \ldots, 19\}$

(d.) $V = \{2, 3, 5, 7, 11, 13, 17, 19, \ldots\}$

(e.) $W = \{\ldots, -3, -1, 1, 3, \ldots\}$

(f.) $X = \left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\right\}$

(g.) $Y = \left\{\frac{1}{9}, -\frac{1}{3}, 1, -3, 9, \ldots\right\}$

(h.) $Z = \{\ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots\}$

Exercise 0.6.2. Express each of the following sets in explicit notation.

(a.) $S = \left\{s \in \mathbb{R} \mid s^2 + \frac{4}{3}s + \frac{1}{3} = 0\right\}$

(b.) $T = \{t \in \mathbb{R} \mid \tan(t) = 0\}$

(c.) $U = \left\{u \in \mathbb{R} : \frac{d}{du}\sqrt{u^2 + 1} = 0\right\}$

(d.) $V = \{v \in \mathbb{N} \mid v^2 + 1 = 26\}$

(e.) $W = \{w \in \mathbb{Z} : w \text{ is odd and } |w| < 10\}$

(f.) $X = \{x \in \mathbb{R} \mid x^3 - 6x^2 + 11x - 6 = 0\}$

(g.) $Y = \{y \in \mathbb{R} \mid y^4 + 3 = 0\}$

(h.) $Z = \left\{ z \in \mathbb{R} : \lim_{x \to z} \dfrac{x^2}{x^4 - 2x^2 + 1} = \infty \right\}$

Exercise 0.6.3. Consider the set $U = \{1, 2, 3, 4, 5\}$ with subsets A and B such that

(a.) $|A| = |B| = 3$;

(b.) 1 lies in A but does not lie in B;

(c.) 2 lies in B but does not lie in A;

(d.) 3 lies in either A or B but not both;

(e.) 4 lies in either A or B but not both; and

(f.) 5 lies in either A or B but not both.

List all possibilities for A in curly brace notation; then, determine the corresponding sets B.

Exercise 0.6.4. Consider the following sets.

$$W = \{1, 2, 3, \ldots, 10\} \qquad \mathbb{E} = \{n \mid n \text{ is an even integer}\}$$
$$X = \{1, 3, 5, 7, 9\} \qquad \mathbb{O} = \{n \mid n \text{ is an odd integer}\}$$
$$Y = \{2, 4, 6, 8, 10\} \qquad \mathbb{Z} = \{n \mid n \text{ is an integer}\}$$

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

Exercise 0.6.5. Consider the sets $W, X, Y, \mathbb{E}, \mathbb{O}$, and $\mathbb{Z}$ defined in Exercise 0.6.4.

(a.) Compute the number of elements of $X \times Y$.

(b.) List at least three distinct elements of $\mathbb{O} \times \mathbb{E}$.

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

(d.) Express the sets $\mathbb{O}$ and $\mathbb{E}$ in set-builder notation.

(e.) Construct a bijective function $f : \mathbb{O} \to \mathbb{E}$. Conclude that $\mathbb{O}$ and $\mathbb{E}$ have "essentially the same" number of elements, i.e., there are "as many" odd integers as there are even integers.

(f.) Construct a bijective function $f : \mathbb{O} \to \mathbb{Z}$. Conclude that $\mathbb{O}$ and $\mathbb{Z}$ have "essentially the same" number of elements, i.e., there are "as many" odd integers as there are integers.

Exercise 0.6.6. Consider the set $\mathbb{Z}$ of integers.

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

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

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

Exercise 0.6.7. Consider the set W consisting of all words in the English language.

(a.) Construct a relation $R \subseteq W \times W$ that is reflexive and symmetric but not transitive.

(b.) 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.

(c.) 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 0.6.8. Consider the sets $\mathbb{Z}$ of integers and $\mathbb{Z}^*$ of nonzero integers. Prove that the relation $R \subseteq \mathbb{Z} \times \mathbb{Z}^*$ defined by $(a, b)\ R\ (c, d)$ if and only if $ad = bc$ for any pair of elements $(a, b), (c, d) \in \mathbb{Z} \times \mathbb{Z}^*$ is an equivalence relation; then, describe the equivalence classes of $\mathbb{Z} \times \mathbb{Z}^*$ modulo R.

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

Exercise 0.6.9. Given any set X, consider the collection $S_X = \{Y \mid Y \subseteq X\}$ of subsets of X. Prove that the inclusion $\subseteq$ defines a partial order P on S_X 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 an explicit counterexample.

Exercise 0.6.10. 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.)

Exercise 0.6.11. Complete the following using modular arithmetic.

(a.) Find the least positive integer x for which $5a + 4 \equiv x \pmod 6$ if $a \equiv 1 \pmod 6$.

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

(c.) Completely reduce 2022^{2023} modulo 10 provided that $2^{2023} \equiv 8 \pmod{10}$.

(d.) Completely reduce 2024^{2025} modulo 10 provided that $2^{2023} \equiv 8 \pmod{10}$.

Exercise 0.6.12. Consider any nonzero integer n and any integers a and b. Prove the following statement or provide an explicit counterexample to demonstrate that it is false.

$$\text{If } ab \equiv 0 \pmod n, \text{ then } a \equiv 0 \pmod n \text{ or } b \equiv 0 \pmod n.$$

Exercise 0.6.13. Consider any prime number p and any integers a and b. Prove the following statement or provide an explicit counterexample to demonstrate that it is false.

$$\text{If } ab \equiv 0 \pmod p, \text{ then } a \equiv 0 \pmod p \text{ or } b \equiv 0 \pmod p.$$

Exercise 0.6.14. Consider any function $f : X \to Y$.

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

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

(c.) Prove that if $f \circ f = \mathrm{id}_X$, then f is bijective.

(d.) Consider the case that $f \circ f \circ f = f$. Prove that f is bijective or provide a counterexample.

Exercise 0.6.15. Explain if each of the following is a statement. Construct the negation of each statement; identify tautologies and contradictions; and write the contrapositive of each implication.

(a.) I yam what I yam.

(b.) If you know, then you know.

(c.) Where there is a will, there is a way.

(d.) Jacob, keep your head down!

(e.) Every four years, there is a Leap Year.

(f.) Does it come in a pint?

(g.) Not all who wander are lost.

(h.) I was and I was not.

(i.) Either it is freezing or Sam wears shorts.

(j.) There exists an irrational number.

Exercise 0.6.16. Consider the following statements.

P: The sun is shining in Kansas City.
Q: Bernard rides his bike to work.

Use the symbols P and Q and logical connectives such as the biconditional $\Leftrightarrow$ conjunction $\wedge$, disjunction $\vee$, implication $\Rightarrow$ and negation $\neg$ to convert each of the following statements into symbols; then, identify all logically equivalent statements, tautologies, and contradictions.

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

(b.) Bernard 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 Bernard rides his bike to work.

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

(e.) Neither the sun is shining in Kansas City nor Bernard rides his bike to work.

(f.) Either the sun is not shining in Kansas City or Bernard does not ride his bike to work.

(g.) The sun is not shining in Kansas City, and Bernard does not ride his bike to work.

(h.) Either Bernard rides his bike to work or Bernard does not ride his bike to work.

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

(j.) Bernard rides his bike to work if and only if the sun is shining in Kansas City.

Exercise 0.6.17. Let P, Q, and R be any statements. Construct an appropriate 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 0.6.18. Use Table 11 to prove that if Bob placed in the top two in a cycling race on Saturday and he did not place second, then Bob must have placed first.

Exercise 0.6.19. Construct a proof by contradiction to demonstrate that if Bob placed in the top two in a cycling race on Saturday and 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.

Exercise 0.6.20. Given any integer $n \geq 0$, prove that $\binom{2n}{n} > 2^n$ using induction.

Exercise 0.6.21. Consider any finite set X with power set $P(X)$.

(a.) Prove that $|P(X)| = 2^{|X|}$ using induction.

(b.) Consider the collection 2^X of all functions $f : X \to X$. Construct an explicit bijection between $P(X)$ and 2^X. Conclude from part (a.) and Proposition 0.1.86 that $|2^X| = 2^{|X|}$.

Exercise 0.6.22. Consider the sequence of **Fibonacci numbers** F_n defined recursively for all non-negative integers 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.

(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 0.6.23. Complete the following two steps to prove that the Principle of Ordinary Induction and the Principle of Complete Induction are materially equivalent to one another.

1.) Given any statement $P(n)$ defined for 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 $P(n)$ defined for 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 0.6.24. Complete the following three steps to prove that the Well-Ordering Principle and the Principle of Ordinary Induction are materially equivalent to one another.

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 0.6.23 in tandem with the proof of the Well-Ordering Principle to conclude conversely that the Principle of Ordinary Induction implies the Well-Ordering Principle.

Exercise 0.6.25. Prove that there is no positive integer less than 1.

Exercise 0.6.26. 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 0.6.27 (Euclid's Lemma)**.** Prove that if $a, b, c \in \mathbb{Z}$, $\gcd(a, b) = 1$, and $a \mid bc$, then $a \mid c$.

Exercise 0.6.28. Prove that there are no positive integers a, b, and c such that $a^2 + b^2 = 3c^2$.

Exercise 0.6.29 (Fundamental Theorem of Arithmetic)**.** Given any positive integer a, prove that

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

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

(**Hint:** Consider the collection N of positive integers that **do not** admit 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 $\leq$. Consider the possible factors of the positive integer n to see that N is empty, hence the existence is established. On the matter of uniqueness, proceed by induction on k.)

Exercise 0.6.30. Prove that every nonzero integer a can be written as $a = \pm p_1^{e_1} \ldots p_n^{e_n}$ for some distinct prime numbers $p_1, \ldots, p_n$ and unique non-negative integers $e_1, \ldots, e_n$ such that $p_i^{e_i+1} \nmid a$.

Given any nonzero 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 $\ell = \operatorname{lcm}(a, b)$ such that

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

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

Exercise 0.6.31. Prove that the least common multiple $\operatorname{lcm}(a, b)$ is unique up to sign.

By the Fundamental Theorem of Arithmetic, for any positive integers a and b, there exist prime numbers $p_1, \ldots, p_k$ and unique non-negative integers $e_1, \ldots, e_k, f_1, \ldots, f_k$ 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 0.6.32. Prove that $\gcd(a, b) = p_1^{\min\{e_1, f_1\}} \cdots p_k^{\min\{e_k, f_k\}}$.

Exercise 0.6.33. Prove that $\operatorname{lcm}(a, b) = p_1^{\max\{e_1, f_1\}} \cdots p_k^{\max\{e_k, f_k\}}$.

Exercise 0.6.34. Conclude from Exercises 0.6.32 and 0.6.33 that $ab = \gcd(a, b) \operatorname{lcm}(a, b)$.

Exercise 0.6.35. Consider any sets W, X, and Y such that $X \subseteq W$ and $Y \subseteq W$.

(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 0.6.36. Prove the second of the Distributive Laws for Sets.

Exercise 0.6.37. Consider any function $f : X \to Y$ from a set X to a set Y.

(a.) Prove that $f(U \cup V) = f(U) \cup f(V)$ for any sets $U, V \subseteq X$.

(b.) Prove that $f(\cup_{i \in I} V_i) = \cup_{i \in I} f(V_i)$ for any index set I and any sets $V_i \subseteq X$.

(c.) Prove that $f(U \cap V) \subseteq f(U) \cap f(V)$ for any sets $U, V \subseteq X$.

(d.) Prove that $f(\cap_{i \in I} V_i) \subseteq \cap_{i \in I} f(V_i)$ for any index set I and any sets $V_i \subseteq X$.

(e.) Construct an explicit counterexample to the superset containment $f(U \cap V) \supseteq f(U) \cap f(V)$.

(f.) Prove that $f^{-1}(V \cup W) = f^{-1}(V) \cup f^{-1}(W)$ for any sets $V, W \subseteq Y$.

(g.) Prove that $f^{-1}(\cup_{i \in I} W_i) = \cup_{i \in I} f^{-1}(W_i)$ for any index set I and any sets $W_i \subseteq Y$.

(h.) Prove that $f^{-1}(V \cap W) = f^{-1}(V) \cap f^{-1}(W)$ for any sets $V, W \subseteq Y$.

(i.) Prove that $f^{-1}(\cap_{i \in I} W_i) = \cap_{i \in I} f^{-1}(W_i)$ for any index set I and any sets $W_i \subseteq Y$.

Exercise 0.6.38. Consider any function $f : X \to Y$ from a set X to a set 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 0.4.53, observe that $f : X \to Y$ 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 0.4.53, observe that $f : X \to Y$ cannot be surjective.)

Exercise 0.6.39. Consider any function $f : X \to Y$ from a set X to a set 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 $f(f^{-1}(W))$.)

Exercise 0.6.40. Given any set X, consider the **diagonal function** $\delta_X : X \to X \times X$ defined by $\delta_X(x) = (x, x)$ and the **diagonal relation** $\Delta_X = \{(x, x) \mid x \in X\}$ on X. Prove that $\Delta_X = \delta_X(X)$.

Exercise 0.6.41. Given any prime number p, 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 $p \nmid a$.

Chapter 1

Essential Topics in Group Theory

Group theory is the study of algebraic structures equipped with associative binary operations that admit distinguished elements called the multiplicative identity and multiplicative inverses. Common examples of groups include cyclic groups, lattice groups, Lie groups, symmetry groups, topological groups, and vector spaces. Groups may be simple to describe and possess uncomplicated arithmetic, but the structure of certain groups is surprisingly complex: indeed, group theory remains an active branch of mathematics. One of the most significant results in group theory is the discovery of the so-called solvable groups by French mathematician Évariste Galois. Crucially, the theory of solvable groups implies that there is no analog to the quadratic formula for polynomials of degree exceeding four with real coefficients. Group theory is useful in coding theory, counting, number theory, and symmetries and in various applications to physical sciences, such as biology, chemistry, and physics.

1.1 Groups: Basic Definitions and Examples

We will assume throughout this chapter that G is a nonempty set. Back in Section 0.1.1, 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 $(G, *)$.

(a.) 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.

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

(c.) 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 1.1.1. Consider the set $\mathbb{Z}$ of integers. Observe that (a.) addition of integers is associative; (b.) the integer 0 satisfies that $0 + n = n = n + 0$ for all integers n; and (c.) 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 1.1.2. Consider the set $\mathbb{Z}_n$ of equivalence classes of integers modulo n. By Proposition 0.4.40, the distinct elements of $\mathbb{Z}_n$ are given by $[r] = \{qn + r \mid q \in \mathbb{Z}\}$ for each integer $0 \leq r \leq n-1$, hence $\mathbb{Z}_n$ is nonempty. Using modular arithmetic, we may define an associative binary operation $+_n$ on $\mathbb{Z}_n$ by setting $[r_1] +_n [r_2] = [r_1 + r_2]$. 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 $[r_1 + r_2]$ as $[x]$,

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 $[0]$; and if $1 \leq r \leq n-1$, then the inverse of $[r]$ is simply $[n-r]$. 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 1.1.3. Consider any regular 3-gon. We denote by ρ_k rotation of the regular 3-gon through the angle $-120k$ degrees for each integer $1 \leq k \leq 3$. We denote by ϕ_k 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 all symmetry-preserving rotations and reflections of a regular 3-gon. By Exercise 1.12.58, for any pair of elements $x, y \in D_3$, composition $y \circ x$ is an associative binary operation on D_3. Even more, we have that ρ_3 satisfies that $x \circ \rho_3 = x = \rho_3 \circ x$ 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 find that $(D_3, \circ)$ is a group under composition: 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 1.1.4. Observe that the group $(\mathbb{Z}, +)$ is abelian because addition of integers is commutative. Likewise, for any elements $[r_1]$ and $[r_2]$ of $\mathbb{Z}_n$, we have that

$$[r_1] +_n [r_2] = [r_1 + r_2] = [r_2 + r_1] = [r_2] +_n [r_1].$$

Consequently, the group $(\mathbb{Z}_n, +_n)$ is also abelian. By Exercise 1.12.58, on the other hand, the group D_3 of Example 1.1.3 is not abelian because we have that $\rho_1 \phi_1 = \phi_3 \neq \phi_2 = \phi_1 \rho_1$.

Example 1.1.5. Consider the set $\mathbb{R}$ of real numbers. Given any integer $n \geq 1$, we denote by $\mathbb{R}^{n \times n}$ the collection of all real $n \times n$ 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, and the inverse of a real $n \times n$ matrix A is the real $n \times n$ 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.

Example 1.1.6. Consider the subset $\mathrm{GL}(n, \mathbb{R})$ of $\mathbb{R}^{n \times n}$ consisting of invertible real $n \times n$ matrices. Under matrix multiplication, $\mathrm{GL}(n, \mathbb{R})$ forms a group: the multiplicative identity of $\mathrm{GL}(n, \mathbb{R})$ is the $n \times n$ identity matrix, and the multiplicative inverse of an invertible $n \times n$ matrix A is A^{-1}. Considering that matrix multiplication is not commutative, the group $(\mathrm{GL}(n, \mathbb{R}), \cdot)$ 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 whose elements define 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 $(\mathrm{GL}(n, \mathbb{R}), \cdot)$ of invertible real $n \times n$ matrices have infinitely many elements, hence they each possess infinite order.

Remark 1.1.7. Unfortunately, even if a nonempty set G admits some associative binary operation $* : G \times G \to G$, it is not immediately true that $(G, *)$ is a group. Explicitly, multiplication of integers is an associative binary operation on the integers; 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 yet, it is not true 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 $n = \pm 1$.

1.2 Groups: Basic Properties and Subgroups

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

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

GP(b.) 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 in this section 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 1.2.1 (Uniqueness of Multiplicative Identity and Multiplicative Inverses of a Group)**.** *Given any group $(G, *)$, the element e_G of first group property above is unique. Even more, for each element $g \in G$, the element $g^{-1} \in G$ of the second group property 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 GP(a.) as the **identity element** of G; the element g^{-1} of GP(b.) is the **inverse** of g. Our next result simplifies the task of finding inverses in a group.

Corollary 1.2.2 (Equality of Left- and Right-Inverses in a Group)**.** *Consider any element g of any group $(G, *)$. Given any element $h \in G$ such that $g * h = e_G$, we have that $h * g = e_G$ and $h = g^{-1}$.*

Proof. By Proposition 1.2.1, 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. $\square$

Often, we will omit the notation $*$ in G and simply use concatenation, e.g., we will write $g_1 g_2$ instead of $g_1 * g_2$. 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 (b.) above, $g_1 g_2$ admits a multiplicative inverse.

Proposition 1.2.3. *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 1.2.2, it suffices to verify that $(g_1 g_2)(g_2^{-1} g_1^{-1}) = e_G$ and $g_1^{-1} g_1 = e_G$. $\square$

Existence of multiplicative inverses implies that groups enjoy the **cancellation property**, i.e., if $g_1 g_2 = g_1 g_3$ for any elements $g_1, g_2, g_3 \in G$ and G is a group, 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" of an identity instead "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 1.2.4 (Group Exponent Laws)**.** *Consider any group G and any integers m and n.*

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 these Group Exponent Laws as Exercise 1.12.12. We point out here the convention that additive notation is used for abelian groups; in this case, this result is clear since

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

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, there may be other proper subgroups.

Example 1.2.5. Consider the abelian group $(\mathbb{Z}, +)$ of integers under addition. Given any integer n, we may define 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 1.2.6. Consider the dihedral group $(D_3, \circ)$ of Example 1.1.3 constructed from the set

$$D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$$

of all symmetry-preserving rotations and reflections of a regular 3-gon with respect to composition. Observe that $\rho_j \circ \rho_i = \rho_{i+j \pmod 3}$, hence $\langle \rho_1 \rangle = \{\rho_1, \rho_2, \rho_3\}$ is a subgroup of D_3: indeed, composition is an associative binary operation on $\langle \rho_1 \rangle$ and every element of $\langle \rho_1 \rangle$ has a multiplicative inverse in $\langle \rho_1 \rangle$. Even more, ρ_3 is the multiplicative identity of D_3, so it is the multiplicative identity of $\langle \rho_1 \rangle$.

Example 1.2.7. Consider the general linear group $\mathrm{GL}(n, \mathbb{R})$ 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

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

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

Remark 1.2.8. We cannot understate the importance of context when discussing the structure of groups and subgroups. Remark 1.1.7 demonstrates that 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 larger group in which it is contained. Concretely, the group $(\mathbb{R}^{n \times n}, +)$ of real $n \times n$ matrices under matrix addition contains $\mathrm{GL}(n, \mathbb{R})$ as a subset; however, $\mathrm{GL}(n, \mathbb{R})$ 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.

We will occasionally use the standard notation $H \leq G$ to denote that H is a subgroup of G, i.e., H is a nonempty subset of G that is a group with respect to the group operation of G. 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 attendant operation of the group.

Proposition 1.2.9 (Subgroup Test)**.** *Given any group $(G, *)$, consider any subset $H \subseteq G$. We have that $(H, *)$ is a subgroup of G if and only if the following three conditions hold.*

(a.) *H contains the identity element e_G of G.*

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

(c.) *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 (b.) holds because H is a group, hence it suffices to check that conditions (a.) and (c.) 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 in G 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 1.2.2 yields that $h' = h^{-1}$. We conclude that $h^{-1} \in H$. □

Corollary 1.2.10 (Two-Step Subgroup Test)**.** *Given any group $(G, *)$, consider any nonempty set $H \subseteq G$. We have that $(H, *)$ is a subgroup of G if and only if the following two conditions hold.*

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

(b.) *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 $h \in H$, hence H is nonempty. □

Corollary 1.2.11 (One-Step Subgroup Test)**.** *Given any group $(G, *)$, consider any 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 $h_1, h_2 \in H$.*

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.

(a.) H contains the identity element e_G of G.

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

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

We verify condition (a.) by noting that $e_G = h_1 h_1^{-1}$ is in H for any element $h_1 \in H$. Consequently, condition (c.) follows because $h^{-1} = e_G h^{-1}$ for all elements $h \in H$ and $e_G \in H$. Last, condition (b.) holds by using Proposition 1.2.3 and condition (c.) to see that $h_1 * h_2 = h_1 * (h_2^{-1})^{-1} \in H$. □

Caution: the Two-Step Subgroup Test and One-Step Subgroup Test require that we begin with a nonempty subset H of a group $(G, *)$. Consequently, there is an implicit requirement to check that H is nonempty. Often, it can be readily deduced from the defining properties of H that the identity element of G lies in H; however, it is possible to prove that H is nonempty in other ways.

Example 1.2.12. Consider the collection $\mathbb{R}^{\mathbb{R}}$ of real univariate functions $f : \mathbb{R} \to \mathbb{R}$ with domain $\mathbb{R}$. One can readily verify that $(\mathbb{R}^{\mathbb{R}}, +)$ is an abelian group with respect to the usual function addition defined by $(f+g)(x) = f(x)+g(x)$ for all real numbers x (see Exercise 1.12.5). We will demonstrate that the set $\mathcal{C}^1(\mathbb{R}) \subseteq \mathbb{R}^{\mathbb{R}}$ of functions $f : \mathbb{R} \to \mathbb{R}$ whose first derivative f' is continuous everywhere constitutes a subgroup of $(\mathbb{R}^{\mathbb{R}}, +)$ by the One-Step Subgroup Test. Crucially, the identity function $\mathrm{id}_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ defined by $\mathrm{id}_{\mathbb{R}}(x) = x$ satisfies that $\mathrm{id}'_{\mathbb{R}}(x) = 1$, hence $\mathcal{C}^1(\mathbb{R})$ is nonempty. Consequently, if $f, g \in \mathcal{C}^1(\mathbb{R})$, then $(f - g)'(x) = f'(x) - g'(x)$ is continuous for all real numbers x.

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

Example 1.2.13. We demonstrate in this example that it is possible to distinguish groups of the same order according to their subgroups, hence the structure of the subgroups of a group provide more refined information than the order of a group. Consider the group $(\mathbb{Z}_4, +_4)$ and the set $\mathbb{Z}_2 \times \mathbb{Z}_2$ with respect to componentwise addition modulo 2. One can readily verify that $\mathbb{Z}_2 \times \mathbb{Z}_2$ is an abelian group under componentwise modular arithmetic, hence we cannot distinguish between $(\mathbb{Z}_4, +_4)$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ according to commutativity of their elements. Even more, both of these groups have order four; however, we will demonstrate that these groups are distinct by showing that $(\mathbb{Z}_4, +_4)$ admits only one non-trivial proper subgroup while $\mathbb{Z}_2 \times \mathbb{Z}_2$ admits three non-trivial proper subgroups. By an abuse of notation, 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

Concretely, the Cayley table of a group $(G, *)$ with n elements is the $(n+1) \times (n+1)$ array whose $(i, 1)$ and $(1, i)$ entries are the elements $g_i \in G$ for each integer $2 \leq i \leq n+1$ and whose (i, j)th entry is $g_i * g_j$ for each pair of integers $2 \leq i, j \leq n+1$. Considering that any subgroup H of $(\mathbb{Z}_4, +_4)$ must contain the identity element 0 and the inverse of any element in H, the only non-trivial subgroup of $(\mathbb{Z}_4, +_4)$ is $H = \{0, 2\}$. On the other hand, observe that $\mathbb{Z}_2 \times \mathbb{Z}_2$ 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 identity element $(0, 0)$ in the above table, we obtain three non-trivial subgroups: namely, they are $\{(0,0), (1,0)\}$, $\{(0,0), (0,1)\}$ and $\{(0,0), (1,1)\}$. We conclude that the order-four abelian groups $(\mathbb{Z}_4, +_4)$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ are distinct; the latter is the **Klein four-group**.

Consequently, it is possible to distinguish between two groups with the same number of elements by demonstrating that the two groups do not admit the same number of (proper) subgroups.

1.3 Cyclic Groups

We continue our exploration into the structure of groups by turning our attention to 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 1.3.1. *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** *of G with* **generator** *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 $g^m(g^n)^{-1} = g^m g^{-n} = g^{m-n}$ according to the Group Exponent Laws so that $g^m(g^n)^{-1} \in \langle g \rangle$. We conclude by the One-Step Subgroup Test that $\langle g \rangle$ is a subgroup of G. □

Example 1.3.2. 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 1.3.3. Observe that for any integer n, the cyclic subgroup $\langle n \rangle$ of $(\mathbb{Z}, +)$ generated by n is given by $\langle n \rangle = \{\ldots, -2n, -n, 0, n, 2n, \ldots\} = \{nk \mid k \in \mathbb{Z}\} = n\mathbb{Z}$, i.e., the integer multiples of n.

Example 1.3.4. Given any positive integer n, we may consider the following subset of $\mathbb{Z}_n$.

$$\mathbb{Z}_n^\times = \{[a] \in \mathbb{Z}_n \mid \gcd(a, n) = 1\}$$

Using the usual Arithmetic Modulo n, we may define a multiplicative group operation on $\mathbb{Z}_n^\times$ via the identification $[a][b] = [ab]$ for each pair of equivalences classes $[a], [b] \in \mathbb{Z}_n^\times$. Crucially, we must verify that this multiplication is well-defined, i.e., we must prove that if $[a], [b] \in \mathbb{Z}_n^\times$, then $[ab] \in \mathbb{Z}_n^\times$. By definition of the set $\mathbb{Z}_n^\times$, we have that $[a], [b] \in \mathbb{Z}_n^\times$ if and only if $\gcd(a, n) = 1$ and $\gcd(b, n) = 1$. Consequently, Bézout's Identity yields integers u, v, x, and y with $ax + ny = 1$ and $bu + nv = 1$. By multiplying these identities, we obtain $1 = 1^2 = (ax + ny)(bu + nv) = abux + n(avx + buy + nvy)$ so that $\gcd(ab, n) = 1$ by Bézout's Identity. We conclude that modular multiplication $[a][b] = [ab]$ is a well-defined associative binary operation on $\mathbb{Z}_n^\times$ for which it holds that $[1][a] = [a] = [a][1]$ for each equivalence class $[a] \in \mathbb{Z}_n^\times$. Last, the identity $ax + ny = 1$ implies that $ax \equiv 1 \pmod{n}$, hence we conclude that $[a][x] = [ax] = [1] = [xa] = [x][a]$ so that $[x] = [a]^{-1}$. We refer to $\mathbb{Z}_n^\times$ as the **multiplicative group of units modulo** n. By Corollary 3.1.14, it follows that $\mathbb{Z}_p^\times$ is a cyclic group for each prime number p! Concretely, the reader should verify by inspection that $\mathbb{Z}_2^\times = \{1\} = \langle 1 \rangle$, $\mathbb{Z}_3 = \{1, 2\} = \langle 2 \rangle$, and $\mathbb{Z}_5 = \{1, 2, 3, 4\} = \langle 2 \rangle = \langle 3 \rangle$; we leave the details as part of Exercise 1.12.28. Conversely, it is possible that $\mathbb{Z}_n^\times$ is cyclic for composite n, hence the primality of n is sufficient but not necessary for the cyclicity of $\mathbb{Z}_n^\times$. We refer the reader to Exercise 1.12.29 for further practice.

Remark 1.3.5. 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 containing g, hence we may view $\langle g \rangle$ as the intersection of all subgroups of G containing g (see Exercise 1.12.34).

We will say that G is a **cyclic group** if G admits an element $g \in G$ such that $G = \langle g \rangle$; in this case, we may also specify that the group G is **generated** by the element g. By definition, the order of the cyclic subgroup $\langle g \rangle$ is the (possibly infinite) number of distinct elements of $\langle g \rangle$. Particularly, if $\langle g \rangle$ is finite, then we may define the **order** of g as the smallest positive integer $r = \text{ord}(g)$ such that $g^r = e_G$. Consequently, the distinct elements of $\langle g \rangle$ are $g^0, g^1, \ldots, g^{r-1}$ so that $\text{ord}(g) = \#\langle g \rangle$.

Example 1.3.6. 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 $(\mathbb{Z}, +) = \langle 1 \rangle = \langle -1 \rangle$ is a cyclic group. We note that the cyclic subgroups $n\mathbb{Z}$ are proper for all integers n such that $|n| \geq 2$.

Example 1.3.7. 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 $\text{ord}(5) = 2$. On the other hand, we have that $\langle 1 \rangle = \{0, 1, 2, \ldots, 9\} = \langle 9 \rangle$, hence both 1 and 9 generate $(\mathbb{Z}_{10}, +_{10})$.

Example 1.3.8. Consider the set $G = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\}$ of all upper-triangular 2×2 matrices with integer entries and 1 along the main diagonal. We claim that G is a group with respect to the usual matrix multiplication. Concretely, observe that for any integers m and n, we have that

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & m+n \\ 0 & 1 \end{pmatrix}$$

so that G is closed under matrix multiplication. Conventionally, any matrix raised to the power of zero is the identity matrix, hence the identity element of G is the identity matrix. Last, according to the above calculation, the inverse of any element $\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$ of G is simply $\left(\begin{smallmatrix} 1 & -n \\ 0 & 1 \end{smallmatrix}\right)$.

We claim that the nonempty subset $H = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}_{\geq 0}\}$ of G is cyclic. By the above calculation, if $n \geq 0$, then we may view $n = n \cdot 1$ as the n-fold sum of copies of 1, i.e., $H = \langle \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \rangle$.

Example 1.3.9. Consider the dihedral group $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of order six. By Example 1.3.2 and Exercise 1.12.58, every subgroup of D_3 is cyclic, but D_3 is not itself a cyclic group.

Caution: Example 1.3.9 exhibits a non-cyclic group whose proper subgroups are all cyclic, hence it is not possible to determine the cyclic property of a group based on its subgroups.

Even though the cyclic nature of a group cannot be deduced according to the cyclic property of its proper subgroups, our next propositions illustrate several important properties of cyclic groups.

Proposition 1.3.10 (Cyclic Groups Are Abelian)**.** *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 1.3.11 (Non-Abelian Groups Are Non-Cyclic)**.** *Every non-abelian group is non-cyclic.*

Proof. We note that this is the contrapositive of Proposition 1.3.10. □

Theorem 1.3.12 (Structure Theorem for Cyclic Groups)**.** *Each subgroup of a cyclic group is cyclic.*

Proof. We will assume that G is a cyclic group that is generated by some element $g \in G$. Concretely, suppose that $G = \{g^n \mid n \in \mathbb{Z}\}$. Consider any subgroup $H \subseteq G$. Certainly, if $H = \{e_G\}$, then H is 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 guarantees that S admits a smallest element s with respect to $\leq$. We will prove in the next paragraph that $H = \langle g^s \rangle$ so that H is cyclic.

By assumption that G is cyclic, if $k \in H$, then 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 $m = qs$ so that every element of H can be written as $g^{qs} = (g^s)^q$ for some unique integer q. □

Corollary 1.3.13. *Every subgroup of $(\mathbb{Z}, +)$ is of the form $n\mathbb{Z}$ for some non-negative integer n.*

Corollary 1.3.14. *Every subgroup of $(\mathbb{Z}_n, +_n)$ is of the form $k\mathbb{Z}_n$ for some integer $0 \leq k \leq n - 1$.*

By the paragraph preceding Example 1.3.6, the order of an element g of a group G is the smallest positive integer $r = \text{ord}(g)$ such that $g^r = e_G$. Our next two results demonstrate that the order of any generator of a cyclic group determines the order of all other elements of the group.

Lemma 1.3.15. *Consider any cyclic group G. If $G = \langle g \rangle$, then $g^n = e_G$ if and only if $\text{ord}(g) \mid n$.*

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

Corollary 1.3.16. *Given any finite-order element g of any group G, if $g^n = e_G$, then $\text{ord}(g) \mid n$.*

Proof. Each element $g \in G$ generates a cyclic group $\langle g \rangle = \{g^k \mid k \in \mathbb{Z}\}$ with $\text{ord}(g)$ elements; thus, if $\text{ord}(g)$ is finite and $g^n = e_G$, then $\text{ord}(g) \mid n$ by applying Lemma 1.3.15 with $H = \langle g \rangle$. □

Proposition 1.3.17 (Order of Powers of a Cyclic Generator)**.** *Consider any cyclic group G that is generated by an element $g \in G$. Given any integer n, the order of g^n is $\text{ord}(g)/\gcd(n, \text{ord}(g))$.*

Proof. We denote $\text{ord}(g) = d$. 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 1.3.16, 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 Euclid's Lemma 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 the following corollary that allows us to determine every generator of a cyclic group using only one known generator and the order of the group.

Corollary 1.3.18 (Generators of a Cyclic Group)**.** *Consider any cyclic group G generated by $g \in G$. Every group generator of G is of the form g^n for some integer n such that $\gcd(n, \text{ord}(g)) = 1$.*

Proof. By Proposition 1.3.17, for any integer n such that $\gcd(n, \text{ord}(g)) = 1$, we have that $\text{ord}(g^n)$ is $\text{ord}(g)$; thus, we conclude that g^n generates G. We leave the converse as Exercise 1.12.41. □

1.4 Complex Numbers as a Group Under Multiplication

Complex numbers arise naturally as solutions to some polynomials with real coefficients. Explicitly, if x is a real number, then $x^2 \geq 0$, hence the quadratic equation $x^2 + 1 = 0$ in the variable x does not admit any real solutions since we have that $x^2 + 1 \geq 1$ for any real number x. Consequently, we may assume that there exists some solution i of the quadratic equation $x^2 + 1 = 0$ so that $i^2 + 1 = 0$. Carrying out the algebra as usual, we find that $i^2 = -1$ so that $i = \sqrt{-1}$ taking the positive square root. Crucially, our assumption that such a root i of the real univariate polynomial $x^2 + 1$ exists and obeys the usual arithmetic of real numbers holds by the Fundamental Theorem of Field Theory.

We define the **complex numbers** as the set of all real linear combinations of 1 and i

$$\mathbb{C} = \mathbb{R}\langle 1, i\rangle = \{a + bi \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}.$$

Consequently, we may view $i = 0 + 1i$ itself as a complex number. We refer to the real number a of the complex number $a + bi$ as the **real part** of $a + bi$; 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 1.4.1. *Let $\mathbb{C}$ denote the complex numbers. Let $\mathbb{C}^* = \mathbb{C} \setminus \{0 + 0i\}$.*

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

2.) *We have that $(\mathbb{C}^*, \cdot)$ 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 1.4.2. *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 $\left|\frac{z}{c}\right| = \frac{|z|}{|c|}$ 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 $r\,\mathrm{cis}\,\theta$.

Example 1.4.3. 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 $\sqrt{2} - i\sqrt{2} = 2\,\mathrm{cis}(7\pi/4)$.

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

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 1.12.49 for the reader.

Proposition 1.4.4. *We have that* $(r_1\,\mathrm{cis}\,\theta_1)(r_2\,\mathrm{cis}\,\theta_2) = r_1 r_2\,\mathrm{cis}(\theta_1 + \theta_2)$.

Corollary 1.4.5 (De Moivre's Theorem)**.** *We have that* $(r\,\mathrm{cis}\,\theta)^n = r^n\,\mathrm{cis}(n\theta)$ *for all integers* $n \geq 0$.

Proof. By the Principle of Ordinary Induction on $n \geq 0$, this follows from Proposition 1.4.4. □

Corollary 1.4.6. *We have that* $|z_1 z_2| = |z_1| \cdot |z_2|$ *for all complex numbers* z_1 *and* z_2.

Corollary 1.4.7. *We have that* $|z^n| = |z|^n$ *for all complex numbers* z *and all integers* n.

Example 1.4.8. One of the benefits of De Moivre'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 $\sqrt{2} - i\sqrt{2} = 2\,\mathrm{cis}(7\pi/4)$ by Example 1.4.3, and De Moivre's Formula gives $(\sqrt{2} - i\sqrt{2})^7 = 2^7\,\mathrm{cis}(49\pi/4) = 128\,\mathrm{cis}(\pi/4) = 64(\sqrt{2} + i\sqrt{2})$.

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 1.4.6, 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 1.4.2, 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 1.4.9. *Let* $(\mathbb{C}^*, \cdot)$ *denote the multiplicative group of complex numbers. We have that* $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ *is a subgroup of* $(\mathbb{C}^*, \cdot)$ *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 1.4.10 (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 1.4.11. *If n is a positive integer, then the nth 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 $\mathbb{T}$.*

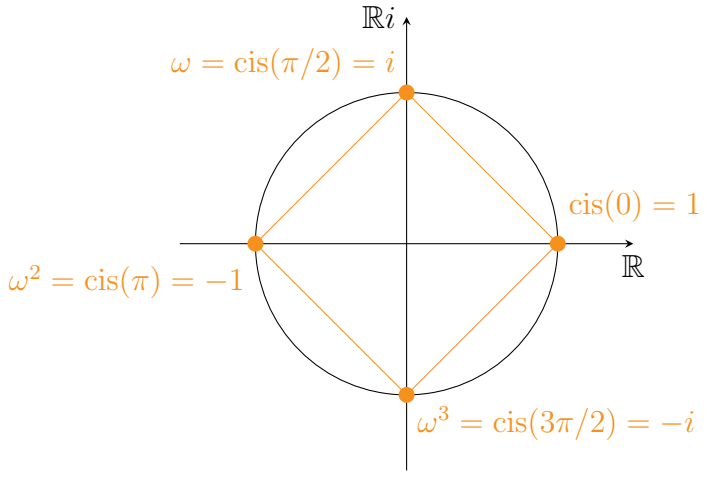
Proof. By the Fundamental Theorem of Algebra, it suffices to prove that $\operatorname{cis}(2k\pi/n)^n = 1$ for each integer $0 \leq k \leq n-1$ because for any pair of integers $0 \leq i < j \leq n-1$, we have that $\operatorname{cis}(2i\pi/n)$ and $\operatorname{cis}(2j\pi/n)$ are distinct. Observe that $\operatorname{cis}(2k\pi/n)^n = \operatorname{cis}(2kn\pi/n) = \operatorname{cis}(2k\pi) = 1$ holds by De Moivre's Theorem. Last, the nth roots of unity form a cyclic subgroup of the circle group $\mathbb{T}$ once again by De Moivre'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 1.4.11 and 1.3.18, we obtain the following.

Corollary 1.4.12. *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 1.4.13. Below are the fourth roots of unity on the unit circle.



1.5 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 $\mathrm{id}_X : X \to X$ defined by $\mathrm{id}_X(x) = x$ 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 $f \circ f^{-1} = \mathrm{id}_X = f^{-1} \circ f$: 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 $\{1, 2, \ldots, |X|\}$ that maps an element from X uniquely to some element of $\{1, 2, \ldots, |X|\}$. 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 $[n] = \{1, 2, \ldots, n\}$ 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 1.5.1. *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 $\sigma(1), \sigma(2), \ldots, \sigma(n)$. 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 $\sigma(1), \sigma(2), \ldots, \sigma(i-1)$ 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 0.6.38, every element σ of $\mathfrak{S}_n$ is uniquely determined by $\sigma(1), \sigma(2), \ldots, \sigma(n)$, 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 $i, \sigma(i), \sigma^2(i), \ldots, \sigma^n(i)$ 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 $i, \sigma(i), \ldots, \sigma^{r_i-1}(i)$ are all distinct. Observe that the columns of the resulting array allow us to easily read off the consecutive integers $i, \sigma(i), \sigma^2(i), \ldots, \sigma^{r_i-1}(i)$. Considering that $\sigma(\sigma^{r_i-1}(i)) = \sigma^{r_i}(i) = i$, it follows that $i, \sigma(i), \sigma^2(i), \ldots, \sigma^{r_i-1}(i)$ constitute a **cycle**; we will refer to the positive integer r_i as the **length** of the cycle $(i, \sigma(i), \sigma^2(i), \ldots, \sigma^{r_i-1}(i))$, 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 $\mathrm{ord}(\sigma) = r_i$. Conventionally, cycles are written without commas, but we will use them when convenient. We will also say that two cycles $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ 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 1.5.2. 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 this section, 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 1.12.58, we have that $\phi_1\rho_1 \neq \rho_1\phi_1$, hence the symmetric group is not necessarily abelian.

Example 1.5.3. 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 1.5.4. *Every permutation can be written as a product of disjoint cycles.*

Proof. Given any permutation σ of $[n]$, observe that the integers $1, \sigma(1), \ldots, \sigma^n(1)$ 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 $1, \sigma(1), \ldots, \sigma^{r_1-1}(1)$ are distinct; and $\sigma_1 = (1, \sigma(1), \ldots, \sigma^{r_1-1}(1))$ 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 $i_2, \sigma(i_2), \ldots, \sigma^n(i_2)$ 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 $\sigma_2 = (i_2, \sigma(i_2), \ldots, \sigma^{r_2-1}(i_2))$ 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 $\sigma_1, \sigma_2, \ldots, \sigma_k$ such that every element of $[n]$ lies in one and only one cycle and $\sigma = \sigma_1 \cdots \sigma_k$. □

Proposition 1.5.5. *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 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 1.5.6. *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 $1, 2, \ldots, n$ appears in one and only one cycle σ_i. We are now able to prove the following.

Proposition 1.5.7. *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* $\mathrm{ord}(\sigma) = \mathrm{lcm}(r_1, \ldots, r_k)$.

Proof. By Proposition 1.5.5, the disjoint cycles $\sigma_1, \ldots, \sigma_k$ commute, hence we have that

$$\mathrm{ord}(\sigma) = \mathrm{ord}(\sigma_1 \cdots \sigma_k) = \min\{r \geq 1 \,|\, (\sigma_1 \cdots \sigma_k)^r = \iota\} = \min\{r \geq 1 \,|\, \sigma_k^r \cdots \sigma_1^r = \iota\}.$$

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 $\sigma_1, \ldots, \sigma_k$ are disjoint. Consequently, we conclude that

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

Explicitly, the second equality follows from Corollary 1.3.16, and the third equality follows from the definition of the least common multiple that precedes Exercise 0.6.31. □

Permutations of order two are called **involutions**. By Propositions 1.5.6 and 1.5.7, a permutation is an involution if and only if its cycle decomposition is the product of disjoint transpositions.

Example 1.5.8. Consider the permutation σ of Example 1.5.3 with disjoint cycles (125), (38), (4), and (67). By Proposition 1.5.6, its cycle decomposition is given by $\sigma = (125)(38)(4)(67)$, hence we have that $\text{ord}(\sigma) = \text{lcm}(3, 2, 1, 2) = \text{lcm}(6, 1, 2) = \text{lcm}(6, 2) = 6$ by Proposition 1.5.7.

Corollary 1.5.6 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 1.5.9. 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 $1, 2, \ldots, n$ 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 $\sigma(1), \sigma(2), \ldots, \sigma(n)$ are all found.

Example 1.5.10. 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 $1, 2, \ldots, 8$ 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 1.5.9 to find the two-line notation for the resulting permutation.

Algorithm 1.5.11. We reconstruct the two-line notation for any permutation $\sigma = \sigma_1 \cdots \sigma_k$ that is a product of (not necessarily disjoint) cycles $\sigma_1, \ldots, \sigma_k$ 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 $1, 2, \ldots, n$ 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 $\sigma(1), \sigma(2), \ldots, \sigma(n)$ 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 1.5.12. 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 $\{4, \ldots, n\}$. Consequently, a permutation is uniquely determined by its cycle decomposition (excluding 1-cycles) regardless of the symmetric group to which it belongs.

Proposition 1.5.13. *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 1.5.14. *Let σ be any permutation with cycle decomposition $\sigma_1 \cdots \sigma_k$ and cycle type $(r_1, \ldots, r_k)$. We have that $\sigma^{-1} = \sigma_1^{r_1 - 1} \cdots \sigma_k^{r_k - 1}$.*

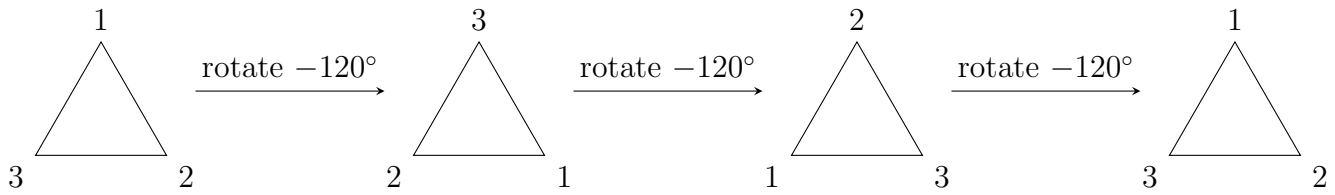
Ultimately, Proposition 1.5.14 makes small work of the matter of finding inverses of permutations written in cycle decomposition: observe that $(a_1, \ldots, a_k)^{-1} = (a_1, a_k, a_{k-1}, \ldots, a_3, a_2)$.

1.6 Rigid Motions and Dihedral Groups

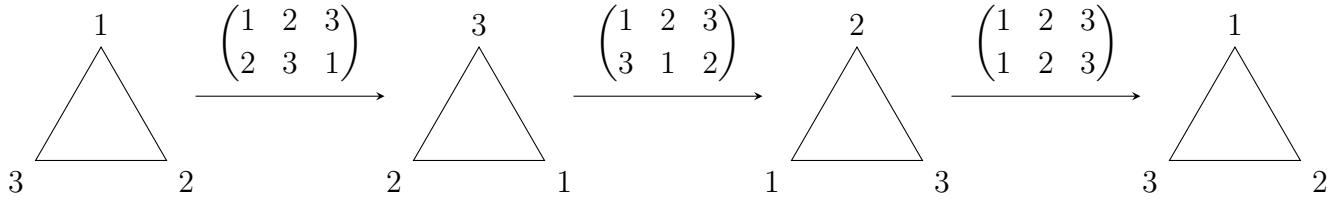
Recall that a **polygon** is a two-dimensional object consisting of straight line segments that intersect to form a closed and bounded region in the plane. Common examples of polygons include triangles, rectangles, and stars. Each of intersection point of a pair of straight line segments is called a **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.

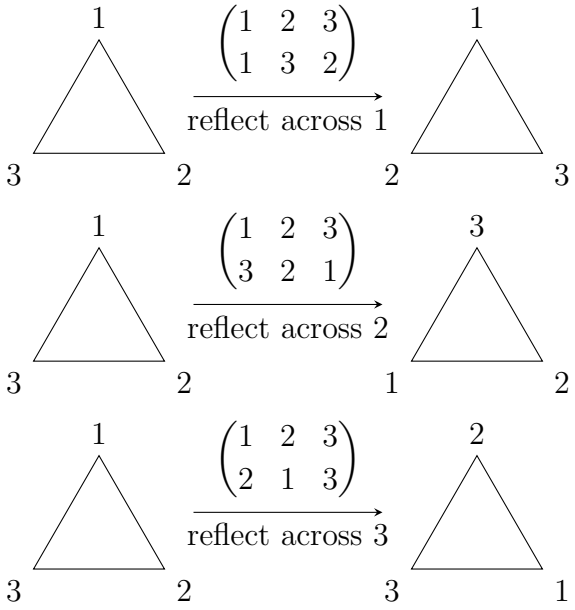
We will fix our attention throughout this section on two specific rigid motions of any regular n-gon. Each of the n vertices of a regular n-gon lies on the circumference of a circle. Consequently, for any integer $1 \leq k \leq n$, a **rotation** of a regular n-gon through an angle of $-360k/n$ degrees 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.



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.



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 consider this concept next in the context of our discussion of groups.

Concretely, we must notice 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

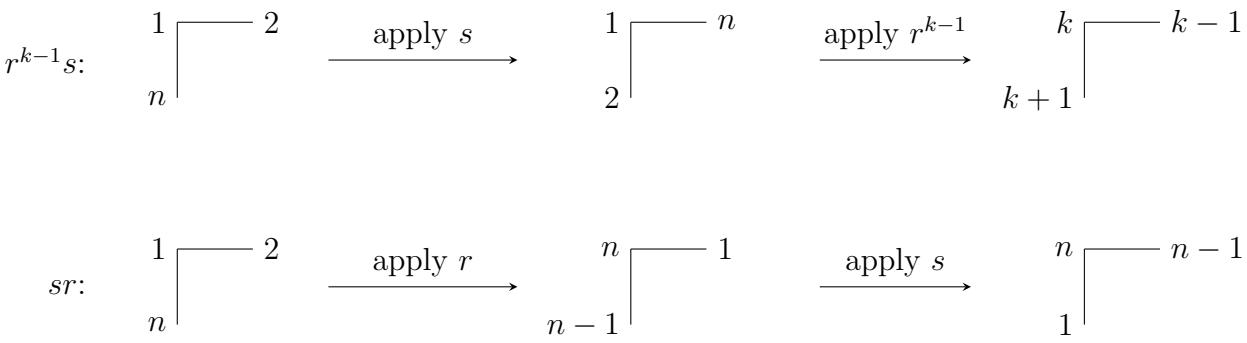
n). Likewise, the reflection of any regular n-gon across a vertex k swaps the labels of the vertices other than k according to the parity of n. Our immediate aim is to establish that the collection D_n of these rotations and reflections of a regular n-gon constitutes a subgroup of order $2n$ of the symmetric group $\mathfrak{S}_n$ on n letters.

Proposition 1.6.1. *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 $1, 2, \ldots, n$ 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 $s, rs, r^2s, \ldots, r^{n-1}s$.

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 $sr = r^{n-1}s$. □



Proposition 1.6.2. *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 $1, \ldots, n$, hence D_n is a subset of $\mathfrak{S}_n$. By Proposition 1.6.1, the distinct elements of D_n are $r, r^2, \ldots, r^n, s, rs, r^2s, \ldots, r^{n-1}s$, 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 1.6.1, 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 $x = r^k s$ and $y = r^\ell$ for some integers $1 \leq k, \ell \leq n$, we conclude that $xy = r^k s r^\ell = r^k r^{\ell(n-1)} s = r^{\ell(n-1)+k} s$ lies in D_n. Last, if $x = r^k s$ and $y = r^\ell s$ for some integers $1 \leq k, \ell \leq s$, then $xy = r^k s r^\ell s = r^k r^{\ell(n-1)} s^2 = r^{\ell(n-1)+k}$ is in D_n. □

We will henceforth refer to D_n as the **dihedral group** of order $2n$ in light of Proposition 1.6.2. 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 1.6.3. Consider the dihedral group D_4 of order 8, i.e., the group of symmetry-preserving rotations and reflections of a square. By Proposition 1.6.1, 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 $r^2 s$ across the vertices 2 and 4; and the reflection $r^3 s$ 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 ca 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.

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

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

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

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

We leave the above details for the reader to check pictorially in Exercise 1.12.59.

Example 1.6.4. 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 1.6.1, 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 $r^2 s$ across vertex 2; the reflection $r^3 s$ across vertex 5; and the reflection $r^4 s$ across vertex 3.

$$\begin{aligned} 1 &= (1) \\ r &= (12345) \\ r^2 &= (13524) \\ r^3 &= (14253) \\ r^4 &= (15432) \end{aligned}$$

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

We leave the above details for the reader to check pictorially in Exercise 1.12.60.

1.7 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 1.2.13 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 1.7.1. 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 1.6.1. 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 1.8.

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 1.7.2. 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 1.7.3. *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 1.7.4. *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 1.12.65 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 1.12.66). 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 1.7.5. 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 1.7.1. We established previously that $[G : H] = 3$ and $[G : K] = 2$.

Example 1.7.6. We established in Example 1.7.2 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 1.7.7. *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 Proposition 0.1.86(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 $g_1H = g_2H$, then $Hg_1^{-1} = Hg_2^{-1}$. (Essentially, this is the converse of the definition of injective.) But this holds by Proposition 1.7.4. Even more, the same proposition illustrates that if $Hg_1^{-1} = Hg_2^{-1}$, then $g_1H = g_2H$, 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 1.7.8. *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 Proposition 0.1.86(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 1.7.9. *Let G be a group with a subgroup H. We have that $g_1H \sim g_2H$ if and only if $g_1g_2^{-1} \in H$ 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 $g_1H \sim g_2H$ if and only if $g_1g_2^{-1} \in H$ is reflexive, symmetric, and transitive.

1.) By assumption that H is a subgroup of G, we have that $e_G = g_1g_1^{-1}$ lies in H for all left cosets g_1H of H in G, hence we have that $g_1H \sim g_1H$, and the relation is reflexive.

2.) If $g_1H \sim g_2H$, then $g_1g_2^{-1}$ lies in H. Once again, by hypothesis that H is a subgroup of G, it follows that $g_2g_1^{-1} = (g_1g_2^{-1})^{-1} \in H$ so that $g_2H \sim g_1H$, i.e., the relation is symmetric.

3.) If $g_1H \sim g_2H$ and $g_2H \sim g_3H$, then both $g_1g_2^{-1}$ and $g_2g_3^{-1}$ lie in H. Consequently, their product $g_1g_3^{-1} = (g_1g_2^{-1})(g_2g_3^{-1})$ lies in H so that $g_1H \sim g_3H$, and the relation is transitive.

By Proposition 1.7.4, the inclusion $g_1g_2^{-1} \in H$ is equivalent to equality of the left cosets $g_1H = g_2H$, hence left coset equality is an equivalence relation on the left cosets of H in G. By Corollary 0.1.51, 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 1.7.10 (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 1.7.9, 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 1.7.8. □

Remark 1.7.11. 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 1.7.9, there exist elements $g_1, \ldots, g_n \in G$ such that $g_1H, \ldots, g_nH$ are pairwise disjoint and $G = g_1H \cup \cdots \cup g_nH$. Consequently, we have that $|G| = \sum_{i=1}^{n} |g_iH|$. Lemma 1.7.8 yields that $|g_iH| = |H|$, and there are $[G : H]$ summands.

Corollary 1.7.12. *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 1.7.13. *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 $\text{ord}(g) = |G|$, i.e., $G = \langle g \rangle$. □

Corollary 1.7.14. *Every group of prime order is abelian.*

Proof. By Corollary 1.7.13, such a group is cyclic and hence abelian by Proposition 1.3.10. □

Corollary 1.7.15. *If G is a finite group, then $\text{ord}(g)$ 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 $\text{ord}(g)$ divides $|G|$. □

Corollary 1.7.16. *If G is a finite group, then $g^{|G|} = e_G$ for every element $g \in G$.*

Proof. By Corollary 1.7.15, there exists a positive integer q such that $|G| = \text{ord}(g)q$. Consequently, by the Group Exponent Laws, it follows that $g^{|G|} = g^{\text{ord}(g)q} = (g^{\text{ord}(g)})^q = (e_G)^q = e_G$. □

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 1.7.17 (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 1.12.54, 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 1.7.11. By Proposition 1.7.4, 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. □

1.8 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 1.8.1. *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 1.7.4, 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 1.8.1 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 1.8.2. *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 1.8.3. 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 1.7.1, we have that $xK = Kx$ for every element $x \in D_3$. Consequently, it follows by Proposition 1.8.1 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 1.8.4. *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 1.8.5. 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 1.8.4, $2\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$, hence we may form the quotient group $\mathbb{Z}/2\mathbb{Z}$. By Example 1.7.2, $(\mathbb{Z}/2\mathbb{Z}, +)$ 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 1.8.4 yields that $n\mathbb{Z}$ is normal, hence Corollary 1.8.2 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 $0+n\mathbb{Z}, 1+n\mathbb{Z}, \ldots, (n-1)+n\mathbb{Z}$ 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 1.8.6. *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. (a.) 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.

(b.) 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 $(g_1H)(g_2H) = g_1g_2H = g_2g_1H = (g_2H)(g_1H)$. □

1.9 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 1.9.1. 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 1.9.2. 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 1.4.4. 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 1.9.3. 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 1.9.4. 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 1.9.5. *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 (a.) 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 1.3.16.

(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 1.9.6. *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 1.9.5, 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 1.9.5, 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 1.9.7. 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 1.9.8. 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 1.9.6 that φ is injective.

Example 1.9.9. 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 1.9.10. 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 1.9.11. 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 1.9.12. 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 1.9.8, we conclude by Proposition 0.1.86 that φ is surjective, hence it is an isomorphism.

Example 1.9.13. 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 1.9.14. 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 1.9.15. 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 1.9.16. 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 1.9.12, 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 1.9.17. *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. (a.) Exercise 0.6.14(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.

(b.) 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

$$\varphi^{-1}(\varphi(g_1) \star \varphi(g_2)) = \varphi^{-1}(\varphi(g_1 * g_2)) = g_1 * g_2 = \varphi^{-1}(\varphi(g_1)) * \varphi^{-1}(\varphi(g_2)).$$

(c.) 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.

(d.) 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.

(e.) By the fifth part of Proposition 1.9.5, 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 1.9.18. *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 1.9.19. *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, Proposition 0.1.86(d.) implies that φ is bijective if and only if it is injective if and only if it is surjective. □

Theorem 1.9.20. *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 1.9.19 that φ is an isomorphism. □

Example 1.9.21. 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 1.9.20, 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 $\varphi(n + 4\mathbb{Z}) = i^n$.

1.10 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 1.12.84 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 1.10.1. We have already seen in Example 1.2.13 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 1.9.17. Every cyclic group of order four is isomorphic to $\mathbb{Z}/4\mathbb{Z}$ by Proposition 1.9.20; 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 1.10.2 (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 1.9.5. Exercise 1.12.83 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 1.7.4, 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 1.10.3 (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 1.12.75, 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 1.8.1, 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 1.12.76, 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 $\varphi(h_1h_2) = h_1h_2N = (h_1N)(h_2N)$ 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 $hN = e_GN$. By Proposition 1.7.4, we have that $hN = e_GN$ 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 1.10.4 (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 1.8.1, 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 1.8.1.

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 $g_1N = g_2N$, then $g_2^{-1}g_1$ is an element of N by Proposition 1.7.4. 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 $\varphi(g_1N) = g_1H = g_2H = \varphi(g_2N)$ 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 $gH = \varphi(gN) = e_GH$ if and only if $g \in H$ by Proposition 1.7.4, hence we conclude that $\ker\varphi = H/N$. By the First Isomorphism Theorem, we conclude that $(G/H)/(H/N) \cong G/N$. □

Theorem 1.10.5 (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 1.12.83; 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 1.12.83. □

Example 1.10.6. Consider the general linear group $\mathrm{GL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) \neq 0\}$ under matrix multiplication. We will use the First Isomorphism Theorem to prove that the multiplicative group $(\mathbb{R}^*, \cdot)$ of nonzero real numbers is isomorphic to a subgroup of $\mathrm{GL}(2, \mathbb{R})$. Consider the function $\varphi : \mathbb{R}^* \to \mathrm{GL}(2, \mathbb{R})$ 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, $(\mathbb{R}^*, \cdot)$ is isomorphic to $\varphi(\mathbb{R}^*) = \{cI \mid c \in \mathbb{R}^*\}$, i.e., the nonzero real multiples of the identity matrix.

Example 1.10.7. We will prove next that multiplicative group $(\mathbb{R}_{>0}, +)$ of positive real numbers is isomorphic to a proper quotient of the multiplicative group $(\mathbb{R}^*, \cdot)$ of nonzero real numbers, hence these groups are not isomorphic. Consider the function $\nu : \mathbb{R}^* \to \mathbb{R}_{>0}$ 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 $(\mathbb{R}^*/\{-1, 1\}, \cdot) \cong (\mathbb{R}_{>0}, \cdot)$ and $(\mathbb{R}_{>0}, \cdot) \ncong (\mathbb{R}^*, \cdot)$.

Example 1.10.8. 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 $\mathbb{Z}$.

1.11 Chapter 1 Overview

Check back at a later date, as this section is currently under construction.

1.12 Chapter 1 Exercises

Exercise 1.12.1. Prove or disprove that $\mathbb{R}$ forms a group with respect to multiplication.

Exercise 1.12.2. Use the definition of a group and the fact that real multiplication is associative and commutative to prove that $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ forms an abelian group with respect to multiplication.

Exercise 1.12.3. Use the definition of a group and the fact that real multiplication is associative and commutative to prove that $G = \{-1, 1\}$ forms an abelian group with respect to multiplication.

Exercise 1.12.4. Consider the complex number $i^2 = -1$. Use the definition of a group and Proposition 1.4.1 to prove that $G = \{-1, 1, -i, i\}$ forms an abelian group with respect to multiplication.

Exercise 1.12.5. Use the definition of a group and the fact that real addition is associative and commutative to prove that the set $\mathbb{R}^{\mathbb{R}}$ of real univariate functions $f : \mathbb{R} \to \mathbb{R}$ forms an abelian group with respect to the function addition defined for all real numbers x by $(f + g)(x) = f(x) + g(x)$.

Exercise 1.12.6. Use the definition of a group and the fact that real multiplication is associative and commutative to prove that the set $\mathbb{R}^{\mathbb{R}}$ of real univariate functions $f : \mathbb{R} \to \mathbb{R}$ forms an abelian group with respect to the function multiplication defined for all real numbers x by $fg(x) = f(x)g(x)$.

Exercise 1.12.7. Use the definition of a group to prove that the set $\mathbb{R}^{\mathbb{R}}$ of real univariate functions $f : \mathbb{R} \to \mathbb{R}$ forms a group with respect to the function composition defined for all real numbers x by $(f \circ g)(x) = f(g(x))$. Prove that $(\mathbb{R}^{\mathbb{R}}, \circ)$ is abelian or provide an explicit counterexample.

Exercise 1.12.8. Consider any group G. 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 1.12.9. Given any positive integer n, consider the set $\mathbb{Z}_n$ of equivalence classes of integers modulo n. Prove or disprove that $\mathbb{Z}_n$ forms an abelian group with respect to multiplication mod n.

(**Hint:** We suggest the reader revisit and complete Exercise 0.6.12 for reference.)

Exercise 1.12.10. Given any prime number p, consider the set $\mathbb{Z}_p$ of equivalence classes of integers modulo p. Prove or disprove that $\mathbb{Z}_p$ forms an abelian group with respect to multiplication mod p.

(**Hint:** We suggest the reader revisit and complete Exercise 0.6.41 for reference.)

Exercise 1.12.11. 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 1.12.12. Prove that the Group Exponent Laws hold for any group G.

Exercise 1.12.13. Prove that a group G is abelian if $ghg^{-1}h^{-1} = e_G$ for all elements $g, h \in G$.

Exercise 1.12.14. Prove that a group G is abelian if $(gh)^2 = g^2h^2$ for all elements $g, h \in G$.

(**Hint:** Give a symbolic simplification of the element $(gh)^2$ in two ways; then, compare your results.)

Exercise 1.12.15. Prove that a group G is abelian if $g^2 = e_G$ for every element $g \in G$.

(**Hint:** Use a previous exercise to derive this as a corollary.)

Exercise 1.12.16. Prove that a group G is abelian if $gh = g^{-1}h^{-1}$ for all elements $g, h \in G$.

(**Hint:** Use a previous exercise to derive this as a corollary.)

Exercise 1.12.17. Prove that a group G is abelian if $g^3 = e_G$ and $g^4h = hg$ for all $g, h \in G$.

Exercise 1.12.18. Prove that any group of order four is abelian.

Exercise 1.12.19. Given any group G, we define the **center** of G as follows.

$$Z(G) = \{x \in G \mid gx = xg \text{ for all elements } g \in G\}$$

Prove that $Z(G)$ is a subgroup of G. (We derive the notation from the German "das Zentrum.")

Exercise 1.12.20. Given any group G, we define the **centralizer** of an element $x \in G$ as follows.

$$Z_G(x) = \{g \in G \mid gx = xg\}$$

Prove that the centralizer $Z_G(x)$ of an element $x \in G$ is a subgroup of G.

Exercise 1.12.21. Given any group G, the **conjugate** of a subgroup H by an element $g \in G$ is

$$gHg^{-1} = \{ghg^{-1} \mid h \in H\}.$$

Prove that the conjugate gHg^{-1} of a subgroup H of G by an element $g \in G$ is a subgroup of G.

Exercise 1.12.22. Given any group G, we define the **normalizer** of a subgroup H of G as follows.

$$N_G(H) = \{g \in G \mid gHg^{-1} = H\}$$

Prove that the normalizer of a subgroup H of G is 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 1.12.23. Construct an explicit counterexample to disprove the following statement: if G is any group and H and K are any subgroups of G, then $H \cup K$ is a subgroup of G.

Exercise 1.12.24. Prove that if G is a group with subgroups H and K, then $H \cap K$ is a subgroup of G. Conclude by induction that the finite intersection of subgroups of G is a subgroup of G.

Exercise 1.12.25. Prove that if G is a group with subgroups $\{G_i\}_{i \in I}$, then $\cap_{i \in I} G_i$ is a subgroup.

Exercise 1.12.26. Prove that if G is a group with subgroups H and K, then the **product** HK of H and K defined by $HK = \{hk \mid h \in H \text{ and } k \in K\}$ is a subgroup of G if and only if $HK = KH$.

Exercise 1.12.27. Prove or disprove that the rational numbers $\mathbb{Q}$ form a cyclic group.

Exercise 1.12.28. Consider the multiplicative group $\mathbb{Z}_p^\times = \{a \in \mathbb{Z}_p \mid \gcd(a, p) = 1\}$ of integers modulo a prime number p. Verify that $\mathbb{Z}_p^\times$ is a cyclic group of order $p - 1$ for $p = 2, 3, 5, 7$, and 11.

Exercise 1.12.29. List the elements of $\mathbb{Z}_n^\times$ for each integer $n = 4, 6, 8, 9$, and 10. Be sure to make note of which of these groups is cyclic, and provide at least one generator for each cyclic group.

Exercise 1.12.30. Consider the multiplicative group $\mathbb{Z}_n^\times = \{a \in \mathbb{Z}_n \mid \gcd(a, n) = 1\}$ of integers modulo n. Prove that there exists a non-identity element $a \in \mathbb{Z}_n^\times$ of order two.

Exercise 1.12.31. Consider the additive group $\mathbb{Z}_p$ of integers modulo a prime number p. Prove that $(\mathbb{Z}_p, +)$ admits no subgroups other than itself and the trivial subgroup (i.e., it is **simple**).

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

Exercise 1.12.33 (Fermat's Little Theorem)**.** Prove that $|\mathbb{Z}_p^\times| = p - 1$ 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 with $p \nmid a$.

Exercise 1.12.34. Prove that if G is a group and g is any element of G, then the cyclic subgroup $\langle g \rangle$ of G generated by g is the intersection of all subgroups of G that contain g.

Exercise 1.12.35. Consider any group G. Prove that the following statements hold.

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

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

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

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

Exercise 1.12.36. Consider any elements g and h of an abelian group G such that the orders of g, h, and gh are finite. Concretely, we will henceforth denote $\text{ord}(g) = r$, $\text{ord}(h) = s$, and $\text{ord}(gh) = t$.

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

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

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

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

(**Hint:** Use Corollary 1.3.16 and make the "inspired substitution" $h^{st} = e_G$ in $g^{st} = g^{st}h^{st}$ of part (b.), to quote the great Lucian Grand. Use Euclid's Lemma and Exercise 0.6.34 for part (c.).)

Exercise 1.12.37. Consider any abelian group G and any prime number p. Prove that for any elements $g, h \in G$ such that $\text{ord}(g) = p^m$ and $\text{ord}(h) = p^n$, we have that $\text{ord}(gh) = \max\{m, n\}$.

Exercise 1.12.38. Consider any cyclic group G of order n. Prove that $\text{ord}(x) \mid n$ for all $x \in G$.

Exercise 1.12.39. Consider any cyclic group G of order n. Prove that for all positive divisors d of the positive integer n, there exists a cyclic subgroup of G of order d.

Exercise 1.12.40. Consider any abelian group G. Prove that the set $G_T = \{g \in G \mid \text{ord}(g) \text{ is finite}\}$ of elements of G of finite order is a subgroup of G; it is aptly called the **torsion subgroup** of G.

Exercise 1.12.41. Consider any cyclic group G that is generated by an element $g \in G$. Prove that if $g^n \in G$ generates G for some integer n, then we must have that $\gcd(n, \text{ord}(g)) = 1$.

Exercise 1.12.42. Prove that if some 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.

Exercise 1.12.43. Graph the fifth roots of unity in the complex plane. List each as a complex number of the form $\text{cis}(\theta)$ 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 1.12.44. Graph the sixth roots of unity in the complex plane. List each as a complex number of the form $\text{cis}(\theta)$ 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 those from Exercise 1.12.43 and explain the differences.

Exercise 1.12.45. 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 1.12.46. Consider the complex number $\omega = \cos(2\pi/5) + i\sin(2\pi/5)$.

(a.) Prove that ω is a primitive fifth root of unity.

(b.) Prove that $\omega^{-1} = \cos(-2\pi/5) + i\sin(-2\pi/5)$.

(c.) Prove that $\omega + \omega^{-1} = 2\cos(2\pi/5)$.

(d.) Prove that $\omega^4 + \omega^{-4} = 2\cos(2\pi/5)$.

(**Hint:** Recall that if $\theta + \phi = 2\pi$, then $\cos(\theta) = \cos(\phi)$; such angles are **explementary**.)

Exercise 1.12.47. Prove that if $z \in \mathbb{C}^*$ 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 1.12.48. Find all elements of finite order in the multiplicative group of complex numbers.

Exercise 1.12.49. 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 $\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1$.)

Exercise 1.12.50. Given any positive integer n, consider the set $[n] = \{1, 2, \ldots, n\}$. Complete the following sequential proof of Proposition 1.5.5. Use the first step provided as a guide for 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 $\sigma\tau(i) = \tau\sigma(i)$ for any disjoint cycles σ and τ.

Exercise 1.12.51. Prove that every k-cycle can be written as a product of transpositions. Conclude by Proposition 1.5.4 that every permutation can be written as a product of transpositions.

(**Hint:** Consider the k-cycle $(a_1, \ldots, a_k)$. 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 $\sigma(a_1, a_i)$ 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 1.12.52. Prove that a permutation σ is either even or odd but not both.

Exercise 1.12.53. Prove that a cycle of odd length is even and a cycle of even length is odd.

Exercise 1.12.54. 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 $\mathfrak{S}_4$.

Exercise 1.12.55. 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 1.12.56. 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 1.12.55 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 1.12.57. Conjecture the formula for the total number of symmetry-preserving rotations and reflections of a regular n-gon; then, prove that your formula holds.

(**Hint:** Use the example of Section 1.6, your work from Exercise 1.12.56, and possibly an additional example to spot the pattern and deduce a formula; then, use the Fundamental Counting Principle.)

Exercise 1.12.58. Consider the regular 3-gon of Section 1.6 whose vertices we labelled 1, 2, 3 in clockwise order. Consider the rotation ρ_k of the regular 3-gon through an angle of $-120k$ degrees. Explicitly, there are three distinct rotations ρ_1, ρ_2, and ρ_3. Consider the reflection ϕ_k 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						

Exercise 1.12.59. Verify the explanation of Example 1.6.3 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 1.12.60. Verify the explanation of Example 1.6.4 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 1.12.61. 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 1.6.1?)

Exercise 1.12.62. 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. Why does this not violate Exercise 1.12.36?

Exercise 1.12.63. 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 the characterization of $Z(D_n)$ below.

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

(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 an element of $Z(D_n)$.

(ii.) Use 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 step, 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 1.3.16, 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 1.12.64. 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?)

Exercise 1.12.65. Prove that (i.) $\implies$ (ii.) $\implies$ (iii.) $\implies$ (iv.) of Proposition 1.7.4 hold.

Exercise 1.12.66. 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 1.12.67. Prove that if H is a subgroup of some group G such that $[G : H] = 2$, then we have that $gH = Hg$ for all elements $g \in G$.

Exercise 1.12.68. Given any group G and any subgroups H and K of G, consider the product

$$HK = \{hk \mid h \in H \text{ and } k \in K\}$$

of Exercise 1.12.26. Prove that if H and K are finite, then $|HK| = \dfrac{|H||K|}{|H \cap K|}$.

Exercise 1.12.69. Consider any cyclic group G of order n. Exercise 1.12.39 implies that 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 $n = \sum_{d|n} \phi(d)$ for Euler's totient function $\phi : \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 1}$ defined before Euler's Theorem.

(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.

(c.) Combine the previous two parts to prove that $n = \sum_{d|n} \phi(d)$.

Exercise 1.12.70. Consider Euler's totient function $\phi : \mathbb{Z}_{\geq 1} \to \mathbb{Z}_{\geq 1}$ defined before Euler's Theorem. Complete the following steps to prove that $\phi(n)$ is an even integer for all integers $n \geq 2$.

(a.) Conclude by Euler's Theorem that $\phi(n) = |\mathbb{Z}_n^\times|$ for every positive integer n.

(b.) Conclude by Exercise 1.12.30 that there exists an element of order two in $\mathbb{Z}_n^\times$.

(c.) Conclude the desired result by Lagrange's Theorem.

Exercise 1.12.71. Consider a group G with a normal subgroup H. Prove that $\text{ord}(gH) \mid \text{ord}(g)$ for every element $g \in G$. Conclude that $\max\{\text{ord}(gH) \mid g \in G\} \leq \max\{\text{ord}(g) \mid g \in G\}$.

Exercise 1.12.72. Consider a group G with a subgroup H. Prove that if the index $[G : H]$ of H in G satisfies that $[G : H] = 2$, then H is a normal subgroup of G.

Exercise 1.12.73. 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 1.8.6 is false.

(**Hint:** By Corollary 1.7.13 and Exercise 1.12.72, 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 1.12.74. 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 1.8.6 is false.

(**Hint:** By Corollary 1.7.14 and Exercise 1.12.72, 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 1.12.75. Given any group G and any subgroups H and K of G, consider the product

$$HK = \{hk \mid h \in H \text{ and } k \in K\}$$

of Exercises 1.12.26 and 1.12.68. 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 1.12.76. Consider any group G 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 1.12.77. Consider any group G. 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 0.6.14(a.).)

Exercise 1.12.78. Consider the center $Z(G)$ of some group G defined in Exercise 1.12.19.

(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 1.12.79. Consider the normalizer $N_G(H)$ of a group $H \leq G$ defined in Exercise 1.12.22.

(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 $Z_G(h)$ of h in G defined in Exercise 1.12.20. We define the **centralizer** of H in G 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 1.8.1, 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 1.12.80. Consider the additive group of rational numbers $\mathbb{Q}$. 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.

Exercise 1.12.81. 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 1.12.82. 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 with the delightful fact 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 1.12.83. Consider any group homomorphism $\varphi : (G, *) \to (H, \star)$.

(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 1.12.84. 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. Particularly, if there are n distinct equivalence classes of groups that satisfy a property $\mathcal{P}$, then we say that there are n groups that satisfy property $\mathcal{P}$ **up to isomorphism**.

Exercise 1.12.85. Prove that there is only one infinite cyclic group up to isomorphism.

Exercise 1.12.86. Prove that there is only one cyclic group of order n up to isomorphism.

Exercise 1.12.87. Consider any group G 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:** By Cayley's Theorem (Theorem 3.4.1) 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 1.12.88. Consider any finite group G 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 1.12.87, 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 $H = N$.)

Exercise 1.12.89. Prove that $(\mathbb{Q}, +)$ and $(\mathbb{Z}, +)$ are not isomorphic.

(**Hint:** Exercise 1.12.27 and Proposition 1.9.18 directly imply this.)

Exercise 1.12.90. 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 1.12.91. Given any group G with a normal subgroup K, consider any group H such that there exists a group homomorphism $\varphi : G \to H$. (Use concatenation for all group operations.)

(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 1.12.92. Consider the general linear group $\mathrm{GL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) \neq 0\}$ and the special linear group $\mathrm{SL}(2, \mathbb{R}) = \{A \in \mathbb{R}^{2\times 2} \mid \det(A) = 1\}$. Let $\mathbb{R}^*$ denote the set of nonzero real numbers. Complete the following steps to prove that $(\mathrm{GL}(2, \mathbb{R})/\mathrm{SL}(2, \mathbb{R}), \cdot) \cong (\mathbb{R}^*, \cdot)$.

(a.) Prove that the determinant function is a group homomorphism from $\mathrm{GL}(2, \mathbb{R})$ to $(\mathbb{R}^*, \cdot)$.

(**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 $\ker \varphi = \mathrm{SL}(2, \mathbb{R})$.

(c.) Conclude the desired result by the First Isomorphism Theorem.

Exercise 1.12.93. 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 1.12.91 that $(\mathbb{R}/\mathbb{Z}, +) \cong (\mathbb{T}, \cdot)$.

Chapter 2

Essential Topics in Ring Theory

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.

2.1 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 2.1.1. 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 2.1.2. 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 2.1.3. 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 1.7.4 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 2.1.4. 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 3.1.

Example 2.1.5. 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 2.1.6. 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 2.1.7. Consider any finite collections of rngs $R_1, \ldots, R_n$. One can readily verify that the external direct product $R_1 \times \cdots \times R_n$ 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 $R_1, \ldots, R_n$. Even more, as before, the properties of $R_1 \times \cdots \times R_n$ are intimately connected with the properties of $R_1, \ldots, R_n$. Explicitly, we have that

1.) $R_1 \times \cdots \times R_n$ is a unital ring if and only if $R_1, \ldots, R_n$ are unital rings and

2.) $R_1 \times \cdots \times R_n$ is commutative if and only if $R_1, \ldots, R_n$ are commutative.

Explicitly, for the first property, the multiplicative identity must be the n-tuple $(1_{R_1}, \ldots, 1_{R_n})$.

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 2.1.8. *Consider any rng R with additive identity element 0_R.*

1.) *We have that $0_R r = 0_R = r0_R$ 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 $r0_R = 0_R$ follows similarly.

2.) Observe that the additive inverse of an element of an additive abelian group is unique by Proposition 1.2.2, 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 $rs + r(-s) = r(s + (-s)) = r0_R = 0_R$.

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 2.1.9 (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 2.7.4 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 2.1.10. We have seen in Example 1.1.7 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 2.1.11. 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 1.7.4 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 1.12.32.

Example 2.1.12. 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 2.1.13. 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 2.1.14. 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 2.1.15. 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 2.1.16. 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 2.7.9 and 2.7.36 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 1.9.6.

Proposition 2.1.17. *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 2.1.18. 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 2.1.19. 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 2.1.16 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 2.1.20. 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 2.1.21. 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 $\operatorname{char}(\mathbb{Z}/n\mathbb{Z}) = n$.

2.2 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 2.2.1 (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 2.1.8, 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 2.7.17 and 2.7.30).

Example 2.2.2. Each of the subset containments $\mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}$ induce subring containments.

Example 2.2.3. 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 2.2.4. *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 1.9.5. Even more, by Proposition 2.1.8, 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 2.7.4, 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 2.2.5. *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 2.2.4, 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 2.1.8; 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 2.2.5 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 2.2.6. 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 2.2.7. 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 2.2.8 (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 2.2.9. *Given any elements $x_1, \ldots, x_n$ of any commutative rng R, we have that*

$$(x_1, \ldots, x_n) = \{r_1x_1 + \cdots + r_nx_n \mid r_1, \ldots, r_n \in R\}$$

is an ideal of R called that is said to be **finitely generated** *by $x_1, \ldots, x_n$.*

Proof. We note that $(x_1, \ldots, x_n)$ contains $0_R = 0_R + \cdots + 0_R = 0_Rx_1 + \cdots + 0_Rx_n$ by Proposition 2.1.8, hence it is nonempty. By the Three-Step Ideal Test, in order to prove that $(x_1, \ldots, x_n)$ 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 $(r_1x_1 + \cdots + r_nx_n) - (s_1x_1 + \cdots + s_nx_n) = (r_1 - s_1)x_1 + \cdots + (r_n - s_n)x_n$ and $r(r_1x_1 + \cdots + r_nx_n) = r(r_1x_1) + \cdots + r(r_nx_n) = (rr_1)x_1 + \cdots + (rr_n)x_n$ for any elements $r \in R$ and $r_1x_1 + \cdots + r_nx_n \in (x_1, \ldots, x_n)$ by the associativity of multiplication. We conclude that $(x_1, \ldots, x_n)$ 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 $(x_1, \ldots, x_n)$ of R generated by $x_1, \ldots, x_n$ contains the elements $x_1, \ldots, x_n$ themselves. Consequently, we will restrict our attention to commutative unital rings when investigating these types of ideals.

Proposition 2.2.10. *Given any elements $x_1, \ldots, x_n$ of any commutative unital ring R and any ideal I of R, we have that $I \supseteq (x_1, \ldots, x_n)$ if and only if $x_1, \ldots, x_n \in I$.*

Proof. Consider any ideal $I \subseteq R$ such that $x_1, \ldots, x_n \in I$. We have that $r_1x_1, \ldots, r_nx_n \in I$ for all possible elements $r_1, \ldots, r_n \in R$ because I is an ideal of R and must therefore be closed under multiplication by elements of R. Even more, we have that $r_1x_1 + \cdots + r_nx_n \in I$ 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 $I \supseteq (x_1, \ldots, x_n)$. Conversely, every ideal $I \supseteq (x_1, \ldots, x_n)$ must contain each of the generators $x_1, \ldots, x_n$ 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 2.2.11. *Given any elements $x_1, \ldots, x_n$ of any commutative unital ring R, $(x_1, \ldots, x_n)$ is the smallest (with respect to inclusion) ideal of R that contains $x_1, \ldots, x_n$.*

Remark 2.2.12. One can define finitely generated left ideals and finitely generated right ideals by mimicking the definition of Proposition 2.2.9. Explicitly, for any elements $x_1, \ldots, x_n$ of a rng R,

$$R(x_1, \ldots, x_n) = \{r_1x_1 + \cdots + r_nx_n \mid r_1, \ldots, r_n \in R\}$$

is a left ideal of R by the proof of the above proposition, and by analogy, one can demonstrate that

$$(x_1, \ldots, x_n)R = \{x_1r_1 + \cdots + x_nr_n \mid r_1, \ldots, r_n \in R\}$$

is a right ideal of R. Of course, if R is commutative, then these ideals are equal. Each of Proposition 2.2.10 and Corollary 2.2.11 can likewise be generalized to the case of a non-commutative unital ring R. Explicitly, a left ideal I of R contains $R(x_1, \ldots, x_n)$ if and only if it contains $x_1, \ldots, x_n$, hence $R(x_1, \ldots, x_n)$ is the smallest (with respect to inclusion) left ideal of R that contains $x_1, \ldots, x_n$.

Example 2.2.13. 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 2.2.14. 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 2.2.9. Even more, we say that a system of generators $x_1, \ldots, x_n$ of an ideal $(x_1, \ldots, x_n)$ is a **minimal system of generators** whenever $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ 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

$$\mu(I) = \inf\{n \geq 0 \mid x_1, \ldots, x_n \text{ form a minimal system of generators of } I\}.$$

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 2.2.15. Considering that every integer n is the n-fold sum of 1, it follows that the abelian group $(\mathbb{Z}, +)$ is finitely generated by 1. 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 $p_1, \ldots, p_n$, the positive integers $x_i = p_1 \cdots p_n / p_i$ satisfy that $\gcd(x_1, \ldots, x_n) = 1$, hence Bézout's Identity yields that $a_1 x_1 + \cdots + a_n x_n = 1$ for some integers $a_1, \ldots, a_n$. Consequently, we may view $\mathbb{Z}$ as an ideal of itself generated by $(x_1, \ldots, x_n)$. Even more, this system of generators is minimal since the greatest common divisor of the integers in the set $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ is in fact the prime number p_i. Put another way, the ideal generated by $\{x_1, \ldots, x_n\} \setminus \{x_i\}$ 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 2.2.16. *Let R be a commutative unital ring with an ideal $I = (x_1, \ldots, x_n)$. Consider the ideal $J = (x_1, \ldots, x_{i-1}, u_1 x_1 + \cdots + u_n x_n, x_{i+1}, \ldots, x_n)$ for some units $u_1, \ldots, u_n \in R$, i.e., the ideal of R generated by the elements of $\{x_1, \ldots, x_n, u_1 x_1 + \cdots + u_n x_n\} \setminus \{x_i\}$. We have that $I = J$.*

Proof. We can immediately verify that $J \subseteq I$ by Proposition 2.2.10 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 2.2.17. 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 2.2.18. 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 2.7.28 demonstrates that the previous examples are indicative of a general phenomenon.

Example 2.2.19. Consider the ideal $I = (x^2 - 1, x^3 - x, x^4 - x^2)$ of $\mathbb{R}[x]$. By Proposition 2.2.16, 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 2.7.22, 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 1.12.26 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** $IJ = \{i_1j_1 + \cdots + i_nj_n \mid n \geq 1, i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J\}$ 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 2.7.23 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 2.2.20. *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 2.7.22 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 $i_1, \ldots, i_n \in I$ and $j_1, \ldots, j_n \in I$, the closure of I and J under addition yields that $i_1j_1 + \cdots + i_nj_n \in I \cap J$ so that $IJ \subseteq I \cap J$. □

Products of finitely generated ideals of a commutative rng are especially simple to describe.

Proposition 2.2.21. *For any elements $x_1, \ldots, x_m, y_1, \ldots, y_n$ of a commutative rng R, we have that*

$$(x_1, \ldots, x_m)(y_1, \ldots, y_n) = (x_iy_j \mid 1 \leq i \leq m \text{ and } 1 \leq j \leq n).$$

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 2.7.22, it follows that $(x_1, \ldots, x_m)(y_1, \ldots, y_n)$ is a two-sided ideal of R. Each of the products x_iy_j with $1 \leq i \leq m$ and $1 \leq j \leq n$ lies in this ideal, hence it follows by Proposition 2.2.10 that the ideal generated by x_iy_j for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$ lies in $(x_1, \ldots, x_m)(y_1, \ldots, y_n)$. Conversely, every element of the product ideal is of the form $i_1j_1 + \cdots + i_\ell j_\ell$ for some elements $i_1, \ldots, i_\ell \in (x_1, \ldots, x_m)$ and $j_1, \ldots, j_\ell \in (y_1, \ldots, y_n)$. Consequently, it suffices to prove that every product ij of an element $i \in (x_1, \ldots, x_m)$ and an element $j \in (y_1, \ldots, y_n)$ is an element of the ideal $(x_iy_j \mid 1 \leq i \leq m \text{ and } 1 \leq j \leq n)$. But this is not difficult: observe that if $i = r_1x_1 + \cdots + r_mx_m$ and $j = s_1y_1 + \cdots + s_ny_n$ for some elements $r_1, \ldots, r_m, s_1, \ldots, s_n \in R$, then

$$ij = (r_1x_1 + \cdots + r_mx_m)(s_1y_1 + \cdots + s_ny_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} r_is_jx_iy_j$$

is an element of the finitely generated ideal $(x_iy_j \mid 1 \leq i \leq m \text{ and } 1 \leq j \leq n)$ 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 2.2.22. 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 2.1.3.

Example 2.2.23. 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 real univariate 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 $a_n, \ldots, a_1, a_0$, 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 2.2.24. 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 2.2.25. *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 1.7.4, from which it follows that $\ker \pi = I$, as desired. Conversely, Proposition 2.2.5 shows that $\ker \varphi$ is a two-sided ideal for any rng homomorphism $\varphi : R \to S$. $\square$

2.3 Ring Isomorphism Theorems

We provide in this section analogs of the Group Isomorphism Theorems of Section 1.10 for rngs.

Theorem 2.3.1 (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 2.2.4, we have that $\varphi(R)$ is a subrng of S. Considering that $\ker\varphi$ is a two-sided ideal of R by Proposition 2.2.5, 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 1.7.4 and 2.2.4, 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 2.3.2 (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 2.3.3 (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 2.3.4 (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 2.1.8. 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 1.7.4; 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 2.3.5. 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 2.2.24. By the First Isomorphism Theorem for Rngs, we have that $R/\ker \varphi \cong \mathbb{R}$.

Example 2.3.6. 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 2.2.23).

Example 2.3.7. 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 $\mathbb{Z}$.

2.4 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 2.4.10). Conversely, a left **zero divisor** is any element $r \in R$ for which $rs = 0_R$ for some nonzero element $s \in R$.

Example 2.4.1. 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 1.7.4. 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 Euclid's Lemma if and only if $a + n\mathbb{Z}$ is a unit by Example 2.1.11. Explicitly, if $\gcd(a, n) = d > 1$, then by Exercise 0.6.34, it follows that $a + n\mathbb{Z}$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$.

Example 2.4.2. 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 2.4.3. 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 2.4.4. 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 2.4.5 (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 2.4.6. 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 2.1.10, we have that $U(\mathbb{Z}) = \{1, -1\}$, but there are infinitely many nonzero integers.

Example 2.4.7. Consider the commutative unital ring $\mathbb{Z}/p\mathbb{Z}$ for any prime number p. By Example 2.1.11, 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 2.4.8. 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 2.4.9. *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 2.1.8. □

Proposition 2.4.10. *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 2.4.11. *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 2.4.9. 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 2.4.12. *Every nonzero unital subring of a field is an integral domain.*

Example 2.4.13. 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 2.4.12 ensures that the Gaussian integers form an integral domain.

Proposition 2.4.14. *Every integral domain has characteristic 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 2.4.15. By Example 2.1.20, 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.

$$\text{finite fields} \subsetneq \text{(skew) fields} \subsetneq \text{(integral) domains} \subsetneq \text{(commutative) unital rings}$$

One can furthermore specialize this hierarchy to discuss different types of integral domains; however, for our purposes, we continue to explore the relationship between (integral) domains and (skew) fields.

Proposition 2.4.16. *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 2.4.10. Considering that R is finite, it follows that φ is surjective by Proposition 0.1.86, 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 2.4.17. *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 2.2.5, 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 2.4.18. *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 2.4.17, we conclude that φ is injective. □

Corollary 2.4.19. *If k is a skew field, then its only two-sided ideals are the zero ideal and itself.*

Proof. By Proposition 2.2.25, every two-sided ideal of k is the kernel of some unital ring homomorphism from k. By Proposition 2.4.17, 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). □

2.5 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 2.5.1. 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 2.4.7.

Example 2.5.2. 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 2.3.6, 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 2.5.3. *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 1.7.4. 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 2.5.4. *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 2.5.3, 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 2.5.5. 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 2.4.7 and Corollary 2.4.19.

Example 2.5.6. Consider the principal ideal (x) of the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. By Proposition 2.2.10, (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 2.3.6, $\mathbb{R}[x]/(x) \cong \mathbb{R}$ is a field.

Proposition 2.5.7. *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 2.7.25, 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 2.4.19, 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 2.5.8. *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 2.4.9, 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 2.5.7 so that it is an integral domain. We conclude by Proposition 2.5.3 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. □

By the exposition preceding Proposition 2.2.20, every pair of ideals I and J of a commutative unital ring R induce a product ideal $IJ = \{i_1j_1 + \cdots + i_nj_n \mid n \geq 1, i_1, \ldots, i_n \in I, j_1, \ldots, j_n \in J\}$. 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 2.5.9. *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 0.6.17, 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 2.5.8, 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 0.6.9, 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 2.5.10. *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 the set $\cup_{n=1}^{\infty} I_n$ 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. □

Our next observation is simple, but it will come to have astonishing significance.

Corollary 2.5.11. *Given any ascending chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of proper ideals of a commutative unital principal ideal ring R, there exists an integer $n \geq 1$ such that $I_k = I_n$ for all integers $k \geq n$.*

Proof. By virtue of Proposition 2.5.10, we have that $\cup_{n=1}^{\infty} I_n$ is a proper ideal of R. Consequently, there exists an element $x \in \cup_{n=1}^{\infty} I_n$ such that $\cup_{n=1}^{\infty} I_n = xR$ by assumption that R is a principal ideal ring. We claim that $I_n = xR$ for the smallest integer $n \geq 1$ such that $x \in I_n$. Explicitly, by Proposition 2.2.10, we have that $xR \subseteq I_n \subseteq \cup_{n=1}^{\infty} I_n = xR$, as desired. Even more, for each integer $k \geq n$, we have that $xR = I_n \subseteq I_k \subseteq \cup_{n=1}^{\infty} I_n = xR$ so that $I_k = I_n$. □

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; the ideas contained in the following proofs are quite common in commutative algebra, so we urge to read them carefully.

Theorem 2.5.12. *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 2.5.10, 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 2.5.13. *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 $\mathcal{P}$ must not be empty 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 2.5.10. 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 $J = R$. □

2.6 Chapter 2 Overview

Check back at a later date, as this section is currently under construction.

2.7 Chapter 2 Exercises

Exercise 2.7.1. Prove the Ring Exponent Laws.

Exercise 2.7.2. 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 2.7.3. 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 2.7.4. 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 2.7.5. 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 2.7.6. 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 2.7.7. Given any unital rings R and S, prove that $U(R \times S) = U(R) \times U(S)$.

Exercise 2.7.8. 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 2.7.9. Determine if the following pairs of commutative rngs are isomorphic.

(a.) $\frac{\mathbb{Z}}{6\mathbb{Z}}$ and $\frac{\mathbb{Z}}{2\mathbb{Z}} \times \frac{\mathbb{Z}}{3\mathbb{Z}}$

(b.) $\frac{\mathbb{Z}}{16\mathbb{Z}}$ and $\frac{\mathbb{Z}}{4\mathbb{Z}} \times \frac{\mathbb{Z}}{4\mathbb{Z}}$

(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 possible orders of elements of the rings in parts (a.) and (b.). Consider the possible square roots of elements of the rings in parts (e.) and (f.).)

Exercise 2.7.10. 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 2.1.16.)

Exercise 2.7.11. Consider the set $\text{End}(R) = \{\varphi : R \to R \mid \varphi \text{ is a rng homomorphism}\}$ of rng endomorphisms of a rng R. Prove that $\text{End}(R)$ is a non-commutative unital ring under composition.

Exercise 2.7.12. 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 $\text{Inn}(R) = \{\chi_u : R \to R \mid u \in U(R)\}$ the set of inner automorphisms of R.

(b.) Prove that $\text{Inn}(R)$ forms a non-abelian group under composition.

(c.) Prove that $\text{Inn}(R)$ is not an additive group. Conclude that $\text{Inn}(R)$ is not a rng.

(d.) Prove that the function $\psi : U(R) \to \text{Inn}(R)$ defined by $\psi(u) = \chi_u$ is a group homomorphism.

(e.) Compute the kernel of the group homomorphism $\psi : U(R) \to \text{Inn}(R)$.

Exercise 2.7.13. Compute the characteristic of each of the following commutative unital rings.

(a.) $\frac{\mathbb{Z}}{2\mathbb{Z}} \times \frac{\mathbb{Z}}{2\mathbb{Z}}$

(b.) $\frac{\mathbb{Z}}{2\mathbb{Z}} \times \frac{\mathbb{Z}}{3\mathbb{Z}}$

(c.) $\frac{\mathbb{Z}}{4\mathbb{Z}} \times \frac{\mathbb{Z}}{6\mathbb{Z}}$

(d.) $\mathbb{Q}$

(e.) $\mathbb{R}$

(f.) $\mathbb{C}$

Exercise 2.7.14. 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 2.7.15. Prove that if R is a finite unital ring, then the characteristic of R is positive.

Exercise 2.7.16. 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 2.7.2 that R is commutative.

Conclude that commutativity of a rng is a stronger condition than commutativity of a group.

Exercise 2.7.17. 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 2.7.18. 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 2.7.19. 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 2.7.20. 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 2.7.21. Consider the commutative unital ring $\mathbb{R}[x]$ of real univariate 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 2.7.22. 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 $IJ = \{i_1j_1 + \cdots + i_nj_n \mid n \geq 1, i_1, \ldots i_n \in I, j_1, \ldots, j_n \in J\}$ 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 2.7.23. 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 2.7.24. Prove that if $R \supseteq S$ are rngs, then $I \cap S$ is an ideal of S for any ideal I of R.

Exercise 2.7.25. 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 2.7.26. 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 2.7.27. 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.

We refer to a rng R for which all (one-sided) ideals are principal as a **principal ideal rng**.

Exercise 2.7.28. 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.

Exercise 2.7.29. Prove that if R is any principal ideal rng and I is any two-sided ideal of R, then the quotient ring R/I is a principal ideal rng.

(**Hint:** Use the Fourth Isomorphism Theorem for Rngs.)

Exercise 2.7.30. 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 2.7.31. 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 2.7.32. 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 2.7.31 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 2.7.33. 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 2.7.32. Consider the Fourth Isomorphism Theorem for Rngs for parts (a.) and (b.). Consider Exercise 2.7.28 for parts (c.), (d.), and (e.). Use Corollary 2.4.19 for part (f.).)

Exercise 2.7.34. Consider the non-commutative unital ring $\mathbb{R}^{n \times n}$ consisting of real $n \times n$ matrices for some positive integer $n \geq 2$. 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 2.7.35. 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 2.7.36. 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 $\mathbb{R}^{2n}$

Exercise 2.7.37. 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 2.7.38. 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 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 $1 \leq n \leq 6$. Conclude that φ is surjective.

(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 2.7.39. 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 2.7.40. 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 2.7.24 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 2.7.41. 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 $(R/J)/(I/J) \cong R/I$.

Exercise 2.7.42. Prove that $\mathbb{Z}[\sqrt{n}\,] = \{a + b\sqrt{n} \mid a, b \in \mathbb{Z}\}$ is an integral domain for any $n \in \mathbb{Z}$.

Exercise 2.7.43. Prove that $\mathbb{Q}(\sqrt{\alpha}) = \{a + b\sqrt{\alpha} \mid a, b \in \mathbb{Q}\}$ is a field for any rational number α.

Exercise 2.7.44. 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 2.7.45. 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 $\text{ind}(r) = \min\{k \geq 1 \mid r^k = 0_R\}$ 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 2.7.46. 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 2.7.47. Prove or disprove that if R and S are domains, then $R \times S$ is a domain.

Exercise 2.7.48. Prove that any domain R whose only ideals are $\{0_R\}$ and R is a skew field.

Exercise 2.7.49. 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 Proposition 0.1.86.)

Exercise 2.7.50. 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 2.7.51. 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 2.7.52. Prove that there are no nonzero unital ring homomorphisms $\varphi : \mathbb{Q} \to \mathbb{Z}$.

Exercise 2.7.53. 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 2.7.54. 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 2.7.35 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 $k^{n\times n}$.

Exercise 2.7.55. 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 2.7.56. 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 2.7.57. 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 2.7.58. 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 2.7.59. 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 2.7.60. 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 2.7.61. 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 2.7.62. Prove that if P is any prime ideal of a commutative unital ring R, then the set complement $R \setminus P = \{r \in R \mid r \notin P\}$ of P in R is a multiplicatively closed subset of R.

Exercise 2.7.63. 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 2.7.64. 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 2.7.65. 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 2.7.64.)

Exercise 2.7.66. 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 $\mathbb{Z}$.)

Chapter 3

Essential Topics in Field Theory

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.

3.1 Polynomial Rings and Polynomial Long Division

Given any rng R, define the collection of **univariate polynomials** in **indeterminate** x over R by

$$R[x] = \{r_n x^n + \cdots + r_1 x + r_0 \mid n \geq 0 \text{ is an integer and } r_0, r_1, \ldots, r_n \in R\}.$$

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 $p(x) + 0_R = p(x) = 0_R + p(x)$ 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 3.1.1. *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 2.1.8, 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 3.1.2. We are quite familiar with real polynomials already; however, we can also restrict our attention to polynomials with integer coefficients. By Proposition 3.1.1, 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 3.1.3. 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 3.7.1 and 3.7.2). 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 3.1.4. *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 3.1.1, 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 3.1.5. *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 $s_0, s_1, \ldots, s_m \in R$ 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 3.1.4 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 3.7.7; however, we will prove this for polynomials with coefficients in a domain.

Theorem 3.1.6 (Polynomial Division Algorithm)**.** *Given any domain R, 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 3.1.4 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 3.1.7. 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 $p(x) = x - 1$.

$$\begin{array}{r|l} & 3x^2 + 5x + 4 \\ \hline x - 1 & 3x^3 + 2x^2 - x + 1 \\ & -3x^3 + 3x^2 \\ \hline & \quad 5x^2 - x \\ & \quad -5x^2 + 5x \\ \hline & \qquad 4x + 1 \\ & \qquad -4x + 4 \\ \hline & \qquad\quad 5 \end{array}$$

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 3.1.8. 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 $p(x) = x + 2$.

$$\begin{array}{r|l} & x^2 - 2x + 5 \\ \hline x + 2 & x^3 + x + 1 \\ & -x^3 - 2x^2 \\ \hline & \quad -2x^2 + x \\ & \quad 2x^2 + 4x \\ \hline & \qquad 5x + 1 \\ & \qquad -5x - 10 \\ \hline & \qquad\quad -9 \end{array}$$

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 2.3.6 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 3.1.9 (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 3.7.7, 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 3.1.10 (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 3.1.11. By Example 3.1.7 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 3.1.12. By Example 3.1.8 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.

Theorem 3.1.13 (Fundamental Theorem of the Roots of Polynomials over a Field)**.** *Every non-constant univariate polynomial $p(x)$ of degree n over a field k admits at most n roots lying in k.*

Proof. Consider any non-constant univariate polynomial $p(x)$ of degree n over a field k. We may denote by r the number of roots of $p(x)$ lying in k. By the Factor Theorem, if there exists a root α_1 of $p(x)$ lying in k, then there exists a unique polynomial $q_1(x)$ in $k[x]$ such that $p(x) = (x-\alpha_1)q_1(x)$. □

Corollary 3.1.14 (The Multiplicative Group of a Finite Field Is Cyclic)**.** *Given any finite field F, we have that $F^* = F \setminus \{0\}$ is a cyclic group with respect to the multiplication defined on F.*

Proof. Use the Fundamental Theorem of Finite Abelian Groups. Prove that the least common multiple $\operatorname{lcm}(a_1, \ldots, a_r)$ of the cyclic factors is the product of the cyclic factors: one way is clear since $\operatorname{lcm}(a_1, \ldots, a_r) \leq a_1 \cdots a_r$ by definition; the other way follows by showing that each element of F is a root of $x^{\operatorname{lcm}(a_1,\ldots,a_r)} - 1$ in $F[x]$. □

3.2 Polynomial Irreducibility

We turn our attention in this section to the question of polynomial factorization. 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 factorizations of a polynomial are 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 $r = ds$. We say that a nonzero non-unit element $p \in R$ is **prime** if and only if the principal ideal $(p) = \{px \mid x \in R\}$ of R generated by p is a prime ideal of R if and only if p satisfies the property that if p divides a product rs of elements $r, s \in R$, then either p divides r or p divides s. Even more, by analogy to the greatest common divisor of integers, we say that for any pair of nonzero elements $r, s \in R$, a **greatest common divisor** of r and s is any element $d \in R$ such that

(1.) $d \mid r$ and $d \mid s$, i.e., d divides both r and s and

(2.) if d' is a nonzero element of R such that $d' \mid r$ and $d' \mid s$, then $d' \mid d$.

One can demonstrate that if the greatest common divisor of some nonzero elements of an arbitrary integral domain exists, then it is unique up to multiplication by a unit of R. We will not concern ourselves at the moment with either the existence or the uniqueness of the greatest common divisor; rather, we will assume that R is an integral domain in which any pair of nonzero elements r and s admits a greatest common divisor $\gcd(r,s)$. Per usual, the greatest common divisor $\gcd(r_0,r_1,\ldots,r_n)$ of any collection of nonzero elements $r_0,r_1,\ldots,r_n \in R$ can be computed recursively via $\gcd(r_0,r_1,\ldots,r_n) = \gcd(\gcd(r_0,r_1),r_2,\ldots,r_n)$. 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 3.1.4, 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 $\text{content}(p) = \gcd(r_0,r_1,\ldots,r_n)$ of R. We say that a polynomial $p(x)$ is **primitive** if $\text{content}(p)$ 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 3.2.1. *Let R be an integral domain in which any pair of nonzero 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 $p(x) = \text{content}(p)q(x)$ 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 $\text{content}(p) = \gcd(r_0,r_1,\ldots,r_n)$ — namely, there exist elements $s_0,s_1,\ldots,s_n$ such that $r_i = \text{content}(p)s_i$ for each integer $0 \leq n$. Consider the polynomial $q(x) = s_n x^n + \cdots + s_1 x + s_0$. Observe that $p(x) = \text{content}(p)q(x)$, hence it suffices to prove that $q(x)$ is primitive, i.e., $\text{content}(q) = \gcd(s_0,s_1,\ldots,s_n)$ is a unit of R. Considering that $r_i = \text{content}(p)s_i$, we may uniquely identify $s_i = r_i/\text{content}(p)$ for each integer $1 \leq i \leq n$. One can prove using general considerations that $\text{content}(q) = \gcd(s_0,s_1,\ldots,s_n)$ is a unit of R 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 3.2.2. Every monic polynomial in $R[x]$ is primitive because its leading coefficient is 1_R.

Example 3.2.3. 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 3.2.4. 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 3.2.5. 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 3.2.6. 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 3.7.11 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 3.2.7. *Consider any integral domain R and any indeterminate x over R. Given any factorization $p(x) = q(x)r(x)$ of polynomials $p(x), q(x)$, and $r(x)$ in $R[x]$, if α is a prime element of R that divides every coefficient of $p(x)$, then either α divides every coefficient of $q(x)$ or α divides every coefficient of $r(x)$. Put another way, if $\alpha \mid q(x)r(x)$, then either $\alpha \mid q(x)$ or $\alpha \mid r(x)$.*

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 $R[x]$. On the contrary, suppose that the prime element α 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 α does not divide the ith coefficient a_i of $q(x)$, and there exists a least integer j such that α does not divide the jth coefficient b_j of $r(x)$ by the Well-Ordering Principle. Considering that α is a prime element, α

cannot divide $a_i b_j$; however, by assumption, α 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 $a_0, \ldots, a_{i-1}$ and $b_0, \ldots, b_{j-1}$ is divisible by α by construction of a_i and b_j, hence α divides $a_i b_j$ — a contradiction. We conclude that α divides $q(x)$ or n divides $r(x)$. □

Proposition 3.2.8. *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 $a_i, \ldots, a_1, a_0$ and some nonzero integers $b_i, \ldots, b_1, b_0$. 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 3.2.7. 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 minimality property that defines m. □

Put another way, Proposition 3.2.8 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 3.2.9 (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 3.2.8, 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 3.2.1, we may factor the polynomial $Q_0(x)$ of

$\mathbb{Z}[x]$ as $Q_0(x) = \text{content}(Q_0)q(x)$ 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

$$p(x) = Q(x)R(x) = \frac{1}{\alpha}Q_0(x)H(x) = \frac{\text{content}(Q_0)}{\alpha}q(x)H(x) = q(x)\left(\frac{\text{content}(Q_0)}{\alpha}R(x)\right).$$

By Proposition 3.2.8, there exists a polynomial $r(x)$ in $\mathbb{Z}[x]$ such that $p(x) = q(x)r(x)$ in $\mathbb{Z}[x]$. □

Theorem 3.2.10 (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 3.2.1, there exists a primitive polynomial $r(x)$ in $\mathbb{Z}[x]$ such that $p(x)q(x) = \text{content}(pq)r(x)$. Observe that as polynomials in $\mathbb{Q}[x]$, we have the following factorization of $r(x)$ by $p(x)$.

$$r(x) = p(x)\left(\frac{1}{\text{content}(pq)}q(x)\right)$$

By assumption that $p(x)$ is a primitive polynomial and $r(x)$ is an integer polynomial, Proposition 3.2.8 yields a nonzero 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 3.1.4, 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 3.2.11. *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 3.2.12. 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 3.2.11 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 $x^4 + 0x^3 + 0x^2 + 1x + 1 = x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd$.

$$\begin{aligned} a + c &= 0 & \qquad ad + bc &= 1 \\ ac + b + d &= 0 & \qquad bd &= 1 \end{aligned}$$

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 3.2.13. 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 3.2.11 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.

$$\begin{aligned} a + c &= 0 & \qquad ad + bc + e &= -4 & \qquad be &= 2 \\ ac + b + d &= 0 & \qquad ae + bd &= 0 \end{aligned}$$

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 3.2.14 (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 3.2.15. Observe that $p(x) = x^3 - 2$ is 2-Eisenstein and hence irreducible in $\mathbb{Q}[x]$.

Example 3.2.16. 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 3.2.17 (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 3.1.4, 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 3.2.18. 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 3.2.17, 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 3.2.19. *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 3.2.20. *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 3.2.21. *Consider the commutative unital ring $\mathbb{R}[x]$ of real univariate 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 3.2.20, real quadratic polynomials with negative discriminant are irreducible. Conversely, by Proposition 3.2.19, 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 2.1.19, 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. □

3.3 Roots of Polynomials and Field Extensions

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.

We will concern ourselves throughout this chapter with the univariate polynomial ring $k[x]$ for some field k. Like in Section 3.2, 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 $a_n, \ldots, a_1, a_0$ 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 3.2.21 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 $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ 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 3.3.1. *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 2.5.7, 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 3.3.1 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 3.3.2 (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 3.3.1 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 3.3.3. 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 2.7.18 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 3.3.4. 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 2.7.58, $\mathbb{R}[x]/(x^2 + 1)$ 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 3.3.5. 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 $\alpha(\alpha+1)=\alpha^2+\alpha=\alpha^2+\alpha+1+1=1$.

3.4 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 3.4.1. 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 3.4.2. 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 3.4.3. 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 3.4.4. Elements of an extension field F of k need not be algebraic: indeed, it is a non-trivial 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 $\alpha_1, \ldots, \alpha_n$ of F over k, we write $k(\alpha_1, \ldots, \alpha_n)$ to denote the smallest extension field of k lying in F that contains k and the elements $\alpha_1, \ldots, \alpha_n$. 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 $k(\alpha_1, \ldots, \alpha_n)$ is called a **finitely generated extension** of k.

Example 3.4.5. By Example 3.4.1, we have that $\mathbb{Q}(\sqrt{2})$ is a simple extension of $\mathbb{Q}$.

Example 3.4.6. By Example 3.4.2, we have that $\mathbb{Q}(i)$ is a simple extension of $\mathbb{Q}$.

Example 3.4.7. We will demonstrate that the field $(\mathbb{Z}/2\mathbb{Z})(\alpha)$ of Example 3.3.5 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 3.3.1 that implies it. Before this, we need the following lemma.

Lemma 3.4.8. *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 3.4.9. *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 3.4.8, 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 3.4.10. 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 3.4.11. 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 3.4.12. We have seen already in Example 3.4.3 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 3.4.13. Example 3.4.10 illustrates that $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.

Example 3.4.14. Example 3.4.11 illustrates that $[\mathbb{Q}(i) : \mathbb{Q}] = 2$.

Example 3.4.15. Example 3.4.12 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 3.4.16. *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 $\{1_k, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ 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 3.4.8 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 $a_n, \ldots, a_1, a_0$ 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 $\{1_k, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ 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 $1_k, \alpha, \alpha^2, \ldots, \alpha^{n-1}$. 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 3.4.8, 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 3.4.17. 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 2.7.18.

Example 3.4.18. 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 3.4.19. 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 $\{1, \sqrt{2 + \sqrt{3}}, (\sqrt{2 + \sqrt{3}})^2, (\sqrt{2 + \sqrt{3}})^3\}$.

3.5 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 3.5.1. *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 $1, \alpha, \alpha^2, \ldots, \alpha^n$ are linearly dependent over k. Consequently, there exist elements $a_0, a_1, a_2, \ldots, a_n \in k$ 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 3.5.2. *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 3.4.16, it follows that $k(\alpha)$ is a finite-dimensional k-vector space. □

Caution: the converse of Proposition 3.5.1 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 3.4.16, 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 3.5.3. *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 $\alpha_1, \ldots, \alpha_m$ form a k-vector space basis of F and $\beta_1, \ldots, \beta_n$ 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 $a_1, \ldots, a_n \in F$ 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 $b_{1i}, \ldots, b_{mi} \in k$. 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 $\alpha_1, \ldots, \alpha_m$ 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 3.5.4. *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 3.5.3 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 3.5.5. *Given any algebraic elements $\alpha_1, \ldots, \alpha_n$ of any extension field F of any field k, we have that $k(\alpha_1, \ldots, \alpha_i)$ is an algebraic extension of $k(\alpha_1, \ldots, \alpha_{i-1})$ 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 $k(\alpha_1, \ldots, \alpha_i) = k(\alpha_1, \ldots, \alpha_{i-1})(\alpha_i)$ 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 3.4.16, hence we find that $k(\alpha_1, \ldots, \alpha_n)$ is a finite extension of k; it must be algebraic by Proposition 3.5.1. □

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 $k(\alpha_1, \ldots, \alpha_n)$ over k. One immediate idea is to view $k(\alpha_1, \ldots, \alpha_n)$ as a simple extension $k(\alpha_1, \ldots, \alpha_{n-1})(\alpha_n)$; 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 3.5.6. *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 3.4.16. Even more, it follows that β is algebraic over k by Corollary 3.5.2, hence the simple extension $k(\beta)$ over k is finite by Proposition 3.4.16. 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 3.5.3 yields that $\deg(\mu_\alpha) = [k(\alpha) : k] = [k(\alpha) : k(\beta)][k(\beta) : k] = [k(\alpha) : k(\beta)] \deg(\mu_\beta)$. □

Example 3.5.7. Consider the finitely generated extension $\mathbb{Q}(\sqrt{2}, i)$ of $\mathbb{Q}$. By Example 3.4.13, 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 3.5.3 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 3.5.8. By Example 3.4.15, 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 3.5.9. 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 3.7.25, 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 3.5.3.

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 3.5.6. 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 3.5.1 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 $b=d=\pm2$ 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 3.7.25. 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 3.5.10. *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 $\alpha_1, \ldots, \alpha_n \in F$ such that $F = k(\alpha_1, \ldots, \alpha_n)$ 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 $\alpha_1, \ldots, \alpha_n$ such that $F = k(\alpha_1, \ldots, \alpha_n)$ and for each integer $1 \leq i \leq n$, we have that $k(\alpha_1, \ldots, \alpha_i) = k(\alpha_1, \ldots, \alpha_{i-1})(\alpha_i)$ is an algebraic extension of $k(\alpha_1, \ldots, \alpha_{i-1})$.*

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 $\alpha_1, \ldots, \alpha_n \in F$ of F as a k-vector space. Every element of F can be written as a k-linear combination of the elements $\alpha_1, \ldots, \alpha_n$, hence F is contained in $k(\alpha_1, \ldots, \alpha_n)$. Considering that $k(\alpha_1, \ldots, \alpha_n)$ is by definition the smallest extension field of k lying in F that contains k and the elements $\alpha_1, \ldots, \alpha_n$, we conclude that $F = k(\alpha_1, \ldots, \alpha_n)$. By Proposition 3.5.1, we must have that F is algebraic over k.

We will assume next that F admits elements $\alpha_1, \ldots, \alpha_n$ such that $F = k(\alpha_1, \ldots, \alpha_n)$ and α_i is algebraic over k for each integer $1 \leq i \leq n$. Corollary 3.5.5 ensures the desired result.

Last, we will assume that there exist elements $\alpha_1, \ldots, \alpha_n$ such that $F = k(\alpha_1, \ldots, \alpha_n)$ and for each integer $1 \leq i \leq n$, we have that $k(\alpha_1, \ldots, \alpha_i)$ is an algebraic extension of $k(\alpha_1, \ldots, \alpha_{i-1})$. Each of the simple extensions $k(\alpha_1, \ldots, \alpha_i)$ over $k(\alpha_1, \ldots, \alpha_{i-1})$ is algebraic by assumption, hence each one is finite by Proposition 3.4.16; we conclude that F is finite over k by Corollary 3.5.4. □

Corollary 3.5.11. *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 $a_0, a_1, \ldots, a_n \in F$ not all of which are zero such that $a_n\alpha^n + \cdots + a_1\alpha + a_0 = 0_F$. Considering that coefficients $a_0, a_1, \ldots, a_n$ induce a finitely generated extension $k(a_0, a_1, \ldots, a_n)$ over k, we find that α is algebraic over $k(a_0, a_1, \ldots, a_n)$ so that $k(a_0, a_1, \ldots, a_n, \alpha)$ is a simple algebraic extension over $k(a_0, a_1, \ldots, a_n)$ by Corollary 3.5.2. Even more, by assumption, each of the elements $a_0, a_1, \ldots, a_n$ is algebraic over k, hence the extension field $k(a_0, a_1, \ldots, a_n, \alpha)$ is obtained from a finite sequence of simple algebraic extensions over k. We conclude by Theorem 3.5.10 that $k(a_0, a_1, \ldots, a_n, \alpha)$ is a finite extension of k, hence there exists an integer $m \geq 0$ such that $1, \alpha, \alpha^2, \ldots, \alpha^m$ 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 $p(\alpha) = 0_k$. □

3.6 Chapter 3 Overview

Check back at a later date, as this section is currently under construction.

3.7 Chapter 3 Exercises

Exercise 3.7.1. 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 3.7.2. 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 3.7.3. 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 3.7.4. 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 3.7.5. 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 3.7.6. 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 3.7.7. 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.

(1.) $f(x)$ is the zero polynomial.

(2.) $p(x)$ 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 3.7.8. Consider the commutative unital ring $\mathbb{R}[x]$ of real polynomials in indeterminate x. Convert Exercise 2.7.28 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 3.7.9 (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 Euclid's Lemma that a divides c_0.

(d.) Conclude from the first step that b divides $c_n a^n$.

(e.) Conclude from Exercise Euclid's Lemma that b divides c_n.

Exercise 3.7.10. 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 3.7.11 (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 3.7.12. 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 3.7.13. Prove that there are infinitely many irreducible polynomials in $\mathbb{Q}[x]$.

Exercise 3.7.14. Prove that there are irreducible polynomials of arbitrary positive degree in $\mathbb{Q}[x]$.

Exercise 3.7.15. 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 3.7.16 (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 3.7.17. 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 3.7.18. 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 3.2.21, 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.)

Exercise 3.7.19. 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 3.7.20. 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 $\mathbb{Q}[x]/(x^2 - 3)$ 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 3.7.21. 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 $\mathbb{Z}/2\mathbb{Z}$ 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.

Exercise 3.7.22. Construct a finite field F with the specified number of elements.

(a.) $|F| = 8 = 2^3$

(b.) $|F| = 9 = 3^2$

(c.) $|F| = 27 = 3^3$

(d.) $|F| = 32 = 2^5$

Exercise 3.7.23. 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.) $\cos\left(\frac{2\pi}{5}\right)$

(**Hint:** On part (f.), it is useful to note that if $u = \frac{2\pi}{5}$, then $\cos(2u) = \cos(3u)$. Express $\cos(2u)$ and $\cos(3u)$ as polynomials in $\cos(u)$; then, use that $\cos(3u) - \cos(2u) = 0$ to conclude the result.)

Exercise 3.7.24. Compute the minimal polynomial of the following complex numbers over $\mathbb{Q}$.

(a.) $\sqrt{3}$

(b.) $\sqrt{1 + \sqrt{3}}$

(c.) $\sqrt{3} + \sqrt{5}$

(d.) $\sqrt[4]{5}$

(e.) $\sqrt[6]{2}$

(f.) $\cos\left(\frac{\pi}{4}\right) + i \sin\left(\frac{\pi}{4}\right)$

(g.) $i\sqrt[6]{2}$

(h.) $\cos\left(\frac{2\pi}{5}\right)$

(**Hint:** On part (h.), find an irreducible quadratic factor of the polynomial from Exercise 3.7.23(f.); then, argue that this irreducible quadratic polynomial is the minimal polynomial.)

Exercise 3.7.25. 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 3.7.26. Prove that if a and b are nonzero rational numbers, then $\mathbb{Q}(\sqrt{a}) = \mathbb{Q}(\sqrt{b})$ if and only if there exists a nonzero rational number c for which $a = c^2 b$.

Exercise 3.7.27. Prove that $\sqrt{\pi}$ is algebraic over $\mathbb{Q}(\pi)$; then, find the degree of $\mathbb{Q}(\sqrt{\pi})$ over $\mathbb{Q}(\pi)$.

Exercise 3.7.28. Prove that π is algebraic over $\mathbb{Q}(\pi^4)$; then, find the degree of $\mathbb{Q}(\pi)$ over $\mathbb{Q}(\pi^4)$.

Exercise 3.7.29. Prove that if α is a positive real transcendental number, then $\alpha^{p/q}$ is not algebraic over $\mathbb{Q}(\alpha)$ for any nonzero relatively prime integers p and q.

Exercise 3.7.30. Consider any element α of an extension field F of any field k.

(a.) Prove that if α is transcendental over k, then $k(\alpha)$ is an infinite-dimensional k-vector space.

(b.) Prove that if α is transcendental over k, then α^n is transcendental over k for all $n \in \mathbb{Z} \setminus \{0\}$.

(c.) Prove that if α is transcendental over k, then for any elements $a_0, a_1, \ldots, a_n$ of k (not all zero), the element $a_n\alpha^n + \cdots + a_1\alpha + a_0$ of $k(\alpha)$ admits a unique multiplicative inverse in $k(\alpha)$.

(d.) Prove that if α is transcendental over k, then for any element β of F such that α is algebraic over $k(\beta)$, it must also hold that β is algebraic over $k(\alpha)$.

(**Hint:** By definition, we have that α is algebraic over $k(\beta)$ if and only if there exist polynomials $p_0(x), p_1(x), \ldots, p_n(x)$ in $k[x]$ such that $p_n(\beta)\alpha^n + \cdots + p_1(\beta)\alpha + p_0(\beta) = 0_k$. Prove that if α is transcendental over k, then some polynomial $p_i(x)$ must be non-constant. Observe that $f(x,y) = p_n(x)y^n + \cdots + p_1(x)y + p_0(x) = q_m(y)x^m + \cdots + q_1(y)x + q_0(y)$ for some polynomials $q_0(y), q_1(y), \ldots, q_m(y)$ in $k[y]$ at least one of which is non-constant and $f(\beta, \alpha) = 0_k$.)

Exercise 3.7.31. Prove that $\mathbb{C}$ is an algebraic extension of $\mathbb{R}$.

(**Hint:** Given any complex number $a + bi$, construct a polynomial with roots $a + bi$ and $a - bi$.)

Exercise 3.7.32. Prove that if k is a finite field of prime order p and α is any algebraic element of any extension field F of k, then $k(\alpha)$ is a finite field of order p^n for some integer $n \geq 1$.

Exercise 3.7.33. Consider any field k.

(a.) Prove that if $\{F_i\}_{i \in I}$ is any nonempty collection of fields indexed by I such that $F_i \subseteq k$ for each index i, then $\cap_{i \in I} F_i$ is the smallest (with respect to inclusion) field contained in k.

(b.) Prove that if $\{F_i\}_{i \in I}$ is any nonempty collection of fields indexed by I such that these fields form an ascending chain $k \subseteq F_1 \subseteq F_2 \subseteq \cdots$, then $\cup_{i \in I} F_i$ is a field that contains k.

(c.) Prove that if $\{F_i\}_{i \in I}$ is any nonempty collection of algebraic extensions of k indexed by I such that these fields form an ascending chain $k \subseteq F_1 \subseteq F_2 \subseteq \cdots$, then $\cup_{i \in I} F_i$ is algebraic over k.

Exercise 3.7.34. Prove or disprove that each of the following pairs of finite extensions are equal.

(a.) $\mathbb{Q}(3+i)$ and $\mathbb{Q}(1-i)$

(b.) $\mathbb{Q}(\sqrt{20})$ and $\mathbb{Q}(\sqrt{5})$

(c.) $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$

(d.) $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$ and $\mathbb{Q}(\sqrt[6]{2})$

(e.) $\mathbb{Q}(\sqrt{2}, \sqrt{8})$ and $\mathbb{Q}(\sqrt[3]{8})$

(f.) $\mathbb{Q}(\sqrt[4]{3})$ and $\mathbb{Q}(i\sqrt[4]{3})$

Exercise 3.7.35. Consider any field k of characteristic other than 2. Given any elements a and b in k, let $\sqrt{a}$ denote a root of the monic polynomial $x^2 - a$ in $k[x]$. Prove that $k(\sqrt{a}+\sqrt{b}) = k(\sqrt{a}, \sqrt{b})$.

Exercise 3.7.36. Prove that if k is an extension field over $\mathbb{Q}$ such that $[k : \mathbb{Q}] = 2$, then $k = \mathbb{Q}(\sqrt{a})$ for some integer a that is not divisible by the square of any prime number (i.e., a is **square-free**).

Exercise 3.7.37. Prove that if α is any algebraic element of any extension field F of any field k such that the dimension $[k(\alpha) : k]$ is odd, then we must have that $k(\alpha) = k(\alpha^2)$.

Exercise 3.7.38. Prove that if F is any finite extension over some field k such that the dimension $[F : k]$ of F as a k-vector space is prime, then F must be a simple extension of k. Explicitly, prove that for every element α of F that does not lie in k, we have that $F = k(\alpha)$.

Exercise 3.7.39. Consider any algebraic elements α and β of any extension field F of any field k.

(a.) Prove that if the degrees of $\mu_\alpha(x)$ and $\mu_\beta(x)$ over k are relatively prime, then it holds that

$$[k(\alpha, \beta) : k] = [k(\alpha) : k][k(\beta) : k].$$

(b.) Give an explicit example of a field k and a pair of algebraic elements α and β over k for which the positive integers $[k(\alpha, \beta) : k]$ and $[k(\alpha) : k][k(\beta) : k]$ are not equal.

(c.) Use the criterion of the first part above to find $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$ without constructing a tower.

Exercise 3.7.40. Consider any pair of monic irreducible polynomials $p(x)$ and $q(x)$ in the univariate polynomial ring $k[x]$ over any field k. Prove that if the degrees of $p(x)$ and $q(x)$ are relatively prime, then for any root α of $p(x)$ in any extension field F of k, we have that $q(x)$ is irreducible over $k(\alpha)$.

(**Hint:** Given any root β of $q(x)$ in any extension field E of k, compute and compare the k-vector space dimension of $k(\alpha, \beta)$ over k in two different ways using Proposition 3.5.11 and Exercise 3.7.39.)

Exercise 3.7.41. 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.

(c.) Prove that the order of the finite field k is p^n for some integer $n \geq 1$.

(**Hint:** Begin by demonstrating that k contains an isomorphic copy of the finite field $\mathbb{Z}/p\mathbb{Z}$, i.e., prove that there exists an injective unital ring homomorphism $\varphi : \mathbb{Z}/p\mathbb{Z} \to k$. Consequently, we may view k as an extension field of $\mathbb{Z}/p\mathbb{Z}$, hence k is a $(\mathbb{Z}/p\mathbb{Z})$-vector space. Refer to the previous part of this problem to deduce that k is a finite extension of $\mathbb{Z}/p\mathbb{Z}$, hence k admits a basis $\{\alpha_1, \ldots, \alpha_n\}$ over $\mathbb{Z}/p\mathbb{Z}$. Use this to determine the form of any element of k.)

(d.) Conclude that every finite field has order p^n for some prime number p and some integer $n \geq 1$.

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/.

[Dai19] R.C. Daileda. *GCDs and Gauss' Lemma*. 2019. URL: http://ramanujan.math.trinity.edu/rdaileda/teach/f19/m4363/gauss_lemma.pdf.

[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.

[kjo] kjo. *Prove that* $\gcd(M,N) \times \operatorname{lcm}(M,N) = M \times N$. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/645590 (version: 2014-01-20). URL: https://math.stackexchange.com/q/645590.

[Mag11] A. Magidin. *Why are two permutations conjugate iff they have the same cycle structure?* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/48137 (version: 2011-06-28). 2011. URL: https://math.stackexchange.com/a/48137/390180.