MA383: Introduction to Modern Algebra

Dylan C. Beck

Contents

Chapter 1

Basic Properties of Sets and Logic

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

1.1 Sets and Set Operations

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

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

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

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

Like with the arithmetic of real numbers, there are mathematical operations on sets that allow us, e.g., to compare them; take their differences; and combine them. For instance, every element of $Y = \{1, 2, 3, 4, 5\}$ is also an element of $X = \{1, 2, 3, 4, 5, 6\}$, but the element 6 of X is not contained

in Y. We express this by saying that Y is a **proper subset** of X: the additional modifier "proper" is used to indicate that X and Y are not the same set (because they do not have the same members). Put into symbols, we write that $Y \subsetneq X$ whenever it is true that (i.) every element of Y is also an element of X and (ii.) there exists an element of X that is not contained in Y; this can be read as "Y is contained in X, but Y does not equal X." We may also say that Y is "included in" X. One other way to indicate that Y is a (proper) subset of X is by saying that X is a (proper) **superset** of Y, in which case we write that $X \supseteq Y$ (or $X \supsetneq Y$). If we could step through the paper and look at the superset containment $X \supseteq Y$ from the other side, it would be nothing more than $Y \subseteq X$.

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

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

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

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

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

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

Each time we define a function $f : X \to Y$, in addition, we implicitly distinguish the collection of elements $y \in Y$ such that $y = f(x)$ for some element $x \in X$; this is called the **image** of X (in Y) with respect to f, and it is denoted by $f(X) = \{y \in Y \mid y = f(x) \text{ for some element } x \in X\}$. Conversely, if W is a subset of Y, then the collection of elements $x \in X$ such that $f(x) \in W$ is the **pre-image** of W (in X) with respect to f. Explicitly, we have that $f^{-1}(W) = \{x \in X \mid f(x) \in W\}$. We note that for any subsets $V \subseteq X$ and $W \subseteq Y$, it is always the case that $V \subseteq f^{-1}(f(V))$ and $f(f^{-1}(W)) \subseteq W$; however, the superset containments $V \supseteq f^{-1}(f(V))$ and $f(f^{-1}(W)) \supseteq W$ do not always hold (cf. Exercise 1.10.6). We introduce two properties of functions that are sufficient to guarantee that these superset inclusions. If $f(x_1) = f(x_2)$ implies that $x_1 = x_2$ for any pair of elements $x_1, x_2 \in X$, then we say that $f : X \to Y$ is **injective**. Essentially, a function $f : X \to Y$ is injective if and only if distinct elements $x_1, x_2 \in X$ induce distinct elements $f(x_1), f(x_2) \in Y$. We will soon verify this formally. Even more, we say that $f : X \to Y$ is **surjective** if for every element $y \in Y$, there exists an element $x \in X$ such that $y = f(x)$. One way to think about the surjective property is that every element of Y is "mapped onto" or "covered" by an element of X. If $f : X \to Y$ is both injective and surjective, then we say that f is **bijective**. We may think about a bijection $f : X \to Y$ as a relabelling of the elements of Y using the names of elements of X; in this way, two sets X and Y are "essentially the same" if there exists a bijection $f : X \to Y$.

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

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

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

Proof. 1.) By Exercise 1.10.6, it suffices to prove that $f^{-1}(f(V)) \subseteq V$. Let x be an arbitrary element of $f^{-1}(f(V))$. By definition of the pre-image $f^{-1}(f(V))$ of $f(V)$, this means that $f(x) \in f(V)$. By definition of the image $f(V)$, we have that $f(x) = f(v)$ for some element $v \in V$. Last, by assumption that f is injective and $V \subseteq X$, we conclude that $x = v$, hence x is an element of V.

(2.) By Exercise 1.10.6, it suffices to prove that $W \subseteq f(f^{-1}(W))$. Let w be any element of W. By assumption that f is surjective and $W \subseteq Y$, there exists an element $x \in X$ such that $w = f(x)$. By definition of the pre-image $f^{-1}(W)$, it follows that $x \in f^{-1}(W)$. By definition of the image $f(f^{-1}(W))$, we conclude that $w = f(x)$ for some element $x \in f^{-1}(W)$ so that $w \in f(f^{-1}(W))$. $\square$

Conversely, if $f^{-1}(f(V)) = V$ holds for any set $V \subseteq X$, then $f : X \to Y$ must be injective; likewise, if $f(f^{-1}(W)) = W$ for any set $W \subseteq Y$, then f must be surjective (cf. Exercise 1.10.7).

1.2 Logic and Truth Tables

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

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

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

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

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

Unfortunately, in some situations, it is difficult to establish the verity of a statement Q from a statement P that is known to be true. Under these circumstances, it is not possible to determine

if the statement $P \implies Q$ is true or false because this depends entirely on whether Q is true or false; however, it is possible in some cases to extract a statement $S(P, Q)$ (depending upon P and Q) that is **logically equivalent** to the implication $P \implies Q$. We say that two statements S and S' are logically equivalent if and only if their values in a truth table are equal. Consequently, if we could demonstrate that the statement $S(P, Q)$ were true, then $P \implies Q$ must be true, as well.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Last, the **proof by contradiction** (or *reductio ad absurdum*) rounds out the tools that we will most often use in mathematical proofs. Essentially, the proof by contradiction constitutes a valid law of inference by a combination of the Law of the Excluded Middle (which asserts that either a statement or its negation must be true), the Law of Non-Contradiction (which asserts that a statement and its negation cannot both be true), and Table 1.3 (which asserts that $\neg(P \implies Q)$ and $P \wedge \neg Q$ are logically equivalent): one of the statements $P \implies Q$ or $\neg(P \implies Q)$ must be true, hence if we can establish that $P \wedge \neg Q$ is not true, then it must be the case that $\neg(P \implies Q)$ is not true (because these two statements are logically equivalent) so that $P \implies Q$ is true. For instance,

if we define an even number to be a whole number that is divisible by two, then we may appeal to a proof by contradiction to establish that twice any whole number is also even. Explicitly, suppose that P is the statement that "x is a whole number" and Q is the statement that "$2x$ is an even number." If we wish to establish the verity of the implication $P \implies Q$ given by the statement, "If x is a whole number, then $2x$ is an even number," then we may assume to the contrary that P is true (i.e., x is a whole number) and $\neg Q$ is also true (i.e., $2x$ is not an even number); all together, we are assuming $P \wedge \neg Q$, i.e., "x is a whole number and $2x$ is not an even number." By definition, $2x$ is an even number because it is twice a whole number, so we have arrived at a contradiction — namely, that $2x$ is an even number (by definition) and $2x$ is not an even number (by assumption). Ultimately, the statement $P \wedge \neg Q$ cannot be true, hence $P \implies Q$ must be true. Generally, a successful proof by contradiction begins by assuming (to the contrary) that P is true and that Q is not true; then, a contradiction of the form (a.) $P \wedge \neg P$ or (b.) $Q \wedge \neg Q$ is derived. Observe that if $\neg P$ can be deduced from $\neg Q$ (i.e., (a.) holds) then a proof by contraposition may be simpler than a proof by contradiction; on the other hand, if Q can be deduced from P (i.e., (b.) holds), then a **direct proof** may be simpler than a proof by contradiction. Bear this in mind always.

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

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

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

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

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

Even more, **logical quantifiers** allow us to symbolically express the concepts of "for all" (or "for every") and "there exists" (or "for at least one" or "for some"). Explicitly, we adopt the **universal quantifier** $\forall$ as the symbolic representation of the phrase "for all" and the **existential quantifier** $\exists$ as the symbolic representation of the phrase "there exists." Using these quantifiers,

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

1.3 Sets and Set Operations, Revisited

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

- We may view the set membership $x \in X$ as the statement "x is an element of X"; its negation is the statement that "x is not an element of X" (or $x \notin X$ in symbols).
- We have that $X \subseteq Y$ ("X is a subset of Y") if and only if for every element $x \in X$, it is true that $x \in Y$, i.e., if and only if it is true that $(\forall x \in X)(x \in Y)$. Consequently, the empty set $\emptyset$ is a subset of every set: there are no elements in $\emptyset$, hence $(\forall e \in \emptyset)(e \in X)$ is vacuously true!
- If it holds that $X \subseteq Y$ and $(\exists y \in Y)(y \notin X)$ ("there exists an element $y \in Y$ such that $y \notin X$"), then we say that X is a proper subset of Y, and we write $X \subsetneq Y$; otherwise, it must be the case that $Y \subseteq X$, hence X and Y are equal, i.e., we must have that $X = Y$.
- Elements of either X or Y comprise the union $X \cup Y$ of X and Y. Put another way, we have that $X \cup Y$ is the superset of both X and Y for which $(w \in X) \vee (w \in Y)$ is true.
- Elements of both X and Y comprise the intersection $X \cap Y$ of X and Y. Put another way, we have that $X \cap Y$ is the subset of both X and Y for which $(w \in X) \wedge (w \in Y)$ is true.
- Elements in Y but not in X comprise the relative complement $Y \setminus X$ of X with respect to Y. Put another way, we have that $Y \setminus X$ is the subset of Y for which $(y \in Y) \wedge (y \notin X)$ is true.
- We may view the Cartesian product $X \times Y$ of X and Y as the collection of all ordered pairs (x, y) for which the statement $(x \in X) \wedge (y \in Y)$ is true.

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

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

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

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

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

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

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

We leave the proofs of De Morgan's Laws for Sets as Exercise 1.10.13.

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

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

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

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

Observe that $w \in \cap_{i=1}^{n} X_i$ if and only if $(\forall i \in \{1, 2, \ldots, n\})(w \in X_i)$. Generally, the following extension of De Morgan's Laws for Sets holds; its proof is left as Exercise 1.10.14.

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

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 $W \setminus (X_1 \cap X_2 \cap \cdots \cap X_n) = (W \setminus X_1) \cup (W \setminus X_2) \cup \cdots \cup (W \setminus X_n)$.*

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

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

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

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

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

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

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

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

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

1.4 Equivalence Relations and Partial Orders

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

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

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

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

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

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

Example 1.4.3. Consider the equivalence relation E defined on the set $\mathcal{C}^1(\mathbb{R})$ of Example 1.4.2. By the Fundamental Theorem of Calculus, if $f'(x) = g'(x)$, then there exists a real number C such

that $f(x) = g(x) + C$. Conversely, if $f(x) = g(x) + C$ for some real number C, then $f'(x) = g'(x)$. We conclude that the equivalence classes of $\mathcal{C}^1(\mathbb{R})$ modulo E are given precisely by the sets

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

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

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

Proof. Consider any pair $[x_1]$ and $[x_2]$ of equivalence classes of X modulo E. By Exercise 1.10.9, it suffices to prove that $[x_1] = [x_2]$ if they are not disjoint. Consequently, we may assume that there exists an element $x \in [x_1] \cap [x_2]$. By definition, this means that $(x, x_1) \in E$ and $(x, x_2) \in E$. By assumption that E is an equivalence relation, it follows that $(x_1, x) \in E$ by symmetry, hence the transitivity of E implies that $(x_1, x_2) \in E$. Given any element $x_0 \in [x_1]$, we have that $(x_0, x_1) \in E$ implies that $(x_0, x_2) \in E$ by transitivity, hence we conclude that $[x_1] \subseteq [x_2]$. Likewise, the symmetry of E implies that $(x_2, x_1) \in E$, hence the same argument as the previous sentences shows that $[x_2] \subseteq [x_1]$. Combined, the containments $[x_1] \subseteq [x_2]$ and $[x_2] \subseteq [x_1]$ yields that $[x_1] = [x_2]$. □

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

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

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

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

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

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

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

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

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

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

1.5 The Principle of Mathematical Induction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(a.) $0 \in S$ and

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

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

By the Principle of Ordinary Induction, we can return to prove Conjectures 1.5.2 and 1.5.4; we leave these as the respective Exercises 1.10.20 and 1.10.21 for the reader. Occasionally, it is desirable to strengthen the hypotheses of the Principle of Ordinary Induction in order to simplify proofs involving induction. Currently, we may view induction as a property of falling dominoes: (a.) if the 0th domino falls and (b.) the nth domino falling causes the $(n+1)$th domino to fall, then all dominoes indexed by the non-negative integers will fall. But suppose that we could knock down all dominoes from the first to the nth domino: this would provide even more power with which to knock down the $(n+1)$th domino! We introduce this as the following.

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

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

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

Even though the hypotheses of the Principle of Complete Induction appear to be stronger than the Principle of Ordinary Induction, the two are in fact equivalent to one another (cf. Exercise 1.10.24); together, they are the Principle of Mathematical Induction. Using complete induction, we may obtain another ubiquitous mathematical tool that will prove crucial in our future endeavors.

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

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

Conversely, the Well-Ordering Principle implies the Principle of Ordinary Induction, hence it is equivalent to both ordinary induction and complete induction (cf. Exercise 1.10.25).

1.6 The Division Algorithm

Even as early as grade school, we learn the process of dividing one integer by another. Each time we divide an integer a by a nonzero integer b, we obtain an integer q and a non-negative integer r that is strictly smaller than $|b|$ such that $a = qb + r$. Explicitly, we say that a is the **dividend**; b is the **divisor**; q is the **quotient**; and r is the **remainder** of the division. Our aim throughout this

section is to establish that this process is well-founded, i.e., the process of division of an integer a by a nonzero integer b always results in *unique* integers q and r such that $a = qb+r$ and $0 \leq r < |b|$. We will also establish an algorithm that will allow us to efficiently find the integers q and r.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Generally, for any nonzero integers a and b, we may determine $\gcd(a,b)$ from the prime factorizations of a and b in the same manner as Example 1.6.8 (cf. Exercise 1.10.31).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 1.6.13. Consider the case that $a = 110$ and $b = 24$. Observe that the unique prime factorizations of 110 and 15 are $110 = 10 \cdot 11 = 2 \cdot 5 \cdot 11$ and $24 = 2^3 \cdot 3$, respectively. By Exercise 1.10.31, it follows that $\gcd(110, 15) = 2$. By successively implementing the Division Algorithm, we may find the integers x and y such that $110x + 24y = 2$, as guaranteed to us by Bézout's Identity. Explicitly, we begin by running the Division Algorithm with $a = 110$ and $b = 24$ to find the unique integers q_1 and $0 \leq r_1 < 24$ such that $110 = 24q_1 + r_1$; then, we run the Division Algorithm with 24 and r_1 to produce the unique integers q_2 and $0 \leq r_2 < r_1$ such that $24 = q_2 r_1 + r_2$. Continuing in this manner produces a strictly decreasing sequence $r_1 > r_2 > \cdots > r_n$ of non-negative integers at the nth step; by the Well-Ordering Principle, this sequence must have a least element, hence the process must eventually terminate. Putting this process to the test, we find that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.7 The Integers Modulo n

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

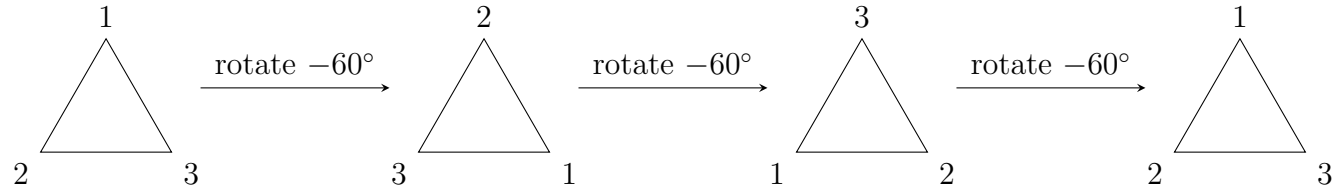
1.8 Rigid Motions

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

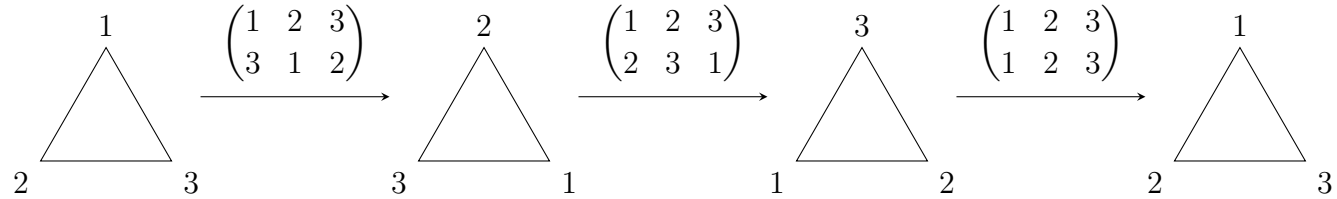
Rigid motions of polygons are those operations that we can perform on polygons without altering the distance between any two vertices of the polygon. For instance, if we have a square in the plane, then we may shift each of the vertices of the square any distance north, south, east, or

west without disturbing the distances between any of the vertices of the square; however, we cannot move just one vertex any nonzero distance north, south, east, or west without altering its distance from another vertex. Put another way, **translation** of a polygon is a rigid motion.

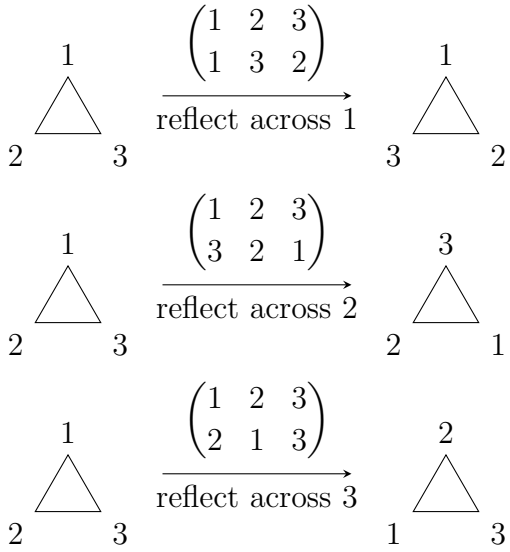
We will fix our attention throughout this section on two specific rigid motions of any regular n-gon. Considering that the sum of the interior angles of a regular n-gon is $(n-2)180$ degrees, it follows that every interior angle of a regular n-gon has a measure of $(n-2)180/n$. Consequently, for any integer $1 \leq k \leq n$, a **rotation** of a regular n-gon through an angle of $(2-n)180k/n$ 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 interior 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 $(2-n)180k/n$ 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 j is a permutation of the vertices of the regular n-gon that fixes the vertex j 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 integer $\{1, 2, 3\}$, hence every sequence of rotations and reflections yields a rotation or a reflection. We will return to this concept soon in our discussion of groups.

1.9 Chapter 1 Overview

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

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

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

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

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

We say that a complete sentence P is a **statement** if it asserts something that can be unambiguously measured as true or false. Examples of statements include "3 is an odd number" and "17 is larger than 38"; the first statement is true, but the second statement is false. Using logical connectives, we can form new statements from given statements P and Q. Explicitly, the **implication** $P \implies Q$ is the statement that "P implies Q" (or equivalently, "If P, then Q"); the implication

is false if and only if P is true and Q is false. If P is false, then $P \implies Q$ is called a **vacuous truth**. We define the **disjunction** $P \vee Q$ ("P or Q"), the **conjunction** $P \wedge Q$ ("P and Q"), and the **negation** $\neg P$ ("not P"). Observe that the disjunction $P \vee Q$ is true if and only if P is true or Q is true; the conjunction $P \vee Q$ is true if and only if P is true and Q is true; and the negation $\neg P$ takes the opposite truth-value of P. The Law of the Excluded Middle asserts that either P or $\neg P$ must be true, and the Law of Non-Contradiction asserts that P and $\neg P$ cannot both be true.

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

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

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

Using logical quantifiers allows us to conveniently state many properties of sets, e.g., the Law of the Excluded Middle for Sets, Law of Non-Contradiction for Sets, and De Morgan's Laws for Sets. Even more, logical quantifiers enable us to extend De Morgan's Laws for Sets to arbitrary unions and arbitrary intersections of sets. Explicitly, we may consider an arbitrary set I as an **index set** for an arbitrary family of sets $\{X_i \mid i \in I\}$ such that each set X_i is a subset of some set W called our **universe**. By definition, the arbitrary union of these sets is simply $\cup_{i \in I} X_i$; membership of an

element $w \in W$ in this arbitrary union is characterized by $w \in \cup_{i \in I} X_i$ if and only if $w \in X_i$ for some index $i \in I$. Likewise, the arbitrary intersection of these sets is $\cap_{i \in I} X_i$ with membership of an element $w \in W$ characterized by $w \in \cap_{i \in I} X_i$ if and only if $w \in X_i$ for all indices $i \in I$. We say that two sets X_i and X_j are **disjoint** if $X_i \cap X_j = \emptyset$; moreover, if $X_i \cap X_j = \emptyset$ for all distinct indices $i, j \in I$, then we say that the sets in $\{X_i \mid i \in I\}$ are **pairwise disjoint** or **mutually exclusive**. We say that the collection $\mathcal{P} = \{X_i \mid i \in I\}$ forms a **partition** of the set W if and only if

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

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

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

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

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

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

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

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

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

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

then $P(n)$ is a true statement for all non-negative integers n. One of the benefits of using complete induction is that its stronger hypotheses allow us more information with which to conveniently write proofs that might otherwise be awkward with ordinary induction (cf. Exercise 1.10.23). Even more, the Principle of Mathematical Induction appears also in the guise of the Well-Ordering Principle for the non-negative integers; this powerful tool guarantees that every nonempty set of non-negative integers admits a smallest element with respect to the total order $\leq$. Put another way, if $S \subseteq \mathbb{Z}_{\geq 0}$ is a nonempty set, then there exists an element $s_0 \in S$ such that $s_0 \leq s$ for all elements $s \in S$.

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

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

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

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

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

Polygons are two-dimensional, closed, and bounded objects determined by the intersection of finitely many straight line segments in the plane; the intersection points are called **vertices**. Examples of polygons include triangles, rectangles, and stars. **Regular** polygons have the additional property that their sides possess equal length and each angle at a vertex has equal measure. Equilateral triangles and squares are regular, but most triangles and rectangles are not regular. Generally,

an n-**gon** is any polygon with n vertices. **Rigid motions** of a polygon are those operations that can be performed on the polygon without altering the distance between any two of its vertices. Regular n-gons have the property that **rotation** by an angle of $(2-n)180k/n$ degrees is a rigid motion for each integer $1 \leq k \leq n$. Likewise, **reflection** of a regular n-gon across any one of its n vertices also constitutes a rigid motion of the regular n-gon. Combined, rotations and reflections of a regular n-gon can be performed in any order to produce another rotation or reflection.

1.10 Chapter 1 Exercises

1.10.1 Sets and Set Operations

Exercise 1.10.1. Consider the sets

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

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

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

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

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

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

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

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

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

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

(b.) Prove that for any subset $Z \subseteq W$ such that $Z \subseteq X$ and $Z \subseteq Y$, it follows that $Z \subseteq X \cap Y$. Conclude that $I = X \cap Y$ is the "largest" subset of W with 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 but that is not contained in X.

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

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

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

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

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

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

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

Exercise 1.10.6. Let $f : X \to Y$ be any function between any two sets X and Y.

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

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

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

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

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

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

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

1.) 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 $V = \{x_1\}$.)

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

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

1.10.2 Logic and Truth Tables

Exercise 1.10.8. Let P be the statement that "The sun is shining in Kansas City." Let Q be the statement that "Bob rides his bike to work." Use the letters P and Q and logical connectives such as $\implies$, $\iff$, $\vee$, $\wedge$, and $\neg$ to convert each of the following statements into symbols; then, identify all of the logically equivalent statements, tautologies, and self-contradictions.

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

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

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

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

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

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

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

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

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

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

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

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

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

Exercise 1.10.9. Let P, Q, and R be any statements. Construct a truth table to prove that the statements "If P, then Q or R" and "If P and not Q, then R" are logically equivalent.

Exercise 1.10.10. Let P and Q be any statements. Construct a truth table to prove that the statement "If P or Q but not Q, then P" is a tautology.

Exercise 1.10.11. Use Exercise 1.10.10 to prove that if Bob placed in the top two in a cycling race on Saturday, but he did not place second, then Bob must have placed first.

Exercise 1.10.12. Use a proof by contradiction to prove that if Bob placed in the top two in a cycling race on Saturday, but he did not place second, then Bob must have placed first. Cite any theorems or laws of inference (by name) that you use in your proof.

1.10.3 Sets and Set Operations, Revisited

Exercise 1.10.13. (De Morgan's Laws for Sets) Let $X \subseteq W$ and $Y \subseteq W$ be arbitrary sets.

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

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

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

Exercise 1.10.14. Let $X_1, X_2, \ldots, X_n \subseteq W$ be arbitrary sets.

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

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

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

Exercise 1.10.15. Let $\mathbb{Z}$ denote the set of integers.

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

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

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

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

1.10.4 Equivalence Relations and Partial Orders

Exercise 1.10.16. Consider the set W consisting of all words in the English language.

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

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

Exercise 1.10.17. Let $\mathbb{Z}$ be the set of integers. Prove that the relation $(a, b) \sim (c, d)$ if and only if $ad = bc$ on $\mathbb{Z} \times \mathbb{Z}$ is an equivalence relation. Describe the collection of distinct equivalence classes.

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

Exercise 1.10.18. Let X be an arbitrary set. Consider the collection $S = \{Y \mid Y \subseteq X\}$. Prove that the inclusion $\subseteq$ defines a partial order P on S such that $(Y_1, Y_2) \in P$ if and only if $Y_1 \subseteq Y_2$; then, either prove that P is a total order on S, or provide a counterexample to show that it is not.

Exercise 1.10.19. List the maximal elements of the subset $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$ of the set $\mathbb{Z}_{\geq 0}$ of non-negative integers with respect to the partial order D of divisibility.

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

1.10.5 The Principle of Mathematical Induction

Exercise 1.10.20. Prove Conjecture 1.5.2 using the Principle of Ordinary Induction.

Exercise 1.10.21. Prove Conjecture 1.5.4 using the Principle of Ordinary Induction.

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

Exercise 1.10.22. Let X be an arbitrary finite set with power set $P(X)$.

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

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

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

Exercise 1.10.23. Let F_n denote the nth Fibonacci number.

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

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

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

Exercise 1.10.24. Prove that the Principle of Ordinary Induction and the Principle of Complete Induction are equivalent to one another by completing the following two steps.

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

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

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

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

Exercise 1.10.25. Prove that the Well-Ordering Principle and the Principle of Ordinary Induction are equivalent to one another by completing the following three steps.

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

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

(3.) Conclude that the Well-Ordering Principle implies the Principle of Ordinary Induction; then, use Exercise 1.10.24 and the proof of the Well-Ordering Principle to conclude that the Principle of Ordinary Induction implies the Well-Ordering Principle.

1.10.6 The Division Algorithm

Exercise 1.10.26. 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 1.10.27. Let a, b, and c be any integers. Prove that if $\gcd(a, b) = 1$ and $a \mid bc$, then $a \mid c$.

Exercise 1.10.28 (Fundamental Theorem of Arithmetic)**.** Let a be a positive integer. 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 just a rearrangement of $p_1, \ldots, p_k$).

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

Exercise 1.10.29. Let a be any positive integer. Prove that there exist distinct prime numbers $p_1, \ldots, p_n$ and unique non-negative integers $e_1, \ldots, e_n$ such that $a = p_1^{e_1} \ldots p_n^{e_n}$.

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

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

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

Exercise 1.10.30. Prove that the $\operatorname{lcm}(a, b)$ is unique.

By the Fundamental Theorem of Arithmetic, for any positive integers a and b, there exist prime numbers $p_1, \ldots, p_k$ and unique integers $e_1, \ldots, e_k, f_1, \ldots, f_k$ such that $a = p_1^{e_1} \cdots p_k^{e_k}$ and $b = p_1^{f_1} \cdots p_k^{f_k}$.

Exercise 1.10.31. Prove that $\gcd(a, b) = p_1^{\min\{e_1, f_1\}} \cdots p_k^{\min\{e_k, f_k\}}$.

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

Exercise 1.10.33. Conclude from Exercises 1.10.31 and 1.10.32 that $ab = \gcd(a, b) \operatorname{lcm}(a, b)$.

1.10.7 The Integers Modulo n

Exercise 1.10.34. Complete the following using modular arithmetic.

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

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

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

Exercise 1.10.35. Consider the collection $\mathbb{Z}_n$ of equivalence classes of the integers modulo n. If $ab \equiv 0 \pmod{n}$, then must it be true that $a \equiv 0 \pmod{n}$ or $b \equiv 0 \pmod{n}$? Explain.

Exercise 1.10.36. Let p be any prime number. Consider the collection $\mathbb{Z}_p$ of equivalence classes of the integers modulo p. Prove that if $ab \equiv 0 \pmod{p}$, then $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$.

Exercise 1.10.37. Let p be any prime number. Consider the collection $\mathbb{Z}_p$ of equivalence classes of the integers modulo p. Prove that $[a]$ admits a multiplicative inverse if and only if $p \nmid a$.

1.10.8 Rigid Motions

Exercise 1.10.38. Prove that for a regular n-gon, there are at most $n! = n(n-1)(n-2)\cdots 2 \cdot 1$ symmetries corresponding to rotation through an angle or reflection across a vertex.

Exercise 1.10.39. List all permutations of the integers $\{1, 2, 3, 4\}$ corresponding to the rotations and reflections of a regular 4-gon. Conclude that the upper bound of Exercise 1.10.38 can be strict.

Exercise 1.10.40. Explicitly compute the number of symmetries of a regular n-gon corresponding to rotation through and angle or reflection across a vertex; then, prove that your formula holds.

(**Hint:** Use the example of Section 1.8, your work from Exercise 1.10.39, and possibly an additional example to spot the pattern and deduce a formula; then, use induction to prove the formula holds.)

Exercise 1.10.41. Consider the regular 3-gon T with vertices 1, 2, and 3. Let ρ_k denote rotation of T through an angle of $-60k$ degrees. Explicitly, there are three distinct rotations ρ_1, ρ_2, and ρ_3. Let ϕ_k denote the reflection of T across the vertex k. Explicitly, there are three distinct reflections ϕ_1, ϕ_2, and ϕ_3. If x and y are elements of $\{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$, then let xy denote the operation of first performing x and subsequently performing y. Explicitly, $\rho_i\phi_j$ is the operation of first rotating through an angle of $-60i$ degrees and then reflecting about the vertex j of the original arrangement of T. Complete the table below by computing xy according to the rows x and columns y.

$x \backslash y$	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			ϕ_1		
ϕ_2						
ϕ_3						

Chapter 2

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

2.1 Groups (Definitions and Examples)

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

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

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

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

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

Example 2.1.2. Consider the collection $\mathbb{Z}_n$ of equivalence classes of integers modulo n. By Proposition 1.7.1, the distinct elements of $\mathbb{Z}_n$ are given by $r \pmod n$ for each integer $0 \leq r \leq n - 1$, hence it is nonempty. Using modular arithmetic, we may define an associative binary operation $+_n$ on $\mathbb{Z}_n$. Explicitly, we accomplish this by setting $r_1 \pmod n +_n r_2 \pmod n = (r_1 + r_2) \pmod n$. Of course, we may reduce $r_1 + r_2$ modulo n by computing the least non-negative integer x for which $r_1 + r_2 \equiv x \pmod n$; then, we may view $(r_1 + r_2) \pmod n$ as $x \pmod n$, hence $+_n$ is a binary operation on $\mathbb{Z}_n$. Considering that addition of integers is associative, $+_n$ is associative; the identity

element of $\mathbb{Z}_n$ is simply 0 (mod n); and if $1 \leq r \leq n-1$, then the inverse of r (mod n) is simply $(n-r)$ (mod n). Ultimately, this goes to show that $(\mathbb{Z}_n, +_n)$ is a group. Once again, observe that we have used additive notation in place of the multiplicative notation of arbitrary groups.

Example 2.1.3. Let T denote any regular 3-gon. Let ρ_k denote rotation through an angle of $-60k$ degrees for each integer $1 \leq k \leq 3$. Let ϕ_k denote reflection across the vertex k for each integer $1 \leq k \leq 3$. Consider the collection $D_3 = \{\rho_1, \rho_2, \rho_3, \phi_1, \phi_2, \phi_3\}$ of the symmetries of a regular 3-gon obtained by rotation or reflection. By Exercise 1.10.41, for any pair of elements $x, y \in D_3$, the assignment $x \circ y = xy$ is a binary operation on D_3. Certainly, it is associative because we have that $(x \circ y) \circ z = xyz = x \circ (y \circ z)$ for any elements $x, y, z \in D_3$. 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 conclude that $(D_3, \circ)$ is a group. We refer to this multiplicative group as the **dihedral group** of order $6 = 2 \cdot 3$.

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

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

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

Consequently, the group $(\mathbb{Z}_n, +_n)$ is abelian, as well. By Exercise 1.10.41, on the other hand, the group $(D_3, \circ)$ of Example 2.1.3 is not abelian because we have that $\rho_1\phi_1 = \phi_3 \neq \phi_2 = \phi_1\rho_1$.

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

Even more, let $\mathrm{GL}(n, \mathbb{R})$ denote the subset of $\mathbb{R}^{n \times n}$ consisting of invertible $n \times n$ 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 $I_{n \times n}$, and the multiplicative inverse of an invertible $n \times n$ matrix A is A^{-1}. Considering that matrix multiplication is not commutative, $(\mathrm{GL}(n, \mathbb{R}), \cdot)$ 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 defining a group as the **order** of the group. Observe that the additive group $(\mathbb{Z}_n, +_n)$ of the integers modulo n has order $|\mathbb{Z}_n| = n$, and the dihedral group $(D_3, \circ)$ of order six has order $|D_3| = 6$. On the other hand, the additive groups $(\mathbb{Z}, +)$ of integers and $(\mathbb{R}^{n \times n}, +)$ of real $n \times n$ matrices and the multiplicative group $(\mathrm{GL}(n, \mathbb{R}), \cdot)$ of real invertible $n \times n$ matrices have infinitely many elements, hence they each possess infinite order.

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

2.2 Groups (Basic Properties and Subgroups)

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

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

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

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

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

Proof. We must show that if e is any element of G satisfying 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 satisfying that $g * h = e_G = h * g$, then $h = g^{-1}$. Considering that $*$ is associative and $g^{-1} * g = e_G$, it follows that

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

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

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

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

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. Usually, we will omit the operation $*$ of G and simply use concatenation, e.g., we write $g_1 * g_2$ as $g_1 g_2$ Consequently, by property (2.) above, it must have a multiplicative inverse.

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

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

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

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

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

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

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

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

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

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

We leave the proofs of the Group Exponent Laws as Exercise **??**.

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

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

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

Example 2.2.7. Consider the general linear group $\mathrm{GL}(n, \mathbb{R})$ 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 2.2.8. We cannot understate the importance of context when discussing the structure of groups and subgroups. Like we mentioned in Remark 2.1.6, a nonempty set with an associative binary operation need not be a group — even if it possesses a multiplicative identity. Likewise, a nonempty subset of a group is not necessarily a subgroup. Crucially, a subgroup must inherit the same binary operation as the group in which it is contained. Observe that the group $(\mathbb{R}^{n\times n}, +)$ of $n \times n$ real matrices contains $\mathrm{GL}(n, \mathbb{R})$ 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.

Often, it is convenient to use the following proposition and its two corollary.

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

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

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

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

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

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

Corollary 2.2.10 (Two-Step Subgroup Test)**.** *Given a group* $(G, *)$ *and a nonempty set* $H \subseteq G$, *we have that* $(H, *)$ *is a subgroup of* G *if and only if* $h_1 * h_2 \in H$ *for all elements* $h_1, h_2 \in H$ *and* $h^{-1} \in H$ *for all elements* $h \in H$.

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

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

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

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

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

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

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

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

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

Example 2.2.12. We will soon come to see that like the order of a group, the subgroups admitted by a group provide a concrete way to distinguish that group from other groups of the same order. Consider the groups $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ with respect to modular addition. Both of these have order four, but we will demonstrate that they are distinct by showing that $\mathbb{Z}_4$ admits only one non-trivial

proper subgroup while $\mathbb{Z}_2 \times \mathbb{Z}_2$ admits three non-trivial proper subgroups. If we drop the modulo n notation for this example, the elements of $\mathbb{Z}_4$ are $\{0, 1, 2, 3\}$, and its **Cayley table** is as follows.

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

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

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

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

References

[DW00] J.P. D'Angelo and D.B. West. *Mathematical Thinking: Problem Solving and Proofs*. Upper Saddle River, NJ: Prentice-Hall, 2000.