Logic II: Set Theory

Gudfit

Contents

Chapter 1

Ideas & Motivations

Welcome to Set theory (with some theory) by me (Gudfit). The point of these notes is to cover everything I think is important as I build up to my current knowledge, while keeping it free and accessible for everyone from kids to adults.

I aim for each set of notes to be max 100 pages, as rigorous as possible, and far-reaching too.

That means I'll cover the axioms and proofs of the most interesting stuff, plus I'll pull in other subjects we've already touched on to show how math builds on itself like funky Lego. These notes build on my existing logic notes; they're aimed at keeping the proofs, ideas, and build-up of set theory as informal as possible.

It'll be a mix of quick ideas and concepts, but in the appendix for each section, I'll go rigorous with the key axioms pulled from a bunch of books.

Chapter 2

An Introduction to Set Theory

With the development of propositional logic, we have a formal language for rigorous communication. This language is sufficiently expressive to construct complex arguments from simple truths. With this language in hand, we are ready to embark on our studies of mathematics proper. The first question we must answer is: what is our universe of discourse? What are mathematical objects?

The most intuitive objects are numbers. The most fundamental kind of number is the natural number, which corresponds to the non-negative whole numbers. We can characterise these with a recurrence: zero is a natural number, and if n is a natural number, then its successor, $s(n)$, is also a natural number. This notion of a successor is an example of a function: an object that maps inputs from a domain to outputs in a codomain in a deterministic way.

Since functions map inputs to outputs, we are driven to ask, "inputs from where?" All roads eventually lead to the idea of a collection of things. In the same way binary numbers form a foundation for the files on your computer, we will build our mathematical universe using collections as our fundamental unit. We will call these collections sets, and refer to the objects they contain as their elements. As the most fundamental object in our universe, we will study sets first and encode their behaviour as axioms. This system is known as Zermelo-Fraenkel set theory.

To facilitate our discussion, we will adopt standard symbols for the common number systems used in mathematics.

Symbol	Name	Description
$\mathbb{N}$	Natural Numbers	$\{0, 1, 2, 3, \ldots\}$
$\mathbb{Z}$	Integers	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{Z}^+$	Positive Integers	$\{1, 2, 3, \ldots\}$
$\mathbb{Q}$	Rational Numbers	$\{p/q \mid p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0\}$
$\mathbb{R}$	Real Numbers	The set of all numbers on the continuous number line.

Remark. As a final note on notation, we will be simplifying our notation from this section forward. We had previously been introduced to the symbols $\rightarrow$ and $\leftrightarrow$ for expressing conditional statements and the symbols $\vdash$ and $\equiv$ in the metalanguage. Given the theorems we proved in the previous notes, the line between these two classes of symbols has been made blurrier. It is typical in mathematical practice to ignore this distinction. So, we now introduce the symbol $\Rightarrow$ to denote entailment as a replacement for the $\rightarrow$ and $\vdash$ symbols. Similarly, we introduce $\Leftrightarrow$ as a replacement for $\leftrightarrow$ and $\equiv$, denoting logical equivalence in all contexts.

2.1 The Language of Sets

A set is an abstraction of the idea of a collection of objects. This idea implies the need to communicate two kinds of relationships: equality and elementhood. These will be the two basic predicate symbols of our

theory.

To identify objects that are the same, we introduce the binary equality predicate, denoted by $=$. We say $x = y$ precisely when x is identical to y. We assume axiomatically that equality is an equivalence relation, meaning it is:

1. Reflexive: $\forall x(x = x)$
2. Symmetric: $\forall x \forall y((x = y) \Rightarrow (y = x))$
3. Transitive: $\forall x \forall y \forall z(((x = y) \wedge (y = z)) \Rightarrow (x = z))$

The second, more interesting, predicate relates sets to the elements they contain. We call this predicate elementhood and denote it with the $\in$ symbol. For a set A, the statement $x \in A$ asserts that x is an element of A, while $x \notin A$ is shorthand for $\neg(x \in A)$.

These two predicates are enough to express anything we need to say about sets. If we wanted to state that a set A contains exactly the elements 0 and 1, we could write $\forall x(x \in A \Leftrightarrow (x = 0 \vee x = 1))$. This would be cumbersome to write every time, so we introduce some notation.

Definition 2.1.1. ***Set-Builder Notation***. Given finitely many terms $x_0, x_1, \ldots, x_{n-1}$, we denote by $\{x_0, x_1, \ldots, x_{n-1}\}$ the set whose elements are exactly these objects. For any object z, the following holds:

$$z \in \{x_0, x_1, \ldots, x_{n-1}\} \Leftrightarrow (z = x_0) \vee (z = x_1) \vee \ldots \vee (z = x_{n-1})$$

This notation is restrictive; it only allows us to describe sets with finitely many elements. To describe larger sets, such as the set of even numbers, we introduce set-comprehension notation.

Definition 2.1.2. ***Set-Comprehension Notation***. Given a property $\mathbf{P}(x)$, we can refer to the collection of all objects x that satisfy this property by writing $\{x \mid \mathbf{P}(x)\}$. The defining equivalence is:

$$z \in \{x \mid \mathbf{P}(x)\} \Leftrightarrow \mathbf{P}(z)$$

For example, the set of even natural numbers could be written as $\{x \mid x \text{ is a natural number } \wedge x \text{ is even}\}$.

The elementhood predicate naturally implies another relationship two sets can share. It is clear that every even natural number is also a natural number. This emergent relationship is captured by the following definition.

Definition 2.1.3. ***Subset***. Given two sets A and B, we say that A is a subset of B, denoted $A \subseteq B$, when every element of A is also an element of B. Formally:

$$A \subseteq B \Leftrightarrow \forall x(x \in A \Rightarrow x \in B)$$

2.2 The Axioms of Set Theory

The notation we have introduced does not assert that any of these sets actually exist. To formally have sets to talk about, we need to introduce them with axioms.

Axiom of Extensionality

Sets are entirely determined by their elements. If we think of sets as abstract collections, then everything we need to know about a set should be determined by the elements it contains. We should expect that two sets are equal precisely when they have the same elements.

***Axiom* 2.2.1. Extensionality**. Two sets are equal if and only if they have the same elements.

$$\forall X \forall Y(X = Y \Leftrightarrow \forall z(z \in X \Leftrightarrow z \in Y))$$

The Axiom of Extensionality formalises our intuition. For example, if $A = \{0, 1, 2\}$ and $B = \{2, 0, 1\}$, we can see they contain the same elements, just in a different order. The axiom allows us to conclude that $A = B$. The order of elements and their repetition do not matter.

This axiom provides the standard method for proving two sets are equal.

***Theorem* 2.2.1.** For any sets X and Y, we have $X = Y \Leftrightarrow (X \subseteq Y) \wedge (Y \subseteq X)$.

Proof. Let X and Y be arbitrary sets. Observe the following chain of equivalences.

$$\begin{aligned} X = Y &\Leftrightarrow \forall z(z \in X \Leftrightarrow z \in Y) && \text{by Extensionality} \\ &\Leftrightarrow \forall z((z \in X \Rightarrow z \in Y) \wedge (z \in Y \Rightarrow z \in X)) && \text{by Biconditional Disintegration} \\ &\Leftrightarrow (\forall z(z \in X \Rightarrow z \in Y)) \wedge (\forall z(z \in Y \Rightarrow z \in X)) && \text{by Distributivity of } \forall \text{ over } \wedge \\ &\Leftrightarrow (X \subseteq Y) \wedge (Y \subseteq X) && \text{by definition of Subset} \end{aligned}$$

Therefore, the equivalence holds. ■

This gives space for two important definitions, the proper subset and the empty set.

Definition 2.2.1. ***Proper Subset***. A set A is a proper subset of a set B, denoted $A \subset B$, if A is a subset of B but is not equal to B.

$$A \subset B \Leftrightarrow (A \subseteq B) \wedge (A \neq B)$$

Definition 2.2.2. ***Empty Set***. A set X is empty if it contains no elements, i.e., $\forall y(y \notin X)$. We define the symbol $\emptyset$ to denote such a set: $\emptyset := \{z \mid z \neq z\}$.

With this definition, we can prove some fundamental properties about $\emptyset$, assuming for a moment that it exists.

***Theorem* 2.2.2. The Empty Set is Empty**. The set $\emptyset$ contains no elements.

$$\forall x(x \notin \emptyset)$$

Proof. Let x be an arbitrary object. Suppose, for the sake of contradiction, that $x \in \emptyset$. By the definition of $\emptyset$, this means $x \in \{z \mid z \neq z\}$. By the definition of set comprehension, this implies that the property defining the set must be true for x, so $x \neq x$. This contradicts the reflexivity of equality ($x = x$). Therefore, our initial assumption must be false, and $x \notin \emptyset$. ■

***Theorem* 2.2.3. The Empty Set is Unique**. There is only one empty set. Any set with no elements is equal to $\emptyset$.

Proof. Let X be any set that is empty, so $\forall y(y \notin X)$. We wish to show $X = \emptyset$. By the Axiom of Extensionality, this requires showing $\forall z(z \in X \Leftrightarrow z \in \emptyset)$. Let z be an arbitrary object. The statement $z \in X$ is false by our assumption about X. The statement $z \in \emptyset$ is false by the previous theorem. Thus, $z \in X \Leftrightarrow z \in \emptyset$ is equivalent to $\bot \Leftrightarrow \bot$, which is true. Since z was arbitrary, we conclude $X = \emptyset$. ■

***Theorem* 2.2.4.** The empty set is a subset of every set.

Proof. Let A be an arbitrary set. We wish to show $\emptyset \subseteq A$, which is equivalent to $\forall x(x \in \emptyset \implies x \in A)$. Let x be an arbitrary object. The antecedent, $x \in \emptyset$, is false by definition of the empty set. Therefore, the implication is vacuously true. Since x was arbitrary, the theorem holds. ■

Axiom Schema of Separation

Our set-comprehension notation $\{x \mid \mathbf{P}(x)\}$ is powerful, but dangerous if used without restriction. Consider the predicate $\mathbf{P}(s) \Leftrightarrow s \notin s$, and define the set $R := \{s \mid s \notin s\}$. This is the "set of all sets that are not elements of themselves". Now, let us ask: is R an element of itself?

- If $R \in R$, then by the definition of R, R must satisfy the property $s \notin s$. This means $R \notin R$, a contradiction.
- If $R \notin R$, then R satisfies the property $s \notin s$. By the definition of R, this means R must be an element of R, so $R \in R$, another contradiction.

This is Russell's Paradox. The mere existence of R is inherently contradictory. The problem arose from our use of unrestricted comprehension. To avoid this, we restrict comprehension to already existing sets. We can form a new set by separating off elements that satisfy a property from a set we already know exists.

***Axiom* 2.2.2. Schema of Separation**. For any property $\mathbf{P}(x)$ and any set A, there exists a set B whose elements are precisely those elements of A that satisfy $\mathbf{P}(x)$.

$$\forall A \exists B(\forall x(x \in B \Leftrightarrow (x \in A \wedge \mathbf{P}(x))))$$

Remark. This is called an axiom schema because it is not a single axiom, but an infinite collection of axioms — one for each possible property $\mathbf{P}(x)$.

We denote the set B by $\{x \in A \mid \mathbf{P}(x)\}$. This axiom now allows us to formally prove the existence of $\emptyset$. The Axiom of Existence, which we have not stated, simply guarantees that at least one set exists. Let A be any such set. Then by the Schema of Separation, the set $B = \{x \in A \mid x \neq x\}$ exists. Since no object x can satisfy $x \neq x$, B has no elements. It is an empty set, and by our uniqueness theorem, it must be $\emptyset$.

The schema also synergises well with other axioms, allowing us to prove that many useful set-theoretic constructions are possible.

Definition 2.2.3. ***Intersection***. The intersection of two sets A and B, denoted $A \cap B$, is the set of all elements they share in common.

$$A \cap B := \{x \mid x \in A \wedge x \in B\}$$

***Theorem* 2.2.5. Existence of Intersections**. For any two sets A and B, the set $A \cap B$ exists.

Proof. Let A and B be sets. Consider the property $\mathbf{P}(x) \Leftrightarrow x \in B$. By the Schema of Separation, the set $C = \{x \in A \mid \mathbf{P}(x)\}$ exists. This is precisely $\{x \in A \mid x \in B\}$, which is the definition of $A \cap B$. ■

Definition 2.2.4. ***Set Difference***. The difference of sets A and B, denoted $A \setminus B$, is the set obtained by removing every element of B from A.

$$A \setminus B := \{x \mid x \in A \wedge x \notin B\}$$

Remark. The proof for the existence of the set difference $A \setminus B$ is analogous to that for intersection and is left as an exercise.

Further Axioms of Construction

***Axiom* 2.2.3. Pairing**. For any two objects A and B, there exists a set that contains exactly A and B.

$$\forall A \forall B \exists C(\forall x(x \in C \Leftrightarrow (x = A \vee x = B)))$$

This set C is denoted $\{A, B\}$. If $A = B$, we get the singleton set $\{A\}$. This axiom guarantees that our set-builder notation is meaningful for pairs and singletons.

***Axiom* 2.2.4. Union**. For any set of sets S, there exists a set U that contains all the elements of all the sets in S.

$$\forall S \exists U(\forall x(x \in U \Leftrightarrow \exists A(A \in S \wedge x \in A)))$$

This set U is called the union of S and is denoted $\bigcup S$. For example, if $S = \{\{0,1\},\{1,2,3\}\}$, then $\bigcup S = \{0,1,2,3\}$.

This axiom allows us to prove the existence of the familiar union of two sets.

Definition 2.2.5. ***Union of Two Sets***. The union of two sets A and B, denoted $A \cup B$, is the set containing all elements from A or B.

$$A \cup B := \{x \mid x \in A \vee x \in B\}$$

***Theorem* 2.2.6. Existence of Unions**. For any two sets A and B, the set $A \cup B$ exists.

Proof. Let A and B be sets. By the Axiom of Pairing, the set $S = \{A, B\}$ exists. By the Axiom of Union, the set $\bigcup S$ exists. An element x is in $\bigcup S$ if and only if there exists a set $C \in S$ such that $x \in C$. Since $S = \{A, B\}$, this is equivalent to $(C = A \vee C = B) \wedge x \in C$, which simplifies to $x \in A \vee x \in B$. This is precisely the definition of $A \cup B$. Thus, $A \cup B = \bigcup\{A, B\}$ and therefore exists. ■

***Axiom* 2.2.5. Power Set**. For any set S, there exists a set containing all the subsets of S.

$$\forall S \exists P(\forall X(X \in P \Leftrightarrow X \subseteq S))$$

This set P is called the power set of S and is denoted $\mathcal{P}(S)$. For example, if $S = \{0,1\}$, then $\mathcal{P}(S) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$.

Axiom of Regularity

You may have wondered whether a set can contain itself as an element. So far, nothing formally prohibits $x \in x$. The final axiom we will consider is designed to prevent such pathological structures and ensure that the elementhood relation is well-founded.

Definition 2.2.6. ***Disjoint Sets***. Two sets A and B are said to be disjoint if their intersection is the empty set, i.e., $A \cap B = \emptyset$.

***Axiom* 2.2.6. Regularity**. Every non-empty set X contains an element y that is disjoint from X.

$$\forall X(X \neq \emptyset \Rightarrow \exists y(y \in X \wedge X \cap y = \emptyset))$$

This axiom has far-reaching consequences, one of which is that there are no infinitely descending $\in$-chains. For our purposes, it establishes that sets do not contain themselves.

***Theorem* 2.2.7. Well-Foundedness of Elementhood**. No set is an element of itself.

$$\forall x(x \notin x)$$

Proof. Let x be an arbitrary set and suppose, for a contradiction, that $x \in x$. Consider the set $A = \{x\}$, which exists by the Axiom of Pairing. Since $x \in A$, A is not empty. By the Axiom of Regularity, there must be an element $y \in A$ such that $A \cap y = \emptyset$. Since x is the only element of A, we must have $y = x$. The condition becomes $A \cap x = \emptyset$. However, we know $x \in A$. We also assumed $x \in x$. Therefore, x is in both A and x, which means $x \in A \cap x$. This implies $A \cap x \neq \emptyset$, contradicting our deduction from the axiom. Thus, our initial assumption must be false, so $x \notin x$. ■

This result can be used to show that a "set of all sets" cannot exist.

***Theorem* 2.2.8. The Universe Does Not Exist**. There is no set that contains every set.

$$\neg\exists U \forall x(x \in U)$$

Proof. Suppose for a contradiction that such a universal set U exists, satisfying $\forall x(x \in U)$. Since this holds for all sets x, it must hold for U itself. This means $U \in U$. However, this contradicts the previous theorem, which states $\forall z(z \notin z)$. Therefore, no such set U can exist. ■

2.3 Elementary Operations on Sets

The axioms allow us to define standard operations for manipulating sets. For any sets A and B, their existence is guaranteed by the axioms.

- **Union**: $A \cup B := \bigcup\{A, B\} = \{x \mid x \in A \vee x \in B\}$.
- **Intersection**: $A \cap B := \{x \in A \mid x \in B\}$.
- **Difference**: $A \setminus B := \{x \in A \mid x \notin B\}$.
- **Symmetric Difference**: $A \triangle B := (A \setminus B) \cup (B \setminus A)$.

We can also define these operations over arbitrary non-empty collections of sets $\mathcal{F}$.

- **Union**: $\bigcup \mathcal{F} = \{x \mid \exists A \in \mathcal{F}, x \in A\}$.
- **Intersection**: $\bigcap \mathcal{F} = \{x \mid \forall A \in \mathcal{F}, x \in A\}$.

Definition 2.3.1. ***Mutually Disjoint Sets***. A collection of sets $\mathcal{F}$ is a system of mutually disjoint sets (or pairwise disjoint sets) if for any two distinct sets $A, B \in \mathcal{F}$, it holds that $A \cap B = \emptyset$. Formally:

$$\forall A \in \mathcal{F}, \forall B \in \mathcal{F}(A \neq B \Rightarrow A \cap B = \emptyset)$$

Properties of Set Operations

The operations of union, intersection, and difference obey a set of algebraic laws that are analogous to the laws of propositional logic.

***Theorem* 2.3.1.** For any sets A, B, C and a universal set U, the following properties hold:

(i) **Commutative Laws:**
- $A \cup B = B \cup A$
- $A \cap B = B \cap A$

(ii) **Associative Laws:**
- $(A \cup B) \cup C = A \cup (B \cup C)$
- $(A \cap B) \cap C = A \cap (B \cap C)$

(iii) **Distributive Laws:**
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(iv) **De Morgan's Laws:**
- $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$
- $U \setminus (A \cap B) = (U \setminus A) \cup (U \setminus B)$

Proof. We prove a selection of these properties. The remaining proofs are left as exercises.

(i) Commutativity of Union: We must show $A \cup B = B \cup A$. By definition, this requires showing that for any x, $x \in A \cup B \Leftrightarrow x \in B \cup A$.

$$\begin{aligned} x \in A \cup B &\Leftrightarrow (x \in A) \vee (x \in B) && \text{by definition of Union} \\ &\Leftrightarrow (x \in B) \vee (x \in A) && \text{by Commutativity of } \vee \\ &\Leftrightarrow x \in B \cup A && \text{by definition of Union} \end{aligned}$$

(iii) Distributivity of Intersection over Union: We must show $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

$$\begin{aligned}
x \in A \cap (B \cup C) &\Leftrightarrow (x \in A) \wedge (x \in B \cup C) && \text{by definition of Intersection} \\
&\Leftrightarrow (x \in A) \wedge ((x \in B) \vee (x \in C)) && \text{by definition of Union} \\
&\Leftrightarrow ((x \in A) \wedge (x \in B)) \vee ((x \in A) \wedge (x \in C)) && \text{by Distributivity of } \wedge \text{ over } \vee \\
&\Leftrightarrow (x \in A \cap B) \vee (x \in A \cap C) && \text{by definition of Intersection} \\
&\Leftrightarrow x \in (A \cap B) \cup (A \cap C) && \text{by definition of Union}
\end{aligned}$$

(iv) De Morgan's Law: We must show $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$.

$$\begin{aligned}
x \in U \setminus (A \cup B) &\Leftrightarrow (x \in U) \wedge (x \notin A \cup B) && \text{by definition of Difference} \\
&\Leftrightarrow (x \in U) \wedge \neg(x \in A \vee x \in B) && \text{by definition of Union} \\
&\Leftrightarrow (x \in U) \wedge (x \notin A \wedge x \notin B) && \text{by De Morgan's Law for logic} \\
&\Leftrightarrow (x \in U \wedge x \notin A) \wedge (x \in U \wedge x \notin B) && \text{by Associativity and Idempotence of } \wedge \\
&\Leftrightarrow (x \in U \setminus A) \wedge (x \in U \setminus B) && \text{by definition of Difference} \\
&\Leftrightarrow x \in (U \setminus A) \cap (U \setminus B) && \text{by definition of Intersection}
\end{aligned}$$

■

2.4 Exercises

Part I: Foundational Concepts & The Axioms

1. Determine whether the following statements are true or false. Justify your answer with a brief explanation or a counterexample. Let $A = \{1, \{2, 3\}\}$.

(a) $1 \in A$
(b) $2 \in A$
(c) $\{1\} \subseteq A$
(d) $\{2, 3\} \in A$
(e) $\{2, 3\} \subseteq A$
(f) $\emptyset \in A$
(g) $\emptyset \subseteq A$

2. Use the Axiom of Extensionality to prove that the sets $A = \{n \in \mathbb{Z} \mid n^2 = 4\}$ and $B = \{-2, 2\}$ are equal.

Remark. You must show that $(x \in A \Rightarrow x \in B)$ and $(x \in B \Rightarrow x \in A)$.

3. List all the elements of the set $\mathcal{P}(\mathcal{P}(\{\emptyset\}))$. How many elements does it contain?

4. Prove that the subset relation is transitive. That is, for any sets A, B, C, if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

5. Disprove the following statement by providing a counterexample: For any sets A and B, $\mathcal{P}(A \cup B) = \mathcal{P}(A) \cup \mathcal{P}(B)$. Under what condition does equality hold?

6. The text proves the existence of $A \cap B$. Using a similar method, prove that for any sets A and B, the set difference $A \setminus B$ exists. Explicitly state which axiom you are using and what property $\mathbf{P}(x)$ you have chosen.

7. Using the Axiom of Regularity, prove that for any two sets x, y, it is not possible to have both $x \in y$ and $y \in x$.

Remark. Assume such sets exist and consider the set $\{x, y\}$. Apply the Axiom of Regularity to this set to derive a contradiction.

8. Let a and b be objects. The ordered pair (a, b) is defined as the set $\{\{a\}, \{a, b\}\}$. Using this definition, compute the sets corresponding to $(1, 2)$ and $(2, 1)$. Then determine the following sets in roster form.

(a) $(1, 2) \cup (2, 1)$

(b) $(1, 2) \cap (2, 1)$

(c) $(1, 2) \setminus (2, 1)$

9. Let $\mathbf{P}(x)$ be a logical formula. Prove the following statements, which connect the logical quantifiers to set operations.

(a) $\forall x(x \in A \cup B \Rightarrow \mathbf{P}(x)) \Leftrightarrow (\forall x(x \in A \Rightarrow \mathbf{P}(x)) \wedge \forall x(x \in B \Rightarrow \mathbf{P}(x)))$

(b) $\exists x(x \in A \cup B \wedge \mathbf{P}(x)) \Leftrightarrow (\exists x(x \in A \wedge \mathbf{P}(x)) \vee \exists x(x \in B \wedge \mathbf{P}(x)))$

10. Find a formula $\mathbf{P}(x)$ and sets A and B to show that the following implication is false:

$$(\exists x(x \in A \wedge \mathbf{P}(x))) \wedge (\exists x(x \in B \wedge \mathbf{P}(x))) \Rightarrow \exists x(x \in A \cap B \wedge \mathbf{P}(x))$$

Part II: Properties of Set Operations

11. The text proves several algebraic laws for set operations. Prove the following remaining laws for arbitrary sets A, B, C.

(a) **Associativity of Intersection:** $(A \cap B) \cap C = A \cap (B \cap C)$

(b) **Second Distributive Law:** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

12. The symmetric difference of two sets A and B is defined as $A \triangle B := (A \setminus B) \cup (B \setminus A)$. Prove that the symmetric difference is associative: $(A \triangle B) \triangle C = A \triangle (B \triangle C)$.

13. Prove the following identities involving the power set.

(a) For any sets A and B, $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$.

(b) For any sets A and B, $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$.

14. For any sets A, B, C, prove that if $A \subseteq B$, then $C \setminus B \subseteq C \setminus A$.

15. Prove that the following three statements are equivalent for any sets A and B.

(i) $A \subseteq B$

(ii) $A \cup B = B$

(iii) $A \setminus B = \emptyset$

16. Using the result from the previous problem, or otherwise, prove that for any sets A and B, we have $A \cup B = A \cap B$ if and only if $A = B$.

17. Prove that for any sets A, B, C, if $A \cap C = B \cap C$ and $A \cup C = B \cup C$, then $A = B$.

Remark. Consider an element $x \in A$. If $x \in C$, what can you conclude? If $x \notin C$, what can you conclude?

Part III: Families of Sets and Advanced Topics

18. The generalised intersection $\bigcap \mathcal{F}$ is defined as $\{x \mid \forall A \in \mathcal{F}, x \in A\}$. Assuming $\mathcal{F}$ is a non-empty collection of sets, prove that $\bigcap \mathcal{F}$ exists.

Remark. Since $\mathcal{F}$ is non-empty, there exists some set $S_0 \in \mathcal{F}$. Show that $\bigcap \mathcal{F}$ is a subset of S_0 and then apply the appropriate axiom.

What problem arises if $\mathcal{F} = \emptyset$?

19. Prove for any sets A and B that $\bigcup\{A, B\} = A \cup B$ and $\bigcap\{A, B\} = A \cap B$. This demonstrates that the binary operations are special cases of the generalised operations on families of sets.

20. Let $\mathcal{F}$ be a collection of sets and let $X \in \mathcal{F}$. Prove that $\bigcap \mathcal{F} \subseteq X \subseteq \bigcup \mathcal{F}$.

21. Provide an example of a family of sets $\mathcal{F}$ that is disjoint (i.e., $\bigcap \mathcal{F} = \emptyset$) but not pairwise disjoint.

22. Let $\mathcal{F} = \{A_i \mid i \in I\}$ be a family of sets. Prove the generalised De Morgan's laws for set difference:

(a) $B \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (B \setminus A_i)$
(b) $B \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (B \setminus A_i)$

23. In the next chapter, an ordered pair (a, b) is defined as $\{\{a\}, \{a, b\}\}$. An alternative definition, proposed by Norbert Wiener, is $(a, b)_W := \{\{\{a\}, \emptyset\}, \{\{b\}\}\}$. Prove that this definition also satisfies the fundamental property of ordered pairs: $(a, b)_W = (a', b')_W \Leftrightarrow (a = a' \wedge b = b')$.

Chapter 3

Relations

Central to mathematical practice is the concept of a function, which formalises the idea of an association between two sets. We have used this notion informally, but to define it within set theory, we must first construct a way to represent a directed link from an "input" to an "output". An unordered pair $\{a, b\}$ is insufficient for this task, as it is indistinguishable from $\{b, a\}$. We require a structure where the order of elements is significant.

3.1 Ordered Pairs and Cartesian Products

We begin by defining an ordered pair using only the axioms of set theory. The following definition, due to Kazimierz Kuratowski, captures the essential property that the order of the coordinates matters.

Definition 3.1.1. ***Ordered Pair***. Given objects a and b, the ordered pair (a, b) is the set defined as:

$$(a, b) := \{\{a\}, \{a, b\}\}$$

The object a is the first coordinate, and b is the second coordinate.

This construction may seem unintuitive, but its sole purpose is to satisfy the fundamental property of ordered pairs, which we now prove.

***Theorem* 3.1.1. Equality of Ordered Pairs**. For any objects a, b, a', b', we have $(a, b) = (a', b') \Leftrightarrow a = a' \wedge b = b'$.

Proof. The implication ($\Leftarrow$) is immediate from the definition. If $a = a'$ and $b = b'$, then the sets $\{\{a\}, \{a, b\}\}$ and $\{\{a'\}, \{a', b'\}\}$ are identical.

For the forward implication ($\Rightarrow$), assume $(a, b) = (a', b')$, which means $\{\{a\}, \{a, b\}\} = \{\{a'\}, \{a', b'\}\}$. We consider two cases.

$a = b$. In this case, $(a, b) = \{\{a\}, \{a, a\}\} = \{\{a\}\}$. Our assumption becomes $\{\{a\}\} = \{\{a'\}, \{a', b'\}\}$. The set on the left contains one element, $\{a\}$. Thus, the set on the right must also contain only one element, which implies $\{a'\} = \{a', b'\}$. This equality holds if and only if $a' = b'$. Therefore, $\{\{a\}\} = \{\{a'\}\}$, which gives $\{a\} = \{a'\}$, and so $a = a'$. Since $a = b$ and $a' = b'$, we have $a = a'$ and $b = b'$.

$a \neq b$. The set $\{\{a\}, \{a, b\}\}$ contains two distinct elements: the singleton $\{a\}$ and the doubleton $\{a, b\}$. The equality $\{\{a\}, \{a, b\}\} = \{\{a'\}, \{a', b'\}\}$ implies that the set on the right must also contain two distinct elements, so $a' \neq b'$. By set equality, the elements must be the same. The singleton $\{a\}$ must equal the singleton $\{a'\}$, because it cannot equal the doubleton $\{a', b'\}$. From $\{a\} = \{a'\}$, we deduce $a = a'$. The remaining elements must also be equal: $\{a, b\} = \{a', b'\}$. Since we have already shown $a = a'$, this simplifies to $\{a, b\} = \{a, b'\}$. As $a \neq b$, it must be that $b = b'$.

In both cases, we conclude $a = a'$ and $b = b'$. ■

Remark. This principle can be extended recursively to define ordered n-tuples. For instance, an ordered triple is defined as $(a, b, c) := ((a, b), c)$.

With a formal definition for ordered pairs, we can define the set of all possible pairs between elements of two sets. This construction, named in honour of René Descartes , is fundamental to defining relations in this chapter and functions in the next.

Definition 3.1.2. ***Cartesian Product***. Let A and B be sets. The Cartesian product of A and B, denoted $A \times B$, is the set of all ordered pairs whose first coordinate is in A and whose second coordinate is in B.

$$A \times B := \{(a, b) \mid a \in A \wedge b \in B\}.$$

We write A^2 as shorthand for $A \times A$.

***Theorem* 3.1.2. Existence of Cartesian Products**. For any two sets A and B, the Cartesian product $A \times B$ exists.

Proof. Let $a \in A$ and $b \in B$. The ordered pair (a, b) is $\{\{a\}, \{a, b\}\}$. Since $a \in A \cup B$ and $b \in A \cup B$, the sets $\{a\}$ and $\{a, b\}$ are both subsets of $A \cup B$. By the Axiom of Power Set, this means $\{a\} \in \mathcal{P}(A \cup B)$ and $\{a, b\} \in \mathcal{P}(A \cup B)$. Consequently, the pair $(a, b) = \{\{a\}, \{a, b\}\}$ is a subset of $\mathcal{P}(A \cup B)$. Applying the Power Set axiom again, this implies $(a, b) \in \mathcal{P}(\mathcal{P}(A \cup B))$. Since every element of $A \times B$ is an element of the set $\mathcal{P}(\mathcal{P}(A \cup B))$, which exists, we can use the Schema of Separation to construct $A \times B$:

$$A \times B = \{p \in \mathcal{P}(\mathcal{P}(A \cup B)) \mid \exists a \in A, \exists b \in B, p = (a, b)\}$$

Therefore, $A \times B$ exists. ■

3.2 Relations

A relation describes an association between objects. Formally, a relation between two sets is defined as a collection of ordered pairs.

Definition 3.2.1. ***Binary Relation***. A set R is a binary relation from a set A to a set B if R is a subset of the Cartesian product $A \times B$. If $A = B$, we say R is a relation on A.

Note. If $(x, y) \in R$, we often write xRy and say that x is related to y by R.

Definition 3.2.2. ***Restriction of a Relation***. Let R be a binary relation on a set A, and let $S \subseteq A$. The restriction of R to S is the relation R_S on S defined as:

$$R_S = R \cap (S \times S) = \{(x, y) \in R \mid x \in S \wedge y \in S\}$$

Example 3.2.1. The less-than relation on the set of natural numbers $\mathbb{N}$ can be formalised as the set of ordered pairs:

$$L = \{(x, y) \in \mathbb{N}^2 \mid x < y\}$$

Thus, $3L5$ because $(3, 5) \in L$, but it is not the case that $7L2$, because $(7, 2) \notin L$.

For any relation, we are interested in the sets of all possible first and second coordinates.

Definition 3.2.3. ***Domain, Range, and Field***. Let R be a binary relation.

- The **domain** of R is the set of all first coordinates:

$$\text{dom}(R) = \{x \mid \exists y, (x, y) \in R\}$$

- The *range* of R is the set of all second coordinates:

$$\text{ran}(R) = \{y \mid \exists x, (x, y) \in R\}$$

- The *field* of R is the union of its domain and range:

$$\text{field}(R) = \text{dom}(R) \cup \text{ran}(R)$$

Example 3.2.2. Consider the relation $R = \{(1, 3), (2, 4), (2, 5)\}$. The set of all first coordinates is $\{1, 2\}$, and the set of all second coordinates is $\{3, 4, 5\}$. Therefore, $\text{dom}(R) = \{1, 2\}$ and $\text{ran}(R) = \{3, 4, 5\}$. The relation is visualised in Figure 3.1.

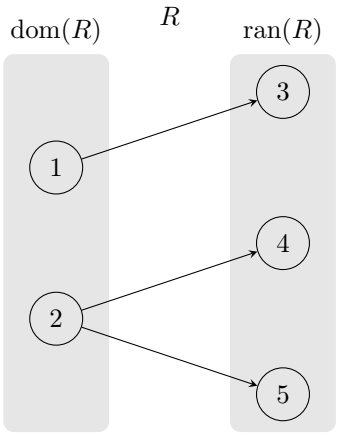


Figure 3.1: A visualisation of the relation $R = \{(1, 3), (2, 4), (2, 5)\}$.

***Proposition* 3.2.1.** Let R be a binary relation. And let $A = \bigcup(\bigcup R)$. Then $(x, y) \in R$ implies $x \in A$ and $y \in A$.

Proof. If $(x, y) = \{\{x\}, \{x, y\}\} \in R$, then $\{x, y\} \in \bigcup R$, and thus $x, y \in A$. ■

***Theorem* 3.2.1. Existence of Domain and Range**. For any binary relation R, the sets $\text{dom}(R)$ and $\text{ran}(R)$ exist.

Proof. Let R be a binary relation. For any $(x, y) \in R$, we have $(x, y) = \{\{x\}, \{x, y\}\}$. By definition of the union of a set, $\{x\}$ and $\{x, y\}$ are elements of $\bigcup R$. Applying the union operation again, x and y are elements of $\bigcup(\bigcup R)$. Thus, both $\text{dom}(R)$ and $\text{ran}(R)$ are subsets of $\bigcup(\bigcup R)$, which exists by the Axiom of Union. By the Schema of Separation, $\text{dom}(R)$ and $\text{ran}(R)$ therefore exist. ■

We can define several operations that construct new relations from existing ones.

Definition 3.2.4. ***Operations on Relations***. Let R and S be binary relations.

- The *inverse* of R, denoted R^{-1}, is the relation obtained by reversing the pairs in R:

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

- The *composition* of R and S, denoted $S \circ R$, links elements through an intermediary:

$$S \circ R = \{(x, z) \mid \exists y, (x, y) \in R \wedge (y, z) \in S\}$$

Example 3.2.3. The inverse of the less-than relation L on $\mathbb{N}$ is:

$$L^{-1} = \{(y, x) \mid (x, y) \in L\} = \{(y, x) \in \mathbb{N}^2 \mid x < y\} = \{(y, x) \in \mathbb{N}^2 \mid y > x\}$$

This is precisely the greater-than relation on $\mathbb{N}$.

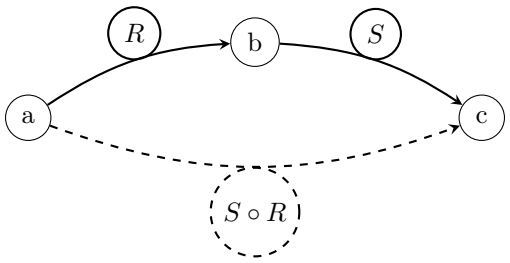


Figure 3.2: The composition $S \circ R$ provides a direct path from a to c.

Example 3.2.4. Let $R = \{(1,3),(2,4),(2,5)\}$ and $S = \{(0,1),(1,0),(0,2),(2,0)\}$. To find the composition $R \circ S$, we look for pairs (x,z) such that $(x,y) \in S$ and $(y,z) \in R$ for some intermediary y.

- $(0,1) \in S$ and $(1,3) \in R$. This gives $(0,3) \in R \circ S$.
- $(0,2) \in S$ and $(2,4) \in R$. This gives $(0,4) \in R \circ S$.
- $(0,2) \in S$ and $(2,5) \in R$. This gives $(0,5) \in R \circ S$.

No other links exist. Thus, $R \circ S = \{(0,3),(0,4),(0,5)\}$. Note that the composition $S \circ R$ is empty because $\text{ran}(R) = \{3,4,5\}$ and $\text{dom}(S) = \{0,1,2\}$ are disjoint.

***Theorem* 3.2.2. Existence of Inverse and Composition**. For any binary relations R and S, the relations R^{-1} and $S \circ R$ exist.

Proof. For the inverse, if $(y,x) \in R^{-1}$ then $y \in \text{ran}(R)$ and $x \in \text{dom}(R)$. Thus, $R^{-1} \subseteq \text{ran}(R) \times \text{dom}(R)$. Since the latter set exists, R^{-1} exists by Separation.

For the composition, if $(x,z) \in S \circ R$, then $x \in \text{dom}(R)$ and $z \in \text{ran}(S)$. Thus, $S \circ R \subseteq \text{dom}(R) \times \text{ran}(S)$. This parent set exists, so $S \circ R$ exists by Separation. ■

***Theorem* 3.2.3. Associativity of Composition**. For any relations R, S, T, composition is associative:

$$T \circ (S \circ R) = (T \circ S) \circ R$$

Proof. Let R, S, T be arbitrary relations. Then,

$$\begin{aligned}
(x,w) \in T \circ (S \circ R) &\Leftrightarrow \exists z, (x,z) \in S \circ R \wedge (z,w) \in T \\
&\Leftrightarrow \exists z, (\exists y, (x,y) \in R \wedge (y,z) \in S) \wedge (z,w) \in T \\
&\Leftrightarrow \exists z, \exists y, (x,y) \in R \wedge (y,z) \in S \wedge (z,w) \in T \\
&\Leftrightarrow \exists y, \exists z, (x,y) \in R \wedge (y,z) \in S \wedge (z,w) \in T \\
&\Leftrightarrow \exists y, (x,y) \in R \wedge (\exists z, (y,z) \in S \wedge (z,w) \in T) \\
&\Leftrightarrow \exists y, (x,y) \in R \wedge (y,w) \in T \circ S \\
&\Leftrightarrow (x,w) \in (T \circ S) \circ R
\end{aligned}$$

Since the conditions for membership are equivalent, the sets are equal. ■

***Theorem* 3.2.4. Inverse and Identity**. For any binary relation R, we have $\text{Id}_{\text{dom}(R)} \subseteq R^{-1} \circ R$ and $\text{Id}_{\text{ran}(R)} \subseteq R \circ R^{-1}$.

Proof. To prove the first inclusion, let $(x,x) \in \text{Id}_{\text{dom}(R)}$. This implies $x \in \text{dom}(R)$. By the definition of domain, there must exist some y such that $(x,y) \in R$. By the definition of the inverse relation, if $(x,y) \in R$, then $(y,x) \in R^{-1}$. We have found an intermediary y such that $(x,y) \in R$ and $(y,x) \in R^{-1}$. By the definition of composition, this implies that $(x,x) \in R^{-1} \circ R$. Therefore, $\text{Id}_{\text{dom}(R)} \subseteq R^{-1} \circ R$.

The proof for the second inclusion is analogous. Let $(y,y) \in \text{Id}_{\text{ran}(R)}$. This implies $y \in \text{ran}(R)$. By the definition of range, there exists some x such that $(x,y) \in R$. This in turn implies $(y,x) \in R^{-1}$. We have found

an x such that $(y, x) \in R^{-1}$ and $(x, y) \in R$. By the definition of composition, this means $(y, y) \in R \circ R^{-1}$. Therefore, $\mathrm{Id}_{\mathrm{ran}(R)} \subseteq R \circ R^{-1}$. ■

Definition 3.2.5. ***Special Relations***. Let A be a set.

- The *identity relation* on A, denoted Id_A, is the set of pairs relating each element of A to itself:

$$\mathrm{Id}_A = \{(a, a) \mid a \in A\}$$

- The *membership relation* on A, denoted $\in_A$, relates elements of A based on set membership:

$$\in_A = \{(a, b) \mid a, b \in A \wedge a \in b\}$$

These relations exist because they are subsets of $A \times A$.

Definition 3.2.6. ***Image and Inverse Image***. Let R be a binary relation and A be a set.

- The *image* of A under R is the set of all elements in the range of R that are related to some element of A:

$$R[A] = \{y \in \mathrm{ran}(R) \mid \exists x \in A, (x, y) \in R\}$$

- The *inverse image* of A under R is the set of all elements in the domain of R that are related to some element of A:

$$R^{-1}[A] = \{x \in \mathrm{dom}(R) \mid \exists y \in A, (x, y) \in R\}$$

Remark. The notation $R^{-1}[A]$ can be ambiguous: does it mean the inverse image of A under R, or the image of A under the relation R^{-1}? The following theorem shows that both interpretations yield the same set.

***Theorem* 3.2.5. Unambiguity of Inverse Notation**. For any binary relation R and any set A, the inverse image of A under R is equal to the image of A under R^{-1}.

Proof. By definition, the inverse image of A under R is $\{x \in \mathrm{dom}(R) \mid \exists y \in A, (x, y) \in R\}$. The image of A under the relation R^{-1} is $(R^{-1})[A] = \{z \in \mathrm{ran}(R^{-1}) \mid \exists y \in A, (y, z) \in R^{-1}\}$. We know that $\mathrm{ran}(R^{-1}) = \mathrm{dom}(R)$ and that $(y, z) \in R^{-1} \Leftrightarrow (z, y) \in R$. Substituting these into the definition of the image gives:

$$(R^{-1})[A] = \{z \in \mathrm{dom}(R) \mid \exists y \in A, (z, y) \in R\}$$

This is identical to the definition of the inverse image of A under R. ■

3.3 Exercises

1. Let $A = \{1, 2\}$, $B = \{x, y, z\}$, and $C = \{1, 3\}$. Let R be a relation from A to B defined by $R = \{(1, y), (1, z), (2, x)\}$. Let S be a relation from B to C defined by $S = \{(y, 1), (y, 3), (z, 3)\}$.

(a) Compute the relations R^{-1} and $S \circ R$. State the domain and range of R, S, and $S \circ R$.
(b) Compute the image $R[A]$ and the image $R[\{1\}]$. Compute the inverse image $S^{-1}[C]$. Is this equal to $\mathrm{dom}(S)$?

2. Prove the following fundamental properties for any binary relation R.

(a) $(R^{-1})^{-1} = R$.
(b) $\mathrm{dom}(R) = \mathrm{ran}(R^{-1})$.
(c) Suppose S is any binary relation for which the composition is defined prove $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.

3. Let R and S be binary relations with the same domain and codomain. Prove that the inverse operation distributes over set union and intersection.

(a) $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$
(b) $(R \cap S)^{-1} = R^{-1} \cap S^{-1}$

4. Let $R \subseteq A \times B$ be a relation. Prove that the identity relations Id_A and Id_B act as identity elements for composition. So $R \circ \mathrm{Id}_A = R$ and $\mathrm{Id}_B \circ R = R$.

5. Let $R \subseteq A \times B$ and $S \subseteq B \times C$ be relations. Prove that $S \circ R = \emptyset$ if and only if $\mathrm{ran}(R)$ and $\mathrm{dom}(S)$ are disjoint.

6. Let R be a binary relation and let A and B be arbitrary sets. Prove the following properties regarding the image of a set under a relation.

(a) $R[A \cup B] = R[A] \cup R[B]$.
(b) $R[A \cap B] \subseteq R[A] \cap R[B]$.
(c) Provide a specific example of a relation R and sets A, B for which the inclusion in part (b) is proper, i.e., $R[A \cap B] \neq R[A] \cap R[B]$.

7. Let R be a relation. Prove that its domain can be expressed as the union of the inverse images of all singleton sets $\{y\}$, where y is an element of the range.

$$\mathrm{dom}(R) = \bigcup_{y \in \mathrm{ran}(R)} R^{-1}[\{y\}]$$

8. Let R be a relation on a set A (i.e., $R \subseteq A \times A$). A relation is often classified by certain properties. We formally define three such properties:

- R is **reflexive** if $\mathrm{Id}_A \subseteq R$.
- R is **symmetric** if $R^{-1} = R$.
- R is **transitive** if $R \circ R \subseteq R$.

(a) For the set $A = \{1, 2, 3\}$, provide an example of a relation that is reflexive and symmetric, but not transitive.
(b) Prove that the definition of transitivity given above is equivalent to the logical statement: $\forall x, y, z \in A, ((x, y) \in R \wedge (y, z) \in R \Rightarrow (x, z) \in R)$.
(c) If a relation R on a non-empty set A is symmetric and transitive, must it also be reflexive? Justify your answer with a proof or a counterexample.

Remark. Consider the domain of R. Does it have to be equal to A?

3.4 Equivalence Relations

Certain relations are of particular interest because they capture a notion of sameness or equivalence between distinct objects. For instance, the fractions $\frac{1}{2}$ and $\frac{2}{4}$ are different as ordered pairs $(1, 2)$ and $(2, 4)$, yet they represent the same rational number. This notion of equivalence is formalised by relations that possess three specific properties.

Definition 3.4.1. ***Properties of a Relation***. Let $\sim$ be a binary relation on a set A.

- $\sim$ is **reflexive** if every element is related to itself.

$$\forall x \in A, (x \sim x)$$

- $\sim$ is **symmetric** if the relation holds in both directions.

$$\forall x, y \in A, (x \sim y \Rightarrow y \sim x)$$

- $\sim$ is **transitive** if a chain of relations implies a direct relation.

$$\forall x, y, z \in A, ((x \sim y \wedge y \sim z) \Rightarrow x \sim z)$$

Definition 3.4.2. ***Equivalence Relation***. A relation $\sim$ on a set A is an equivalence relation if it is reflexive, symmetric, and transitive.

Example 3.4.1. Consider the divisibility relation '$|$' on the set of integers $\mathbb{Z}$, where $a|b$ if there exists an integer k such that $b = ak$.

- **Reflexive**: For any $a \in \mathbb{Z}$, $a = a \cdot 1$, so $a|a$. The relation is reflexive.
- **Symmetric**: $2|4$ because $4 = 2 \cdot 2$, but $4 \nmid 2$ because there is no integer k such that $2 = 4k$. The relation is not symmetric.
- **Transitive**: If $a|b$ and $b|c$, then $b = ak_1$ and $c = bk_2$ for some integers k_1, k_2. Substituting gives $c = (ak_1)k_2 = a(k_1k_2)$. Since k_1k_2 is an integer, $a|c$. The relation is transitive.

Since divisibility is not symmetric, it is not an equivalence relation.

Example 3.4.2. Congruence Modulo n. Let $n \in \mathbb{Z}^+$. For integers a and b, we say a is congruent to b modulo n, written $a \equiv b \pmod{n}$, if n divides their difference, i.e., $n \mid (a - b)$. We will prove this is an equivalence relation on $\mathbb{Z}$. Let $a, b, c \in \mathbb{Z}$.

- **Reflexive**: $a - a = 0$. Since $n \mid 0$ (as $0 = n \cdot 0$), we have $a \equiv a \pmod{n}$.
- **Symmetric**: Suppose $a \equiv b \pmod{n}$. Then $n \mid (a - b)$, so $a - b = nk$ for some integer k. It follows that $b - a = n(-k)$. Since $-k$ is an integer, $n \mid (b - a)$, so $b \equiv a \pmod{n}$.
- **Transitive**: Suppose $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$. Then $n \mid (a-b)$ and $n \mid (b-c)$. So, $a-b = nk_1$ and $b - c = nk_2$ for some integers k_1, k_2. Adding these equations gives $(a - b) + (b - c) = nk_1 + nk_2$, which simplifies to $a - c = n(k_1 + k_2)$. Since $k_1 + k_2$ is an integer, $n \mid (a - c)$, so $a \equiv c \pmod{n}$.

Congruence modulo n is an equivalence relation.

Equivalence Classes and Partitions

An equivalence relation on a set naturally groups elements together. All elements that are equivalent to each other form a single group or class.

Definition 3.4.3. ***Equivalence Class***. Let $\sim$ be an equivalence relation on a set A. For any element $a \in A$, the **equivalence class** of a, denoted $[a]$, is the set of all elements in A that are equivalent to a.

$$[a] = \{x \in A \mid x \sim a\}$$

Example 3.4.3. For the relation of congruence modulo 3 on $\mathbb{Z}$:

- $[0] = \{x \in \mathbb{Z} \mid x \equiv 0 \pmod{3}\} = \{\ldots, -6, -3, 0, 3, 6, \ldots\}$
- $[1] = \{x \in \mathbb{Z} \mid x \equiv 1 \pmod{3}\} = \{\ldots, -5, -2, 1, 4, 7, \ldots\}$
- $[2] = \{x \in \mathbb{Z} \mid x \equiv 2 \pmod{3}\} = \{\ldots, -4, -1, 2, 5, 8, \ldots\}$

Note that $[3] = [0]$ and $[-1] = [2]$, as $3 \equiv 0 \pmod{3}$ and $-1 \equiv 2 \pmod{3}$. There are only three distinct equivalence classes.

Equivalence classes have a fundamental property: they divide the set into a collection of non-overlapping subsets.

***Theorem* 3.4.1. Properties of Equivalence Classes**. Let $\sim$ be an equivalence relation on a set A, and let $a, b \in A$.

(i) $a \in [a]$.

(ii) $a \sim b \Leftrightarrow [a] = [b]$.

(iii) $a \not\sim b \Leftrightarrow [a] \cap [b] = \emptyset$.

Proof.

$a \in [a]$. By reflexivity, $a \sim a$, so by definition of $[a]$, $a \in [a]$. This also implies that no equivalence class is empty.

$a \sim b \Leftrightarrow [a] = [b]$**.** ($\Rightarrow$) Assume $a \sim b$. We must show $[a] = [b]$ by mutual inclusion. Let $x \in [a]$, so $x \sim a$. Since $a \sim b$ and $\sim$ is transitive, $x \sim b$. Thus $x \in [b]$, proving $[a] \subseteq [b]$. Now let $y \in [b]$, so $y \sim b$. By symmetry, from $a \sim b$ we have $b \sim a$. By transitivity, $y \sim a$. Thus $y \in [a]$, proving $[b] \subseteq [a]$. Therefore, $[a] = [b]$.

($\Leftarrow$) Assume $[a] = [b]$. Since $a \in [a]$, it must be that $a \in [b]$. By definition, this means $a \sim b$.

$a \not\sim b \Leftrightarrow [a] \cap [b] = \emptyset$**.** This is the contrapositive of the previous statement. Two equivalence classes are either identical or completely disjoint. Suppose their intersection is not empty, so there exists some c such that $c \in [a]$ and $c \in [b]$. This means $c \sim a$ and $c \sim b$. By symmetry, $a \sim c$. By transitivity, $a \sim b$. From (ii), this implies $[a] = [b]$. Therefore, if $[a]$ and $[b]$ are not disjoint, they must be equal. The contrapositive is that if they are not equal (i.e., $a \not\sim b$), they must be disjoint.

■

This theorem shows that an equivalence relation on a set A splits A into a collection of mutually disjoint, non-empty subsets whose union is A. Such a collection is called a *partition* (see the formal definition in Section 3.5).

The set of all equivalence classes of a relation forms a partition of the underlying set. This set of classes is important enough to have its own name.

Definition 3.4.4. ***Quotient Set***. Let $\sim$ be an equivalence relation on a set A. The quotient set of A by $\sim$, denoted $A/\sim$, is the set of all equivalence classes of $\sim$.

$$A/\sim = \{[a] \mid a \in A\}$$

The quotient set $A/\sim$ is a partition of A.

Remark. The process of forming a quotient set gives rise to a canonical projection map (or quotient map), $p : A \to A/\sim$, defined by $p(a) = [a]$. By construction, this map is surjective.

Example 3.4.4. For congruence modulo 3, the quotient set is $\mathbb{Z}/\equiv$, usually written as $\mathbb{Z}_3$.

$$\mathbb{Z}_3 = \{[0], [1], [2]\}$$

This is a set containing three elements, where each element is itself an infinite set of integers.

Example 3.4.5. Define a relation $\sim$ on $\mathbb{R}^2$ such that for (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$,

$$(x_1, y_1) \sim (x_2, y_2) \Leftrightarrow y_1 - x_1 = y_2 - x_2$$

This is an equivalence relation. The equivalence class of a point (a, b) is:

$$\begin{aligned}[(a, b)] &= \{(x, y) \in \mathbb{R}^2 \mid y - x = b - a\} \\ &= \{(x, y) \in \mathbb{R}^2 \mid y = x + (b - a)\}\end{aligned}$$

Each equivalence class is a line with a slope of 1. The quotient set $\mathbb{R}^2/\sim$ is the set of all such lines, which partition the Cartesian plane, as shown in Figure 3.3.

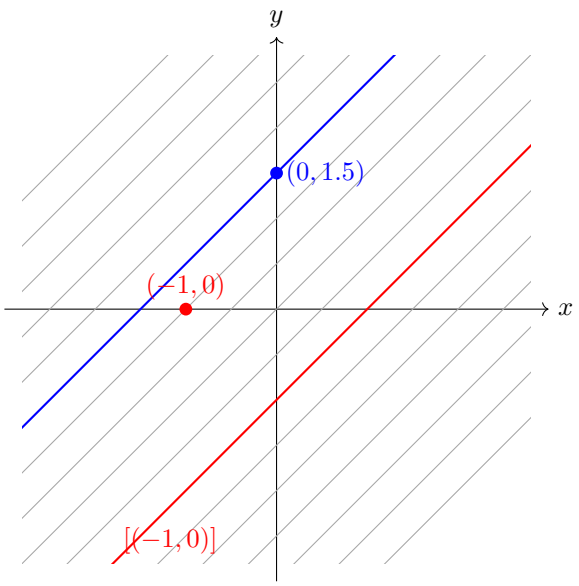


Figure 3.3: The plane $\mathbb{R}^2$ partitioned by the equivalence relation $y - x = c$. Each line is an equivalence class.

3.5 Partitions and Equivalence

The previous section concluded with the observation that the set of equivalence classes partitions the underlying set. We now formalise this relationship, showing that partitions and equivalence relations are two perspectives on the same fundamental structure.

Definition 3.5.1. ***Partition***. A partition of a non-empty set A is a collection of subsets $\mathcal{P} \subseteq \mathcal{P}(A)$ with the following properties:

1. No subset in the partition is empty ($\forall S \in \mathcal{P}, S \neq \emptyset$).
2. The union of all subsets in the partition is the original set A ($\bigcup \mathcal{P} = A$).
3. The subsets in the partition are mutually disjoint ($\forall S_1, S_2 \in \mathcal{P}, (S_1 \neq S_2 \Rightarrow S_1 \cap S_2 = \emptyset)$).

Example 3.5.1. Let $A = \{1, 2, 3, 4, 5, 6\}$. The collection of sets $\mathcal{P} = \{\{1, 5\}, \{2\}, \{3, 4, 6\}\}$ is a partition of A. Each element of A is in exactly one of these subsets. This is visualised in Figure 3.4.

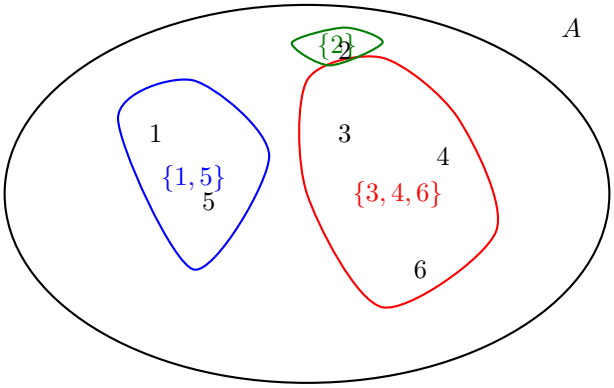


Figure 3.4: A partition of the set $A = \{1, 2, 3, 4, 5, 6\}$ into three disjoint subsets.

Example 3.5.2. Consider the set $\mathbb{R}^2$. For each non-negative real number $r \geq 0$, define C_r to be the circle of radius r centred at the origin:

$$C_r = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$$

The collection $\mathcal{P} = \{C_r \mid r \geq 0\}$ is a partition of $\mathbb{R}^2$. Every point (x, y) in the plane lies on exactly one such circle, namely the one with radius $r = \sqrt{x^2 + y^2}$.

Definition 3.5.2. ***Set of Representatives***. A set $X \subseteq A$ is a set of representatives for the equivalence E (or for the partition S of A) if $\forall C \in S$, $|X \cap C| = 1$.

Note. Notation: $|Y|$ denotes the cardinality (size) of the set Y. For finite sets, this is simply the count of elements. For example, $|\{0, 1, 2\}| = 3$.

The Fundamental Correspondence

We now establish the main result: an equivalence relation induces a partition, and conversely, a partition induces an equivalence relation.

***Theorem* 3.5.1. Equivalence Relations Induce Partitions**. If $\sim$ is an equivalence relation on a non-empty set A, then the quotient set $A/\sim$ is a partition of A.

Proof. Let $\sim$ be an equivalence relation on A. We must verify that the collection of equivalence classes, $A/\sim$, satisfies the three properties of a partition.

1. **Non-empty subsets**: For any class $[a] \in A/\sim$, we know $a \in [a]$ by reflexivity. Thus, no equivalence class is empty.
2. **Union is A**: For any element $a \in A$, we know a belongs to the equivalence class $[a]$, which is an element of $A/\sim$. Therefore, $\bigcup(A/\sim) = A$.
3. **Mutually disjoint**: Let $[a]$ and $[b]$ be two distinct equivalence classes in $A/\sim$. From the properties of equivalence classes, we know that if two classes are not identical, their intersection must be empty. Therefore, the elements of $A/\sim$ are mutually disjoint.

Since all three conditions are met, $A/\sim$ is a partition of A. ■

This theorem shows how to construct a partition from an equivalence relation. The reverse is also true.

***Theorem* 3.5.2. Partitions Induce Equivalence Relations**. If $\mathcal{P}$ is a partition of a non-empty set A, then the relation $\sim_{\mathcal{P}}$ defined by

$$x \sim_{\mathcal{P}} y \Leftrightarrow \exists S \in \mathcal{P}, (x \in S \wedge y \in S)$$

is an equivalence relation on A.

Proof. Let $\mathcal{P}$ be a partition of A, and define $\sim_{\mathcal{P}}$ as the relation where two elements are related if they belong to the same subset in the partition. We verify the three required properties.

1. **Reflexive**: Let $x \in A$. Since $\mathcal{P}$ is a partition, $\bigcup \mathcal{P} = A$, so x must belong to some subset $S \in \mathcal{P}$. Thus, $x \in S$ and $x \in S$, which means $x \sim_{\mathcal{P}} x$.
2. **Symmetric**: Suppose $x \sim_{\mathcal{P}} y$. By definition, there is a set $S \in \mathcal{P}$ such that $x \in S$ and $y \in S$. This statement is symmetric with respect to x and y, so we can also say $y \in S$ and $x \in S$, which implies $y \sim_{\mathcal{P}} x$.
3. **Transitive**: Suppose $x \sim_{\mathcal{P}} y$ and $y \sim_{\mathcal{P}} z$. Then there exists a set $S_1 \in \mathcal{P}$ such that $x, y \in S_1$, and a set $S_2 \in \mathcal{P}$ such that $y, z \in S_2$. Since y is an element of both S_1 and S_2, their intersection is non-empty. As $\mathcal{P}$ is a partition, its subsets are mutually disjoint, so if their intersection is non-empty, they must be the same set: $S_1 = S_2$. Therefore, x, y, and z all belong to the same subset, which implies $x \sim_{\mathcal{P}} z$.

Thus, $\sim_{\mathcal{P}}$ is an equivalence relation on A. ■

These two theorems establish a fundamental duality: every equivalence relation on a set corresponds to a unique partition of that set, and every partition corresponds to a unique equivalence relation. The equivalence classes of the induced relation are precisely the sets in the original partition.

3.6 Exercises

1. **The Kernel Relation.** Let A and B be sets and let f be a function mapping A to B. Define a relation $\sim$ on A by $x \sim y \Leftrightarrow f(x) = f(y)$. Prove that $\sim$ is an equivalence relation on A. If f maps words to their first letter (e.g., $f(\text{"Cat"}) = \text{"C"}$), describe the equivalence class ["Dog"].

2. Let the relation $\sim$ be defined on $\mathbb{R} \setminus \{0\}$ such that for all $x, y \in \mathbb{R} \setminus \{0\}$, $x \sim y$ if and only if $xy > 0$.
   (a) Show that $\sim$ is an equivalence relation.
   (b) Identify the distinct equivalence classes that form the partition of $\mathbb{R} \setminus \{0\}$.

3. Let the relation $\sim$ be defined on $\mathbb{Z}$ such that for all $x, y \in \mathbb{Z}$, $x \sim y$ if and only if $|x| = |y|$.
   (a) Prove that $\sim$ is an equivalence relation.
   (b) Describe the partition of $\mathbb{Z}$ induced by this relation. How many elements are in each equivalence class?

4. Let A be a set. Prove that the given relations on $\mathcal{P}(A)$ are not equivalence relations by identifying which properties fail.
   (a) For $X, Y \in \mathcal{P}(A)$, define $X \sim Y$ if and only if $X \cap Y \neq \emptyset$.
   (b) For $X, Y \in \mathcal{P}(A)$, and a fixed element $a \in A$, define $X \sim Y$ if and only if $a \in X \cap Y$.

5. Let $A = \{1, 2, 3, 4, 5\}$. Consider the partition $\mathcal{P} = \{\{1, 4\}, \{2, 3, 5\}\}$ of A. List all the ordered pairs in the equivalence relation $\sim_{\mathcal{P}}$ induced by this partition.

6. Consider the set $\mathbb{R}^2$. For each of the following collections of subsets, determine if it constitutes a partition of $\mathbb{R}^2$ and justify your answer.
   (a) $\mathcal{P}_1 = \{L_c \mid c \in \mathbb{R}\}$, where $L_c = \{(x, y) \in \mathbb{R}^2 \mid x = c\}$.
   (b) $\mathcal{P}_2 = \{I_n \mid n \in \mathbb{Z}\}$, where $I_n = \{(x, y) \in \mathbb{R}^2 \mid n < y \leq n + 1\}$.
   (c) $\mathcal{P}_3 = \{J_{n,m} \mid n, m \in \mathbb{Z}\}$, where $J_{n,m} = \{(x, y) \in \mathbb{R}^2 \mid n \leq x < n + 1 \wedge m \leq y < m + 1\}$.

7. Let $\sim_1$ and $\sim_2$ be two equivalence relations on a set A.
   (a) Prove that their intersection, $\sim_1 \cap \sim_2$, is also an equivalence relation on A.
   (b) Provide a counterexample to show that their union, $\sim_1 \cup \sim_2$, is not necessarily an equivalence relation.

8. Let R be a binary relation on a set A. Prove the following:
   (a) R is reflexive if and only if its inverse R^{-1} is reflexive.
   (b) R is symmetric if and only if $R = R^{-1}$.
   (c) R is transitive if and only if its inverse R^{-1} is transitive.

9. Let $\sim$ be a symmetric and transitive relation on a set A. Prove that $\sim$ is an equivalence relation if and only if its domain is the entire set, i.e., $\text{dom}(\sim) = A$.

10. Let $\sim$ be an equivalence relation on a set A. Prove that $A = \bigcup_{a \in A} [a]$.

11. Let the relation $\sim$ be defined on $\mathbb{R} \times \mathbb{R}$ such that $(a, b) \sim (x, y)$ if and only if $a^2 + b^2 = x^2 + y^2$.
   (a) Show that $\sim$ is an equivalence relation.
   (b) Describe the partition of the plane induced by $\sim$.

12. A relation R on a set A is said to be *circular* if for all $x, y, z \in A$, $(xRy \wedge yRz) \Rightarrow zRx$. Prove that a relation is an equivalence relation if and only if it is reflexive and circular.

13. Let R and S be relations on A. The symmetric closure of R is defined as the relation S such that $R \subseteq S$, S is symmetric, and for any symmetric relation T on A, if $R \subseteq T$ then $S \subseteq T$. Prove that $R \cup R^{-1}$ is the symmetric closure of R and that the symmetric closure is unique.

3.7 Partial Orders

Equivalence relations capture a notion of sameness, generalising equality. We now consider relations that capture a notion of ordering or hierarchy, generalising the "$\leq$" relation on numbers. To do so, we introduce further properties a relation may possess.

Definition 3.7.1. Let R be a binary relation on a set A.

- R is **irreflexive** if no element is related to itself.

$$\forall x \in A, (x, x) \notin R$$

- R is **asymmetric** if the relation cannot hold in both directions for distinct elements.

$$\forall x, y \in A, ((x, y) \in R \Rightarrow (y, x) \notin R)$$

- R is **antisymmetric** if whenever the relation holds in both directions, the elements must be identical.

$$\forall x, y \in A, (((x, y) \in R \wedge (y, x) \in R) \Rightarrow x = y)$$

Remark. A relation on a non-empty set cannot be both reflexive and irreflexive, nor can it be both symmetric and asymmetric. However, a relation may have neither property. For instance, $R = \{(1, 1)\}$ on $\{1, 2\}$ is neither reflexive (since $(2, 2) \notin R$) nor irreflexive (since $(1, 1) \in R$).

Example 3.7.1. The less-than relation, $<$, on $\mathbb{Z}$ is irreflexive and asymmetric. It is also vacuously antisymmetric: the premise of the implication 'if $x < y$ and $y < x$, then $x = y$' is always false, making the statement true. The less-than-or-equal-to relation, $\leq$, is also antisymmetric, but it is neither irreflexive nor asymmetric since, for example, $3 \leq 3$.

Example 3.7.2. Let $R = \{(1, 2)\}$ be a relation on $\{1, 2\}$. This relation is asymmetric since $(1, 2) \in R$ but $(2, 1) \notin R$. It is also antisymmetric and irreflexive. The relation $S = \{(1, 2), (2, 1)\}$ on the same set is symmetric, but not antisymmetric since $(1, 2) \in S$ and $(2, 1) \in S$ but $1 \neq 2$.

***Theorem* 3.7.1.** A relation R on a set A is antisymmetric if and only if $R \cap R^{-1} \subseteq \mathrm{Id}_A$.

Proof. ($\Rightarrow$) Assume R is antisymmetric. Let $(x, y) \in R \cap R^{-1}$. By definition of intersection and inverse, this means $(x, y) \in R$ and $(y, x) \in R$. Since R is antisymmetric, it must be that $x = y$. Therefore, the pair is of the form (x, x), which is an element of Id_A. Thus, $R \cap R^{-1} \subseteq \mathrm{Id}_A$.

($\Leftarrow$) Assume $R \cap R^{-1} \subseteq \mathrm{Id}_A$. Let $(x, y) \in R$ and $(y, x) \in R$. From $(y, x) \in R$, we have $(x, y) \in R^{-1}$. Therefore, $(x, y) \in R \cap R^{-1}$. By our assumption, this implies $(x, y) \in \mathrm{Id}_A$, which means $x = y$. Thus, R is antisymmetric. ■

With these properties, we can define a new structure that generalises familiar orderings. Such a relation is often denoted by a symbol like $\preceq$.

Definition 3.7.2. ***Partial Order***. A relation $\preceq$ on a set A is a partial order if it is reflexive, antisymmetric, and transitive. The ordered pair $(A, \preceq)$ is called a *partially ordered set*, or poset. For elements $x, y \in A$, the notation $x \prec y$ means $x \preceq y$ but $x \neq y$.

Note. Given a partial order $\preceq$, we also define the following relations:

- $x \succeq y$ if and only if $y \preceq x$.
- $x \succ y$ if and only if $y \prec x$.

Example 3.7.3. The relations $\leq$ and $=$ are partial orders on $\mathbb{R}$. However, $<$ is not a partial order on $\mathbb{R}$ because it is not reflexive.

Example 3.7.4. Divisibility. The divisibility relation '$|$' is a partial order on the set of positive integers $\mathbb{Z}^+$. Let $a, b, c \in \mathbb{Z}^+$.

- **Reflexive**: $a|a$ since $a = a \cdot 1$.
- **Antisymmetric**: If $a|b$ and $b|a$, then $b = ak_1$ and $a = bk_2$ for some $k_1, k_2 \in \mathbb{Z}^+$. This implies $a = (ak_1)k_2 = a(k_1k_2)$. Since $a \neq 0$, we can divide by a to get $1 = k_1k_2$. As k_1 and k_2 are positive integers, it must be that $k_1 = k_2 = 1$, which implies $a = b$.
- **Transitive**: If $a|b$ and $b|c$, then $b = ak_1$ and $c = bk_2$ for some $k_1, k_2 \in \mathbb{Z}^+$. By substitution, $c = (ak_1)k_2 = a(k_1k_2)$, so $a|c$.

Example 3.7.5. The Prefix Order. We can use partial order to define an ordering on strings (something from automata theory), but first we establish the domain of discourse. Let Σ be a finite set of symbols, referred to as an *alphabet*. We construct the set of all strings of length k, denoted Σ^k, by taking the k-fold Cartesian product of Σ with itself. The set of all finite strings over Σ, denoted by Σ^* and known as the *Kleene star* of Σ, is the countable union of these sets:

$$\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$$

Here, Σ^0 contains a unique element of length zero, the empty string ε. For example, if $\Sigma = \{a, b\}$, then $\Sigma^1 = \{a, b\}$, $\Sigma^2 = \Sigma \times \Sigma = \{aa, ab, ba, bb\}$ (Strings of length 2), $\Sigma^3 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$, and $\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$.

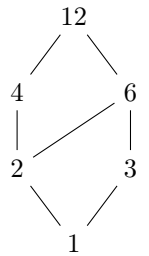
We define the prefix relation $\preceq$ on Σ^* algebraically using concatenation. For strings $s, t \in \Sigma^*$, we say s is a prefix of t if s forms the initial segment of t. Formally:

$$s \preceq t \iff \exists u \in \Sigma^* \text{ such that } t = s \frown u.$$

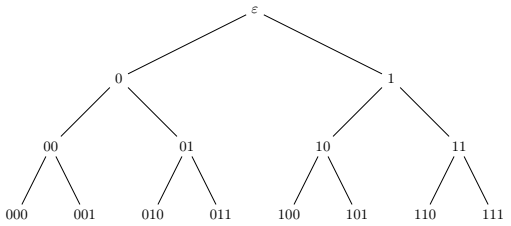
Consider the concrete case where $s = \texttt{ab}$ and $t = \texttt{abcde}$. Since t can be decomposed as $\texttt{ab} \frown \texttt{cde}$, we have $s \preceq t$. Conversely, if $x = \texttt{ab}$ and $y = \texttt{ac}$, there exists no u such that $\texttt{ac} = \texttt{ab} \frown u$; thus $x \not\preceq y$. Since neither is a prefix of the other, x and y are termed incomparable (this concept is properly defined later).

This relation $\preceq$ satisfies the axioms of a partial order. Reflexivity is immediate as $s = s \frown \varepsilon$. Transitivity follows from the associativity of concatenation. Antisymmetry is enforced by the length of the strings: if $s \preceq t$ and $t \preceq s$, their lengths must be equal, implying the suffix u is ε, and hence $s = t$.

To visualise such structures, we employ *Hasse diagrams*. In these diagrams, we represent elements as vertices. An edge is drawn from x up to y if y strictly covers x — that is, $x \prec y$ and there is no element z such that $x \prec z \prec y$. This method produces a transitive reduction of the graph, removing redundant lines to reveal the hierarchical structure. In Figure 3.5, we present two such diagrams: the prefix order on binary strings (forming an infinite tree rooted at ε) and the divisibility relation on the set $D_{12} = \{1, 2, 3, 4, 6, 12\}$.



(a) The poset $(D_{12}, |)$.



(b) A Hasse diagram for the prefix partial order on $\{0, 1\}^*$.

Figure 3.5: Visualising partial orders using Hasse diagrams.

The partial order $\leq$ on $\mathbb{R}$ has the property that for any two numbers x, y, either $x \leq y$ or $y \leq x$. A partial order with this property is called a total order (or linear order). This property does not hold for all posets.

Example 3.7.6. Order on Function Spaces. Let A be a set and let $(B, \preceq_B)$ be a poset. The set of all functions from A to B, denoted $\text{Funct}(A, B)$, can be equipped with a partial order $\preceq$ defined pointwise: for any two functions $f, g \in \text{Funct}(A, B)$,

$$f \preceq g \Leftrightarrow \forall x \in A, f(x) \preceq_B g(x)$$

This relation is a partial order on Funct(A, B). Note that even if $(B, \preceq_B)$ is a totally ordered set, Funct(A, B) is generally not totally ordered unless A is a singleton.

***Theorem* 3.7.2. Weak-Trichotomy Law**. If $\preceq$ is a partial order on a set A, then for any $x, y \in A$, at most one of the following is true: $x \prec y$, $y \prec x$, or $x = y$.

Proof. Let $(A, \preceq)$ be a poset and let $x, y \in A$. Suppose $x \prec y$. This means $x \preceq y$ and $x \neq y$. If we also had $y \prec x$, it would mean $y \preceq x$. By antisymmetry, $x \preceq y$ and $y \preceq x$ implies $x = y$, which contradicts our assumption that $x \neq y$. Thus, $y \prec x$ cannot be true. By definition, if $x = y$, then neither $x \prec y$ nor $y \prec x$ can be true. The cases are mutually exclusive. ■

Example 3.7.7. Subset Relation Let A be a set. The subset relation on its power set, $\mathcal{P}(A)$, is the relation $\{(X, Y) \in \mathcal{P}(A) \times \mathcal{P}(A) \mid X \subseteq Y\}$. We show that $(\mathcal{P}(A), \subseteq)$ is a poset. Let $X, Y, Z \in \mathcal{P}(A)$.

- **Reflexive**: $X \subseteq X$ for any set X.
- **Antisymmetric**: If $X \subseteq Y$ and $Y \subseteq X$, then by the Axiom of Extensionality, $X = Y$.
- **Transitive**: If $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$.

This poset aligns with the Weak-Trichotomy Law. For instance, if $X \subset Y$, then it cannot be that $Y \subset X$.

Before we finish off this section one final theorem.

***Theorem* 3.7.3. Correspondence of Partial and Strict Orders**. Let A be a set.

(i) If $\preceq$ is a partial order on A, then the relation $\prec$ defined by $x \prec y \Leftrightarrow (x \preceq y \wedge x \neq y)$ is a strict order.
(ii) If $\prec$ is a strict order on A, then the relation $\preceq$ defined by $x \preceq y \Leftrightarrow (x \prec y \vee x = y)$ is a partial order.

Proof. (i) Let $\preceq$ be a partial order.

- **Irreflexive**: Suppose $x \prec x$. This means $x \preceq x$ and $x \neq x$, a contradiction. Thus $\prec$ is irreflexive.
- **Asymmetric**: Suppose $x \prec y$ and $y \prec x$. This implies $x \preceq y$ and $y \preceq x$. By antisymmetry of $\preceq$, we must have $x = y$. This contradicts $x \prec y \Rightarrow x \neq y$. Thus $\prec$ is asymmetric.
- **Transitive**: Suppose $x \prec y$ and $y \prec z$. This means $x \preceq y$, $x \neq y$, $y \preceq z$, and $y \neq z$. By transitivity of $\preceq$, $x \preceq z$. If $x = z$, then from $x \preceq y$ we have $z \preceq y$. With $y \preceq z$, antisymmetry implies $y = z$, which is a contradiction. Thus $x \neq z$. Therefore, $x \prec z$.

(ii) The proof for the second part is analogous and is left as an exercise for the reader. ■

Bounds

Let $(A, \preceq)$ be a poset. If elements l and l' are both least elements of A, then by definition, $l \preceq a$ and $l' \preceq a$ for all $a \in A$. In particular, $l \preceq l'$ (since $l' \in A$) and $l' \preceq l$ (since $l \in A$). By antisymmetry, $l = l'$. A similar argument holds for the greatest element. Thus, if they exist, the least and greatest elements of a poset are unique.

Definition 3.7.3. ***Least and Greatest Elements***. Let $(A, \preceq)$ be a poset. An element $l \in A$ is the *least element* of A if $l \preceq x$ for all $x \in A$. An element $g \in A$ is the *greatest element* of A if $x \preceq g$ for all $x \in A$.

Example 3.7.8. For the poset $(\mathcal{P}(A), \subseteq)$, the empty set $\emptyset$ is the least element and the set A itself is the greatest element. For the prefix order on Σ^* (**??**), the empty string ε is the least element, but there is no greatest element. With respect to $\leq$, the set $\mathbb{Z}^+$ has a least element (1) but no greatest element. The set $\mathbb{Z}^-$ has a greatest element (-1) but no least element.

For subsets of a poset, we can define a more general notion of bounding elements.

Definition 3.7.4. ***Upper and Lower Bounds***. Let $(A, \preceq)$ be a poset and let $S \subseteq A$.

- An element $u \in A$ is an upper bound of S if $s \preceq u$ for all $s \in S$.
- The element $u_0 \in A$ is the least upper bound (or **supremum**) of S if it is an upper bound and for any other upper bound u' of S, we have $u_0 \preceq u'$.
- An element $l \in A$ is a lower bound of S if $l \preceq s$ for all $s \in S$.
- The element $l_0 \in A$ is the greatest lower bound (or **infimum**) of S if it is a lower bound and for any other lower bound l' of S, we have $l' \preceq l_0$.

Note. The supremum of a two-element set $\{x, y\}$ is often denoted $x \vee y$ (read "x join y"), and the infimum is denoted $x \wedge y$ (read "x meet y").

Definition 3.7.5. ***Maximum and Minimum of a Subset***. Let $(A, \preceq)$ be a poset and let $S \subseteq A$.

- An element $m \in S$ is the maximum of S, denoted $\max(S)$, if it is an upper bound of S.
- An element $m \in S$ is the minimum of S, denoted $\min(S)$, if it is a lower bound of S.

By antisymmetry, the maximum and minimum of a set, if they exist, are unique.

Remark. It is crucial to distinguish the supremum from the maximum. If $\sup(S)$ exists and is an element of S, then $\sup(S) = \max(S)$. However, a set may have a supremum that is not in the set itself, in which case it has no maximum. A symmetric statement holds for the infimum and minimum.

Example 3.7.9. Consider the interval $(3, 5)$ as a subset of the poset $(\mathbb{R}, \leq)$. The set of upper bounds is $[5, \infty)$, and the least upper bound is 5. The set of lower bounds is $(-\infty, 3]$, and the greatest lower bound is 3. Note that neither the supremum nor the infimum are elements of the set $(3, 5)$.

Example 3.7.10. Let $\mathcal{F}$ be a collection of subsets of a set A, so $\mathcal{F} \subseteq \mathcal{P}(A)$. Consider $\mathcal{F}$ as a subset of the poset $(\mathcal{P}(A), \subseteq)$. The union $\bigcup \mathcal{F}$ is an upper bound for $\mathcal{F}$, since for any set $S \in \mathcal{F}$, we have $S \subseteq \bigcup \mathcal{F}$. To show it is the least upper bound, let U be any other upper bound of $\mathcal{F}$. This means $S \subseteq U$ for all $S \in \mathcal{F}$. Let $x \in \bigcup \mathcal{F}$. By definition, there exists some $S_0 \in \mathcal{F}$ such that $x \in S_0$. Since U is an upper bound, $S_0 \subseteq U$, which implies $x \in U$. As x was arbitrary, we conclude $\bigcup \mathcal{F} \subseteq U$. Therefore, $\bigcup \mathcal{F}$ is the supremum of $\mathcal{F}$.

An analogous argument shows that the intersection $\bigcap \mathcal{F}$ is the greatest lower bound (infimum) of $\mathcal{F}$.

***Theorem* 3.7.4. Basic Properties of Infimum and Supremum**. Let $(A, \leq)$ be an ordered set and $B \subseteq A$.

(i) B has at most one infimum.
(ii) If b is the least element of B, then b is the infimum of B.
(iii) If $b \in B$ is the infimum of B, then b is the least element of B.

Proof. (i) Follows from uniqueness of maximal lower bound. (ii) Least b is a lower bound; any other lower bound x satisfies $x \leq b$ (since $b \in B$), so b greatest. (iii) Infimum $b \in B$ is a lower bound, hence least in B. ■

Remark. This is Symmetric for supremum and thus left as an exercise.

3.8 Structure within Posets

In a partially ordered set, it is not required for every pair of elements to be related. This leads to a rich structure of comparable and incomparable elements, which we can use to classify posets and their subsets.

Comparability, Maxima, and Minima

We begin by formalising the notion of two elements being related within a poset.

Definition 3.8.1. ***Comparable Elements***. Let $(A, \preceq)$ be a poset. Two elements $x, y \in A$ are comparable if either $x \preceq y$ or $y \preceq x$. If they are not comparable, they are incomparable.

Example 3.8.1. In the poset of strings $(\{0,1\}^*, \preceq)$ from our example above with the prefix order, the strings '01' and '0110' are comparable because '01' is a prefix of '0110'. However, the strings '010' and '011' are incomparable, as neither is a prefix of the other.

Example 3.8.2. In the poset $(\mathbb{Z}^+, |)$, the integers 4 and 12 are comparable because $4|12$. The integers 5 and 7 are incomparable because $5 \nmid 7$ and $7 \nmid 5$.

This distinction allows us to refine the concepts of least and greatest elements. A poset may not have a single "smallest" element, but it can have elements that have nothing smaller than them.

Definition 3.8.2. ***Minimal and Maximal Elements***. Let $(A, \preceq)$ be a poset and let $m \in A$.

- m is a minimal element of A if there is no element $x \in A$ such that $x \prec m$.
- m is a maximal element of A if there is no element $x \in A$ such that $m \prec x$.

Note. It is crucial to distinguish a *maximal* element from a *maximum* (or greatest) element. A maximum element must be comparable to and greater than every other element in the set. A maximal element simply has no element strictly greater than it. While every greatest element is maximal, the converse is not true. For example, in the poset of non-empty, proper subsets of $\{1,2,3\}$ ordered by $\subseteq$, the sets $\{1,2\}$ and $\{1,3\}$ are both maximal, but neither is a greatest element because they are incomparable. A symmetric distinction applies to minimal and least elements.

Remark. Every least element is minimal and every greatest element is maximal. However, the converse is not true. A poset can have multiple minimal or maximal elements, but it can have at most one least and at most one greatest element.

Example 3.8.3. Let $S = \mathcal{P}(\{1,2,3\}) \setminus \{\emptyset, \{1,2,3\}\}$ be the set of all non-empty proper subsets of $\{1,2,3\}$, ordered by $\subseteq$.

- The minimal elements are $\{\{1\},\{2\},\{3\}\}$. None of these can have a subset within S that is smaller. There is no least element, as these three are incomparable.
- The maximal elements are $\{\{1,2\},\{1,3\},\{2,3\}\}$. None of these is a subset of any other element in S. There is no greatest element.

Example 3.8.4. Consider the set $A = \mathbb{Z}^+ \setminus \{1\} = \{2,3,4,\dots\}$ with the divisibility relation '|'. The minimal elements of this poset are precisely the prime numbers, since no prime can be divided by any other element in A. This poset has no maximal elements.

***Theorem* 3.8.1. Basic Properties of Least and Minimal Elements**. Let $(A, \leq)$ be an ordered set and $B \subseteq A$.

(i) B has at most one least element.
(ii) The least element of B (if it exists) is also minimal.
(iii) If B is a chain, every minimal element of B is also least.

Proof. (i) If b, b' are least, then $b \leq b'$ and $b' \leq b$, so $b = b'$ by antisymmetry. (ii) Let b be least. For $x \in B$ with $x \leq b$, also $b \leq x$, so $x = b$ by antisymmetry; b minimal. (iii) Let b be minimal. For $x \in B$, since chain, $x \leq b$ or $b \leq x$. If $x \leq b$, then $x = b$ (minimal), so $b \leq x$ always. ■

Note. Analogous results hold for greatest and maximal elements, and thus left as an exercise.

Chains and Antichains

The properties of comparability can be extended to entire subsets of a poset.

Definition 3.8.3. ***Chain***. A subset C of a poset $(A, \preceq)$ is a chain if every pair of elements in C is comparable.

Example 3.8.5.

- In $(\mathbb{Z}, \leq)$, any subset of integers, such as the set of even numbers $\{\ldots, -2, 0, 2, \ldots\}$, forms a chain.
- In $(\mathcal{P}(\mathbb{N}), \subseteq)$, the collection $\mathcal{C} = \{\emptyset, \{0\}, \{0,1\}, \{0,1,2\}, \ldots\}$ is a chain because for any two sets in $\mathcal{C}$, one is a subset of the other.
- In $(\mathbb{Z}^+, |)$, the set of powers of two, $\{1, 2, 4, 8, 16, \ldots\}$, is a chain.

Some posets have the special property that the entire set is a chain.

Definition 3.8.4. ***Linearly Ordered Set***. A poset $(A, \preceq)$ is a linearly ordered set (or *totally ordered set*) if A itself is a chain. The relation $\preceq$ is then called a linear order (or *total order*).

For such sets, the Weak-Trichotomy Law can be strengthened.

***Theorem* 3.8.2. Trichotomy Law**. If $\preceq$ is a linear order on a set A, then for any two elements $x, y \in A$, exactly one of the following holds:

$$x \prec y, \quad y \prec x, \quad \text{or} \quad x = y$$

The opposite of a chain is a set where no two distinct elements are comparable.

Definition 3.8.5. ***Antichain***. A subset S of a poset $(A, \preceq)$ is an antichain if every pair of distinct elements in S is incomparable.

Example 3.8.6. In the poset $(\mathcal{P}(\{1,2,3,4\}), \subseteq)$:

- The set of all two-element subsets, $\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}$, is an antichain.
- The set $\{\{1\}, \{2,3\}, \{4\}\}$ is an antichain.

In the poset $(\mathbb{Z}^+, |)$, the set of prime numbers is an antichain.

Monotone Functions

A function between two posets is of particular interest if it preserves the order structure.

Definition 3.8.6. ***Monotone Function***. Let $(A, \preceq_A)$ and $(B, \preceq_B)$ be two posets. A function $f : A \to B$ is:

- **increasing** (or order-preserving) if for all $x, y \in A$, $x \preceq_A y \Rightarrow f(x) \preceq_B f(y)$.
- **decreasing** (or order-reversing) if for all $x, y \in A$, $x \preceq_A y \Rightarrow f(y) \preceq_B f(x)$.
- **strictly increasing** if for all $x, y \in A$, $x \prec_A y \Rightarrow f(x) \prec_B f(y)$.
- **monotone** if it is either increasing or decreasing.

Example 3.8.7. Let S be a set. The function $f : \mathcal{P}(S) \to \mathcal{P}(S)$ defined by $f(X) = S \setminus X$ (the complement) is a decreasing function with respect to the subset order $\subseteq$, since $X \subseteq Y \Rightarrow S \setminus Y \subseteq S \setminus X$.

3.9 Exercises

1. Let $D_{30} = \{1, 2, 3, 5, 6, 10, 15, 30\}$ be the set of positive divisors of 30.

(a) Draw the Hasse diagram for the poset $(D_{30}, |)$, where '|' denotes the divisibility relation.
(b) Identify all minimal, maximal, least, and greatest elements of this poset.
(c) Find a chain of length 4.
(d) Find an antichain of size 3.

2. Let $(A, \preceq_A)$ and $(B, \preceq_B)$ be posets. The *product order* $\preceq$ on $A \times B$ is defined by:

$$(a_1, b_1) \preceq (a_2, b_2) \Leftrightarrow (a_1 \preceq_A a_2 \text{ and } b_1 \preceq_B b_2)$$

Prove that $(A \times B, \preceq)$ is a poset.

3. Let $(\mathbb{Z}, \leq)$ be the set of integers with the standard order. The *lexicographical order* $\preceq_{lex}$ on $\mathbb{Z}^2$ is defined by:

$$(a, b) \preceq_{lex} (c, d) \Leftrightarrow (a < c) \text{ or } (a = c \text{ and } b \leq d)$$

Prove that $\preceq_{lex}$ is a linear order on $\mathbb{Z}^2$.

4. Let A be a set with at least two elements. Prove that the poset $(\mathcal{P}(A), \subseteq)$ is not a linear order by providing an example of two incomparable elements.

5. Let S be the set of all functions from $\{a, b\}$ to $\{0, 1\}$. Define a relation $\preceq$ on S by $f \preceq g$ if $f(x) \leq g(x)$ for all $x \in \{a, b\}$.

(a) Prove that $(S, \preceq)$ is a poset.
(b) Draw the Hasse diagram for this poset. Is it a linear order?

6. Let R be a binary relation on a set A. Prove that R is asymmetric if and only if $R \cap R^{-1} = \emptyset$. Compare this to the property that R is antisymmetric if and only if $R \cap R^{-1} \subseteq \mathrm{Id}_A$.

7. Let A be a set. Prove that if $\mathcal{F} \subseteq \mathcal{P}(A) \setminus \{\emptyset\}$ is a collection of mutually disjoint sets, then $\mathcal{F}$ is an antichain with respect to $\subseteq$. Provide a counterexample to show that the converse is false (i.e., find an antichain where the sets are not disjoint).

8. In the poset $(\mathbb{R}, \leq)$, find the supremum and infimum (if they exist) of the following sets. State whether these bounds are elements of the set itself.

(a) $S_1 = \{\frac{n}{n+1} \mid n \in \mathbb{Z}^+\}$.
(b) $S_2 = \{x \in \mathbb{Q} \mid x^2 < 2\}$.

9. Let $A = \{1, 2, 3\}$. Provide an example of a relation on A that is:

(a) Transitive and reflexive, but not antisymmetric.
(b) Transitive and antisymmetric, but not reflexive.

10. Well-Ordering and Infinite Chains. A linearly ordered set $(A, \preceq)$ is said to be *well-ordered* if every non-empty subset of A contains a least element. Prove that a linearly ordered set $(A, \preceq)$ is well-ordered if and only if it contains no infinite descending chain (a sequence $x_1, x_2, x_3, \ldots$ such that $x_{n+1} \prec x_n$ for all $n \in \mathbb{Z}^+$).

(a) ($\Rightarrow$) Assume $(A, \preceq)$ is well-ordered. Prove by contradiction that it cannot contain an infinite descending chain. (Hint: consider the set of elements in such a chain.)
(b) ($\Leftarrow$) Assume $(A, \preceq)$ is a linear order with no infinite descending chains. Let S be any non-empty subset of A. Prove by contradiction that S must have a least element. (Hint: if S has no least element, construct an infinite descending chain.)

11. Lattices. A poset $(A, \preceq)$ is a lattice if for every pair of elements $x, y \in A$, the set $\{x, y\}$ has both a least upper bound (supremum, denoted $x \vee y$) and a greatest lower bound (infimum, denoted $x \wedge y$).

(a) Show that for any set S, the poset $(\mathcal{P}(S), \subseteq)$ is a lattice. What familiar set operations correspond to $X \vee Y$ and $X \wedge Y$?
(b) Show that every linearly ordered set is a lattice. What are $x \vee y$ and $x \wedge y$ in this case?
(c) Show that the poset $(\{1, 2, 3, 4, 5, 6\}, |)$ is not a lattice by finding a pair of elements that lacks a supremum or an infimum.

Chapter 4

Functions

The concept of a function, familiar from algebra and calculus as a rule that assigns a unique output to each input, is formalised in set theory as a specific type of relation. This formalisation allows us to rigorously define and analyse the properties of functions, which are the fundamental building blocks for nearly all areas of mathematics.

4.1 The Definition of a Function

The defining characteristic of a function is that each input is associated with exactly one output. This intuition, often visualised with the vertical line test for graphs in the Cartesian plane, is captured precisely within the language of relations.

Definition 4.1.1. ***Function***. A binary relation f is a function if for every element in its domain, there is a unique corresponding element in its range. Formally, for any x, y_1, y_2,

$$((x, y_1) \in f \land (x, y_2) \in f) \Rightarrow y_1 = y_2$$

A function is *n-ary* if its domain is a Cartesian product of n sets. A function is unary if $n = 1$ and binary if $n = 2$.

In other words, a function is a relation that does not contain two distinct ordered pairs with the same first coordinate.

Example 4.1.1.

- The relation $f = \{(1, 2), (4, 5), (6, 5)\}$ is a function. The first coordinate 1 maps only to 2, 4 maps only to 5, and 6 maps only to 5. It is permissible for different inputs to map to the same output.
- The relation $R = \{(1, 2), (1, 5), (6, 5)\}$ is not a function because the input 1 is associated with two different outputs, 2 and 5.
- The empty set, $\emptyset$, is a function. The condition in the definition is vacuously true, as there are no elements (x, y_1) or (x, y_2) in $\emptyset$ to test.

Remark. In many contexts, especially within set theory, a function is identified with its graph (the set of ordered pairs). However, in other areas of mathematics, it is common to define a function as an ordered triple (A, f, B), where A is the domain, B is the codomain, and $f \subseteq A \times B$ is the graph. This alternative definition explicitly includes the codomain as part of the function's identity.

The standard arithmetic operations can be formalised as functions. For example, addition on the integers is the binary function $+ \subseteq (\mathbb{Z} \times \mathbb{Z}) \times \mathbb{Z}$ defined by:

$$+ = \{((x, y), z) \mid x, y, z \in \mathbb{Z} \land z = x + y\}$$

Notation 4.1.1. Other notations for the function F include:

$$\langle F(a) \mid a \in A\rangle, \quad \langle F_a \mid a \in A\rangle, \quad \langle F_a\rangle_{a\in A}.$$

The range of F can be denoted by $\{F(a) \mid a \in A\}$ or $\{F_a\}_{a\in A}$.

Function Notation and Terminology

Since the output for any given input in a function's domain is unique, we can simplify our notation.

Definition 4.1.2. ***Function Value***. Let f be a function. For any element $x \in \text{dom}(f)$, there exists a unique element y such that $(x, y) \in f$. We define $f(x) = y$. If $x \notin \text{dom}(f)$, we say that $f(x)$ is *undefined*. We call $f(x)$ the value of the function f at x.

The statement $y = f(x)$ is equivalent to $(x, y) \in f$. When discussing a function, we must specify the set of allowed inputs (the domain) and a set that contains all possible outputs (the codomain).

Definition 4.1.3. ***Domain, Codomain, and Range***. Let f be a function.

- The **domain** of f, $\text{dom}(f)$, is the set of all first coordinates.
- A **codomain** of f is any set B such that $\text{ran}(f) \subseteq B$.
- The **range** of f, $\text{ran}(f)$, is the set of all second coordinates.

We write $f : A \to B$ to denote that f is a function with domain A and codomain B. A function is also called a *map* or *mapping*.

Note. If g is also a function with domain A and codomain B, we can use the abbreviation $f, g : A \to B$.

Remark. The range is a property of the function itself, whereas the codomain is part of the function's definition and can be any superset of the range.

If $y = f(x)$, we say that y is the **image** of x under f, and x is a **pre-image** of y. The set of all images of elements in a subset $S \subseteq A$ is denoted $f[S]$, consistent with the notation for relations. The set of all pre-images of a set $T \subseteq B$ is the inverse image, $f^{-1}[T]$. This relationship is illustrated in Figure 4.1.

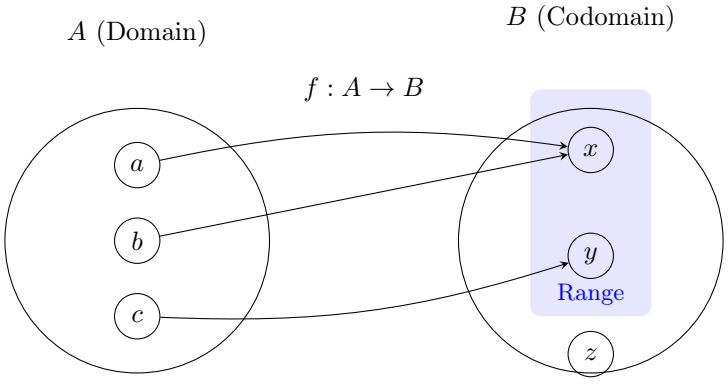


Figure 4.1: A function f mapping elements from its domain A to its codomain B. The range of f is $\{x, y\}$, a subset of B. Here, $f(a) = x$, so x is the image of a. Both a and b are pre-images of x.

Example 4.1.2. Let A be any set. The identity relation Id_A is a function, called the identity map on A. For any $x \in A$, $\text{Id}_A(x) = x$.

Example 4.1.3. Greatest Integer FunctionLet $x \in \mathbb{R}$. The greatest integer function, denoted $\lfloor x \rfloor$, is defined as the greatest integer less than or equal to x. For example, $\lfloor 5 \rfloor = 5$, $\lfloor 1.4 \rfloor = 1$, and $\lfloor -3.4 \rfloor = -4$. This defines a function $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$.

Example 4.1.4. Let A and B be non-empty sets.

- A **constant function** is a function $f : A \to B$ such that for a fixed element $c \in B$, we have $f(x) = c$ for all $x \in A$.
- If $A \subseteq B$, the **inclusion map** is the function $i : A \to B$ defined by $i(x) = x$. Note that $i = \mathrm{Id}_A$ as sets of pairs, but they are distinct as functions if $A \neq B$ because their codomains differ.
- The **characteristic function** of a subset $S \subseteq A$ is the function $\chi_S : A \to \{0, 1\}$ defined by:

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

- For a Cartesian product $A \times B$, the **projections** are the functions $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$ defined by $\pi_1(a, b) = a$ and $\pi_2(a, b) = b$.

Well-Defined Functions

When defining a function with a rule, particularly on a set of equivalence classes, we must ensure that the rule produces a unique output regardless of the representation of the input. A rule that satisfies this condition is said to be **well-defined**. Proving a function is well-defined is equivalent to proving it is a function according to our definition.

Example 4.1.5. Let $n, m \in \mathbb{Z}^+$ such that $m \mid n$. Define $\phi : \mathbb{Z}_n \to \mathbb{Z}_m$ by $\phi([a]_n) = [a]_m$. To show ϕ is well-defined, assume $[a]_n = [b]_n$ for some integers a, b.

$$\begin{aligned} [a]_n = [b]_n &\Rightarrow n \mid (a - b) \\ &\Rightarrow a - b = nk \quad \text{for some } k \in \mathbb{Z} \end{aligned}$$

Since $m \mid n$, we know $n = mj$ for some $j \in \mathbb{Z}$. Substituting gives:

$$a - b = (mj)k = m(jk)$$

Since jk is an integer, this implies $m \mid (a - b)$, which means $[a]_m = [b]_m$. The output is independent of the representative chosen, and ϕ is a well-defined function.

Example 4.1.6. Define $\psi : \mathbb{Z}_3 \to \mathbb{Z}_2$ by $\psi([a]_3) = [a]_2$. This rule is not well-defined. Consider the equivalence class $[1]_3$. We have $[1]_3 = [4]_3$ since $3 \mid (4 - 1)$. According to the rule, $\psi([1]_3) = [1]_2$. However, using the representative 4, the rule gives $\psi([4]_3) = [4]_2 = [0]_2$. Since $[1]_2 \neq [0]_2$, the same input $[1]_3$ yields two different outputs. Therefore, ψ is not a function.

Equality and the Space of Functions

As functions are sets, two functions are equal if and only if they contain the same set of ordered pairs. This leads to a more practical criterion for function equality.

Definition 4.1.4. ***Set of Functions***. If A and B are sets, the set of all functions from A to B is denoted by B^A.

$$B^A = \{f \mid f \text{ is a function } f : A \to B\}$$

Example 4.1.7. Let $A = \{1, 2\}$ and $B = \{0, 1\}$. Then B^A is the set of all functions from A to B. A function $f : A \to B$ must assign to each element of A either 0 or 1. There are four such functions, which we can describe as sets of ordered pairs:

$$f_1 = \{(1, 0), (2, 0)\}, \quad f_2 = \{(1, 0), (2, 1)\}, \quad f_3 = \{(1, 1), (2, 0)\}, \quad f_4 = \{(1, 1), (2, 1)\}$$

So in this case, $B^A = \{f_1, f_2, f_3, f_4\}$.

Example 4.1.8. A sequence of real numbers $(a_0, a_1, a_2, \ldots)$ can be seen as a function $f : \mathbb{N} \to \mathbb{R}$ where $f(n) = a_n$. Thus, the set of all real sequences is the function space $\mathbb{R}^{\mathbb{N}}$. For example, the sequence $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ corresponds to the function $f : \mathbb{Z}^+ \to \mathbb{R}$ defined by $f(n) = 1/n$.

***Proposition* 4.1.1.** If A and B are sets, then B^A exists.

Proof. If f is a function from A into B, then $f \subseteq A \times B$, so $f \in \mathcal{P}(A \times B)$. ■

***Theorem* 4.1.1. Equality of Functions**. Two functions $f : A \to B$ and $g : C \to D$ are equal if and only if they have the same domain ($A = C$) and for every element x in that domain, $f(x) = g(x)$.

Proof. ($\Rightarrow$) Assume $f = g$. Since they are the same set of ordered pairs, their domains must be identical, so $A = C$. For any $x \in A$, let $y = f(x)$. This means $(x, y) \in f$. Since $f = g$, we also have $(x, y) \in g$, which means $g(x) = y$. Thus, $f(x) = g(x)$.

($\Leftarrow$) Assume $A = C$ and $f(x) = g(x)$ for all $x \in A$. Let $(x, y) \in f$. By definition, $y = f(x)$. By our assumption, $g(x) = f(x) = y$, which implies $(x, y) \in g$. Thus $f \subseteq g$. A symmetric argument shows $g \subseteq f$, so $f = g$. ■

Example 4.1.9. Let $f, g : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = (x - 3)^2 + 2$ and $g(x) = x^2 - 6x + 11$. For any $x \in \mathbb{R}$,

$$f(x) = (x - 3)^2 + 2 = (x^2 - 6x + 9) + 2 = x^2 - 6x + 11 = g(x)$$

Since they have the same domain and their values agree for all inputs, $f = g$.

Example 4.1.10. Let $f, g : \mathbb{Z} \to \mathbb{Z}_6$ be defined by $f(n) = [n]_6$ and $g(n) = [n + 12]_6$. We claim that $f = g$. For any $a \in \mathbb{Z}$, we have $g(a) = [a + 12]_6$. Since $6 \mid 12$, it follows that $6 \mid ((a + 12) - a)$, which means $[a + 12]_6 = [a]_6$. Thus, $g(a) = [a]_6 = f(a)$ for all $a \in \mathbb{Z}$. As the domains are identical and the values agree for every input, $f = g$.

4.2 Operations on Functions

With a formal definition of a function, we can now define standard operations such as composition, restriction, and extension, which allow for the construction and manipulation of functions.

Composition of Functions

The composition of relations is a particularly meaningful operation when the relations are functions. It represents the application of one function followed by another.

***Theorem* 4.2.1.** If $f : A \to B$ and $g : C \to D$ are functions such that $\text{ran}(f) \subseteq C$, then the composition $g \circ f$ is a function from A to D, and for all $x \in A$, $(g \circ f)(x) = g(f(x))$.

Proof. First, we show that $g \circ f$ is a function. Let $(x, z_1) \in g \circ f$ and $(x, z_2) \in g \circ f$. By the definition of composition, there must exist $y_1, y_2 \in \text{ran}(f)$ such that:

- $(x, y_1) \in f$ and $(y_1, z_1) \in g$
- $(x, y_2) \in f$ and $(y_2, z_2) \in g$

Since f is a function, $(x, y_1) \in f$ and $(x, y_2) \in f$ implies $y_1 = y_2$. Now, since $y_1 = y_2$ and g is a function, $(y_1, z_1) \in g$ and $(y_1, z_2) \in g$ implies $z_1 = z_2$. Therefore, $g \circ f$ satisfies the definition of a function.

The domain of $g \circ f$ is the set of all x for which the composition is defined. For any $x \in A$, $f(x)$ exists and is in $\text{ran}(f)$. Since $\text{ran}(f) \subseteq C = \text{dom}(g)$, $g(f(x))$ is also defined. Thus, the domain of $g \circ f$ is A. The range of $g \circ f$ is a subset of the range of g, which is a subset of D. Therefore, $g \circ f : A \to D$.

Finally, for any $x \in A$, let $y = f(x)$. Then $(x, y) \in f$. Since $y \in \text{ran}(f) \subseteq C$, we have $(y, g(y)) \in g$. By definition of composition, $(x, g(y)) \in g \circ f$. Thus, $(g \circ f)(x) = g(y) = g(f(x))$. ■

Example 4.2.1. Let $f : \mathbb{Z} \to \mathbb{Z}$ be $f(x) = x+2$ and $g : \mathbb{Z} \to \mathbb{Z}_5$ be $g(y) = [y^2]_5$. Since $\text{ran}(f) = \mathbb{Z} = \text{dom}(g)$, the composition $g \circ f$ is well-defined. For any $x \in \mathbb{Z}$,

$$(g \circ f)(x) = g(f(x)) = g(x + 2) = [(x + 2)^2]_5 = [x^2 + 4x + 4]_5$$

Example 4.2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x - 1$ and $g : \mathbb{R}^+ \to \mathbb{R}$ be $g(x) = \ln(x)$. The range of f is $\mathbb{R}$, which is not a subset of the domain of g, $\mathbb{R}^+$. The composition $g \circ f$ is not defined on all of $\mathbb{R}$. For instance, $(g \circ f)(0) = g(f(0)) = g(-1)$, which is undefined.

Commutative Diagrams

Relationships between functions, particularly those involving composition, can be visualised using commutative diagrams. In such a diagram, sets are represented by vertices and functions by arrows. A diagram is said to **commute** if following any two paths of arrows from one vertex to another yields the same result via composition.

For example, given functions $f : A \to B$, $g : B \to C$, and $h : A \to C$, the diagram in Figure 4.2 commutes if and only if $h = g \circ f$.

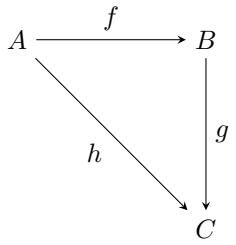


Figure 4.2: A commutative diagram illustrating $h = g \circ f$.

Restrictions and Extensions

It is often useful to consider a function's behaviour on a smaller domain or to define a larger function that agrees with a smaller one.

Definition 4.2.1. ***Restriction and Extension***. Let $f : A \to B$ be a function and let $S \subseteq A$.

- The **restriction** of f to S, denoted $f|_S$, is the function $f|_S : S \to B$ defined by $f|_S(x) = f(x)$ for all $x \in S$. Formally, $f|_S = f \cap (S \times B)$.
- A function $g : C \to D$ is an **extension** of f if $A \subseteq C$ and $f = g|_A$.

Example 4.2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = |x|$. The function $g : [0, \infty) \to \mathbb{R}$ defined by $g(x) = x$ is equal to the restriction of f to the non-negative reals, i.e., $g = f|_{[0,\infty)}$. Conversely, f is an extension of g.

***Theorem* 4.2.2.** Let $f : A \to B$ be a function and let $S, T \subseteq A$. Then $f|_{S \cup T} = (f|_S) \cup (f|_T)$.

Proof. We show the sets are equal by showing an element is in one if and only if it is in the other.

$$\begin{aligned}
(x, y) \in f|_{S \cup T} &\Leftrightarrow x \in S \cup T \wedge y = f(x) \\
&\Leftrightarrow (x \in S \vee x \in T) \wedge y = f(x) \\
&\Leftrightarrow (x \in S \wedge y = f(x)) \vee (x \in T \wedge y = f(x)) \\
&\Leftrightarrow (x, y) \in f|_S \vee (x, y) \in f|_T \\
&\Leftrightarrow (x, y) \in (f|_S) \cup (f|_T)
\end{aligned}$$

■

Compatible Functions and Function Unions

When constructing a new function from several existing ones, it is essential that their definitions do not conflict. This leads to the idea of compatibility.

Definition 4.2.2. ***Compatible Functions***. Two functions f and g are compatible if their values agree on the intersection of their domains. That is, for all $x \in \text{dom}(f) \cap \text{dom}(g)$, we have $f(x) = g(x)$. A collection of functions $\mathcal{F}$ is a compatible system if every pair of functions in $\mathcal{F}$ is compatible.

Compatibility is the precise condition required to ensure that the union of two or more functions is itself a function.

***Theorem* 4.2.3.** If $\mathcal{F}$ is a compatible system of functions, then its union, $\bigcup \mathcal{F}$, is also a function. The domain of this new function is the union of the domains of the functions in $\mathcal{F}$.

$$\text{dom}(\bigcup \mathcal{F}) = \bigcup_{f \in \mathcal{F}} \text{dom}(f)$$

Proof. Let $\mathcal{F}$ be a compatible system of functions. The set $h = \bigcup \mathcal{F}$ is a set of ordered pairs, and is thus a binary relation. To show that h is a function, we must demonstrate that each element in its domain maps to a unique value.

Let $(x, y_1) \in h$ and $(x, y_2) \in h$. By the definition of a set union, there must exist functions $f_1, f_2 \in \mathcal{F}$ such that $(x, y_1) \in f_1$ and $(x, y_2) \in f_2$. This implies $x \in \text{dom}(f_1)$ and $x \in \text{dom}(f_2)$, so x is in the intersection of their domains. Since $\mathcal{F}$ is a compatible system, f_1 and f_2 are compatible, meaning $f_1(x) = f_2(x)$. As $y_1 = f_1(x)$ and $y_2 = f_2(x)$, it follows that $y_1 = y_2$.

Therefore, $h = \bigcup \mathcal{F}$ is a function. The statement regarding its domain follows directly from the properties of set union. ■

Remark. This theorem provides the formal justification for defining functions piecewise. For instance, the absolute value function can be defined as the union of two compatible functions: $f(x) = x$ restricted to $[0, \infty)$ and $g(x) = -x$ restricted to $(-\infty, 0)$.

Binary Operations

A binary operation is a special type of function that takes two elements from a set and produces a single element within that same set.

Definition 4.2.3. ***Binary Operation***. A binary operation $*$ on a non-empty set A is a function $* : A \times A \to A$.

Remark. To prove that a rule $*$ defines a binary operation on a set A, one must show two things:

1. $*$ is well-defined: equal inputs produce equal outputs.
2. A is closed under $*$: for all $a, b \in A$, the result $a * b$ is also in A.

For a binary operation $*$, we typically use infix notation $a * b$ instead of prefix function notation $*(a, b)$.

Example 4.2.4. Union on a Power Set. Let S be a set. The union operation, $\cup$, is a binary operation on the power set $\mathcal{P}(S)$.

- **Well-defined**: If $A_1 = A_2$ and $B_1 = B_2$ are subsets of S, then $A_1 \cup B_1 = A_2 \cup B_2$. This holds by the definition of set equality.
- **Closure**: For any $A, B \in \mathcal{P}(S)$, we have $A \subseteq S$ and $B \subseteq S$. Their union $A \cup B$ is also a subset of S, and is therefore an element of $\mathcal{P}(S)$.

Thus, $\cup : \mathcal{P}(S) \times \mathcal{P}(S) \to \mathcal{P}(S)$ is a binary operation.

Example 4.2.5. Operation via Cayley Table. A binary operation on a finite set can be completely specified using a multiplication table, known as a Cayley table. Let $S = \{e, a, b, c\}$. The following table defines a binary operation $*$ on S:

$*$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

The entry in the row corresponding to x and the column corresponding to y represents the result $x * y$. For instance, $a * b = c$. The operation is well-defined because each cell has exactly one entry. Closure is guaranteed because every entry in the table is an element of S.

Example 4.2.6. Addition on $\mathbb{Z}_n$. Let $n \in \mathbb{Z}^+$. Addition on $\mathbb{Z}_n$ is defined by $[a]_n + [b]_n = [a+b]_n$. We show this is a binary operation.

- **Well-defined**: We must show the result is independent of the representatives a and b. Suppose $[a_1]_n = [a_2]_n$ and $[b_1]_n = [b_2]_n$. This means $n \mid (a_1 - a_2)$ and $n \mid (b_1 - b_2)$. So, $a_1 - a_2 = nk_1$ and $b_1 - b_2 = nk_2$ for some integers k_1, k_2. Adding these gives $(a_1 + b_1) - (a_2 + b_2) = n(k_1 + k_2)$. This implies $n \mid ((a_1 + b_1) - (a_2 + b_2))$, so $[a_1 + b_1]_n = [a_2 + b_2]_n$. The operation is well-defined.
- **Closure**: For any $[a]_n, [b]_n \in \mathbb{Z}_n$, $a + b$ is an integer, so $[a + b]_n$ is an element of $\mathbb{Z}_n$.

Properties of Binary Operations

Binary operations can be classified by properties that they may or may not possess.

Definition 4.2.4. . Let $*$ be a binary operation on a set A.

- $*$ is *associative* if $(a * b) * c = a * (b * c)$ for all $a, b, c \in A$.
- $*$ is *commutative* if $a * b = b * a$ for all $a, b \in A$.
- An element $e \in A$ is an *identity element* if $a * e = e * a = a$ for all $a \in A$.
- If A has an identity element e, an element $a' \in A$ is an *inverse* of $a \in A$ if $a * a' = a' * a = e$.

***Theorem* 4.2.4.** If an identity element exists for a binary operation, it is unique. If an operation is associative and an element has an inverse, that inverse is unique.

Proof. Let $*$ be a binary operation on A. Suppose e_1 and e_2 are both identity elements. Then $e_1 = e_1 * e_2$ (since e_2 is an identity) and $e_1 * e_2 = e_2$ (since e_1 is an identity). Therefore, $e_1 = e_2$.

Now assume $*$ is associative and has identity e. Suppose $a \in A$ has two inverses, a_1' and a_2'. Consider the expression $a_1' * a * a_2'$.

$$a_1' * (a * a_2') = a_1' * e = a_1'$$
$$(a_1' * a) * a_2' = e * a_2' = a_2'$$

By associativity, these two results must be equal. Therefore, $a_1' = a_2'$. ■

Example 4.2.7. Consider the binary operation of addition on $\mathbb{Z}_n$.

- **Associative**: $([a] + [b]) + [c] = [a + b] + [c] = [(a + b) + c] = [a + (b + c)] = [a] + ([b] + [c])$.
- **Commutative**: $[a] + [b] = [a + b] = [b + a] = [b] + [a]$.
- **Identity**: The class $[0]_n$ is the identity, since $[a] + [0] = [a + 0] = [a]$.
- **Inverse**: For any class $[a]_n$, its additive inverse is $[-a]_n$, since $[a] + [-a] = [a - a] = [0]$.

Example 4.2.8. Consider the binary operation of union, $\cup$, on $\mathcal{P}(S)$ for some set S.

- **Associative**: $(A \cup B) \cup C = A \cup (B \cup C)$.
- **Commutative**: $A \cup B = B \cup A$.
- **Identity**: The empty set $\emptyset$ is the identity, since $A \cup \emptyset = A$.
- **Inverse**: Only $\emptyset$ has an inverse, which is itself. For any non-empty set $A \in \mathcal{P}(S)$, there is no set $A' \in \mathcal{P}(S)$ such that $A \cup A' = \emptyset$.

4.3 Exercises

Part I: Foundational Concepts

1. Determine which of the following relations are functions. For those that are not, provide a specific reason.

(a) $f \subseteq \mathbb{R} \times \mathbb{R}$, where $f = \{(x, y) \mid x^2 + y^2 = 9\}$.

(b) $g \subseteq \mathbb{Z} \times \mathbb{Z}$, where $g = \{(x, y) \mid y = x^3 - x\}$.

(c) $h \subseteq \mathbb{Z}_6 \times \mathbb{Z}_6$, where $h = \{([a], [b]) \mid [a]^2 = [b]\}$.

(d) $k \subseteq \mathcal{P}(\mathbb{Z}) \times \mathbb{Z}$, where $(A, n) \in k$ if and only if n is the smallest element of A.

2. For each of the following rules, determine the largest possible subset of $\mathbb{R}$ that can serve as the domain (the *maximal domain*) and find the corresponding range.

(a) $f(x) = \sqrt{4 - x^2}$.

(b) $g(x) = \frac{x+2}{x^2-4}$.

(c) $h(x) = \frac{1}{\sqrt{|x|-x}}$.

3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2 - 2x$. Find the following sets:

(a) The image $f[\{0, 1, 2, 3\}]$.

(b) The image $f[[-1, 3]]$.

(c) The pre-image $f^{-1}[\{0, 3\}]$.

(d) The pre-image $f^{-1}[[-10, -1]]$.

4. A rule $f : \mathbb{Q} \to \mathbb{Z}$ is proposed as follows: for any rational number represented as a fraction a/b (where $a, b \in \mathbb{Z}, b \neq 0$), $f(a/b) = a - b$. Is this rule a well-defined function? Justify your answer.

Remark. Consider the rational number 1. It can be written as $1/1$ or $2/2$. Do these representations yield the same output?

5. Let $f, g : \mathbb{R} \to \mathbb{R}$. Are the functions $f(x) = \frac{x^2-9}{x+3}$ and $g(x) = x - 3$ equal? Justify your answer by carefully applying the theorem on equality of functions.

Part II: Function Construction and Spaces

6. Let $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{Z}$, and $h : \mathbb{Z} \to \mathbb{Z}_4$ be defined by $f(x) = x^2 + 1$, $g(x) = \lfloor x \rfloor$ (the floor function), and $h(n) = [n + 1]_4$.

(a) Compute $(g \circ f)(2.5)$ and $(h \circ g)(-1)$.

(b) Find an expression for $(g \circ f)(x)$ and state its domain and range.

(c) Find an expression for $(h \circ g)(x)$. Is $f \circ h$ a meaningful composition? Explain.

7. Associativity of Composition. Let $f : A \to B$, $g : B \to C$, and $h : C \to D$ be functions. Prove that function composition is associative; that is, prove $h \circ (g \circ f) = (h \circ g) \circ f$ as functions from A to D.

8. If $|A| = n$ and $|B| = m$ are finite sets, what is the cardinality of the function space B^A? Prove your claim.

Remark. For each element $a \in A$, how many choices are there for its image $f(a)$ in B? Apply the Multiplication Principle.

9. The *identity function* on a set A is $\mathrm{Id}_A : A \to A$ defined by $\mathrm{Id}_A(x) = x$. Let $f : A \to B$ be any function. Prove that $f \circ \mathrm{Id}_A = f$ and $\mathrm{Id}_B \circ f = f$. This shows that identity functions act as identity elements for the operation of composition.

10. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \sin(\pi x)$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = \cos(\pi x)$. Find a non-empty set $S \subset \mathbb{R}$ such that the restriction $f|_S$ is equal to the restriction $g|_S$. Can you describe all such sets?

Part III: Binary Operations

11. For each of the following sets and rules, determine if the rule defines a binary operation on the set. If so, determine whether it is associative, commutative, has an identity element, and which elements have inverses.

(a) The set $\mathbb{Z}$ with the operation $a * b = a + b - ab$.

(b) The set $\mathbb{R} \setminus \{-1\}$ with the operation $a * b = a + b + ab$.

(c) The set of all functions $f : \mathbb{R} \to \mathbb{R}$ with the operation of pointwise addition: $(f + g)(x) = f(x) + g(x)$.

12. Let A be a set. Consider the function space A^A, which is the set of all functions from A to A.

(a) Prove that function composition, $\circ$, is a binary operation on A^A.

(b) We know from Exercise 7 that $\circ$ is associative. Prove that $\circ$ is not, in general, commutative by providing a counterexample.

Remark. Let $A = \{1, 2\}$ and construct two functions $f, g \in A^A$ such that $f \circ g \neq g \circ f$.

13. Cancellation Laws. Let $*$ be an associative binary operation on a set S with identity element e. Let $a \in S$ be an element that has an inverse, a'. Prove that for any $b, c \in S$:

(a) If $a * b = a * c$, then $b = c$ (Left Cancellation).

(b) If $b * a = c * a$, then $b = c$ (Right Cancellation).

14. $\star$ Let $S = \mathbb{R} \times \mathbb{R}$. Define a binary operation $*$ on S by $(a, b) * (c, d) = (ac - bd, ad + bc)$. This operation defines multiplication of complex numbers. Show that this operation is associative and commutative, find the identity element, and find the inverse for any non-identity element $(a, b) \neq (0, 0)$.

15. $\star$ **Characteristic Functions.** For any set U, and any subset $S \subseteq U$, the *characteristic function* of S is the function $\chi_S : U \to \{0, 1\}$ defined by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

Let A, B be subsets of U. Prove the following identities relating set operations to arithmetic operations on their characteristic functions:

(a) $\chi_{A \cap B} = \chi_A \cdot \chi_B$ (pointwise product).

(b) $\chi_{A \cup B} = \chi_A + \chi_B - \chi_A \cdot \chi_B$.

(c) $\chi_{A \setminus B} = \chi_A(1 - \chi_B)$.

4.4 Injective, Surjective, and Bijective Functions

The inverse of a relation R, denoted R^{-1}, is formed by reversing the coordinates of each ordered pair in R. While this operation is always defined for relations, the inverse of a function is not necessarily a function itself. For example, if $f = \{(1,3),(2,3)\}$, its inverse is the relation $f^{-1} = \{(3,1),(3,2)\}$. This relation fails the definition of a function because the input 3 is mapped to two distinct outputs. This observation motivates the classification of functions based on properties that govern the behaviour of their inverses.

Injective Functions (Injections)

For an inverse relation f^{-1} to be a function, it must not map any element of its domain to more than one output. This implies that the original function f must not map more than one element of its domain to the same output. This property is known as injectivity.

Definition 4.4.1. ***Injective Function***. A function $f : A \to B$ is injective (or one-to-one) if for every pair of distinct elements in the domain, their images are also distinct. Formally:

$$\forall x_1, x_2 \in A, (x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2))$$

The logically equivalent contrapositive is often used in proofs:

$$\forall x_1, x_2 \in A, (f(x_1) = f(x_2) \Rightarrow x_1 = x_2)$$

An injective function is called an injection.

Geometrically, an injection is a function whose graph passes the horizontal line test: any horizontal line intersects its graph at most once.

Example 4.4.1. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 5x + 1$. To show f is injective, let $x_1, x_2 \in \mathbb{R}$ and assume $f(x_1) = f(x_2)$.

$$5x_1 + 1 = 5x_2 + 1 \Rightarrow 5x_1 = 5x_2 \Rightarrow x_1 = x_2$$

Therefore, f is an injection.

Example 4.4.2. Define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x^2$. This function is not injective because we can find distinct inputs that map to the same output. For instance, $g(2) = 4$ and $g(-2) = 4$, but $2 \neq -2$.

Remark. A function that is not injective may become injective if its domain is restricted. For the function $g(x) = x^2$, the restriction $g|_{[0,\infty)}$ is an injection.

Surjective Functions (Surjections)

Even if a function is injective, its inverse may not be defined on the entirety of the original codomain. For f^{-1} to be a function from B to A, its domain must be B. This requires that the range of the original function f is equal to its codomain B.

Definition 4.4.2. ***Surjective Function***. A function $f : A \to B$ is surjective (or onto) if every element in the codomain B is the image of at least one element in the domain A. Formally:

$$\forall y \in B, \exists x \in A, f(x) = y$$

This is equivalent to the statement $\text{ran}(f) = B$. A surjective function is called a surjection.

Example 4.4.3. Any non-constant linear function $f : \mathbb{R} \to \mathbb{R}$ of the form $f(x) = ax + b$ with $a \neq 0$ is a surjection. To prove this, let $y \in \mathbb{R}$ be an arbitrary element of the codomain. We must find an $x \in \mathbb{R}$ such that $f(x) = y$. We solve for x:

$$ax + b = y \Rightarrow ax = y - b \Rightarrow x = \frac{y - b}{a}$$

Since $a \neq 0$, this value of x exists and is in the domain $\mathbb{R}$. Therefore, f is surjective.

Example 4.4.4. Define a projection map $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ by $\pi(x, y, z) = (x, y)$. This function is surjective. For any arbitrary $(a, b) \in \mathbb{R}^2$, we can choose the element $(a, b, 0) \in \mathbb{R}^3$. Then $\pi(a, b, 0) = (a, b)$. However, π is not injective since, for example, $\pi(1, 2, 3) = \pi(1, 2, 4) = (1, 2)$.

Bijective Functions and Invertibility

A function whose inverse is also a function with a domain equal to the original codomain must possess both of the properties discussed above.

Definition 4.4.3. ***Bijective Function***. A function $f : A \to B$ is bijective if it is both injective and surjective. A bijection is also known as a one-to-one correspondence.

Example 4.4.5.

- The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = ax + b$ with $a \neq 0$ is a bijection.
- The function $g : (-\pi/2, \pi/2) \to \mathbb{R}$ defined by $g(x) = \tan(x)$ is a bijection.
- The function $h : \mathbb{R} \to (0, \infty)$ defined by $h(x) = e^x$ is a bijection.

Note. In mathematical literature, special arrows are often used to denote the type of function being described.

- An injection $f : A \to B$ may be written as $f : A \hookrightarrow B$.
- A surjection $f : A \to B$ may be written as $f : A \twoheadrightarrow B$.
- A bijection $f : A \to B$ may be written as $f : A \hookrightarrow\!\!\!\!\twoheadrightarrow B$.

We are now prepared to formally connect these properties to the concept of invertibility.

Definition 4.4.4. ***Invertible Function***. A function $f : A \to B$ is invertible if its inverse relation f^{-1} is a function from B to A.

***Theorem* 4.4.1. Invertibility Theorem**. A function $f : A \to B$ is invertible if and only if it is a bijection.

Proof. ($\Rightarrow$) Assume f is invertible, meaning $f^{-1} : B \to A$ is a function.

- **Injectivity**: Let $x_1, x_2 \in A$ and assume $f(x_1) = f(x_2) = y$. By definition of the inverse relation, $(y, x_1) \in f^{-1}$ and $(y, x_2) \in f^{-1}$. Since f^{-1} is a function, it must be that $x_1 = x_2$. Thus, f is injective.
- **Surjectivity**: Let y be an arbitrary element of B. Since B is the domain of f^{-1}, there exists some $x \in A$ such that $f^{-1}(y) = x$. This implies $(y, x) \in f^{-1}$, which means $(x, y) \in f$, so $f(x) = y$. Thus, f is surjective.

Since f is both injective and surjective, it is a bijection.

($\Leftarrow$) Assume f is a bijection. We must show that the relation f^{-1} is a function from B to A.

- The domain of f^{-1} is the range of f. Since f is surjective, $\text{ran}(f) = B$, so the domain of f^{-1} is B.
- To show f^{-1} is a function, let $(y, x_1) \in f^{-1}$ and $(y, x_2) \in f^{-1}$. This means $(x_1, y) \in f$ and $(x_2, y) \in f$, so $f(x_1) = y$ and $f(x_2) = y$. Since f is injective, we must have $x_1 = x_2$.

Thus, f^{-1} is a function from B to A, and f is invertible. ■

An equivalent characterisation of invertibility involves the identity map. If $f : A \to B$ is a bijection, its inverse function $f^{-1} : B \to A$ satisfies $f^{-1}(y) = x \Leftrightarrow f(x) = y$. This leads to the following theorem.

***Theorem* 4.4.2. Characterisation of Invertibility**. A function $f : A \to B$ is a bijection if and only if there exists a function $g : B \to A$ such that $g \circ f = \text{Id}_A$ and $f \circ g = \text{Id}_B$. Furthermore, this function g is unique and is equal to f^{-1}.

Proof. ($\Rightarrow$) Assume $f : A \to B$ is a bijection. By the Invertibility Theorem, f is invertible and its inverse $f^{-1} : B \to A$ is a function. Let $g = f^{-1}$.

- For any $a \in A$, let $b = f(a)$. Then $g(b) = f^{-1}(b) = a$. So, $(g \circ f)(a) = g(f(a)) = g(b) = a$. Thus, $g \circ f = \mathrm{Id}_A$.
- For any $b \in B$, let $a = g(b)$. Then $f(a) = f(g(b))$. Since $a = f^{-1}(b)$, it follows that $f(a) = b$. So, $(f \circ g)(b) = f(g(b)) = b$. Thus, $f \circ g = \mathrm{Id}_B$.

($\Leftarrow$) Assume such a function $g : B \to A$ exists.

- **Injectivity**: Let $x_1, x_2 \in A$ and assume $f(x_1) = f(x_2)$. Applying g to both sides gives $g(f(x_1)) = g(f(x_2))$. This is $(g \circ f)(x_1) = (g \circ f)(x_2)$. Since $g \circ f = \mathrm{Id}_A$, we have $\mathrm{Id}_A(x_1) = \mathrm{Id}_A(x_2)$, which implies $x_1 = x_2$. Thus, f is injective.
- **Surjectivity**: Let $y \in B$. Consider the element $x = g(y) \in A$. Then $f(x) = f(g(y)) = (f \circ g)(y) = \mathrm{Id}_B(y) = y$. We have found an element $x \in A$ that maps to y. Thus, f is surjective.

Since f is a bijection, it is invertible. The uniqueness of g follows because if another function h existed with the same properties, we would have $h = h \circ \mathrm{Id}_B = h \circ (f \circ g) = (h \circ f) \circ g = \mathrm{Id}_A \circ g = g$. ■

Composition and Function Properties

The properties of being injective, surjective, or bijective are preserved under composition.

***Theorem* 4.4.3. Preservation of Properties under Composition**. Let $f : A \to B$ and $g : B \to C$ be functions.

(i) If f and g are both injective, then $g \circ f$ is injective.
(ii) If f and g are both surjective, then $g \circ f$ is surjective.
(iii) If f and g are both bijective, then $g \circ f$ is bijective.

Proof.

(i) Assume f and g are injective. Let $x_1, x_2 \in A$ and assume $(g \circ f)(x_1) = (g \circ f)(x_2)$. This means $g(f(x_1)) = g(f(x_2))$. Since g is injective, we must have $f(x_1) = f(x_2)$. Since f is injective, this implies $x_1 = x_2$. Thus, $g \circ f$ is injective.

(ii) Assume f and g are surjective. Let c be an arbitrary element in C. Since g is surjective, there exists a $b \in B$ such that $g(b) = c$. Since f is surjective, there exists an $a \in A$ such that $f(a) = b$. Therefore, $(g \circ f)(a) = g(f(a)) = g(b) = c$. Thus, $g \circ f$ is surjective.

(iii) If f and g are bijections, they are both injective and surjective. By (i) and (ii), their composition $g \circ f$ is also both injective and surjective, and is therefore a bijection.

■

***Theorem* 4.4.4. Inverse of a Composition**. If $f : A \to B$ and $g : B \to C$ are bijections, then the composition $g \circ f$ is invertible, and its inverse is given by:

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}$$

Proof. From Theorem 4.4.3, since f and g are bijections, their composition $g \circ f : A \to C$ is also a bijection and is therefore invertible. To verify the formula for the inverse, we use the identity property from the previous theorem. We must show that $(f^{-1} \circ g^{-1})$ acts as the inverse to $(g \circ f)$. Consider the composition $(f^{-1} \circ g^{-1}) \circ (g \circ f)$. Using the associativity of composition:

$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = f^{-1} \circ (g^{-1} \circ g) \circ f = f^{-1} \circ \mathrm{Id}_B \circ f = f^{-1} \circ f = \mathrm{Id}_A.$$

Similarly, consider the composition in the other order:

$$(g \circ f) \circ (f^{-1} \circ g^{-1}) = g \circ (f \circ f^{-1}) \circ g^{-1} = g \circ \mathrm{Id}_B \circ g^{-1} = g \circ g^{-1} = \mathrm{Id}_C.$$

Since $f^{-1} \circ g^{-1}$ satisfies the conditions for being the inverse of $g \circ f$, and since the inverse is unique, we conclude that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. ■

Example 4.4.6. The Canonical Factorisation. Let $f : A \to B$ be a function. We can define an equivalence relation $\sim_f$ on the domain A by:

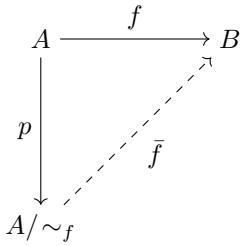
$$x \sim_f y \Leftrightarrow f(x) = f(y)$$

This relation is known as the *kernel* of f. The equivalence class of an element $a \in A$ is precisely the set of all elements mapping to the same output, $[a] = f^{-1}(\{f(a)\})$.

We can construct a function $\bar{f} : A/\sim_f \to B$ from the quotient set to the codomain defined by $\bar{f}([a]) = f(a)$. This function is well-defined. Furthermore:

- $\bar{f}$ is **injective** (since $\bar{f}([a]) = \bar{f}([b]) \Rightarrow f(a) = f(b) \Rightarrow a \sim_f b \Rightarrow [a] = [b]$).
- If f is **surjective**, then $\bar{f}$ is also surjective, and thus a **bijection**.

This result is known as the Canonical Factorisation of a function. It decomposes any function f into the composition $\bar{f} \circ p$, where $p : A \to A/\sim_f$ is the canonical projection mapping $a \mapsto [a]$.



4.5 Exercises

1. For each of the following functions, determine if it is injective, surjective, or bijective. Justify your answer by providing a proof or a specific counterexample.

(a) $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = 2n + 3$.

(b) $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = \sin(x)$.

(c) $h : \mathcal{P}(\mathbb{Z}) \to \mathcal{P}(\mathbb{Z})$ defined by $h(S) = S \cup \{0\}$.

(d) $p : \mathbb{Z} \to \mathbb{Z}_5$ defined by $p(n) = [n^2]_5$.

2. Let the function $f : \mathbb{R} \setminus \{3\} \to \mathbb{R} \setminus \{2\}$ be defined by $f(x) = \frac{2x+1}{x-3}$. Prove that f is a bijection and find a formula for its inverse function f^{-1}.

3. Let $f : A \to B$ and $g : B \to C$ be functions. Prove the following two statements:

(a) If the composition $g \circ f$ is injective, then f must be injective.
(b) If the composition $g \circ f$ is surjective, then g must be surjective.

For each case, provide a counterexample to show that the other function (g in part (a), f in part (b)) is not necessarily injective/surjective.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the quadratic $f(x) = x^2 - 6x + 11$.

(a) By completing the square, find the vertex of the parabola $y = f(x)$.
(b) Find a maximal domain $A \subseteq \mathbb{R}$ and a codomain $B \subseteq \mathbb{R}$ such that the restriction $f|_A : A \to B$ is a bijection.

5. A function $f : A \to B$ has a left inverse if there exists a function $g : B \to A$ such that $g \circ f = \mathrm{Id}_A$. Prove that a function has a left inverse if and only if it is injective.

6. Let A and B be non-empty sets. Consider the function $\pi_A : A \times B \to A$ defined by $\pi_A(a, b) = a$. This function is known as the projection onto A.

(a) Prove that π_A is surjective.
(b) Under what conditions on the set B is π_A injective?

7. Let $f : A \to B$ be a function. Prove that the following two statements are equivalent:

(i) f is injective.
(ii) For every subset $S \subseteq A$, $f^{-1}[f[S]] = S$.

8. Let $f : A \to C$ and $g : B \to D$ be bijections. Define a new function $H : A \times B \to C \times D$ by the rule $H(a, b) = (f(a), g(b))$. Prove that H is a bijection.

9. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be **periodic** if there exists a constant $P > 0$ (called the period) such that $f(x + P) = f(x)$ for all $x \in \mathbb{R}$. Prove that no periodic function can be injective.

10. Prove that the inverse of a bijection is itself a bijection. That is, if $f : A \to B$ is a bijection, prove that its inverse function $f^{-1} : B \to A$ is both injective and surjective.

4.6 Order Isomorphisms

A bijection establishes that two sets are of the same size. We can extend this idea to partially ordered sets to capture the notion of having the same structure. A function that preserves the order relations between two posets reveals that they are structurally identical.

Consider the sets $A = \mathbb{R} \times \{0\}$ and $B = \{0\} \times \mathbb{R}$. Let $(A, \preceq_A)$ be the poset where $(x_1, 0) \preceq_A (x_2, 0) \Leftrightarrow x_1 \leq x_2$, and let $(B, \preceq_B)$ be the poset where $(0, y_1) \preceq_B (0, y_2) \Leftrightarrow y_1 \leq y_2$. The function $\phi : A \to B$ defined by $\phi(x, 0) = (0, x)$ is a bijection. Furthermore, it preserves the ordering:

$$(x_1, 0) \preceq_A (x_2, 0) \Leftrightarrow x_1 \leq x_2 \Leftrightarrow \phi(x_1, 0) \preceq_B \phi(x_2, 0)$$

The function ϕ demonstrates that these two posets, while composed of different elements, have the exact same order structure. This leads to the following definitions.

Definition 4.6.1. ***Order-Preserving Function***. Let $(A, \preceq_A)$ and $(B, \preceq_B)$ be posets. A function $f : A \to B$ is **order-preserving** if for all $x_1, x_2 \in A$,

$$x_1 \preceq_A x_2 \Leftrightarrow f(x_1) \preceq_B f(x_2)$$

Definition 4.6.2. ***Order Isomorphism***. An order-preserving function $f : A \to B$ that is also a bijection is called an order isomorphism. If such a function exists, the posets $(A, \preceq_A)$ and $(B, \preceq_B)$ are said to be order isomorphic, denoted $(A, \preceq_A) \cong (B, \preceq_B)$. Posets that are order isomorphic are said to have the same order type.

Example 4.6.1. Define $f : \mathbb{Z}^+ \to \mathbb{Z}^-$ by $f(x) = -x$. This is a bijection. Let us consider the posets $(\mathbb{Z}^+, \leq)$ and $(\mathbb{Z}^-, \geq)$. For any $x_1, x_2 \in \mathbb{Z}^+$, we have:

$$x_1 \leq x_2 \Leftrightarrow -x_1 \geq -x_2 \Leftrightarrow f(x_1) \geq f(x_2)$$

Thus, f is an order isomorphism, and $(\mathbb{Z}^+, \leq) \cong (\mathbb{Z}^-, \geq)$. This shows that an isomorphism can relate different ordering relations.

Example 4.6.2. The function $f : \mathbb{R} \to (0, \infty)$ defined by $f(x) = e^x$ is an order isomorphism between the posets $(\mathbb{R}, \leq)$ and $((0, \infty), \leq)$.

***Lemma* 4.6.1.** Let $(P, \leq)$ be a totally ordered set and $(Q, \preceq)$ an ordered set. Let $h : P \to Q$ be a bijection such that $\forall p_1, p_2 \in P, (p_1 \leq p_2 \implies h(p_1) \preceq h(p_2))$. Then h is an order isomorphism, and $(Q, \preceq)$ is totally ordered.

Proof. Take any $p_1, p_2 \in P$ and assume $h(p_1) \preceq h(p_2)$. Suppose $p_2 < p_1$. Then $h(p_2) \preceq h(p_1)$, implying $h(p_1) = h(p_2)$ by antisymmetry, but h injective gives $p_1 = p_2$, contradiction. Thus, $p_1 \leq p_2$.

For totality of $(Q, \preceq)$, take $q_1, q_2 \in Q$. Since surjective, $q_1 = h(p_1)$, $q_2 = h(p_2)$. As P total, $p_1 \leq p_2$ or vice versa, so $q_1 \preceq q_2$ or reverse. ■

***Theorem* 4.6.1.** The inverse of an order isomorphism is an order isomorphism.

Proof. Let $f : (A, \preceq_A) \to (B, \preceq_B)$ be an order isomorphism. We know from the Invertibility Theorem that $f^{-1} : B \to A$ is a bijection. We must show it is order-preserving. Let $y_1, y_2 \in B$. Since f is surjective, there exist unique $x_1, x_2 \in A$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$, which implies $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$. Because f is an order isomorphism, we know that $x_1 \preceq_A x_2 \Leftrightarrow f(x_1) \preceq_B f(x_2)$. Substituting for x_1, x_2 and $f(x_1), f(x_2)$ gives:

$$f^{-1}(y_1) \preceq_A f^{-1}(y_2) \Leftrightarrow y_1 \preceq_B y_2$$

This is precisely the condition for f^{-1} to be order-preserving. Thus, f^{-1} is an order isomorphism. ■

***Theorem* 4.6.2.** The composition of order isomorphisms is an order isomorphism.

Proof. Let $f : (A, \preceq_A) \to (B, \preceq_B)$ and $g : (B, \preceq_B) \to (C, \preceq_C)$ be order isomorphisms. We know from Theorem 4.4.3 that their composition $g \circ f : A \to C$ is a bijection. To show it is order-preserving, let $x_1, x_2 \in A$.

$$\begin{aligned} x_1 \preceq_A x_2 &\Leftrightarrow f(x_1) \preceq_B f(x_2) && \text{since } f \text{ is an isomorphism} \\ &\Leftrightarrow g(f(x_1)) \preceq_C g(f(x_2)) && \text{since } g \text{ is an isomorphism} \\ &\Leftrightarrow (g \circ f)(x_1) \preceq_C (g \circ f)(x_2) \end{aligned}$$

Therefore, $g \circ f$ is an order isomorphism. ■

These theorems imply that $\cong$ is an equivalence relation on the class of all posets. Reflexivity holds via the identity map, symmetry via the inverse, and transitivity via composition.

Sometimes a poset may be structurally identical to a sub-poset of another. This idea is captured by an embedding.

Definition 4.6.3. ***Order Embedding***. A function $f : (A, \preceq_A) \to (B, \preceq_B)$ is an order embedding if it is an order isomorphism from $(A, \preceq_A)$ onto $(\text{ran}(f), \preceq_B)$. That is, for all $x_1, x_2 \in A$,

$$x_1 \preceq_A x_2 \Leftrightarrow f(x_1) \preceq_B f(x_2)$$

Note that such a function is necessarily injective.

Example 4.6.3. The function $f : \mathbb{Q} \to \mathbb{R}$ defined by $f(x) = x$ is an order embedding of $(\mathbb{Q}, \leq)$ into $(\mathbb{R}, \leq)$.

Example 4.6.4. Let $\preceq_{lex}$ denote the lexicographical order. The function $\pi : (\mathbb{R}^2, \preceq_{lex}) \to (\mathbb{R}^3, \preceq_{lex})$ defined by $\pi(x, y) = (x, y, 0)$ is an order embedding. Although $\mathbb{R}^2$ is not a subset of $\mathbb{R}^3$, the image of π is a copy of $\mathbb{R}^2$ within $\mathbb{R}^3$ that preserves the order structure.

4.7 Exercises

1. Determine whether the following pairs of posets are order isomorphic. If they are, construct an explicit order isomorphism and prove it is so. If not, explain why no such isomorphism can exist.
   (a) $(\mathcal{P}(\{1, 2\}), \subseteq)$ and $(\{1, 2, 3, 6\}, |)$, where $|$ denotes the "divides" relation.
   (b) $(\mathbb{N}, \leq)$ and $(\mathbb{Z}, \leq)$.
   (c) The set of positive even integers $\{2, 4, 6, \dots\}$ with the standard order $\leq$, and the set of positive integers $\mathbb{Z}^+$ with the standard order $\leq$.

2. A strictly increasing function $f : (A, \leq_A) \to (B, \leq_B)$ between two totally ordered sets is a function such that for all $x_1, x_2 \in A$, if $x_1 <_A x_2$ then $f(x_1) <_B f(x_2)$.

(a) Prove that any strictly increasing function is injective.

(b) Prove that if f is a strictly increasing surjection, then it is an order isomorphism.

(c) Use this to construct an order isomorphism between the posets $(\,(0,1), \leq)$ and $(\mathbb{R}, \leq)$.

3. Let $f : (A, \preceq_A) \to (B, \preceq_B)$ be an order isomorphism. Prove that an element $a \in A$ is a minimal element of A if and only if its image $f(a)$ is a minimal element of B.

4. Duality and Anti-isomorphisms. Given a poset $(A, \preceq)$, its dual poset is $(A, \succeq)$, where $x \succeq y$ if and only if $y \preceq x$. A poset is said to be self-dual if it is order isomorphic to its own dual.

(a) Define a function $f : (\mathcal{P}(S), \subseteq) \to (\mathcal{P}(S), \supseteq)$ for a non-empty set S using the complement operation.

(b) Prove that your function is an order isomorphism, and therefore that any power set poset is self-dual.

5. $\star$ **Classification of Finite Total Orders.** Prove that any finite, totally ordered set $(A, \preceq)$ with n elements is order isomorphic to the poset $(\{1, 2, \ldots, n\}, \leq)$.

Remark. Construct the isomorphism by mapping the minimal element of A to 1, the minimal element of the remaining set to 2, and so on. Formally, this can be done by induction on n.

4.8 Images and Inverse Images of Sets

We have defined the image of a single element under a function. This concept can be extended to consider the set of all images of the elements within a given subset of the domain.

Definition 4.8.1. ***Image of a Set***. Let $f : A \to B$ be a function and let $S \subseteq A$. The image of S under f is the set of all images of elements in S:

$$f[S] = \{f(x) \mid x \in S\}$$

It follows that $f[S] \subseteq \text{ran}(f) \subseteq B$ and that $f[A] = \text{ran}(f)$.

A similar concept applies to subsets of the codomain, where we consider the set of all elements in the domain that map into the given subset.

Definition 4.8.2. ***Inverse Image of a Set***. Let $f : A \to B$ be a function and let $T \subseteq B$. The inverse image of T under f is the set of all pre-images of elements in T:

$$f^{-1}[T] = \{x \in A \mid f(x) \in T\}$$

Observe that $f^{-1}[T] \subseteq A$ and $f^{-1}[B] = A$.

Notation 4.8.1. When the set T is a singleton, $T = \{y\}$, we often abbreviate the notation for its inverse image. The set $f^{-1}[\{y\}]$ is called the **fibre** of f over y, and may be written as $f^{-1}(y)$. The fibre is the set of all solutions to the equation $f(x) = y$.

Remark. The notation $f^{-1}[T]$ is used even if the function f is not invertible. It denotes the set of pre-images, not the application of an inverse function.

Example 4.8.1. Let $f : \{1, 2, 3, 4\} \to \{2, 4, 5\}$ be the function $f = \{(1, 2), (2, 4), (3, 5), (4, 5)\}$.

- The image of the set $\{1, 3\}$ is $f[\{1, 3\}] = \{f(1), f(3)\} = \{2, 5\}$.
- The inverse image of the set $\{5\}$ is $f^{-1}[\{5\}] = \{x \mid f(x) \in \{5\}\} = \{3, 4\}$.

Example 4.8.2. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2 + 1$.

- To find the image $f[(1,2)]$, we note that if $1 < x < 2$, then $1 < x^2 < 4$, which implies $2 < x^2 + 1 < 5$. Thus, $f[(1,2)] = (2,5)$.
- To find the inverse image $f^{-1}[(2,5)]$, we solve $2 < f(x) < 5$:

$$2 < x^2 + 1 < 5 \Leftrightarrow 1 < x^2 < 4 \Leftrightarrow 1 < |x| < 2$$

  This is true when $x \in (-2,-1)$ or $x \in (1,2)$. Therefore, $f^{-1}[(2,5)] = (-2,-1) \cup (1,2)$.
- To find $f^{-1}[(-2,-1)]$, we need to find x such that $-2 < x^2+1 < -1$, which simplifies to $-3 < x^2 < -2$. Since x^2 cannot be negative, no such x exists. Thus, $f^{-1}[(-2,-1)] = \emptyset$.

The operations of taking images and inverse images interact predictably with set operations like union and intersection.

***Theorem* 4.8.1. Properties of Image and Inverse Image**. Let $f : A \to B$ be a function. Let S, S_1, S_2 be subsets of A, and let T, T_1, T_2 be subsets of B.

(i) If $S_1 \subseteq S_2$, then $f[S_1] \subseteq f[S_2]$.

(ii) $f[S_1 \cup S_2] = f[S_1] \cup f[S_2]$.

(iii) $f[S_1 \cap S_2] \subseteq f[S_1] \cap f[S_2]$.

(iv) If $T_1 \subseteq T_2$, then $f^{-1}[T_1] \subseteq f^{-1}[T_2]$.

(v) $f^{-1}[T_1 \cup T_2] = f^{-1}[T_1] \cup f^{-1}[T_2]$.

(vi) $f^{-1}[T_1 \cap T_2] = f^{-1}[T_1] \cap f^{-1}[T_2]$.

(vii) $f^{-1}[B \setminus T] = A \setminus f^{-1}[T]$.

The statements concerning unions and intersections generalise to arbitrary families of sets.

Proof. We prove (i) and (iv). The remaining proofs are analogous.

Proof of (i):

$$\begin{aligned} y \in f\left[\bigcup_{i\in I} S_i\right] &\Leftrightarrow \exists x \in \bigcup_{i\in I} S_i \text{ such that } f(x) = y \\ &\Leftrightarrow \exists i \in I, \exists x \in S_i \text{ such that } f(x) = y \\ &\Leftrightarrow \exists i \in I \text{ such that } y \in f[S_i] \\ &\Leftrightarrow y \in \bigcup_{i\in I} f[S_i] \end{aligned}$$

Proof of (iv):

$$\begin{aligned} x \in f^{-1}\left[\bigcap_{j\in J} T_j\right] &\Leftrightarrow f(x) \in \bigcap_{j\in J} T_j \\ &\Leftrightarrow \forall j \in J, f(x) \in T_j \\ &\Leftrightarrow \forall j \in J, x \in f^{-1}[T_j] \\ &\Leftrightarrow x \in \bigcap_{j\in J} f^{-1}[T_j] \end{aligned}$$

■

The inclusion in property (ii) cannot be strengthened to an equality in general. Let $f : \{1, 2\} \to \{3\}$ be defined by $f = \{(1, 3), (2, 3)\}$. Let $S_1 = \{1\}$ and $S_2 = \{2\}$. Then $S_1 \cap S_2 = \emptyset$, so $f[S_1 \cap S_2] = f[\emptyset] = \emptyset$. However, $f[S_1] = \{3\}$ and $f[S_2] = \{3\}$, so $f[S_1] \cap f[S_2] = \{3\}$. Thus, $f[S_1 \cap S_2] \neq f[S_1] \cap f[S_2]$. Equality holds if the function is injective.

Finally, we examine the composition of image and inverse image operations.

***Theorem* 4.8.2.** Let $f : A \to B$ be a function. For any $S \subseteq A$ and $T \subseteq B$:

(i) $S \subseteq f^{-1}[f[S]]$. Equality holds for all $S \subseteq A$ if and only if f is injective.

(ii) $f[f^{-1}[T]] \subseteq T$. Equality holds for all $T \subseteq B$ if and only if f is surjective.

Proof.

Proof of (i): Let $x \in S$. Then $f(x) \in f[S]$. By definition of the inverse image, this means $x \in f^{-1}[f[S]]$. Thus, $S \subseteq f^{-1}[f[S]]$. Now, assume f is injective. Let $x \in f^{-1}[f[S]]$. This means $f(x) \in f[S]$, so there exists some $s \in S$ such that $f(x) = f(s)$. Since f is injective, $x = s$, which implies $x \in S$. Thus, $f^{-1}[f[S]] \subseteq S$, giving equality. Conversely, assume $S = f^{-1}[f[S]]$ for all $S \subseteq A$. Let $x_1, x_2 \in A$ with $f(x_1) = f(x_2)$. Let $S = \{x_1\}$. Then $f[S] = \{f(x_1)\}$. The inverse image is $f^{-1}[f[S]] = \{z \in A \mid f(z) = f(x_1)\}$. Since $f(x_2) = f(x_1)$, we know $x_2 \in f^{-1}[f[S]]$. By our assumption, this means $x_2 \in S$, so $x_2 = x_1$. Thus, f is injective.

Proof of (ii): Let $y \in f[f^{-1}[T]]$. By definition, there exists an $x \in f^{-1}[T]$ such that $f(x) = y$. The fact that $x \in f^{-1}[T]$ means that $f(x) \in T$. Therefore, $y \in T$. Thus, $f[f^{-1}[T]] \subseteq T$. Now, assume f is surjective. Let $y \in T$. Since f is surjective, there exists an $x \in A$ such that $f(x) = y$. As $y \in T$, we have $f(x) \in T$, which means $x \in f^{-1}[T]$. Since we have found an element x in $f^{-1}[T]$, its image, $f(x) = y$, must be in $f[f^{-1}[T]]$. Thus, $T \subseteq f[f^{-1}[T]]$, giving equality. Conversely, assume $f[f^{-1}[T]] = T$ for all $T \subseteq B$. In particular, this holds for $T = B$. Then $f[f^{-1}[B]] = B$. Since $f^{-1}[B] = A$, this becomes $f[A] = B$, which is the definition of surjectivity.

■

4.9 Exercises

1. Let the function $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \cos(x)$. Find the following sets, providing brief justification.

(a) The image $f\left[\left[0, \frac{3\pi}{4}\right]\right]$.

(b) The inverse image $f^{-1}[\{1\}]$.

(c) The inverse image $f^{-1}\left[\left(\frac{1}{\sqrt{2}}, \infty\right)\right]$.

2. Let $f : A \to B$ be a function, let S_1, S_2 be subsets of A, and let T_1, T_2 be subsets of B. The provided text proves that the image operation distributes over arbitrary unions and that the inverse image operation distributes over arbitrary intersections. Prove the other two fundamental properties for pairs of sets:

(a) $f[S_1 \cap S_2] \subseteq f[S_1] \cap f[S_2]$.

(b) $f^{-1}[T_1 \cup T_2] = f^{-1}[T_1] \cup f^{-1}[T_2]$.

3. As established in the text and the previous exercise, the inclusion $f[S_1 \cap S_2] \subseteq f[S_1] \cap f[S_2]$ holds for any function. Prove that a function $f : A \to B$ is injective if and only if this inclusion is an equality for all subsets $S_1, S_2 \subseteq A$.

4. Let $f : A \to B$ be a function and let $T_1, T_2 \subseteq B$. Prove that the inverse image operation is well-behaved with respect to set difference, that is:

$$f^{-1}[T_1 \setminus T_2] = f^{-1}[T_1] \setminus f^{-1}[T_2]$$

5. The forward image is less well-behaved with respect to set difference. Let $f : A \to B$ be a function and $S \subseteq A$.

(a) Prove that $f[A] \setminus f[S] \subseteq f[A \setminus S]$.
(b) Provide a specific counterexample using a function $f : \mathbb{Z} \to \mathbb{Z}$ to show that equality does not generally hold.

6. Let $f : A \to B$ and $g : B \to C$ be functions. The composition $g \circ f$ is a function from A to C. Prove the following identities relating composition to images and inverse images:

(a) For any $S \subseteq A$, $(g \circ f)[S] = g[f[S]]$.
(b) For any $T \subseteq C$, $(g \circ f)^{-1}[T] = f^{-1}[g^{-1}[T]]$.

7. $\star$ A function is injective if and only if it keeps distinct elements separate. We can extend this idea to sets. Prove that a function $f : A \to B$ is injective if and only if for any two disjoint subsets S_1, S_2 of A (i.e., $S_1 \cap S_2 = \emptyset$), their images are also disjoint (i.e., $f[S_1] \cap f[S_2] = \emptyset$).

Chapter 5

Infinite Sets and the Axiom of Choice

The axioms introduced thus far provide a robust framework for constructing finite sets and defining fundamental mathematical structures. However, they do not guarantee the existence of any infinite sets. To formalise concepts like the natural numbers, and subsequently calculus, we must introduce axioms that explicitly posit the existence of infinite collections.

5.1 The Axiom of Infinity

To construct an infinite set, we can envision a process that begins with the empty set and iteratively generates new, larger sets. This process is formalised using the concept of a successor.

Definition 5.1.1. ***Successor of a Set***. The successor of a set x, denoted $S(x)$, is the set defined as:

$$S(x) = x \cup \{x\}$$

Applying this operation repeatedly starting from $\emptyset$ generates a sequence of distinct sets:

- $S(\emptyset) = \emptyset \cup \{\emptyset\} = \{\emptyset\}$
- $S(\{\emptyset\}) = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}$
- $S(\{\emptyset, \{\emptyset\}\}) = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$

This sequence never repeats, suggesting an infinite collection. We capture the essence of such a collection with the following definition.

Definition 5.1.2. ***Inductive Set***. A set I is called inductive if it contains the empty set and is closed under the successor operation. Formally:

$$(\emptyset \in I) \wedge (\forall x(x \in I \Rightarrow S(x) \in I))$$

To ensure such a set exists, we introduce a new axiom.

***Axiom* 5.1.1. Infinity**. There exists at least one inductive set.

This axiom guarantees the existence of a set A such that $\emptyset \in A$, and if $x \in A$, then $x \cup \{x\} \in A$. Such a set necessarily contains the infinite sequence of successors starting from $\emptyset$. The Axiom of Infinity does not state that there is only one such set. To define the natural numbers uniquely, we take the intersection of all such sets.

Definition 5.1.3. ***The Set of Natural Numbers***. The set of natural numbers, denoted $\mathbb{N}$, is the set containing all elements that belong to every inductive set.

$$\mathbb{N} := \{x \mid \forall I(I \text{ is an inductive set } \Rightarrow x \in I)\}$$

The existence of $\mathbb{N}$ is guaranteed. By the Axiom of Infinity, there is at least one inductive set, say A. We can then use the Schema of Separation to construct $\mathbb{N}$ as a subset of A:

$$\mathbb{N} = \{x \in A \mid \forall I(I \text{ is an inductive set } \Rightarrow x \in I)\}$$

This construction defines $\mathbb{N}$ as the smallest inductive set.

***Theorem* 5.1.1. Properties of the Natural Numbers**. The set $\mathbb{N}$ is inductive. Furthermore, for any inductive set I, we have $\mathbb{N} \subseteq I$.

Proof. By definition, $\emptyset$ is an element of every inductive set I, so $\emptyset \in \mathbb{N}$. Now, let $n \in \mathbb{N}$. This means n is in every inductive set I. By the definition of an inductive set, the successor $S(n)$ must also be in every inductive set I. Therefore, $S(n) \in \mathbb{N}$. Since $\mathbb{N}$ contains $\emptyset$ and is closed under the successor operation, it is an inductive set. The inclusion $\mathbb{N} \subseteq I$ for any inductive set I follows directly from the definition of $\mathbb{N}$. ■

This construction provides a set-theoretic foundation for the natural numbers. We adopt the following standard notation.

Note. The natural numbers are defined as:

- $0 := \emptyset$
- $1 := S(0) = \{\emptyset\} = \{0\}$
- $2 := S(1) = \{\emptyset, \{\emptyset\}\} = \{0, 1\}$
- In general, $n + 1 := S(n) = n \cup \{n\} = \{0, 1, \ldots, n\}$.

This representation naturally induces an ordering on $\mathbb{N}$.

Definition 5.1.4. ***Order on*** $\mathbb{N}$. The strict order relation $<$ on $\mathbb{N}$ is defined by set membership:

$$m < n \Leftrightarrow m \in n$$

The non-strict order $\leq$ is defined as $m \leq n \Leftrightarrow (m < n \vee m = n)$, which is equivalent to $m \subseteq n$.

It can be shown that this definition makes $(\mathbb{N}, \leq)$ a linearly ordered set.

5.2 The Axiom Schema of Replacement

The Axiom of Infinity guarantees the existence of one infinite set. The Axiom Schema of Replacement provides a powerful principle for constructing new sets from existing ones. It formalises the intuition that if the domain of a function-like rule is a set, then its range should also be a set. The Schema of Separation allows us to form a subset of an existing set, whereas Replacement allows us to form a new set by applying a transformation to each element of an existing set.

***Axiom* 5.2.1. Schema of Replacement**. Let $\mathbf{P}(x, y)$ be a formula in the language of set theory whose free variables are among x, y and some other variables $w_1, \ldots, w_k$. If for any set A, the formula $\mathbf{P}$ acts like a function on the elements of A, then the collection of all outputs y forms a set. Formally:

$$\forall A\,([\forall x \in A, \exists! y, \mathbf{P}(x, y)] \Rightarrow \exists B, \forall y\,(y \in B \Leftrightarrow \exists x \in A, \mathbf{P}(x, y)))$$

The notation $\exists! y$ is shorthand for "there exists a unique y".

Remark. This is an axiom schema because it generates a separate axiom for every possible formula $\mathbf{P}(x, y)$.

The primary application of this axiom is to guarantee the existence of sets that are constructed by indexing over another set. For instance, let I be a set (the index set), and for each $i \in I$, let A_i be a set. We can use Replacement to prove that the indexed family $\{A_i \mid i \in I\}$ is a set. Let $\mathbf{P}(i, Y)$ be the formula $Y = A_i$. For any $i \in I$, the output Y is uniquely determined. By the Axiom Schema of Replacement, there exists a set

B whose elements are precisely the sets A_i for $i \in I$. This justifies operations over arbitrary families of sets, such as $\bigcup_{i \in I} A_i$.

The Axiom Schema of Replacement also provides a more direct proof for the existence of the Cartesian product. Given sets A and B, for each $a \in A$, the set $\{a\} \times B$ exists. The formula $\mathbf{P}(a, Y) \Leftrightarrow Y = \{a\} \times B$ defines a unique output for each $a \in A$. By Replacement, the set $\mathcal{F} = \{\{a\} \times B \mid a \in A\}$ exists. The Cartesian product $A \times B$ is then simply the union $\bigcup \mathcal{F}$, whose existence is guaranteed by the Axiom of Union.

5.3 The Axiom of Choice

The preceding axioms allow for the construction of sets in a deterministic manner. However, many areas of mathematics require the ability to make an infinite number of simultaneous selections, even when no explicit rule for selection exists. The Axiom of Choice provides the formal basis for such non-constructive existence proofs.

***Axiom* 5.3.1. Choice**. For any family $\mathcal{F}$ of non-empty, mutually disjoint sets, there exists a set C that contains exactly one element from each set in $\mathcal{F}$. Formally:

$$\forall S \in \mathcal{F}, |C \cap S| = 1$$

The set C is called a choice set or selector for $\mathcal{F}$.

The axiom asserts the existence of C without specifying how its elements are to be chosen. This is particularly powerful when $\mathcal{F}$ is an infinite family. A more common and equivalent formulation involves the concept of a choice function.

Definition 5.3.1. ***Choice Function***. Let $\mathcal{F}$ be a family of non-empty sets. A function $f : \mathcal{F} \to \bigcup \mathcal{F}$ is a choice function for $\mathcal{F}$ if for every set $S \in \mathcal{F}$, we have $f(S) \in S$.

The Axiom of Choice is equivalent to stating that every family of non-empty sets has a choice function. This form leads to another powerful, equivalent principle.

***Theorem* 5.3.1.** The Axiom of Choice is equivalent to the statement that for every binary relation R, there exists a function $f \subseteq R$ such that $\text{dom}(f) = \text{dom}(R)$.

Proof. Assume the Axiom of Choice. Let R be a binary relation. For each $x \in \text{dom}(R)$, the image set $R[\{x\}] = \{y \mid (x, y) \in R\}$ is non-empty. Consider the family of sets $\mathcal{F} = \{\{x\} \times R[\{x\}] \mid x \in \text{dom}(R)\}$. Each set in $\mathcal{F}$ is a collection of ordered pairs starting with x. These sets are non-empty and mutually disjoint. By the Axiom of Choice, there exists a choice set f for $\mathcal{F}$. For each $x \in \text{dom}(R)$, f contains exactly one element from $\{x\} \times R[\{x\}]$. This element must be of the form (x, y) for some unique y. Thus, f is a function with domain $\text{dom}(R)$ and $f \subseteq R$.

The converse, showing this principle implies the existence of a choice function, is left as an exercise. ■

Zorn's Lemma

Perhaps the most frequently used equivalent of the Axiom of Choice in higher mathematics is Zorn's Lemma. It provides a condition for guaranteeing the existence of maximal elements in a partially ordered set.

***Theorem* 5.3.2. Zorn's Lemma**. Let $(A, \preceq)$ be a non-empty partially ordered set. If every chain in A has an upper bound in A, then A contains at least one maximal element.

The proof that the Axiom of Choice implies Zorn's Lemma is highly non-trivial. Conversely, we can prove that Zorn's Lemma implies the Axiom of Choice.

***Theorem* 5.3.3. Equivalence of Zorn's Lemma and the Axiom of Choice**. Zorn's Lemma is equivalent to the Axiom of Choice.

Proof that Zorn's Lemma ⇒ Axiom of Choice. We will prove the equivalent form from the previous theorem. Let R be a binary relation. We want to show there exists a function $f \subseteq R$ with $\text{dom}(f) = \text{dom}(R)$.

Let $\mathcal{A}$ be the collection of all functions that are subsets of R:

$$\mathcal{A} = \{g \mid g \text{ is a function } \wedge g \subseteq R\}$$

The collection $(\mathcal{A}, \subseteq)$ is a partially ordered set. We wish to apply Zorn's Lemma to find a maximal element in $\mathcal{A}$.

Let $\mathcal{C}$ be an arbitrary chain in $(\mathcal{A}, \subseteq)$. We must show that $\mathcal{C}$ has an upper bound in $\mathcal{A}$. Let $h = \bigcup \mathcal{C}$.

- Since each $g \in \mathcal{C}$ is a subset of R, their union h is also a subset of R.
- We must show h is a function. Let $(x, y_1) \in h$ and $(x, y_2) \in h$. By definition of union, there exist $g_1, g_2 \in \mathcal{C}$ such that $(x, y_1) \in g_1$ and $(x, y_2) \in g_2$. Since $\mathcal{C}$ is a chain, g_1 and g_2 are comparable. Without loss of generality, assume $g_1 \subseteq g_2$. Then both (x, y_1) and (x, y_2) are in g_2. Since g_2 is a function, it must be that $y_1 = y_2$. Thus, h is a function.

Since h is a function and $h \subseteq R$, we have $h \in \mathcal{A}$. Furthermore, for any $g \in \mathcal{C}$, we have $g \subseteq h$, so h is an upper bound for $\mathcal{C}$.

The conditions of Zorn's Lemma are satisfied. Therefore, $\mathcal{A}$ has a maximal element, let's call it f. This f is a function and $f \subseteq R$. We claim $\text{dom}(f) = \text{dom}(R)$. Suppose, for a contradiction, that $\text{dom}(f) \neq \text{dom}(R)$. Then there exists an element $x_0 \in \text{dom}(R)$ such that $x_0 \notin \text{dom}(f)$. Since $x_0 \in \text{dom}(R)$, there is some y_0 such that $(x_0, y_0) \in R$. Consider the set $f' = f \cup \{(x_0, y_0)\}$. Since $x_0 \notin \text{dom}(f)$, f' is a function. Also, $f' \subseteq R$. Thus, $f' \in \mathcal{A}$. However, $f \subset f'$, which contradicts the maximality of f. Therefore, our assumption must be false, and $\text{dom}(f) = \text{dom}(R)$. This completes the proof. ■

The Axiom of Choice was once controversial due to its non-constructive nature. Unlike other axioms, it asserts the existence of a set without providing a rule for its construction. Over time, it has been accepted by the majority of mathematicians, as its omission would invalidate a vast number of important results across many fields of mathematics.

5.4 Recursion on the Natural Numbers

Many functions on the natural numbers are defined by specifying a base case, $f(0)$, and a recursive rule that defines $f(S(n))$ in terms of $f(n)$. The Recursion Theorem provides the formal justification that such a rule uniquely specifies a function, which is a set of ordered pairs.

***Theorem* 5.4.1. Recursion**. Let A be a set, let $a \in A$, and let $g : A \to A$ be a function. There exists a unique function $f : \mathbb{N} \to A$ such that:

(i) $f(0) = a$
(ii) $f(S(n)) = g(f(n))$ for all $n \in \mathbb{N}$.

Proof. We first prove its Existence. We construct f as the intersection of all relations satisfying the recursive property. Let $\mathcal{F}$ be the collection of all relations $h \subseteq \mathbb{N} \times A$ such that:

(a) $(0, a) \in h$
(b) $\forall n \in \mathbb{N}, \forall y \in A, ((n, y) \in h \Rightarrow (S(n), g(y)) \in h)$

The collection $\mathcal{F}$ is a set, and it is non-empty as $\mathbb{N} \times A \in \mathcal{F}$. Let $f = \bigcap \mathcal{F}$. The relation f also satisfies properties (a) and (b).

We now show that f is a function with domain $\mathbb{N}$. Let $I = \{n \in \mathbb{N} \mid \exists! y \in A, (n, y) \in f\}$. We prove that I is an inductive set.

- **Base Case**: We know $(0, a) \in f$. If there were another pair $(0, a') \in f$ with $a' \neq a$, then the relation $f' = f \setminus \{(0, a')\}$ would still be in $\mathcal{F}$. This would mean $f' \subset f$, contradicting that f is the intersection of all sets in $\mathcal{F}$. Thus, the element associated with 0 is unique, and $0 \in I$.
- **Successor Step**: Assume $n \in I$. Then there is a unique $y \in A$ such that $(n, y) \in f$. By property (b), $(S(n), g(y)) \in f$. A similar argument to the base case shows that if there were another pair $(S(n), z) \in f$, we could remove it to create a smaller relation in $\mathcal{F}$, a contradiction. Thus, the element associated with $S(n)$ is unique, and $S(n) \in I$.

Since I is an inductive set and $I \subseteq \mathbb{N}$, it follows from the definition of $\mathbb{N}$ as the smallest inductive set that $I = \mathbb{N}$. Therefore, f is a function with domain $\mathbb{N}$.

Now we prove its Uniqueness. Suppose $h : \mathbb{N} \to A$ is another function satisfying the properties. Let $J = \{n \in \mathbb{N} \mid f(n) = h(n)\}$.

- $f(0) = a = h(0)$, so $0 \in J$.
- Assume $n \in J$, so $f(n) = h(n)$. Then $f(S(n)) = g(f(n)) = g(h(n)) = h(S(n))$. Thus, $S(n) \in J$.

As J is an inductive subset of $\mathbb{N}$, we conclude $J = \mathbb{N}$, which implies $f = h$. ■

Application: Arithmetic Operations

The Recursion Theorem provides the formal tool needed to define the standard arithmetic operations on the natural numbers.

Definition 5.4.1. ***Addition on*** $\mathbb{N}$. For any fixed $m \in \mathbb{N}$, we define the function "add m", denoted $add_m : \mathbb{N} \to \mathbb{N}$. We apply the Recursion Theorem with $A = \mathbb{N}$, base case $a = m$, and the successor function $g = S$. The theorem guarantees a unique function add_m such that:

- $add_m(0) = m$
- $add_m(S(n)) = S(add_m(n))$

The binary operation of addition is then defined as $m + n := add_m(n)$. This gives the familiar recursive properties of addition: $m + 0 = m$ and $m + S(n) = S(m + n)$.

Definition 5.4.2. ***Multiplication on*** $\mathbb{N}$. With addition defined, we can define multiplication. For any fixed $m \in \mathbb{N}$, we define the function "multiply by m", $mult_m : \mathbb{N} \to \mathbb{N}$. We apply the Recursion Theorem with $A = \mathbb{N}$, base case $a = 0$, and the function $g(y) = y + m$. The theorem guarantees a unique function $mult_m$ such that:

- $mult_m(0) = 0$
- $mult_m(S(n)) = mult_m(n) + m$

The binary operation of multiplication is then defined as $m \cdot n := mult_m(n)$. This yields $m \cdot 0 = 0$ and $m \cdot S(n) = (m \cdot n) + m$.

5.5 Exercises

Part I: The Natural Numbers and Infinity

1. Using the set-theoretic definition $0 = \emptyset$ and $S(n) = n \cup \{n\}$, write out the explicit set representations for the natural numbers 3 and 4.

2. Consider the successor function $S : \mathbb{N} \to \mathbb{N}$ defined by $S(n) = n \cup \{n\}$.

(a) Prove that S is an injective function.

(b) Prove that S is not a surjective function. What is the one element of $\mathbb{N}$ that is not in the range of S?

3. Using the definition of order on $\mathbb{N}$ ($m < n \Leftrightarrow m \in n$), prove that for any $n \in \mathbb{N}$, we have $n < S(n)$.

4. Let $k \in \mathbb{N}$ be a fixed natural number. Is the set $A = \mathbb{N} \setminus \{k\}$ an inductive set? Justify your answer.

5. Prove that for any two natural numbers $m, n \in \mathbb{N}$, if $m \in n$, then it is also true that $m \subset n$ (i.e., m is a proper subset of n).

Remark. Use induction on n. The base case $n = 0$ is vacuously true. For the inductive step, assume the property holds for n and consider an element $m \in S(n)$.

Part II: The Axiom of Choice and Zorn's Lemma

6. For which of the following families of sets $\mathcal{F}$ is the Axiom of Choice required to guarantee the existence of a choice function? For those where it is not needed, explicitly define a choice function.

(a) $\mathcal{F}$ is a finite family of non-empty subsets of $\mathbb{R}$.
(b) $\mathcal{F}$ is the family of all non-empty subsets of $\mathbb{N}$.
(c) $\mathcal{F}$ is the family of all open intervals (a, b) in $\mathbb{R}$ where $a < b$.

7. The text mentions that the existence of a choice function for any family of non-empty sets is equivalent to the Axiom of Choice as stated for disjoint families. Prove the forward direction: Assume that for any family $\mathcal{G}$ of non-empty sets, there exists a choice function $f : \mathcal{G} \to \bigcup \mathcal{G}$. Use this to prove that for any family $\mathcal{F}$ of non-empty, mutually disjoint sets, there exists a choice set C.

8. Let $(A, \preceq)$ be a partially ordered set. A *chain* in A is a subset $C \subseteq A$ that is totally ordered by $\preceq$. Use Zorn's Lemma to prove that every non-empty poset contains at least one maximal chain.

Remark. Consider the collection $\mathcal{C}$ of all chains in A, ordered by set inclusion $\subseteq$. Show that this new poset $(\mathcal{C}, \subseteq)$ satisfies the condition of Zorn's Lemma.

9. Determine if Zorn's Lemma can be applied to the following posets. If so, identify a maximal element. If not, explain which condition fails.

(a) The set $\mathcal{P}(\mathbb{N}) \setminus \{\mathbb{N}\}$ of all proper subsets of $\mathbb{N}$, ordered by set inclusion $\subseteq$.
(b) The set of all finite subsets of $\mathbb{R}$, ordered by set inclusion $\subseteq$.

Part III: Recursion and Arithmetic

10. Using the Recursion Theorem, provide a formal definition for the exponentiation function $exp_m(n) = m^n$ for $m, n \in \mathbb{N}$ with $m \neq 0$. Specify the set A, the base case element a, and the function g used in the theorem.

Remark. For a fixed m, define m^0 and then define $m^{S(n)}$ in terms of m^n and multiplication.

11. Using only the recursive definition of addition ($m + 0 = m$ and $m + S(n) = S(m + n)$) and the fact that $\mathbb{N}$ is the smallest inductive set (i.e., proof by induction), prove that $0 + n = n$ for all $n \in \mathbb{N}$.

12. Following on from the previous exercise, prove the associativity of addition in $\mathbb{N}$. That is, prove that for all $m, n, k \in \mathbb{N}$,

$$(m + n) + k = m + (n + k)$$

Remark. Fix m and n and proceed by induction on k.

13. $\star$ The order relation $<$ on $\mathbb{N}$ was defined set-theoretically as $m < n \Leftrightarrow m \in n$. We can also define an arithmetic order $m <_a n$ to mean that there exists a non-zero natural number k such that $m + k = n$. Prove that these two definitions are equivalent for all $m, n \in \mathbb{N}$.

5.6 Arithmetic Operations on $\mathbb{N}$

Having justified the existence of arithmetic operations using the Recursion Theorem, we now examine their properties.

Addition

Recall that for any $m, n \in \mathbb{N}$, the sum $m + n$ is determined by the recursive rules:

$$m + 0 = m \quad \text{and} \quad m + S(n) = S(m + n)$$

This definition immediately gives $m + 1 = m + S(0) = S(m + 0) = S(m)$. The fundamental properties of addition can be established from this definition. The proofs rely on showing that the set of numbers for which a property holds is an inductive set.

***Lemma* 5.6.1.** For all $m, n \in \mathbb{N}$:

(i) $0 + n = n$
(ii) $S(m) + n = S(m + n)$

Proof. (i) Let $I = \{n \in \mathbb{N} \mid 0 + n = n\}$. The base case $0 + 0 = 0$ holds by definition, so $0 \in I$. Assume $k \in I$, so $0 + k = k$. Then $0 + S(k) = S(0 + k) = S(k)$, which implies $S(k) \in I$. Thus, I is an inductive set, and $I = \mathbb{N}$.

(ii) Fix $m \in \mathbb{N}$ and let $I = \{n \in \mathbb{N} \mid S(m) + n = S(m + n)\}$. For the base case, $S(m) + 0 = S(m)$ by definition, and $S(m+0) = S(m)$. Thus, $0 \in I$. Assume $k \in I$, so $S(m)+k = S(m+k)$. Then $S(m)+S(k) = S(S(m) + k) = S(S(m + k))$. Also, $S(m + S(k)) = S(S(m + k))$. Therefore, $S(m) + S(k) = S(m + S(k))$, which means $S(k) \in I$. Thus, $I = \mathbb{N}$. ■

***Theorem* 5.6.1. Properties of Addition**. For all $m, n, p \in \mathbb{N}$:

(i) **Associativity:** $(m + n) + p = m + (n + p)$
(ii) **Commutativity:** $m + n = n + m$

The number 0 is the unique additive identity.

Proof. (i) The proof for associativity is left as an exercise.

(ii) To prove commutativity, we fix $m \in \mathbb{N}$ and let $I = \{n \in \mathbb{N} \mid m + n = n + m\}$. The base case $m + 0 = m$ and $0 + m = m$ (by the previous lemma) shows $0 \in I$. Assume $k \in I$, so $m + k = k + m$. We must show $m + S(k) = S(k) + m$.

$$\begin{aligned} m + S(k) &= S(m + k) && \text{by definition of addition} \\ &= S(k + m) && \text{by the inductive hypothesis} \\ &= S(k) + m && \text{by the previous lemma} \end{aligned}$$

Thus, $S(k) \in I$. This proves $I = \mathbb{N}$, establishing commutativity. The role of 0 as the identity follows from the definition and the lemma. ■

Multiplication

Similarly, multiplication is defined recursively. Recall that for any $m, n \in \mathbb{N}$, the product $m \cdot n$ satisfies:

$$m \cdot 0 = 0 \quad \text{and} \quad m \cdot S(n) = (m \cdot n) + m$$

Analogous to addition, we establish key properties needed for the main proofs.

***Lemma* 5.6.2.** For all $m, n \in \mathbb{N}$:

(i) $0 \cdot n = 0$
(ii) $S(m) \cdot n = (m \cdot n) + n$

Proof. (i) Let $I = \{n \in \mathbb{N} \mid 0 \cdot n = 0\}$. The base case $0 \cdot 0 = 0$ holds, so $0 \in I$. Assume $k \in I$. Then $0 \cdot S(k) = (0 \cdot k) + 0 = 0 + 0 = 0$. Thus, $S(k) \in I$, which implies $I = \mathbb{N}$.

(ii) This proof is similar and is left as an exercise. ■

***Theorem* 5.6.2. Properties of Multiplication and Distributivity**. For all $m, n, p \in \mathbb{N}$:

(i) **Associativity:** $(m \cdot n) \cdot p = m \cdot (n \cdot p)$
(ii) **Commutativity:** $m \cdot n = n \cdot m$
(iii) **Distributivity:** $m \cdot (n + p) = (m \cdot n) + (m \cdot p)$

The number $1 = S(0)$ is the unique multiplicative identity.

Proof. We prove distributivity. The proofs for associativity, commutativity, and the identity property are left as exercises. Fix $m, n \in \mathbb{N}$ and let $I = \{p \in \mathbb{N} \mid m \cdot (n + p) = (m \cdot n) + (m \cdot p)\}$. For the base case $p = 0$, $m \cdot (n + 0) = m \cdot n$. Also, $(m \cdot n) + (m \cdot 0) = (m \cdot n) + 0 = m \cdot n$. Thus, $0 \in I$. Assume $k \in I$, so $m \cdot (n + k) = (m \cdot n) + (m \cdot k)$. We examine the case for $S(k)$.

$$\begin{aligned} m \cdot (n + S(k)) &= m \cdot S(n + k) && \text{by definition of addition} \\ &= (m \cdot (n + k)) + m && \text{by definition of multiplication} \\ &= ((m \cdot n) + (m \cdot k)) + m && \text{by inductive hypothesis} \\ &= (m \cdot n) + ((m \cdot k) + m) && \text{by associativity of addition} \\ &= (m \cdot n) + (m \cdot S(k)) && \text{by definition of multiplication} \end{aligned}$$

Therefore, $S(k) \in I$, which implies $I = \mathbb{N}$. ■

Finally, we establish the cancellation laws, which are crucial for solving equations.

***Theorem* 5.6.3. Cancellation Laws**. Let $m, n, p \in \mathbb{N}$.

(i) If $m + p = n + p$, then $m = n$.
(ii) If $m \cdot p = n \cdot p$ and $p \neq 0$, then $m = n$.

Proof. (i) Let $I = \{p \in \mathbb{N} \mid \forall m, n(m + p = n + p \Rightarrow m = n)\}$. The base case $p = 0$ is trivial, as $m + 0 = n + 0 \Rightarrow m = n$. So $0 \in I$. Assume $k \in I$. Let $m + S(k) = n + S(k)$. By definition, this is $S(m + k) = S(n + k)$. Since the successor function is injective (a consequence of the properties of $\mathbb{N}$), we have $m + k = n + k$. By the inductive hypothesis, $m = n$. Thus $S(k) \in I$, and the result follows.

(ii) The proof for multiplication is left as an exercise. ■

5.7 Exercises

14. Prove the following properties of multiplication in $\mathbb{N}$.

(i) Commutativity property of multiplication in $\mathbb{N}$. That is, for all $m, n \in \mathbb{N}$, prove that $m \cdot n = n \cdot m$.

Remark. Fix m and proceed by induction on n. You will first need to prove the lemma that $S(m) \cdot n = (m \cdot n) + n$, which was left as an exercise in the text.

(ii) Associativity property of multiplication in $\mathbb{N}$. That is, for all $m, n, p \in \mathbb{N}$, prove that $(m \cdot n) \cdot p = m \cdot (n \cdot p)$.

Remark. Fix m and n and proceed by induction on p. The distributive law, which was proven in the text, will be essential.

15. Using the arithmetic definition of order from the previous exercise set ($m < n$ if there exists $k \in \mathbb{N} \setminus \{0\}$ such that $m + k = n$), prove that the order is compatible with the arithmetic operations:

(a) If $m < n$, then $m + p < n + p$ for all $p \in \mathbb{N}$.

(b) If $m < n$ and $p \neq 0$, then $m \cdot p < n \cdot p$.

16. Prove that the natural numbers have no zero divisors. That is, for any $m, n \in \mathbb{N}$, if $m \cdot n = 0$, then $m = 0$ or $n = 0$.

Remark. Assume for contradiction that $m \neq 0$ and $n \neq 0$. This implies $m = S(k_1)$ and $n = S(k_2)$ for some $k_1, k_2 \in \mathbb{N}$. Show that their product cannot be 0.

17. $\star$ Complete the proof of the Cancellation Laws by proving the law for multiplication. For all $m, n, p \in \mathbb{N}$, if $m \cdot p = n \cdot p$ and $p \neq 0$, then $m = n$.

Remark. Assume for contradiction that $m \neq n$. By the trichotomy of order (which you may assume), one must be larger, say $m < n$. Then apply the result from exercise 16.

5.8 Constructing the Integers

The set of natural numbers $\mathbb{N}$ is closed under addition and multiplication. However, an equation of the form $x + n = m$ may have no solution within $\mathbb{N}$ (for instance, $x + 5 = 2$). To provide a system where such equations are always solvable, we must construct the negative integers and formalise the concept of zero.

The intuition is to represent an integer as a formal difference $m - n$ for some $m, n \in \mathbb{N}$. Since multiple pairs can represent the same integer (e.g., $5 - 2$ and $4 - 1$ both correspond to 3), we define an equivalence relation on the set of pairs $\mathbb{N} \times \mathbb{N}$ to group them appropriately. The condition $m - n = p - q$ is rewritten using only addition on $\mathbb{N}$ as $m + q = n + p$.

Definition 5.8.1. ***Equivalence Relation for Integers***. Let $\sim$ be a binary relation on $\mathbb{N} \times \mathbb{N}$ defined such that for any pairs (m, n) and (p, q),

$$(m, n) \sim (p, q) \Leftrightarrow m + q = n + p$$

Remark. It is a straightforward exercise to show that $\sim$ is an equivalence relation on $\mathbb{N} \times \mathbb{N}$.

Definition 5.8.2. ***The Set of Integers***. The set of integers, denoted $\mathbb{Z}$, is the quotient set of $\mathbb{N} \times \mathbb{N}$ by the equivalence relation $\sim$.

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\sim$$

An element of $\mathbb{Z}$ is an equivalence class, denoted $[(m, n)]$.

Note. Each equivalence class corresponds to an integer:

- Positive integers are classes of the form $[(n, 0)]$ for $n > 0$. For example, $+3 = [(3, 0)]$.
- Negative integers are classes of the form $[(0, n)]$ for $n > 0$. For example, $-3 = [(0, 3)]$.
- The integer zero is the class $[(0, 0)] = \{(k, k) \mid k \in \mathbb{N}\}$.

Example 5.8.1. The class $[(1, 4)]$ represents the integer -3, as does $[(0, 3)]$ and $[(5, 8)]$, since $1 + 3 = 4 + 0$ and $1 + 8 = 4 + 5$.

Order and Arithmetic on the Integers

The order and arithmetic operations on $\mathbb{Z}$ are defined in terms of the corresponding structures on $\mathbb{N}$.

Definition 5.8.3. ***Order on*** $\mathbb{Z}$. Let $\alpha = [(m, n)]$ and $\beta = [(p, q)]$ be integers. The order relation $\leq$ on $\mathbb{Z}$ is defined by:

$$\alpha \leq \beta \Leftrightarrow m + q \leq n + p$$

where the $\leq$ on the right is the order on $\mathbb{N}$.

***Theorem* 5.8.1.** The pair $(\mathbb{Z}, \leq)$ is a linearly ordered set.

Proof. Reflexivity, antisymmetry, and transitivity follow from the corresponding properties of $\leq$ on $\mathbb{N}$. To show it is a linear order, we must show any two elements are comparable. For any $\alpha = [(m,n)]$ and $\beta = [(p,q)]$, consider the natural numbers $m+q$ and $n+p$. Since $(\mathbb{N}, \leq)$ is a linear order, either $m+q \leq n+p$ or $n+p \leq m+q$. This implies either $\alpha \leq \beta$ or $\beta \leq \alpha$. ■

Definition 5.8.4. ***Arithmetic on*** $\mathbb{Z}$. Let $\alpha = [(m,n)]$ and $\beta = [(p,q)]$ be integers.

- **Addition**: $\alpha + \beta := [(m+p, n+q)]$
- **Multiplication**: $\alpha \cdot \beta := [(mp+nq, mq+np)]$

Remark. For these definitions to be valid, we must show they are well-defined; that is, the result of an operation is independent of the choice of representatives from the equivalence classes. We prove this for addition.

Suppose $(m,n) \sim (m',n')$ and $(p,q) \sim (p',q')$. This means $m+n' = n+m'$ and $p+q' = q+p'$. We must show that $[(m+p, n+q)] = [(m'+p', n'+q')]$, which requires showing $(m+p)+(n'+q') = (n+q)+(m'+p')$. By associativity and commutativity of addition on $\mathbb{N}$:

$$(m+p)+(n'+q') = (m+n')+(p+q') = (n+m')+(q+p') = (n+q)+(m'+p')$$

The operation is well-defined. The proof for multiplication is analogous.

These operations inherit their algebraic properties from the operations on $\mathbb{N}$.

***Theorem* 5.8.2. Properties of Integer Arithmetic**. Let $\alpha, \beta, \gamma \in \mathbb{Z}$.

(i) Addition and multiplication are associative and commutative.
(ii) The additive identity is $[(0,0)]$, and the multiplicative identity is $[(1,0)]$.
(iii) Every integer $\alpha = [(m,n)]$ has a unique additive inverse, $-\alpha = [(n,m)]$.
(iv) The distributive law holds: $\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$.
(v) $\mathbb{Z}$ is an integral domain.

Proof of (iii) and (v). (iii) Let $\alpha = [(m,n)]$. Its additive inverse is claimed to be $-\alpha = [(n,m)]$. We verify this:

$$\alpha + (-\alpha) = [(m,n)] + [(n,m)] = [(m+n, n+m)]$$

Since $m+n = n+m$ in $\mathbb{N}$, we have $(m+n, n+m) \sim (0,0)$. Thus, $[(m+n, n+m)] = [(0,0)]$, which is the additive identity. The other properties follow similarly from the properties of arithmetic on $\mathbb{N}$.

(v) The preceding points establish that $(\mathbb{Z}, +, \cdot)$ is a commutative ring with unity. To show it is an integral domain, we must prove it has no zero divisors. Let $\alpha = [(m,n)]$ and $\beta = [(p,q)]$ be elements of $\mathbb{Z}$ such that their product is the additive identity, $\alpha \cdot \beta = [(0,0)]$. By the definition of multiplication on $\mathbb{Z}$, we have:

$$\alpha \cdot \beta = [(mp+nq, mq+np)] = [(0,0)]$$

This equivalence implies $mp+nq = mq+np$, an equality within the natural numbers $\mathbb{N}$. We analyse this relation under the assumption that $\alpha \neq [(0,0)]$, which means $m \neq n$. Since $(\mathbb{N}, \leq)$ is a linear order, we have two sub-cases.

- **Case 1:** $m > n$**.** There exists a non-zero natural number d such that $m = n+d$. Substituting this into our equality:

$$\begin{aligned} (n+d)p + nq &= (n+d)q + np \\ np + dp + nq &= nq + dq + np && \text{by distributivity in } \mathbb{N} \\ (np+nq) + dp &= (np+nq) + dq && \text{by associativity and commutativity in } \mathbb{N} \end{aligned}$$

  By the cancellation law for addition on $\mathbb{N}$, we conclude $dp = dq$. As $d \neq 0$, the cancellation law for multiplication on $\mathbb{N}$ implies $p = q$. This means $\beta = [(p,p)] = [(0,0)]$.

- **Case 2:** $n > m$**.** A symmetric argument holds. There exists $d \in \mathbb{N} \setminus \{0\}$ such that $n = m + d$. Substituting this leads to $mp + (m + d)q = mq + (m + d)p$, which simplifies to $dq = dp$, and again yields $p = q$. Thus, $\beta = [(0, 0)]$.

In all scenarios where $\alpha \neq [(0, 0)]$, we are forced to conclude that $\beta = [(0, 0)]$. Therefore, if $\alpha \cdot \beta = 0$, it must be that either $\alpha = 0$ or $\beta = 0$. $\mathbb{Z}$ has no zero divisors and is an integral domain. ■

The existence of additive inverses allows for a formal definition of subtraction.

Algebraic Structures: Rings and Fields

The properties established for $\mathbb{Z}$ are characteristic of a general mathematical structure known as a ring. Formalising this concept provides a powerful language for discussing number systems.

Definition 5.8.5. ***Ring***. A set R equipped with two binary operations, addition $(+)$ and multiplication $(\cdot)$, is a ring if for all $a, b, c \in R$:

(i) **Additive Group Properties:** Addition is associative $((a + b) + c = a + (b + c))$, commutative $(a + b = b + a)$, possesses an identity element 0, and every element a has an additive inverse $-a$.
(ii) **Multiplicative Associativity:** Multiplication is associative $((a \cdot b) \cdot c = a \cdot (b \cdot c))$.
(iii) **Distributive Law:** Multiplication distributes over addition $(a \cdot (b+c) = (a \cdot b)+(a \cdot c)$ and $(a+b) \cdot c = (a \cdot c) + (b \cdot c))$.

A ring is commutative if its multiplication is commutative. A ring with unity possesses a multiplicative identity 1.

Remark. From the properties previously established, the set of integers $(\mathbb{Z}, +, \cdot)$ forms a commutative ring with unity.

***Lemma* 5.8.1.** *Properties of Rings.* Let R be a ring with additive identity 0. For all $a, b \in R$:

(i) $a \cdot 0 = 0 \cdot a = 0$
(ii) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$
(iii) $(-a) \cdot (-b) = a \cdot b$

Proof. (i) We have $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$. Adding the inverse $-(a \cdot 0)$ to both sides gives $0 = a \cdot 0$. The proof for $0 \cdot a = 0$ is analogous. (ii) From $b + (-b) = 0$, the distributive law gives $a \cdot (b + (-b)) = a \cdot 0$. This implies $a \cdot b + a \cdot (-b) = 0$. By the uniqueness of the additive inverse, $a \cdot (-b) = -(a \cdot b)$. (iii) Using the previous result twice, we find $(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b))$. Since the inverse of an inverse is the element itself, this simplifies to $a \cdot b$. ■

Remark. An important property of some rings is the absence of zero divisors. An element $a \neq 0$ in a ring R is a zero divisor if there exists a non-zero element $b \in R$ such that $a \cdot b = 0$ or $b \cdot a = 0$.

Definition 5.8.6. ***Integral Domain***. An integral domain is a commutative ring with unity $1 \neq 0$ that has no zero divisors. That is, for any a, b in the ring, if $a \cdot b = 0$, then either $a = 0$ or $b = 0$.

While every element in $\mathbb{Z}$ has an additive inverse, the same is not true for multiplication. This motivates the definition of a field, where division by non-zero elements is always possible.

Definition 5.8.7. ***Field***. A field is a commutative ring with unity $1 \neq 0$ in which every non-zero element possesses a multiplicative inverse. That is, for every $a \neq 0$, there exists an element a^{-1} such that $a \cdot a^{-1} = 1$.

The existence of multiplicative inverses for all non-zero elements ensures that a field can have no zero divisors. Suppose $a \cdot b = 0$ and $a \neq 0$. Since a has an inverse a^{-1}, we can write $a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0$. By associativity, this becomes $(a^{-1} \cdot a) \cdot b = 0$, which simplifies to $1 \cdot b = 0$, or $b = 0$. Thus, every field is an integral domain.

Remark. The integers $\mathbb{Z}$ do not form a field, as only 1 and -1 have multiplicative inverses within $\mathbb{Z}$. The construction of the rational numbers remedies this deficiency.

Definition 5.8.8. ***Subtraction***. For any integers α, β, subtraction is defined as the addition of the inverse: $\alpha - \beta := \alpha + (-\beta)$.

Rational Numbers

The set of integers $\mathbb{Z}$ is closed under addition, subtraction, and multiplication. However, it is not closed under division; an equation like $2x = 1$ has no solution in $\mathbb{Z}$. To create a number system where division by non-zero elements is always possible, we construct the rational numbers.

A rational number is conceived as a fraction m/n, where $m, n \in \mathbb{Z}$ and $n \neq 0$. This representation is not unique, as, for example, $1/2 = 2/4$. The general rule for equivalence is $m/n = p/q \Leftrightarrow mq = np$. We use this rule to define an equivalence relation on pairs of integers.

Definition 5.8.9. ***Equivalence Relation for Rationals***. Let $\approx$ be a binary relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ defined such that for any pairs (m, n) and (p, q),

$$(m, n) \approx (p, q) \Leftrightarrow m \cdot q = n \cdot p$$

Definition 5.8.10. ***The Set of Rational Numbers***. The set of rational numbers, denoted $\mathbb{Q}$, is the quotient set of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by the equivalence relation $\approx$.

$$\mathbb{Q} := (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\approx$$

An element of $\mathbb{Q}$, denoted $[(m, n)]$, is commonly written as the fraction m/n.

The structures on $\mathbb{Q}$ are defined in a way that mimics the familiar rules of fraction arithmetic.

Definition 5.8.11. ***Order on*** $\mathbb{Q}$. Let $\alpha = [(m, n)]$ and $\beta = [(p, q)]$ be rational numbers. We can always choose representatives such that $n > 0$ and $q > 0$, since $[(m, n)] = [(-m, -n)]$. With this convention, the order relation $\leq$ on $\mathbb{Q}$ is defined by:

$$\alpha \leq \beta \Leftrightarrow m \cdot q \leq n \cdot p$$

where the $\leq$ on the right is the order on $\mathbb{Z}$.

It can be shown that $(\mathbb{Q}, \leq)$ is a linearly ordered set.

Definition 5.8.12. ***Arithmetic on*** $\mathbb{Q}$. Let $\alpha = [(m, n)]$ and $\beta = [(p, q)]$ be rational numbers.

- **Addition**: $\alpha + \beta := [(mq + np, nq)]$
- **Multiplication**: $\alpha \cdot \beta := [(mp, nq)]$

These operations are well-defined, meaning the result is independent of the choice of representatives for the equivalence classes.

***Theorem* 5.8.3. Properties of Rational Arithmetic**. The set $\mathbb{Q}$ with the operations of addition and multiplication has the following properties:

(i) Addition and multiplication are associative and commutative.
(ii) The additive identity is $0 = [(0, 1)]$, and the multiplicative identity is $1 = [(1, 1)]$.
(iii) Every rational number α has a unique additive inverse, $-\alpha$.
(iv) Every non-zero rational number $\alpha = [(m, n)]$ has a unique multiplicative inverse, $\alpha^{-1} = [(n, m)]$.
(v) The distributive law holds.

Proof. The proofs of these properties follow from the definitions and the corresponding properties of the integers. We prove the existence of the multiplicative inverse. Let $\alpha = [(m, n)]$ be a non-zero rational

number, which implies $m \neq 0$ and $n \neq 0$. Its inverse is $\alpha^{-1} = [(n, m)]$, which is a valid rational number as $m \neq 0$. Then:

$$\alpha \cdot \alpha^{-1} = [(m, n)] \cdot [(n, m)] = [(mn, nm)]$$

Since $mn = nm$ in $\mathbb{Z}$, we have $(mn, nm) \approx (1, 1)$, so $[(mn, nm)] = [(1, 1)]$, which is the multiplicative identity. ■

The properties established demonstrate that $(\mathbb{Q}, +, \cdot)$ is a field. Furthermore, the order relation $\leq$ is compatible with these algebraic operations. This structure is known as an ordered field.

Definition 5.8.13. ***Ordered Field***. An ordered field is a field K together with a total order relation $\leq$ such that for all $x, y, z \in K$:

(OR1) If $x \leq y$, then $x + z \leq y + z$.
(OR2) If $x \geq 0$ and $y \geq 0$, then $x \cdot y \geq 0$.

***Theorem* 5.8.4. Properties of Ordered Fields**. Let K be an ordered field. For all $x, y, a \in K$:

(i) $x > y \Leftrightarrow x - y > 0$.
(ii) If $x > y$ and $a > b$, then $x + a > y + b$.
(iii) If $x > 0$, then $-x < 0$. If $x < 0$, then $-x > 0$.
(iv) If $x > 0$ and $y > 0$, then $x \cdot y > 0$. If $x > 0$ and $y < 0$, then $x \cdot y < 0$.
(v) If $a > 0$ and $x > y$, then $ax > ay$.
(vi) If $a < 0$ and $x > y$, then $ax < ay$.
(vii) $x^2 \geq 0$ for all $x \in K$. In particular, $1 = 1^2 > 0$.
(viii) If $x > 0$, then $x^{-1} > 0$.
(ix) If $x > y > 0$, then $0 < x^{-1} < y^{-1}$.

Proof. These properties are direct consequences of the axioms. As an example, we prove (iii). If $a < 0$, then $-a > 0$. Given $x > y$, we have $x - y > 0$. By (OR2), $(-a)(x - y) > 0$, which simplifies to $-ax + ay > 0$, or $ay > ax$. The other proofs are left as exercises. ■

Definition 5.8.14. ***Complete Ordered Field***. An ordered field $(K, +, \cdot, \leq)$ is complete if every non-empty subset $S \subseteq K$ that is bounded above has a least upper bound in K. A complete ordered field is an ordered field with this property.

Remark. The existence of multiplicative inverses for all non-zero elements means that $\mathbb{Q}$ is a field. With these constructions, we have built the number systems $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ from the ground up using only the axioms of set theory. Since the defined operations and orderings on these sets possess the standard properties we associate with these number systems, we hereafter treat them as the familiar sets of numbers used throughout mathematics.

5.9 Exercises

18. The text states that the relation $\sim$ on $\mathbb{N} \times \mathbb{N}$, defined by $(m, n) \sim (p, q) \Leftrightarrow m + q = n + p$, is an equivalence relation. Prove this by showing that $\sim$ is reflexive, symmetric, and transitive.

19. The text provides a proof that addition on $\mathbb{Z}$ is well-defined. Prove that multiplication on $\mathbb{Z}$ is also well-defined. That is, if $(m, n) \sim (m', n')$ and $(p, q) \sim (p', q')$, show that

$$[(mp + nq, mq + np)] = [(m'p' + n'q', m'q' + n'p')]$$

20. Prove the distributive law for the integers. Using the definitions of addition and multiplication on $\mathbb{Z}$, show that for any $\alpha, \beta, \gamma \in \mathbb{Z}$,

$$\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$$

21. The set of natural numbers $\mathbb{N}$ can be seen as a subset of the integers $\mathbb{Z}$ via an embedding. Consider the function $i : \mathbb{N} \to \mathbb{Z}$ defined by $i(n) = [(n, 0)]$.

(a) Prove that this function i is injective.

(b) Prove that this function preserves addition; that is, show that for all $m, n \in \mathbb{N}$, $i(m+n) = i(m) + i(n)$.

22. $\star$ The construction of the rational numbers $\mathbb{Q}$ uses the set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, explicitly forbidding zero in the second coordinate. Suppose we had ignored this and tried to define the same equivalence relation $\approx$ on the set $\mathbb{Z} \times \mathbb{Z}$. Prove that this relation would no longer be an equivalence relation on this larger set.

Remark. Investigate which property of an equivalence relation fails. Consider pairs of the form $(m, 0)$.

23. Well-Definedness of Rational Addition. In the text, we stated that arithmetic operations on $\mathbb{Q}$ are well-defined. Prove this for addition. Suppose $\alpha = [(m, n)]$ and $\beta = [(p, q)]$. Let $(m, n) \approx (m', n')$ and $(p, q) \approx (p', q')$. Show that the sum is independent of the representative chosen:

$$(m, n) + (p, q) \approx (m', n') + (p', q')$$

Remark. You need to show that if $mn' = nm'$ and $pq' = qp'$, then $(mq+np)(n'q') = (nq)(m'q'+n'p')$.

24. Properties of Ordered Fields I. Using the axioms (OR1) and (OR2) and the properties of fields, prove the following statements from the *Properties of Ordered Fields* theorem:

(a) **Additivity of Order**: If $x > y$ and $a > b$, prove that $x + a > y + b$.

Remark. Use (OR1) twice: first add a to $x > y$, then add y to $a > b$.

(b) **Positivity of Squares**: Prove that for any $x \neq 0$, $x^2 > 0$. Conclude that $1 > 0$.

Remark. Consider two cases: $x > 0$ and $x < 0$ (where $-x > 0$). Use (OR2).

25. Properties of Ordered Fields II. Continuing with the properties of ordered fields, prove the following regarding inverses:

(a) **Positivity of Inverses**: If $x > 0$, prove that $x^{-1} > 0$.

Remark. Proof by contradiction: Assume $x^{-1} \leq 0$. If $x^{-1} = 0$, $x \cdot x^{-1} = 0 \neq 1$. If $x^{-1} < 0$, multiply by x (which is positive) and check the sign of 1.

(b) **Ordering of Inverses**: If $x > y > 0$, prove that $x^{-1} < y^{-1}$.

Remark. Multiply the inequality $x > y$ by the positive quantity $x^{-1}y^{-1}$.

Chapter 6

Mathematical Induction

To prove a formula $\mathbf{P}(n)$ for all integers n greater than or equal to some integer n_0, one might attempt to prove it for each n individually. This approach is impossible as it requires infinitely many steps. An alternative, rigorous method is provided by the principle of mathematical induction.

6.1 The Principle of Induction

The validity of proofs by induction rests upon the following theorem, which formalises the structure of such arguments.

***Theorem* 6.1.1. Principle of Mathematical Induction**. Let $\mathbf{P}(n)$ be a formula. For any integer n_0, if

$$\mathbf{P}(n_0) \wedge \forall n \in \mathbb{Z}\,((n \geq n_0 \wedge \mathbf{P}(n)) \Rightarrow \mathbf{P}(n+1))$$

then

$$\forall n \in \mathbb{Z}(n \geq n_0 \Rightarrow \mathbf{P}(n)).$$

Proof. Assume $\mathbf{P}(n_0)$ and that for all integers $n \geq n_0$, $\mathbf{P}(n) \Rightarrow \mathbf{P}(n+1)$. Define the set

$$S = \{k \in \mathbb{N} \mid \mathbf{P}(n_0 + k)\}$$

We demonstrate that S is an inductive set.

- **Base Case:** By hypothesis, $\mathbf{P}(n_0)$ is true, which implies $0 \in S$.
- **Inductive Step:** Assume $k \in S$. By definition of S, this means $\mathbf{P}(n_0 + k)$ is true. Let $n = n_0 + k$. Since $k \in \mathbb{N}$, we have $n \geq n_0$. Our initial assumption states that for such n, $\mathbf{P}(n) \Rightarrow \mathbf{P}(n+1)$. Therefore, $\mathbf{P}(n_0 + k) \Rightarrow \mathbf{P}(n_0 + k + 1)$. As the antecedent is true, $\mathbf{P}(n_0 + k + 1)$ must be true. This implies $k + 1 \in S$.

Since S contains 0 and is closed under the successor operation, $S = \mathbb{N}$. This means $\mathbf{P}(n_0 + k)$ holds for all $k \in \mathbb{N}$, which is equivalent to stating that $\mathbf{P}(n)$ holds for all integers $n \geq n_0$. ■

An analogy is often drawn with a row of dominoes. Proving the first part of the premise, $\mathbf{P}(n_0)$, is akin to tipping the first domino. Proving the second part, the implication, ensures that the dominoes are positioned such that each one will knock over the next. The conclusion follows by modus ponens: $\mathbf{P}(n_0)$ implies $\mathbf{P}(n_0 + 1)$, which in turn implies $\mathbf{P}(n_0 + 2)$, and so on, with each proposition falling in sequence. This two-step process is characteristic of proofs by induction.

Basis Case The proof of $\mathbf{P}(n_0)$.

Induction Step The proof that for an arbitrary integer $n \geq n_0$, the implication $\mathbf{P}(n) \Rightarrow \mathbf{P}(n+1)$ holds. The assumption that $\mathbf{P}(n)$ is true for the purpose of proving this implication is known as the *induction hypothesis*.

Applications of Induction

We now demonstrate the application of this principle with several examples.

Example 6.1.1. Let x be a positive rational number. We prove by induction that for any $n \in \mathbb{Z}^+$, $(1+x)^n \geq 1 + nx$.

- **Basis Case:** For $n = 1$, the statement is $(1+x)^1 \geq 1 + (1)x$, which is an equality and therefore true.
- **Induction Step:** Let $k \in \mathbb{Z}^+$ and assume the induction hypothesis: $(1+x)^k \geq 1 + kx$. We aim to show $(1+x)^{k+1} \geq 1 + (k+1)x$. We begin with the left-hand side of the target inequality:

$$\begin{aligned}
(1+x)^{k+1} &= (1+x)^k(1+x) \\
&\geq (1+kx)(1+x) && \text{by the induction hypothesis, since } 1+x > 0 \\
&= 1 + x + kx + kx^2 \\
&= 1 + (k+1)x + kx^2 \\
&\geq 1 + (k+1)x && \text{since } k \in \mathbb{Z}^+ \text{ and } x \in \mathbb{Q}^+, kx^2 \geq 0
\end{aligned}$$

Thus, $(1+x)^{k+1} \geq 1 + (k+1)x$. By the principle of mathematical induction, the inequality holds for all $n \in \mathbb{Z}^+$.

Example 6.1.2. We prove by induction that $n^3 < n!$ for all integers $n \geq 6$.

- **Basis Case:** For $n = 6$, we have $6^3 = 216$ and $6! = 720$. The inequality $216 < 720$ is true.
- **Induction Step:** Let $k \geq 6$ be an integer and assume the induction hypothesis $k^3 < k!$. We must show that $(k+1)^3 < (k+1)!$. We expand the left-hand side:

$$(k+1)^3 = k^3 + 3k^2 + 3k + 1$$

By the induction hypothesis, $k^3 < k!$, so we have:

$$(k+1)^3 < k! + 3k^2 + 3k + 1$$

To complete the proof, we must show that $k! + 3k^2 + 3k + 1 < (k+1)!$. We observe that for $k \geq 6$, the terms $3k^2$, $3k$, and 1 are all smaller than $k!$. Specifically, $3k^2 < k^3 < k!$ for $k > 3$, and $3k < k^3 < k!$ for $k > \sqrt{3}$. Therefore:

$$k! + 3k^2 + 3k + 1 < k! + k! + k! + k! = 4k!$$

Now we need to show that $4k! < (k+1)!$. This is equivalent to $4k! < (k+1)k!$, which simplifies to $4 < k+1$, or $k > 3$. Since our assumption is $k \geq 6$, this condition is satisfied. We have established the chain of inequalities:

$$(k+1)^3 < k! + 3k^2 + 3k + 1 < 4k! < (k+1)!$$

By the principle of induction, the statement holds for all integers $n \geq 6$.

6.2 Exercises

1. Let B, A_1, A_2, …, A_n be sets. Prove the generalised distributive law by induction for all $n \in \mathbb{Z}^+$:

$$B \cap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (B \cap A_i)$$

You may assume the law for two sets, $B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap A_2)$, as a known property.

2. Let $A_1, A_2, \ldots, A_n$ be a finite collection of sets within some universal set U. Prove the generalised De Morgan's Law by induction for all $n \in \mathbb{Z}^+$:

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

You may assume the law for two sets, $(A_1 \cup A_2)^c = A_1^c \cap A_2^c$, as a known property.

3. Let $f_1, f_2, \ldots, f_n$ be a sequence of bijective functions, where each function f_i maps a set A to itself ($f_i : A \to A$). Prove by induction that the composition $F_n = f_n \circ f_{n-1} \circ \cdots \circ f_1$ is also a bijection from A to A for all $n \in \mathbb{Z}^+$.

Remark. You may use the theorem stating that the composition of two bijections is a bijection.

4. Let R be a symmetric binary relation on a set A. Define the powers of R recursively by $R^1 = R$ and $R^{n+1} = R^n \circ R$ for $n \in \mathbb{Z}^+$. Prove by induction that R^n is a symmetric relation for all $n \in \mathbb{Z}^+$.

5. A sequence of sets is defined recursively as follows:

- $S_0 = \{\emptyset\}$
- $S_{n+1} = S_n \cup \mathcal{P}(S_n)$ for all $n \in \mathbb{N}$

Prove by induction that for every $n \in \mathbb{N}$, every element of S_n is also a subset of S_n. That is, prove $\forall n \in \mathbb{N}, (\forall x \in S_n \Rightarrow x \subseteq S_n)$.

6.3 Advanced Induction

The principle of mathematical induction discussed in section 6.1 is sometimes called weak induction. Its induction step involves assuming $\mathbf{P}(k)$ to prove $\mathbf{P}(k+1)$. However, in some contexts, particularly with recursively defined sequences where a term depends on multiple predecessors, a stronger assumption is required. The principle of strong induction provides this more powerful hypothesis.

Strong Induction

The intuition behind strong induction is that by the time we consider the $(k+1)$-th case, we have already established the truth of all preceding cases, from the basis up to the k-th case. We are therefore justified in assuming all of them to prove the next one.

***Theorem* 6.3.1. Principle of Strong Induction**. Let $\mathbf{P}(n)$ be a formula. For any integer n_0, if

$$\mathbf{P}(n_0) \wedge \forall k \in \mathbb{Z}\left((k > n_0 \wedge \forall i \in \mathbb{Z}(n_0 \leq i < k \Rightarrow \mathbf{P}(i))) \Rightarrow \mathbf{P}(k)\right)$$

then

$$\forall n \in \mathbb{Z}(n \geq n_0 \Rightarrow \mathbf{P}(n)).$$

Proof. Assume the premise of the theorem holds. Let $\mathbf{Q}(k)$ be the formula $\forall i \in \mathbb{Z}(n_0 \leq i \leq k \Rightarrow \mathbf{P}(i))$. We will use the principle of weak induction to prove that $\mathbf{Q}(k)$ holds for all $k \geq n_0$.

- **Basis Case:** We show $\mathbf{Q}(n_0)$. This is the statement $\mathbf{P}(n_0)$, which is true by the first part of our assumption.
- **Induction Step:** Assume $\mathbf{Q}(k)$ for some $k \geq n_0$. This means $\mathbf{P}(i)$ is true for all integers i such that $n_0 \leq i \leq k$. By the second part of our assumption, this conjunction implies that $\mathbf{P}(k+1)$ is true. Therefore, $\mathbf{P}(i)$ is true for all $n_0 \leq i \leq k$ and for $i = k+1$. This is precisely the statement $\mathbf{Q}(k+1)$.

By the principle of weak induction, $\mathbf{Q}(k)$ is true for all $k \geq n_0$. Since $\mathbf{Q}(k)$ implies $\mathbf{P}(k)$, it follows that $\mathbf{P}(n)$ is true for all integers $n \geq n_0$. ■

The Well-Ordering Principle

The principles of weak and strong induction are not the only ways to formalise proofs concerning the natural numbers. An equivalent, and often more direct, axiom is the Well-Ordering Principle, which states that the standard order on $\mathbb{N}$ has a crucial property.

***Theorem* 6.3.2. The Well-Ordering Principle**. Every non-empty subset of the natural numbers $\mathbb{N}$ contains a least element.

Proof. Let S be a non-empty subset of $\mathbb{N}$. We prove by contradiction that S must have a least element. Assume S does not have a least element. Let $\mathbf{P}(n)$ be the statement "$n \notin S$". We use induction to prove that $\mathbf{P}(n)$ is true for all $n \in \mathbb{N}$.

- **Basis Case**: Consider $n = 0$. If $0 \in S$, then since 0 is the smallest natural number, it would be the least element of S. This contradicts our assumption. Therefore, $0 \notin S$, so $\mathbf{P}(0)$ is true.
- **Induction Step**: Assume $\mathbf{P}(k)$ is true for all $k \leq n$. This means no natural number from 0 to n is in S. Consider the number $n+1$. If $n+1 \in S$, then since all numbers smaller than it are not in S, $n+1$ would be the least element of S. This is a contradiction. Therefore, $n+1 \notin S$, so $\mathbf{P}(n+1)$ is true.

By the principle of strong induction, $\mathbf{P}(n)$ is true for all $n \in \mathbb{N}$. This implies that S contains no elements, i.e., $S = \emptyset$. This contradicts our initial premise that S is a non-empty set. Therefore, the assumption that S has no least element must be false. ■

The Well-Ordering Principle provides an elegant method for proofs. A common application is to prove the existence of a prime factorisation for every natural number greater than one.

***Theorem* 6.3.3. The Division Algorithm**. For each $m \in \mathbb{N} \setminus \{0\}$ and $n \in \mathbb{N}$, there exist unique numbers $k, l \in \mathbb{N}$ such that

$$n = km + l \quad \text{and} \quad l < m$$

The number k is the quotient and l is the remainder.

Proof. We first prove existence by induction on n. Fix $m \in \mathbb{N} \setminus \{0\}$. Let I be the set of natural numbers for which the statement holds:

$$I = \{n \in \mathbb{N} \mid \exists k, l \in \mathbb{N} \text{ such that } n = km + l \wedge l < m\}$$

- **Base Case**: For $n = 0$, we may choose $k = 0$ and $l = 0$. Since $m \neq 0$, we have $0 < m$, so $0 = 0m + 0$ is a valid representation. Thus, $0 \in I$.
- **Inductive Step**: Assume $n \in I$. Then there exist $k, l \in \mathbb{N}$ such that $n = km + l$ and $l < m$. Consider $n + 1$:

  $$n + 1 = km + (l + 1)$$

  If $l + 1 < m$, then we have found the required representation for $n + 1$, with quotient k and remainder $l + 1$. Thus $n + 1 \in I$. If $l + 1 = m$, then $n + 1 = km + m = (k + 1)m$. We can write this as $n + 1 = (k + 1)m + 0$. Since $0 < m$, this is a valid representation with quotient $k + 1$ and remainder 0. Thus $n + 1 \in I$.

Since I is an inductive subset of $\mathbb{N}$, we have $I = \mathbb{N}$, proving existence.

To prove uniqueness, suppose $n = k_1 m + l_1$ and $n = k_2 m + l_2$ where $l_1 < m$ and $l_2 < m$. Assume, without loss of generality, that $l_1 \leq l_2$. Then

$$k_1 m + l_1 = k_2 m + l_2 \Rightarrow (k_1 - k_2)m = l_2 - l_1$$

Since $l_1 \leq l_2 < m$, we have $0 \leq l_2 - l_1 < m$. The equation implies that $l_2 - l_1$ is a multiple of m. As the only multiple of m that is smaller than m is 0, it must be that $l_2 - l_1 = 0$, so $l_1 = l_2$. This gives $(k_1 - k_2)m = 0$. Since $m \neq 0$, the cancellation law for multiplication implies $k_1 - k_2 = 0$, so $k_1 = k_2$. The representation is unique. ■

Definition 6.3.1. ***Prime Number***. A natural number $p \in \mathbb{N}$ is prime if $p \geq 2$ and its only divisors are 1 and p.

***Theorem* 6.3.4. The Fundamental Theorem of Arithmetic**. Every natural number $n > 1$ can be written as a product of prime numbers. This factorisation is unique up to the order of the factors.

Proof. **Existence**: Let S be the set of all natural numbers $n > 1$ that cannot be written as a product of primes. Assume for contradiction that S is non-empty. By the Well-Ordering Principle, S must have a least element, let us call it m. Since $m \in S$, m cannot be prime (otherwise it would be a product of one prime, itself). Thus, m must be composite, meaning there exist natural numbers a, b such that $m = ab$ where $1 < a \leq b < m$. Because a and b are smaller than m, they cannot be in S. Therefore, both a and b can be written as a product of primes. But if this is so, their product $m = ab$ can also be written as a product of primes (by concatenating their prime factorisations). This contradicts the fact that $m \in S$. The contradiction implies our assumption was false; the set S must be empty. Thus, every natural number greater than 1 has a prime factorisation.

Uniqueness: Let T be the set of natural numbers $n > 1$ that have more than one distinct prime factorisation. Assume for contradiction that T is non-empty, and let p be its least element. So,

$$p = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$$

where $\{p_i\}$ and $\{q_j\}$ are two different sets of prime factors. We can assume $p_1 \leq p_2 \leq \cdots \leq p_k$ and $q_1 \leq q_2 \leq \cdots \leq q_l$. Since the factorisations are different, $p_1 \neq q_1$. Assume without loss of generality that $p_1 < q_1$. Consider the number $p' = p - p_1 q_2 \cdots q_l$. We can factor out p_1:

$$p' = p_1(p_2 \cdots p_k - q_2 \cdots q_l)$$

So p_1 divides p'. We can also express p' using the second factorisation of p:

$$p' = (q_1 - p_1)q_2 \cdots q_l$$

Since $p_1 < q_1$, $q_1 - p_1 > 0$, so $p' > 0$. Also, $p' < p$, because $p_1 q_2 \cdots q_l < q_1 q_2 \cdots q_l = p$. Since $p' < p$, p' must have a unique prime factorisation. From the second expression for p', we see that its prime factors must be the prime factors of $(q_1 - p_1)$ together with the primes $q_2, \ldots, q_l$. However, we know that p_1 divides p'. Since p_1 is prime, it must be equal to one of the prime factors in the factorisation of p'.

- p_1 cannot be equal to any of $q_2, \ldots, q_l$, because we assumed the two original factorisations of p were distinct and sorted, so $p_1 < q_1 \leq q_j$ for all j.
- Therefore, p_1 must divide $q_1 - p_1$. If p_1 divides $q_1 - p_1$, then it must also divide $(q_1 - p_1) + p_1 = q_1$. But since q_1 is prime, its only divisors are 1 and q_1. Since $p_1 > 1$, this implies $p_1 = q_1$.

This contradicts our assumption that $p_1 < q_1$. The set T must be empty, proving uniqueness. ■

6.4 Exercises

1. The Fibonacci sequence is defined by $F_0 = 0$, $F_1 = 1$, and the recursive relation $F_n = F_{n-1} + F_{n-2}$ for all integers $n \geq 2$.

(a) Explain why the principle of weak induction is insufficient to prove a property of F_n that relies on its recursive definition.

(b) Let $\phi = \frac{1+\sqrt{5}}{2}$ (the golden ratio). Prove by strong induction that $F_n > \phi^{n-2}$ for all integers $n \geq 3$.

2. A prime number is an integer greater than 1 that has no positive divisors other than 1 and itself. Use the principle of strong induction to prove that every integer $n \geq 2$ can be written as a product of one or more prime numbers.

Remark. For the induction step, consider an integer k. There are two cases: either k is prime, or k is composite. If it is composite, it can be factored as $k = ab$, where a and b are smaller than k.

3. Consider a chocolate bar made of n squares arranged in a single line. A "break" consists of choosing one piece of chocolate and breaking it into two smaller pieces along one of the dividing lines. Use strong induction to prove that for any integer $n \geq 1$, it takes exactly $n-1$ breaks to separate the bar into n individual squares.

4. A sequence is defined by $a_1 = 1$, $a_2 = 3$, and $a_n = 2a_{n-1} - a_{n-2}$ for all $n \geq 3$. Use strong induction to prove that $a_n = 2n - 1$ for all integers $n \geq 1$.

5. Use strong induction to prove that any integer amount of postage $n \geq 8$ cents can be formed using only 3-cent and 5-cent stamps.

Remark. For the induction step, consider an amount k. You can form this amount if you could previously form the amount $k-3$. This logic does not work for all initial values of k, so you will need to establish more than one basis case.

6.5 The Real Numbers

The sets of natural numbers, integers, and rational numbers have been constructed using set-theoretic principles. However, the set of rational numbers, $\mathbb{Q}$, is fundamentally incomplete. There are "gaps" in the rational number line. For instance, the set of rational numbers $S = \{x \in \mathbb{Q} \mid x^2 < 2\}$ is non-empty and bounded above in $\mathbb{Q}$ (by 2, for example), but it has no least upper bound within $\mathbb{Q}$. The object that "should" be the supremum, $\sqrt{2}$, is not rational.

Example 6.5.1. Proof that $\mathbb{Q}$ is not complete. Let us formally prove that the set $S = \{x \in \mathbb{Q} \mid x^2 < 2\}$ has no supremum in $\mathbb{Q}$. Assume for contradiction that a least upper bound $c \in \mathbb{Q}$ for S exists. First, note that $1 \in S$ (since $1^2 < 2$). Since c is an upper bound for S, it must be greater than or equal to every element in S, so $c \geq 1$. By the law of trichotomy, exactly one of $c^2 < 2$, $c^2 > 2$, or $c^2 = 2$ must be true.

- We know that no rational number squares to 2, so $c^2 = 2$ is impossible.
- Suppose $c^2 < 2$. We will construct a rational number $c + \delta$ with $\delta > 0$ such that $(c+\delta)^2 < 2$. This would mean $c + \delta \in S$, contradicting the assumption that c is an upper bound for S. Consider the number $\delta = \frac{2-c^2}{c+2}$. Since $c \geq 1$, the denominator $c + 2$ is positive, and since $c^2 < 2$, the numerator is positive. Thus, δ is a positive rational number. Let's examine the square of $c + \delta$:

$$(c+\delta)^2 = c^2 + 2c\delta + \delta^2 = c^2 + 2c\left(\frac{2-c^2}{c+2}\right) + \left(\frac{2-c^2}{c+2}\right)^2$$

To avoid complex algebra, consider a simpler approach. We want to find a small rational $h > 0$ such that $(c+h)^2 = c^2 + 2ch + h^2 < 2$. This requires $h(2c+h) < 2 - c^2$. If we choose $h < 1$, then $2c + h < 2c + 1$, so we need $h(2c+1) < 2 - c^2$, or $h < \frac{2-c^2}{2c+1}$. Since $2 - c^2 > 0$ and $2c + 1 > 3$, such a positive rational h exists. This contradicts c being an upper bound.

Since all three possibilities lead to a contradiction, our initial assumption that S has a least upper bound in $\mathbb{Q}$ must be false.

The goal of this chapter is to construct a set that formalises the intuitive idea of the real number line, a set that is a complete ordered field. The construction, due to Richard Dedekind, builds the real numbers from the rational numbers by characterising each real number by the set of all rational numbers less than it.

Initial Segments and Dedekind Cuts

The construction of the real numbers relies on specific subsets of partially ordered sets.

Definition 6.5.1. ***Initial Segment***. Let $(A, \preceq)$ be a poset. A subset $S \subseteq A$ is an initial segment of A if it is downward closed, meaning that for all $x, y \in A$, if $y \in S$ and $x \preceq y$, then $x \in S$. An initial segment S is proper if $S \neq A$.

For any element a in a poset $(A, \preceq)$, the set of all elements strictly less than a forms a canonical initial segment.

Definition 6.5.2. . Let $(A, \preceq)$ be a poset and let $a \in A$. Define $\preceq (A, a) = \{x \in A \mid x \prec a\}$.

While any set of this form is an initial segment, not every initial segment must be of this form. However, for well-ordered sets, this is the only possibility.

***Lemma* 6.5.1.** If $(A, \preceq)$ is a well-ordered set and S is a proper initial segment of A, then there exists a unique $a \in A$ such that $S = \preceq (A, a)$.

Proof. Let $(A, \preceq)$ be well-ordered and let S be a proper initial segment. Since S is proper, the set difference $A \setminus S$ is non-empty. As A is well-ordered, $A \setminus S$ must contain a least element; let this element be a. We will show $S = \preceq (A, a)$.

- Let $x \in S$. Since a is the least element of $A \setminus S$, x cannot be in $A \setminus S$, so $x \notin A \setminus S$. If we had $a \preceq x$, then since S is downward closed, we would have $a \in S$, a contradiction. Therefore, by trichotomy, it must be that $x \prec a$. This implies $x \in \preceq (A, a)$, so $S \subseteq \preceq (A, a)$.
- Let $x \in \preceq (A, a)$, which means $x \prec a$. If x were an element of $A \setminus S$, this would contradict the fact that a is the least element of $A \setminus S$. Thus, x must be an element of S. This implies $\preceq (A, a) \subseteq S$.

We have shown $S = \preceq (A, a)$. To prove uniqueness, suppose there exists another element $a' \in A$ such that $S = \preceq (A, a')$. If $a \neq a'$, then either $a \prec a'$ or $a' \prec a$. If $a \prec a'$, then $a \in \preceq (A, a') = S$, which is impossible as $a \in A \setminus S$. A symmetric argument holds if $a' \prec a$. Thus, $a = a'$. ■

We now define a real number as a special type of initial segment of the rational numbers.

Definition 6.5.3. ***Dedekind Cut***. A set $x \subseteq \mathbb{Q}$ is a Dedekind cut (or a *real number*) if it satisfies three properties:

1. x is a non-empty, proper subset of $\mathbb{Q}$ ($x \neq \emptyset$ and $x \neq \mathbb{Q}$).
2. x is an initial segment of $(\mathbb{Q}, \leq)$.
3. x does not have a greatest element.

The set of all Dedekind cuts is denoted by $\mathbb{R}$.

The rational numbers can be embedded into this new set. For any $q \in \mathbb{Q}$, the set $\mathbf{q} = \{r \in \mathbb{Q} \mid r < q\}$ is a Dedekind cut. The function $\phi : \mathbb{Q} \to \mathbb{R}$ defined by $\phi(q) = \mathbf{q}$ is an order embedding. The elements of $\mathbb{R} \setminus \text{ran}(\phi)$ are the irrational numbers. For instance, the number π corresponds to the Dedekind cut $\{q \in \mathbb{Q} \mid q < \pi\}$.

Order and Completeness

To show that $\mathbb{R}$ is a suitable model for the real numbers, we must define an order on it and prove that it is complete.

Definition 6.5.4. . Let $x, y \in \mathbb{R}$. We define $x \leq y$ if and only if $x \subseteq y$. The strict order $x < y$ means $x \subset y$.

This definition naturally extends the order from $\mathbb{Q}$ to $\mathbb{R}$. For $q_1, q_2 \in \mathbb{Q}$, $q_1 < q_2 \Leftrightarrow \mathbf{q_1} \subset \mathbf{q_2}$.

***Theorem* 6.5.1.** The relation $\leq$ is a linear order on $\mathbb{R}$, but it is not a well-order.

The defining property of the real numbers is that they do not suffer from the gaps present in the rationals. This is formalised by the completeness property.

***Theorem* 6.5.2. Completeness of the Real Numbers**. Every non-empty subset of $\mathbb{R}$ that has an upper bound has a least upper bound in $\mathbb{R}$.

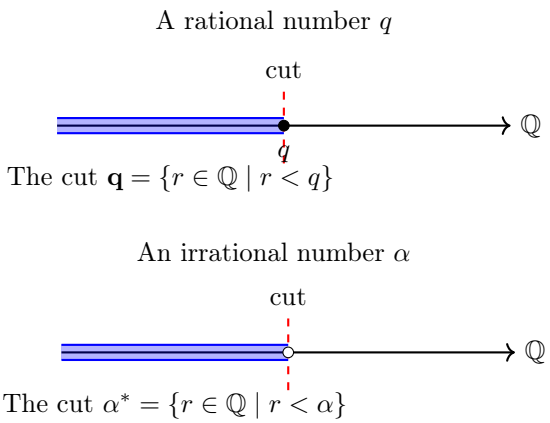


Figure 6.1: A Dedekind cut partitions $\mathbb{Q}$ into two sets. If the cut corresponds to a rational number q, then q is the least element of $\mathbb{Q} \setminus \mathbf{q}$. If it corresponds to an irrational number α, then $\mathbb{Q} \setminus \alpha^*$ has no least element, representing a "gap" in the rationals.

Proof. Let $\mathcal{F}$ be a non-empty subset of $\mathbb{R}$ that is bounded above by some $m \in \mathbb{R}$. The supremum of $\mathcal{F}$ with respect to the subset relation $\subseteq$ is its union, $\bigcup \mathcal{F}$. We must show that this union is itself a Dedekind cut, i.e., an element of $\mathbb{R}$. Let $U = \bigcup \mathcal{F}$.

1. Since $\mathcal{F}$ is non-empty, it contains at least one cut $x \in \mathcal{F}$. Since x is a Dedekind cut, $x \neq \emptyset$, which implies $U \neq \emptyset$. Every $x \in \mathcal{F}$ is a subset of the upper bound m, so $U \subseteq m$. Since m is a proper subset of $\mathbb{Q}$, so is U. Thus, U is a non-empty, proper subset of $\mathbb{Q}$.
2. Let $u \in U$ and let $q \in \mathbb{Q}$ with $q < u$. Since $u \in U$, there must be some cut $x \in \mathcal{F}$ such that $u \in x$. As x is a cut, it is downward closed, so $q \in x$. This implies $q \in U$. Therefore, U is an initial segment of $(\mathbb{Q}, \leq)$.
3. Let $u \in U$. Then u belongs to some cut $x \in \mathcal{F}$. Since x has no greatest element, there exists an element $r \in x$ such that $u < r$. Since $r \in x$, we also have $r \in U$. This shows that no element $u \in U$ can be its greatest element.

We have shown that $U = \bigcup \mathcal{F}$ is a Dedekind cut, and thus an element of $\mathbb{R}$. As the union of a collection of sets ordered by inclusion, it is their least upper bound. Therefore, $\mathbb{R}$ is complete. ■

Definition 6.5.5. ***Ordered and Complete Ordered Field***. An *ordered field* is a field $(K, +, \cdot)$ together with a total order $\leq$ on K such that for all $a, b, c \in K$:

1. If $a \leq b$, then $a + c \leq b + c$.
2. If $0 \leq a$ and $0 \leq b$, then $0 \leq ab$.

An ordered field is *complete* if every non-empty subset $S \subseteq K$ that is bounded above has a least upper bound in K. A *complete ordered field* is an ordered field that is complete in this sense.

6.6 Arithmetic on the Real Numbers

Having constructed the set $\mathbb{R}$ of Dedekind cuts and established its order properties, we now define the arithmetic operations that will give $\mathbb{R}$ the structure of a complete ordered field.

- $\mathbf{0} := \{q \in \mathbb{Q} \mid q < 0\}$
- $\mathbf{1} := \{q \in \mathbb{Q} \mid q < 1\}$

Henceforth, we will use the symbols 0 and 1 to denote these specific cuts.

Definition 6.6.1. ***Arithmetic Operations on*** $\mathbb{R}$. Let $x, y \in \mathbb{R}$.

(i) The **sum** of x and y is defined as:

$$x + y := \{p + q \mid p \in x \wedge q \in y\}$$

(ii) The **product** of x and y is defined piecewise:

$$x \cdot y := \begin{cases} \{pq \mid p \in x, q \in y, p \geq 0, q \geq 0\} \cup \mathbf{0} & \text{if } x \geq 0 \wedge y \geq 0 \\ \{pq \mid p \in x, q \in y, p < 0, q < 0\} \cup \mathbf{0} & \text{if } x < 0 \wedge y < 0 \\ \{pq \mid p \in x, q \in y, q \geq 0\} & \text{if } x < 0 \wedge y \geq 0 \\ \{pq \mid p \in x, q \in y, p \geq 0\} & \text{if } x \geq 0 \wedge y < 0 \end{cases}$$

It can be verified that the sets resulting from these operations are themselves Dedekind cuts. For addition, this is left as an exercise. For multiplication, the argument is analogous to that for showing a set is a Dedekind cut, involving proofs of non-emptiness, properness, downward closure, and the absence of a greatest element.

Field Properties of the Real Numbers

The operations of addition and multiplication on $\mathbb{R}$ satisfy the axioms for a field.

***Theorem* 6.6.1.** Addition and multiplication of real numbers are associative and commutative.

Proof. Let $x, y, z \in \mathbb{R}$. We prove that addition is associative. The remaining properties are established by similar arguments.

$$\begin{aligned} x + (y + z) &= x + \{q + r \mid q \in y \wedge r \in z\} \\ &= \{p + s \mid p \in x \wedge s \in \{q + r \mid q \in y \wedge r \in z\}\} \\ &= \{p + (q + r) \mid p \in x \wedge q \in y \wedge r \in z\} \\ &= \{(p + q) + r \mid p \in x \wedge q \in y \wedge r \in z\} && \text{by associativity of } + \text{ in } \mathbb{Q} \\ &= \{s + r \mid s \in \{p + q \mid p \in x \wedge q \in y\} \wedge r \in z\} \\ &= (x + y) + z \end{aligned}$$

■

The cuts $\mathbf{0}$ and $\mathbf{1}$ function as the identity elements.

***Theorem* 6.6.2.** The set $\mathbb{R}$ has additive and multiplicative identities.

Proof. Let $x \in \mathbb{R}$. We first show $x + 0 = x$. Let $p \in x$. Since x has no greatest element, there exists $p' \in x$ such that $p < p'$. We can write $p = p' + (p - p')$. Since $p - p' < 0$, $p - p' \in 0$. Thus, $p \in x + 0$, which shows $x \subseteq x + 0$. Conversely, let $p + q \in x + 0$, where $p \in x$ and $q \in 0$. Since $q < 0$, we have $p + q < p$. As x is downward closed, $p + q \in x$. This shows $x + 0 \subseteq x$. Therefore, $x + 0 = x$.

Next, we show $x \cdot 1 = x$. We consider the case where $0 \leq x$. By definition, $x \cdot 1 = \{pq \mid p \in x, q \in 1, p \geq 0, q \geq 0\} \cup 0$. Let $s \in x \cdot 1$. If $s \leq 0$, then as $0 \leq x$, $0 \subseteq x$, so $s \in x$. If $s > 0$, then $s = pq$ for some $p \in x, p \geq 0$ and $q \in 1, q \geq 0$. Since $q < 1$, $pq < p$, so $s < p$. As $p \in x$ and x is downward closed, $s \in x$. Thus $x \cdot 1 \subseteq x$. Conversely, take $p \in x$. If $p < 0$, then $p \in 0 \subseteq x \cdot 1$. If $p \geq 0$, since x has no greatest element, there exists $p' \in x$ with $p < p'$. Then $p/p' < 1$, so $p/p' \in 1$. We can write $p = p' \cdot (p/p')$. Since $p' \in x$ and $p/p' \in 1$, $p \in x \cdot 1$. Thus $x \subseteq x \cdot 1$. The case $x < 0$ is analogous. ■

The remaining field axioms also hold for the set of Dedekind cuts.

***Theorem* 6.6.3.** For the set $\mathbb{R}$ with the defined operations:

(i) Every element has an additive inverse.
(ii) Every non-zero element has a multiplicative inverse.
(iii) The distributive law of multiplication over addition holds.

Proof. The proofs are constructive but lengthy and are omitted here (and way above these notes pay grade). But they follow from the properties of the rational numbers. ■

***Corollary* 6.6.1.** . With the operations and order defined above, $(\mathbb{R}, +, \cdot, \leq)$ is a complete ordered field.

Proof. We have shown that $(\mathbb{R}, +, \cdot, \leq)$ is an ordered field, and by the Completeness of the Real Numbers theorem above, every non-empty subset of $\mathbb{R}$ that is bounded above has a least upper bound in $\mathbb{R}$. Hence $(\mathbb{R}, +, \cdot, \leq)$ is a complete ordered field. ■

The Extended Real Number Line

The completeness property ensures that every bounded subset of $\mathbb{R}$ has a supremum. To formalise the concept of "unboundedness" and simplify the treatment of limits, we extend the real numbers by adjoining two elements at infinity.

Definition 6.6.2. ***Extended Real Numbers***. The extended real number line, denoted $\overline{\mathbb{R}}$, is the set $\mathbb{R} \cup \{-\infty, \infty\}$. The order $\leq$ on $\mathbb{R}$ is extended such that for all $x \in \mathbb{R}$:

$$-\infty < x < \infty$$

This makes $\overline{\mathbb{R}}$ a totally ordered set with a minimum $-\infty$ and a maximum ∞.

We partially extend the arithmetic operations to $\overline{\mathbb{R}}$. These definitions are chosen to be consistent with limits in calculus.

Definition 6.6.3. ***Arithmetic in*** $\overline{\mathbb{R}}$. Let $x \in \mathbb{R}$.

(i) **Addition:** $x + \infty = \infty$ and $x + (-\infty) = -\infty$. Furthermore, $\infty + \infty = \infty$ and $-\infty + (-\infty) = -\infty$.
(ii) **Multiplication:** If $x > 0$, then $x \cdot \infty = \infty$ and $x \cdot (-\infty) = -\infty$. If $x < 0$, then $x \cdot \infty = -\infty$ and $x \cdot (-\infty) = \infty$.
(iii) **Inversion:** $\frac{x}{\infty} = \frac{x}{-\infty} = 0$.

Note: The expressions $\infty - \infty$, $0 \cdot \infty$, $\frac{\infty}{\infty}$, and $\frac{x}{0}$ remain undefined. Consequently, $\overline{\mathbb{R}}$ is not a field.

The primary utility of $\overline{\mathbb{R}}$ lies in the existence of suprema and infima for *all* subsets.

***Theorem* 6.6.4.** Every subset $A \subseteq \mathbb{R}$ has a supremum and an infimum in $\overline{\mathbb{R}}$.

- If A is not bounded above in $\mathbb{R}$, $\sup(A) = \infty$.
- If A is not bounded below in $\mathbb{R}$, $\inf(A) = -\infty$.
- By convention, $\sup(\emptyset) = -\infty$ and $\inf(\emptyset) = \infty$.

The Complex Numbers

The final number system we introduce is the set of complex numbers, which extends the real numbers to provide solutions for equations such as $x^2 + 1 = 0$.

Definition 6.6.4. ***Complex Numbers***. The set of complex numbers, denoted $\mathbb{C}$, is the Cartesian product $\mathbb{R} \times \mathbb{R}$. An element $(a, b) \in \mathbb{C}$ is typically written as $a + bi$.

The function $\phi : \mathbb{R} \to \mathbb{C}$ defined by $\phi(x) = (x, 0)$ is an embedding that allows us to consider $\mathbb{R}$ as a subset of $\mathbb{C}$.

Definition 6.6.5. ***Arithmetic on*** $\mathbb{C}$. Let $a + bi$ and $c + di$ be complex numbers.

- **Addition**: $(a + bi) + (c + di) := (a + c) + (b + d)i$

- **Multiplication**: $(a+bi)\cdot(c+di) := (ac-bd)+(ad+bc)i$

With these operations, $\mathbb{C}$ forms a field.

***Theorem* 6.6.5.** The set of complex numbers $\mathbb{C}$ with the operations of addition and multiplication satisfies the field axioms. In particular:

(i) If we define $i := 0+1i$, then $i^2 = -1+0i$.
(ii) Addition and multiplication are associative and commutative.
(iii) The additive identity is $0+0i$ and the multiplicative identity is $1+0i$.
(iv) Every element has an additive inverse, and every non-zero element has a multiplicative inverse.

Proof. These properties are verified by direct computation using the definitions. For instance, $i^2 = (0+1i)\cdot(0+1i) = (0\cdot 0 - 1\cdot 1) + (0\cdot 1 + 1\cdot 0)i = -1+0i$. The remaining proofs are left as an exercise. ■

6.7 Exercises

1. Another canonical example of an irrational Dedekind cut is the one corresponding to $\sqrt{3}$. Let the set $x \subseteq \mathbb{Q}$ be defined as:

$$x = \{q \in \mathbb{Q} \mid q < 0 \vee q^2 < 3\}$$

Prove from the definition that x is a Dedekind cut. Show that for any prime p that $\sqrt{p}$ is an irrational Dedekind cut.

2. The text proves that for a bounded-above set of cuts $\mathcal{F}$, its union $\bigcup \mathcal{F}$ is a Dedekind cut. A similar argument is needed for the arithmetic operations. Let $x, y \in \mathbb{R}$ be two Dedekind cuts. Prove that their sum, defined as

$$x + y := \{p + q \mid p \in x \wedge q \in y\}$$

is also a Dedekind cut.

3. Let $x, y, z \in \mathbb{R}$. The text proves that addition is associative. Prove that the order on $\mathbb{R}$ is transitive. That is, using the definition $x \leq y \Leftrightarrow x \subseteq y$, prove that if $x \leq y$ and $y \leq z$, then $x \leq z$.

4. For any real number $x \in \mathbb{R}$, its additive inverse can be constructed as the set

$$-x := \{q \in \mathbb{Q} \mid \exists r \in \mathbb{Q}(r > 0 \wedge -q - r \notin x)\}$$

Let $x = \{q \in \mathbb{Q} \mid q < 2\}$. Describe the set $-x$ and show that it corresponds to the real number -2.

5. Suppose we relax the definition of a Dedekind cut by removing the third condition (that it has no greatest element). Let us call such a set a *generalised cut*. Prove that if a generalised cut x has a greatest element, then that greatest element must be a rational number.

6. Let $a+bi$ and $c+di$ be two complex numbers. Verify by direct computation from the definitions of the operations on $\mathbb{C}$ that:

(a) Addition is commutative: $(a+bi)+(c+di) = (c+di)+(a+bi)$.
(b) Multiplication is distributive over addition.

7. ⋆ For any non-zero complex number $z = a+bi$, its multiplicative inverse is given by

$$z^{-1} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

Using the definition of multiplication on $\mathbb{C}$, prove that $z \cdot z^{-1} = 1+0i$.

8. Let $(A, \preceq)$ be a poset. Assume that S_1 and S_2 are initial segments of A. Prove that $S_1 \cap S_2$ is an initial segment. Then, either prove that $S_1 \cup S_2$ is always an initial segment or provide a counterexample.

9. Prove from the definition that the sets corresponding to the additive and multiplicative identities, $\mathbf{0} = \{q \in \mathbb{Q} \mid q < 0\}$ and $\mathbf{1} = \{q \in \mathbb{Q} \mid q < 1\}$, are Dedekind cuts.

10. Let $x, y \in \mathbb{R}$ be non-negative real numbers ($x \geq \mathbf{0}$ and $y \geq \mathbf{0}$). Using the definition of multiplication for non-negative cuts, prove that if $x \cdot y = \mathbf{0}$, then $x = \mathbf{0}$ or $y = \mathbf{0}$.

11. Prove that the real numbers are dense. That is, for any two real numbers $x, y \in \mathbb{R}$ with $x < y$, there exists a real number z such that $x < z < y$.

12. Prove that the rational numbers are dense in the real numbers. That is, for any two real numbers $x, y \in \mathbb{R}$ with $x < y$, there exists a rational number q such that $x < \mathbf{q} < y$, where $\mathbf{q}$ is the cut corresponding to q.

13. The function $\phi : \mathbb{Q} \to \mathbb{R}$ defined by $\phi(q) = \{r \in \mathbb{Q} \mid r < q\}$ embeds the rationals into the reals. Prove that ϕ is an order embedding; that is, prove it is injective and that for all $p, q \in \mathbb{Q}$, $p \leq q \Leftrightarrow \phi(p) \subseteq \phi(q)$.

14. The **absolute value** of a real number $x \in \mathbb{R}$ is defined as $|x| = \max\{x, -x\}$, where the maximum is taken with respect to the order $\leq$.

(a) Prove from this definition that for any $x \in \mathbb{R}$, $x \leq |x|$ and $-x \leq |x|$.

(b) Let $a \in \mathbb{R}$ with $a > \mathbf{0}$. Prove that $|x| < a$ if and only if $-a < x < a$.

6.8 Abstract Algebraic Structures

In the preceding section, we constructed the number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ to satisfy specific arithmetic requirements. We observed that these systems share common properties, such as associativity and the existence of identities. In this section, we abstract these properties to define general algebraic structures. This abstraction allows us to derive theorems that apply to any system satisfying the axioms, not just the specific number systems we have built.

Groups

The most fundamental algebraic structure captures the essence of symmetry and reversible operations.

Definition 6.8.1. ***Group***. A group is a pair $(G, \star)$ consisting of a non-empty set G and a binary operation $\star : G \times G \to G$ satisfying three axioms:

(G1) **Associativity:** For all $g, h, k \in G$, $(g \star h) \star k = g \star (h \star k)$.
(G2) **Identity Element:** There exists an element $e \in G$ such that for all $g \in G$, $g \star e = e \star g = g$.
(G3) **Inverse Element:** For each $g \in G$, there exists an element $g^{-1} \in G$ such that $g \star g^{-1} = g^{-1} \star g = e$.

If the operation $\star$ is commutative (i.e., $g \star h = h \star g$ for all $g, h \in G$), the group is called Abelian.

Remark. Consider the set of integers under addition, $(\mathbb{Z}, +)$. It forms an Abelian group with identity 0 and inverse $-n$. Similarly, the non-zero rationals under multiplication, $(\mathbb{Q} \setminus \{0\}, \cdot)$, form an Abelian group with identity 1 and inverse $1/q$. In contrast, the integers under multiplication, $(\mathbb{Z}, \cdot)$, do not form a group because most elements (like 2) have no multiplicative inverse within the set of integers.

***Theorem* 6.8.1. Uniqueness Properties**. Let $(G, \star)$ be a group.

(i) The identity element e is unique.
(ii) For every $g \in G$, the inverse g^{-1} is unique.
(iii) **Cancellation Laws:** For any $a, b, x \in G$, if $a \star x = b \star x$, then $a = b$. Similarly, if $x \star a = x \star b$, then $a = b$.

Proof. (i) Suppose e and e' are identity elements. Then $e = e \star e' = e'$. (ii) Suppose h and k are inverses of g. Then $h = h \star e = h \star (g \star k) = (h \star g) \star k = e \star k = k$. (iii) Multiply both sides by x^{-1} from the right (or left). ■

Examples of Groups

- **Trivial Group:** Let $G = \{e\}$. With the operation $e \star e = e$, this is the smallest possible group.
- **Permutation Groups:** Let X be a non-empty set. The set S_X of all bijections from X to itself forms a group under function composition, denoted $(S_X, \circ)$. The identity is the identity function id_X, and the inverse is the inverse function f^{-1}. If X has n elements, this group is denoted S_n and is called the symmetric group of degree n. Note that for $n \geq 3$, S_n is non-Abelian.
- **Direct Products:** If $(G, \star)$ and $(H, \diamond)$ are groups, their direct product $G \times H$ is a group with the operation defined component-wise: $(g_1, h_1) \cdot (g_2, h_2) = (g_1 \star g_2, h_1 \diamond h_2)$.

Subgroups and Cosets

Just as $\mathbb{Z}$ is a subset of $\mathbb{Q}$ and both are groups under addition, groups often contain smaller groups within them.

Definition 6.8.2. ***Subgroup***. A subset $H \subseteq G$ is a subgroup of $(G, \star)$ if H forms a group under the restricted operation $\star$.

***Theorem* 6.8.2. Subgroup Test**. A non-empty subset $H \subseteq G$ is a subgroup if and only if it is closed under the group operation (i.e., $x, y \in H \Rightarrow x \star y \in H$) and closed under inverses (i.e., $x \in H \Rightarrow x^{-1} \in H$).

Proof. ($\Rightarrow$) If H is a subgroup, it is a group by definition. Thus, the operation $\star$ must be a binary operation on H (closure) and every element must have an inverse in H.

($\Leftarrow$) Assume H is non-empty, closed under $\star$, and closed under inverses. We check the group axioms:

- **Associativity:** Since elements of H are also elements of G, the operation is associative for them.
- **Identity:** Since H is non-empty, let $h \in H$. By hypothesis, $h^{-1} \in H$. By closure, $h \star h^{-1} \in H$. Since $h \star h^{-1} = e$, we have $e \in H$.
- **Inverses:** This is given by hypothesis.

Thus, H satisfies all group axioms and is a subgroup. ■

Given a subgroup N of G and an element $g \in G$, we can define the coset of g with respect to N.

$$g \star N := \{g \star n \mid n \in N\}$$

This concept generalises the arithmetic modulo n. For instance, in $\mathbb{Z}$, if we take the subgroup of multiples of 3, $3\mathbb{Z}$, the coset $1 + 3\mathbb{Z}$ is the set $\{\dots, -2, 1, 4, 7, \dots\}$.

Definition 6.8.3. ***Quotient Group***. If N is a subgroup satisfying $g \star N = N \star g$ for all $g \in G$ (called a *normal subgroup*), the set of cosets $G/N = \{g \star N \mid g \in G\}$ forms a group under the induced operation:

$$(g \star N) \cdot (h \star N) = (g \star h) \star N$$

This group is called the quotient group of G modulo N.

The construction of $\mathbb{Z}$ from $\mathbb{N} \times \mathbb{N}$ and $\mathbb{Q}$ from $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ in previous chapters can be rigorously viewed through the lens of quotient structures.

Homomorphisms and Isomorphisms

To compare two groups, we examine functions that preserve the group structure.

Definition 6.8.4. ***Homomorphism***. Let $(G, \star)$ and $(H, \diamond)$ be groups. A function $\phi : G \to H$ is a homomorphism if for all $x, y \in G$:

$$\phi(x \star y) = \phi(x) \diamond \phi(y)$$

Remark. A homomorphism maps the identity of G to the identity of H and inverses to inverses: $\phi(e_G) = e_H$ and $\phi(g^{-1}) = \phi(g)^{-1}$.

The kernel of a homomorphism ϕ is the set of elements in G that map to the identity in H: $\ker(\phi) = \{g \in G \mid \phi(g) = e_H\}$. The kernel is always a normal subgroup of G.

Definition 6.8.5. ***Isomorphism***. A homomorphism that is bijective is called an isomorphism. If there exists an isomorphism between groups G and H, we say they are isomorphic, denoted $G \cong H$.

Isomorphic groups are structurally identical; they differ only in the labelling of their elements. For example, the group of symmetries of an equilateral triangle is isomorphic to S_3, the group of permutations of three elements.

Polynomial Rings

We previously defined a ring as a set equipped with two operations, addition and multiplication, satisfying properties (R1)-(R3) (Abelian group under addition, associative multiplication, and distributivity). We conclude by introducing a crucial ring in analysis and algebra.

Definition 6.8.6. ***Polynomial Ring***. Let K be a field (such as $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$). The polynomial ring in one variable over K, denoted $K[X]$, is the set of sequences $(a_0, a_1, a_2, \dots)$ of elements from K where $a_n = 0$ for almost all n (i.e., only finitely many coefficients are non-zero). We typically represent an element $P \in K[X]$ as a formal sum:

$$P(X) = \sum_{i=0}^{n} a_i X^i = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$$

Addition and multiplication are defined in the standard way:

- **Addition:** Sum corresponding coefficients.
- **Multiplication:** Convolution of coefficients (distributive expansion).

Properties of Polynomial Rings

To prove that $K[X]$ is an integral domain, we must show that the product of two non-zero polynomials is never zero. The primary tool for this analysis is the degree of a polynomial.

Definition 6.8.7. ***Degree of a Polynomial***. Let $P(X) = \sum_{i=0}^{n} a_i X^i$ be a non-zero polynomial in $K[X]$ with $a_n \neq 0$. The integer n is called the degree of P, denoted $\deg(P)$.

- The degree of a non-zero constant is 0.
- The zero polynomial, 0, does not have a standard integer degree (often defined as $-\infty$).

The coefficient a_n is called the leading coefficient.

***Lemma* 6.8.1.** *Degree of a Product.* Let $P, Q \in K[X]$ be non-zero polynomials. Then:

$$\deg(P \cdot Q) = \deg(P) + \deg(Q)$$

Proof. Let $n = \deg(P)$ and $m = \deg(Q)$. We can write:

$$P(X) = a_n X^n + \cdots + a_0 \quad \text{and} \quad Q(X) = b_m X^m + \cdots + b_0$$

where $a_n \neq 0$ and $b_m \neq 0$. The product $P(X) \cdot Q(X)$ is defined by the convolution of coefficients. Consider the coefficient of the term X^{n+m}. The only way to form X^{n+m} by multiplying a term $a_i X^i$ from P and $b_j X^j$ from Q is if $i + j = n + m$. Since $i \leq n$ and $j \leq m$, the only solution is $i = n$ and $j = m$. Thus, the coefficient of X^{n+m} in the product is $a_n \cdot b_m$. Since K is a field (or an integral domain), and $a_n \neq 0, b_m \neq 0$, their product $a_n b_m$ must be non-zero. Therefore, the highest power of X in $P \cdot Q$ with a non-zero coefficient is X^{n+m}, which proves the lemma. ■

***Theorem* 6.8.3. Integral Domain Property**. If K is a field, then $K[X]$ is an integral domain.

Proof. We must verify the axioms of an integral domain:

(i) **Commutative Ring with Unity:** The polynomial ring inherits commutativity, associativity, and distributivity from the coefficient field K. The unity is the constant polynomial 1, which is distinct from 0.

(ii) **No Zero Divisors:** We must show that if $P \cdot Q = 0$, then either $P = 0$ or $Q = 0$. We prove the contrapositive: if $P \neq 0$ and $Q \neq 0$, then $P \cdot Q \neq 0$. Let P and Q be non-zero polynomials. By the previous lemma, the degree of their product is $\deg(P) + \deg(Q)$. Since degrees are non-negative integers, $\deg(P \cdot Q) \geq 0$. A polynomial with a non-negative degree cannot be the zero polynomial (which has undefined or negative degree). Thus, $P \cdot Q \neq 0$.

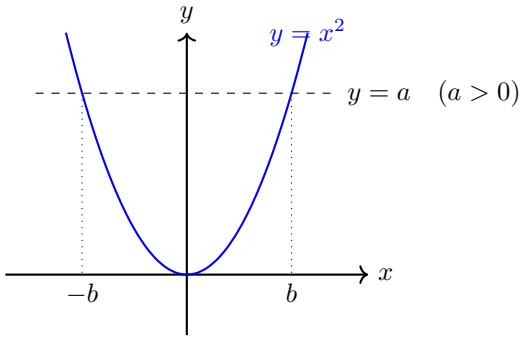
Since $K[X]$ is a commutative ring with unity and has no zero divisors, it is an integral domain. ■

Square Roots in Ordered Fields

Having established the structure of polynomial rings, we turn our attention to the roots of specific polynomials. The solvability of algebraic equations is a central theme in field theory. We consider the specific quadratic equation $x^2 = a$ within an arbitrary ordered field K. This investigation bridges the gap between algebra and the analytic properties of real numbers.

***Lemma* 6.8.2.** *Solutions to $x^2 = a$.* Let K be an ordered field and $a \in K \setminus \{0\}$. If the equation $x^2 = a$ possesses a solution in K, then $a > 0$. Furthermore, if $b \in K$ is a solution, then the equation has exactly two solutions in K, namely b and $-b$.

Proof. First, assume a solution b exists. Since $a \neq 0$, it follows that $b \neq 0$. In an ordered field, the square of any non-zero element is positive; thus $a = b^2 > 0$. Secondly, observe that $(-b)^2 = (-1 \cdot b)^2 = (-1)^2 \cdot b^2 = 1 \cdot b^2 = b^2 = a$. Hence, $-b$ is also a solution. To prove uniqueness, suppose x is any solution. Then $x^2 = b^2$, implying $x^2 - b^2 = 0$. Factoring yields $(x - b)(x + b) = 0$. Since K is a field and contains no zero divisors, it must be that $x - b = 0$ or $x + b = 0$, so $x = b$ or $x = -b$. Since $b \neq 0$, $b \neq -b$, establishing exactly two distinct solutions. ■



This lemma allows us to define the square root function rigorously by selecting the positive solution.

Definition 6.8.8. ***Square Root***. Let K be an ordered field and $a \in K$ with $a > 0$. If the equation $x^2 = a$ has a solution in K, the unique positive solution is called the square root of a, denoted $\sqrt{a}$. Additionally, we define $\sqrt{0} := 0$.

Remark. Properties of Square Roots:

(a) **Multiplicativity:** If $\sqrt{a}$ and $\sqrt{b}$ exist for some $a, b \geq 0$, then $\sqrt{ab}$ exists and $\sqrt{ab} = \sqrt{a}\sqrt{b}$. *Proof:* Let $x = \sqrt{a}$ and $y = \sqrt{b}$. Then $x^2 = a$ and $y^2 = b$. Consequently, $(xy)^2 = x^2y^2 = ab$. Since $x, y \geq 0$, their product $xy \geq 0$. Thus, xy is the unique non-negative number squaring to ab.
(b) **Relation to Absolute Value:** For all $x \in K$, $|x| = \sqrt{x^2}$. *Proof:* If $x \geq 0$, then x is the non-negative solution to $y^2 = x^2$, so $\sqrt{x^2} = x = |x|$. If $x < 0$, then $-x > 0$ and $(-x)^2 = x^2$, so $\sqrt{x^2} = -x = |x|$.
(c) **Square Roots in $\mathbb{Q}$:** For an integer $a \in \mathbb{Z}$, $\sqrt{a}$ exists in $\mathbb{Q}$ if and only if a is the square of a natural number. If a is not a perfect square (e.g., $a = 2$), $\sqrt{a}$ exists in $\mathbb{R}$ but not in $\mathbb{Q}$.

6.9 Exercises

1. **The "Shoes and Socks" Property.** Let $(G, \star)$ be a group. Prove that for any elements $a, b \in G$, the inverse of the product $a \star b$ is the product of the inverses in reverse order. That is:

$$(a \star b)^{-1} = b^{-1} \star a^{-1}$$

Remark. Multiply $(a \star b)$ by $(b^{-1} \star a^{-1})$ and use the associative property to show the result is the identity e.

2. **Uniqueness of the Identity.** The text proves the uniqueness of the identity element assuming it acts as an identity from both the left and the right. Suppose an element $e \in G$ satisfies only $e \star g = g$ for all $g \in G$ (a left identity). Suppose $e' \in G$ satisfies only $g \star e' = g$ for all $g \in G$ (a right identity). Prove that $e = e'$.

3. **Sudoku Property (Latin Squares).** Let $(G, \star)$ be a finite group. Prove that in the multiplication table (Cayley table) of the group, every element of G appears exactly once in each row and exactly once in each column.

Remark. Use the Cancellation Laws proved in the text. To show an element appears at most once, assume it appears twice and derive a contradiction. To show it appears at least once, consider the equation $a \star x = b$ and use the existence of inverses.

4. **Intersection of Subgroups.** Let H and K be two subgroups of a group G. Prove that their intersection $H \cap K$ is also a subgroup of G.

Remark. You must verify the three conditions for a subgroup: non-emptiness (does it contain e?), closure under the operation, and closure under inverses.

5. **The Exponential Map.** Consider the additive group of real numbers $(\mathbb{R}, +)$ and the multiplicative group of positive real numbers $(\mathbb{R}^+, \cdot)$. Let $\phi : \mathbb{R} \to \mathbb{R}^+$ be defined by $\phi(x) = e^x$.
   (a) Prove that ϕ is a group homomorphism.
   (b) Prove that ϕ is an isomorphism.

6. **Polynomial Units.** Let K be a field. An element $P \in K[X]$ is called a *unit* (or invertible) if there exists a polynomial $Q \in K[X]$ such that $P \cdot Q = 1$. Show that the units of $K[X]$ are exactly the non-zero constant polynomials.

Remark. Recall that $\deg(1) = 0$. Use the degree formula $\deg(P \cdot Q) = \deg(P) + \deg(Q)$. Does this formula hold if P or Q is the zero polynomial? Recall that $K[X]$ is an integral domain.

Chapter 7

Ordinal Numbers

In chapter 4, we established that two sets have the same size if a bijection exists between them. For partially ordered sets, the concept of an order isomorphism allows for a more refined comparison, identifying when two posets share the same order structure, or order type. This chapter generalises this idea, seeking to classify all well-ordered sets. We aim to define a canonical representative for each possible order type of a well-ordered set. These representatives will be called ordinal numbers.

7.1 Ordinals

A critical tool in this endeavour is a generalisation of the principle of induction to any well-ordered set, known as transfinite induction.

***Theorem* 7.1.1. Transfinite Induction**. Let $(A, \preceq)$ be a well-ordered set. If $S \subseteq A$ has the property that for every $a \in A$,

$$(\forall x \in A, (x \prec a \implies x \in S)) \implies a \in S$$

then $S = A$.

Proof. Let $(A, \preceq)$ be a well-ordered set and let $S \subseteq A$ be a set satisfying the given property. Suppose, for a contradiction, that $S \neq A$. Then the set difference $A \setminus S$ is non-empty. Since A is well-ordered, $A \setminus S$ must have a least element, which we will call m.

By the definition of m, for any element $x \in A$, if $x \prec m$, then $x \notin A \setminus S$, which means $x \in S$. We have thus shown that $\forall x \in A, (x \prec m \implies x \in S)$. By the hypothesis of the theorem, this implication forces the conclusion that $m \in S$. However, this contradicts our choice of m as the least element of $A \setminus S$. The initial assumption must be false; therefore, $S = A$. ■

Remark. The principle of strong induction is a special case of transfinite induction where the well-ordered set is $(\mathbb{N}, \leq)$.

Our first application of this principle is to demonstrate that an order-preserving function from a well-ordered set to itself cannot map an element to something strictly smaller.

***Lemma* 7.1.1.** Let $(A, \preceq)$ be a well-ordered set. If $f : A \to A$ is an order-preserving function, then $x \preceq f(x)$ for all $x \in A$.

Proof. Let $S = \{x \in A \mid x \preceq f(x)\}$. We use transfinite induction to show that $S = A$. Let $a \in A$ and assume that for all $x \prec a$, we have $x \in S$, meaning $x \preceq f(x)$. We must show $a \in S$.

Suppose, for a contradiction, that $f(a) \prec a$. Since f is order-preserving, applying f to this inequality gives $f(f(a)) \prec f(a)$. Let $y = f(a)$. We have found an element $y \in A$ such that $y \prec a$ and $f(y) \prec y$. However,

since $y \prec a$, our inductive hypothesis states that $y \in S$, which means $y \preceq f(y)$. This is a contradiction. Therefore, our assumption that $f(a) \prec a$ must be false. By trichotomy, it must be that $a \preceq f(a)$, which means $a \in S$.

By the principle of transfinite induction, $S = A$. ■

This lemma leads to several fundamental results about the structure of well-ordered sets.

***Lemma* 7.1.2.** For any well-ordered sets $(A, \preceq_A)$ and $(B, \preceq_B)$, there exists at most one order isomorphism from A to B.

Proof. Let $f : A \to B$ and $g : A \to B$ be two order isomorphisms. The inverse function $g^{-1} : B \to A$ is also an order isomorphism. Consequently, the composition $h = g^{-1} \circ f$ is an order isomorphism from A to A. As a composition of order-preserving functions, h is itself order-preserving. By the preceding lemma, we must have $x \preceq_A h(x)$ for all $x \in A$. Let us apply this same argument to the inverse function, $h^{-1} = f^{-1} \circ g$, which is also an order-preserving map from A to A. This gives $x \preceq_A h^{-1}(x)$ for all $x \in A$. Applying the order-preserving function h to this inequality yields $h(x) \preceq_A h(h^{-1}(x))$, which simplifies to $h(x) \preceq_A x$. We have established both $x \preceq_A h(x)$ and $h(x) \preceq_A x$. By the antisymmetry of $\preceq_A$, we conclude that $h(x) = x$ for all $x \in A$. Thus, $h = \mathrm{Id}_A$. This implies $g^{-1} \circ f = \mathrm{Id}_A$, and composing with g on the left gives $f = g$. ■

***Lemma* 7.1.3.** No well-ordered set is order isomorphic to any of its proper initial segments.

Proof. Let $(A, \preceq)$ be a well-ordered set and let S be a proper initial segment of A. Suppose, for a contradiction, that an order isomorphism $f : A \to S$ exists. Since f is an order-preserving function from A to A (as $S \subset A$), the lemma states that $x \preceq f(x)$ for all $x \in A$. However, since S is a proper initial segment, there exists some element $a \in A \setminus S$. The image of this element, $f(a)$, must be in the range of f, which is S. As S is an initial segment and $a \notin S$, it cannot be that $a \preceq f(a)$. In fact, we must have $f(a) \prec a$, which contradicts the lemma. Therefore, no such isomorphism can exist. ■

These results culminate in a trichotomy theorem for well-ordered sets, which asserts that any two such sets are comparable in a precise sense.

***Theorem* 7.1.2. Trichotomy for Well-Ordered Sets**. If $(A, \preceq_A)$ and $(B, \preceq_B)$ are well-ordered sets, then exactly one of the following holds:

(i) $(A, \preceq_A)$ is order isomorphic to $(B, \preceq_B)$.
(ii) $(A, \preceq_A)$ is order isomorphic to a proper initial segment of $(B, \preceq_B)$.
(iii) $(B, \preceq_B)$ is order isomorphic to a proper initial segment of $(A, \preceq_A)$.

This theorem suggests that we can select canonical representatives for the different order types of well-ordered sets. We now define these representatives.

Definition 7.1.1. ***Ordinal Number***. A set α is an ordinal number if it is a transitive set that is well-ordered by the elementhood relation $\in$.

Note. The natural numbers as we have defined them ($0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, etc.) are all ordinals. For any natural number n, the set n is transitive and is well-ordered by $\in$.

We now establish a sequence of fundamental properties of ordinals.

***Theorem* 7.1.3.** Every element of an ordinal is also an ordinal.

Proof. Let α be an ordinal and let $\beta \in \alpha$. Since α is a transitive set, $\beta \subseteq \alpha$. As a subset of a set that is well-ordered by $\in$, β is also well-ordered by $\in$. We must show that β is transitive. Let $\gamma \in \beta$ and $\delta \in \gamma$. Since $\beta \in \alpha$ and $\gamma \in \beta$, the transitivity of α implies $\gamma \in \alpha$. Similarly, since $\gamma \in \alpha$ and $\delta \in \gamma$, transitivity implies $\delta \in \alpha$. Now we have δ, γ, β are all elements of α, with $\delta \in \gamma$ and $\gamma \in \beta$. Since α is well-ordered by $\in$, and this relation is transitive on α, we conclude that $\delta \in \beta$. Therefore, β is a transitive set. As β is a transitive set that is well-ordered by $\in$, it is an ordinal. ■

A crucial property linking the subset relation to the elementhood relation for ordinals is the following.

***Theorem* 7.1.4.** Let α and β be ordinals. Then $\alpha \subset \beta$ if and only if $\alpha \in \beta$.

Proof. ($\Leftarrow$) Assume $\alpha \in \beta$. Since β is a transitive set, every element of β is also a subset of β. Thus $\alpha \subseteq \beta$. As no set can be an element of itself, $\alpha \neq \beta$, so $\alpha \subset \beta$.

($\Rightarrow$) Assume $\alpha \subset \beta$. Since α is a proper subset of β, the set difference $\beta \setminus \alpha$ is non-empty. As β is well-ordered by $\in$, this non-empty subset must have a least element; let us call it γ. We will show that $\alpha = \gamma$. First, we show $\alpha \subseteq \gamma$. Let $x \in \alpha$. Since γ is the least element of $\beta \setminus \alpha$, we have $x \notin \beta \setminus \alpha$. If we had $\gamma = x$ or $\gamma \in x$, then x would not be in α, a contradiction. As β is totally ordered by $\in$, the only remaining possibility is $x \in \gamma$. Thus $\alpha \subseteq \gamma$. Next, we show $\gamma \subseteq \alpha$. Let $x \in \gamma$. Since $\gamma \in \beta$ and β is transitive, we have $\gamma \subseteq \beta$, so $x \in \beta$. Since γ is the least element of $\beta \setminus \alpha$, x cannot be in $\beta \setminus \alpha$, so $x \in \alpha$. Thus $\gamma \subseteq \alpha$. We have shown $\alpha = \gamma$. Since $\gamma \in \beta$, we conclude $\alpha \in \beta$. ■

This theorem, combined with the fact that ordinals are well-ordered by $\in$, allows us to prove a trichotomy law for ordinals.

***Theorem* 7.1.5. Trichotomy for Ordinals**. For any two ordinals α and β, exactly one of the following holds:

$$\alpha \in \beta, \quad \beta \in \alpha, \quad \text{or} \quad \alpha = \beta$$

Proof. Consider the set $\alpha \cap \beta$. This set is well-ordered by $\in$ because it is a subset of the ordinal α. Let us show it is transitive. If $x \in y$ and $y \in \alpha \cap \beta$, then $y \in \alpha$ and $y \in \beta$. Since α and β are transitive, $x \in \alpha$ and $x \in \beta$, so $x \in \alpha \cap \beta$. Thus, $\alpha \cap \beta$ is an ordinal.

Let $\gamma = \alpha \cap \beta$. We have $\gamma \subseteq \alpha$ and $\gamma \subseteq \beta$. Suppose $\gamma \neq \alpha$ and $\gamma \neq \beta$. Then $\gamma \subset \alpha$ and $\gamma \subset \beta$. By the previous theorem, this implies $\gamma \in \alpha$ and $\gamma \in \beta$. Therefore, $\gamma \in \alpha \cap \beta$, which means $\gamma \in \gamma$, a contradiction. Thus, our supposition must be false. At least one of $\gamma = \alpha$ or $\gamma = \beta$ must be true.

- If $\gamma = \alpha$, then $\alpha = \alpha \cap \beta$, which implies $\alpha \subseteq \beta$.
- If $\gamma = \beta$, then $\beta = \alpha \cap \beta$, which implies $\beta \subseteq \alpha$.

So, for any two ordinals α, β, we have $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$. If both hold, then $\alpha = \beta$. If $\alpha \subset \beta$, then $\alpha \in \beta$. If $\beta \subset \alpha$, then $\beta \in \alpha$. These three cases are mutually exclusive. ■

Finally, we show that the class of ordinals is closed under the operation of taking arbitrary unions.

***Theorem* 7.1.6.** If $\mathcal{F}$ is a set of ordinals, then its union, $\bigcup \mathcal{F}$, is also an ordinal.

Proof. Let $\alpha = \bigcup \mathcal{F}$. We must show that α is a transitive set well-ordered by $\in$.

- **Transitivity**: Let $\beta \in \alpha$ and $\gamma \in \beta$. Since $\beta \in \alpha$, there must exist some ordinal $\delta \in \mathcal{F}$ such that $\beta \in \delta$. Since δ is an ordinal, it is transitive, so $\beta \subseteq \delta$. As $\gamma \in \beta$, it follows that $\gamma \in \delta$. Since $\delta \subseteq \alpha$, we have $\gamma \in \alpha$. Thus, α is transitive.
- **Well-ordering by** $\in$: Let S be a non-empty subset of α. Let β be any element of S. Since $S \subseteq \alpha$, we have $\beta \in \alpha$. The set $S \cap \beta$ is either empty or non-empty. If $S \cap \beta$ is non-empty, then as a non-empty subset of the ordinal β, it has a least element γ with respect to $\in$. This γ is the least element of S. If $S \cap \beta$ is empty, then for every $x \in S$, we have $x \notin \beta$. Since β and x are ordinals, the trichotomy law implies that for each $x \in S$, either $x = \beta$ or $\beta \in x$. As $S \cap \beta$ is empty, we must have $\beta \in x$ for all $x \in S$. Therefore, β is the least element of S.

In both cases, S has a least element, so α is well-ordered by $\in$. As α is a transitive set well-ordered by $\in$, it is an ordinal. ■

Remark. This theorem shows that for any set of ordinals, there exists a least upper bound, which is simply their union. This property is crucial for constructing larger and larger ordinals.

The Classification of Ordinals

Having established the fundamental properties of ordinals, we can now classify them into three disjoint categories: the zero ordinal, successor ordinals, and limit ordinals. This classification is exhaustive and provides the basis for a more structured form of transfinite induction.

For any ordinal α, we can construct its successor, denoted α^+, by taking the union of α with the singleton set containing it.

Definition 7.1.2. ***Successor Ordinal***. $\alpha^+ := \alpha \cup \{\alpha\}$. An ordinal β is a successor ordinal if there exists an ordinal α such that $\beta = \alpha^+$.

Every positive natural number is a successor ordinal. For example, $3 = 2 \cup \{2\}$, so $3 = 2^+$. It can be verified that if α is an ordinal, then α^+ is also an ordinal. The successor operation provides a way to generate a new, larger ordinal from any given one.

Ordinals that are not zero and cannot be reached by the successor operation form the second category.

Definition 7.1.3. ***Limit Ordinal***. A non-zero ordinal λ is a limit ordinal if it is not a successor ordinal.

Limit ordinals are characterised by the property that they are the union of all the ordinals they contain.

***Theorem* 7.1.7.** An ordinal λ is a limit ordinal if and only if $\lambda = \bigcup \lambda$.

Proof. Let λ be an ordinal. ($\Rightarrow$) Assume λ is a limit ordinal. Since λ is transitive, for any $\beta \in \lambda$, we have $\beta \subseteq \lambda$. Therefore, the union of all elements of λ must be a subset of λ, so $\bigcup \lambda \subseteq \lambda$. For the reverse inclusion, let $\beta \in \lambda$. Since λ is not a successor, $\beta \neq \lambda$ implies $\beta^+ \subset \lambda$. By a previous theorem, this means $\beta^+ \in \lambda$. Since $\beta \in \beta^+$, we have found an element of λ (namely, β^+) that contains β. Therefore, $\beta \in \bigcup \lambda$, which shows $\lambda \subseteq \bigcup \lambda$.

($\Leftarrow$) Assume $\lambda = \bigcup \lambda$. Suppose, for a contradiction, that λ is a successor ordinal, so $\lambda = \alpha^+$ for some ordinal α. Then α is the greatest element of λ with respect to $\in$. However, $\bigcup \lambda$ cannot have a greatest element. This is a contradiction, so λ must be a limit ordinal. ■

The smallest limit ordinal is ω, the set of all natural numbers. We can now state the fundamental classification theorem for ordinals.

***Theorem* 7.1.8.** Every ordinal is either the zero ordinal, a successor ordinal, or a limit ordinal.

This classification allows us to view the ordinals as a transfinite sequence, generated by repeatedly applying the successor operation and taking unions at limit stages:

$$0 \in 1 \in 2 \in \cdots \in \omega \in \omega^+ \in \omega^{++} \in \dots$$

This structure gives rise to a more practical version of transfinite induction.

***Theorem* 7.1.9. Transfinite Induction, Second Form**. Let $\mathbf{P}(\alpha)$ be a property of ordinals. If the following three conditions hold, then $\mathbf{P}(\alpha)$ is true for all ordinals α.

(i) $\mathbf{P}(0)$ is true.
(ii) For any ordinal α, if $\mathbf{P}(\alpha)$ is true, then $\mathbf{P}(\alpha^+)$ is true.
(iii) For any limit ordinal λ, if $\mathbf{P}(\beta)$ is true for all $\beta \in \lambda$, then $\mathbf{P}(\lambda)$ is true.

We are now equipped to prove the main theorem of this chapter: that ordinals serve as the canonical representatives for all well-ordered sets.

***Theorem* 7.1.10. Representation of Well-Ordered Sets**. Every well-ordered set is order isomorphic to a unique ordinal.

Proof. The proof of uniqueness follows from the fact that no two distinct ordinals can be order isomorphic. For existence, let $(A, \preceq)$ be a well-ordered set. We construct an order isomorphism f from an ordinal to A. The idea is to map 0 to the least element of A, 1 to the least element of the remainder, and so on. Formally, we define a function f on the class of ordinals by transfinite recursion:

$$f(\alpha) = \text{the least element of } A \setminus \{f(\beta) \mid \beta \in \alpha\}$$

The domain of this function is the set of all ordinals α for which the set $A \setminus \{f(\beta) \mid \beta \in \alpha\}$ is non-empty. Let this domain be γ. It can be shown that γ is an ordinal and that the function $f : \gamma \to A$ is an order isomorphism. Since the range of f must exhaust A, we have $\text{ran}(f) = A$, and therefore $(A, \preceq) \cong (\gamma, \in)$. ■

This unique ordinal is called the order type of the well-ordered set.

The Burali-Forti Paradox

The theory of ordinals, while powerful, leads to some profound limitations of set theory. One might be tempted to consider the "set of all ordinals". However, as discovered by Cesare Burali-Forti, the existence of such a set leads to a direct contradiction.

We first establish a preparatory lemma.

***Theorem* 7.1.11.** Any set of ordinals has a strict upper bound which is also an ordinal.

Proof. Let $\mathcal{F}$ be a set of ordinals. Let $\alpha = \bigcup \mathcal{F}$. We have already proven that α is an ordinal. Consider its successor, α^+. For any $\beta \in \mathcal{F}$, we have $\beta \subseteq \alpha$. By the trichotomy for ordinals, this means either $\beta \in \alpha$ or $\beta = \alpha$. In either case, $\beta \in \alpha^+$. Therefore, α^+ is an ordinal that is a strict upper bound for the set $\mathcal{F}$. ■

This result leads directly to the paradox.

***Theorem* 7.1.12. Burali-Forti Paradox**. There is no set of all ordinals.

Proof. Suppose, for a contradiction, that there exists a set $\mathcal{O}$ containing all ordinals. Since $\mathcal{O}$ is a set of ordinals, we can take its union, $\Omega = \bigcup \mathcal{O}$, which must itself be an ordinal. As Ω is an ordinal, it must be an element of the set of all ordinals, so $\Omega \in \mathcal{O}$. Since Ω is an element of $\mathcal{O}$, by the definition of a union, Ω must be a subset of $\bigcup \mathcal{O}$. This gives $\Omega \subseteq \Omega$. However, consider the successor ordinal Ω^+. As an ordinal, Ω^+ must be in $\mathcal{O}$. But $\Omega \in \Omega^+$ and $\Omega^+ \subseteq \bigcup \mathcal{O} = \Omega$, which implies $\Omega \in \Omega$. This contradicts the Axiom of Regularity, which states that no set can be an element of itself. Therefore, the initial assumption must be false, and no set of all ordinals can exist. ■

An important consequence of this paradox is Hartogs' theorem, which states that for any set, there is an ordinal that cannot be injected into it, meaning there is always an ordinal that is "larger" in some sense than any given set.

***Theorem* 7.1.13. Hartogs' Theorem**. For every set A, there exists an ordinal α such that there is no injective function from α to A.

Proof. Let A be a set. Define the collection of ordinals

$$\mathcal{E} = \{\alpha \mid \alpha \text{ is an ordinal and there exists an injection } f : \alpha \to A\}$$

Our first objective is to demonstrate that $\mathcal{E}$ is a set.

For each $\alpha \in \mathcal{E}$, let $f_\alpha : \alpha \to A$ be an injective function. The range of this function, $\text{ran}(f_\alpha)$, is a subset of A. We can use the bijection between α and its range to transport the well-ordering $\in$ from α to a well-ordering $\preceq_\alpha$ on $\text{ran}(f_\alpha)$. Specifically, for any $x, y \in \text{ran}(f_\alpha)$, we define $x \preceq_\alpha y \Leftrightarrow f_\alpha^{-1}(x) \in f_\alpha^{-1}(y)$. This makes $(\text{ran}(f_\alpha), \preceq_\alpha)$ a well-ordered set that is, by construction, order isomorphic to $(\alpha, \in)$.

Now, consider the collection of all possible well-orderings on subsets of A:

$$\mathcal{F} = \{(B, \preceq) \mid B \subseteq A \text{ and } \preceq \text{ is a well-ordering of } B\}$$

Each element $(B, \preceq)$ of $\mathcal{F}$ is a pair where $B \in \mathcal{P}(A)$ and $\preceq$ is a subset of $B \times B$, and therefore also a subset of $A \times A$. Consequently, $\mathcal{F}$ is a subset of the Cartesian product $\mathcal{P}(A) \times \mathcal{P}(A \times A)$. Since A is a set, the Power Set axiom guarantees that $\mathcal{P}(A)$ and $\mathcal{P}(A \times A)$ are sets. Their product is a set, and by the Schema of Separation, $\mathcal{F}$ is a set.

By the Representation of Well-Ordered Sets theorem, for each pair $(B, \preceq) \in \mathcal{F}$, there exists a unique ordinal, its order type. This defines a function $G : \mathcal{F} \to$ Ordinals where $G((B, \preceq))$ is the unique ordinal isomorphic to $(B, \preceq)$. Since the domain $\mathcal{F}$ is a set, the Axiom of Replacement ensures that the range of this function, $\text{ran}(G)$, is also a set.

For any ordinal $\alpha \in \mathcal{E}$, we constructed a well-ordered set $(\text{ran}(f_\alpha), \preceq_\alpha) \in \mathcal{F}$ which is order isomorphic to α. This means that every ordinal in $\mathcal{E}$ is the order type of some element of $\mathcal{F}$, and is therefore in the range of G. Thus, $\mathcal{E} \subseteq \text{ran}(G)$. Since $\text{ran}(G)$ is a set, $\mathcal{E}$ must also be a set by the Schema of Separation.

As $\mathcal{E}$ is a set of ordinals, the Burali-Forti Paradox implies that $\mathcal{E}$ cannot contain all ordinals. Therefore, there must exist an ordinal α such that $\alpha \notin \mathcal{E}$. By the definition of $\mathcal{E}$, this means there is no injective function from α to A. ■

7.2 Transfinite Recursion and the Well-Ordering Theorem

The principle of transfinite induction provides a method for proving that a property holds for all elements of a well-ordered set. A related principle, transfinite recursion, provides a method for defining functions on well-ordered sets. Just as standard recursion on the natural numbers defines a function's value at $n+1$ based on its value at n, transfinite recursion defines a function's value at an ordinal α based on its values for all ordinals smaller than α.

To state the theorem formally, we require some notation. For a set A and an ordinal α, let $A^{<\alpha}$ denote the set of all functions whose domain is an ordinal smaller than α and whose codomain is A.

$$A^{<\alpha} = \bigcup_{\beta \in \alpha} A^\beta = \{f \mid \exists \beta \in \alpha, f : \beta \to A\}$$

***Theorem* 7.2.1. Transfinite Recursion**. Let α be an ordinal and let $G : A^{<\alpha} \to A$ be a function. There exists a unique function $F : \alpha \to A$ such that for every $\beta \in \alpha$,

$$F(\beta) = G(F|_\beta)$$

where $F|_\beta$ is the restriction of F to the domain β.

Proof. We first prove existence and then uniqueness.

Consider the collection of all functions that satisfy the recursive definition on some initial segment of α:

$$\mathcal{H} = \{f \mid \exists \delta \in \alpha \cup \{\alpha\}, (f : \delta \to A \wedge \forall \beta \in \delta, f(\beta) = G(f|_\beta))\}$$

The condition defining $\mathcal{H}$ is a first-order formula, and $\mathcal{H} \subseteq \mathcal{P}(\alpha \times A)$. By the Axiom of Power Set and the Schema of Separation, $\mathcal{H}$ is a set. Note that the empty function $\emptyset$ (with domain 0) is in $\mathcal{H}$, so $\mathcal{H}$ is non-empty.

Let $f, g \in \mathcal{H}$ with domains δ and ε respectively. By the trichotomy for ordinals, either $\delta \subseteq \varepsilon$ or $\varepsilon \subseteq \delta$. Assume without loss of generality that $\delta \subseteq \varepsilon$. We claim that $f \subseteq g$. If not, the set $S = \{\beta \in \delta \mid f(\beta) \neq g(\beta)\}$ is non-empty and has a least element, β_0. By minimality, $f(\gamma) = g(\gamma)$ for all $\gamma \in \beta_0$, so $f|_{\beta_0} = g|_{\beta_0}$. The recursive definition then implies $f(\beta_0) = G(f|_{\beta_0}) = G(g|_{\beta_0}) = g(\beta_0)$, a contradiction. Thus, S is empty and $f \subseteq g$. This shows that $(\mathcal{H}, \subseteq)$ is a chain.

Let $F = \bigcup \mathcal{H}$. Since $\mathcal{H}$ is a chain of compatible functions, their union F is also a function. The domain of F, $\text{dom}(F) = \bigcup\{\text{dom}(f) \mid f \in \mathcal{H}\}$, is a union of ordinals and therefore is itself an ordinal, which we will call δ_F. It can be shown, by the same logic as above, that F satisfies the recursive definition for all $\beta \in \delta_F$.

We must show that the domain of F is all of α. Suppose, for a contradiction, that $\delta_F \subset \alpha$. Then δ_F is an ordinal and an element of α. Since $F : \delta_F \to A$, $F \in A^{<\alpha}$, so $G(F)$ is defined. We can construct a new function $F' = F \cup \{(\delta_F, G(F))\}$. The domain of F' is δ_F^+, and F' satisfies the recursive definition on its domain. Therefore, $F' \in \mathcal{H}$. But $F \subset F'$, which contradicts the fact that F is the union of all functions in $\mathcal{H}$. Thus, our assumption was false, and $\text{dom}(F) = \alpha$.

As for Uniqueness. Suppose $F_1, F_2 : \alpha \to A$ both satisfy the recursion equation. Let $S = \{\beta \in \alpha \mid F_1(\beta) \neq F_2(\beta)\}$. If S is non-empty, it has a least element β_0. By minimality, $F_1(\gamma) = F_2(\gamma)$ for all $\gamma \in \beta_0$, so the restrictions $F_1|_{\beta_0}$ and $F_2|_{\beta_0}$ are equal. This implies $F_1(\beta_0) = G(F_1|_{\beta_0}) = G(F_2|_{\beta_0}) = F_2(\beta_0)$, which contradicts $\beta_0 \in S$. Therefore, S must be empty, and $F_1 = F_2$. ■

A common application of this theorem is to define sequences indexed by ordinals.

***Proposition* 7.2.1.** Let A be a set, $a_0 \in A$ be a starting element, and $g : A \to A$ be a successor function. For any ordinal α, there exists a unique function $F : \alpha \to A$ such that:

- $F(0) = a_0$.
- $F(\beta^+) = g(F(\beta))$ for all β such that $\beta^+ \in \alpha$.
- $F(\lambda) = \bigcup_{\beta\in\lambda}\{F(\beta)\}$ for all limit ordinals $\lambda \in \alpha$, provided a suitable notion of union or limit is defined on A.

The Well-Ordering Theorem

The Representation Theorem established that every well-ordered set has the order type of an ordinal. However, it does not apply to sets like $(\mathbb{Z}, \leq)$ or $(\mathbb{R}, \leq)$ which are not well-ordered. A landmark result, first proved by Ernst Zermelo, states that the Axiom of Choice implies that *every* set can be equipped with a well-ordering.

***Theorem* 7.2.2. Well-Ordering Theorem**. Every set can be well-ordered.

Proof. Let A be an arbitrary set. We will show there exists a relation $\preceq$ on A such that $(A, \preceq)$ is a well-ordered set. This theorem is equivalent to the Axiom of Choice. We prove it using Zorn's Lemma.

Let $\mathcal{W}$ be the collection of all well-orderings on subsets of A. An element of $\mathcal{W}$ is a pair $(S, \preceq_S)$ where $S \subseteq A$ and $\preceq_S$ is a well-ordering of S. We define a partial order $\sqsubseteq$ on $\mathcal{W}$ as follows:

$$(S_1, \preceq_1) \sqsubseteq (S_2, \preceq_2)$$

if S_1 is an initial segment of S_2 and the ordering $\preceq_1$ is the restriction of $\preceq_2$ to S_1.

Let $\mathcal{C} = \{(S_i, \preceq_i) \mid i \in I\}$ be a chain in $(\mathcal{W}, \sqsubseteq)$. We must show it has an upper bound in $\mathcal{W}$. Let $U = \bigcup_{i\in I} S_i$ and let $\preceq_U = \bigcup_{i\in I} \preceq_i$. It can be verified that $(U, \preceq_U)$ is a well-ordered set and is an upper bound for the chain $\mathcal{C}$.

Since every chain has an upper bound, by Zorn's Lemma, there exists a maximal element in $\mathcal{W}$; let this be $(M, \preceq_M)$. We claim that $M = A$. Suppose, for a contradiction, that $M \neq A$. Then there exists an element $x \in A \setminus M$. We can construct a new well-ordered set $(M', \preceq_{M'})$ by defining $M' = M \cup \{x\}$ and extending the order $\preceq_M$ such that $m \prec_{M'} x$ for all $m \in M$. This makes M a proper initial segment of M'. The pair $(M', \preceq_{M'})$ is an element of $\mathcal{W}$, and we have $(M, \preceq_M) \sqsubset (M', \preceq_{M'})$. This contradicts the maximality of $(M, \preceq_M)$. Therefore, our assumption that $M \neq A$ must be false.

We have found a maximal well-ordering $(A, \preceq_A)$ on the entirety of A. ■

The Well-Ordering Theorem is often considered the most counter-intuitive consequence of the Axiom of Choice, as it implies the existence of a well-ordering on sets like the real numbers, even though no such ordering can be explicitly constructed. This theorem, along with Zorn's Lemma, completes what is often called the "grand equivalency" of fundamental principles in set theory.

7.3 Exercises

1. Use the principle of transfinite induction to prove that for any ordinal α, it is not the case that $\alpha \in \alpha$.

Remark. Let S be the collection of all ordinals β such that $\beta \notin \beta$. Assume there is an ordinal not in S, and consider the least such ordinal.

2. Prove that there is no largest ordinal. That is, for any ordinal α, prove there exists an ordinal β such that $\alpha \in \beta$.

3. The text classifies ordinals as zero, successor, or limit.

(a) Prove that if α is a successor ordinal, say $\alpha = \gamma^+$, then $\bigcup \alpha = \gamma$.
(b) Conversely, prove that if α is an ordinal and $\bigcup \alpha \in \alpha$, then α must be a successor ordinal.

4. Ordinal Addition. We can define addition of ordinals using transfinite recursion. For a fixed ordinal α, we define a function f_α on the class of all ordinals as follows:

(i) $f_\alpha(0) = \alpha$
(ii) $f_\alpha(\beta^+) = (f_\alpha(\beta))^+$ for any ordinal β.
(iii) $f_\alpha(\lambda) = \bigcup_{\gamma \in \lambda} f_\alpha(\gamma)$ for any limit ordinal λ.

We then define $\alpha + \beta := f_\alpha(\beta)$. Using this definition, compute $1 + \omega$ and $\omega + 1$. Are they equal?

Note. Recall that ω is the set of natural numbers, so it is the limit of the sequence $0, 1, 2, \ldots$. Thus, for the calculation of $1 + \omega$, you will need to evaluate the union of $1 + n$ for all $n < \omega$.

5. Let $(A, \preceq)$ be a well-ordered set. Use the Representation of Well-Ordered Sets theorem to prove that every element $a \in A$, except for a possible greatest element, has an immediate successor.

Remark. Consider the order isomorphism $f : \alpha \to A$ for some ordinal α. What ordinal corresponds to the successor of an element a?

6. Let α and β be ordinals. Prove that if there exists a surjective function $f : \alpha \to \beta$, then $\beta \subseteq \alpha$, which by the trichotomy for ordinals means $\beta \in \alpha$ or $\beta = \alpha$.

Remark. For each $\gamma \in \beta$, the pre-image $f^{-1}[\{\gamma\}]$ is a non-empty subset of α. Construct an injective function from β to α.

7. The text proves the Well-Ordering Theorem using Zorn's Lemma. Prove the converse for a specific case: Show how the Well-Ordering Theorem implies the existence of a choice function for any family $\mathcal{F}$ of non-empty sets.

8. $\star$ Use Hartogs' Theorem to prove that there can be no "set of all sets".

Remark. Suppose a set V containing all sets exists. Apply Hartogs' Theorem to V to find an ordinal α that cannot be injected into V. Since α is itself a set, what does this imply?

7.4 Equinumerosity

The concept of an order type classifies well-ordered sets according to their structure. We now introduce a more general criterion for comparing the size of any two sets, finite or infinite. The foundational idea, developed by Georg Cantor, is that two sets have the same size if their elements can be placed into a one-to-one correspondence. This is formalised using the concept of a bijection.

Definition 7.4.1. ***Equinumerous Sets***. Two sets A and B are equinumerous, denoted $A \approx B$, if there exists a bijection $f : A \to B$. If no such bijection exists, we write $A \not\approx B$.

Example 7.4.1.

- Let $m \in \mathbb{Z}$ with $m \neq 0$. The set of all multiples of m, denoted $m\mathbb{Z}$, is equinumerous with $\mathbb{Z}$. The function $f : \mathbb{Z} \to m\mathbb{Z}$ defined by $f(n) = mn$ is a bijection. It is injective because if $f(n_1) = f(n_2)$, then $mn_1 = mn_2$, which implies $n_1 = n_2$ as $m \neq 0$. It is surjective because any element of $m\mathbb{Z}$ is of the form mk for some $k \in \mathbb{Z}$, which is the image of k under f.
- The set of integers $\mathbb{Z}$ is equinumerous with the set of positive integers $\mathbb{Z}^+$. A bijection $g : \mathbb{Z}^+ \to \mathbb{Z}$ can be defined piecewise:
$$g(n) = \begin{cases} (n-1)/2 & \text{if } n \text{ is odd} \\ -n/2 & \text{if } n \text{ is even} \end{cases}$$
This function maps the odd positive integers to the non-negative integers $\{0, 1, 2, \dots\}$ and the even positive integers to the negative integers $\{-1, -2, -3, \dots\}$, covering all of $\mathbb{Z}$.

The relation of equinumerosity behaves like an equivalence relation, although it cannot be one in the formal sense.

***Theorem* 7.4.1.** For any sets A, B, C:

(i) $A \approx A$ (Reflexivity).
(ii) If $A \approx B$, then $B \approx A$ (Symmetry).
(iii) If $A \approx B$ and $B \approx C$, then $A \approx C$ (Transitivity).

Proof.

(i) The identity map $\mathrm{Id}_A : A \to A$ is a bijection.
(ii) If $f : A \to B$ is a bijection, then its inverse $f^{-1} : B \to A$ is also a bijection.
(iii) If $f : A \to B$ and $g : B \to C$ are bijections, their composition $g \circ f : A \to C$ is a bijection.

■

Remark. These properties do not make $\approx$ a formal equivalence relation because its domain would have to be a "set of all sets", which does not exist by the Burali-Forti Paradox.

Example 7.4.2. The open interval $(0, 1)$ is equinumerous with the set of all real numbers $\mathbb{R}$. First, the linear function $f : (0, 1) \to (-\pi/2, \pi/2)$ defined by $f(x) = \pi x - \pi/2$ is a bijection. Second, the function $g : (-\pi/2, \pi/2) \to \mathbb{R}$ defined by $g(x) = \tan(x)$ is a bijection. By transitivity, the composition $g \circ f$ is a bijection from $(0, 1)$ to $\mathbb{R}$, so $(0, 1) \approx \mathbb{R}$.

Cardinal Dominance

If equinumerosity corresponds to equality of size, the existence of an injection from one set to another corresponds to an inequality.

Definition 7.4.2. ***Cardinal Dominance***. A set B dominates a set A, denoted $A \preceq B$, if there exists an injective function $f : A \to B$. If $A \preceq B$ but $A \not\approx B$, we write $A \prec B$.

Example 7.4.3. If $A \subseteq B$, then $A \preceq B$, since the inclusion map $i : A \to B$ defined by $i(x) = x$ is an injection. For instance, $\mathbb{N} \preceq \mathbb{Z} \preceq \mathbb{Q} \preceq \mathbb{R}$.

The existence of a surjection in the opposite direction is an equivalent condition for dominance.

***Theorem* 7.4.2.** For any non-empty sets A and B, $A \preceq B$ if and only if there exists a surjective function $g : B \to A$.

Proof. ($\Rightarrow$) Assume $A \preceq B$, so there is an injection $f : A \to B$. Since A is non-empty, fix an element $a_0 \in A$. Define $g : B \to A$ by

$$g(y) = \begin{cases} f^{-1}(y) & \text{if } y \in \text{ran}(f) \\ a_0 & \text{if } y \notin \text{ran}(f) \end{cases}$$

This function is surjective because for any $a \in A$, its image $f(a)$ is in ran(f), and $g(f(a)) = f^{-1}(f(a)) = a$.

($\Leftarrow$) Assume there is a surjection $g : B \to A$. For each $a \in A$, the pre-image set $g^{-1}[\{a\}]$ is non-empty. By the Axiom of Choice, we can choose one element from each of these pre-image sets. This defines a function $f : A \to B$ such that for every $a \in A$, $g(f(a)) = a$. To show f is injective, let $a_1, a_2 \in A$ and assume $f(a_1) = f(a_2)$. Applying g gives $g(f(a_1)) = g(f(a_2))$, which implies $a_1 = a_2$. Thus, f is an injection and $A \preceq B$. ■

A fundamental theorem in set theory states that this dominance relation is antisymmetric when viewed through the lens of equinumerosity.

***Theorem* 7.4.3. Cantor–Schröder–Bernstein**. For any sets A and B, if $A \preceq B$ and $B \preceq A$, then $A \approx B$.

Proof. Let $f : A \to B$ and $g : B \to A$ be injective functions. We partition the set A into three disjoint subsets based on the "ancestry" of its elements. An element $x \in A$ can be traced back through alternating applications of g^{-1} and f^{-1}. The chain of ancestors for x is the sequence $x, g^{-1}(x), f^{-1}(g^{-1}(x)), \ldots$.

- Let A_A be the set of elements whose chain of ancestors terminates in A (i.e., at an element of $A \backslash \text{ran}(g)$).
- Let A_B be the set of elements whose chain of ancestors terminates in B (i.e., at an element of $B \backslash \text{ran}(f)$).
- Let A_∞ be the set of elements whose chain of ancestors is infinite.

These three sets form a partition of A. Similarly, we can partition B into B_A, B_B, B_∞. The function f maps A_A to B_A and A_∞ to B_∞. The function g maps B_B to A_B. We can now define a bijection $h : A \to B$ piecewise:

$$h(x) = \begin{cases} f(x) & \text{if } x \in A_A \cup A_\infty \\ g^{-1}(x) & \text{if } x \in A_B \end{cases}$$

The function f provides a bijection from $A_A \cup A_\infty$ to $B_A \cup B_\infty$. The function g^{-1} provides a bijection from A_B to B_B. Combining these gives a bijection from A to B. ■

Example 7.4.4. The closed interval $[0, 1]$ is equinumerous with the open interval $(0, 1)$.

- The inclusion map $i : (0, 1) \to [0, 1]$ where $i(x) = x$ is an injection, so $(0, 1) \preceq [0, 1]$.
- The linear function $f : [0, 1] \to (0, 1)$ defined by $f(x) = \frac{1}{2}x + \frac{1}{4}$ maps the interval $[0, 1]$ to the interval $[\frac{1}{4}, \frac{3}{4}]$, which is a subset of $(0, 1)$. This function is an injection, so $[0, 1] \preceq (0, 1)$.

By the Cantor–Schröder–Bernstein theorem, we conclude that $[0, 1] \approx (0, 1)$.

Diagonalization

To prove that two sets are equinumerous, we must construct a bijection. To prove that one set dominates another, we must construct an injection. Proving a strict dominance, $A \prec B$, is more difficult, as it requires showing not only that an injection from A to B exists, but also that no bijection between them can possibly exist. Cantor's diagonalization argument is a powerful and elegant proof technique for establishing such non-equinumerosity.

The method is best illustrated with its classic application: proving that the set of natural numbers is not equinumerous with the set of all infinite sequences of binary digits. Let S be the set of all functions from $\mathbb{N}$ to $\{0, 1\}$. Each such function can be viewed as an infinite binary sequence. We will show $\mathbb{N} \prec S$.

First, $\mathbb{N} \preceq S$ is straightforward to establish. The function that maps each natural number n to the sequence containing a 1 at position n and 0s elsewhere is an injection. The core of the argument is to show that $\mathbb{N} \not\approx S$. We do this by demonstrating that no function from $\mathbb{N}$ to S can be surjective.

Let $f : \mathbb{N} \to S$ be any arbitrary function. We can create a list of the sequences in the range of f:

$$\begin{aligned} f(0) &= (a_{0,0}, a_{0,1}, a_{0,2}, \dots) \\ f(1) &= (a_{1,0}, a_{1,1}, a_{1,2}, \dots) \\ f(2) &= (a_{2,0}, a_{2,1}, a_{2,2}, \dots) \\ &\vdots \\ f(n) &= (a_{n,0}, a_{n,1}, a_{n,2}, \dots, a_{n,n}, \dots) \\ &\vdots \end{aligned}$$

We now construct a new binary sequence, s_d, which is not in this list. The construction proceeds "down the diagonal" of the matrix of $a_{i,j}$ values. We define the n-th term of s_d to be the opposite of the n-th term of the sequence $f(n)$. Formally, let $s_d = (d_0, d_1, d_2, \dots)$, where for each $n \in \mathbb{N}$,

$$d_n = 1 - a_{n,n}$$

The sequence s_d is an element of S. However, by its very construction, s_d cannot be in the range of f. For any $n \in \mathbb{N}$, the sequence s_d differs from the sequence $f(n)$ in the n-th position, since $d_n \neq a_{n,n}$. Therefore, $s_d \notin \text{ran}(f)$. Since our choice of f was arbitrary, we have shown that no function from $\mathbb{N}$ to S can be surjective. Consequently, no bijection exists, and $\mathbb{N} \not\approx S$. We conclude that $\mathbb{N} \prec S$.

As the set of real numbers in $[0, 1]$ can be put into one-to-one correspondence with the set of infinite binary sequences, this argument also shows $\mathbb{N} \prec [0, 1]$, and by extension, $\mathbb{N} \prec \mathbb{R}$.

Cantor's Theorem

The diagonalization argument can be generalised to prove a profound result about the relationship between any set and its power set. This result, known as Cantor's Theorem, establishes that there is an infinite hierarchy of sizes of infinity.

The proof uses the concept of a characteristic function. For any subset $S \subseteq A$, its characteristic function $\chi_S : A \to \{0, 1\}$ is defined by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

The set of all such functions, $\{0, 1\}^A$, is equinumerous with the power set $\mathcal{P}(A)$.

***Theorem* 7.4.4. Cantor's Theorem**. For any set A, $A \prec \mathcal{P}(A)$.

Proof. We must prove two claims: $A \preceq \mathcal{P}(A)$ and $A \not\approx \mathcal{P}(A)$.

(i) To show $A \preceq \mathcal{P}(A)$, we must find an injection from A to its power set. The function $f : A \to \mathcal{P}(A)$ defined by $f(a) = \{a\}$ is an injection. If $f(a_1) = f(a_2)$, then $\{a_1\} = \{a_2\}$, which implies $a_1 = a_2$.

(ii) To show $A \not\approx \mathcal{P}(A)$, we prove that no function from A to $\mathcal{P}(A)$ can be surjective. Let $g : A \to \mathcal{P}(A)$ be an arbitrary function. We will construct a set $D \in \mathcal{P}(A)$ that is not in the range of g.
Consider the following subset of A, which is constructed by a "diagonal" argument:

$$D = \{x \in A \mid x \notin g(x)\}$$

This set contains every element of A that is not a member of the subset to which it is mapped by g. Since $D \subseteq A$, D is an element of $\mathcal{P}(A)$. We claim that D is not in the range of g. Suppose, for a

contradiction, that it is. Then there must exist some element $d \in A$ such that $g(d) = D$. We now ask: is the element d in the set D?

- If $d \in D$, then by the definition of D, we must have $d \notin g(d)$. But since $g(d) = D$, this means $d \notin D$. This is a contradiction.
- If $d \notin D$, then by the definition of D, it is not the case that $d \notin g(d)$, which means $d \in g(d)$. But since $g(d) = D$, this means $d \in D$. This is also a contradiction.

Both possibilities lead to a contradiction. Therefore, our initial assumption must be false: there is no element $d \in A$ such that $g(d) = D$. The function g is not surjective.

Since we have shown that an injection exists but no surjection (and therefore no bijection) can exist, we conclude that $A \prec \mathcal{P}(A)$. ■

Cantor's Theorem has a staggering implication: there is no "largest set". For any set, we can always find a set that is strictly larger in cardinality, namely its power set. This gives rise to an endless hierarchy of infinite magnitudes:

$$\mathbb{N} \prec \mathcal{P}(\mathbb{N}) \prec \mathcal{P}(\mathcal{P}(\mathbb{N})) \prec \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}))) \prec \dots$$

7.5 Exercises

Part I: Equinumerosity and Cardinal Dominance

1. Prove that the following pairs of sets are equinumerous by constructing an explicit bijection between them.

(a) The set of positive integers $\mathbb{Z}^+$ and the set of negative integers $\mathbb{Z}^-$.

(b) Any two open intervals of real numbers, (a, b) and (c, d), where $a < b$ and $c < d$.

(c) The set of points on a line, $\{(x, y) \in \mathbb{R}^2 \mid y = 2x + 4\}$, and the set $\mathbb{R}$.

2. Let A and B be non-empty sets. Prove that $A \preceq A \times B$.

3. Prove the following properties for any non-empty sets A, B, C, D.

(a) If $A \preceq B$ and $B \approx C$, then $A \preceq C$.

(b) If $A \approx B$ and $C \approx D$, then $A \times C \approx B \times D$.

4. Let A be an infinite set and let $S = \{a_1, a_2, \dots, a_n\}$ be a finite set of distinct elements from A. Prove that $A \approx A \setminus S$.

Remark. This is a generalisation of Hilbert's Hotel problem. Recall that every infinite set contains a countably infinite subset (a subset equinumerous with $\mathbb{N}$). Use this subset to absorb the finite set S.

5. Prove that if $A \preceq B$, then $\mathcal{P}(A) \preceq \mathcal{P}(B)$.

Remark. Let $f : A \to B$ be an injection. Use this function to define a new function $g : \mathcal{P}(A) \to \mathcal{P}(B)$ by considering the image of subsets under f.

Part II: The Cantor-Schröder-Bernstein Theorem

6. Use the Cantor-Schröder-Bernstein theorem to prove that any two closed intervals $[a, b]$ and $[c, d]$ (with $a < b, c < d$) are equinumerous.

7. Use the Cantor-Schröder-Bernstein theorem to prove that the set of all infinite binary sequences is equinumerous with the power set of the natural numbers, $\mathcal{P}(\mathbb{N})$.

8. A classic and powerful result is that $\mathbb{R} \approx \mathbb{R} \times \mathbb{R}$.

(a) Prove that $\mathbb{R} \preceq \mathbb{R} \times \mathbb{R}$.

(b) Prove that $\mathbb{R} \times \mathbb{R} \preceq \mathbb{R}$.

Remark. Consider representing each real number in $(0, 1)$ by its unique, non-terminating decimal expansion. An injection $f : (0, 1) \times (0, 1) \to (0, 1)$ can be constructed by interleaving the digits of the two input numbers.

(c) Conclude that $\mathbb{R} \approx \mathbb{R} \times \mathbb{R}$ and, by extension, that $\mathbb{R} \approx \mathbb{C}$.

9. $\star$ Without using the Cantor-Schröder-Bernstein theorem, prove that the open interval $(0, 1)$ is equinumerous with the closed interval $[0, 1]$ by constructing an explicit bijection.

Part III: Diagonalization and Cantor's Theorem

10. The text showed $\mathbb{N} \prec \{0, 1\}^{\mathbb{N}}$ using binary sequences. Use a diagonalization argument to prove directly that the set of natural numbers $\mathbb{N}$ is not equinumerous with the open interval $(0, 1)$.

Remark. Construct a decimal number $0.d_1d_2d_3\ldots$ where each digit d_n is chosen to differ from the n-th digit of the n-th number in the list. To avoid issues with non-unique decimal representations (like $0.199\cdots = 0.200\ldots$), avoid using the digits 0 and 9 in your construction.

11. Prove that $\mathbb{N} \approx \mathbb{N} \times \mathbb{N}$.

Remark. Consider arranging the pairs (m, n) in an infinite grid and finding a path that visits every pair exactly once.

12. A number is called *algebraic* if it is a root of a polynomial with integer coefficients. Prove that the set of all algebraic numbers is equinumerous with $\mathbb{N}$, and use this to show that there exist transcendental (non-algebraic) numbers.

13. Let A be any set. Prove that the set of its power set, $\mathcal{P}(A)$, is equinumerous with the set of all functions from A to $\{0, 1\}$, denoted $\{0, 1\}^A$.

Remark. Use the concept of the characteristic function.

14. Use Cantor's Theorem to prove that there is no "set of all sets".

Remark. Suppose a set V containing all sets exists. Apply Cantor's Theorem to V to derive a contradiction.

15. Does there exist a set A such that $\mathcal{P}(A) \preceq A$? Justify your answer using the theorems from this chapter.

7.6 Cardinal Numbers

Having established that every well-ordered set is order isomorphic to a unique ordinal, we now seek a more general method for comparing the size of any two sets. The concept of equinumerosity, defined by the existence of a bijection, partitions the class of all sets into equivalence classes. We aim to select a canonical representative from each of these classes, which will serve as a measure of a set's size.

By the Well-Ordering Theorem, any set A can be well-ordered, and is therefore order isomorphic to some ordinal α. Consequently, the collection of ordinals equinumerous to A,

$$\mathcal{H}_A = \{\alpha \mid \alpha \text{ is an ordinal and } \alpha \approx A\}$$

is non-empty. As this is a collection of ordinals, it has a least element. This least element is the initial ordinal equinumerous to A, meaning it is not equinumerous to any smaller ordinal. This motivates the following definition.

Definition 7.6.1. ***Cardinal Number***. An ordinal κ is a cardinal number if it is not equinumerous to any of its elements. That is, for every $\alpha \in \kappa$, $\kappa \not\approx \alpha$.

The natural numbers $0, 1, 2, \ldots$ are all cardinal numbers. An infinite ordinal α that is a cardinal must be a limit ordinal; if α were a successor ordinal, say $\alpha = \beta^+$, then $\alpha \approx \beta$ for infinite β, which would contradict the definition of a cardinal. However, not every limit ordinal is a cardinal. For example, $\omega + \omega$ is a limit ordinal, but it is equinumerous with ω, an element of $\omega + \omega$.

The crucial property of cardinals is that they serve as unique representatives for each equinumerosity class.

***Theorem* 7.6.1.** Every set is equinumerous to exactly one cardinal number.

Proof. Let A be a set. The existence of at least one such cardinal follows from the argument that the collection $\mathcal{H}_A$ has a least element, which is a cardinal. For uniqueness, suppose $A \approx \kappa_1$ and $A \approx \kappa_2$ where κ_1 and κ_2 are cardinals. By transitivity, $\kappa_1 \approx \kappa_2$. Since they are both ordinals, the Trichotomy for Ordinals states that either $\kappa_1 \in \kappa_2$, $\kappa_2 \in \kappa_1$, or $\kappa_1 = \kappa_2$. If $\kappa_1 \in \kappa_2$, then κ_2 would be equinumerous to one of its elements, contradicting that it is a cardinal. A symmetric argument holds if $\kappa_2 \in \kappa_1$. The only remaining possibility is $\kappa_1 = \kappa_2$. ■

Definition 7.6.2. ***Cardinality***. The cardinality of a set A, denoted $|A|$, is the unique cardinal equinumerous to A.

Note. By this definition, the cardinality of a cardinal κ is κ itself.

Finite Sets

The concept of a finite set, intuitively understood as one whose elements can be counted, is formalised by identifying its cardinality with a natural number.

Definition 7.6.3. ***Finite and Infinite Sets***. A set A is finite if its cardinality is a natural number, i.e., $|A| \in \omega$. If a set is not finite, it is infinite.

A fundamental property, which distinguishes finite from infinite sets, is that a finite set cannot be put into one-to-one correspondence with a part of itself. We establish this first for the natural numbers. A preparatory lemma is required.

***Lemma* 7.6.1.** Let n be a positive natural number. If $k \in n$, then $n \setminus \{k\} \approx n - 1$.

Proof. We proceed by induction on n. The basis case $n = 1$ is trivial. Assume the statement holds for some $n \geq 1$, and consider $n + 1$. Let $k \in n + 1$. If $k = n$, then $(n + 1) \setminus \{n\} = n$, which is equinumerous to $(n + 1) - 1$. If $k < n$, by the induction hypothesis there is a bijection $g : n \setminus \{k\} \to n - 1$. We define a bijection $h : (n + 1) \setminus \{k\} \to n$ by $h(x) = g(x)$ for all $x \in n \setminus \{k\}$ and $h(n) = n - 1$. ■

***Theorem* 7.6.2.** No natural number is equinumerous to a proper subset of itself.

Proof. Let $n \in \omega$ be the smallest natural number for which there exists a proper subset $S \subset n$ and a bijection $f : n \to S$. The cases $n = 0$ and $n = 1$ are trivial, so $n \geq 2$. Let k be any element in $n \setminus S$. Since f is a bijection, there exists an element $j \in n$ such that $f(j) = k$. This contradicts the fact that the range of f is S. This proof seems too simple. Let us reconsider the structure.

Let $n \in \omega$ be minimal such that there exists $S \subset n$ and a bijection $f : n \to S$. Let $k \in S$. The restriction of f to $n \setminus \{f^{-1}(k)\}$ is a bijection to $S \setminus \{k\}$. By the previous lemma, $n \setminus \{f^{-1}(k)\} \approx n - 1$. The set $S \setminus \{k\}$ is a proper subset of $n - 1$, because if it were equal, then S would equal n, which is not the case. Therefore, we have found that $n - 1$ is equinumerous to a proper subset of itself, which contradicts the minimality of n. ■

This theorem immediately implies that the cardinality of a finite set is unique.

***Corollary* 7.6.1.** Every finite set is equinumerous to exactly one natural number.

***Proposition* 7.6.1.** A bijection from a finite set to itself is called a permutation. If $|A| = n$ for some $n \in \omega$, the number of distinct permutations of A is $n!$.

Proof. We proceed by induction on n.

- **Base Case ($n = 0$)**: If $A = \emptyset$, there is exactly one function $\emptyset \to \emptyset$, the empty function, which is a bijection. Since $0! = 1$, the statement holds.
- **Inductive Step**: Assume that for any set of size n, there are $n!$ permutations. Let A be a set with $|A| = n + 1$. Let us fix an element $a_0 \in A$. A permutation of A must map a_0 to some element $a \in A$. There are $n+1$ choices for this element a. Once the image of a_0 is fixed, say $f(a_0) = a$, the permutation must map the remaining n elements of $A \setminus \{a_0\}$ bijectively to the remaining n elements of $A \setminus \{a\}$. By the induction hypothesis, there are $n!$ ways to do this.

By the rule of product, the total number of permutations of A is $(n + 1) \cdot n! = (n + 1)!$. The result follows by induction. ■

Further consequences follow, including the Pigeonhole Principle.

***Corollary* 7.6.2.** *Pigeonhole Principle.* Let A and B be finite sets with $|A| > |B|$. There is no injective function from A to B.

Proof. Let $|A| = n$ and $|B| = m$ with $n, m \in \omega$ and $n > m$. Assume an injection $f : A \to B$ exists. Then A is equinumerous with the subset $f[A] \subseteq B$, so $|f[A]| = n$. However, $f[A]$ is a subset of B, so its cardinality cannot exceed that of B. This would imply $n \leq m$, a contradiction. ■

***Corollary* 7.6.3.** No finite set is equinumerous to a proper subset of itself.

This corollary's contrapositive provides the defining characteristic of an infinite set, often known as Dedekind-infinitude.

***Corollary* 7.6.4.** A set equinumerous to a proper subset of itself is infinite.

Example 7.6.1. The set of natural numbers, ω, is infinite. The function $f : \omega \to \omega \setminus \{0\}$ defined by $f(n) = n + 1$ is a bijection from ω to a proper subset of itself.

Finally, the properties of finiteness are downward-hereditary, while infiniteness is upward-hereditary.

***Corollary* 7.6.5.** Let $A \subseteq B$. If B is finite, then A is finite. If A is infinite, then B is infinite.

Countable Sets and the Aleph Hierarchy

The natural numbers, ω, constitute the smallest infinite ordinal and, as it is not equinumerous with any of its elements, it is also the smallest infinite cardinal. Cantor introduced the notation $\aleph_0$ (aleph-nought) for this fundamental cardinality. Sets that are no larger than ω are considered the "smallest" infinite sets and are given a special classification.

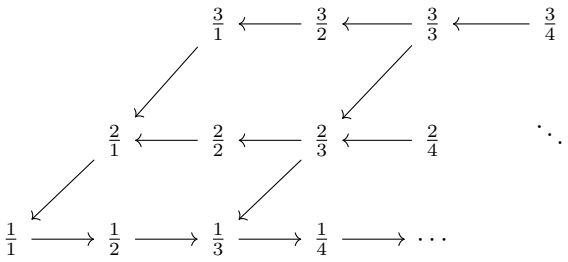
Definition 7.6.4. ***Countable Set***. A set A is countable if $A \preceq \omega$. If a set is countable but not finite, it is countably infinite.

A non-empty set is countable if its elements can be arranged in a list (finite or infinite), indexed by the natural numbers. For instance, the set of integers $\mathbb{Z}$ is countably infinite, as demonstrated by the bijection $g : \omega \to \mathbb{Z}$ defined by $g(n) = (n + 1)/2$ for odd n and $g(n) = -n/2$ for even n. More surprisingly, the set of rational numbers is also countable.

***Theorem* 7.6.3.** A subset of a countable set is countable.

Proof. Let A be a countable set and let $S \subseteq A$. By definition, there exists an injection $f : A \to \omega$. The restriction of this function, $f|_S : S \to \omega$, is also an injection. Therefore, $S \preceq \omega$, and S is countable. ■

Example 7.6.2. The set of rational numbers $\mathbb{Q}$ is countable. We can construct a bijection from ω to $\mathbb{Q}$ by systematically listing all rational numbers. Arrange the positive rationals p/q in an infinite grid where the rows are indexed by the numerator p and the columns by the denominator q. We can then traverse this grid along diagonals where $p + q$ is constant, starting with $p + q = 2$, then $p + q = 3$, and so on. By skipping any fraction that is not in its lowest terms (e.g., skipping $2/2$ as we have already listed $1/1$), we create an exhaustive list of all positive rational numbers. A similar process can be used for the negative rationals, and by interleaving these two lists with 0, we can list all of $\mathbb{Q}$.



The existence of a surjection from the natural numbers provides an equivalent characterisation of countability.

***Theorem* 7.6.4.** A non-empty set A is countable if and only if there exists a surjective function from ω onto A.

Proof. If A is countable, then $A \preceq \omega$. This means there is an injection $f : A \to \omega$. If A is infinite, f can be extended to a bijection, whose inverse is a surjection from ω to A. If A is finite, say $A = \{a_0, \ldots, a_{n-1}\}$, the function $g : \omega \to A$ defined by $g(k) = a_k$ for $k < n$ and $g(k) = a_{n-1}$ for $k \geq n$ is a surjection. Conversely, if a surjection $g : \omega \to A$ exists, then by the Axiom of Choice, there is an injection $f : A \to \omega$, so $A \preceq \omega$. ■

This criterion simplifies proofs about the closure of countability under certain operations.

***Theorem* 7.6.5.** The union of a countable family of countable sets is countable.

Proof. Let $\{\mathcal{A}_i\}_{i \in I}$ be a family of sets where the index set I is countable and each set $\mathcal{A}_i$ is countable. Since I is countable, there is a surjection $g : \omega \to I$. For each $i \in I$, since $\mathcal{A}_i$ is countable, there is a surjection $f_i : \omega \to \mathcal{A}_i$. We know that $\omega \times \omega \approx \omega$, so there exists a bijection $h : \omega \to \omega \times \omega$. We can define a function $F : \omega \times \omega \to \bigcup_{i \in I} \mathcal{A}_i$ by $F(n, m) = f_{g(n)}(m)$. This function is surjective. Let $x \in \bigcup_{i \in I} \mathcal{A}_i$. Then $x \in \mathcal{A}_i$ for some $i \in I$. Since g is surjective, there is an $n \in \omega$ with $g(n) = i$. Since f_i is surjective, there is an $m \in \omega$ with $f_i(m) = x$. Thus, $F(n, m) = x$. The composition $F \circ h : \omega \to \bigcup_{i \in I} \mathcal{A}_i$ is a surjection from ω to the union, which proves the union is countable. ■

***Theorem* 7.6.6.** The Cartesian product of a finite number of countable sets is countable.

Proof. We proceed by induction on the number of sets, n.

- **Base Case ($n = 2$)**: Let A_1 and A_2 be countable sets. There exist surjections $f_1 : \omega \to A_1$ and $f_2 : \omega \to A_2$. We can define a function $g : \omega \times \omega \to A_1 \times A_2$ by $g(m, n) = (f_1(m), f_2(n))$. This function is surjective. Since there exists a bijection $h : \omega \to \omega \times \omega$, the composition $g \circ h : \omega \to A_1 \times A_2$ is a surjection, which proves $A_1 \times A_2$ is countable.
- **Inductive Step**: Assume that for some $n \geq 2$, the product of n countable sets is countable. Consider a collection of $n + 1$ countable sets, $A_1, \ldots, A_{n+1}$. We can write their product as $A_1 \times \cdots \times A_{n+1} = (A_1 \times \cdots \times A_n) \times A_{n+1}$. Let $B = A_1 \times \cdots \times A_n$. By the induction hypothesis, B is countable. By the base case, the product $B \times A_{n+1}$ is countable.

Therefore, by the principle of induction, the theorem holds for any finite $n \geq 2$. The case $n = 1$ is trivial. ■

The existence of uncountable sets, such as $\mathbb{R}$, demonstrated by Cantor's diagonalization, implies the existence of infinite cardinals larger than $\aleph_0$. This leads to an infinite hierarchy of cardinal numbers, indexed by the ordinals. This sequence is defined by transfinite recursion.

Definition 7.6.5. ***The Aleph Sequence***. For any ordinal α, the cardinal $\aleph_\alpha$ is defined by the recursion:

$$\aleph_\alpha = \text{the least infinite cardinal not in } \{\aleph_\beta \mid \beta \in \alpha\}$$

This definition generates the sequence of infinite cardinals:

- $\aleph_0$ is the least infinite cardinal, which is ω.
- $\aleph_1$ is the least infinite cardinal strictly greater than $\aleph_0$.
- $\aleph_\omega$ is the least infinite cardinal strictly greater than all $\aleph_n$ for $n \in \omega$.

A fundamental result, which follows from the Well-Ordering Theorem, is that this sequence exhausts all possible infinite cardinalities.

***Theorem* 7.6.7.** Every infinite cardinal is equal to $\aleph_\alpha$ for some ordinal α.

Proof. Suppose there is an infinite cardinal κ that is not in the aleph sequence. By the Well-Ordering Principle applied to the class of infinite cardinals, there must be a least such κ. Consequently, every infinite cardinal smaller than κ is an aleph. The set of all infinite cardinals strictly smaller than κ is thus $S = \{\aleph_\beta \mid \aleph_\beta < \kappa\}$. The supremum of this set is κ. However, the definition of the aleph sequence for limit ordinals ensures that the supremum of a set of alephs is itself an aleph (specifically, $\sup S = \aleph_\gamma$ where $\gamma = \sup\{\beta \mid \aleph_\beta \in S\}$). Therefore, κ must be an aleph, which contradicts our initial assumption. ■

The aleph hierarchy gives rise to one of the most famous open questions in mathematics. Cantor conjectured that there are no cardinals between $\aleph_0$ and the cardinality of the real numbers, $|\mathbb{R}|$.

Definition 7.6.6. ***Continuum Hypothesis (CH)***. The Continuum Hypothesis is the statement $\aleph_1 = |\mathbb{R}|$.

It was later proven that CH is independent of the standard axioms of set theory (ZFC); it can be neither proved nor disproved within that system. A more general version extends this idea to the entire aleph hierarchy.

Definition 7.6.7. ***Generalised Continuum Hypothesis (GCH)***. The Generalised Continuum Hypothesis is the statement that for every ordinal α, $|\mathcal{P}(\aleph_\alpha)| = \aleph_{\alpha+1}$.

Finally, just as ordinals are classified as successor or limit, so too are the cardinals.

Definition 7.6.8. ***Successor and Limit Cardinals***. A cardinal κ is a successor cardinal if it is of the form $\aleph_{\alpha^+}$ for some ordinal α. Otherwise, it is a limit cardinal.

The cardinals $\aleph_1, \aleph_2, \ldots$ are successor cardinals. The cardinals $\aleph_0, \aleph_\omega, \aleph_{\omega+\omega}, \ldots$ are limit cardinals.

7.7 Exercises

Part I: Finite and Countable Sets

1. Let A and B be finite sets. Using the principle of induction or other established properties, prove that:
   (a) $|A \cup B| = |A| + |B| - |A \cap B|$
   (b) $|A \times B| = |A| \cdot |B|$

2. A set A is called *Dedekind-finite* if it is not equinumerous to any of its proper subsets. The text shows that all finite sets are Dedekind-finite. Prove the converse for a specific case: if R and R^{-1} are both well-orderings of a set A, prove that A must be finite.

3. Prove that the union and Cartesian product of any two countable sets are also countable.

Remark. For the product, you may use the fact that $\omega \times \omega \approx \omega$.

4. Let A be an infinite set. Prove that A can be partitioned into two disjoint infinite subsets, B and C, such that $A = B \cup C$.

Remark. Since A is infinite, it has a countably infinite subset. Partition this subset into two disjoint infinite sets.

5. Prove that the set of all algebraic numbers is countable. An algebraic number is a real number that is a root of a non-zero polynomial with integer coefficients.

Remark. First, argue that for any fixed degree n, the set of polynomials of degree n is countable. Then, use the fact that a countable union of countable sets is countable.

6. Use the result from the previous exercise to prove that the set of transcendental numbers (real numbers that are not algebraic) is uncountable.

Part II: Cardinal Arithmetic and the Aleph Hierarchy

7. Let α and β be ordinals. Prove that if $\alpha \in \beta$, then $\aleph_\alpha \prec \aleph_\beta$.

8. Prove that for any set of cardinals $\mathcal{F}$, their union $\bigcup \mathcal{F}$ is also a cardinal number.

9. Show that while removing a single element from an infinite set does not change its cardinality, the same is not true for ordinals. Let α be an infinite successor ordinal. Prove that $|\alpha| = |\alpha \setminus \{\beta\}|$ for any $\beta \in \alpha$. As a challenge prove this is true for any infinite set (or ordinal), not just successors.

10. Let A be a set. The cardinal $\beth(A)$ (Hartogs' number of A) is defined as the least ordinal α such that there is no injection from α into A.

(a) Prove that $\beth(A)$ is a cardinal number.
(b) Prove that for any cardinal κ, if $\kappa \preceq |A|$, then $\kappa \prec \beth(A)$.

11. Assuming the Generalised Continuum Hypothesis (GCH), what is the cardinality of $\mathcal{P}(\mathcal{P}(\mathcal{P}(\aleph_\omega)))$? Express your answer in the Aleph hierarchy.

12. The Beth sequence is defined by transfinite recursion: $\beth_0 = \aleph_0$, $\beth_{\alpha^+} = |\mathcal{P}(\beth_\alpha)|$, and $\beth_\lambda = \bigcup_{\gamma \in \lambda} \beth_\gamma$ for a limit ordinal λ. Restate the Generalised Continuum Hypothesis using the Aleph and Beth hierarchies.

13. Prove that for any countable set A, its power set $\mathcal{P}(A)$ is equinumerous with the set of real numbers $\mathbb{R}$.

14. Let A and B be sets. If there is a surjection $f : A \to B$, what relationship can you deduce between $|A|$ and $|B|$? Prove your assertion.

7.8 Arithmetic on Ordinals

The arithmetic operations on the natural numbers can be generalised to the class of all ordinals. As the class of all ordinals is not a set, these operations cannot be formalised as single functions. Instead, we define them by transfinite recursion, with the understanding that for any two ordinals α, β, their sum $\alpha + \beta$ is a uniquely defined ordinal, independent of any particular bounding set.

Ordinal Addition

Let α be an ordinal. We define the sum $\alpha + \beta$ for all ordinals β by transfinite recursion:

(i) $\alpha + 0 = \alpha$
(ii) $\alpha + \beta^+ = (\alpha + \beta)^+$ for any ordinal β.
(iii) $\alpha + \lambda = \bigcup_{\gamma \in \lambda}(\alpha + \gamma)$ for any limit ordinal λ.

By the Transfinite Recursion theorem, this uniquely defines the sum for any two ordinals. Intuitively, the ordinal $\alpha + \beta$ corresponds to the order type of a well-ordered set of type α followed by a well-ordered set of type β.

Ordinal addition is associative and has 0 as its identity element. However, it is not commutative. For example, let us compute $1 + \omega$ and $\omega + 1$.

- $1 + \omega = \bigcup_{n \in \omega}(1 + n)$. Since $1 + n$ is just the natural number $n + 1$, this union is $\{1, 2, 3, \dots\}$, which is equinumerous and order isomorphic to ω. Thus, $1 + \omega = \omega$.
- $\omega + 1 = \omega + 0^+ = (\omega + 0)^+ = \omega^+$.

Since $\omega \neq \omega^+$, we have $1 + \omega \neq \omega + 1$.

Ordinal Multiplication

Similarly, we define the product $\alpha \cdot \beta$ for a fixed ordinal α by transfinite recursion on β:

(i) $\alpha \cdot 0 = 0$
(ii) $\alpha \cdot \beta^+ = (\alpha \cdot \beta) + \alpha$ for any ordinal β.
(iii) $\alpha \cdot \lambda = \bigcup_{\gamma \in \lambda}(\alpha \cdot \gamma)$ for any limit ordinal λ.

Ordinal multiplication is associative, has 1 as its identity, and distributes over addition from the left. However, it is not commutative.

- $2 \cdot \omega = \bigcup_{n \in \omega}(2 \cdot n)$. Since $2n$ is a finite ordinal for each $n \in \omega$, their union is ω. Thus, $2 \cdot \omega = \omega$.
- $\omega \cdot 2 = \omega \cdot 1^+ = (\omega \cdot 1) + \omega = \omega + \omega$.

Since $\omega \neq \omega + \omega$, multiplication is not commutative. The right distributive law also fails in general.

***Theorem* 7.8.1.** For any ordinals α, β, γ:

(i) $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ (Associativity of $+$).
(ii) $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$ (Associativity of $\cdot$).
(iii) $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$ (Left Distributivity).

The operations interact with the ordinal ordering in a predictable way from the left, but less so from the right. For example, $1 < 2$ but $1 + \omega = 2 + \omega = \omega$.

***Theorem* 7.8.2.** Let α, β, γ be ordinals.

- $\alpha < \beta \Leftrightarrow \gamma + \alpha < \gamma + \beta$.
- $\alpha \leq \beta \Rightarrow \alpha + \gamma \leq \beta + \gamma$.
- For $\gamma > 0$, $\alpha < \beta \Leftrightarrow \gamma \cdot \alpha < \gamma \cdot \beta$.
- $\alpha \leq \beta \Rightarrow \alpha \cdot \gamma \leq \beta \cdot \gamma$.

Ordinal Exponentiation

Finally, we define exponentiation α^β by transfinite recursion on the exponent β:

(i) $\alpha^0 = 1$
(ii) $\alpha^{\beta^+} = \alpha^\beta \cdot \alpha$ for any ordinal β.
(iii) $\alpha^\lambda = \bigcup_{\gamma\in\lambda} \alpha^\gamma$ for any limit ordinal $\lambda > 0$.

This definition preserves the standard laws of exponents.

***Theorem* 7.8.3.** Let α, β, γ be ordinals.

- $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$.
- $(\alpha^\beta)^\gamma = \alpha^{\beta\cdot\gamma}$.

As with the other operations, exponentiation is not commutative. For example, $2^\omega = \bigcup_{n\in\omega} 2^n = \omega$, whereas $\omega^2 = \omega \cdot \omega = \omega + \omega + \ldots$ is a much larger ordinal.

Cardinal Arithmetic

While every finite cardinal is also a finite ordinal, their arithmetic operations are generalised to the infinite case in fundamentally different ways. Ordinal arithmetic is recursive and sensitive to order, reflecting the concatenation of well-ordered sets. Cardinal arithmetic, in contrast, is defined directly in terms of set operations and reflects the size of the resulting sets, irrespective of order.

Definition 7.8.1. ***Cardinal Arithmetic Operations***. Let κ and λ be cardinal numbers.

(i) **Addition**: $\kappa + \lambda := |(\kappa \times \{0\}) \cup (\lambda \times \{1\})|$. This represents the cardinality of the disjoint union of two sets with cardinalities κ and λ.
(ii) **Multiplication**: $\kappa \cdot \lambda := |\kappa \times \lambda|$. This is the cardinality of the Cartesian product.
(iii) **Exponentiation**: $\kappa^\lambda := |\kappa^\lambda|$. This is the cardinality of the set of all functions from a set of size λ to a set of size κ.

These definitions are consistent with arithmetic on the natural numbers. For example, $2 + 3 = |(\{0, 1\} \times \{0\}) \cup (\{0, 1, 2\} \times \{1\})| = |\{(0, 0), (1, 0), (0, 1), (1, 1), (2, 1)\}| = 5$. Unlike their ordinal counterparts, these operations possess the familiar properties of commutativity and associativity.

***Theorem* 7.8.4.** For any cardinals κ, λ, μ:

(i) $\kappa + \lambda = \lambda + \kappa$ and $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$. The identity is 0.
(ii) $\kappa \cdot \lambda = \lambda \cdot \kappa$ and $(\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$. The identity is 1.
(iii) $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$ (Distributivity).

Proof. These properties follow directly from the existence of bijections between the corresponding set constructions. For example, commutativity of addition holds because there is a simple bijection between $(\kappa \times \{0\}) \cup (\lambda \times \{1\})$ and $(\lambda \times \{0\}) \cup (\kappa \times \{1\})$. ■

Cardinal exponentiation also satisfies the standard laws of exponents.

***Theorem* 7.8.5.** For any cardinals κ, λ, μ:

(i) $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$.
(ii) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$.
(iii) $(\kappa^\lambda)^\mu = \kappa^{\lambda\cdot\mu}$.

Proof. Each identity is proven by constructing a bijection between the corresponding sets of functions. For (iii), we establish a bijection between the set of functions $(\kappa^\lambda)^\mu$ and the set $\kappa^{\lambda\times\mu}$. ■

The arithmetic of infinite cardinals is governed by a powerful absorption property.

***Theorem* 7.8.6.** For any infinite cardinal κ and any cardinal λ such that $\lambda \leq \kappa$,

$$\kappa + \lambda = \kappa \quad \text{and} \quad \kappa \cdot \lambda = \kappa \quad (\text{for } \lambda > 0)$$

A particularly important consequence of this is that for any infinite cardinal κ, $\kappa + \kappa = \kappa$ and $\kappa \cdot \kappa = \kappa$. This simplifies the arithmetic of alephs significantly.

***Theorem* 7.8.7.** For any infinite cardinal κ, $\kappa \cdot \kappa = \kappa$.

Proof. We prove this by transfinite induction on κ. The base case $\aleph_0 \cdot \aleph_0 = \aleph_0$ is known. Assume the property holds for all infinite cardinals $\lambda < \kappa$.Define a well-ordering $\lhd$ on the set of pairs $\kappa \times \kappa$ (called the canonical well-ordering) by:

$$(\alpha, \beta) \lhd (\gamma, \delta) \iff \max(\alpha, \beta) < \max(\gamma, \delta) \vee (\dots)$$

It can be shown that for any infinite cardinal κ, the order type of $(\kappa \times \kappa, \lhd)$ is exactly κ. Since the set $\kappa \times \kappa$ has cardinality $\kappa \cdot \kappa$, and its order type is κ, it follows that $\kappa \cdot \kappa = \kappa$. ■

***Theorem* 7.8.8.** For any ordinals α and β,

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_{\max\{\alpha,\beta\}}$$

Proof. Let $\kappa = \aleph_{\max\{\alpha,\beta\}}$. Then $\aleph_\alpha \leq \kappa$ and $\aleph_\beta \leq \kappa$.The sum $\aleph_\alpha + \aleph_\beta$ is at least κ and at most $\kappa + \kappa = 2 \cdot \kappa \leq \kappa \cdot \kappa = \kappa$. By the Cantor–Schröder–Bernstein theorem, $\aleph_\alpha + \aleph_\beta = \kappa$.The product $\aleph_\alpha \cdot \aleph_\beta$ is at least κ and at most $\kappa \cdot \kappa = \kappa$. By the same theorem, $\aleph_\alpha \cdot \aleph_\beta = \kappa$. ■

For example, $\aleph_5 + \aleph_9 = \aleph_9$ and $\aleph_0 \cdot \aleph_\omega = \aleph_\omega$. This simple arithmetic stands in stark contrast to the complexity of ordinal arithmetic.

7.9 Exercises

Part I: Ordinal Arithmetic

1. Using the recursive definitions, compute the following ordinal sums and products.

(a) $(\omega + 1) + 2$

(b) $2 \cdot (\omega + 1)$

(c) $(\omega + 1) \cdot 2$

(d) 2^ω

2. Prove the associativity of ordinal addition: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ for all ordinals α, β, γ.

Remark. Fix α and β and proceed by transfinite induction on γ.

3. The text states that the right distributive law fails for ordinals. Provide a specific counterexample by computing $(\omega + 1) \cdot \omega$ and $(\omega \cdot \omega) + (1 \cdot \omega)$ and showing they are not equal.

4. Prove the left cancellation law for ordinal addition: For any ordinals α, β, γ, if $\gamma + \alpha = \gamma + \beta$, then $\alpha = \beta$.

Remark. Use the property that $\alpha < \beta \Leftrightarrow \gamma + \alpha < \gamma + \beta$ (proved in the text) along with the trichotomy law. If $\alpha \neq \beta$, say $\alpha < \beta$, what does this imply about the sums?

5. Show by counterexample that the right cancellation law for ordinal addition does not hold. That is, find ordinals α, β, γ such that $\alpha \neq \beta$ but $\alpha + \gamma = \beta + \gamma$.

6. Find ordinals α, β, γ such that $\alpha \in \beta$ but $\alpha \cdot \gamma = \beta \cdot \gamma$, demonstrating that right cancellation for multiplication also fails.

Part II: Cardinal Arithmetic

7. Let κ, λ, μ be cardinals. Prove the associativity of cardinal addition, $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$, by constructing an explicit bijection between the corresponding sets.

8. Prove the law of exponents $(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}$ by constructing an explicit bijection between the set of functions $(\kappa^\lambda)^\mu$ and the set $\kappa^{\lambda \times \mu}$.

Remark. A function $f \in (\kappa^\lambda)^\mu$ is a function $f : \mu \to \kappa^\lambda$. For an element $z \in \mu$, $f(z)$ is itself a function from λ to κ. Use this to define a function whose domain is $\lambda \times \mu$.

9. Using the absorption properties of infinite cardinals, compute the following:

(a) $\aleph_{17} + \aleph_{32}$

(b) $\aleph_0 \cdot \aleph_\omega$

(c) $\aleph_{\omega+5} \cdot \aleph_{\omega+2}$

(d) $|\mathbb{R}| + \aleph_1$ (assuming CH is false and $|\mathbb{R}| > \aleph_1$)

10. Let n be a finite cardinal ($n \in \omega, n > 0$) and let κ be an infinite cardinal. Prove from the definition of cardinal multiplication that $n \cdot \kappa = \kappa$.

Part III: Comparing Ordinal and Cardinal Arithmetic

11. Prove that for any natural number $n \geq 2$, the ordinals $n + \omega$ and $n \cdot \omega$ are not equal, but the cardinals $|n + \omega|$ and $|n \cdot \omega|$ are equal.

12. Prove that the ordinal $\omega + \omega$ is not a cardinal number.

13. Does a general cancellation law for multiplication hold for cardinals? That is, if $\kappa \cdot \mu = \lambda \cdot \mu$ and $\mu \neq 0$, does it follow that $\kappa = \lambda$? Justify your answer.

14. Prove that for any infinite cardinal κ, the set of all finite sequences of elements from κ has cardinality κ.

Remark. The set of all such sequences can be written as $\bigcup_{n \in \omega} \kappa^n$. Apply the properties of cardinal arithmetic.

15. $\star$ The proof that $\kappa \cdot \kappa = \kappa$ for any infinite cardinal κ is a cornerstone of cardinal arithmetic. While the full proof is complex, it can be outlined for the specific case $\kappa = \aleph_0$. Provide an explicit bijection to prove that $|\omega \times \omega| = |\omega|$, thereby showing $\aleph_0 \cdot \aleph_0 = \aleph_0$.

7.10 Large Cardinals

The arithmetic of infinite cardinals simplifies to an absorption rule where the larger cardinal dominates. This simplicity masks a rich underlying structure, which can be investigated by considering how a cardinal can be constructed as a limit of smaller ordinals. This leads to the concept of cofinality, a measure that distinguishes different "types" of limit ordinals and cardinals, and ultimately points towards cardinalities whose existence cannot be proven within the standard framework of set theory.

Cofinality

Every infinite cardinal is a limit ordinal, meaning it is the union of all smaller ordinals it contains: $\kappa = \bigcup \kappa$. However, it is often possible to express κ as a union of a much smaller collection of ordinals. The cofinality of a limit ordinal measures the smallest possible size of such a collection.

Definition 7.10.1. ***Cofinality***. Let λ be a limit ordinal. A subset $\mathcal{F} \subseteq \lambda$ is cofinal in λ if $\bigcup \mathcal{F} = \lambda$. The cofinality of λ, denoted $\text{cf}(\lambda)$, is the least cardinal κ such that there exists a cofinal subset $\mathcal{F} \subseteq \lambda$ with $|\mathcal{F}| = \kappa$.

By definition, $\text{cf}(\lambda) \preceq |\lambda|$, since λ is itself a cofinal subset of λ.

Example 7.10.1.

- The limit ordinal ω can be expressed as the union of the finite ordinals it contains, $\omega = \bigcup_{n\in\omega} n$. The index set $\{n \mid n \in \omega\}$ has cardinality $\aleph_0$. No smaller index set can have a union equal to ω, as the union of a finite number of finite ordinals is finite. Therefore, $\text{cf}(\omega) = \aleph_0$.
- Consider the limit cardinal $\aleph_\omega$. It can be expressed as the union of the preceding alephs: $\aleph_\omega = \bigcup_{n\in\omega} \aleph_n$. The index set has cardinality $\aleph_0$. Thus, $\text{cf}(\aleph_\omega) \leq \aleph_0$. Since the cofinality of any infinite ordinal must be infinite, we conclude $\text{cf}(\aleph_\omega) = \aleph_0$.
- For the cardinal $\aleph_1$, its cofinality must be $\aleph_1$. If $\mathcal{F}$ were a cofinal subset of $\aleph_1$ with $|\mathcal{F}| = \aleph_0$, then $\aleph_1 = \bigcup \mathcal{F}$. Each element of $\mathcal{F}$ is an ordinal smaller than $\aleph_1$, meaning each element is a countable ordinal. The union of a countable collection of countable sets is countable. This would imply $\aleph_1$ is countable, a contradiction. Therefore, $\text{cf}(\aleph_1) = \aleph_1$.

Regular and Singular Cardinals

The cofinality of an infinite cardinal provides a fundamental classification.

Definition 7.10.2. ***Regular and Singular Cardinals***. An infinite cardinal κ is regular if $\text{cf}(\kappa) = \kappa$. Otherwise, if $\text{cf}(\kappa) < \kappa$, it is singular.

From the preceding examples, $\aleph_0$ and $\aleph_1$ are regular cardinals, whereas $\aleph_\omega$ is a singular cardinal. The following theorems further clarify this classification.

***Theorem* 7.10.1.** Every successor cardinal is regular.

Proof. Let $\aleph_{\alpha+1}$ be a successor cardinal. Let $\mathcal{F}$ be a cofinal subset of $\aleph_{\alpha+1}$, so $\aleph_{\alpha+1} = \bigcup \mathcal{F}$. For any ordinal $\beta \in \mathcal{F}$, we must have $\beta < \aleph_{\alpha+1}$, which implies $|\beta| \leq \aleph_\alpha$. Using properties of cardinal arithmetic, we have:

$$\aleph_{\alpha+1} = \left|\bigcup \mathcal{F}\right| \preceq \sum_{\beta\in\mathcal{F}} |\beta| \preceq |\mathcal{F}| \cdot \sup_{\beta\in\mathcal{F}} |\beta| \preceq |\mathcal{F}| \cdot \aleph_\alpha$$

If $|\mathcal{F}| \leq \aleph_\alpha$, this would imply $\aleph_{\alpha+1} \preceq \aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$, a contradiction. Therefore, we must have $|\mathcal{F}| > \aleph_\alpha$. Since $|\mathcal{F}|$ is a cardinal and $\aleph_{\alpha+1}$ is the immediate successor of $\aleph_\alpha$, we must have $|\mathcal{F}| \geq \aleph_{\alpha+1}$. As $|\mathcal{F}|$ can be $\aleph_{\alpha+1}$ (by taking $\mathcal{F} = \aleph_{\alpha+1}$), the least possible cardinality for a cofinal subset is $\aleph_{\alpha+1}$. Thus, $\text{cf}(\aleph_{\alpha+1}) = \aleph_{\alpha+1}$, and $\aleph_{\alpha+1}$ is regular. ■

Since $\aleph_0$ is also regular (as a successor cardinal if we consider the finite cardinals), it follows that any singular cardinal must be an uncountable limit cardinal. The next theorems provide a deeper connection between the cofinality of a limit ordinal and its corresponding aleph.

***Theorem* 7.10.2.** If α is a limit ordinal, then $\text{cf}(\aleph_\alpha) = \text{cf}(\alpha)$.

This theorem confirms our earlier calculation that $\text{cf}(\aleph_\omega) = \text{cf}(\omega) = \aleph_0$. It also implies that for any limit ordinal α, its cofinality must be a regular cardinal.

***Theorem* 7.10.3.** For any limit ordinal α, $\text{cf}(\alpha)$ is a regular cardinal.

König's Theorem and Inaccessible Cardinals

A powerful result due to Julius König places a strict bound on cardinal exponentiation, connecting it to cofinality. Its proof is a generalisation of Cantor's diagonal argument.

***Theorem* 7.10.4. König's Theorem**. For any infinite cardinal κ, $\kappa < \text{cf}(2^\kappa)$.

An immediate and significant consequence is a non-trivial fact about the cofinality of the continuum.

***Corollary* 7.10.1.** The cofinality of the continuum, $\text{cf}(2^{\aleph_0})$, is uncountable.

Proof. By Theorem 7.10.4, we have $\aleph_0 < \text{cf}(2^{\aleph_0})$. Since $\aleph_1$ is the first uncountable cardinal, this implies $\text{cf}(2^{\aleph_0}) \geq \aleph_1$. ■

This result allows us to rule out certain values for the cardinality of the continuum. For example, it proves that $2^{\aleph_0} \neq \aleph_\omega$, because $\text{cf}(\aleph_\omega) = \aleph_0$, which contradicts the corollary.

The study of cofinality reveals that $\aleph_0$ is the only regular limit cardinal whose existence is provable in ZFC. The potential existence of others leads to the notion of large cardinals.

Definition 7.10.3. ***Inaccessible Cardinal***. An uncountable cardinal κ is weakly inaccessible if it is a regular limit cardinal. It is strongly inaccessible if it is regular and for every cardinal $\lambda < \kappa$, we have $2^\lambda < \kappa$.

Every strongly inaccessible cardinal is also weakly inaccessible. The existence of either type of inaccessible cardinal cannot be proven from the axioms of ZFC. These cardinals mark the beginning of the large cardinal hierarchy, a sequence of increasingly strong axioms asserting the existence of cardinals with properties that make them vastly larger than those constructible within ZFC alone. The consistency of these axioms is a central topic in modern set theory.

7.11 Exercises

1. Compute the cofinality of the following limit ordinals and state whether the corresponding cardinal is regular or singular.
   (a) $\omega \cdot 2$ (also written as $\omega + \omega$)
   (b) $\aleph_{\omega+\omega}$
   (c) $\aleph_{\aleph_1}$

2. Prove that for any limit ordinal λ, its cofinality, $\text{cf}(\lambda)$, must be a regular cardinal.

   Remark. Let $\kappa = \text{cf}(\lambda)$ and suppose, for a contradiction, that κ is singular. Then κ can be written as a union of a smaller number of smaller ordinals. Use this to show that λ could also be expressed as a union of a set of ordinals with cardinality less than κ, contradicting the definition of cofinality.

3. A cardinal κ is said to have the *countable cofinality property* if $\text{cf}(\kappa) = \aleph_0$. Use König's theorem to prove that the cardinality of the continuum, $2^{\aleph_0}$, cannot be equal to $\aleph_\omega$.

   Remark. Recall that $\text{cf}(\aleph_\omega) = \aleph_0$. What does König's theorem say about the cofinality of $2^{\aleph_0}$?

4. The Generalised Continuum Hypothesis (GCH) states that $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for every ordinal α. Prove that GCH implies that every uncountable regular limit cardinal is strongly inaccessible.

5. ⋆ Let κ be an infinite regular cardinal. Prove that if $\{A_i\}_{i \in I}$ is a family of sets such that $|I| < \kappa$ and $|A_i| < \kappa$ for all $i \in I$, then the cardinality of their union is also less than κ. That is, $|\bigcup_{i \in I} A_i| < \kappa$.

   Remark. Show that the cardinality of the union is bounded by $|I| \cdot \sup_{i \in I} |A_i|$. Since κ is regular and the number of sets $|I|$ is less than κ, the supremum of their cardinalities must also be less than κ.