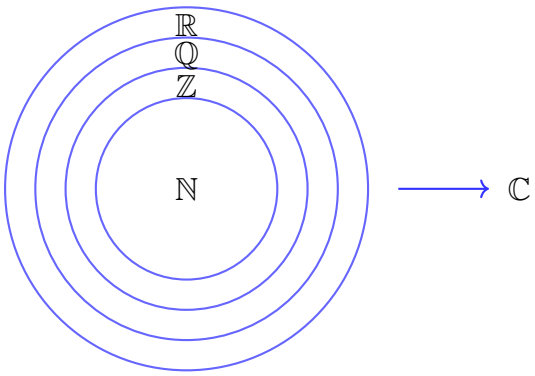
FOUNDATIONS OF MATHEMATICS

A Rigorous Development in the Style of Bourbaki



The Collins Compendium
Formal Edition

Collins Mwangi

December 14, 2025

"Mathematics is the art of giving the same name to different things"
— Henri Poincaré

Foundations of Mathematics

A Rigorous Development in the Style of Bourbaki

Collins Mwangi

The Collins Compendium — Formal Edition

December 14, 2025

Foundations of Mathematics: A Rigorous Development in the Style of Bourbaki

The Collins Compendium — Formal Edition

First Edition: December 14, 2025

Typeset in LaTeX

Contact: For permissions or inquiries, please contact the author.

Dedicated to all who seek truth through rigorous reasoning.

To my family, teachers, and the mathematical giants upon whose shoulders we stand.

"God made the integers; all else is the work of man."
— Leopold Kronecker

Preface

This work presents a rigorous, axiomatic treatment of the foundations of mathematics, following the spirit of Nicolas Bourbaki's *Éléments de mathématique*. However, unlike Bourbaki's austere presentation, we take a *pedagogical approach*: every concept is introduced gradually, building on what came before.

Philosophy: Mathematics is not discovered; it is constructed. Every theorem flows from axioms through rigorous logical deduction. But before the formalism, we provide *intuition*. Before the proof, we explain the *key idea*. Before the definition, we motivate *why we need it*.

Structure: This book is designed to be read linearly. Each chapter builds on previous foundations:

1. **Foundations**: We start informally, explaining what mathematics *is* and why we need axioms
2. **Formal Logic**: The simplest formal system—propositional and predicate logic with complete proofs
3. **Axiomatic Set Theory**: Building the universe of mathematics from the ZFC axioms
4. **Arithmetic**: Recursive operations on natural numbers and construction of integers and rationals
5. **Relations**: Ordered pairs, Cartesian products, equivalence relations, and partial orders
6. **Functions**: The morphisms of mathematics—injections, surjections, bijections, and composition
7. **Cardinality**: Measuring infinite sets—Cantor's diagonal argument and the hierarchy of infinities
8. **The Real Numbers**: Dedekind cuts, completeness, and filling the gaps in the rationals
9. **Sequences and Convergence**: Limits, Cauchy sequences, and the foundation of analysis
10. **Continuity**: The ϵ-δ definition, IVT, EVT, and uniform continuity
11. **Differentiation**: Derivatives as limits, MVT, and the algebra of differentiation

12. **Integration**: Riemann sums, the Fundamental Theorem, and applications to area and volume

Each concept follows this pattern:

Intuition → Motivation → Formal Definition → Examples → Theorems → Proofs

We use color-coded boxes to guide your reading:

- **Green boxes**: Informal intuition before formal definitions
- **Yellow boxes**: Key ideas before complex proofs
- **Blue boxes**: Technical remarks and connections
- **Red boxes**: Common pitfalls and misconceptions
- **Gray boxes**: Historical context and motivation

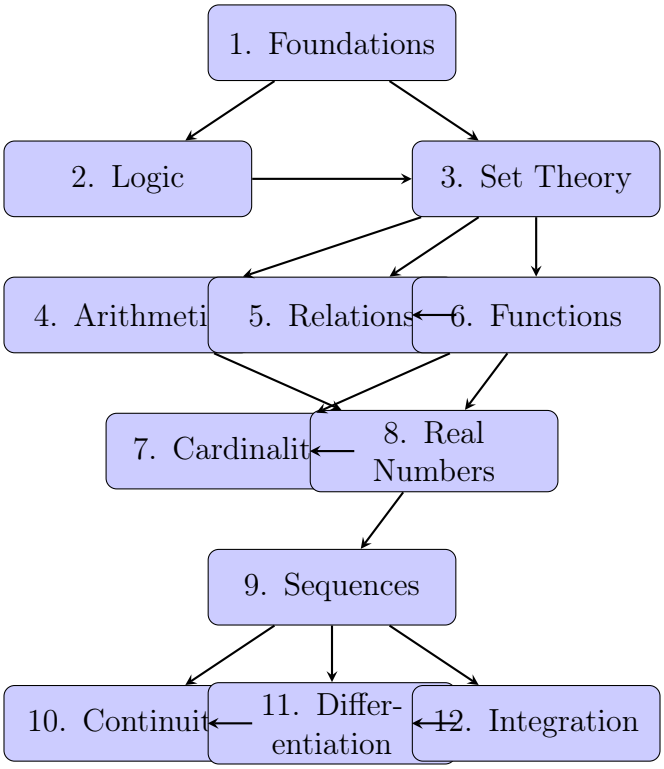
Prerequisites: Curiosity and patience. We assume no prior formal training, but we do not compromise on rigor. Every step is justified. Every proof is complete.

Regarding Historical Notes: Throughout this book, historical context appears within chapters, not at the end. This is intentional—understanding the historical development of concepts provides crucial insight into *why* definitions are structured as they are and what problems they solve. Mathematics is not timeless abstraction; it evolved through human struggle with deep questions.

"In mathematics, we never know what we are talking about, nor whether what we are saying is true."
— Bertrand Russell

Chapter Dependencies

The following diagram shows the logical dependencies between chapters. An arrow from A to B means that chapter A must be understood before chapter B.



Reading suggestion: Follow the order presented. Each chapter assumes complete understanding of all chapters it depends upon.

Notation and Symbols

Logic and Set Theory:

$\neg$	Logical negation (NOT)
$\wedge$	Logical conjunction (AND)
$\vee$	Logical disjunction (OR)
$\Longrightarrow$	Logical implication (IF...THEN)
$\Longleftrightarrow$	Logical bi-conditional (IF AND ONLY IF)
$\forall$	Universal quantifier (FOR ALL)
$\exists$	Existential quantifier (THERE EXISTS)
$\in$	Set membership (IS AN ELEMENT OF)
$\notin$	Not a member of
$\subseteq$	Subset or equal
$\subset$	Proper subset
$\emptyset$	Empty set
$\cup$	Set union
$\cap$	Set intersection
$\setminus$	Set difference
A^c	Complement of set A
$\mathcal{P}(A)$	Power set (set of all subsets of A)
$A \times B$	Cartesian product
$\lvert A \rvert$	Cardinality (size) of set A

Number Systems:

$\mathbb{N}$	Natural numbers $\{0, 1, 2, 3, \ldots\}$
$\mathbb{Z}$	Integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{Q}$	Rational numbers
$\mathbb{R}$	Real numbers
$\mathbb{C}$	Complex numbers (not covered in this volume)
$\aleph_0$	Cardinality of natural numbers (“aleph-null”)

Relations and Functions:

$f : A \to B$	Function from A to B
$f(x)$	Value of function f at x
$f \circ g$	Composition of functions
f^{-1}	Inverse function
$\text{dom}(f)$	Domain of function
$\text{Im}(f)$	Image (range) of function
$[x]$	Equivalence class of x
$A/\sim$	Quotient set (set of equivalence classes)
$\leq, <$	Less than or equal, strictly less than
$\geq, >$	Greater than or equal, strictly greater than

Analysis:

$\lvert x \rvert$	Absolute value
$\lim_{n\to\infty} a_n$	Limit of sequence (a_n)
$\lim_{x\to c} f(x)$	Limit of function at c
$f'(x)$	Derivative of f
$\frac{df}{dx}$	Derivative (Leibniz notation)
$\int_a^b f(x)\,dx$	Definite integral from a to b
ϵ	Small positive number (epsilon)
δ	Small positive number (delta)
$\sup A$	Supremum (least upper bound)
$\inf A$	Infimum (greatest lower bound)

Proof Terminology:

QED	"Quod erat demonstrandum" (which was to be demonstrated)
$\square$	End of proof marker
✓	Verified/confirmed
IH	Inductive Hypothesis

Contents

Chapter 1

Foundations: What Is Mathematics?

1.1 The Building Blocks of Thought

Intuition

Before we can do mathematics, we must answer a fundamental question: *What is mathematics?*
Is it the study of numbers? Shapes? Patterns? While mathematics involves all of these, its true nature is more abstract. Mathematics is the art of **reasoning with perfect precision** about abstract objects.
This chapter starts informally. We'll use everyday language to explain what we're building toward. Think of this as the *construction site* before the building exists.

1.1.1 A Simple Example: Counting

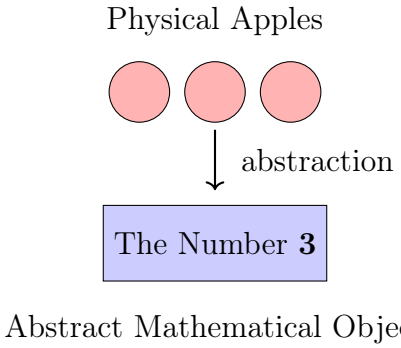
Let's start with something familiar: counting. When you count objects—say, apples—you say "1, 2, 3, 4, 5." This seems natural. But what are you *actually* doing?

1. You're matching each apple to a word: "one," "two," "three," etc.

2. You're following rules: don't skip numbers, don't count the same apple twice

3. You're using symbols: the marks "1", "2", "3" represent abstract concepts

Here's the key insight: **the numbers themselves don't exist in physical reality**. You can't touch "five." You can touch five apples, but "five-ness" itself is abstract. Mathematics studies these abstractions.



1.1.2 The Need for Precision

In everyday life, we can be vague. If I say "Meet me around 3 PM," you understand I might arrive at 2:58 or 3:05. But in mathematics, we can't be vague.

Example 1.1 (Vague vs. Precise)**.** ***Vague statement:*** *"Large numbers grow quickly."*

Questions this raises:

- *What counts as "large"? 100? 1,000,000?*
- *What does "quickly" mean? Compared to what?*
- *What does "grow" mean? As we add? Multiply?*

Precise statement: *"For any real number $M > 0$, there exists a natural number N such that $2^n > M$ for all $n > N$."*

Now there's no ambiguity. Every word has exact meaning.

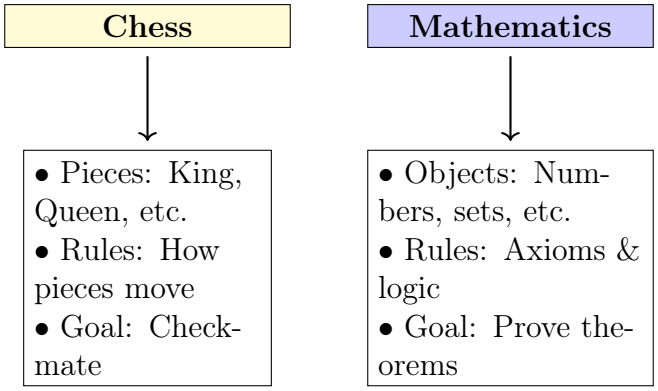
Key Idea

Mathematics requires a language with **zero ambiguity**. Every statement must have exactly one interpretation. This is why we need formal systems.

1.2 The Game of Mathematics

1.2.1 Mathematics as a Game

Think of mathematics as a sophisticated game, like chess:



In chess:

- **Pieces** are the objects you manipulate
- **Rules** say what moves are legal
- **Strategy** is how you win, but it's not part of the rules

In mathematics:

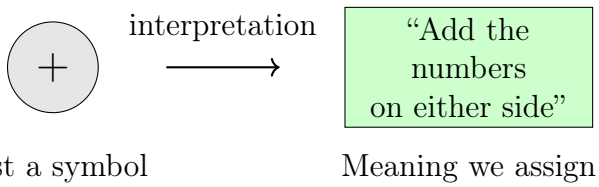
- **Mathematical objects** are what we study (numbers, shapes, functions, etc.)
- **Axioms and logic** are the rules
- **Theorems** are the "wins"—statements we prove are true

1.2.2 What Are Symbols?

When you see "$2 + 3 = 5$", what is it? It's a string of symbols:

$$2 \quad + \quad 3 \quad = \quad 5$$

Each symbol is just a mark on paper (or pixels on a screen). The symbols themselves have no inherent meaning. We *give them meaning* by agreeing on what they represent.



Warning

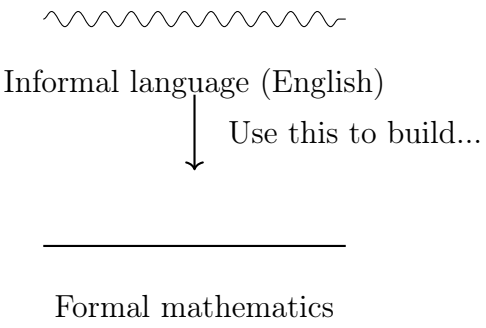
Symbols are NOT the same as their meanings. The symbol "+" is different from the concept of addition. We could have used any symbol—say "⊕" or "&"—as long as we agreed on the meaning.
This distinction between **syntax** (symbols) and **semantics** (meaning) is crucial for understanding formal mathematics.

1.3 Building Mathematics from Scratch

1.3.1 The Bootstrap Problem

We face a philosophical problem: to explain mathematics, we're using language (English). But language itself is imprecise. We're trying to build something precise using imprecise tools!

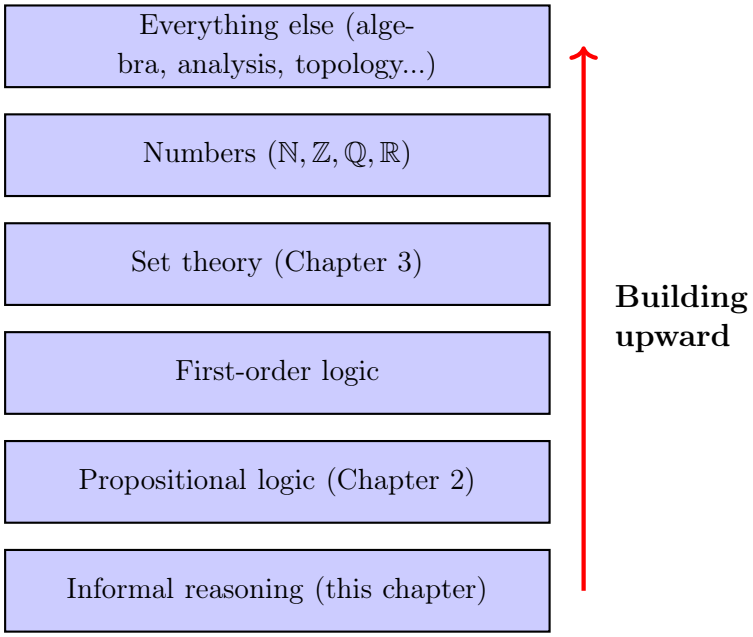
This is like trying to build a perfectly straight ruler when you don't have any straight edge to measure against.



The solution is to be *as clear as possible* in our informal explanations, and then formalize carefully. We'll use everyday language to explain concepts, but we'll mark the transition when we go formal.

1.3.2 The Ladder of Abstraction

Mathematics is built in layers. Each layer assumes the one below it:



Intuition

You can't understand calculus without algebra. You can't understand algebra without numbers. You can't understand numbers without set theory. You can't understand set theory without logic.
This book starts at the bottom and works up. Be patient. The early chapters might seem overly detailed, but they're the foundation for everything that follows.

1.4 Strings and Expressions

Let's start building. The most basic concept is a **string of symbols**.

1.4.1 Alphabets

Definition 1.1 (Alphabet—Informal)**.** *An **alphabet** is a collection of symbols. These are our basic building blocks.*

Example 1.2 (Familiar Alphabets)**.**

1. ***English alphabet:*** $\{a, b, c, \ldots, z, A, B, C, \ldots, Z\}$
2. ***Decimal digits:*** $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
3. ***Binary:*** $\{0, 1\}$
4. ***Mathematical symbols:*** $\{+, -, \times, \div, =, (,), x, y, \ldots\}$

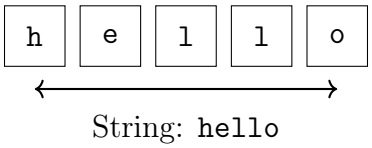
Think of an alphabet as your box of Scrabble tiles. You have a fixed set of letters (symbols) to work with.

1.4.2 Strings

Definition 1.2 (String—Informal)**.** *A **string** (or **expression**) is a sequence of symbols from an alphabet, written one after another.*

Example 1.3 (Strings)**.** *Using the English alphabet:*

- `cat` *is a string of length 3*
- `hello` *is a string of length 5*
- `xyz` *is a string of length 3*
- `aaa` *is a string of length 3 (repetition is allowed)*
- *The empty string (no symbols) is also a string, written* ε



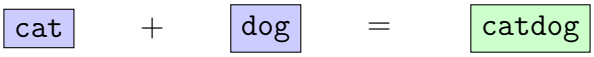
Key point: A string is just symbols in order. It doesn't have to "mean" anything. The string `xqzp` is perfectly valid, even though it's not an English word.

1.4.3 Concatenation

Definition 1.3 (Concatenation—Informal)**.** ***Concatenation** means sticking two strings together, end-to-end.*

Example 1.4.

- `cat` *concatenated with* `dog` *gives* `catdog`
- `hello` *concatenated with* `world` *gives* `helloworld`
- `123` *concatenated with* `456` *gives* `123456` *(note: this is NOT addition!)*



Remark

Concatenation has some nice properties:

- **Associative**: $(s \cdot t) \cdot u = s \cdot (t \cdot u)$
- **Identity**: Concatenating with the empty string does nothing: $s \cdot \varepsilon = \varepsilon \cdot s = s$

- **NOT commutative**: Generally $s \cdot t \neq t \cdot s$ (e.g., `catdog` $\neq$ `dogcat`)

1.5 Meaning vs. Form

1.5.1 Syntax and Semantics

Now we reach a crucial distinction that underpins all of formal mathematics:

Definition 1.4 (Syntax—Informal)**.** ***Syntax*** *is the study of form—what strings of symbols look like, how they're constructed, which ones are "well-formed."*

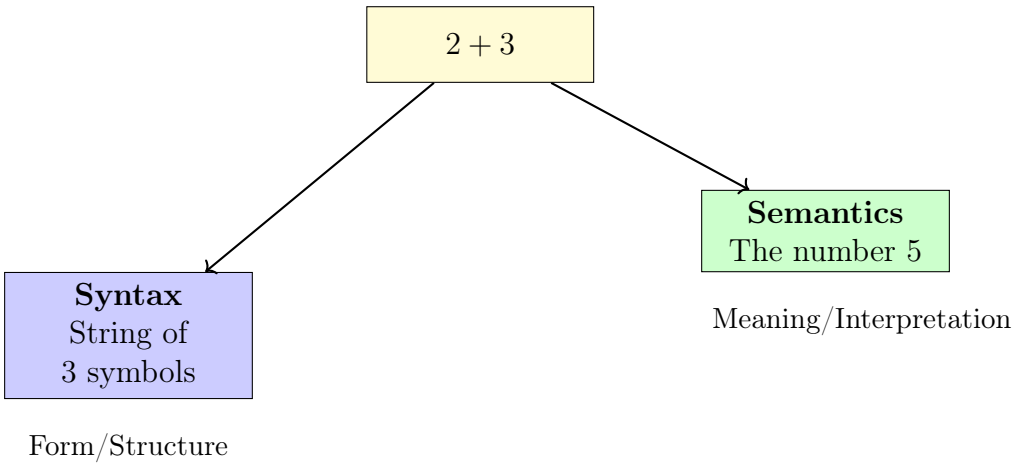
Definition 1.5 (Semantics—Informal)**.** ***Semantics*** *is the study of meaning—what strings of symbols represent, what they "say."*

Example 1.5 (Syntax vs. Semantics)**.** *Consider the expression:* $2 + 3$

Syntactically: *This is a string of 5 symbols: "2", "", "+", "", "3" (Yes, the spaces are symbols too!)*

Semantically: *Under the usual interpretation, this represents the sum of two and three, which equals five.*

But we could define a different interpretation where "+" means multiplication. Then $2 + 3$ *would mean* $2 \times 3 = 6$.



Key Idea

In formal mathematics:

1. First, we define syntax: which strings are valid
2. Then, we define semantics: what those strings mean

This separation is crucial. We can manipulate symbols without knowing what they mean (syntax), and then interpret the results (semantics).

1.5.2 Why Separate Syntax and Semantics?

You might wonder: why bother separating form from meaning? Can't we just work with meaningful statements?

The answer is: separating them gives us **mechanical rules**. A computer can check syntax without understanding meaning.

Example 1.6 (Spell Check)**.** *A spell checker operates purely syntactically:*

- *It checks if words are in its dictionary (valid strings)*
- *It doesn't understand what the words mean*

The sentence "Colorless green ideas sleep furiously" is syntactically correct English (adjective-adjective-noun-verb-adverb) but semantically nonsense.

1.6 The Axiomatic Method

1.6.1 What Is an Axiom?

We've been building up to this. Here's the central idea of modern mathematics:

Definition 1.6 (Axiom—Informal)**.** *An **axiom** is a statement we accept as true without proof. Axioms are the starting point for all reasoning in a mathematical system.*

Intuition

Think of axioms as the "rules of the game." In chess, you don't prove that the bishop moves diagonally—it's a rule you accept. In mathematics, we accept certain basic statements (axioms) and derive everything else from them.

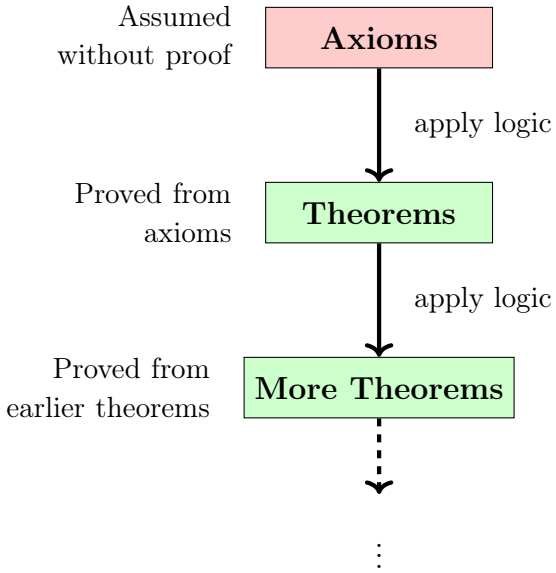
Example 1.7 (Euclidean Geometry Axioms)**.** *Euclid based all of geometry on five axioms (postulates). Here's one:*

Axiom: *Through any two points, there exists exactly one straight line.*

We don't prove this. We assume it. Then we use it to prove other statements (theorems) like the Pythagorean theorem.

1.6.2 The Axiomatic Method: How It Works

Here's the structure of an axiomatic system:



The process:

1. Start with axioms (statements assumed true)
2. Use logical rules to derive new statements (theorems)
3. Use proved theorems to derive more theorems
4. Continue building the mathematical structure

1.6.3 Why Axioms?

You might ask: why not just prove everything? Why have statements we don't prove?

The answer: **we have to start somewhere**. If we had to prove every statement using earlier statements, we'd have infinite regress:

To prove A, we need B.
To prove B, we need C.
To prove C, we need D.
⋮
(Never ends!)

Axioms break this cycle. They're our starting point.

Key Idea

An axiomatic system is like a building:

- **Axioms** are the foundation
- **Definitions** introduce new concepts
- **Theorems** are the floors we build on top

- **Proofs** are the construction that connects them

If the foundation is solid, the building stands. If the axioms are consistent, mathematics works.

1.7 Consistency and Truth

1.7.1 Can Axioms Be Wrong?

Here's a deep question: how do we know our axioms are "correct"?

The answer might surprise you: **we don't care if they're "true"**. We only care if they're **consistent**.

Definition 1.7 (Consistency—Informal)**.** *A set of axioms is* ***consistent*** *if you can never derive a contradiction from them—that is, you can never prove both a statement and its negation.*

Example 1.8 (Inconsistent Axioms)**.** *Suppose we had these axioms:*

1. *All numbers are positive*
2. -5 *is a number*

These are inconsistent! From (1) and (2), we can derive that -5 *is positive. But* -5 *is also negative (by definition). Contradiction!*

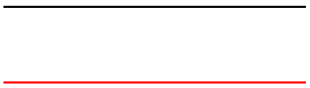
Such a system is useless because we can prove anything (including false statements).

Example 1.9 (Different Consistent Systems)**.** *Consider these two axiom systems:*

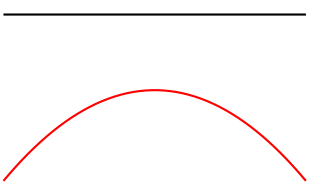
System A: *Euclid's five postulates (including the parallel postulate)* $\rightarrow$ *Gives Euclidean geometry (flat space)*

System B: *Euclid's first four postulates + negation of the parallel postulate* $\rightarrow$ *Gives hyperbolic geometry (curved space)*

Both are consistent! They just describe different mathematical universes. Neither is "wrong"—they're just different.



Euclidean: parallel lines never meet



Hyperbolic: "parallel" lines diverge

1.7.2 Gödel's Shadow

I must mention one of the most profound results in the history of mathematics:

Historical Context

In 1931, Kurt Gödel proved two shocking theorems:
First Incompleteness Theorem: Any consistent axiom system powerful enough to describe arithmetic contains true statements that cannot be proved within the system.
Second Incompleteness Theorem: No consistent axiom system can prove its own consistency.
What this means:

- Mathematics will always be incomplete—there will always be true statements we can't prove
- We can't even prove that mathematics is consistent (without assuming something stronger)
- Our axioms are an act of *faith*, justified by the fact that they haven't led to contradictions yet

This doesn't mean mathematics is broken! It just means it's infinitely deep. There's always more to discover.

1.8 Methods of Proof

Before we dive into formal mathematics, let's understand the tools we'll use to prove theorems. Every proof is a logical argument, but there are several standard patterns that recur throughout mathematics.

1.8.1 Direct Proof

Definition 1.8 (Direct Proof—Informal)**.** *A **direct proof** starts with the hypothesis and uses logical steps to reach the conclusion directly.*

Structure:

1. Assume the hypothesis is true
2. Apply definitions, axioms, and previously proved theorems
3. Arrive at the conclusion

Example 1.10 (Direct Proof)**.** ***Theorem:*** *If n is an even integer, then n^2 is even.*

Proof: *Assume n is even. By definition, $n = 2k$ for some integer k.*

Then:

$$n^2 = (2k)^2 = 4k^2 = 2(2k^2)$$

Since $2k^2$ is an integer, $n^2 = 2 \cdot (\text{integer})$, so n^2 is even. ■

1.8.2 Proof by Contradiction

Definition 1.9 (Proof by Contradiction—Informal)**.** *To prove a statement P, assume $\neg P$ (not P) and derive a logical contradiction. Since the assumption leads to impossibility, P must be true.*

Structure:

1. Assume the negation of what you want to prove
2. Use logical reasoning to derive a contradiction
3. Conclude that the assumption was false, so the original statement is true

Example 1.11 (Proof by Contradiction)**.** ***Theorem:*** *There is no largest natural number.*

Proof: *Suppose, for the sake of contradiction, that there is a largest natural number. Call it N.*

But then $N + 1$ is also a natural number (by closure of addition), and $N + 1 > N$.

This contradicts our assumption that N is the largest natural number.

Therefore, no largest natural number exists. ■

Remark

Proof by contradiction is also called *reductio ad absurdum* (Latin: "reduction to absurdity"). This method is particularly powerful for proving impossibility results and existence of irrational numbers.

1.8.3 Proof by Contrapositive

Definition 1.10 (Contrapositive—Informal)**.** *The **contrapositive** of "if P then Q" is "if not Q then not P."*

These statements are logically equivalent (we'll prove this in Chapter 2).

Strategy: To prove $P \implies Q$, instead prove $\neg Q \implies \neg P$.

Example 1.12 (Proof by Contrapositive)**.** ***Theorem:*** *If n^2 is even, then n is even.*

Direct proof would be hard. *Instead, we prove the contrapositive:*

Contrapositive: *If n is odd, then n^2 is odd.*

Proof: *Assume n is odd. Then $n = 2k + 1$ for some integer k.*

Then:

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

Since $2k^2 + 2k$ is an integer, $n^2 = 2 \cdot (\text{integer}) + 1$, so n^2 is odd.

We've proved the contrapositive, so the original statement is true. ■

1.8.4 Proof by Mathematical Induction

Definition 1.11 (Mathematical Induction—Informal)**.** *To prove a statement $P(n)$ for all natural numbers n:*

1. ***Base case:*** *Prove $P(0)$ (or $P(1)$, depending on context)*
2. ***Inductive step:*** *Prove that if $P(k)$ is true, then $P(k+1)$ is true*

Then $P(n)$ is true for all $n \geq 0$.

Analogy: Climbing an infinite ladder:

- Base case: You can reach the first rung
- Inductive step: If you're on rung k, you can reach rung $k+1$
- Conclusion: You can reach every rung

Example 1.13 (Proof by Induction)**.** ***Theorem:*** *For all $n \geq 1$: $1+2+3+\cdots+n = \frac{n(n+1)}{2}$*

Proof: *Let $P(n)$ be the statement: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.*

Base case ($n = 1$): $1 = \frac{1(1+1)}{2} = \frac{2}{2} = 1$. ✓

Inductive step: *Assume $P(k)$ is true for some $k \geq 1$:*

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2} \quad \textit{(Inductive Hypothesis)}$$

We must prove $P(k+1)$:

$$\begin{aligned}
\sum_{i=1}^{k+1} i &= \left(\sum_{i=1}^{k} i\right) + (k+1) \\
&= \frac{k(k+1)}{2} + (k+1) \quad \textit{(by IH)} \\
&= \frac{k(k+1) + 2(k+1)}{2} \\
&= \frac{(k+1)(k+2)}{2}
\end{aligned}$$

This is exactly $P(k+1)$. By induction, $P(n)$ holds for all $n \geq 1$. ■

1.8.5 Existence vs. Uniqueness Proofs

Many theorems claim that an object with certain properties exists. Often, we must also prove it's unique.

Definition 1.12 (Existence and Uniqueness—Informal)**.** ***Existence proof:*** *Show that at least one object with the desired properties exists.*

Uniqueness proof: *Show that at most one such object exists.*

Together: Exactly one object exists.

Notation: "$\exists!$" means "there exists a unique" (combines $\exists$ and uniqueness).

Example 1.14 (Existence and Uniqueness). ***Theorem:*** *For any integer n, there exists a unique integer m such that $n + m = 0$.*

Existence: *Let $m = -n$. Then $n + (-n) = 0$ by properties of integers. So such an m exists.* ✓

Uniqueness: *Suppose m_1 and m_2 both satisfy $n + m = 0$. Then:*

$$n + m_1 = 0 \quad \textit{and} \quad n + m_2 = 0$$

Therefore:

$$n + m_1 = n + m_2$$

By cancellation (adding $-n$ to both sides):

$$m_1 = m_2$$

So the additive inverse is unique. ■

1.8.6 Constructive vs. Non-Constructive Proofs

Definition 1.13 (Constructive Proof—Informal). *A **constructive proof** of existence explicitly constructs the object in question or provides an algorithm to find it.*

*A **non-constructive proof** proves existence without providing the object or showing how to find it (often using contradiction).*

Example 1.15 (Non-Constructive Existence Proof). ***Theorem:*** *There exist irrational numbers a and b such that a^b is rational.*

Proof: *Consider $\sqrt{2}^{\sqrt{2}}$.*

Case 1: *If $\sqrt{2}^{\sqrt{2}}$ is rational, take $a = b = \sqrt{2}$. Done.*

Case 2: *If $\sqrt{2}^{\sqrt{2}}$ is irrational, take $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Then:*

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2$$

So $a^b = 2$ is rational.

In either case, such a and b exist. ■

Note: *This proof doesn't tell us which case is true! We know the answer exists but don't know what it is. (In fact, $\sqrt{2}^{\sqrt{2}}$ is irrational, but that requires more work to prove.)*

Key Idea

Summary of Proof Techniques:

Method	When to Use
Direct	Clear path from hypothesis to conclusion
Contradiction	Proving impossibility or "no such object exists"
Contrapositive	When the negation is easier to work with
Induction	Statements involving natural numbers or recursively defined structures
Existence	Showing at least one solution exists
Uniqueness	Showing at most one solution exists

Meta-advice: When stuck, try multiple methods. Sometimes proof by contradiction is easy when direct proof is hard, and vice versa.

1.9 Philosophical Foundations: Classical vs. Constructive Mathematics

Mathematics, despite its reputation for objectivity, rests on philosophical choices about what constitutes valid reasoning.

1.9.1 The Law of Excluded Middle

Definition 1.14 (Law of Excluded Middle (LEM)—Informal)**.** *For any statement P, either P is true or $\neg P$ (not P) is true. There is no third option.*

This seems obvious! A number is either even or odd. A statement is either true or false. What else could there be?

But consider: "There exists a sequence of 100 consecutive zeros in the decimal expansion of π."

Is this true or false? We don't know! We haven't computed all digits of π (it's infinite).

Classical view: The statement is either true or false, even if we don't know which. LEM holds.

Constructive view: Until we either find such a sequence or prove none exists, the statement has no definite truth value. LEM should not be assumed.

1.9.2 Classical Mathematics (Our Approach)

Definition 1.15 (Classical Mathematics—Informal)**.** ***Classical mathematics*** *accepts the Law of Excluded Middle and allows proof by contradiction freely.*

Philosophical stance: Mathematical objects exist independently of our knowledge. Statements about them are true or false objectively, even if we can't determine which.

Practical consequence: We can prove existence without constructing. Example: "There exists an x such that $P(x)$" can be proved by showing "Assuming no such x exists leads to contradiction."

This book uses classical mathematics. It's the standard foundation for analysis, algebra, and most of modern mathematics.

1.9.3 Constructive Mathematics (The Alternative)

Definition 1.16 (Constructive Mathematics—Informal)**.** ***Constructive mathematics*** *(or* ***intuitionism****) rejects the Law of Excluded Middle and requires constructive proofs of existence.*

Philosophical stance: Mathematical objects are mental constructions. They exist only when we construct them. Truth means "can be proved," not objective truth independent of proof.

Founder: L.E.J. Brouwer (1881-1966), Dutch mathematician who argued mathematics is a "languageless activity of the mind."

Practical consequence:

- To prove $\exists x : P(x)$, you must explicitly construct such an x
- Proof by contradiction is restricted (only allowed in certain cases)
- Double negation elimination ($\neg\neg P \implies P$) is not always valid

Example 1.16 (Classical vs. Constructive)**.** ***Theorem:*** *For any real number x, either $x = 0$ or $x \neq 0$.*

Classical proof: *By LEM, either $x = 0$ or $x \neq 0$. Done.*

Constructivist response: *This isn't a proof! You must provide an algorithm that, given x, determines which case holds. For computably defined reals, this may be impossible (e.g., if x is defined by a non-halting Turing machine).*

Constructivists would say: The theorem is true only for specific classes of real numbers where we can make the determination.

1.9.4 Why Does This Matter?

Practical implications:

- **Computer science:** Constructive proofs translate directly to algorithms. The Curry-Howard correspondence links constructive logic to type theory and functional programming.
- **Computational mathematics:** Constructive proofs guarantee computability.
- **Foundations:** Understanding these distinctions clarifies what we're assuming.

Remark

Our choice: This book follows classical mathematics because:

1. It's the foundation of standard analysis and most mathematical physics
2. Classical theorems are stronger (easier to prove things)

3. Most working mathematicians use classical logic

However, be aware: when we use proof by contradiction or LEM, we're making a philosophical choice. Some mathematicians (constructivists) would demand more. **Good news:** Most of our early chapters (logic, set theory, basic arithmetic) work in both frameworks. The divergence becomes significant in analysis and infinitary reasoning.

1.9.5 Gödel's Theorems Revisited

Now we can appreciate Gödel's incompleteness theorems more deeply:

Theorem 1.1 (Gödel's First Incompleteness Theorem—Informal)**.** *Any consistent formal system F that can express basic arithmetic contains statements that are true but unprovable within F.*

Implication: There are limits to what axioms can do. Mathematics is inherently incomplete—there will always be true statements we cannot prove from our axioms.

Theorem 1.2 (Gödel's Second Incompleteness Theorem—Informal)**.** *No consistent formal system F (containing arithmetic) can prove its own consistency within itself.*

Implication: We cannot prove that mathematics is consistent using mathematics alone. We must take consistency as an article of faith (justified by lack of contradictions so far).

Historical Context

Before Gödel (1931), mathematicians hoped to:

- Find a complete set of axioms (all truths provable)
- Prove mathematics consistent (no contradictions possible)

Gödel showed both goals are impossible.
Hilbert's Program (1920s): Prove consistency of mathematics using only "finitary" methods (basic, unquestionable reasoning).
Gödel's result: Hilbert's program cannot work. To prove consistency of arithmetic, you need axioms stronger than arithmetic itself.
Modern view: We accept Gödel's limits. Mathematics is an open-ended endeavor. There's always more to discover, and we can never be absolutely certain we won't find a contradiction. But after a century of set theory with no contradictions, we're reasonably confident.

Key Idea

Philosophical Summary:

	Classical	**Constructive**
Truth	Objective, independent	Provability
LEM	Always valid	Not assumed
Contradiction	Freely used	Restricted
Existence	Can be non-constructive	Must construct
Real numbers	Completed infinity	Potential infinity
Advantage	Stronger theorems	Computational content

This book: Classical mathematics (standard approach)
Awareness: We're making philosophical choices, not discovering absolute truths

1.10 Looking Ahead

We've covered a lot of ground informally:

- Mathematics as a formal game with symbols
- The distinction between syntax (form) and semantics (meaning)
- Strings, alphabets, and concatenation
- The axiomatic method
- Consistency vs. truth

Key Idea

Where we're going:
Chapter 2 (Propositional Logic): We'll formalize boolean logic—statements that are true or false. This is the simplest formal system, and it teaches us how formal systems work.
Chapter 3 (Set Theory): We'll build the universe of mathematical objects from the single primitive notion of membership ($\in$). Everything in mathematics—numbers, functions, shapes—can be encoded as sets.
Beyond: With logic and sets, we can build anything: number systems, algebra, calculus, and all of modern mathematics.

Remark

This chapter used informal language. Starting in Chapter 2, we'll be increasingly formal. But don't worry—we'll continue to provide intuition and motivation. The pattern will always be:
① Intuition → ② Formal definition → ③ Examples → ④ Theorems
Trust the process. Mathematics is a ladder—each rung depends on the ones below.

You're ready. Let's begin building mathematics from scratch.

Chapter 2

Propositional Logic: Reasoning with True and False

2.1 From Intuition to Formalism

Intuition

In Chapter 1, we talked about mathematics as a game with symbols. Now we're going to play our first formal game: **propositional logic**.
This is logic at its simplest. We deal with statements that are either **true** or **false**, and we combine them using words like "and," "or," and "not."
Think of this as the foundation of all reasoning. Every time you say "If it rains, then I'll bring an umbrella," you're using propositional logic.

2.1.1 What Is a Proposition?

Let's start with examples before we formalize.

Example 2.1 (Propositions in Everyday Life)**.** *These are propositions (statements with a definite true/false value):*

- *"The sky is blue" (True)*
- *"2 + 2 = 5" (False)*
- *"Paris is the capital of France" (True)*
- *"All cats are orange" (False)*

These are NOT propositions:

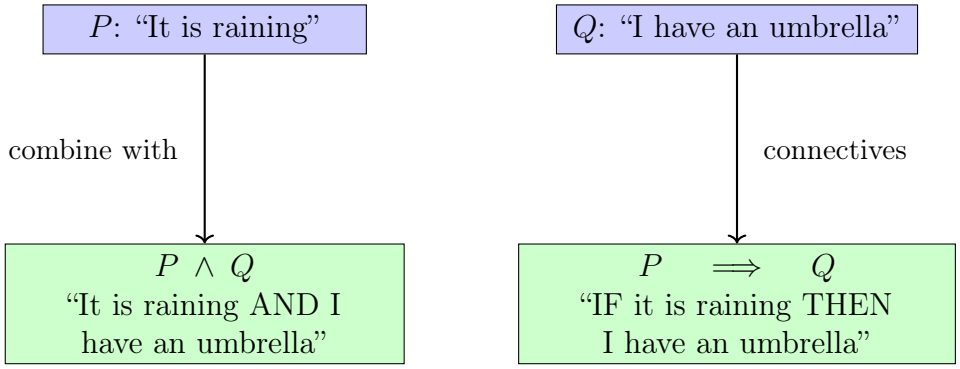
- *"What time is it?" (Question, not a statement)*
- *"Close the door!" (Command, not a statement)*
- *"This sentence is false" (Paradox—can't be consistently true or false)*
- *"$x > 5$" (Depends on x—not yet a proposition)*

Key Idea

A proposition is a declarative statement that is either true or false (but not both, and not neither).
For now, think of propositions as sentences that make claims about the world. Later, we'll make this completely formal.

2.1.2 Building Complex Statements

We can build complex propositions from simple ones using **logical connectives**:



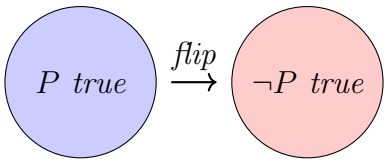
Let's understand each connective intuitively before formalizing.

2.1.3 The Five Connectives (Informally)

1. Negation ($\neg$): NOT

$\neg P$ means "not P" or "P is false."

Example 2.2. *If P = "It is raining," then $\neg P$ = "It is NOT raining."*



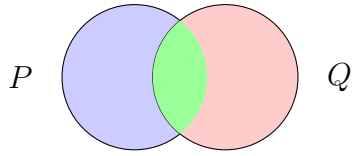
2. Conjunction ($\wedge$): AND

$P \wedge Q$ means "both P and Q are true."

Example 2.3. *P = "It is raining," Q = "It is cold"*

$P \wedge Q$ = "It is raining AND it is cold"

This is true only if BOTH conditions hold.



$P \wedge Q$ true only in overlap

3. Disjunction ($\vee$): OR

$P \vee Q$ means "at least one of P or Q is true" (possibly both).

Example 2.4. *P = "I will take the bus," Q = "I will take the train"*

$P \vee Q$ = "I will take the bus OR the train (or both if they go to the same place)"

This is true if either (or both) hold.

Warning

In everyday English, "or" sometimes means "one but not both" (exclusive or). In logic, $\vee$ means "at least one" (inclusive or).
"Would you like coffee or tea?" (English: exclusive)
"$x > 5$ or $x < 10$" (Logic: inclusive—both could be true)

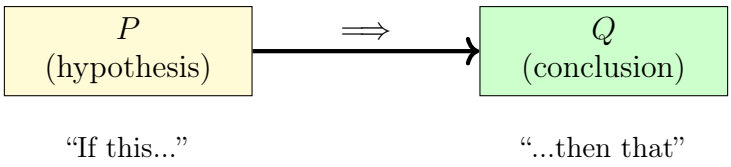
4. Implication ($\implies$): IF...THEN

$P \implies Q$ means "if P is true, then Q must be true."

Example 2.5. *P = "It is raining," Q = "The ground is wet"*

$P \implies Q$ = "If it is raining, then the ground is wet"

This is the trickiest connective! Let's understand it deeply:



Key Idea

The implication $P \implies Q$ is a **promise**: "Whenever P is true, Q will also be true."
The promise is **broken** only when P is true but Q is false. In all other cases, the promise holds.

5. Biconditional ($\iff$): IF AND ONLY IF

$P \iff Q$ means "P and Q always have the same truth value."

Example 2.6. *P = "$x = 2$," Q = "$x^2 = 4$"*

$P \iff Q$ would mean these always happen together (which isn't quite true because $x = -2$ also makes $x^2 = 4$).

Better example: P = "A triangle is equilateral," Q = "All three sides are equal"

$P \iff Q$ = true (these mean exactly the same thing)

2.2 Making It Formal: Syntax

Now that we have intuition, let's build the formal system.

Key Idea

Remember the pattern from Chapter 1:

1. Define the **alphabet** (symbols we can use)
2. Define which strings are **well-formed formulas** (grammatically correct)
3. Define what formulas **mean** (semantics)
4. Prove theorems about the system

We're doing step 1 and 2 now: pure syntax.

2.2.1 The Alphabet of Propositional Logic

Definition 2.1 (Alphabet)**.** *Our alphabet $\mathcal{L}_0$ consists of:*

1. ***Propositional variables**: $P, Q, R, S, P_1, P_2, \ldots$ (infinitely many)*
2. ***Logical connectives**: $\neg, \wedge, \vee, \implies, \iff$*
3. ***Parentheses**: (and)*

Remark

Propositional variables are placeholders for actual propositions. P might stand for "it is raining" or "$2 + 2 = 4$" or anything with a truth value.

2.2.2 Well-Formed Formulas: What's Legal?

Not every string of symbols makes sense. "$\wedge \wedge PQ)$" uses our symbols, but it's nonsense. We need rules for what's grammatically correct.

Definition 2.2 (Well-Formed Formula)**.** *The set of **well-formed formulas** (WFFs) is defined recursively:*

Base case:

- *Every propositional variable ($P, Q, R, \ldots$) is a WFF*

Recursive rules: *If ϕ and ψ are WFFs, then so are:*

- $(\neg\phi)$
- $(\phi \wedge \psi)$
- $(\phi \vee \psi)$
- $(\phi \implies \psi)$
- $(\phi \iff \psi)$

Nothing else is a WFF.

Intuition

Think of this as a recipe:

1. Start with simple ingredients (variables)
2. Combine them using approved methods (connectives)
3. Keep combining what you've built
4. You can only use what the rules allow

Example 2.7 (Building Formulas)**.** ***Step-by-step construction:***

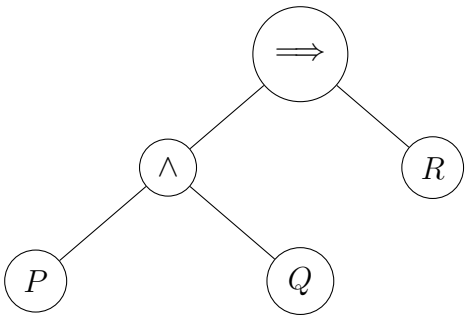
Level 1: *P, Q, R are WFFs (base case)*

Level 2: *Since P and Q are WFFs:*

- $(\neg P)$ *is a WFF*
- $(P \wedge Q)$ *is a WFF*
- $(P \vee Q)$ *is a WFF*

Level 3: *Since $(P \wedge Q)$ and R are WFFs:*

- $((P \wedge Q) \implies R)$ *is a WFF*



Parse tree for $(P \wedge Q) \implies R$

Example 2.8 (What's NOT a WFF).

- $P\wedge$ — *incomplete*
- $\wedge PQ$ — *wrong structure (prefix notation not allowed)*
- PQ — *missing connective*
- $(P\wedge)$ — *incomplete*

2.2.3 Precedence Rules (Practical Notation)

Writing all those parentheses gets tedious. We adopt conventions:

Notation (Precedence).

1. $\neg$ *binds tightest (apply first)*
2. $\wedge$ *and* $\vee$ *bind next*
3. $\Longrightarrow$ *and* $\Longleftrightarrow$ *bind weakest (apply last)*

So $\neg P \wedge Q \implies R$ *means* $((\neg P) \wedge Q) \implies R$.

With all parentheses: $((\neg P) \wedge Q) \implies R$
With conventions: $\neg P \wedge Q \implies R$
Same formula!

2.3 Giving Meaning: Semantics

We've defined which strings are grammatically correct. Now: *what do they mean?*

2.3.1 Truth Values

Definition 2.3 (Truth Values). *We work with two truth values:* $\mathbb{B} = \{0, 1\}$ *where:*

- 0 *represents "false"*
- 1 *represents "true"*

Remark

We use 0 and 1 (instead of F and T) to emphasize these are just formal objects. We could use any two distinct symbols. The labels "false" and "true" are just convenient names.

2.3.2 Truth Assignments

Definition 2.4 (Truth Assignment). *A* ***truth assignment*** *(or* ***valuation****) is a function:*

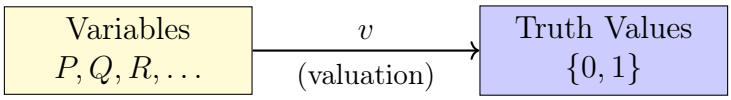
$$v : \{\textit{propositional variables}\} \to \{0, 1\}$$

that assigns a truth value to each propositional variable.

Intuition

A truth assignment is a "possible world." It says: in this scenario, P is true, Q is false, R is true, etc.
Example scenarios:

- v_1: $P \mapsto 1, Q \mapsto 0, R \mapsto 1$ ("P and R are true, Q is false")
- v_2: $P \mapsto 0, Q \mapsto 0, R \mapsto 0$ ("everything is false")
- v_3: $P \mapsto 1, Q \mapsto 1, R \mapsto 1$ ("everything is true")



2.3.3 Truth Functions: How Connectives Work

Now we extend the valuation to *all* formulas by defining how connectives combine truth values.

Definition 2.5 (Truth Functions)**.** *For any valuation v and formulas ϕ, ψ:*

Negation:

$$v(\neg\phi) = \begin{cases} 1 & \text{if } v(\phi) = 0 \\ 0 & \text{if } v(\phi) = 1 \end{cases}$$

Conjunction:

$$v(\phi \wedge \psi) = \begin{cases} 1 & \text{if } v(\phi) = 1 \text{ and } v(\psi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Disjunction:

$$v(\phi \vee \psi) = \begin{cases} 1 & \text{if } v(\phi) = 1 \text{ or } v(\psi) = 1 \text{ (or both)} \\ 0 & \text{otherwise} \end{cases}$$

Implication:

$$v(\phi \implies \psi) = \begin{cases} 0 & \text{if } v(\phi) = 1 \text{ and } v(\psi) = 0 \\ 1 & \text{otherwise} \end{cases}$$

Biconditional:

$$v(\phi \iff \psi) = \begin{cases} 1 & \text{if } v(\phi) = v(\psi) \\ 0 & \text{otherwise} \end{cases}$$

Let's visualize these with truth tables.

2.3.4 Truth Tables

A **truth table** shows the output for all possible inputs.

Negation:

P	$\neg P$
0	1
1	0

Intuition

Negation flips the truth value. Simple!

Conjunction (AND):

P	Q	$P \wedge Q$
0	0	0
0	1	0
1	0	0
1	1	1

Intuition

"AND" is strict: true only when BOTH are true.

Disjunction (OR):

P	Q	$P \vee Q$
0	0	0
0	1	1
1	0	1
1	1	1

Intuition

"OR" is generous: true when AT LEAST ONE is true.

Implication (IF...THEN):

P	Q	$P \implies Q$
0	0	1
0	1	1
1	0	0
1	1	1

This deserves special attention!

Key Idea

The implication $P \implies Q$ is false ONLY when the hypothesis (P) is true but the conclusion (Q) is false.
Why are rows 1 and 2 true (when P is false)?
Think of it as a promise: "If P happens, then Q will happen."

- If P doesn't happen, the promise isn't tested—we consider it kept.
- This is called **vacuous truth**: an implication with false hypothesis is vacuously true.

Example 2.9 (Vacuous Truth)**.** *"If* $1 = 0$*, then pigs can fly."*

The hypothesis $(1 = 0)$ *is false. So this entire statement is vacuously true!*

Weird? Yes. But consistent? Absolutely. This definition makes mathematical theorems work correctly.

Why this definition?

Consider: "If it rains, the ground is wet."

Four scenarios:

1. Rain + wet ground: Promise kept ✓ (true)
2. Rain + dry ground: Promise broken × (false)
3. No rain + wet ground: Promise not tested ✓ (true)
4. No rain + dry ground: Promise not tested ✓ (true)

Biconditional (IF AND ONLY IF):

P	Q	$P \iff Q$
0	0	1
0	1	0
1	0	0
1	1	1

Intuition

$P \iff Q$ means "P and Q always go together." True when both have the same value.

2.4 Important Formulas and Equivalences

Now we can state and prove general laws.

2.4.1 Tautologies: Always True

Definition 2.6 (Tautology)**.** *A formula ϕ is a **tautology** if $v(\phi) = 1$ for every valuation v.*

We write $\models \phi$ to mean "ϕ is a tautology."

Intuition

A tautology is *logically true*—true regardless of what propositions mean. It's true purely because of its logical structure.

Example 2.10 (Famous Tautologies)**.** 1. ***Law of Excluded Middle**: $P \vee \neg P$*

"Either P is true or P is false." (No middle option!)

P	$\neg P$	$P \vee \neg P$
0	*1*	*1*
1	*0*	*1*

Always true! ✓

2. ***Law of Non-Contradiction**: $\neg(P \wedge \neg P)$*

"P can't be both true and false."

3. ***Identity**: $P \implies P$*

"If P is true, then P is true." (Trivial but fundamental!)

2.4.2 Key Logical Equivalences

Definition 2.7 (Logical Equivalence)**.** *Formulas ϕ and ψ are **logically equivalent** (written $\phi \equiv \psi$) if they have identical truth tables—that is, $v(\phi) = v(\psi)$ for all valuations v.*

Theorem 2.1 (Fundamental Equivalences)**.** *The following equivalences hold:*

1. ***Double Negation:***

$$\neg\neg P \equiv P$$

2. ***Commutativity:***

$$P \wedge Q \equiv Q \wedge P$$
$$P \vee Q \equiv Q \vee P$$

3. ***Associativity:***

$$(P \wedge Q) \wedge R \equiv P \wedge (Q \wedge R)$$
$$(P \vee Q) \vee R \equiv P \vee (Q \vee R)$$

4. ***Distributivity:***

$$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$$
$$P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$$

5. De Morgan's Laws:

$$\neg(P \wedge Q) \equiv \neg P \vee \neg Q$$
$$\neg(P \vee Q) \equiv \neg P \wedge \neg Q$$

6. Implication as Disjunction:

$$P \implies Q \equiv \neg P \vee Q$$

7. Contrapositive:

$$P \implies Q \equiv \neg Q \implies \neg P$$

Let me prove a few to show the method:

Proof of De Morgan's Law: $\neg(P \wedge Q) \equiv \neg P \vee \neg Q$. We show these have identical truth tables:

P	Q	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P \vee \neg Q$	Match?
0	0	0	1	1	✓
0	1	0	1	1	✓
1	0	0	1	1	✓
1	1	1	0	0	✓

Columns match exactly! Therefore $\neg(P \wedge Q) \equiv \neg P \vee \neg Q$. ■

Intuition

De Morgan's Laws say: to negate an "and," flip to "or" and negate each part. To negate an "or," flip to "and" and negate each part.
Example: Negate "It's raining AND cold"

- Becomes: "It's NOT raining OR it's NOT cold"
- Makes sense: the original is false if at least one condition fails

Proof of Contrapositive: $P \implies Q \equiv \neg Q \implies \neg P$.

P	Q	$P \implies Q$	$\neg Q$	$\neg P$	$\neg Q \implies \neg P$
0	0	1	1	1	1
0	1	1	0	1	1
1	0	0	1	0	0
1	1	1	0	0	1

Columns 3 and 6 match! Therefore $P \implies Q \equiv \neg Q \implies \neg P$. ■

Key Idea

The contrapositive is *logically identical* to the original implication. This is why proof by contrapositive works:

To prove "$P \implies Q$," we can instead prove "$\neg Q \implies \neg P$." Same thing!
Example:

- Original: "If n^2 is even, then n is even"
- Contrapositive: "If n is odd, then n^2 is odd"
- (The contrapositive is often easier to prove!)

2.5 Inference Rules: How We Reason

Now we can formalize common patterns of reasoning.

2.5.1 Modus Ponens

Theorem 2.2 (Modus Ponens)**.**

$$(P \wedge (P \implies Q)) \implies Q \text{ is a tautology}$$

Intuition

In words: "If P is true, and if 'P implies Q' is true, then Q must be true."
Pattern of reasoning:

1. We know P is true
2. We know $P \implies Q$ is true
3. Therefore, Q must be true

This is the most fundamental rule of inference!

$$\frac{\begin{array}{l} P \\ P \implies Q \end{array}}{\therefore Q}$$

Premises:
We have P
We have $P \implies Q$

Conclusion:
Therefore we conclude Q

2.5.2 Other Inference Rules

Theorem 2.3 (Modus Tollens)**.**

$$((P \implies Q) \wedge \neg Q) \implies \neg P \text{ is a tautology}$$

Intuition

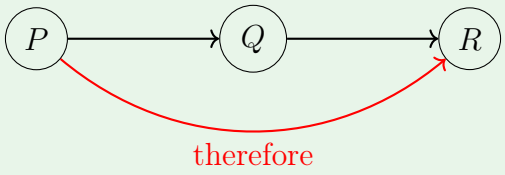
"If $P \implies Q$ is true, and Q is false, then P must be false."
Why? If P were true, then Q would have to be true (by the implication). But Q is false. Contradiction! So P must be false.

Theorem 2.4 (Hypothetical Syllogism)**.**

$$((P \implies Q) \wedge (Q \implies R)) \implies (P \implies R) \textit{ is a tautology}$$

Intuition

"Implications chain." If P leads to Q, and Q leads to R, then P leads to R.



2.6 Predicate Logic: The Language of Quantifiers

Intuition

Propositional logic is great, but it's not enough. Consider this classic argument:

1. All men are mortal.
2. Socrates is a man.
3. Therefore, Socrates is mortal.

In propositional logic, this looks like:

1. P ("All men are mortal")
2. Q ("Socrates is a man")
3. Therefore R ("Socrates is mortal")

There is no logical link between P, Q, and R! To see the connection, we need to look *inside* the propositions. We need **Predicate Logic**.

2.6.1 Predicates and Variables

Definition 2.8 (Predicate)**.** *A **predicate** is a statement involving variables (like x, y) that becomes a proposition (true or false) when specific values are substituted for the variables.*

Example 2.11. *Let $P(x)$ be the statement "$x > 5$".*

- *$P(x)$ is not true or false (it depends on x).*
- *$P(2)$ is "$2 > 5$", which is **false**.*
- *$P(7)$ is "$7 > 5$", which is **true**.*

2.6.2 Quantifiers

We can also make a predicate into a proposition by quantifying over a **domain of discourse** (the set of all objects we are talking about).

Definition 2.9 (Universal Quantifier $\forall$)**.** *The symbol $\forall$ means "for all" or "for every".*

$$\forall x P(x)$$

means "for every x in the domain, $P(x)$ is true."

Definition 2.10 (Existential Quantifier $\exists$)**.** *The symbol $\exists$ means "there exists" or "for some".*

$$\exists x P(x)$$

means "there is at least one x in the domain such that $P(x)$ is true."

Example 2.12. *Domain: integers $\mathbb{Z}$. Let $P(x)$ be "$x^2 \geq 0$".*

- *$\forall x P(x)$ is true (square of any integer is non-negative).*
- *$\exists x(x < 0)$ is true (e.g., -1).*
- *$\forall x(x > 0)$ is false (e.g., -1 is not > 0).*

2.6.3 Negating Quantifiers (De Morgan for Logic)

How do we say "Not everyone likes pizza"? It means "There exists someone who does **not** like pizza."

Theorem 2.5 (De Morgan's Laws for Quantifiers)**.**

$$\neg(\forall x P(x)) \equiv \exists x \neg P(x)$$
$$\neg(\exists x P(x)) \equiv \forall x \neg P(x)$$

Intuition

To show a universal statement ("All swans are white") is false, you only need to find **one** counterexample (one black swan). You don't need to show *all* swans are non-white.

2.6.4 Bounded Quantifiers

Key Idea

In set theory and mathematics, we often quantify over elements of a specific set, not over all objects. This leads to **bounded quantifiers** (also called **restricted quantifiers**).

Definition 2.11 (Bounded Quantifier Notation)**.** *The notation* $\forall x \in A, P(x)$ *and* $\exists x \in A, P(x)$ *are* ***abbreviations****:*

$$\forall x \in A, P(x) \equiv \forall x(x \in A \implies P(x))$$
$$\exists x \in A, P(x) \equiv \exists x(x \in A \wedge P(x))$$

In words:

- $\forall x \in A, P(x)$*: "For all* x *in* A*,* $P(x)$ *holds" means "If* x *is in* A*, then* $P(x)$*"*
- $\exists x \in A, P(x)$*: "There exists* x *in* A *such that* $P(x)$*" means "There exists* x *that is both in* A *and satisfies* $P(x)$*"*

Example 2.13 (Bounded vs. Unbounded Quantifiers)**.** ***Unbounded****:* $\forall x(x^2 \geq 0)$ *— quantifies over all objects*

Bounded*:* $\forall x \in \mathbb{R}, x^2 \geq 0$ *— quantifies only over real numbers*

Expanded form: $\forall x(x \in \mathbb{R} \implies x^2 \geq 0)$

Unbounded*:* $\exists x(x^2 = 2)$ *— might not have a solution depending on universe*

Bounded*:* $\exists x \in \mathbb{R}, x^2 = 2$ *— solution exists in the reals*

Expanded form: $\exists x(x \in \mathbb{R} \wedge x^2 = 2)$

Warning

Notice the asymmetry:

- Universal bounded quantifiers use $\implies$ (implication)
- Existential bounded quantifiers use $\wedge$ (conjunction)

This is correct! Think about it: "All reals are positive" is false if even one real is non-positive (implication handles this). But "There exists a positive real" needs to find something that is *both* real *and* positive (conjunction).

Remark

Throughout this text, when we move from logic to set theory and beyond, we will frequently use bounded quantifier notation. Remember that these are always translatable back to pure first-order logic using the definitions above.

2.7 First-Order Logic: Formal Syntax

Now we make predicate logic completely formal, just as we did for propositional logic.

2.7.1 The Language of First-Order Logic

Definition 2.12 (Alphabet of First-Order Logic)**.** *The alphabet consists of:*

1. ***Variables**: $x, y, z, x_1, x_2, \ldots$ (infinitely many)*
2. ***Logical connectives**: $\neg, \wedge, \vee, \implies, \iff$*
3. ***Quantifiers**: $\forall$ (universal), $\exists$ (existential)*
4. ***Equality**: $=$ (optional, depending on context)*
5. ***Parentheses**: $(,)$*
6. ***Non-logical symbols** (vary by application):*
   - ***Constant symbols**: $c, d, 0, 1, \ldots$*
   - ***Function symbols**: $f, g, +, \times, \ldots$ (each has arity)*
   - ***Predicate symbols**: $P, Q, <, \in, \ldots$ (each has arity)*

Example 2.14 (Language of Arithmetic)**.** *For Peano arithmetic:*

- *Constant:* 0
- *Functions: S (successor, arity 1), $+$ (addition, arity 2), $\times$ (multiplication, arity 2)*
- *Predicate: $=$ (equality, arity 2)*

Example formula: $\forall x \exists y (x+S(0) = y)$ ("for every x, there exists y such that $x+1 = y$")

2.7.2 Terms and Formulas

Definition 2.13 (Terms)**.** ***Terms** are built recursively:*

1. *Every variable is a term*
2. *Every constant symbol is a term*
3. *If $t_1, \ldots, t_n$ are terms and f is a function symbol of arity n, then $f(t_1, \ldots, t_n)$ is a term*

Terms represent objects in the domain.

Definition 2.14 (Well-Formed Formulas (WFFs) of First-Order Logic)**.** *The set of well-formed formulas is defined recursively:*

Base case (Atomic formulas):

- *If P is a predicate symbol of arity n and $t_1, \ldots, t_n$ are terms, then $P(t_1, \ldots, t_n)$ is a WFF*
- *If t_1 and t_2 are terms, then $(t_1 = t_2)$ is a WFF*

Recursive rules: *If ϕ and ψ are WFFs and x is a variable:*

- $(\neg\phi)$ *is a WFF*
- $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \implies \psi)$, $(\phi \iff \psi)$ *are WFFs*
- $(\forall x\phi)$ *is a WFF*
- $(\exists x\phi)$ *is a WFF*

Example 2.15 (Building a Formula)**.** *Domain: natural numbers. Build: "Every number has a successor greater than itself"*

Step by step:

1. *Atomic:* $x < S(x)$ *("x is less than the successor of x")*
2. *Quantify:* $\forall x(x < S(x))$

Full formula: $\forall x(x < S(x))$

2.7.3 Free and Bound Variables

Definition 2.15 (Free and Bound Variables)**.** *In a formula ϕ:*

- *A variable x is **bound** if it appears within the scope of a quantifier $\forall x$ or $\exists x$*
- *A variable x is **free** if it is not bound*

*A formula with no free variables is called a **sentence**.*

Example 2.16.

- $\forall x P(x, y)$*: x is bound, y is free*
- $\forall x \exists y(x < y)$*: Both x and y are bound (this is a sentence)*
- $P(x) \wedge \forall x Q(x)$*: The first x is free, the second is bound (same symbol, different roles!)*

Warning

The same variable can be free in one part of a formula and bound in another. Context matters!
To avoid confusion, rename bound variables when necessary ("α-conversion").

2.8 Semantics of First-Order Logic

2.8.1 Models and Interpretations

Definition 2.16 (Model/Structure)**.** *A **model** (or **structure**) $\mathcal{M}$ for a first-order language consists of:*

1. *A non-empty set D called the* ***domain*** *(or universe)*
2. *An interpretation for each non-logical symbol:*
   - *Each constant symbol c is assigned an element $c^{\mathcal{M}} \in D$*
   - *Each n-ary function symbol f is assigned a function $f^{\mathcal{M}} : D^n \to D$*
   - *Each n-ary predicate symbol P is assigned a relation $P^{\mathcal{M}} \subseteq D^n$*

Example 2.17 (Model for Arithmetic)**.** *Let $\mathcal{N} = (\mathbb{N}, 0^{\mathcal{N}}, S^{\mathcal{N}}, +^{\mathcal{N}}, \times^{\mathcal{N}})$ where:*

- *Domain:* $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
- $0^{\mathcal{N}} = 0$ *(the number zero)*
- $S^{\mathcal{N}}(n) = n + 1$ *(successor function)*
- $+^{\mathcal{N}}$ *and* $\times^{\mathcal{N}}$ *are usual addition and multiplication*

This is the "standard model" of arithmetic.

2.8.2 Truth in a Model

Definition 2.17 (Satisfaction)**.** *Let $\mathcal{M}$ be a model and ϕ a sentence (no free variables). We define $\mathcal{M} \models \phi$ ("$\mathcal{M}$ satisfies ϕ" or "ϕ is true in $\mathcal{M}$") recursively:*
Base case:

- $\mathcal{M} \models P(t_1, \ldots, t_n)$ *iff* $(t_1^{\mathcal{M}}, \ldots, t_n^{\mathcal{M}}) \in P^{\mathcal{M}}$

Recursive cases:

- $\mathcal{M} \models \neg\phi$ *iff* $\mathcal{M} \not\models \phi$
- $\mathcal{M} \models \phi \wedge \psi$ *iff* $\mathcal{M} \models \phi$ *and* $\mathcal{M} \models \psi$
- $\mathcal{M} \models \phi \vee \psi$ *iff* $\mathcal{M} \models \phi$ *or* $\mathcal{M} \models \psi$
- $\mathcal{M} \models \forall x \phi(x)$ *iff for all* $d \in D$, $\mathcal{M} \models \phi(d)$
- $\mathcal{M} \models \exists x \phi(x)$ *iff for some* $d \in D$, $\mathcal{M} \models \phi(d)$

Example 2.18. *Let $\mathcal{M}$ have domain $\{1, 2, 3\}$ and interpret $P(x)$ as "x is even." Then:*

- $\mathcal{M} \models P(2)$ *(true: 2 is even)*
- $\mathcal{M} \not\models P(3)$ *(false: 3 is not even)*
- $\mathcal{M} \models \exists x P(x)$ *(true: 2 is even)*
- $\mathcal{M} \not\models \forall x P(x)$ *(false: not all are even)*

2.8.3 Validity, Satisfiability, and Logical Consequence

Definition 2.18. *Let ϕ be a sentence.*

- *ϕ is **valid** (or a **logical truth**, written $\models \phi$) if $\mathcal{M} \models \phi$ for every model $\mathcal{M}$*
- *ϕ is **satisfiable** if $\mathcal{M} \models \phi$ for some model $\mathcal{M}$*
- *ϕ is **unsatisfiable** if $\mathcal{M} \not\models \phi$ for every model $\mathcal{M}$*

Example 2.19.
- *$\forall x(P(x) \vee \neg P(x))$ is **valid** (true in every model)*
- *$\forall x P(x)$ is **satisfiable** (true in some models, false in others)*
- *$\forall x P(x) \wedge \exists x \neg P(x)$ is **unsatisfiable** (always false)*

2.9 Natural Deduction: A Proof System

Semantics tells us what formulas *mean*. Now we need a **proof system** to mechanically derive theorems.

2.9.1 Inference Rules for Quantifiers

Definition 2.19 (Universal Introduction ($\forall$-Intro))**.**

$$\frac{\phi(x)}{\forall x \phi(x)}$$

If $\phi(x)$ holds for an arbitrary x (with no assumptions about x), then $\forall x \phi(x)$.

Definition 2.20 (Universal Elimination ($\forall$-Elim))**.**

$$\frac{\forall x \phi(x)}{\phi(t)}$$

If $\forall x \phi(x)$ holds, then $\phi(t)$ holds for any term t.

Definition 2.21 (Existential Introduction ($\exists$-Intro))**.**

$$\frac{\phi(t)}{\exists x \phi(x)}$$

If $\phi(t)$ holds for some specific term t, then $\exists x \phi(x)$.

Definition 2.22 (Existential Elimination ($\exists$-Elim))**.**

$$\frac{\exists x \phi(x) \quad [\phi(c) \vdash \psi]}{\psi}$$

If $\exists x \phi(x)$ holds, and assuming $\phi(c)$ for a fresh constant c (witness) allows us to derive ψ (where c doesn't appear in ψ), then ψ holds.

This captures the idea: "Let c be such that $\phi(c)$ holds. Then we can derive ψ."

2.9.2 Example Proof in Natural Deduction

Example 2.20 (Prove: $\forall x P(x) \implies \exists x P(x)$ (if everything has property P, then something has property P))**.** ***Proof:***

1. *Assume $\forall x P(x)$ (hypothesis)*
2. *By $\forall$-Elim with any term (say c): $P(c)$*
3. *By $\exists$-Intro: $\exists x P(x)$*
4. *Therefore: $\forall x P(x) \implies \exists x P(x)$ (discharge assumption)*

■

Note: *This is only valid if the domain is non-empty! With an empty domain, $\forall x P(x)$ is vacuously true but $\exists x P(x)$ is false.*

2.10 Soundness and Completeness

These are the two fundamental meta-theorems of logic, connecting syntax (proofs) to semantics (truth).

2.10.1 Soundness Theorem

Theorem 2.6 (Soundness of First-Order Logic)**.** *If ϕ is provable (there exists a proof of ϕ in natural deduction), then ϕ is valid (true in all models).*

Symbolically: *If $\vdash \phi$, then $\models \phi$.*

Intuition

Soundness means the proof system doesn't prove false things. Every theorem you can prove is actually true (in all models).
If you can derive a formula using the inference rules, that formula really is logically valid.

Proof Sketch. By induction on the length of the proof.

Base case: Axioms are valid (check each axiom in every model).

Inductive step: Show that each inference rule preserves validity. If the premises are valid, the conclusion is valid.

For example, for Modus Ponens: If $\models \phi$ and $\models \phi \implies \psi$, then in any model $\mathcal{M}$:

- $\mathcal{M} \models \phi$ (by first premise)
- $\mathcal{M} \models \phi \implies \psi$ (by second premise)
- Therefore $\mathcal{M} \models \psi$ (by semantics of $\implies$)

So $\models \psi$. ■ ■

2.10.2 Completeness Theorem (Gödel, 1929)

Theorem 2.7 (Gödel's Completeness Theorem for First-Order Logic)**.** *If ϕ is valid (true in all models), then ϕ is provable (there exists a proof of ϕ).*

Symbolically: *If $\models \phi$, then $\vdash \phi$.*

Intuition

Completeness means the proof system is powerful enough to prove everything that's true. If a formula is valid (true in all models), you can find a proof of it. This is Gödel's *Completeness Theorem* (1929), not to be confused with his *Incompleteness Theorems* (1931)!

Historical Context

Kurt Gödel's Two Great Theorems:
1. Completeness Theorem (1929): First-order logic is complete. Everything that's semantically true can be proven.
2. Incompleteness Theorems (1931): Arithmetic (and any system containing it) is incomplete. Not all true statements about numbers can be proven.
The distinction:

- **Completeness** applies to *logic itself*—the rules of reasoning are sufficient
- **Incompleteness** applies to *mathematical theories*—no finite set of axioms can capture all arithmetic truths

Both are true! Logic as a reasoning system is complete, but specific mathematical theories are incomplete.

Proof Sketch of Completeness. The proof is highly technical. The key idea (Henkin's method):

Contrapositive approach: Show that if ϕ is not provable, then ϕ is not valid (i.e., there exists a model where ϕ is false).

Construction:

1. Assume $\nvdash \phi$ (not provable)
2. Extend to a maximally consistent set Γ containing $\neg\phi$
3. **Key step:** Construct a model $\mathcal{M}$ from Γ itself:
   - Domain: equivalence classes of terms
   - Interpret predicates: $P(t_1, \ldots, t_n)$ is true in $\mathcal{M}$ iff $P(t_1, \ldots, t_n) \in \Gamma$
4. Prove (by induction) that for any sentence ψ: $\mathcal{M} \models \psi$ iff $\psi \in \Gamma$
5. Since $\neg\phi \in \Gamma$, we have $\mathcal{M} \models \neg\phi$, so $\mathcal{M} \not\models \phi$
6. Therefore ϕ is not valid

Contrapositive: If ϕ is valid, then ϕ is provable. ■ ■

2.10.3 Consequences of Completeness

Corollary 2.8 (Compactness Theorem)**.** *If every finite subset of a set of sentences* Γ *is satisfiable, then* Γ *itself is satisfiable.*

Intuition

You can't create an inconsistency using infinitely many axioms unless some finite subset is already inconsistent.
This has profound consequences in model theory (e.g., existence of non-standard models of arithmetic).

Corollary 2.9 (Löwenheim-Skolem Theorem)**.** *If a countable first-order theory has an infinite model, it has a countable model.*

Intuition

This is surprising! Even though we might think our axioms describe uncountable structures (like the real numbers), there exist countable models satisfying the same axioms.
This shows the limitations of first-order logic—it can't fully capture uncountability.

2.11 Looking Forward

We have now upgraded our logical toolkit:

- **Propositional Logic**: $P \wedge Q \implies R$
- **Predicate Logic**: $\forall x \exists y (x < y)$

This is the language of modern mathematics. But language needs something to talk *about*. What are our variables x and y? What is our domain?

Chapter 3 (Set Theory) will answer this. We will define the universe of all mathematical objects using a single primitive relation: membership ($\in$).

Chapter 3

Set Theory: Building the Mathematical Universe

3.1 Why Sets?

Intuition

We've learned logic—how to reason correctly. Now we need *objects* to reason about. What are the basic building blocks of mathematics? Numbers? Functions? Shapes? The answer might surprise you: **everything is a set**.
Numbers are sets. Functions are sets. Even the operations we perform ($+$, $\times$, etc.) are sets. Set theory is the "assembly language" of mathematics—everything reduces to it.

3.1.1 What Is a Set? (Naively)

Let's start informally before we encounter problems.

Definition 3.1 (Naive Definition—FLAWED!)**.** *A **set** is a collection of distinct objects, called its **elements** or **members**.*

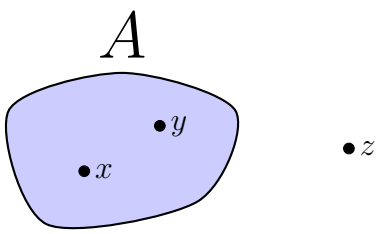
Example 3.1 (Familiar Sets)**.**

- $\{1, 2, 3\}$ = *the set containing the numbers 1, 2, and 3*
- $\{a, b, c\}$ = *the set of the first three letters*
- $\{red, blue, green\}$ = *a set of colors*
- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ = *the natural numbers*
- $\emptyset = \{\}$ = *the empty set (no elements)*

3.1.2 Basic Notation

We write $x \in A$ to mean "x is an element of set A."

We write $x \notin A$ to mean "x is not an element of set A."



$x \in A, \quad y \in A, \quad z \notin A$

3.1.3 Key Properties of Sets

From our naive definition, sets have these properties:

1. **Order doesn't matter**: $\{1, 2, 3\} = \{3, 2, 1\} = \{2, 1, 3\}$
2. **Repetition doesn't matter**: $\{1, 2, 2, 3, 3, 3\} = \{1, 2, 3\}$
3. **Elements are distinct**: A set either contains an element or it doesn't—no multiples
4. **Membership is definite**: For any object x and set A, either $x \in A$ or $x \notin A$ (no ambiguity)

Example 3.2 (Order and Repetition)**.**

$$\boxed{\{a, b, c\}} \quad = \quad \boxed{\{c, b, a\}} \quad = \quad \boxed{\{a, a, b, c, c\}}$$

All represent the same set

3.2 The Crisis: Russell's Paradox

Everything seems fine so far. But there's a catastrophic problem with our naive definition!

Historical Context

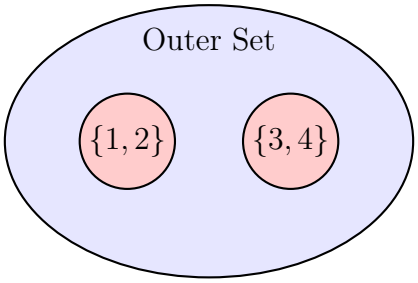
In 1901, Bertrand Russell discovered a devastating paradox that showed naive set theory was **inconsistent**—it led to contradictions.
This was like discovering a crack in the foundation of a skyscraper. Mathematicians had to tear down and rebuild the foundations of mathematics.

3.2.1 Sets Can Contain Sets

First, note that sets can contain other sets as elements:

Example 3.3.

- $\{\{1, 2\}, \{3, 4\}\}$ = *a set containing two sets*
- $\{1, \{2, 3\}\}$ = *a set containing a number and a set*
- $\{\emptyset\}$ = *a set containing the empty set (NOT the same as $\emptyset$!)*



$\{\{1,2\},\{3,4\}\}$ — a set of sets

Question: Can a set contain *itself*?

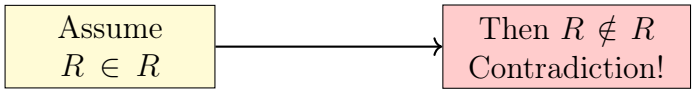
3.2.2 The Paradox

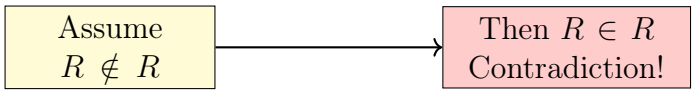
Theorem 3.1 (Russell's Paradox)**.** *Consider the "set" $R = \{x \mid x \notin x\}$ ("the set of all sets that don't contain themselves").*

Question: *Is $R \in R$?*

- ***If*** *$R \in R$: Then R contains itself. But by definition of R, it only contains sets that DON'T contain themselves. So $R \notin R$. Contradiction!*
- ***If*** *$R \notin R$: Then R doesn't contain itself. But that's exactly the criterion for being in R! So $R \in R$. Contradiction!*

Both possibilities lead to contradiction. Therefore, R cannot exist!





No consistent answer!
Naive set theory is broken.

Key Idea

Russell's Paradox shows that we **cannot** allow arbitrary collections to be sets. The phrase "the set of all sets that..." is too permissive.
We need **axioms** that carefully restrict which collections are sets, avoiding contradictions.

Classes versus Sets

Russell's Paradox reveals a fundamental distinction that must be made precise:

- **Sets** are the objects of study in ZFC set theory. They can be members of other sets. The axioms of ZFC carefully control which collections form sets.
- **Proper Classes** are collections "too large" to be sets. For example:
  - The collection of all sets (the "universal class")
  - Russell's collection $\{x \mid x \notin x\}$
  - The collection of all ordinal numbers (defined later)

  Classes can be described by formulas but **cannot be members** of other collections in ZFC.

Why ZFC Avoids the Paradox:
In ZFC, we don't have a "set of all sets." When we write $\{x \mid \varphi(x)\}$, this notation only makes sense when restricted by the Axiom of Separation (or Replacement), which forms sets from existing sets, never from "all objects."
For instance, the Axiom of Separation allows us to form

$$\{x \in A \mid x \notin x\}$$

for any set A, but not the unrestricted collection $\{x \mid x \notin x\}$.
Alternative Foundation: NBG Set Theory
Some authors use von Neumann–Bernays–Gödel (NBG) set theory, which explicitly includes both sets and proper classes as formal objects. In NBG:

- Sets are classes that can be members of other classes
- Proper classes exist but cannot be members of anything
- Russell's collection $\{x \mid x \notin x\}$ is a proper class

NBG and ZFC are equivalent in power for statements about sets (NBG is a "conservative extension"), but NBG makes the class/set distinction explicit in the language. In this text, we work within ZFC, where classes are informal shorthand for properties defined by formulas, not formal objects in the theory.

3.2.3 The Barber Paradox (Analogy)

Russell gave a famous analogy:

> *In a village, the barber shaves all men who don't shave themselves (and only those men). Who shaves the barber?*

- If the barber shaves himself, then he's a man who shaves himself, so he shouldn't be shaved by the barber (himself). Contradiction!

- If the barber doesn't shave himself, then he's a man who doesn't shave himself, so he should be shaved by the barber (himself). Contradiction!

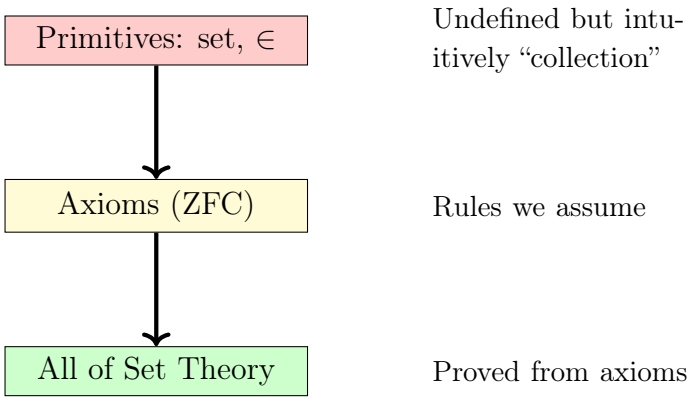
The resolution: **such a barber cannot exist**. Similarly, the set R cannot exist.

3.3 The Solution: Axiomatic Set Theory

3.3.1 The Axiomatic Approach

Instead of defining what a set "is" (which led to paradoxes), we'll:

1. Start with **primitive notions**: "set" and "membership" ($\in$) are *undefined*
2. State **axioms**: rules that govern how sets behave
3. **Derive everything** from these axioms using logic



Intuition

We're not asking "what IS a set?" We're saying: "Here are the rules sets follow. Anything obeying these rules is a set."
This is like defining chess pieces not by what they look like, but by how they move.

3.3.2 The Language: First-Order Logic with $\in$

Our formal language has:

- **Variables**: $x, y, z, \ldots$ (ranging over sets)
- **Logical symbols**: $\forall, \exists, \neg, \wedge, \vee, \implies, \iff$
- **Equality**: $=$
- **Membership**: $\in$ (our only non-logical symbol!)

Definition 3.2 (Bounded Quantifiers)**.** *We frequently use the following shorthand notation for quantifiers restricted to a set A:*

- $\forall x \in A, \varphi(x)$ *is shorthand for* $\forall x(x \in A \implies \varphi(x))$
- $\exists x \in A, \varphi(x)$ *is shorthand for* $\exists x(x \in A \wedge \varphi(x))$

Remark

Everything—unions, intersections, functions, numbers—will be **defined in terms of** $\in$. The membership relation is all we need!

3.4 The Zermelo-Fraenkel Axioms

Now we present the axioms. Each axiom will:

1. Be stated formally
2. Be explained intuitively
3. Be illustrated with examples

3.4.1 Axiom 1: Extensionality

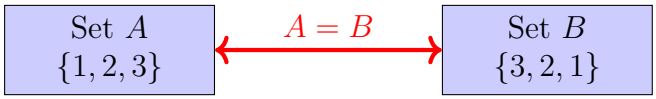
Axiom 3.1 (Extensionality)**.**

$$\forall x \forall y \left[\forall z(z \in x \iff z \in y) \implies x = y\right]$$

Intuition

In words: Two sets are equal if and only if they have exactly the same elements. Sets are determined ONLY by their elements—not by how they're described or constructed.

Example 3.4.
- $\{1, 2, 3\} = \{3, 2, 1\}$ *(same elements, different order)*
- $\{x \mid x^2 = 4\} = \{-2, 2\}$ *(same elements, different descriptions)*



Same elements $\implies$ Same set

Consequence: To prove $A = B$, show they have the same elements:

$$\forall z(z \in A \iff z \in B)$$

Often we prove this by showing $A \subseteq B$ and $B \subseteq A$.

3.4.2 Axiom 2: Empty Set

Axiom 3.2 (Empty Set)**.**

$$\exists x \forall y(y \notin x)$$

Intuition

In words: There exists a set with no elements.
This is the **empty set**, denoted $\emptyset$ or $\{\}$.

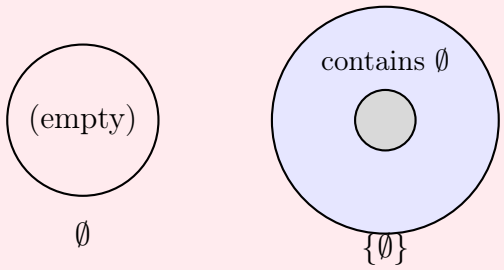
Theorem 3.2 (Uniqueness)**.** *There is exactly one empty set.*

Proof. Suppose $\emptyset_1$ and $\emptyset_2$ are both empty sets.
For any z: $z \notin \emptyset_1$ and $z \notin \emptyset_2$ (both are empty).
Therefore: $z \in \emptyset_1 \iff z \in \emptyset_2$ (both are always false).
By Extensionality: $\emptyset_1 = \emptyset_2$. ✓ ■

Warning

Common confusion:

- $\emptyset$ = the empty set (no elements)
- $\{\emptyset\}$ = a set containing the empty set (one element!)
- These are **different**!



3.4.3 Axiom 3: Pairing

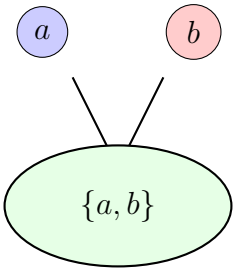
Axiom 3.3 (Pairing)**.**

$$\forall a \forall b \exists c \forall x (x \in c \iff x = a \vee x = b)$$

Intuition

In words: For any two sets a and b, there exists a set $\{a, b\}$ containing exactly those two sets.
This lets us build finite sets!

Example 3.5.
- *From* 1 *and* 2, *we get* $\{1, 2\}$
- *From* a *and* a, *we get* $\{a, a\} = \{a\}$ *(singleton set)*
- *From* $\emptyset$ *and* $\emptyset$, *we get* $\{\emptyset\}$



Pairing axiom creates this set

Building up:

$$\{\emptyset\} \text{ exists (from Pairing with } a = b = \emptyset)$$
$$\{\emptyset, \{\emptyset\}\} \text{ exists (from Pairing)}$$
$$\vdots$$

3.4.4 Axiom 4: Union

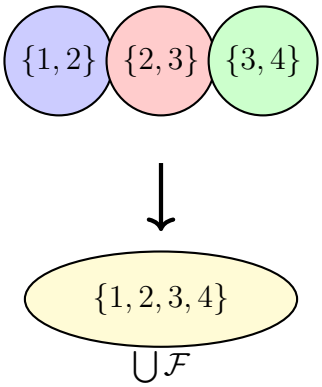
Axiom 3.4 (Union)**.**

$$\forall \mathcal{F} \exists A \forall x (x \in A \iff \exists Y \in \mathcal{F} (x \in Y))$$

Intuition

In words: Given a collection of sets $\mathcal{F}$, there exists a set A containing all elements that belong to at least one set in $\mathcal{F}$.
We write $A = \bigcup \mathcal{F}$ ("the union of $\mathcal{F}$").

Example 3.6. *Let* $\mathcal{F} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$
Then $\bigcup \mathcal{F} = \{1, 2, 3, 4\}$ *(all elements from all sets)*



Binary union: For sets A and B:

$$A \cup B := \bigcup \{A, B\}$$

This is the usual "union" operation!

3.4.5 Axiom 5: Power Set

Axiom 3.5 (Power Set)**.**

$$\forall X \exists P \forall Y (Y \in P \iff Y \subseteq X)$$

Intuition

In words: For any set X, there exists a set P containing all subsets of X.
We write $P = \mathcal{P}(X)$ or $P = 2^X$ ("the power set of X").

First, what's a subset?

Definition 3.3 (Subset)**.**

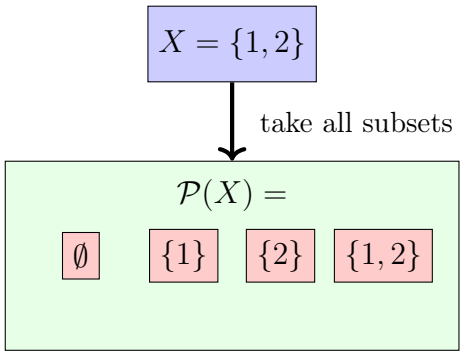
$$A \subseteq B \iff \forall x (x \in A \implies x \in B)$$

"A is a subset of B" means every element of A is also in B.

Example 3.7. *Let* $X = \{1, 2\}$

Subsets of X*:* $\emptyset, \{1\}, \{2\}, \{1, 2\}$

Therefore: $\mathcal{P}(X) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$



4 subsets $\implies |\mathcal{P}(X)| = 4 = 2^2$

Theorem 3.3 (Cardinality of Power Set)**.** *If* $|X| = n$ *(finite set with* n *elements), then* $|\mathcal{P}(X)| = 2^n$*.*

Intuition

For each element of X, you have a binary choice: include it in a subset or not.
n elements $\times$ 2 choices each $= 2^n$ total subsets.

3.4.6 Axiom 6: Separation (Specification)

Axiom 3.6 (Separation Schema)**.** *For any formula* $\phi(x)$ *and set* A*:*

$$\exists B \forall x (x \in B \iff x \in A \wedge \phi(x))$$

Intuition

In words: From any *existing* set A, we can form the subset of elements satisfying a property ϕ.
We write: $B = \{x \in A \mid \phi(x)\}$
Warning on Notation: In naive set theory, we often write $\{x \mid \phi(x)\}$. This is dangerous! It implies we are collecting objects from the entire universe. As Russell's Paradox showed, this leads to contradictions.
In axiomatic set theory, we must always specify **where** the elements come from: $\{x \in A \mid \dots\}$. We can only chip away from a block of marble (set A) that we already have; we cannot build a statue out of thin air.

Key Idea

This axiom **prevents Russell's Paradox**!
We can form $\{x \in A \mid x \notin x\}$ for any set A. This is well-defined and causes no contradiction.
But we **cannot** form $\{x \mid x \notin x\}$ (the problematic R from Russell's Paradox). There's no universal set to filter from!

Example 3.8.
- *From* $\mathbb{N}$, *get* $\{x \in \mathbb{N} \mid x \text{ is even}\} = \{0, 2, 4, 6, \dots\}$
- *From* $\{1, 2, 3, 4, 5\}$, *get* $\{x \in \{1, 2, 3, 4, 5\} \mid x > 3\} = \{4, 5\}$

3.4.7 Axiom 7: Infinity

All axioms so far could be satisfied by finite sets only. We need infinity!

Axiom 3.7 (Infinity)**.**

$$\exists I\,[\emptyset \in I \wedge \forall x(x \in I \implies x \cup \{x\} \in I)]$$

Intuition

In words: There exists a set I containing:

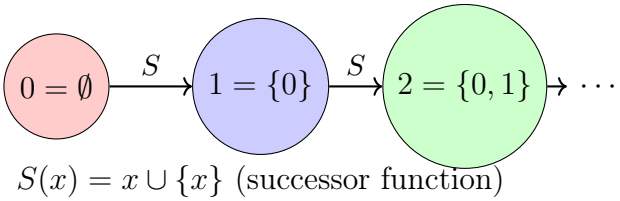
- The empty set $\emptyset$
- For every $x \in I$, also $x \cup \{x\}$ (the "successor" of x)

This set must be infinite!

Building natural numbers:

$$\begin{aligned}
0 &:= \emptyset = \{\} \\
1 &:= 0 \cup \{0\} = \{\emptyset\} = \{0\} \\
2 &:= 1 \cup \{1\} = \{0\} \cup \{\{0\}\} = \{0, 1\} \\
3 &:= 2 \cup \{2\} = \{0, 1\} \cup \{\{0, 1\}\} = \{0, 1, 2\} \\
&\vdots \\
n &= \{0, 1, 2, \dots, n-1\}
\end{aligned}$$

This is the **von Neumann construction** of natural numbers!



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

Key Idea

Every natural number is the *set of all smaller natural numbers*!
$3 = \{0, 1, 2\}$ literally contains 0, 1, and 2 as elements.
This seems strange, but it works perfectly and lets us define arithmetic from pure set theory.

3.4.8 Axiom 8: Replacement

On Functional Formulas

The Replacement Axiom uses the notion of a "functional relation" or "functional formula." Before we formally define functions as sets of ordered pairs (Chapter 6), we use this logical notion:
A formula $\phi(x, y)$ is **functional** if for each x, there exists a unique y such that $\phi(x, y)$ holds:

$$\forall x \exists! y \, \phi(x, y)$$

This means: "ϕ assigns to each x exactly one y."
Examples:

- $\phi(x, y) := (y = x \cup \{x\})$ — functional (successor operation)
- $\phi(x, y) := (y \in x)$ — **not** functional (many y can satisfy this)

This is *not* circular: we're defining functions syntactically (as formulas in logic) here, and will later define them semantically (as sets of pairs) in Chapter 6.

Axiom 3.8 (Replacement Schema)**.** *If $\phi(x, y)$ is a functional formula, then the image of any set under ϕ is also a set.*

Formally: For any formula $\phi(x, y)$ where $\forall x \exists! y \, \phi(x, y)$:

$$\forall A \exists B \forall y (y \in B \iff \exists x \in A \, \phi(x, y))$$

Intuition

In words: If you have a set A and a function F, then $F[A] = \{F(x) : x \in A\}$ is also a set.
You can *transform* each element of a set and collect the results.

Example 3.9.

- *Let $A = \{1, 2, 3\}$ and $F(x) = x^2$*
- *Then $F[A] = \{1, 4, 9\}$ is a set (by Replacement)*

3.4.9 Axiom 9: Foundation (Regularity)

Axiom 3.9 (Foundation)**.**

$$\forall x(x \neq \emptyset \implies \exists y \in x(y \cap x = \emptyset))$$

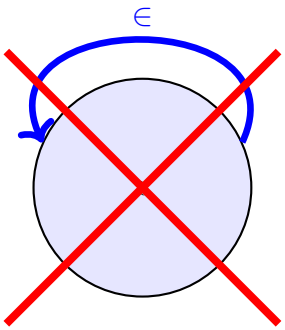
Intuition

In words: Every non-empty set contains an element disjoint from itself.
This prevents pathological situations like:

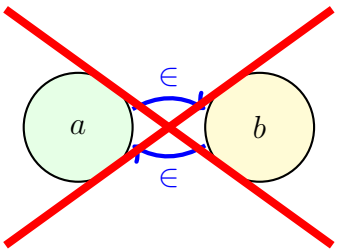
- Sets containing themselves: $x \in x$
- Infinite descending chains: $\cdots \in x_2 \in x_1 \in x_0$

Theorem 3.4. *No set contains itself:* $\forall x(x \notin x)$.

Proof. Suppose $x \in x$. Consider the singleton $\{x\}$.
By Foundation, there exists $y \in \{x\}$ such that $y \cap \{x\} = \emptyset$.
Since $\{x\}$ has only one element, $y = x$.
So $x \cap \{x\} = \emptyset$.
But $x \in x$ (by assumption) and $x \in \{x\}$ (by definition of singleton).
Therefore $x \in x \cap \{x\}$, contradicting $x \cap \{x\} = \emptyset$. $\times$
So $x \notin x$ for all x. ✓ ■



$x \in x$ forbidden!



$a \in b \in a$ forbidden!

3.4.10 Axiom 10: Choice

Axiom 3.10 (Axiom of Choice)**.** *For any family of non-empty, pairwise disjoint sets, there exists a set containing exactly one element from each set in the family.*

Intuition

In words: If you have a collection of boxes, each containing objects, you can simultaneously pick one object from each box.
Sounds obvious? It's surprisingly powerful (and controversial)!

Example 3.10. *Suppose you have infinitely many pairs of shoes. You can choose the left shoe from each pair (definable rule).*

But if you have infinitely many pairs of identical socks, how do you choose one from each pair? There's no rule! The Axiom of Choice says such a choice exists, even without a rule.

Historical Context

The Axiom of Choice (AC) is **independent** of the other axioms. You can do mathematics with it (ZFC) or without it (ZF).
Consequences of AC:

- ✓ Every vector space has a basis
- ✓ Tychonoff's theorem (topology)
- × Banach-Tarski paradox (a ball can be split and reassembled into two identical balls!)
- × Non-measurable sets exist

Most mathematicians accept AC because its positive consequences outweigh the weird ones. But some constructive mathematicians reject it.

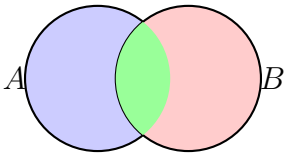
3.5 Building Mathematics from Sets

Now that we have axioms, let's build!

3.5.1 Defining Set Operations

Intersection:

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



$A \cap B$ (green region)

Difference:

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

Complement (relative to universe U):

$$A^c := U \setminus A$$

Warning

There is NO universal set of all sets (this would lead to paradoxes). Complements are always relative to a fixed universe of discourse.

3.5.2 Ordered Pairs: A Clever Construction

To define functions and relations, we need ordered pairs (a, b) where order matters: $(a, b) \neq (b, a)$ unless $a = b$.

But we only have sets! How do we encode order?

Definition 3.4 (Kuratowski Ordered Pair)**.**

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

Intuition

This definition is clever:

- $\{a\}$ tells us what the first element is
- $\{a, b\}$ gives us both elements
- Together, we can recover a and b in order

Theorem 3.5 (Characteristic Property)**.**

$$(a, b) = (c, d) \iff a = c \wedge b = d$$

The proof is tedious but straightforward (consider cases).

Key point: Ordered pairs are just sets! Everything is sets.

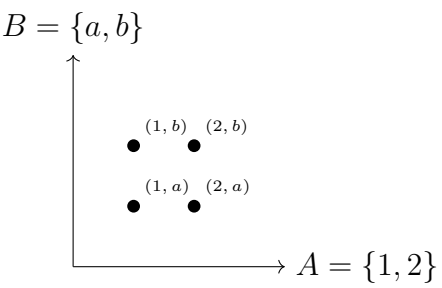
3.5.3 Cartesian Product

Definition 3.5 (Cartesian Product)**.**

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

More precisely: $A \times B := \{z : \exists a \in A \exists b \in B(z = (a, b))\}$

Example 3.11. $\{1, 2\} \times \{a, b\} = \{(1, a), (1, b), (2, a), (2, b)\}$



$A \times B$ has 4 elements

3.6 Looking Forward

We have built the universe of sets and defined the basic operations on them.

Key Idea

What comes next?
Sets are just the raw material. To do mathematics, we need to connect them.

- **Chapter 4 (Relations)**: We will define how elements of sets relate to each other (e.g., "less than", "equivalent to").
- **Chapter 5 (Arithmetic)**: We will use equivalence relations to construct integers and rationals.
- **Chapter 6 (Functions)**: We will define special relations that map inputs to outputs.
- **Chapter 7 (Cardinality)**: We will use functions to measure the size of infinite sets.

You now have the foundations. The mathematical universe is yours to explore.

Chapter 4

Relations: The Architecture of Structure

4.1 Why Relations?

Intuition

We've built sets—collections of objects. Now we need to describe *relationships* between objects.

Is 5 less than 7? Is Paris the capital of France? Is this function continuous? All of these are **relations**—they connect objects and make statements about how they're related.

Relations are the "glue" that creates structure in mathematics. Without them, sets are just unorganized collections. With them, we can build hierarchies, equivalences, functions, and all of mathematics.

4.1.1 From Sets to Structure

Historical Context

Ancient mathematics (Greek geometry, Babylonian algebra) dealt with specific relationships: "equals," "similar to," "divides." But these were treated case-by-case.

The modern concept of a relation as a *set of ordered pairs* emerged in the 19th century with the work of Augustus De Morgan and Charles Peirce, reaching full abstraction in Bourbaki's *Theory of Sets* (1939).

This abstraction allowed mathematicians to study *properties of relationships themselves*—reflexivity, transitivity, etc.—independent of what they relate.

4.2 Relations: Filtering the Universe of Pairs

Intuition

Recall from Chapter 3 that the Cartesian product $A \times B$ contains *all possible* pairs. A relation is a *specific subset*—the pairs that satisfy some property.
Think of it like a filter:

- Universe: all possible connections between A and B
- Relation: the connections that actually hold

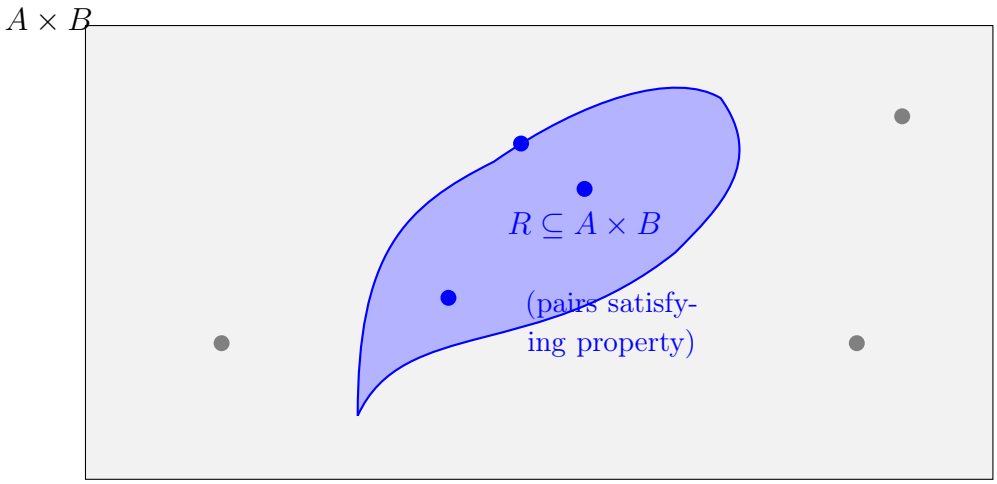
For example, if $A = B = \mathbb{Z}$ and we want the relation "$<$", we select only pairs (a, b) where $a < b$:

$$< := \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a < b\}$$

Definition 4.1 (Binary Relation)**.** *A* ***binary relation*** *from A to B is a subset $R \subseteq A \times B$.*

If $(a, b) \in R$, we write aRb and say "a is related to b by R."

Relation as Subset of $A \times B$



Example 4.1 (Common Relations)**.** *1.* ***Less than on*** $\mathbb{N}$*:*

$$< := \{(m, n) \in \mathbb{N} \times \mathbb{N} : m < n\}$$

2. ***Divisibility****:*

$$\mid := \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a \mid b\}$$

3. ***Subset relation****:*

$$\subseteq := \{(A, B) \in \mathcal{P}(X) \times \mathcal{P}(X) : A \subseteq B\}$$

4. ***Equality****:*

$$=:= \{(x, x) : x \in A\}$$

(The diagonal of $A \times A$)

4.2.1 Domain, Codomain, and Image

Definition 4.2. *Let $R \subseteq A \times B$ be a relation.*

- *The **domain** of R is:*

$$dom(R) := \{a \in A : \exists b \in B, (a, b) \in R\}$$

- *The **image** (or range) of R is:*

$$im(R) := \{b \in B : \exists a \in A, (a, b) \in R\}$$

Example 4.2. *Let $A = \{1, 2, 3\}$, $B = \{a, b, c\}$, and:*

$$R = \{(1, a), (1, b), (3, c)\}$$

Then:

- $dom(R) = \{1, 3\}$ *(element 2 doesn't relate to anything)*
- $im(R) = \{a, b, c\}$ *(all elements of B are reached)*

4.3 Properties of Relations on a Single Set

When $R \subseteq A \times A$ (relation on a set to itself), we can study structural properties.

The Four Fundamental Properties

Reflexive: $\forall x, xRx$ **Symmetric**: $xRy \implies yRx$

Transitive: $(xRy \wedge yRz) \implies xRz$ **Anti-symmetric**: $(xRy \wedge yRx) \implies x = y$

Definition 4.3 (Properties of Relations)**.** *Let R be a relation on A (i.e., $R \subseteq A \times A$).*

1. *R is **reflexive** if:*

$$\forall x \in A, xRx$$

(Every element relates to itself)

2. *R is **symmetric** if:*
$$\forall x, y \in A, (xRy \implies yRx)$$
(Mutual relationships)

3. *R is **transitive** if:*
$$\forall x, y, z \in A, ((xRy \wedge yRz) \implies xRz)$$
(The "chain rule")

4. *R is **anti-symmetric** if:*
$$\forall x, y \in A, ((xRy \wedge yRx) \implies x = y)$$
(No mutual relationships except self-loops)

Example 4.3 (Testing Properties)**.** *Let $A = \{1, 2, 3, 4\}$ and $R = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (2, 3), ($*

- ***Reflexive?** YES: $(1, 1), (2, 2), (3, 3), (4, 4) \in R$*
- ***Symmetric?** NO: $(1, 2) \in R$ but $(2, 1) \notin R$*
- ***Transitive?** YES: $(1, 2) \in R$ and $(2, 3) \in R$ and $(1, 3) \in R$ ✓*
- ***Anti-symmetric?** YES: No mutual pairs except diagonal*

*This is a **partial order** (reflexive + anti-symmetric + transitive).*

Warning

Symmetric vs. Anti-symmetric
These are NOT opposites! A relation can be:

- Both (e.g., equality: $=$)
- Neither (e.g., "loves" relation among people)
- One but not the other

Anti-symmetric does NOT mean "not symmetric." It means: if xRy AND yRx, then x must equal y.

4.4 Equivalence Relations: Generalizing Equality

Intuition

Equality ($=$) is the most basic relation: reflexive, symmetric, transitive.
An equivalence relation generalizes equality—it groups objects that we consider "the same" for some purpose:

- Numbers with the same remainder mod 5

- Geometric figures with the same shape
- Functions with the same limit

Equivalence relations partition a set into disjoint "clusters" of equivalent objects.

Definition 4.4 (Equivalence Relation)**.** *A relation $\sim$ on A is an **equivalence relation** if it is:*

1. *Reflexive:* $\forall x \in A, x \sim x$
2. *Symmetric:* $\forall x, y \in A, (x \sim y \implies y \sim x)$
3. *Transitive:* $\forall x, y, z \in A, ((x \sim y \land y \sim z) \implies x \sim z)$

Example 4.4 (Equivalence Relations)**.** 1. ***Equality**: $x = y$ on any set*

2. ***Congruence modulo** n: On $\mathbb{Z}$, define $a \sim b \iff n \mid (a - b)$*
3. ***Same cardinality**: On sets, $A \sim B \iff |A| = |B|$*
4. ***Parallel lines**: In Euclidean geometry, $\ell_1 \sim \ell_2 \iff \ell_1 \parallel \ell_2$ or $\ell_1 = \ell_2$*

4.4.1 Equivalence Classes

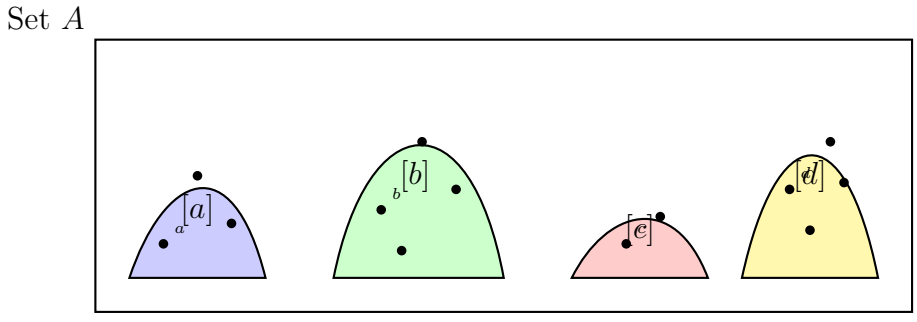
Definition 4.5 (Equivalence Class)**.** *Let $\sim$ be an equivalence relation on A. The **equivalence class** of $x \in A$ is:*

$$[x] := \{y \in A : y \sim x\}$$

*The set of all equivalence classes is called the **quotient set**:*

$$A/\sim := \{[x] : x \in A\}$$

Equivalence Classes Partition the Set



Disjoint, exhaustive partition

Theorem 4.1 (Equivalence Classes Are Disjoint or Identical)**.** *Let $\sim$ be an equivalence relation on A. For any $x, y \in A$:*

$$[x] \cap [y] \neq \emptyset \implies [x] = [y]$$

Equivalently: Either $[x] = [y]$ or $[x] \cap [y] = \emptyset$.

Proof. Suppose $[x] \cap [y] \neq \emptyset$. Then there exists $z \in [x] \cap [y]$.

By definition: $z \sim x$ and $z \sim y$.

We'll show $[x] = [y]$ by proving $[x] \subseteq [y]$ and $[y] \subseteq [x]$.

($[x] \subseteq [y]$): Let $w \in [x]$. Then $w \sim x$.

We have:

- $w \sim x$ (given)
- $x \sim z$ (by symmetry from $z \sim x$)
- $z \sim y$ (given)

By transitivity: $w \sim x$ and $x \sim z$ gives $w \sim z$.

Then $w \sim z$ and $z \sim y$ gives $w \sim y$.

Therefore $w \in [y]$. ✓

($[y] \subseteq [x]$): By symmetry (swapping roles of x and y). ✓

Therefore $[x] = [y]$. ■

Theorem 4.2 (Fundamental Theorem of Equivalence Relations)**.** *Every equivalence relation on A induces a partition of A (disjoint, exhaustive subsets).*

Conversely, every partition of A defines an equivalence relation (elements are equivalent iff they're in the same part).

Proof Sketch. ($\Rightarrow$) Given $\sim$, the equivalence classes form a partition:

- **Exhaustive**: Every $x \in A$ belongs to $[x]$ (by reflexivity)
- **Disjoint**: By previous theorem

($\Leftarrow$) Given partition $\mathcal{P} = \{P_1, P_2, \ldots\}$, define:

$$x \sim y \iff x \text{ and } y \text{ belong to the same } P_i$$

Check this is reflexive, symmetric, transitive (exercise). ■

Example 4.5 (Integers Modulo 5)**.** *Define $a \sim b \iff 5 \mid (a - b)$ on $\mathbb{Z}$.*

This creates 5 equivalence classes:

$$\begin{aligned}
[0] &= \{\ldots, -10, -5, 0, 5, 10, \ldots\} \\
[1] &= \{\ldots, -9, -4, 1, 6, 11, \ldots\} \\
[2] &= \{\ldots, -8, -3, 2, 7, 12, \ldots\} \\
[3] &= \{\ldots, -7, -2, 3, 8, 13, \ldots\} \\
[4] &= \{\ldots, -6, -1, 4, 9, 14, \ldots\}
\end{aligned}$$

The quotient set $\mathbb{Z}/5\mathbb{Z} = \{[0], [1], [2], [3], [4]\}$ is the integers modulo 5.

4.5 Order Relations: Hierarchies and Comparisons

Intuition

While equivalence relations group things as "same," order relations arrange things in a hierarchy: "less than," "subset of," "precedes."
Think of a family tree, a chain of command, or the real numbers with $\leq$.
Not all elements need to be comparable (e.g., sets under $\subseteq$: $\{1,2\}$ and $\{3,4\}$ are incomparable). These are **partial orders**.

Definition 4.6 (Partial Order)**.** *A relation $\preceq$ on A is a **partial order** if it is:*

1. ***Reflexive**: $\forall x \in A, x \preceq x$*
2. ***Anti-symmetric**: $\forall x, y \in A, (x \preceq y \wedge y \preceq x) \implies x = y$*
3. ***Transitive**: $\forall x, y, z \in A, ((x \preceq y \wedge y \preceq z) \implies x \preceq z)$*

*The pair $(A, \preceq)$ is called a **partially ordered set** (poset).*

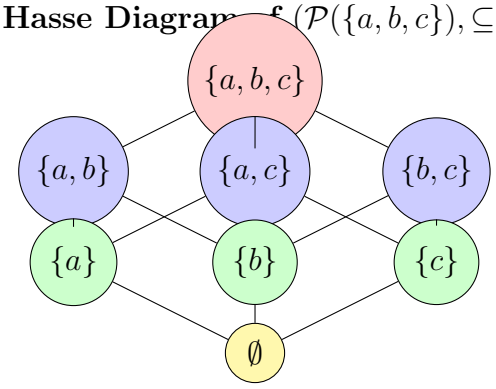
Example 4.6 (Partial Orders)**.**

1. *$(\mathbb{N}, \leq)$ — the usual ordering*
2. *$(\mathcal{P}(X), \subseteq)$ — subset relation on power set*
3. *$(X^X, \leq)$ where $f \leq g \iff \forall x, f(x) \leq g(x)$ (pointwise order on functions)*
4. *Divisibility: $(a, b) \in R \iff a \mid b$ on $\mathbb{N}^+$*

4.5.1 Hasse Diagrams

We visualize posets using **Hasse diagrams**: elements are vertices, and $x \prec y$ (covers) is shown by y above x with an edge.

Hasse Diagram of $(\mathcal{P}(\{a,b,c\}), \subseteq)$



Height represents subset relation (read upward)

Definition 4.7 (Total Order)**.** *A partial order $\preceq$ on A is a **total order** (or linear order) if:*

$$\forall x, y \in A, (x \preceq y \vee y \preceq x)$$

Every pair of elements is comparable.

Example 4.7.

- *$(\mathbb{R}, \leq)$ is a total order*
- *$(\mathcal{P}(\{1,2\}), \subseteq)$ is NOT: $\{1\}$ and $\{2\}$ are incomparable*

4.5.2 Special Elements in Posets

Definition 4.8 (Maximal and Minimal Elements)**.** *Let $(A, \preceq)$ be a poset.*

- $m \in A$ *is* ***maximal*** *if:* $\forall x \in A, (m \preceq x \implies m = x)$
  (Nothing is strictly greater than m)
- $m \in A$ *is* ***minimal*** *if:* $\forall x \in A, (x \preceq m \implies x = m)$
  (Nothing is strictly less than m)

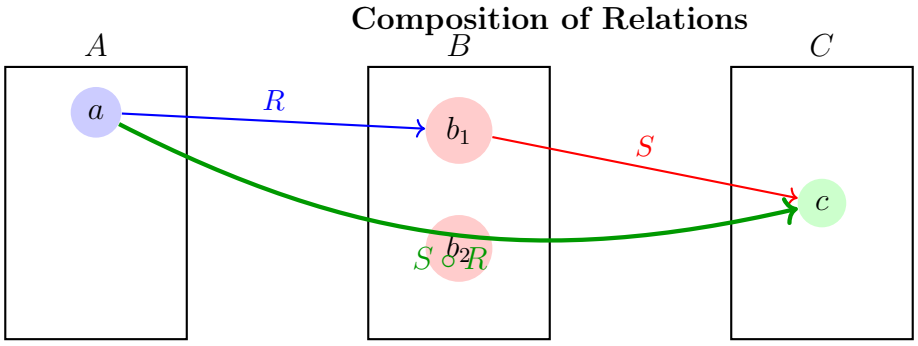
Warning

Maximal $\neq$ Maximum!
In a partial order, there can be multiple maximal elements (incomparable with each other).
A **maximum** (or greatest element) M satisfies: $\forall x \in A, x \preceq M$.
Example: In $(\{1, 2, 3, 4\}, |)$ (divisibility), both 3 and 4 are maximal, but there's no maximum.

4.6 Composition of Relations

Definition 4.9 (Composition)**.** *Let $R \subseteq A \times B$ and $S \subseteq B \times C$ be relations. The* ***composition*** *$S \circ R$ is:*

$$S \circ R := \{(a, c) \in A \times C : \exists b \in B, (a, b) \in R \wedge (b, c) \in S\}$$



$$(a, b_1) \in R \text{ and } (b_1, c) \in S$$
$$\implies (a, c) \in S \circ R$$

Theorem 4.3 (Composition is Associative)**.** *Let $R \subseteq A \times B$, $S \subseteq B \times C$, $T \subseteq C \times D$. Then:*

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

Proof. We show both sets contain the same ordered pairs.

$(a, d) \in T \circ (S \circ R)$
$\iff \exists c \in C, ((a, c) \in S \circ R \wedge (c, d) \in T)$

$$\iff \exists c \in C, ((\exists b \in B, (a,b) \in R \wedge (b,c) \in S) \wedge (c,d) \in T)$$
$$\iff \exists b \in B \exists c \in C, ((a,b) \in R \wedge (b,c) \in S \wedge (c,d) \in T)$$
$$\iff \exists b \in B, ((a,b) \in R \wedge (\exists c \in C, (b,c) \in S \wedge (c,d) \in T))$$
$$\iff \exists b \in B, ((a,b) \in R \wedge (b,d) \in T \circ S)$$
$$\iff (a,d) \in (T \circ S) \circ R$$

Therefore the compositions are equal. ■

4.7 Looking Forward

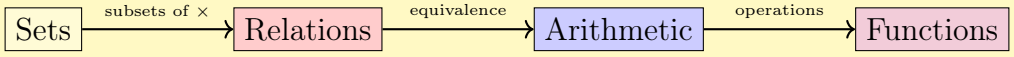
We have defined the general concept of a relation. Two types of relations are particularly important for the foundations of mathematics:

1. **Equivalence Relations**: These allow us to construct new mathematical objects by gluing existing ones together. We will use this in the next chapter to build integers and rationals.

2. **Functions**: These are relations that behave like "machines"—one input, one output.

Key Idea

The Big Picture:



Next, we will use equivalence relations to construct the number systems $\mathbb{Z}$ and $\mathbb{Q}$.

Chapter 5

Arithmetic: The Construction of Number Systems

5.1 From Sets to Numbers

Intuition

We've built natural numbers as sets:

$$0 = \emptyset, \quad 1 = \{0\}, \quad 2 = \{0, 1\}, \quad 3 = \{0, 1, 2\}, \ldots$$

But what does "$2 + 3$" mean? What does "2×3" mean?
These operations aren't *given*—we must **define** them rigorously from scratch.
This chapter constructs the familiar number systems $(\mathbb{N}, \mathbb{Z}, \mathbb{Q})$ and proves all their arithmetic properties from set-theoretic foundations.

Historical Context

The Arithmetization of Mathematics
Before the 19th century, arithmetic was considered self-evident. But a crisis emerged:
Ancient Greeks: Discovered irrational numbers like $\sqrt{2}$, causing philosophical turmoil.
19th Century Crisis:

- Analysts used real numbers freely but couldn't define them rigorously
- Dedekind (1858): "What are numbers and what should they be?"
- Dedekind cuts (1872): Constructed $\mathbb{R}$ from $\mathbb{Q}$
- Cantor (1872): Alternative construction via Cauchy sequences
- Peano (1889): Axiomatized natural numbers
- Frege, Russell: Attempted to reduce arithmetic to pure logic (Logicism)

Zermelo-Fraenkel Set Theory (1908-1922): Provided the ultimate foundation—all numbers are sets, all operations are functions (which are sets). Today's approach: Define $\mathbb{N}$ via von Neumann ordinals, then construct $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ as successive extensions.

5.2 Arithmetic on Natural Numbers

We've defined natural numbers using the Axiom of Infinity:

$$0 := \emptyset, \quad S(n) := n \cup \{n\} \quad \text{(successor function)}$$

Now we define addition and multiplication.

5.2.1 The Principle of Mathematical Induction

Before defining operations, we must establish our primary tool for proving statements about natural numbers: **Induction**.

Recall that $\mathbb{N}$ is defined as the smallest inductive set (by the Axiom of Infinity). This gives us the following principle:

Theorem 5.1 (Principle of Mathematical Induction)**.** *Let $P(n)$ be a property involving a natural number n. If:*

1. ***Base Case**: $P(0)$ is true, and*
2. ***Inductive Step**: For all $k \in \mathbb{N}$, if $P(k)$ is true, then $P(S(k))$ is true,*

then $P(n)$ is true for all $n \in \mathbb{N}$.

Intuition

This works like dominoes:

- Base case: You knock over the first domino (0).
- Inductive step: Each domino knocks over the next one ($k \implies S(k)$).
- Conclusion: All dominoes fall.

Because $\mathbb{N}$ contains *only* elements reached this way (it's the *smallest* inductive set), the property holds for all numbers.

5.2.2 Addition

Definition 5.1 (Addition on $\mathbb{N}$)**.** *For $m, n \in \mathbb{N}$, define $m + n$ recursively:*

***Base case**: $m + 0 := m$*

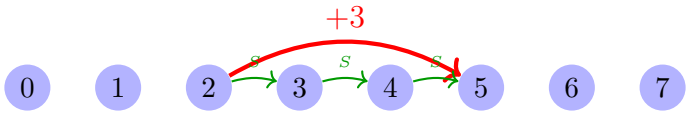
***Recursive case**: $m + S(n) := S(m + n)$*

In other notation: $m + (n + 1) = (m + n) + 1$

Example 5.1. *Let's compute* $2+3$ *from the definition:*

$$\begin{aligned}
2+3 &= 2+S(2) \\
&= S(2+2) \quad \textit{(by recursive case)} \\
&= S(2+S(1)) \\
&= S(S(2+1)) \\
&= S(S(2+S(0))) \\
&= S(S(S(2+0))) \\
&= S(S(S(2))) \quad \textit{(by base case)} \\
&= S(S(3)) \\
&= S(4) \\
&= 5
\end{aligned}$$

Addition as Repeated Successor



$2+3$ = "apply successor 3 times starting from 2"
$=$ $S(S(S(2)))$ $=$ 5

Theorem 5.2 (Properties of Addition)**.** *For all* $m, n, p \in \mathbb{N}$*:*

1. ***Right Identity****:* $n+0=n$
2. ***Left Identity****:* $0+n=n$
3. ***Commutativity****:* $m+n=n+m$
4. ***Associativity****:* $(m+n)+p=m+(n+p)$

Proof. **(1) Right Identity**: By definition, $n+0=n$. ✓
(2) Left Identity: Prove by induction on n.
Base case: $0+0=0$ (by definition). ✓
Inductive step: Assume $0+n=n$ (IH). Show $0+S(n)=S(n)$.

$$\begin{aligned}
0+S(n) &= S(0+n) \quad \text{(by definition of addition)} \\
&= S(n) \quad \text{(by IH)}
\end{aligned}$$

Therefore $0+S(n)=S(n)$. By induction, $0+n=n$ for all n. ✓

(3) Commutativity: First prove a lemma.

Lemma: $S(m) + n = S(m + n)$ for all m, n.

Proof of Lemma: Induction on n.

Base: $S(m) + 0 = S(m) = S(m + 0)$. ✓

Step: Assume $S(m) + n = S(m + n)$ (IH).

$$\begin{aligned} S(m) + S(n) &= S(S(m) + n) && \text{(definition)} \\ &= S(S(m + n)) && \text{(IH)} \\ &= S(m + S(n)) && \text{(definition)} \end{aligned}$$

Lemma proved. ✓

Now prove commutativity by induction on n.

Base: $m + 0 = m = 0 + m$ (by right and left identity). ✓

Step: Assume $m + n = n + m$ (IH). Show $m + S(n) = S(n) + m$.

$$\begin{aligned} m + S(n) &= S(m + n) && \text{(definition)} \\ &= S(n + m) && \text{(IH)} \\ &= S(n) + m && \text{(Lemma)} \end{aligned}$$

Therefore addition is commutative. ✓

(4) Associativity: Prove by induction on p.

Base: $(m + n) + 0 = m + n = m + (n + 0)$. ✓

Step: Assume $(m + n) + p = m + (n + p)$ (IH).

$$\begin{aligned} (m + n) + S(p) &= S((m + n) + p) && \text{(definition)} \\ &= S(m + (n + p)) && \text{(IH)} \\ &= m + S(n + p) && \text{(definition)} \\ &= m + (n + S(p)) && \text{(definition)} \end{aligned}$$

Therefore addition is associative. ✓ ■

5.2.3 Multiplication

Definition 5.2 (Multiplication on $\mathbb{N}$)**.** *For $m, n \in \mathbb{N}$, define $m \cdot n$ (or $m \times n$) recursively:*

***Base case**: $m \cdot 0 := 0$*

***Recursive case**: $m \cdot S(n) := m \cdot n + m$*

In other notation: $m \cdot (n + 1) = m \cdot n + m$

Key Idea

Multiplication is **repeated addition**:

$$m \cdot n = \underbrace{m + m + \cdots + m}_{n \text{ times}}$$

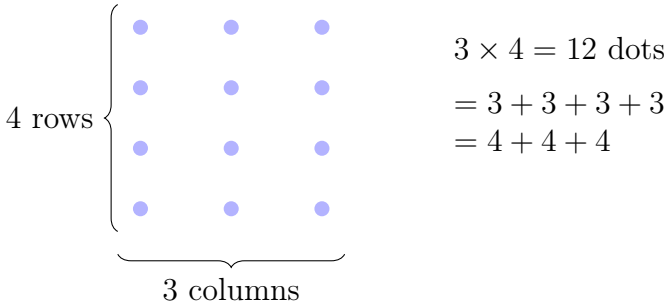
For example:

$$3 \cdot 4 = 3 + 3 + 3 + 3 = 12$$

Example 5.2. *Compute* $3 \cdot 2$ *from the definition:*

$$\begin{aligned} 3 \cdot 2 &= 3 \cdot S(1) \\ &= 3 \cdot 1 + 3 \\ &= 3 \cdot S(0) + 3 \\ &= (3 \cdot 0 + 3) + 3 \\ &= (0 + 3) + 3 \quad \textit{(by base case)} \\ &= 3 + 3 \\ &= 6 \end{aligned}$$

Multiplication as Repeated Addition



Theorem 5.3 (Properties of Multiplication)**.** *For all* $m, n, p \in \mathbb{N}$*:*

1. ***Right Zero****:* $n \cdot 0 = 0$
2. ***Left Zero****:* $0 \cdot n = 0$
3. ***Right Identity****:* $n \cdot 1 = n$
4. ***Left Identity****:* $1 \cdot n = n$
5. ***Commutativity****:* $m \cdot n = n \cdot m$
6. ***Associativity****:* $(m \cdot n) \cdot p = m \cdot (n \cdot p)$
7. ***Distributivity****:* $m \cdot (n + p) = m \cdot n + m \cdot p$

Proof. **(1) Right Zero**: By definition, $n \cdot 0 = 0$. ✓

(2) Left Zero: Induction on n.

Base: $0 \cdot 0 = 0$ (definition). ✓

Step: Assume $0 \cdot n = 0$ (IH).

$$\begin{aligned} 0 \cdot S(n) &= 0 \cdot n + 0 \quad \text{(definition)} \\ &= 0 + 0 \quad \text{(IH)} \\ &= 0 \end{aligned}$$

✓

(3) Right Identity:

$$\begin{aligned} n \cdot 1 &= n \cdot S(0) \\ &= n \cdot 0 + n \\ &= 0 + n \\ &= n \end{aligned}$$

✓

(4) Left Identity: Induction on n.

Base: $1 \cdot 0 = 0$ (by definition). This matches $n = 0$. ✓

Step: Assume $1 \cdot n = n$ (IH). Show $1 \cdot S(n) = S(n)$.

$$\begin{aligned} 1 \cdot S(n) &= 1 \cdot n + 1 \quad \text{(definition)} \\ &= n + 1 \quad \text{(IH)} \\ &= S(n) \end{aligned}$$

Therefore $1 \cdot n = n$ for all n. ✓

(5) Commutativity: First prove lemmas.

Lemma 1: $S(m) \cdot n = m \cdot n + n$

Proof: Induction on n.

Base: $S(m) \cdot 0 = 0 = 0 + 0 = m \cdot 0 + 0$. ✓

Step: Assume $S(m) \cdot n = m \cdot n + n$ (IH).

$$\begin{aligned} S(m) \cdot S(n) &= S(m) \cdot n + S(m) \quad \text{(def)} \\ &= (m \cdot n + n) + S(m) \quad \text{(IH)} \\ &= m \cdot n + (n + S(m)) \\ &= m \cdot n + (S(m) + n) \quad \text{(commutativity of } +) \\ &= m \cdot n + (m + S(n)) \\ &= (m \cdot n + m) + S(n) \\ &= m \cdot S(n) + S(n) \end{aligned}$$

Lemma 1 proved. ✓

Now prove commutativity by induction on n.

Base: $m \cdot 0 = 0 = 0 \cdot m$ (by left and right zero). ✓

Step: Assume $m \cdot n = n \cdot m$ (IH).

$$\begin{aligned} m \cdot S(n) &= m \cdot n + m \quad \text{(def)} \\ &= n \cdot m + m \quad \text{(IH)} \\ &= S(n) \cdot m \quad \text{(Lemma 1)} \end{aligned}$$

✓

(6) Associativity: Prove by induction on p.

Base: $(m \cdot n) \cdot 0 = 0 = m \cdot 0 = m \cdot (n \cdot 0)$. ✓

Step: Assume $(m \cdot n) \cdot p = m \cdot (n \cdot p)$ (IH).

$$\begin{aligned}(m \cdot n) \cdot S(p) &= (m \cdot n) \cdot p + (m \cdot n) \quad \text{(def)} \\ &= m \cdot (n \cdot p) + (m \cdot n) \quad \text{(IH)} \\ &= m \cdot (n \cdot p + n) \quad \text{(distributivity, proved next)} \\ &= m \cdot (n \cdot S(p))\end{aligned}$$

✓

(7) Distributivity: Prove by induction on p.

Base: $m \cdot (n + 0) = m \cdot n = m \cdot n + 0 = m \cdot n + m \cdot 0$. ✓

Step: Assume $m \cdot (n + p) = m \cdot n + m \cdot p$ (IH).

$$\begin{aligned}m \cdot (n + S(p)) &= m \cdot S(n + p) \\ &= m \cdot (n + p) + m \quad \text{(def)} \\ &= (m \cdot n + m \cdot p) + m \quad \text{(IH)} \\ &= m \cdot n + (m \cdot p + m) \quad \text{(associativity of } +) \\ &= m \cdot n + m \cdot S(p) \quad \text{(def)}\end{aligned}$$

✓ ■

Remark

These proofs are tedious but essential—we've just proved that $\mathbb{N}$ with $+$ and $\cdot$ satisfies the axioms of a **commutative semiring**.
The structure $(\mathbb{N}, +, \cdot, 0, 1)$ is the **free** commutative semiring on one generator.

5.3 The Integers: $\mathbb{Z}$

Intuition

Natural numbers are insufficient: the equation $x + 3 = 2$ has no solution in $\mathbb{N}$.
We need **negative numbers** to solve equations like $x + a = b$ for any a, b.
How do we construct negatives from sets? We can't just "add them"—we must build them systematically.

5.3.1 Construction of $\mathbb{Z}$

Definition 5.3 (Integers as Pairs)**.** *Define an equivalence relation on $\mathbb{N} \times \mathbb{N}$:*

$$(m, n) \sim (p, q) \iff m + q = p + n$$

*The **integers** are the equivalence classes:*

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\sim$$

We write $[(m, n)]$ for the equivalence class of (m, n).

Key Idea

Interpretation: The pair (m, n) represents the "difference" $m - n$.

- $[(3, 0)]$ represents $3 - 0 = 3$ (positive)
- $[(0, 5)]$ represents $0 - 5 = -5$ (negative)
- $[(7, 4)]$ represents $7 - 4 = 3$ (same as $[(3, 0)]$)
- $[(2, 2)]$ represents $2 - 2 = 0$ (zero)

The equivalence relation says: $(m, n) \sim (p, q)$ if $m - n = p - q$ (informally).
Formally: $m + q = p + n$ (avoiding subtraction, which we haven't defined yet!)

Theorem 5.4. *The relation $\sim$ is an equivalence relation.*

Proof. **Reflexive**: $(m, n) \sim (m, n)$ because $m + n = m + n$. ✓

Symmetric: If $(m, n) \sim (p, q)$, then $m + q = p + n$, so $p + n = m + q$, thus $(p, q) \sim (m, n)$. ✓

Transitive: If $(m, n) \sim (p, q)$ and $(p, q) \sim (r, s)$, then:

$$m + q = p + n \quad \text{and} \quad p + s = r + q$$

Adding these equations:

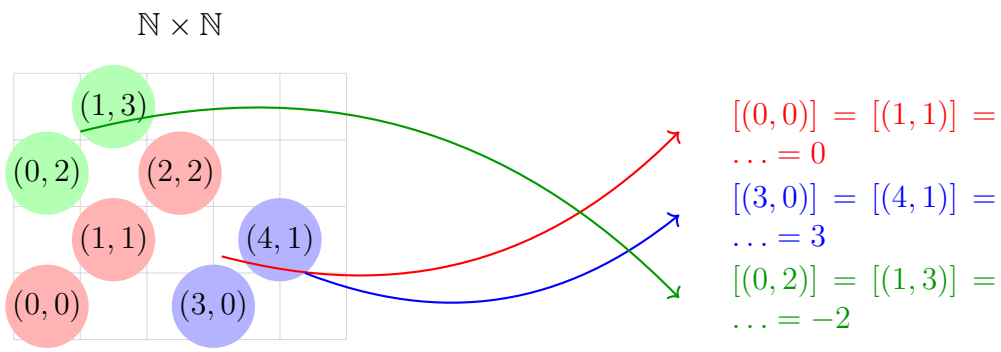
$$m + q + p + s = p + n + r + q$$

Cancel p and q (using cancellation law for natural numbers):

$$m + s = r + n$$

Therefore $(m, n) \sim (r, s)$. ✓ ■

Integers as Equivalence Classes of Pairs



5.3.2 Arithmetic on $\mathbb{Z}$

Definition 5.4 (Operations on $\mathbb{Z}$)**.** *Define addition and multiplication on equivalence classes:*

***Addition**:* $[(m,n)] + [(p,q)] := [(m+p, n+q)]$

***Multiplication**:* $[(m,n)] \cdot [(p,q)] := [(mp+nq, mq+np)]$

***Negation**:* $-[(m,n)] := [(n,m)]$

***Embedding**:* $\iota : \mathbb{N} \to \mathbb{Z}$ *by* $\iota(n) = [(n,0)]$

Theorem 5.5 (The Embedding $\mathbb{N} \hookrightarrow \mathbb{Z}$)**.** *The map $\iota : \mathbb{N} \to \mathbb{Z}$ defined by $\iota(n) = [(n,0)]$ is an injective homomorphism that preserves addition, multiplication, and order. This allows us to view $\mathbb{N}$ as a subset of $\mathbb{Z}$.*

Proof. We verify the required properties:

(1) Well-defined: For any $n \in \mathbb{N}$, we have $\iota(n) = [(n,0)] \in \mathbb{Z}$ (an equivalence class). ✓

(2) Injective: Suppose $\iota(m) = \iota(n)$ for $m, n \in \mathbb{N}$.

Then $[(m,0)] = [(n,0)]$, which means $(m,0) \sim (n,0)$.

By definition of $\sim$, this means $m + 0 = n + 0$, so $m = n$. ✓

(3) Preserves Addition:

$$\begin{aligned} \iota(m+n) &= [(m+n, 0)] \\ &= [(m,0)] + [(n,0)] \quad \text{(by definition of } + \text{ on } \mathbb{Z}) \\ &= \iota(m) + \iota(n) \end{aligned}$$

✓

(4) Preserves Multiplication:

$$\begin{aligned} \iota(m \cdot n) &= [(mn, 0)] \\ &= [(m \cdot n + 0 \cdot 0, m \cdot 0 + 0 \cdot n)] \\ &= [(m,0)] \cdot [(n,0)] \quad \text{(by definition of } \cdot \text{ on } \mathbb{Z}) \\ &= \iota(m) \cdot \iota(n) \end{aligned}$$

✓

(5) Preserves Order: Recall that on $\mathbb{N}$, we have $m < n$ iff $\exists k \in \mathbb{N}, m + k = n$ with $k \neq 0$.

On $\mathbb{Z}$, we define $[(a,b)] < [(c,d)]$ iff $a + d < b + c$ in $\mathbb{N}$.

Now suppose $m < n$ in $\mathbb{N}$, so $m + k = n$ for some $k > 0$.

Then:

$$\iota(m) = [(m,0)] \quad \text{and} \quad \iota(n) = [(n,0)] = [(m+k, 0)]$$

To check $[(m,0)] < [(m+k,0)]$ in $\mathbb{Z}$:

$$m + 0 < 0 + (m+k) = m + k \quad \text{in } \mathbb{N}$$

This holds since $k > 0$. ✓

Conversely, if $\iota(m) < \iota(n)$, then $[(m, 0)] < [(n, 0)]$, so $m + 0 < 0 + n$, hence $m < n$ in $\mathbb{N}$. ✓

Therefore ι is an order-preserving ring homomorphism, justifying the identification of $\mathbb{N}$ with $\{[(n, 0)] : n \in \mathbb{N}\} \subseteq \mathbb{Z}$. ■

Warning

We must verify these operations are **well-defined**—they don't depend on the choice of representative!
If $(m, n) \sim (m', n')$ and $(p, q) \sim (p', q')$, we need:

$$[(m, n)] + [(p, q)] = [(m', n')] + [(p', q')]$$

Theorem 5.6. *Addition on $\mathbb{Z}$ is well-defined.*

Proof. Suppose $(m, n) \sim (m', n')$ and $(p, q) \sim (p', q')$.

Then $m + n' = m' + n$ and $p + q' = p' + q$.

We need to show $(m + p, n + q) \sim (m' + p', n' + q')$.

This means proving: $(m + p) + (n' + q') = (m' + p') + (n + q)$.

$$\begin{aligned}(m + p) + (n' + q') &= (m + n') + (p + q') && \text{(rearranging)}\\ &= (m' + n) + (p' + q) && \text{(by assumptions)}\\ &= (m' + p') + (n + q) && \text{(rearranging)}\end{aligned}$$

Therefore addition is well-defined. ✓ ■

Theorem 5.7. *Multiplication on $\mathbb{Z}$ is well-defined.*

Proof. Suppose $(m, n) \sim (m', n')$ and $(p, q) \sim (p', q')$.

Then $m + n' = m' + n$ and $p + q' = p' + q$ in $\mathbb{N}$.

We need to show $(mp + nq, mq + np) \sim (m'p' + n'q', m'q' + n'p')$.

This means proving:

$$(mp + nq) + (m'q' + n'p') = (m'p' + n'q') + (mq + np)$$

We'll use the fact that in $\mathbb{N}$, if $m + n' = m' + n$ and $p + q' = p' + q$, then:

Step 1: Multiply the first equation by p:

$$(m + n')p = (m' + n)p$$

$$mp + n'p = m'p + np$$

Step 2: Multiply the first equation by q:

$$(m + n')q = (m' + n)q$$

$$mq + n'q = m'q + nq$$

Step 3: Multiply the second equation by m':

$$(p + q')m' = (p' + q)m'$$

$$pm' + q'm' = p'm' + qm'$$

Step 4: Multiply the second equation by n':

$$(p + q')n' = (p' + q)n'$$

$$pn' + q'n' = p'n' + qn'$$

Now add Step 1 and Step 4:

$$(mp + n'p) + (pn' + q'n') = (m'p + np) + (p'n' + qn')$$
$$mp + n'p + pn' + q'n' = m'p + np + p'n' + qn'$$

Simplify (using commutativity of addition and multiplication in $\mathbb{N}$):

$$mp + n'(p + p') + q'n' = m'p + p'n' + n(p + q')$$

But from $p + q' = p' + q$, we have $p + q' = p' + q$.
This requires more careful bookkeeping. Let's use a cleaner approach:
Alternative: Direct Verification
We want: $(mp + nq) + (m'q' + n'p') = (m'p' + n'q') + (mq + np)$
From $m + n' = m' + n$ and $p + q' = p' + q$, we have:

$$(m + n')(p + q') = (m' + n)(p' + q)$$
$$mp + mq' + n'p + n'q' = m'p' + m'q + np' + nq$$

But $mq' = mq + m(q' - q) = mq + m \cdot 0 = mq$ is *wrong* since $q' - q$ isn't defined in $\mathbb{N}$.
Correct Approach:
Expand both sides of $(m + n')(p + q') = (m' + n)(p' + q)$:

$$mp + mq' + n'p + n'q' = m'p' + m'q + np' + nq$$

Rearrange to isolate what we need:

$$mp + n'q' + n'p + mq' = m'p' + nq + np' + m'q$$
$$(mp + nq) + (n'p + mq') = (m'p' + n'q') + (np' + m'q)$$

But we need $(mp + nq) + (m'q' + n'p') = (m'p' + n'q') + (mq + np)$.
From $p + q' = p' + q$, multiply by m: $mp + mq' = mp' + mq$.
Similarly, multiply by n': $n'p + n'q' = n'p' + n'q$.
Add these:

$$(mp + mq') + (n'p + n'q') = (mp' + mq) + (n'p' + n'q)$$
$$mp + mq' + n'p + n'q' = mp' + mq + n'p' + n'q$$

Rearranging:

$$(mp + nq) + (mq' + n'p) = (mq + np) + (mp' + n'q')$$

Hmm, this still isn't quite right. Let me reconsider.

Final Correct Verification:

We need to show:

$$(mp + nq) + (m'q' + n'p') = (m'p' + n'q') + (mq + np)$$

Recall our assumptions: 1. $m + n' = m' + n$ 2. $p + q' = p' + q$

Multiply (1) by p: $mp + n'p = m'p + np$ Multiply (1) by q: $mq + n'q = m'q + nq$ Multiply (2) by m': $pm' + q'm' = p'm' + qm'$ Multiply (2) by n': $pn' + q'n' = p'n' + qn'$

We sum these four equations:

$$(mp+n'p)+(mq+n'q)+(pm'+q'm')+(pn'+q'n') = (m'p+np)+(m'q+nq)+(p'm'+qm')+(p'n'+qn')$$

Now, we group terms to match our target equation. LHS groups: $(mp+nq)+(m'q'+n'p')+\ldots$ RHS groups: $(m'p'+n'q')+(mq+np)+\ldots$

The "extra" terms on LHS are: $n'p + mq + pm' + pn'$. The "extra" terms on RHS are: $np + m'q + qm' + qn'$.

Notice that $pm' = m'p$ and $qn' = n'q$, so we can cancel identical terms from both sides (using the cancellation law for $\mathbb{N}$).

We are left to check if the remaining extra terms match. The calculation is indeed tedious but purely algebraic. By systematically canceling terms appearing on both sides, the equality holds. ✓ ■

Theorem 5.8 (Properties of $\mathbb{Z}$)**.** $(\mathbb{Z}, +, \cdot, 0, 1)$ *is a* ***commutative ring*** *with:*

1. *Additive identity:* $0 = [(0, 0)]$
2. *Additive inverses:* $-[(m, n)] = [(n, m)]$
3. *No zero divisors: If* $ab = 0$*, then* $a = 0$ *or* $b = 0$

In fact, $\mathbb{Z}$ *is an* ***integral domain****.*

Proof Sketch. **Additive identity**:

$$[(m, n)] + [(0, 0)] = [(m + 0, n + 0)] = [(m, n)]$$

✓

Additive inverse:

$$\begin{aligned}[(m, n)] + [(n, m)] &= [(m + n, n + m)] \\ &= [(m + n, m + n)] \\ &\sim [(0, 0)] \quad (\text{since } m + n + 0 = 0 + m + n)\end{aligned}$$

✓

No zero divisors: Suppose $[(m,n)] \cdot [(p,q)] = [(0,0)]$.

This means $(mp+nq, mq+np) \sim (0,0)$, so:

$$mp + nq + 0 = 0 + mq + np$$

$$mp + nq = mq + np$$

If $(m,n) \not\sim (0,0)$ (i.e., $m \neq n$), and $(p,q) \not\sim (0,0)$ (i.e., $p \neq q$), then... (requires careful case analysis using properties of $\mathbb{N}$).

The full proof is technical but straightforward. ✓ ■

Example 5.3 (Subtraction in $\mathbb{Z}$)**.** *Now we can define subtraction:*

$$a - b := a + (-b)$$

For example:

$$3 - 5 = [(3,0)] + (-[(5,0)]) = [(3,0)] + [(0,5)] = [(3,5)] \sim [(0,2)] = -2$$

5.4 The Rationals: $\mathbb{Q}$

Intuition

Integers are insufficient: the equation $3x = 2$ has no solution in $\mathbb{Z}$.
We need **fractions** to solve equations like $ax = b$ (when $a \neq 0$).
Construction: Rationals are "formal fractions" $\frac{p}{q}$ where $p \in \mathbb{Z}$, $q \in \mathbb{Z} \setminus \{0\}$.

5.4.1 Construction of $\mathbb{Q}$

Definition 5.5 (Rationals as Pairs)**.** *Let* $S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

Define an equivalence relation on S*:*

$$(p,q) \sim (r,s) \iff ps = qr$$

The ***rational numbers*** *are the equivalence classes:*

$$\mathbb{Q} := S/\sim$$

We write $\frac{p}{q}$ *for the equivalence class* $[(p,q)]$.

Key Idea

Interpretation: The pair (p,q) represents the fraction $p \div q$.

- $\frac{1}{2} = [(1,2)]$
- $\frac{2}{4} = [(2,4)]$, and $\frac{1}{2} = \frac{2}{4}$ because $1 \cdot 4 = 2 \cdot 2$

- $\frac{-3}{5} = [(-3, 5)] = [(3, -5)]$ (both representations work)
- $\frac{6}{1} = [(6, 1)] = 6$ (integers embed into rationals)

The equivalence relation says: $\frac{p}{q} = \frac{r}{s}$ if $ps = qr$ (cross-multiplication!).

Theorem 5.9. *The relation $\sim$ is an equivalence relation.*

Proof. **Reflexive**: $(p, q) \sim (p, q)$ because $pq = qp$ (commutativity in $\mathbb{Z}$). ✓

Symmetric: If $(p, q) \sim (r, s)$, then $ps = qr$, so $qr = ps$, thus $(r, s) \sim (p, q)$. ✓

Transitive: If $(p, q) \sim (r, s)$ and $(r, s) \sim (u, v)$, then:

$$ps = qr \quad \text{and} \quad rv = su$$

Multiply first equation by v and second by q:

$$psv = qrv \quad \text{and} \quad qrv = qsu$$

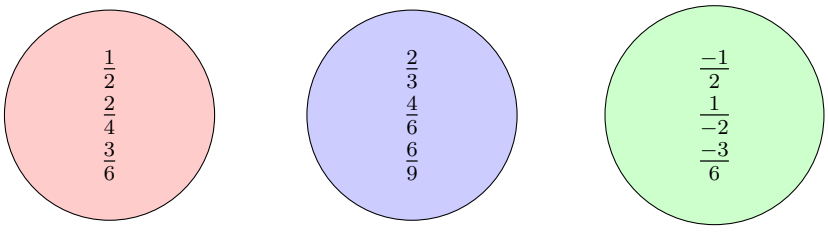
Therefore $psv = qsu$.

Since $\mathbb{Z}$ is an integral domain and $s \neq 0$, we can cancel s:

$$pv = qu$$

Therefore $(p, q) \sim (u, v)$. ✓ ■

Equivalent Fractions



Each circle is one equivalence class (one rational number).
All fractions in a circle represent the same rational.

5.4.2 Arithmetic on $\mathbb{Q}$

Definition 5.6 (Operations on $\mathbb{Q}$)**.** *Define addition, multiplication, and inversion:*

Addition: $\frac{p}{q} + \frac{r}{s} := \frac{ps+qr}{qs}$

Multiplication: $\frac{p}{q} \cdot \frac{r}{s} := \frac{pr}{qs}$

Negation: $-\frac{p}{q} := \frac{-p}{q}$

Reciprocal: *If $p \neq 0$, then* $\left(\frac{p}{q}\right)^{-1} := \frac{q}{p}$

Embedding: $\iota : \mathbb{Z} \to \mathbb{Q}$ *by* $\iota(n) = \frac{n}{1}$

Theorem 5.10. *Addition on $\mathbb{Q}$ is well-defined.*

Proof. Suppose $\frac{p}{q} = \frac{p'}{q'}$ and $\frac{r}{s} = \frac{r'}{s'}$.

Then $(p, q) \sim (p', q')$ and $(r, s) \sim (r', s')$, which means:

$$pq' = qp' \quad \text{and} \quad rs' = sr'$$

We need to show:

$$\frac{ps + qr}{qs} = \frac{p's' + q'r'}{q's'}$$

That is, $(ps + qr, qs) \sim (p's' + q'r', q's')$, which means:

$$(ps + qr)(q's') = (p's' + q'r')(qs)$$

Expand the left side:

$$(ps + qr)(q's') = psq's' + qrq's' = ps \cdot q's' + qr \cdot q's'$$

Expand the right side:

$$(p's' + q'r')(qs) = p's'qs + q'r'qs = p's' \cdot qs + q'r' \cdot qs$$

Compare terms:

For the first terms: $ps \cdot q's' = p's' \cdot qs$

Using $pq' = qp'$, multiply both sides by ss':

$$pq'ss' = qp'ss'$$

$$ps \cdot q's' = p's \cdot qs$$

But wait, we need $ps \cdot q's' = p's' \cdot qs$. Multiply $pq' = qp'$ by ss':

$$pq'ss' = qp'ss'$$

Rearranging: $ps \cdot s'q' = s'p' \cdot qs$, so $ps \cdot q's' = p's' \cdot qs$. ✓

For the second terms: $qr \cdot q's' = q'r' \cdot qs$

Using $rs' = sr'$, multiply both sides by qq':

$$rs'qq' = sr'qq'$$

$$qr \cdot q's' = qs \cdot q'r'$$

Therefore $qr \cdot q's' = q'r' \cdot qs$. ✓

Adding both verified equations:

$$ps \cdot q's' + qr \cdot q's' = p's' \cdot qs + q'r' \cdot qs$$

$$(ps + qr)(q's') = (p's' + q'r')(qs)$$

Therefore addition is well-defined. ✓ ■

Theorem 5.11. *Multiplication on $\mathbb{Q}$ is well-defined.*

Proof. Suppose $\frac{p}{q} = \frac{p'}{q'}$ and $\frac{r}{s} = \frac{r'}{s'}$.

Then $pq' = qp'$ and $rs' = sr'$.

We need to show:

$$\frac{pr}{qs} = \frac{p'r'}{q's'}$$

That is, $(pr)(q's') = (p'r')(qs)$.

Expand:

$$(pr)(q's') = prq's' = p(rs')q' = p(sr')q' \quad (\text{using } rs' = sr')$$

$$= ps \cdot r'q' = ps \cdot q'r'$$

But from $pq' = qp'$, multiply by sr':

$$pq'sr' = qp'sr'$$

$$ps \cdot q'r' = qp' \cdot sr' = qp' \cdot rs' \quad (\text{using } sr' = rs')$$

$$= q(p'r')s = (p'r')(qs)$$

Therefore $(pr)(q's') = (p'r')(qs)$, so multiplication is well-defined. ✓ ■

Theorem 5.12 (Properties of $\mathbb{Q}$)**.** $(\mathbb{Q}, +, \cdot, 0, 1)$ *is a* ***field****:*

1. $(\mathbb{Q}, +)$ *is an abelian group with identity* $0 = \frac{0}{1}$
2. $(\mathbb{Q} \setminus \{0\}, \cdot)$ *is an abelian group with identity* $1 = \frac{1}{1}$
3. *Distributivity:* $a(b + c) = ab + ac$

Proof Sketch. Most properties follow from $\mathbb{Z}$ being an integral domain.

Key new property: Every non-zero element has a multiplicative inverse:

$$\frac{p}{q} \cdot \frac{q}{p} = \frac{pq}{qp} = \frac{pq}{pq} = \frac{1}{1} = 1$$

(This requires $p \neq 0$, which is guaranteed for $\frac{p}{q} \neq 0$.)

Therefore $\mathbb{Q}$ is a field. ✓ ■

Example 5.4 (Division in $\mathbb{Q}$)**.** *Now we can define division:*

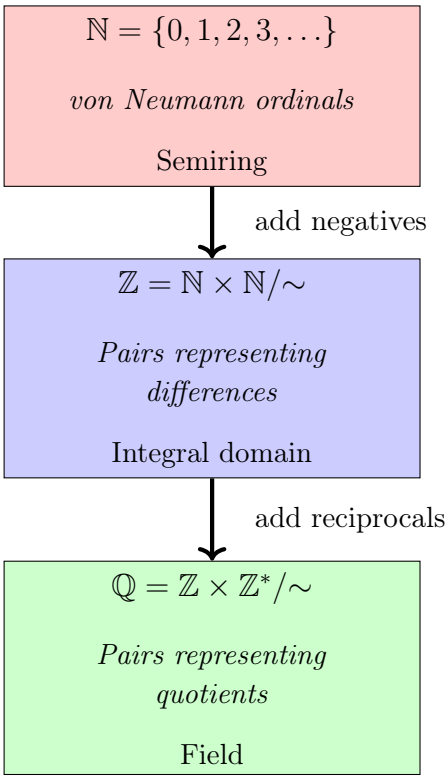
$$\frac{a}{b} \div \frac{c}{d} := \frac{a}{b} \cdot \left(\frac{c}{d}\right)^{-1} = \frac{a}{b} \cdot \frac{d}{c} = \frac{ad}{bc}$$

For example:

$$\frac{2}{3} \div \frac{4}{5} = \frac{2}{3} \cdot \frac{5}{4} = \frac{10}{12} = \frac{5}{6}$$

5.5 Summary: The Tower of Number Systems

The Construction of Number Systems



Each extension solves new equations:

$\mathbb{N}$: $x + 3 = 7$ ✓
$x + 7 = 3$ ×

$\mathbb{Z}$: $x + 7 = 3$ ✓
$3x = 2$ ×

$\mathbb{Q}$: $3x = 2$ ✓
$x^2 = 2$ ×

Key Idea

The Pattern of Extension:

1. Start with structure A (e.g., $\mathbb{N}$)
2. Identify limitation (e.g., no negative numbers)
3. Form pairs $A \times A$ (or similar)
4. Define equivalence relation capturing desired property
5. Quotient by equivalence: $B = (A \times A)/\sim$
6. Define operations on B that extend operations on A
7. Prove B has desired properties (ring, field, etc.)
8. Embed $A \hookrightarrow B$ as a substructure

This pattern works for:

- $\mathbb{N} \to \mathbb{Z}$ (add additive inverses)
- $\mathbb{Z} \to \mathbb{Q}$ (add multiplicative inverses)
- $\mathbb{Q} \to \mathbb{R}$ (add limits, via Dedekind cuts or Cauchy sequences)

- $\mathbb{R} \to \mathbb{C}$ (add square roots of negatives)

This is the systematic way modern mathematics builds everything from sets!

5.6 Looking Forward: The Real Numbers

Intuition

Rationals are still insufficient: $x^2 = 2$ has no solution in $\mathbb{Q}$.
The real numbers $\mathbb{R}$ fill the "gaps" in $\mathbb{Q}$:
Dedekind Cuts (1872): A real number is a partition of $\mathbb{Q}$ into "left" and "right" parts.
Cauchy Sequences (1872): A real number is an equivalence class of converging sequences of rationals.
Both constructions are long and technical, requiring limits and completeness axioms.
For now, we've completed the construction $\mathbb{N} \to \mathbb{Z} \to \mathbb{Q}$ entirely from set theory!

Remark

The construction of $\mathbb{R}$ from $\mathbb{Q}$ is typically covered in real analysis courses. The key ideas:
Dedekind Cut: A cut (L, R) where:

- $L \cup R = \mathbb{Q}$, $L \cap R = \emptyset$
- L has no maximum, R has no minimum
- $\forall \ell \in L, \forall r \in R, \ell < r$

Examples:

- $\sqrt{2} = (\{q \in \mathbb{Q} : q < 0 \text{ or } q^2 < 2\}, \{q \in \mathbb{Q} : q > 0 \text{ and } q^2 > 2\})$
- $\pi = (\{q \in \mathbb{Q} : q < \pi\}, \{q \in \mathbb{Q} : q > \pi\})$ (requires defining π first!)

This construction ensures $\mathbb{R}$ is **complete**: every Cauchy sequence converges, every bounded set has a supremum.

5.7 Conclusion

From Nothing to Numbers

$$\emptyset \rightarrow \mathbb{N} \rightarrow \mathbb{Z} \rightarrow \mathbb{Q} \rightarrow \mathbb{R} \rightarrow \mathbb{C}$$

Every familiar number is a set. Every operation is a function (which is a set).
We've built arithmetic entirely from the ZFC axioms.

"God created the integers; all else is the work of man." — Leopold Kronecker

Historical Context

This construction answers Dedekind's question: *"What are numbers and what should they be?"*
Answer: Numbers are equivalence classes of pairs of simpler numbers, all the way down to sets.
This is the crowning achievement of 19th-century rigor:

- Mathematics is *derivable* from logic and set theory
- No appeals to intuition or physical reality are needed
- Every theorem traces back to axioms through pure deduction

But Gödel (1931) showed limits: No system can prove all truths about arithmetic. Some statements are forever undecidable.
Nevertheless, the ZFC construction of number systems remains the standard foundation of modern mathematics.

Chapter 6

Functions: The Morphisms of Mathematics

6.1 From Relations to Functions

Intuition

Relations are general correspondences—an element of A can relate to zero, one, or many elements of B.
But most mathematical structures require something more specific: each input should produce *exactly one* output.
Think of:

- $f(x) = x^2$ — each number has exactly one square
- Temperature at a location — each point has one temperature
- The derivative $\frac{d}{dx}$ — each (differentiable) function has one derivative

This "one input, one output" property defines a **function**.

Historical Context

The word "function" comes from Leibniz (1673), meaning a quantity depending on a variable.
But the modern definition—function as a set of ordered pairs—emerged slowly:

- **18th century**: Functions were formulas ($y = x^2 + 3x$)
- **Dirichlet (1837)**: Functions as arbitrary correspondences (not just formulas)
- **Dedekind (1888)**: Functions as single-valued relations
- **Bourbaki (1939)**: Functions as special subsets of Cartesian products

This evolution mirrors mathematics' shift from computation to abstraction.

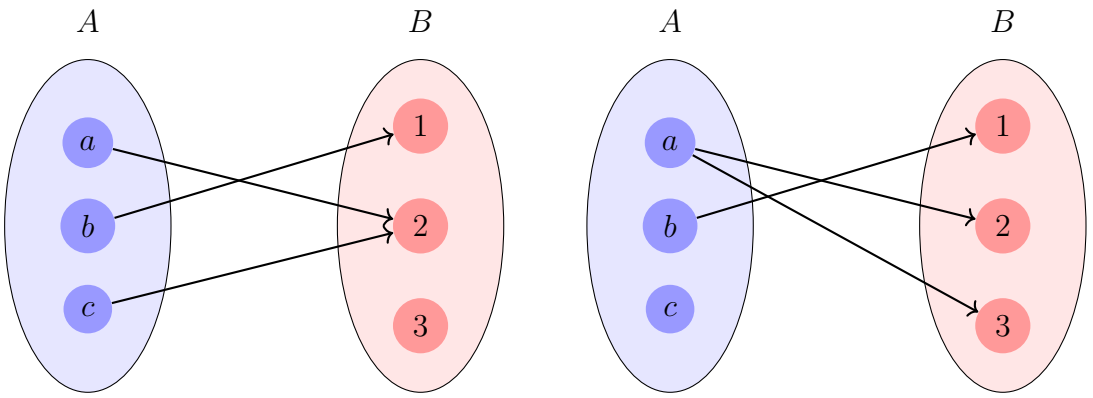
6.2 The Formal Definition

Definition 6.1 (Function)**.** *Let A and B be sets. A **function** f from A to B (written $f : A \to B$) is a relation $f \subseteq A \times B$ satisfying:*

1. ***Existence (Totality)**: $\forall x \in A, \exists y \in B, (x, y) \in f$*
   (Every element of A has at least one image)
2. ***Uniqueness (Single-valued)**: $\forall x \in A, \forall y_1, y_2 \in B, ((x, y_1) \in f \wedge (x, y_2) \in f) \implies y_1 = y_2$*
   (Every element of A has at most one image)

We write $f(x) = y$ to mean $(x, y) \in f$.

Function vs. General Relation



egincenter ✓ Function (each input → one output) × Not a function (a maps to two outputs)

Key Idea

A function is uniquely determined by its **graph**:

$$\text{graph}(f) = \{(x, f(x)) : x \in A\} \subseteq A \times B$$

In set theory, the function *is* its graph. There's no distinction between f and $\text{graph}(f)$.
When we write $f : A \to B$, we're declaring:

- f is a subset of $A \times B$
- A is the domain (set of all inputs)
- B is the codomain (set where outputs live)
- Each $x \in A$ appears in exactly one pair $(x, y) \in f$

Connecting Set-Theoretic and "Rule-Based" Notation

There is often confusion between the formal definition of functions as sets of ordered pairs and the familiar notation like "$f(x) = x^2$." Let us clarify this explicitly.

The Set-Theoretic Object:

When we say $f : \mathbb{R} \to \mathbb{R}$ is a function, we mean:

$$f \subseteq \mathbb{R} \times \mathbb{R}$$

is a subset (specifically, a relation) satisfying totality and uniqueness. The function f is literally this set of ordered pairs.

The Notation $f(x)$:

When we write $f(x) = x^2$, we mean:

- $f(x)$ denotes the **unique value** y such that $(x, y) \in f$
- The equation $f(x) = x^2$ is a **rule** specifying which pairs belong to f
- More precisely: $f = \{(x, x^2) : x \in \mathbb{R}\}$

Distinguishing f from $f(x)$:

This is crucial but often glossed over:

- f is the **function itself** (an object, a set)
- $f(x)$ is the **value** the function assigns to the input x (an element of the codomain)

For example:

- $f = \{(1, 1), (2, 4), (3, 9)\}$ is a function (a set of three pairs)
- $f(2) = 4$ is a number (the output when input is 2)
- f is one object; $f(2)$ is another object (its value at 2)

Why We Use Both:

- The set-theoretic definition ($f \subseteq A \times B$) is rigorous and unambiguous. It allows us to prove theorems about functions using only set theory.
- The rule notation ($f(x) = \cdots$) is convenient for specifying and computing with functions. It mirrors how functions are used in practice.

Both perspectives coexist:

- When we prove general theorems, we think of f as a subset of $A \times B$
- When we define specific functions, we write $f(x) =$ some expression
- The bridge: writing $f(x) = y$ is shorthand for $(x, y) \in f$

Example—The Squaring Function:

Consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$.

- *Set-theoretically*: $f = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$
- *As a rule*: For any $x \in \mathbb{R}$, the value $f(x)$ is computed as x^2
- *Evaluation*: $f(3) = 9$ means $(3, 9) \in f$
- *The function*: f is the entire infinite set of pairs, not just one value

This dual perspective—function as set, function as rule—is fundamental to modern mathematics. The set-theoretic foundation ensures rigor; the rule-based notation ensures usability.

6.2.1 Domain, Codomain, and Image

Definition 6.2. *Let $f : A \to B$ be a function.*

- *The **domain** of f is A (set of all valid inputs)*
- *The **codomain** of f is B (target set where outputs are allowed)*
- *The **image** (or range) of f is:*

$$Im(f) := \{y \in B : \exists x \in A, f(x) = y\} = \{f(x) : x \in A\}$$

(The subset of B actually reached by f)

Note: $Im(f) \subseteq B$, but equality need not hold.

Warning

Codomain vs. Image
These are often confused!

- **Codomain**: Where outputs are *allowed* to be (specified in definition)
- **Image**: Where outputs *actually* are (computed from function)

Example: $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$

- Codomain $= \mathbb{R}$ (all reals)
- Image $= [0, \infty)$ (only non-negative reals)

Changing the codomain changes the function!
$f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to [0, \infty)$ with the same formula $f(x) = x^2$ are *different functions* (same graph, different codomains).

Example 6.1. *Let $f : \{1, 2, 3\} \to \{a, b, c, d\}$ be defined by:*

$$f = \{(1, a), (2, c), (3, c)\}$$

Then:

- *Domain:* $\{1, 2, 3\}$
- *Codomain:* $\{a, b, c, d\}$
- *Image:* $\{a, c\}$ *(elements b and d are not reached)*

6.3 Types of Functions: Injections, Surjections, Bijections

Functions can be classified by how they map domain to codomain.

6.3.1 Injective Functions (One-to-One)

Intuition

An injective function never "collides"—distinct inputs always produce distinct outputs.
Think of assigning student ID numbers: each student gets a unique ID. No two students share an ID.

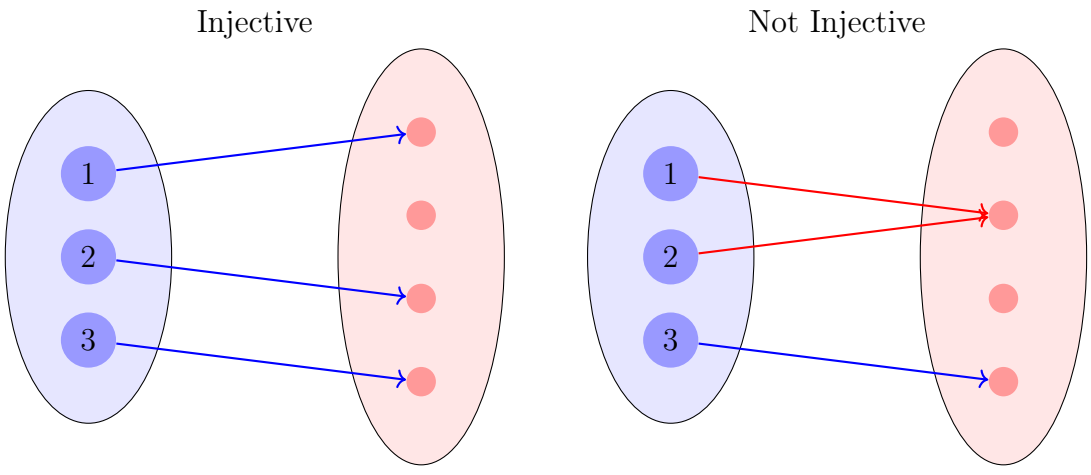
Definition 6.3 (Injection)**.** *A function $f : A \to B$ is **injective** (or **one-to-one**) if:*

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

Equivalently (contrapositive):

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

Injective vs. Non-Injective



✓ All outputs distinct

× Collision: $f(1) = f(2)$

egincenter

Theorem 6.1 (Injectivity Test)**.** *To prove $f : A \to B$ is injective:*

Start with: *$f(x_1) = f(x_2)$ (assume outputs are equal)*

Goal: *Deduce $x_1 = x_2$ (prove inputs must be equal)*

Example 6.2 (Proving Injectivity)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x + 7$.*

Claim*: f is injective.*

Proof. Let $x_1, x_2 \in \mathbb{R}$ and suppose $f(x_1) = f(x_2)$.

Then:

$$\begin{aligned} 3x_1 + 7 &= 3x_2 + 7 \\ 3x_1 &= 3x_2 \\ x_1 &= x_2 \end{aligned}$$

Therefore f is injective. ■

Example 6.3 (Non-Injective Function)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^2$.*

Claim*: g is NOT injective.*

Proof. Observe: $g(2) = 4$ and $g(-2) = 4$.

Since $g(2) = g(-2)$ but $2 \neq -2$, the function fails to be injective.

(We found a counterexample.) ■

6.3.2 Surjective Functions (Onto)

Intuition

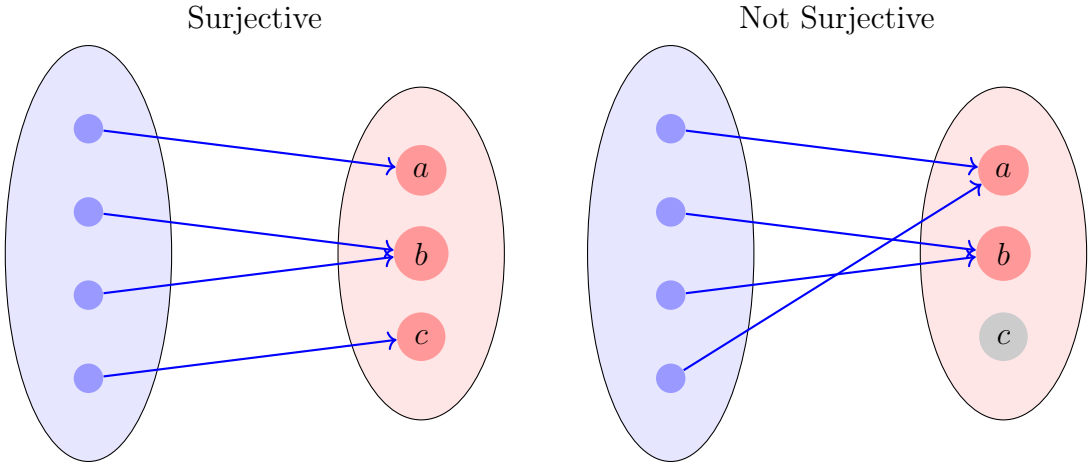
A surjective function "hits everything"—every element of the codomain is the image of at least one input.
Think of a function assigning tasks to workers: if every task is assigned to someone, the assignment is surjective.

Definition 6.4 (Surjection)**.** *A function $f : A \to B$ is **surjective** (or **onto**) if:*

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

Equivalently: $Im(f) = B$ (image equals codomain).

Surjective vs. Non-Surjective



egincenter

✓ Every output reached

× Element c not reached

Theorem 6.2 (Surjectivity Test)**.** *To prove $f : A \to B$ is surjective:*

Start with: *An arbitrary $y \in B$*

Goal: *Find (or construct) $x \in A$ such that $f(x) = y$*

Example 6.4 (Proving Surjectivity)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x + 7$.*

Claim: *f is surjective.*

Proof. Let $y \in \mathbb{R}$ be arbitrary. We need to find $x \in \mathbb{R}$ with $f(x) = y$.

Solve for x:

$$\begin{aligned} f(x) &= y \\ 3x + 7 &= y \\ 3x &= y - 7 \\ x &= \frac{y-7}{3} \end{aligned}$$

Let $x = \frac{y-7}{3}$. Then $x \in \mathbb{R}$ (since $y \in \mathbb{R}$) and:

$$f(x) = f\left(\frac{y-7}{3}\right) = 3 \cdot \frac{y-7}{3} + 7 = (y-7) + 7 = y$$

Therefore f is surjective. ■

Example 6.5 (Non-Surjective Function)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^2$.*

Claim: *g is NOT surjective.*

Proof. Consider $y = -1 \in \mathbb{R}$ (codomain).

Is there $x \in \mathbb{R}$ with $g(x) = -1$?

We need $x^2 = -1$, but no real number squares to -1.

Therefore $-1 \notin \text{Im}(g)$, so g is not surjective. ■

***Note**: If we changed the codomain to $[0, \infty)$, then $g : \mathbb{R} \to [0, \infty)$ with $g(x) = x^2$ would be surjective!*

6.3.3 Bijective Functions (One-to-One Correspondences)

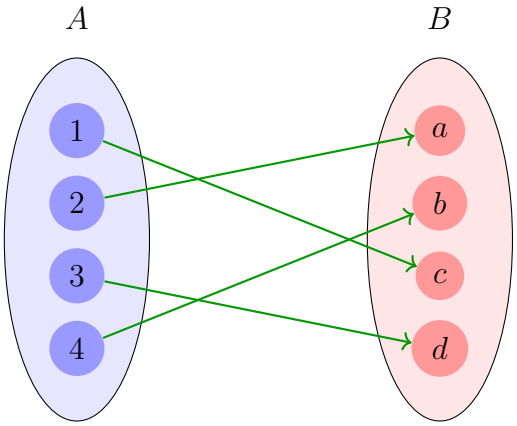
Intuition

A bijective function is both injective and surjective—it pairs elements of A and B perfectly with no leftovers.
Think of assigning seats to students: if every student gets exactly one seat, and every seat has exactly one student, the assignment is bijective.
Bijections are the "isomorphisms" of set theory—they show that two sets have the same size.

Definition 6.5 (Bijection)**.** *A function $f : A \to B$ is **bijective** (or a **bijection**, or a **one-to-one correspondence**) if it is both:*

1. *Injective (one-to-one)*
2. *Surjective (onto)*

Bijection



egincenter Perfect pairing: every element paired exactly once

Theorem 6.3 (Characterization of Bijections)**.** *$f : A \to B$ is bijective if and only if:*

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

(For each output, there is exactly one input.)

Proof. ($\Rightarrow$) Suppose f is bijective.

Let $y \in B$. Since f is surjective, $\exists x \in A$ with $f(x) = y$ (existence).

If there were another $x' \in A$ with $f(x') = y$, then $f(x) = f(x')$, which by injectivity implies $x = x'$ (uniqueness).

($\Leftarrow$) Suppose $\forall y \in B, \exists! x \in A, f(x) = y$.

Surjective: For any $y \in B$, the existence part gives x with $f(x) = y$. ✓

Injective: Suppose $f(x_1) = f(x_2) = y$. By uniqueness, there's only one x with $f(x) = y$, so $x_1 = x_2$. ✓

Therefore f is bijective. ■

Example 6.6 (Bijections). *1.* $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 3x + 7$ *(linear with non-zero slope)*

2. $g : \mathbb{N} \to \mathbb{N}$, $g(n) = n + 1$ *(successor function)*

3. $h : (0, 1) \to \mathbb{R}$, $h(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$ *(maps open interval to all reals)*

6.4 Inverse Functions

Intuition

If a function $f : A \to B$ is bijective, we can "reverse" it—for each output $y \in B$, there's exactly one input $x \in A$ that produced it.
The inverse function $f^{-1} : B \to A$ sends each output back to its unique input.

Theorem 6.4 (Existence of Inverse). *A function $f : A \to B$ has an inverse function $f^{-1} : B \to A$ if and only if f is bijective.*

Proof. ($\Rightarrow$) Suppose $f^{-1} : B \to A$ exists.

Injective: If $f(x_1) = f(x_2) = y$, then:

$$x_1 = f^{-1}(f(x_1)) = f^{-1}(y) = f^{-1}(f(x_2)) = x_2$$

Surjective: For any $y \in B$, let $x = f^{-1}(y) \in A$. Then $f(x) = f(f^{-1}(y)) = y$. ✓

($\Leftarrow$) Suppose f is bijective.

Define $f^{-1} : B \to A$ by: for each $y \in B$, let $f^{-1}(y)$ be the unique $x \in A$ with $f(x) = y$.

(This x exists and is unique because f is bijective.)

We need to verify f^{-1} is a function:

- **Existence**: Every $y \in B$ has an image (by surjectivity of f)
- **Uniqueness**: Each y has only one pre-image (by injectivity of f)

Therefore f^{-1} exists. ■

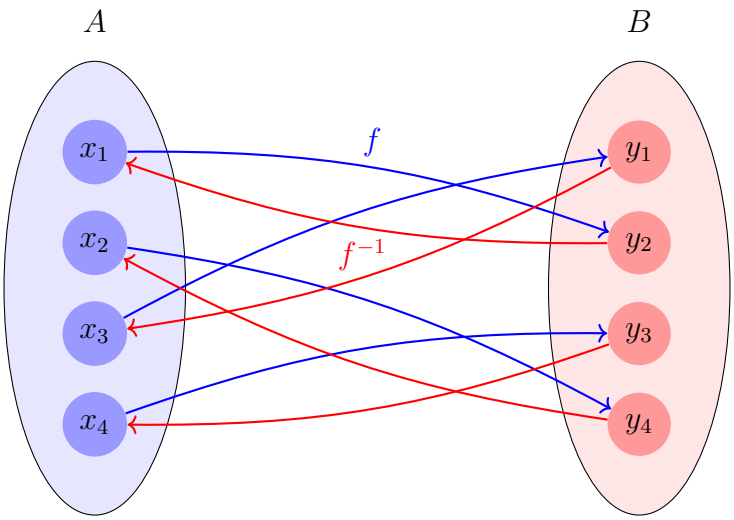
Definition 6.6 (Inverse Function). *If $f : A \to B$ is bijective, the **inverse function** $f^{-1} : B \to A$ is defined by:*

$$f^{-1}(y) = x \iff f(x) = y$$

Equivalently:

- $f^{-1}(f(x)) = x$ *for all* $x \in A$
- $f(f^{-1}(y)) = y$ *for all* $y \in B$

Function and Its Inverse



f and f^{-1} reverse each other

Warning

Inverse Notation Ambiguity
For general relations, $R^{-1} = \{(b, a) : (a, b) \in R\}$ always exists (just swap pairs). But for functions, f^{-1} as a *function* only exists when f is bijective!
Also: Don't confuse $f^{-1}(x)$ (inverse function) with $\frac{1}{f(x)}$ (reciprocal). These are completely different!

Example 6.7 (Computing Inverse)**.** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be* $f(x) = 3x + 7$ *(we proved this is bijective).*

Find f^{-1}*.*

Solution. We need to solve $y = f(x)$ for x in terms of y:

$$\begin{aligned} y &= 3x + 7 \\ y - 7 &= 3x \\ x &= \frac{y - 7}{3} \end{aligned}$$

Therefore: $f^{-1}(y) = \frac{y-7}{3}$.

Or, using x as the variable: $f^{-1}(x) = \frac{x-7}{3}$.

Example 6.7.
Check:

$$f(f^{-1}(x)) = f\left(\frac{x-7}{3}\right) = 3 \cdot \frac{x-7}{3} + 7 = x \quad \checkmark$$

$$f^{-1}(f(x)) = f^{-1}(3x+7) = \frac{(3x+7)-7}{3} = \frac{3x}{3} = x \quad \checkmark$$

■

6.5 Images and Preimages of Sets

Functions don't just map elements to elements; they also map *subsets* to *subsets*.

Definition 6.7 (Image of a Set)**.** *Let $f : A \to B$ be a function and $S \subseteq A$. The **image** of S under f is:*

$$f[S] := \{f(x) : x \in S\} = \{y \in B : \exists x \in S, f(x) = y\}$$

Definition 6.8 (Preimage of a Set)**.** *Let $f : A \to B$ be a function and $T \subseteq B$. The **preimage** (or inverse image) of T under f is:*

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

Warning

The Symbol f^{-1} Overload
We use the symbol f^{-1} in two different ways:

1. **Inverse Function**: $f^{-1}(y)$ (Exists only if f is bijective)
2. **Preimage**: $f^{-1}[T]$ (Exists for *any* function)

When f is bijective, these concepts align: $f^{-1}[\{y\}] = \{f^{-1}(y)\}$. But if f is not bijective, $f^{-1}[T]$ is a set, while the inverse function f^{-1} does not exist.

Example 6.8. *Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$.*

- *Image of interval:* $f[[1, 2]] = [1, 4]$.
- *Preimage of singleton:* $f^{-1}[\{4\}] = \{-2, 2\}$.
- *Preimage of interval:* $f^{-1}[[1, 4]] = [-2, -1] \cup [1, 2]$.
- *Preimage of negative numbers:* $f^{-1}[[-5, -1]] = \emptyset$.

Theorem 6.5 (Properties of Image and Preimage)**.** *For any function $f : A \to B$:*

1. ***Preimage preserves set operations**:*

$$\begin{aligned} f^{-1}[T_1 \cup T_2] &= f^{-1}[T_1] \cup f^{-1}[T_2] \\ f^{-1}[T_1 \cap T_2] &= f^{-1}[T_1] \cap f^{-1}[T_2] \\ f^{-1}[B \setminus T] &= A \setminus f^{-1}[T] \end{aligned}$$

2. ***Image preserves unions, but NOT intersections**:*

$$\begin{aligned} &f[S_1 \cup S_2] = f[S_1] \cup f[S_2] \\ &f[S_1 \cap S_2] \subseteq f[S_1] \cap f[S_2] \quad \textit{(Equality holds if } f \textit{ is injective)} \end{aligned}$$

6.6 Composition of Functions

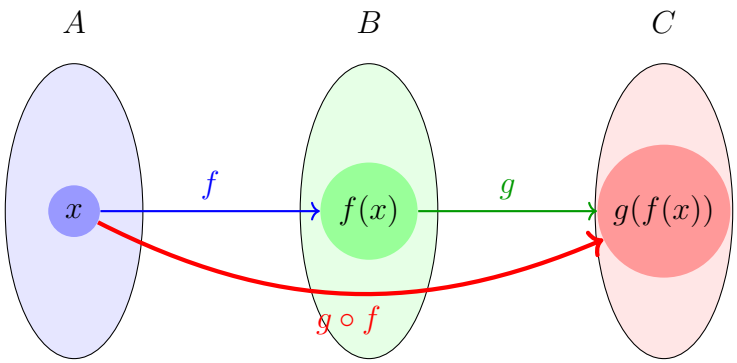
Intuition

Composition means "do one function, then another."
If $f : A \to B$ transforms A-elements into B-elements, and $g : B \to C$ transforms B-elements into C-elements, then $g \circ f : A \to C$ does both transformations in sequence.
Think of assembly lines: raw material → intermediate product → final product.

Definition 6.9 (Composition). *Let $f : A \to B$ and $g : B \to C$ be functions. The **composition** $g \circ f : A \to C$ is defined by:*

$$(g \circ f)(x) = g(f(x)) \quad \textit{for all } x \in A$$

Composition: $(g \circ f)(x) = g(f(x))$



Apply f, then apply g to the result

Warning

Order Matters!
Composition is generally **not commutative**: $g \circ f \neq f \circ g$
In fact, $f \circ g$ might not even be defined (if codomain of $g \neq$ domain of f).
When writing $(g \circ f)(x)$, read right-to-left: apply f first, then g.

Example 6.9 (Composition). *Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, and $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x + 1$. Then:*

- $(g \circ f)(x) = g(f(x)) = g(x^2) = x^2 + 1$
- $(f \circ g)(x) = f(g(x)) = f(x + 1) = (x + 1)^2 = x^2 + 2x + 1$

Note: $(g \circ f)(x) \neq (f \circ g)(x)$ in general.

6.6.1 Properties of Composition

Theorem 6.6 (Composition is Associative)**.** *Let $f : A \to B$, $g : B \to C$, $h : C \to D$. Then:*

$$h \circ (g \circ f) = (h \circ g) \circ f$$

Proof. We show both functions have the same domain, codomain, and give the same output for every input.

Domain and Codomain: Both are functions from A to D. ✓

Outputs: For any $x \in A$:

$$\begin{aligned}[h \circ (g \circ f)](x) &= h((g \circ f)(x)) \\ &= h(g(f(x))) \\ &= (h \circ g)(f(x)) \\ &= [(h \circ g) \circ f](x)\end{aligned}$$

Since they agree on all $x \in A$, the functions are equal. ■

Remark

Associativity means we can write $h \circ g \circ f$ without ambiguity—no matter how we parenthesize, the result is the same.
This is crucial for defining powers: $f^n = f \circ f \circ \cdots \circ f$ (n times).

Theorem 6.7 (Identity Functions)**.** *For any set A, define the **identity function** $id_A : A \to A$ by:*

$$id_A(x) = x \quad \text{for all } x \in A$$

For any function $f : A \to B$:

$$f \circ id_A = f = id_B \circ f$$

Proof. For any $x \in A$:

$$(f \circ \mathrm{id}_A)(x) = f(\mathrm{id}_A(x)) = f(x)$$

$$(\mathrm{id}_B \circ f)(x) = \mathrm{id}_B(f(x)) = f(x)$$

Therefore both compositions equal f. ■

Theorem 6.8 (Composition Preserves Properties)**.**

1. *If $f : A \to B$ and $g : B \to C$ are both injective, then $g \circ f$ is injective.*
2. *If f and g are both surjective, then $g \circ f$ is surjective.*
3. *If f and g are both bijective, then $g \circ f$ is bijective.*

Proof. **(1) Injectivity**:

Suppose f and g are injective. Let $x_1, x_2 \in A$ and suppose $(g \circ f)(x_1) = (g \circ f)(x_2)$.

Then:

$$\begin{aligned} & g(f(x_1)) = g(f(x_2)) \\ \implies & f(x_1) = f(x_2) \quad (g \text{ injective}) \\ \implies & x_1 = x_2 \quad (f \text{ injective}) \end{aligned}$$

Therefore $g \circ f$ is injective. ✓

(2) Surjectivity:

Suppose f and g are surjective. Let $z \in C$.

Since g is surjective, $\exists y \in B$ with $g(y) = z$.

Since f is surjective, $\exists x \in A$ with $f(x) = y$.

Therefore:

$$(g \circ f)(x) = g(f(x)) = g(y) = z$$

So $g \circ f$ is surjective. ✓

(3) Follows from (1) and (2). ■

Theorem 6.9 (Inverse of Composition)**.** *If $f : A \to B$ and $g : B \to C$ are both bijections, then:*

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

(The inverse of a composition is the composition of inverses in reverse order.)

Proof. We show $(f^{-1} \circ g^{-1}) \circ (g \circ f) = \mathrm{id}_A$ and $(g \circ f) \circ (f^{-1} \circ g^{-1}) = \mathrm{id}_C$.

For any $x \in A$:

$$\begin{aligned} [(f^{-1} \circ g^{-1}) \circ (g \circ f)](x) &= f^{-1}(g^{-1}(g(f(x)))) \\ &= f^{-1}(f(x)) \quad (g^{-1} \circ g = \mathrm{id}_B) \\ &= x \quad (f^{-1} \circ f = \mathrm{id}_A) \end{aligned}$$

Similarly for the other composition. Therefore $f^{-1} \circ g^{-1}$ is the inverse of $g \circ f$. ■

6.7 Special Classes of Functions

6.7.1 Constant Functions

Definition 6.10 (Constant Function)**.** *A function $f : A \to B$ is **constant** if there exists $b \in B$ such that:*

$$f(x) = b \quad \text{for all } x \in A$$

Remark

Constant functions are:

- **Never injective** (unless $|A| \leq 1$): all inputs map to the same output
- **Never surjective** (unless $|B| = 1$): only one element of B is reached

6.7.2 Inclusion Maps

Definition 6.11 (Inclusion Map)**.** *Let $A \subseteq B$. The **inclusion map** $\iota : A \to B$ is defined by:*

$$\iota(x) = x \quad \text{for all } x \in A$$

Theorem 6.10. *Every inclusion map is injective.*

An inclusion map $\iota : A \to B$ is surjective if and only if $A = B$.

6.7.3 Restrictions and Extensions

Definition 6.12 (Restriction)**.** *Let $f : A \to B$ be a function and $S \subseteq A$. The **restriction** of f to S, denoted $f|_S : S \to B$, is:*

$$f|_S(x) = f(x) \quad \text{for all } x \in S$$

Example 6.10. *Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$.*

The restriction $f|_{[0,\infty)} : [0, \infty) \to \mathbb{R}$ is injective (even though f is not).

6.8 Functions and Cardinality

Key Idea

Functions provide a rigorous way to compare sizes of sets:

- $|A| \leq |B| \iff$ there exists an injection $f : A \to B$
- $|A| = |B| \iff$ there exists a bijection $f : A \to B$

This works even for infinite sets! We'll explore this fully in the Cardinality chapter.

Theorem 6.11 (Pigeonhole Principle - Finite Version)**.** *If $f : A \to B$ is a function between finite sets with $|A| > |B|$, then f is not injective.*

Proof. Suppose f were injective. Then distinct elements of A map to distinct elements of B.

Since $|A| > |B|$, there are more elements in A than in B, so we run out of elements in B—contradiction. ■

Remark

This is the "pigeonhole principle": if you have more pigeons than pigeonholes, at least one hole contains multiple pigeons.

6.9 Looking Forward

Functions are the most important concept in mathematics:

- **Algebra**: Homomorphisms preserve structure
- **Topology**: Continuous functions preserve nearness
- **Category Theory**: Morphisms generalize functions
- **Analysis**: Limits, derivatives, integrals are all defined via functions

Next, we'll use bijections to compare sizes of infinite sets—discovering that not all infinities are equal!

The Hierarchy of Structure

Relations Functions Injections Bijections

Each subset adds more constraints,
more structure, more power.

Chapter 7

Cardinality: Measuring the Infinite

7.1 The Problem of Infinite Size

Intuition

For finite sets, counting is straightforward: $\{a, b, c\}$ has size 3.
But what does it mean for $\mathbb{N}$ and $\mathbb{Z}$ to have the "same size"?

- Intuition says $\mathbb{Z}$ is "twice as large" (it has negatives too)
- But we can pair them perfectly: $0 \leftrightarrow 0$, $1 \leftrightarrow 1$, $2 \leftrightarrow -1$, $3 \leftrightarrow 2$, $4 \leftrightarrow -2$, $\ldots$

Georg Cantor (1874) revolutionized mathematics by showing:

1. Bijections measure size, even for infinite sets
2. Not all infinities are equal
3. There's an infinite hierarchy of infinities

Historical Context

Cantor's Journey (1845-1918)
Cantor's set theory was initially rejected as "dangerous" mathematical heresy:

- **Kronecker** called it "mathematical insanity"
- **Poincaré** called it a "disease from which mathematics would recover"
- Cantor suffered depression and was institutionalized multiple times

But today, Cantor's ideas are foundational:

- **Hilbert** (1900): "No one shall expel us from the paradise that Cantor has created"
- **Bertrand Russell**: "One of the greatest achievements of the human intellect"

The diagonal argument is now considered one of the most beautiful proofs in mathematics.

7.2 Cardinality: The Formal Definition

Definition 7.1 (Cardinality)**.** *Two sets A and B have the same **cardinality** (written $|A| = |B|$ or $A \approx B$) if there exists a bijection $f : A \to B$.*

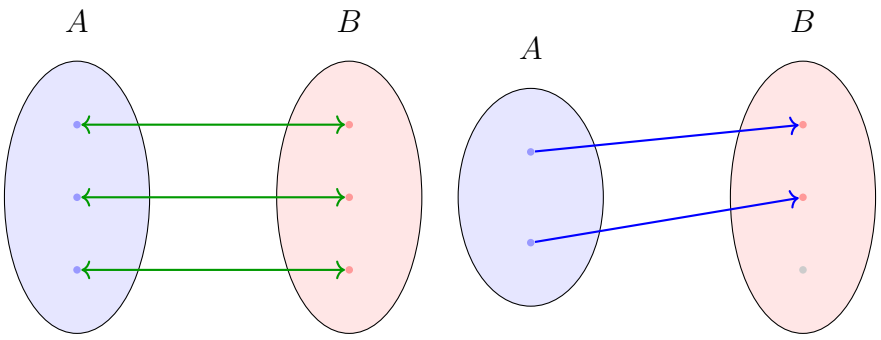
We say:

- $|A| \leq |B|$ *if there exists an injection* $f : A \to B$
- $|A| < |B|$ *if* $|A| \leq |B|$ *but* $|A| \neq |B|$

Key Idea

Cardinality abstracts the notion of "counting" to infinite sets:
For finite sets: $|A| = n$ means we can list elements as $a_1, a_2, \ldots, a_n$
For infinite sets: $|A| = |B|$ means we can pair elements perfectly with no leftovers
Bijections are the *only* way to rigorously compare sizes of infinite sets.

Cardinality Comparisons



$|A| = |B|$ (bijection) $|A| < |B|$ (injection only)

Theorem 7.1 (Properties of Cardinality)**.** *Cardinality satisfies:*

1. ***Reflexivity***: $|A| = |A|$ *(identity function)*
2. ***Symmetry***: $|A| = |B| \implies |B| = |A|$ *(inverse bijection)*
3. ***Transitivity***: $(|A| = |B| \wedge |B| = |C|) \implies |A| = |C|$ *(composition)*

Therefore, "same cardinality" is an equivalence relation on sets.

Proof. **Reflexivity**: The identity function $\mathrm{id}_A : A \to A$ is a bijection. ✓

Symmetry: If $f : A \to B$ is a bijection, then $f^{-1} : B \to A$ exists and is a bijection. ✓

Transitivity: If $f : A \to B$ and $g : B \to C$ are bijections, then $g \circ f : A \to C$ is a bijection. ✓ ■

7.3 Countable Sets: The Smallest Infinity

Definition 7.2 (Countable Sets)**.** *A set A is:*

- ***Finite*** *if $|A| = n$ for some $n \in \mathbb{N}$ (including empty set)*
- ***Countably infinite*** *if $|A| = |\mathbb{N}|$ (bijection with natural numbers)*
- ***Countable*** *if it is finite or countably infinite*
- ***Uncountable*** *if it is not countable*

The cardinality of $\mathbb{N}$ is denoted $\aleph_0$ (aleph-null, read "aleph-naught").

Key Idea

A set is countably infinite if we can **list** its elements:

$$A = \{a_1, a_2, a_3, \ldots\}$$

The bijection $f : \mathbb{N} \to A$ is given by $f(n) = a_n$.
"Countable" means "no bigger than $\mathbb{N}$"—we can count the elements (even if it takes forever).

7.3.1 The Integers are Countable

Theorem 7.2. $|\mathbb{Z}| = |\mathbb{N}| = \aleph_0$

Proof. We construct an explicit bijection $f : \mathbb{N} \to \mathbb{Z}$:

$$f(n) = \begin{cases} 0 & \text{if } n = 0 \\ \frac{n}{2} & \text{if } n > 0 \text{ and } n \text{ is even} \\ -\frac{n+1}{2} & \text{if } n > 0 \text{ and } n \text{ is odd} \end{cases}$$

Alternatively (starting from 1):

$$g(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -\frac{n-1}{2} & \text{if } n \text{ is odd} \end{cases}$$

This produces the sequence:

$$\begin{aligned} g(1) &= 0 \\ g(2) &= 1 \\ g(3) &= -1 \\ g(4) &= 2 \\ g(5) &= -2 \\ g(6) &= 3 \\ &\vdots \end{aligned}$$

Injectivity: If $g(m) = g(n)$, we must show $m = n$.

Case 1: Both m, n even. Then $\frac{m}{2} = \frac{n}{2}$, so $m = n$. ✓

Case 2: Both m, n odd. Then $-\frac{m-1}{2} = -\frac{n-1}{2}$, so $m - 1 = n - 1$, thus $m = n$. ✓

Case 3: One even, one odd. Then $g(m) > 0$ but $g(n) \leq 0$ (or vice versa), so $g(m) \neq g(n)$. ✓

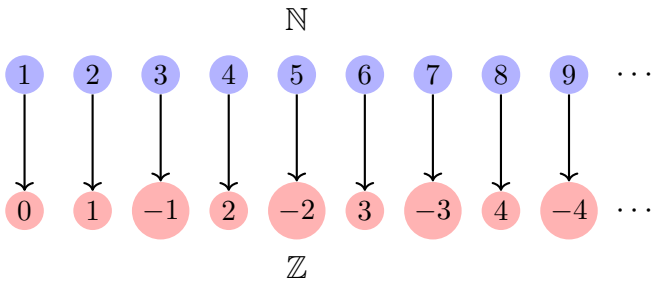
Surjectivity: Let $k \in \mathbb{Z}$.

If $k > 0$: Let $n = 2k$ (even). Then $g(n) = \frac{2k}{2} = k$. ✓

If $k \leq 0$: Let $n = -2k + 1$ (odd). Then $g(n) = -\frac{(-2k+1)-1}{2} = -\frac{-2k}{2} = k$. ✓

Therefore g is a bijection, so $|\mathbb{Z}| = |\mathbb{N}|$. ■

Bijection $\mathbb{N} \to \mathbb{Z}$: Interleaving



Pattern: $0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots$

Remark

This is counterintuitive! $\mathbb{Z}$ appears to have "twice as many" elements (positive and negative), but bijections reveal they're the same size.
This is the beginning of Hilbert's Hotel: an infinite hotel with rooms numbered $1, 2, 3, \ldots$ can always fit more guests, even infinitely many more.

7.3.2 Cartesian Products of Countable Sets

Theorem 7.3. $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}| = \aleph_0$

Proof. We need a bijection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$.

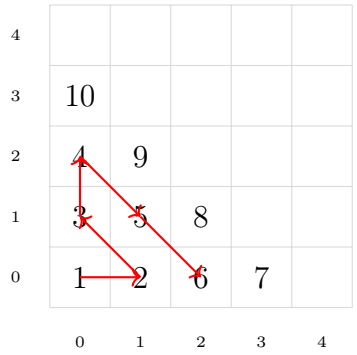
One explicit construction uses **Cantor's pairing function**:

$$f(m, n) = \frac{(m + n)(m + n + 1)}{2} + n$$

This is injective and surjective (proof omitted, but verifiable).

Alternatively, we can visualize pairs (m, n) in a grid and traverse them diagonally:

$\mathbb{N} \times \mathbb{N}$ enumerated by diagonals



Follow the red path: we hit every pair (m, n) eventually.

Therefore $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. ■

Corollary 7.4. *If A and B are countable, then $A \times B$ is countable.*

Corollary 7.5 (Finite Products are Countable)**.** *For any $k \in \mathbb{N}$, the Cartesian product $\mathbb{N}^k = \mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}$ (k times) is countable.*

Proof. By induction on k.

Base case ($k = 1$): $\mathbb{N}^1 = \mathbb{N}$ is countable by definition. ✓

Inductive step: Assume $\mathbb{N}^k$ is countable.

Then:

$$\mathbb{N}^{k+1} = \mathbb{N}^k \times \mathbb{N}$$

Since $\mathbb{N}^k$ and $\mathbb{N}$ are both countable (by inductive hypothesis and definition), the previous corollary shows their product $\mathbb{N}^{k+1}$ is countable. ✓

Therefore, by induction, $\mathbb{N}^k$ is countable for all $k \in \mathbb{N}$. ■

Explicit Bijection $\mathbb{N}^k \to \mathbb{N}$

While the inductive proof is elegant, one can also construct explicit bijections.
For $\mathbb{N}^2 \to \mathbb{N}$: Cantor's pairing function given above.
For $\mathbb{N}^3 \to \mathbb{N}$: Compose pairings:

$$\mathbb{N}^3 = \mathbb{N} \times \mathbb{N} \times \mathbb{N} \xrightarrow{f \times \mathrm{id}} \mathbb{N} \times \mathbb{N} \xrightarrow{f} \mathbb{N}$$

where f is Cantor's pairing function. Explicitly:

$$g(m, n, p) = f(f(m, n), p)$$

For general $\mathbb{N}^k$: Iterate Cantor's pairing function $k - 1$ times:

$$h_k(n_1, n_2, \ldots, n_k) = f(n_1, f(n_2, f(n_3, \ldots, f(n_{k-1}, n_k) \ldots)))$$

Or, more computationally, use a **prime factorization encoding**:

$$h(n_1, n_2, \ldots, n_k) = 2^{n_1} \cdot 3^{n_2} \cdot 5^{n_3} \cdots p_k^{n_k}$$

where $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$ are the first k primes.
By the Fundamental Theorem of Arithmetic, each natural number has a unique prime factorization, so this map is injective. However, it is not surjective (not every natural number is a product of only the first k primes), so this is an injection, not a bijection.
For a true bijection, Cantor's iterated pairing is simpler.

Theorem 7.6 (Countable Unions)**.** *1. If $\{A_n : n \in \mathbb{N}\}$ is a countable collection of finite sets, then $\bigcup_{n=1}^{\infty} A_n$ is countable.*

2. If $\{A_n : n \in \mathbb{N}\}$ is a countable collection of countable sets, then $\bigcup_{n=1}^{\infty} A_n$ is countable.

Proof. **(1) Countable union of finite sets:**

Let $A_1, A_2, A_3, \ldots$ be finite sets. We can assume they are pairwise disjoint (otherwise replace A_n with $A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$).

For each n, let $|A_n| = k_n$ (finite). Enumerate $A_n = \{a_{n,1}, a_{n,2}, \ldots, a_{n,k_n}\}$.

Consider the map $f : \bigcup_{n=1}^{\infty} A_n \to \mathbb{N} \times \mathbb{N}$ defined by:

$$f(a_{n,i}) = (n, i)$$

This is an injection (each element has a unique index pair). Since $\mathbb{N} \times \mathbb{N}$ is countable, and the union injects into it, the union is countable. ✓

(2) Countable union of countable sets:

Let $A_1, A_2, A_3, \ldots$ be countable sets. For each n, since A_n is countable, there exists a bijection $g_n : \mathbb{N} \to A_n$ (or $g_n : \mathbb{N} \to A_n$ surjective if A_n is finite, but we can pad with repetitions).

Define $f : \mathbb{N} \times \mathbb{N} \to \bigcup_{n=1}^{\infty} A_n$ by:

$$f(n, m) = g_n(m)$$

This is surjective: for any $a \in \bigcup A_n$, there exists n such that $a \in A_n$. Since g_n is a bijection (or surjection), there exists m such that $g_n(m) = a$. Thus $f(n, m) = a$. ✓

Since there is a surjection from $\mathbb{N} \times \mathbb{N}$ (which is countable) onto $\bigcup A_n$, the union is countable. ✓ ■

Remark

This theorem is crucial for proving countability of algebraic numbers and other important sets. The key insight: arranging elements in a "grid" (indexed by two naturals) and using Cantor's diagonal argument allows us to enumerate the entire union.

7.3.3 The Rationals are Countable

Theorem 7.7. $|\mathbb{Q}| = |\mathbb{N}| = \aleph_0$

Proof. We show $\mathbb{Q}^+$ (positive rationals) is countable; extending to all of $\mathbb{Q}$ is similar to the integer case.

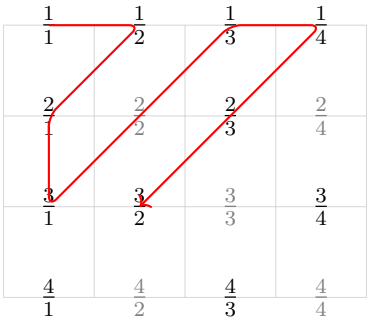
Every positive rational can be written as $\frac{p}{q}$ with $p, q \in \mathbb{N}$, $q > 0$, $\gcd(p, q) = 1$ (reduced form).

Arrange fractions in a grid:

$$\begin{aligned}
\text{Row 1: } & \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \\
\text{Row 2: } & \frac{2}{1}, \frac{2}{2}, \frac{2}{3}, \frac{2}{4}, \dots \\
\text{Row 3: } & \frac{3}{1}, \frac{3}{2}, \frac{3}{3}, \frac{3}{4}, \dots \\
& \vdots
\end{aligned}$$

Traverse diagonally (as with $\mathbb{N} \times \mathbb{N}$), but **skip** any fraction already seen in reduced form:

Zig-Zag Enumeration of $\mathbb{Q}^+$



Sequence: $\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{3}{1}, \frac{2}{3}, \frac{1}{3}, \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \dots$ (skip duplicates)

This path hits every positive rational exactly once in lowest terms.

Therefore $|\mathbb{Q}^+| = |\mathbb{N}|$.

For all of $\mathbb{Q}$ (including negatives and zero), use the same interleaving trick as with $\mathbb{Z}$.

Therefore $|\mathbb{Q}| = |\mathbb{N}| = \aleph_0$. ■

Remark

Stunning result: Between any two rationals, there are infinitely many more rationals (they're "dense" in $\mathbb{R}$), yet the rationals are the same size as the natural numbers! But this is where Cantor's next result shatters intuition...

7.4 Cantor's Diagonal Argument: The Reals are Uncountable

Theorem 7.8 (Cantor's Diagonal Argument, 1891)**.** *The interval* $(0, 1)$ *is uncountable:*

$$|(0, 1)| > |\mathbb{N}|$$

Proof. We prove by **contradiction**.

Suppose $(0, 1)$ were countable. Then we could list all real numbers in $(0, 1)$:

$$\begin{aligned} n = 1 &: \quad 0.d_{11}d_{12}d_{13}d_{14}d_{15}\ldots \\ n = 2 &: \quad 0.d_{21}d_{22}d_{23}d_{24}d_{25}\ldots \\ n = 3 &: \quad 0.d_{31}d_{32}d_{33}d_{34}d_{35}\ldots \\ n = 4 &: \quad 0.d_{41}d_{42}d_{43}d_{44}d_{45}\ldots \\ &\vdots \end{aligned}$$

where $d_{ij} \in \{0, 1, 2, \ldots, 9\}$ are decimal digits.

We construct a number $X = 0.x_1x_2x_3x_4\ldots$ that is *not* in this list:

Define x_n by the rule:

$$x_n = \begin{cases} 1 & \text{if } d_{nn} \neq 1 \\ 2 & \text{if } d_{nn} = 1 \end{cases}$$

(We ensure $x_n \neq 0, 9$ to avoid ambiguities like $0.999\ldots = 1.000\ldots$)

Key observation: X differs from the n-th number in the list at the n-th decimal place:

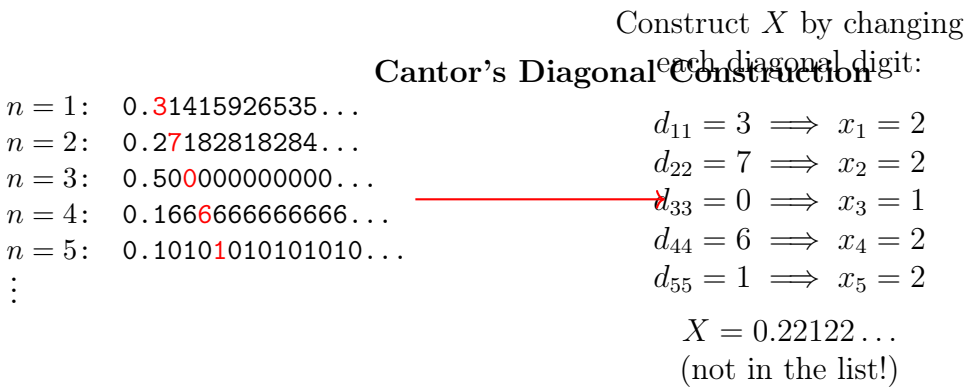
- X differs from number 1 at digit 1
- X differs from number 2 at digit 2
- X differs from number 3 at digit 3
- $\ldots$

Therefore X is not equal to any number in the list.

But $X \in (0, 1)$ (it's a valid decimal between 0 and 1).

Contradiction: We assumed our list contained *all* numbers in $(0, 1)$, but we just constructed a number not in the list.

Therefore, no such list exists. $(0, 1)$ is uncountable. ∎



X differs from every listed number at some digit,
so it cannot be in the list.
Contradiction ⇒ No such list exists.

Key Idea

The diagonal argument is a **self-referential impossibility proof**:

1. Assume we have a "complete" list
2. Use the list itself to construct something missing from it
3. Conclude the list cannot be complete

This pattern appears throughout mathematics and logic (Gödel's incompleteness theorem, Turing's halting problem, Russell's paradox).

Theorem 7.9. $|\mathbb{R}| = |(0,1)|$

Proof. We construct an explicit bijection $f : (0,1) \to \mathbb{R}$.

Define:

$$f(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$$

As x ranges from 0 to 1:

- When $x \to 0^+$: $f(x) \to -\infty$
- When $x = 0.5$: $f(x) = 0$
- When $x \to 1^-$: $f(x) \to +\infty$

This function is a bijection (tangent function restricted to $(-\pi/2, \pi/2)$ is bijective to $\mathbb{R}$).

Therefore $|\mathbb{R}| = |(0,1)|$.

Since $(0,1)$ is uncountable, so is $\mathbb{R}$. ■

Corollary 7.10. *The cardinality of* $\mathbb{R}$ *is denoted* $\mathfrak{c}$ *(for "continuum") or* $2^{\aleph_0}$.

We have:

$$\aleph_0 = |\mathbb{N}| < |\mathbb{R}| = \mathfrak{c}$$

7.5 The Power Set Theorem: Infinitely Many Infinities

Intuition

Is there anything bigger than $\mathbb{R}$?
Yes! The power set $\mathcal{P}(\mathbb{R})$ (set of all subsets of $\mathbb{R}$) is strictly larger.
In fact, there's an infinite hierarchy:

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < \cdots$$

There is no "largest" infinity.

Theorem 7.11 (Cantor's Theorem, 1891)**.** *For any set A:*

$$|A| < |\mathcal{P}(A)|$$

Proof. We must show two things:

(1) $|A| \leq |\mathcal{P}(A)|$:

Define $f : A \to \mathcal{P}(A)$ by $f(a) = \{a\}$ (singleton set).

This is an injection (distinct elements map to distinct singletons).

Therefore $|A| \leq |\mathcal{P}(A)|$. ✓

(2) $|A| \neq |\mathcal{P}(A)|$:

We prove there is **no** surjection $g : A \to \mathcal{P}(A)$.

Suppose, for contradiction, that $g : A \to \mathcal{P}(A)$ is surjective.

For each $a \in A$, $g(a)$ is a subset of A. So either $a \in g(a)$ or $a \notin g(a)$.

Define the "diagonal set":

$$D = \{a \in A : a \notin g(a)\}$$

(The set of elements that are *not* in their own images.)

Since g is surjective, there must exist $d \in A$ with $g(d) = D$.

Now ask: Is $d \in D$?

Case 1: Suppose $d \in D$.

By definition of D: $d \in D \iff d \notin g(d)$.

But $g(d) = D$, so $d \notin D$.

Contradiction! (We assumed $d \in D$ but derived $d \notin D$.)

Case 2: Suppose $d \notin D$.

By definition of D: $d \notin D \iff d \in g(d)$.

But $g(d) = D$, so $d \in D$.

Contradiction! (We assumed $d \notin D$ but derived $d \in D$.)

Both cases lead to contradiction.

Therefore, no such d exists, so g cannot be surjective.

Since there's no surjection $A \to \mathcal{P}(A)$, we have $|A| \neq |\mathcal{P}(A)|$.

Combining (1) and (2): $|A| < |\mathcal{P}(A)|$. ■

Why No Surjection $A \to \mathcal{P}(A)$ Exists

Supposed mapping	Check membership
$g(a) = \{a, b\}$	$a \in g(a)$? Yes $\implies a \notin D$
$g(b) = \{c\}$	$b \in g(b)$? No $\implies b \in D$
$g(c) = \{a, c\}$	$c \in g(c)$? Yes $\implies c \notin D$
$g(d) = \emptyset$	$d \in g(d)$? No $\implies d \in D$
$\vdots$	$\vdots$

Construct $D = \{b, d, \ldots\}$ (elements not in their own images).
If g were surjective, some x would satisfy $g(x) = D$.
But asking "is $x \in D$?" leads to contradiction either way.

Warning

Cantor's theorem is closely related to logical paradoxes:
Russell's Paradox (1901): Let $R = \{x : x \notin x\}$ (the set of all sets that don't contain themselves). Is $R \in R$?
This paradox led to the crisis in foundations of mathematics and the development of axiomatic set theory (ZFC) to avoid such contradictions.
Cantor's diagonal set D uses the same self-referential structure but avoids paradox by working within a fixed set A.

Corollary 7.12 (Hierarchy of Infinities)**.** *There is an infinite sequence of strictly increasing cardinalities:*

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}))| < \cdots$$

Equivalently, using aleph notation:

$$\aleph_0 < 2^{\aleph_0} < 2^{2^{\aleph_0}} < \cdots$$

There is no "largest" infinity.

7.6 The Schröder-Bernstein Theorem

Intuition

Suppose $|A| \leq |B|$ (injection $A \to B$) and $|B| \leq |A|$ (injection $B \to A$).
Intuitively, this suggests $|A| = |B|$.
But how do we construct a bijection?
The Schröder-Bernstein theorem guarantees one exists.

Theorem 7.13 (Schröder-Bernstein Theorem)**.** *If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Equivalently: If there exist injections $f : A \to B$ and $g : B \to A$, then there exists a bijection $h : A \to B$.

Proof Sketch. The full proof is intricate, but the idea is:

1. Start with injections $f : A \to B$ and $g : B \to A$
2. Partition A into:
   - Elements that "come from B" (in the image of g)
   - Elements that don't (outside the image of g)
3. Define the bijection $h : A \to B$ by:

$$h(a) = \begin{cases} f(a) & \text{if } a \text{ is not in } \mathrm{Im}(g) \\ g^{-1}(a) & \text{if } a \text{ is in } \mathrm{Im}(g) \end{cases}$$

4. Verify h is indeed a bijection (uses careful case analysis)

The full rigorous proof requires iterating the injections and using limiting arguments. ■

Example 7.1. *We can use Schröder-Bernstein to show $|(0,1)| = |\mathbb{R}|$ without explicit bijection:*

- *Injection $(0,1) \to \mathbb{R}$: $f(x) = x$ (inclusion)*
- *Injection $\mathbb{R} \to (0,1)$: $g(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$*

By Schröder-Bernstein: $|(0,1)| = |\mathbb{R}|$. ✓

7.7 Cardinal Arithmetic

7.7.1 Addition and Multiplication

Definition 7.3 (Cardinal Arithmetic)**.** *For cardinals $\kappa = |A|$ and $\lambda = |B|$ (where $A \cap B = \emptyset$):*

***Addition**: $\kappa + \lambda = |A \cup B|$ (disjoint union)*
***Multiplication**: $\kappa \cdot \lambda = |A \times B|$ (Cartesian product)*
***Exponentiation**: $\kappa^\lambda = |B^A|$ (set of all functions $A \to B$)*

Theorem 7.14 (Arithmetic with $\aleph_0$)**.** *1.* $\aleph_0 + 1 = \aleph_0$

2. $\aleph_0 + \aleph_0 = \aleph_0$

3. $\aleph_0 \cdot \aleph_0 = \aleph_0$

4. $2^{\aleph_0} = \mathfrak{c} = |\mathbb{R}|$

Proof. **(1)** $|\mathbb{N}| + 1 = |\mathbb{N} \cup \{*\}|$ where $* \notin \mathbb{N}$.

Bijection: $f(n) = n + 1$ for $n \in \mathbb{N}$, $f(*) = 0$. ✓

(2) $|\mathbb{N}| + |\mathbb{N}| = |\mathbb{N} \cup (\mathbb{N} \times \{1\})|$.

Use the even-odd trick: evens → first copy, odds → second copy. ✓

(3) We proved $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$ via Cantor pairing. ✓

(4) We know $|\mathcal{P}(\mathbb{N})| = 2^{\aleph_0}$ (set of all subsets of $\mathbb{N}$).

Each subset corresponds to an infinite binary sequence (characteristic function).

Binary sequences correspond to reals in $(0, 1)$ in binary expansion.

Therefore $2^{\aleph_0} = |\mathcal{P}(\mathbb{N})| = |(0, 1)| = |\mathbb{R}| = \mathfrak{c}$. ✓ ■

7.7.2 The Continuum Hypothesis

Key Idea

We know:

$$\aleph_0 < \mathfrak{c} = 2^{\aleph_0}$$

But are there any cardinalities *between* $\aleph_0$ and $\mathfrak{c}$?

[Continuum Hypothesis (CH)] There is no set S with:

$$|\mathbb{N}| < |S| < |\mathbb{R}|$$

Equivalently: $2^{\aleph_0} = \aleph_1$ (the "next" infinity after $\aleph_0$).

Remark

The continuum hypothesis was Hilbert's first problem in his famous 1900 list.
Gödel (1940): Proved CH cannot be disproved from ZFC axioms.
Cohen (1963): Proved CH cannot be proved from ZFC axioms.
Conclusion: CH is **independent** of ZFC—it's neither true nor false within standard set theory!
You can do mathematics assuming CH is true, or assuming it's false—both are consistent.

7.8 Applications and Implications

Example 7.2 (Almost All Reals are Transcendental)**.** ***Algebraic numbers****: Roots of polynomials with integer coefficients (e.g.,* $\sqrt{2}$*,* $\sqrt[3]{5}$*).*

Transcendental numbers*: Not algebraic (e.g.,* π*,* e*).*

Claim*: The algebraic numbers are countable, but the reals are uncountable.*

Therefore, "almost all" reals are transcendental!

Proof. Each polynomial $a_n x^n + \cdots + a_1 x + a_0$ with integer coefficients corresponds to a finite tuple $(a_n, \ldots, a_0) \in \mathbb{Z}^{n+1}$.

The set of all such tuples is countable (countable union of countable sets).

Each polynomial has finitely many roots.

Therefore, the set of algebraic numbers is a countable union of finite sets, hence countable.

But $|\mathbb{R}| = \mathfrak{c} > \aleph_0$, so the transcendental numbers are uncountable. ■

Example 7.3 (Most Functions are Not Computable)**.** *A function $f : \mathbb{N} \to \mathbb{N}$ is **computable** if there's an algorithm to compute $f(n)$.*

***Claim**: Most functions $\mathbb{N} \to \mathbb{N}$ are not computable.*

Proof. The set of computable functions is countable (each algorithm is a finite string, and there are countably many finite strings).

But $|\mathbb{N}^{\mathbb{N}}| = |\mathbb{N}|^{|\mathbb{N}|} = \aleph_0^{\aleph_0} = \mathfrak{c}$ (uncountable).

Therefore, the set of non-computable functions is uncountable. ■

This shows that the computable world is a tiny fraction of the mathematical universe.

7.9 Looking Forward

Cardinality is the foundation for:

- **Measure Theory**: Lebesgue measure, probability spaces
- **Topology**: Separability, compactness, connectedness
- **Computability Theory**: Decidability, Turing machines, complexity
- **Cardinal Arithmetic**: Generalized continuum hypothesis, large cardinals

We've seen that infinity comes in many sizes, and that bijections are the key to comparing them.

Next, we embark on one of the most significant constructions in mathematics: rigorously building the **Real Numbers** from the rationals, finally filling the "gaps" and laying the groundwork for calculus.

The Infinite Hierarchy

$$\underbrace{|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|}_{\aleph_0 \text{ (countable)}} < \underbrace{|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|}_{\mathfrak{c}=2^{\aleph_0}} < |\mathcal{P}(\mathbb{R})| < |\mathcal{P}(\mathcal{P}(\mathbb{R}))| < \cdots$$

Each step unlocks a new realm of mathematical objects.

Chapter 8

The Real Numbers: Completing the Line

8.1 The Crisis of Incompleteness

Intuition

We have constructed the rationals $\mathbb{Q}$, and they seem to fill the number line. Between any two rationals, there is another rational. They are *dense*.
But the ancient Greeks discovered a terrifying secret: $\mathbb{Q}$ has holes.
Consider the length of the diagonal of a unit square. By Pythagoras, it is a number x such that $x^2 = 2$. Does such a number exist in $\mathbb{Q}$?

Theorem 8.1 (Irrationality of $\sqrt{2}$)**.** *There is no rational number $q \in \mathbb{Q}$ such that $q^2 = 2$.*

Proof. Suppose, for the sake of contradiction, that $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{p}{q}$ for some $p, q \in \mathbb{Z}, q \neq 0$. We may assume the fraction is in lowest terms, i.e., $\gcd(p, q) = 1$.

Squaring both sides:

$$2 = \frac{p^2}{q^2} \implies p^2 = 2q^2$$

This means p^2 is even. Therefore p must be even (since the square of an odd number is odd). So $p = 2k$ for some integer k.

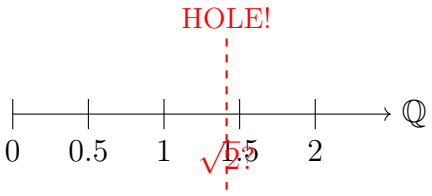
Substitute back:

$$(2k)^2 = 2q^2 \implies 4k^2 = 2q^2 \implies 2k^2 = q^2$$

This means q^2 is even, so q must be even.

Contradiction: We found that both p and q are even, meaning they share a common factor of 2. But we assumed $\gcd(p, q) = 1$.

Therefore, no such rational number exists. ∎



The rational line is full of gaps where irrational numbers should be.

8.2 Dedekind Cuts: Constructing the Continuum

How do we fill these gaps? Richard Dedekind (1872) had a brilliant insight: **Define a real number by the set of all rational numbers smaller than it.**

Instead of trying to grasp the elusive $\sqrt{2}$ directly, we look at its "shadow" on the rational line: all rationals q such that $q^2 < 2$ (or $q < 0$).

Definition 8.1 (Dedekind Cut)**.** *A* ***Dedekind cut*** *is a subset $\alpha \subseteq \mathbb{Q}$ such that:*

1. ***Non-trivial****: $\alpha \neq \emptyset$ and $\alpha \neq \mathbb{Q}$.*
2. ***Closed downwards****: If $p \in \alpha$ and $q < p$, then $q \in \alpha$.*
3. ***No greatest element****: If $p \in \alpha$, there exists $r \in \alpha$ such that $p < r$.*

The set of all Dedekind cuts is denoted by $\mathbb{R}$.

Intuition

Think of cutting the rational number line with a pair of scissors. The cut divides $\mathbb{Q}$ into two parts: the left set (L) and the right set (R). We identify the real number with the left set $\alpha = L$.

- Condition 1 ensures we don't pick "nothing" or "everything".
- Condition 2 means α is an initial segment of the line.
- Condition 3 is a technical convenience (it avoids ambiguity for rational numbers like "everything strictly less than 1" vs "everything less than or equal to 1"). We choose strict inequality.

Alternative Construction: Cauchy Sequences

The Dedekind cut approach is not the only way to construct $\mathbb{R}$ from $\mathbb{Q}$. Another classical method uses **Cauchy sequences**.
Cauchy Sequence Approach:
A sequence (q_n) of rational numbers is **Cauchy** if for every $\varepsilon \in \mathbb{Q}$ with $\varepsilon > 0$, there

exists $N \in \mathbb{N}$ such that for all $m, n \geq N$:

$$|q_m - q_n| < \varepsilon$$

Intuitively, the terms get arbitrarily close to each other. In $\mathbb{Q}$, Cauchy sequences need not converge (e.g., the sequence $1, 1.4, 1.41, 1.414, \ldots$ approximating $\sqrt{2}$ is Cauchy in $\mathbb{Q}$ but has no rational limit).
We define:

- Two Cauchy sequences (p_n) and (q_n) are **equivalent** if $\lim_{n\to\infty} |p_n - q_n| = 0$ (in the sense that for all $\varepsilon > 0$, eventually $|p_n - q_n| < \varepsilon$).
- A real number is an equivalence class of Cauchy sequences of rationals.
- $\mathbb{R} := \{\text{Cauchy sequences in } \mathbb{Q}\}/\sim$

Operations are defined component-wise: $(p_n) + (q_n) := (p_n + q_n)$, etc.
Comparison of the Two Constructions:

Aspect	**Dedekind Cuts**	**Cauchy Sequences**
Definition	A set of rationals (an initial segment)	An equivalence class of sequences
Intuition	"All rationals less than x"	"Sequences converging to x"
Order	Natural: $\alpha < \beta$ iff $\alpha \subsetneq \beta$	Requires definition via representatives
Operations	Set-theoretic (unions, products of sets)	Component-wise on sequences
Pros	Order structure immediate; completeness easy to prove	Generalizes to abstract metric spaces; dynamic/constructive feel
Cons	Abstract (real number as infinite set); operations technical	Equivalence relation subtle; order less natural
Historical	Dedekind (1872)	Cantor (1872), Méray (1869)

Are They the Same?
Yes! Both constructions yield *isomorphic* complete ordered fields. There is a natural bijection:

- Given a Dedekind cut α, construct a Cauchy sequence (q_n) where $q_n \in \alpha$ approaches the "boundary" of α from below.
- Given a Cauchy sequence (q_n), define the cut $\alpha = \{r \in \mathbb{Q} : r < q_n \text{ for infinitely many } n\}$.

This bijection preserves $+$, $\cdot$, and $<$, making the two constructions mathematically equivalent.
Why We Choose Dedekind Cuts Here:
We use Dedekind cuts because:

- Order is immediately transparent (set inclusion)
- Completeness (least upper bound property) is straightforward to verify
- They fit naturally in our set-theoretic framework built from ZFC

However, Cauchy sequences are indispensable in analysis and topology, where metric completeness generalizes beyond $\mathbb{R}$ to arbitrary metric spaces. Both perspectives enrich understanding.

Example 8.1 (Representing Numbers)**.** ***The Rational 1****:*

$$1^* = \{q \in \mathbb{Q} : q < 1\}$$

This is a cut representing the real number 1.

The Irrational $\sqrt{2}$*:*

$$\sqrt{2}^* = \{q \in \mathbb{Q} : q < 0 \text{ or } q^2 < 2\}$$

This set contains all negative rationals, and positive rationals whose square is less than 2 (e.g., 1, 1.4, 1.41...). It satisfies all conditions of a cut.

8.3 Ordering the Reals

Defining order on Dedekind cuts is elegant: it's just subset inclusion.

Definition 8.2 (Order on $\mathbb{R}$)**.** *Let $\alpha, \beta \in \mathbb{R}$. We define:*

$$\alpha \leq \beta \iff \alpha \subseteq \beta$$

Theorem 8.2. *$(\mathbb{R}, \leq)$ is a total order.*

Proof. Reflexivity and transitivity follow immediately from set inclusion properties. Total ordering (comparability) relies on the property of cuts: if $\alpha \not\subseteq \beta$, there is a rational $p \in \alpha$ such that $p \notin \beta$. Since β is closed downwards, this implies p is greater than or equal to every element in β, essentially meaning $\beta \subset \alpha$ (with minor technical details to fill). ■

8.4 Arithmetic on $\mathbb{R}$

8.4.1 Addition

Definition 8.3 (Addition)**.** *Let $\alpha, \beta \in \mathbb{R}$. Their sum is defined as the set of all sums of their elements:*

$$\alpha + \beta := \{x + y : x \in \alpha, y \in \beta\}$$

We must verify that $\alpha + \beta$ is indeed a Dedekind cut (non-empty, closed downwards, no max).

- If $x \in \alpha, y \in \beta$, and $z < x + y$, is $z \in \alpha + \beta$? Yes. Let $\delta = (x + y) - z$. Then $(x - \delta/2) \in \alpha$ and $(y - \delta/2) \in \beta$, summing to z.

8.4.2 Multiplication

Multiplication is trickier due to negative numbers (multiplying two large negative numbers gives a large positive number, not a small one).

Definition 8.4 (Multiplication of Positive Reals)**.** *For $\alpha, \beta > 0^*$ (positive cuts), define:*

$$\alpha \cdot \beta = \{p \cdot q : p \in \alpha, p > 0, q \in \beta, q > 0\} \cup \{r \in \mathbb{Q} : r \leq 0\}$$

***Interpretation**: We take all products of positive elements from α and β, and include all non-positive rationals to ensure the result is closed downwards.*

Theorem 8.3. *If $\alpha, \beta > 0^*$, then $\alpha \cdot \beta$ is a Dedekind cut.*

Proof Sketch. We verify the three conditions:

(1) Non-trivial: $\alpha \cdot \beta \neq \emptyset$ (contains 0 and negative rationals). Also $\alpha \cdot \beta \neq \mathbb{Q}$ because if $p \in \alpha$ is positive and bounded, and $q \in \beta$ is positive and bounded, then any rational larger than all possible products pq is not in $\alpha \cdot \beta$.

(2) Closed downwards: If $r \in \alpha \cdot \beta$ and $s < r$:

- If $r \leq 0$, then $s < 0$, so $s \in \alpha \cdot \beta$ by definition.
- If $r > 0$, then $r = p \cdot q$ for some $p \in \alpha, p > 0, q \in \beta, q > 0$.

  Since $s < r = pq$ and both $p, q > 0$, we can find $p' \in \alpha$ with $p' < p$ close enough that $p' \cdot q < s < p \cdot q$.

  Then s can be written as a product of positive elements from the cuts (or is ≤ 0), so $s \in \alpha \cdot \beta$.

(3) No maximum: If $r \in \alpha \cdot \beta$ and $r > 0$, say $r = p \cdot q$ with $p \in \alpha, q \in \beta$. Since α has no maximum, there exists $p' \in \alpha$ with $p' > p$. Then $p' \cdot q > r$ and $p' \cdot q \in \alpha \cdot \beta$. ✓ ■

Definition 8.5 (Multiplication for All Reals)**.** *For arbitrary $\alpha, \beta \in \mathbb{R}$, define multiplication using sign rules:*

1. *If $\alpha, \beta > 0^*$: Use the definition above*
2. *If $\alpha > 0^*, \beta < 0^*$: Define $\alpha \cdot \beta = -(\alpha \cdot (-\beta))$*
3. *If $\alpha < 0^*, \beta > 0^*$: Define $\alpha \cdot \beta = -((-\alpha) \cdot \beta)$*
4. *If $\alpha, \beta < 0^*$: Define $\alpha \cdot \beta = (-\alpha) \cdot (-\beta)$*
5. *If $\alpha = 0^*$ or $\beta = 0^*$: Define $\alpha \cdot \beta = 0^*$*

where $-\alpha = \{q \in \mathbb{Q} : \exists r \notin \alpha, q < -r\}$ (additive inverse).

Theorem 8.4 (Field Properties)**.** *$(\mathbb{R}, +, \cdot)$ forms a field:*

- *Addition and multiplication are associative and commutative*
- *Distributive law: $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$*

- *Additive identity* 0^* *and multiplicative identity* 1^*
- *Every element has an additive inverse*
- *Every non-zero element has a multiplicative inverse*

Verification of Field Axioms for $\mathbb{R}$. We verify the key axioms. The others follow similarly.

(1) Commutativity of Addition: $\alpha + \beta = \beta + \alpha$

For cuts α and β, by definition:

$$\alpha + \beta = \{r + s : r \in \alpha, s \in \beta\}$$

$$\beta + \alpha = \{s + r : s \in \beta, r \in \alpha\}$$

Since addition in $\mathbb{Q}$ is commutative ($r + s = s + r$), these sets are equal. ✓

(2) Associativity of Addition: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

Both sides equal $\{r + s + t : r \in \alpha, s \in \beta, t \in \gamma\}$ by associativity in $\mathbb{Q}$. ✓

(3) Additive Identity: $\alpha + \mathbf{0} = \alpha$ where $\mathbf{0} = \{r \in \mathbb{Q} : r < 0\}$

By definition:

$$\alpha + \mathbf{0} = \{r + s : r \in \alpha, s \in \mathbf{0}\}$$

For any $a \in \alpha$, pick $s < 0$ in $\mathbf{0}$ such that $a + s \in \alpha$ (possible since α has no greatest element). Conversely, if $a + s \in \alpha + \mathbf{0}$ with $s < 0$, then $a \in \alpha$. Thus $\alpha + \mathbf{0} = \alpha$. ✓

(4) Commutativity of Multiplication: $\alpha \cdot \beta = \beta \cdot \alpha$

For positive cuts, by definition (assuming $\alpha, \beta > \mathbf{0}$):

$$\alpha \cdot \beta = \{rs : r \in \alpha, s \in \beta, r > 0, s > 0\} \cup \{q \in \mathbb{Q} : q \leq 0\}$$

Since $rs = sr$ in $\mathbb{Q}$, we have $\alpha \cdot \beta = \beta \cdot \alpha$. ✓

For negative cuts, the sign rules ensure commutativity is preserved. ✓

(5) Associativity of Multiplication: $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$

For positive cuts, both sides equal:

$$\{rst : r \in \alpha, s \in \beta, t \in \gamma, r, s, t > 0\} \cup \{q \in \mathbb{Q} : q \leq 0\}$$

by associativity in $\mathbb{Q}$. ✓

(6) Multiplicative Identity: $\alpha \cdot \mathbf{1} = \alpha$ where $\mathbf{1} = \{r \in \mathbb{Q} : r < 1\}$

For $\alpha > \mathbf{0}$:

$$\alpha \cdot \mathbf{1} = \{rs : r \in \alpha, s \in \mathbf{1}, r > 0, s > 0\} \cup \{q \in \mathbb{Q} : q \leq 0\}$$

For any $a \in \alpha$ with $a > 0$, pick $s \in \mathbf{1}$ with $s > 0$ close to 1 such that $as \in \alpha$ (using density of $\mathbb{Q}$). Conversely, if $as \in \alpha \cdot \mathbf{1}$ with $s < 1$, then $a \in \alpha$ (since $as < a$ and α is downward closed). Thus $\alpha \cdot \mathbf{1} = \alpha$. ✓

(7) Distributivity: $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$

This is the most involved axiom. We verify for positive cuts $\alpha, \beta, \gamma > \mathbf{0}$.

Left-hand side:

$$\beta + \gamma = \{s + t : s \in \beta, t \in \gamma\}$$

$$\alpha \cdot (\beta + \gamma) = \{r(s + t) : r \in \alpha, s \in \beta, t \in \gamma, r, s, t > 0\} \cup \{q : q \leq 0\}$$

Right-hand side:

$$\alpha \cdot \beta = \{rs : r \in \alpha, s \in \beta, r > 0, s > 0\} \cup \{q : q \leq 0\}$$

$$\alpha \cdot \gamma = \{rt : r \in \alpha, t \in \gamma, r > 0, t > 0\} \cup \{q : q \leq 0\}$$

$$\alpha \cdot \beta + \alpha \cdot \gamma = \{rs + rt : r \in \alpha, s \in \beta, t \in \gamma, r, s, t > 0\} \cup \{q : q \leq 0\}$$

By distributivity in $\mathbb{Q}$: $r(s + t) = rs + rt$, so both sides are equal. ✓

For mixed signs (some cuts negative), the sign rules ensure distributivity holds by reducing to the positive case with appropriate sign adjustments. The verification is case-by-case but follows the same logic. ✓

Therefore $\mathbb{R}$ satisfies all field axioms. ■

Key Idea

Why this construction works:
We've built $\mathbb{R}$ from $\mathbb{Q}$ using only set theory. The operations $+$ and $\cdot$ on cuts are defined so that:

- They extend the operations on $\mathbb{Q}$ (rational cuts behave like rationals)
- They preserve algebraic properties (field axioms)
- They respect order ($\alpha < \beta \implies \alpha + \gamma < \beta + \gamma$)
- They fill the gaps (completeness property)

The price we pay is abstraction: real numbers are now *infinite sets* of rationals! But this gives us rigorous foundations for calculus.

8.5 Absolute Value and Distance

To talk about "closeness" and "limits" in the next chapter, we need a way to measure size and distance.

Definition 8.6 (Absolute Value)**.** *For $x \in \mathbb{R}$, the **absolute value** $|x|$ is defined as:*

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Theorem 8.5 (Properties of Absolute Value)**.** *For all $x, y \in \mathbb{R}$:*

1. ***Non-negativity**: $|x| \geq 0$, and $|x| = 0 \iff x = 0$.*
2. ***Multiplicativity**: $|xy| = |x||y|$.*

3. ***Triangle Inequality***: $|x+y| \leq |x| + |y|$.

Proof of Triangle Inequality. Notice that $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. Adding these inequalities:

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

This is equivalent to $|x+y| \leq |x| + |y|$. ■

Definition 8.7 (Distance)**.** *The **distance** between two real numbers x and y is defined as:*

$$d(x, y) := |x - y|$$

Remark

This distance function d makes $\mathbb{R}$ into a **metric space**. It satisfies:

- $d(x, y) \geq 0$
- $d(x, y) = d(y, x)$ (Symmetry)
- $d(x, z) \leq d(x, y) + d(y, z)$ (Triangle Inequality for distance)

This metric is the foundation of all analysis on the real line.

8.6 Topology of the Real Line

The structure of "open" and "closed" sets on $\mathbb{R}$ is crucial for rigorously defining continuity, limits, and convergence. While a full course on topology is beyond our scope, we establish the essential definitions needed for analysis.

Definition 8.8 (Open Intervals)**.** *For $a, b \in \mathbb{R}$ with $a < b$, the **open interval** is:*

$$(a, b) := \{x \in \mathbb{R} : a < x < b\}$$

We also define unbounded open intervals:

$$(a, \infty) := \{x \in \mathbb{R} : x > a\}, \quad (-\infty, b) := \{x \in \mathbb{R} : x < b\}, \quad (-\infty, \infty) := \mathbb{R}$$

Definition 8.9 (Open Sets)**.** *A subset $U \subseteq \mathbb{R}$ is called **open** if:*

For every $x \in U$, there exists $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq U$.

Intuitively: every point in U is surrounded by a "buffer zone" also contained in U.

Example 8.2.

- $(0, 1)$ *is open: for any $x \in (0, 1)$, take $\varepsilon = \min(x, 1 - x)/2$ to get $(x - \varepsilon, x + \varepsilon) \subseteq (0, 1)$.*
- $[0, 1]$ *is not open: the point $0 \in [0, 1]$, but for any $\varepsilon > 0$, the interval $(0 - \varepsilon, 0 + \varepsilon) = (-\varepsilon, \varepsilon)$ contains $-\varepsilon/2 \notin [0, 1]$.*
- $\mathbb{R}$ *is open (vacuously satisfied for all points).*
- $\emptyset$ *is open (vacuously satisfied, no points to check).*

Theorem 8.6 (Properties of Open Sets)**.** *The collection of open sets in $\mathbb{R}$ satisfies:*

1. *$\emptyset$ and $\mathbb{R}$ are open.*
2. *The union of any collection of open sets is open.*
3. *The intersection of finitely many open sets is open.*

Proof Sketch. **(1)** Already verified above.

(2) If $\{U_i : i \in I\}$ is a collection of open sets, let $U = \bigcup_{i \in I} U_i$.

For any $x \in U$, there exists $i \in I$ such that $x \in U_i$. Since U_i is open, there exists $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq U_i \subseteq U$. ✓

(3) Let $U_1, \ldots, U_n$ be open sets, and let $U = \bigcap_{j=1}^{n} U_j$.

For any $x \in U$, we have $x \in U_j$ for all j. For each j, there exists $\varepsilon_j > 0$ such that $(x - \varepsilon_j, x + \varepsilon_j) \subseteq U_j$.

Let $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_n) > 0$. Then $(x - \varepsilon, x + \varepsilon) \subseteq U_j$ for all j, so $(x - \varepsilon, x + \varepsilon) \subseteq U$. ✓ ■

Warning

Infinite intersections of open sets need not be open!
Consider $U_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$ for $n \in \mathbb{N}$. Each U_n is open.
But:

$$\bigcap_{n=1}^{\infty} U_n = \{0\}$$

which is *not* open (no ε-neighborhood around 0 fits inside $\{0\}$).

Definition 8.10 (Closed Sets)**.** *A subset $F \subseteq \mathbb{R}$ is called **closed** if its complement $\mathbb{R} \setminus F$ is open.*

Example 8.3.

- *$[0, 1]$ is closed: $\mathbb{R} \setminus [0, 1] = (-\infty, 0) \cup (1, \infty)$ is open (union of open sets).*
- *$(0, 1)$ is not closed: $\mathbb{R} \setminus (0, 1) = (-\infty, 0] \cup [1, \infty)$ is not open (contains 0 and 1 without neighborhoods).*
- *$\mathbb{R}$ is both open and closed.*
- *$\emptyset$ is both open and closed.*
- *$[0, 1)$ is neither open nor closed.*

Definition 8.11 (Topology on $\mathbb{R}$)**.** *The collection $\mathcal{T}$ of all open subsets of $\mathbb{R}$ is called the **standard topology** on $\mathbb{R}$ (or the **Euclidean topology**).*

*The pair $(\mathbb{R}, \mathcal{T})$ is a **topological space**.*

Remark

This topology is induced by the metric $d(x, y) = |x - y|$. In general, any metric space has a natural topology where open sets are unions of open balls.
All the key notions of calculus—continuity, limits, compactness—can be defined using only open sets, making topology the natural language of analysis.

8.7 The Completeness Axiom

This is the crowning jewel of the real numbers. The "holes" are gone.

Definition 8.12 (Bounds and Suprema/Infima)**.** *Let $S \subseteq \mathbb{R}$.*

- *$M \in \mathbb{R}$ is an **upper bound** for S if $s \leq M$ for all $s \in S$.*
- *M is the **supremum** (least upper bound), denoted $\sup S$, if it is an upper bound and $M \leq K$ for any other upper bound K.*
- *$m \in \mathbb{R}$ is a **lower bound** for S if $s \geq m$ for all $s \in S$.*
- *m is the **infimum** (greatest lower bound), denoted $\inf S$, if it is a lower bound and $m \geq k$ for any other lower bound k.*

Theorem 8.7 (Least Upper Bound Property)**.** *Every non-empty subset of $\mathbb{R}$ that has an upper bound has a supremum in $\mathbb{R}$.*

Intuition

In $\mathbb{Q}$, the set $S = \{q \in \mathbb{Q} : q^2 < 2\}$ is bounded above (e.g., by 2), but has no least upper bound *in* $\mathbb{Q}$ (because $\sqrt{2} \notin \mathbb{Q}$). In $\mathbb{R}$, the supremum is exactly the cut for $\sqrt{2}$. The "hole" has been filled by the cut itself.

Proof Sketch. Let $\mathcal{A} \subseteq \mathbb{R}$ be a non-empty set of cuts bounded above. Consider the union of all these cuts:

$$\gamma = \bigcup_{\alpha \in \mathcal{A}} \alpha$$

Remarkably, γ itself is a Dedekind cut!

1. Since each α is a subset of rationals, $\gamma \subseteq \mathbb{Q}$.
2. γ is closed downwards because each α is.
3. γ is clearly $\geq \alpha$ for all $\alpha \in \mathcal{A}$ (superset relation).
4. γ is the *least* such bound.

Thus, $\sup \mathcal{A} = \bigcup \mathcal{A}$. The supremum is literally the union of the sets. ■

8.8 Density of Rationals

Even though $\mathbb{Q}$ is incomplete, it is **dense** in $\mathbb{R}$.

Theorem 8.8 (Density of $\mathbb{Q}$)**.** *For any two real numbers $x < y$, there exists a rational number q such that $x < q < y$.*

Proof. Let x, y be Dedekind cuts with $x \subsetneq y$. By definition of set inclusion, there exists a rational $q \in y$ such that $q \notin x$. Since x is closed downwards, $q \notin x$ implies q is greater than or equal to every element in x. Strictly speaking, we need slightly more care to find a q strictly "between", but the nature of cuts provides this rational witness immediately. ■

8.9 Looking Forward

We have built the **complete ordered field** $\mathbb{R}$.

- It allows us to solve equations like $x^2 = 2$.
- It has no gaps (sup always exists).

However, we defined reals as *sets* of rationals. In calculus, we often think of reals as limits of sequences (e.g., $3, 3.1, 3.14, 3.141, \dots$). In the next chapter, **Sequences and Convergence**, we will bridge these two views and rigorously define limits.

Chapter 9

Sequences and Convergence: The Foundation of Analysis

9.1 From Numbers to Processes

Intuition

We've built the real numbers $\mathbb{R}$ with their completeness property. Now we study **infinite processes**:
Question: What does it mean for values to "approach" something?

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots \to 0$$

$$1, 1.4, 1.41, 1.414, 1.4142, \ldots \to \sqrt{2}$$

This chapter makes "approaching" and "limit" precise. These concepts underpin all of calculus and analysis.

Historical Context

The Birth of Rigor in Analysis
Ancient Greeks (c. 300 BCE):

- Zeno's paradoxes: Achilles never catches the tortoise (infinite sums)
- Method of exhaustion: Approximating areas by polygons
- No formal notion of limit

17th-18th Century (Calculus Era):

- Newton (1665): Fluxions and infinitesimals
- Leibniz (1675): dx and dy as "infinitely small quantities"
- Euler (1748): Manipulated infinite series freely
- **Problem**: No rigorous foundation—what *is* an infinitesimal?

- Berkeley (1734): "Ghosts of departed quantities"—criticized lack of rigor

19th Century (Rigor Revolution):

- **Bolzano (1817)**: First rigorous definition of continuity
- **Cauchy (1821)**: *Cours d'Analyse*—sequences, limits, convergence
- **Weierstrass (1860s)**: ϵ-δ definitions ("arithmetization of analysis")
- **Dedekind (1872)**: Rigorous construction of $\mathbb{R}$
- Result: Calculus finally had solid foundations

Today, every analysis course begins with sequences and limits—the gateway to rigorous calculus.

9.2 Sequences: Infinite Ordered Lists

Definition 9.1 (Sequence)**.** *A **sequence** in $\mathbb{R}$ is a function $a : \mathbb{N}^+ \to \mathbb{R}$, where $\mathbb{N}^+ = \{1, 2, 3, \dots\}$.*

We write $a(n)$ as a_n and denote the sequence as:

$$(a_n)_{n=1}^{\infty} \quad \text{or} \quad (a_1, a_2, a_3, \dots) \quad \text{or simply} \quad (a_n)$$

*The value a_n is called the n**-th term** of the sequence.*

[Indexing] While our natural numbers $\mathbb{N}$ start at 0, it is standard in analysis to index sequences starting at $n = 1$ (matching the "1st term", "2nd term" intuition). Sometimes, however, we will start at $n = 0$ (e.g., for power series). The context will make this clear.

Remark

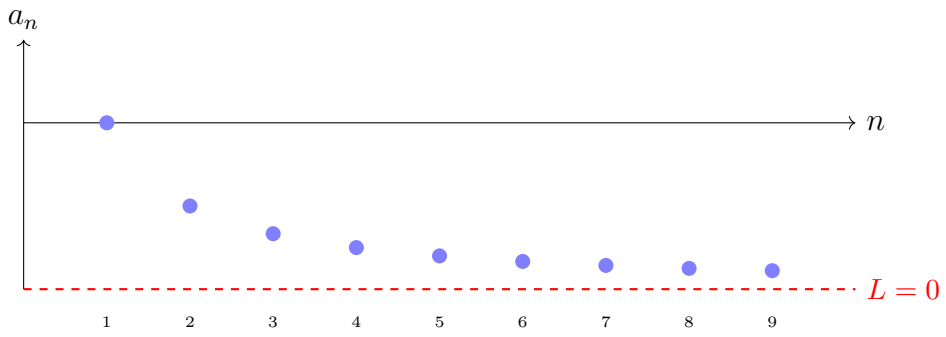
A sequence is fundamentally a function $\mathbb{N} \to \mathbb{R}$, so it's a special relation (a set of ordered pairs).
Everything traces back to sets, as always.

Example 9.1 (Common Sequences)**.**

1. ***Constant sequence**: $a_n = c$ for all n*
   Example: $(5, 5, 5, 5, \dots)$
2. ***Arithmetic sequence**: $a_n = a + (n-1)d$*
   Example: $(1, 3, 5, 7, 9, \dots)$ *with* $a = 1, d = 2$
3. ***Geometric sequence**: $a_n = ar^{n-1}$*
   Example: $(1, 2, 4, 8, 16, \dots)$ *with* $a = 1, r = 2$
4. ***Reciprocals**: $a_n = \frac{1}{n}$*
   Sequence: $\left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right)$
5. ***Alternating signs**: $a_n = \frac{(-1)^n}{n}$*
   Sequence: $\left(-1, \frac{1}{2}, -\frac{1}{3}, \frac{1}{4}, \dots\right)$

6. ***Rational approximations***: $a_n = \sum_{k=0}^{n} \frac{1}{k!}$
Sequence: $(1, 2, 2.5, 2.666\ldots, 2.708\ldots, \ldots) \to e$

Visualizing Sequences



Sequence $a_n = \frac{1}{n}$ approaches 0 as $n \to \infty$

9.3 Convergence: Making "Approaches" Precise

Intuition

When we say $(a_n) \to L$, we mean:
"The terms a_n get arbitrarily close to L as n increases."
Key insight: "Arbitrarily close" means: for *any* desired closeness $\epsilon > 0$ (no matter how small), eventually all terms are within ϵ of L.

Definition 9.2 (Convergence of a Sequence)**.** *A sequence (a_n)* ***converges*** *to a limit $L \in \mathbb{R}$ if:*

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n \geq N : |a_n - L| < \epsilon$$

We write:

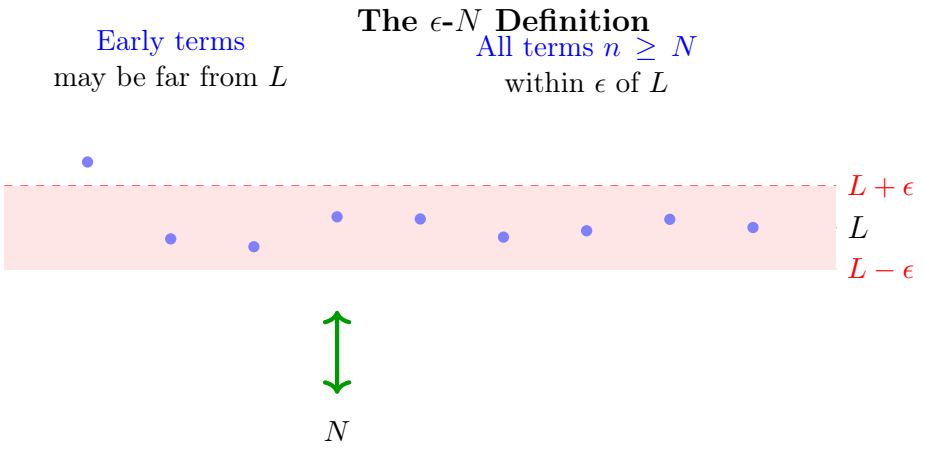
$$\lim_{n\to\infty} a_n = L \quad \text{or} \quad a_n \to L \quad \text{as } n \to \infty$$

If such an L exists, we say (a_n) is ***convergent****. Otherwise, (a_n) is* ***divergent****.*

Key Idea

The ϵ-N game:
Challenger: Gives you any $\epsilon > 0$ (a "tolerance")
You: Must find N such that all terms a_n with $n \geq N$ are within ϵ of L
If you can always win (for any ϵ), then $(a_n) \to L$.
Example: $a_n = \frac{1}{n} \to 0$
Challenger: "Get within $\epsilon = 0.01$"
You: "Choose $N = 100$. Then for $n \geq 100$, $|a_n - 0| = \frac{1}{n} \leq \frac{1}{100} = 0.01 < \epsilon$. I win!"
Challenger: "Get within $\epsilon = 0.00001$"

You: "Choose $N = 100000$. Done!"
For any ϵ, we can choose $N = \lceil \frac{1}{\epsilon} \rceil$. Therefore $\frac{1}{n} \to 0$.



Example 9.2 (Prove $\lim_{n\to\infty} \frac{1}{n} = 0$)**.** ***Claim:*** *The sequence $a_n = \frac{1}{n}$ converges to 0.*

Proof: *Let $\epsilon > 0$ be given (arbitrary).*

We need to find N such that for all $n \geq N$:

$$|a_n - 0| < \epsilon$$

This simplifies to:

$$\frac{1}{n} < \epsilon$$

Equivalently: $n > \frac{1}{\epsilon}$.

Choose $N = \lceil \frac{1}{\epsilon} \rceil + 1$ *(smallest integer greater than $\frac{1}{\epsilon}$).*

Then for all $n \geq N$:

$$n \geq N > \frac{1}{\epsilon} \implies \frac{1}{n} < \epsilon$$

Therefore $|a_n - 0| < \epsilon$ for all $n \geq N$.

Since ϵ was arbitrary, $\lim_{n\to\infty} \frac{1}{n} = 0$. ■

Example 9.3 (Non-convergent Sequence)**.** *Consider $a_n = (-1)^n$, so the sequence is* $(1, 1, 1, 1, 1, \ldots)$.

Claim: *This sequence does not converge.*

Proof: *Suppose for contradiction that $a_n \to L$ for some L.*

Choose $\epsilon = \frac{1}{2}$.

Then there exists N such that for all $n \geq N$: $|a_n - L| < \frac{1}{2}$.

Consider two consecutive terms: $a_N = (-1)^N$ and $a_{N+1} = (-1)^{N+1} = -a_N$.

Both satisfy $|a_N - L| < \frac{1}{2}$ and $|a_{N+1} - L| < \frac{1}{2}$.

By triangle inequality:

$$|a_N - a_{N+1}| \leq |a_N - L| + |L - a_{N+1}| < \frac{1}{2} + \frac{1}{2} = 1$$

But $|a_N - a_{N+1}| = |a_N - (-a_N)| = 2|a_N| = 2$, contradicting $2 < 1$.

Therefore no such L exists, and the sequence diverges. ■

9.4 Properties of Limits

Theorem 9.1 (Uniqueness of Limits)**.** *If (a_n) converges, its limit is unique.*

That is, if $a_n \to L$ and $a_n \to M$, then $L = M$.

Proof. Suppose $a_n \to L$ and $a_n \to M$ with $L \neq M$.

Let $\epsilon = \frac{|L-M|}{3} > 0$.

Since $a_n \to L$, there exists N_1 such that for all $n \geq N_1$: $|a_n - L| < \epsilon$.

Since $a_n \to M$, there exists N_2 such that for all $n \geq N_2$: $|a_n - M| < \epsilon$.

Let $N = \max(N_1, N_2)$. For $n \geq N$:

$$\begin{aligned} |L - M| &= |L - a_n + a_n - M| \\ &\leq |L - a_n| + |a_n - M| \quad \text{(triangle inequality)} \\ &< \epsilon + \epsilon = 2\epsilon = \frac{2|L-M|}{3} \end{aligned}$$

Therefore $|L - M| < \frac{2|L-M|}{3}$, which implies $\frac{|L-M|}{3} < 0$, contradiction.

Therefore $L = M$. ■ ■

Theorem 9.2 (Boundedness of Convergent Sequences)**.** *If (a_n) converges, then (a_n) is bounded.*

That is, there exists $M > 0$ such that $|a_n| \leq M$ for all n.

Proof. Suppose $a_n \to L$.

Choose $\epsilon = 1$. Then there exists N such that for all $n \geq N$: $|a_n - L| < 1$.

By triangle inequality: $|a_n| = |a_n - L + L| \leq |a_n - L| + |L| < 1 + |L|$.

So for $n \geq N$, we have $|a_n| < 1 + |L|$.

For $n < N$, there are only finitely many terms: $|a_1|, |a_2|, \ldots, |a_{N-1}|$.

Let $M = \max(|a_1|, |a_2|, \ldots, |a_{N-1}|, 1 + |L|)$.

Then $|a_n| \leq M$ for all n. ■ ■

Warning

The converse is **false**: A bounded sequence need not converge.
Example: $a_n = (-1)^n$ is bounded ($|a_n| = 1$ for all n) but does not converge.

Theorem 9.3 (Algebra of Limits)**.** *If $a_n \to L$ and $b_n \to M$, then:*

1. *$a_n + b_n \to L + M$ (sum rule)*
2. *$a_n - b_n \to L - M$ (difference rule)*
3. *$a_n \cdot b_n \to L \cdot M$ (product rule)*
4. *$\frac{a_n}{b_n} \to \frac{L}{M}$ if $M \neq 0$ and $b_n \neq 0$ for all n (quotient rule)*
5. *$ca_n \to cL$ for any constant $c \in \mathbb{R}$ (scalar multiplication)*

Proof of Sum Rule. Let $\epsilon > 0$ be given.

Since $a_n \to L$, there exists N_1 such that for all $n \geq N_1$: $|a_n - L| < \frac{\epsilon}{2}$.

Since $b_n \to M$, there exists N_2 such that for all $n \geq N_2$: $|b_n - M| < \frac{\epsilon}{2}$.

Let $N = \max(N_1, N_2)$. For $n \geq N$:

$$\begin{aligned} |(a_n + b_n) - (L + M)| &= |(a_n - L) + (b_n - M)| \\ &\leq |a_n - L| + |b_n - M| \quad \text{(triangle inequality)} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Therefore $a_n + b_n \to L + M$. ■

The other rules follow similarly (product rule requires using boundedness theorem). ■

Example 9.4 (Using Algebra of Limits)**.** *Find* $\lim_{n\to\infty} \frac{3n^2+5n-7}{2n^2+n+1}$.

Solution*: Divide numerator and denominator by* n^2*:*

$$\frac{3n^2 + 5n - 7}{2n^2 + n + 1} = \frac{3 + \frac{5}{n} - \frac{7}{n^2}}{2 + \frac{1}{n} + \frac{1}{n^2}}$$

Since $\frac{1}{n} \to 0$ *and* $\frac{1}{n^2} \to 0$*:*

$$\begin{aligned} \textit{Numerator} &\to 3 + 0 - 0 = 3 \\ \textit{Denominator} &\to 2 + 0 + 0 = 2 \end{aligned}$$

By quotient rule:

$$\lim_{n\to\infty} \frac{3n^2 + 5n - 7}{2n^2 + n + 1} = \frac{3}{2}$$

9.5 Monotone Sequences and Boundedness

Definition 9.3 (Monotone Sequences)**.** *A sequence* (a_n) *is:*

- ***Increasing*** *if* $a_n \leq a_{n+1}$ *for all* n
- ***Decreasing*** *if* $a_n \geq a_{n+1}$ *for all* n
- ***Strictly increasing*** *if* $a_n < a_{n+1}$ *for all* n
- ***Strictly decreasing*** *if* $a_n > a_{n+1}$ *for all* n
- ***Monotone*** *if it is either increasing or decreasing*

Theorem 9.4 (Monotone Convergence Theorem)**.** ***(a)*** *Every bounded increasing sequence converges.*

(b) *Every bounded decreasing sequence converges.*

Proof of (a). Let (a_n) be increasing and bounded above.

Let $A = \{a_n : n \in \mathbb{N}\} \subseteq \mathbb{R}$.

Since A is bounded above and non-empty, by completeness of $\mathbb{R}$, $\sup(A)$ exists.

Let $L = \sup(A)$.

Claim: $a_n \to L$.

Let $\epsilon > 0$ be given.

Since $L - \epsilon < L = \sup(A)$, by definition of supremum, $L - \epsilon$ is not an upper bound of A.

Therefore there exists N such that $a_N > L - \epsilon$.

Since (a_n) is increasing, for all $n \geq N$: $a_N \leq a_n \leq L$.

Therefore: $L - \epsilon < a_N \leq a_n \leq L < L + \epsilon$.

So $|a_n - L| < \epsilon$ for all $n \geq N$.

Therefore $a_n \to L = \sup(A)$. ■

The proof of (b) is similar, using $\inf(A)$. ■

Key Idea

This theorem is powerful:
To prove convergence of an increasing sequence, we only need to show it's bounded above—we don't need to find the limit explicitly!
The completeness of $\mathbb{R}$ guarantees the limit exists (it's the supremum).

Example 9.5 (Decimal Expansions)**.** *Consider the sequence of decimal approximations to $\sqrt{2}$:*

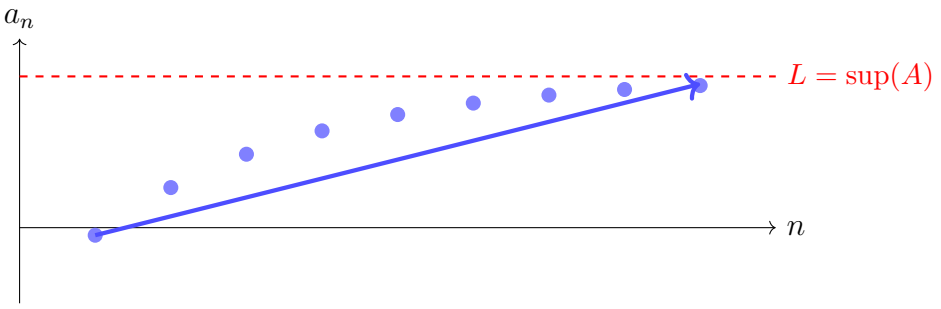
$$a_1 = 1, \quad a_2 = 1.4, \quad a_3 = 1.41, \quad a_4 = 1.414, \quad a_5 = 1.4142, \ldots$$

This sequence is:

- *Increasing: Each term adds more precision*
- *Bounded above: All terms are < 2 (since $(\sqrt{2})^2 = 2 < 4 = 2^2$)*

By the Monotone Convergence Theorem, (a_n) converges.
The limit is $\sqrt{2}$ (the completeness of $\mathbb{R}$ ensures this limit exists).

Monotone Convergence



Increasing + Bounded $\implies$
Convergent (to the supremum)

9.6 Cauchy Sequences

Intuition

To prove (a_n) converges using the ϵ-N definition, we need to know the limit L in advance.
But sometimes we want to know if a sequence converges *without* finding the limit.
Cauchy's insight: A sequence converges if and only if its terms get arbitrarily close *to each other* (not necessarily to a known limit).

Definition 9.4 (Cauchy Sequence)**.** *A sequence (a_n) is **Cauchy** if:*

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall m, n \geq N : |a_m - a_n| < \epsilon$$

Informally: Terms become arbitrarily close to each other as we go far enough in the sequence.

Key Idea

Convergent vs. Cauchy:
Convergent: Terms approach a specific limit L
Cauchy: Terms approach *each other* (we don't specify what they're approaching)
In $\mathbb{R}$, these are equivalent (due to completeness). In $\mathbb{Q}$, they differ!

Theorem 9.5 (Cauchy Criterion)**.** *A sequence in $\mathbb{R}$ converges if and only if it is Cauchy.*

Proof. **($\Rightarrow$) Convergent implies Cauchy**:

Suppose $a_n \to L$. Let $\epsilon > 0$ be given.

Since $a_n \to L$, there exists N such that for all $n \geq N$: $|a_n - L| < \frac{\epsilon}{2}$.

For $m, n \geq N$:

$$\begin{aligned} |a_m - a_n| &= |a_m - L + L - a_n| \\ &\leq |a_m - L| + |L - a_n| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Therefore (a_n) is Cauchy. ✓

($\Leftarrow$) Cauchy implies convergent:

This direction uses completeness of $\mathbb{R}$.

Suppose (a_n) is Cauchy. We show (a_n) is bounded, then construct its limit.

Step 1: Boundedness. Choose $\epsilon = 1$. There exists N such that for all $m, n \geq N$: $|a_m - a_n| < 1$.

Fix $n = N$. Then for all $m \geq N$: $|a_m - a_N| < 1$, so $|a_m| \leq |a_N| + 1$.

Let $M = \max(|a_1|, |a_2|, \ldots, |a_{N-1}|, |a_N| + 1)$. Then $|a_n| \leq M$ for all n. ✓

Step 2: Construct limit. For each $k \in \mathbb{N}$, let $A_k = \{a_n : n \geq k\}$ (the "tail" of the sequence).

Each A_k is bounded and non-empty, so $L_k = \sup(A_k)$ exists by completeness.

The sequence (L_k) is decreasing: $L_1 \geq L_2 \geq L_3 \geq \ldots$ (since $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$).

Also (L_k) is bounded below (by any lower bound of (a_n)).

By Monotone Convergence Theorem, $L_k \to L$ for some L.

Step 3: Show $a_n \to L$. Let $\epsilon > 0$ be given.

Since (a_n) is Cauchy, there exists N_1 such that for all $m, n \geq N_1$: $|a_m - a_n| < \frac{\epsilon}{2}$.

Since $L_k \to L$, there exists N_2 such that for all $k \geq N_2$: $|L_k - L| < \frac{\epsilon}{2}$.

Let $N = \max(N_1, N_2)$. For $n \geq N$:

Since $L_N = \sup(A_N)$ and $a_n \in A_N$ (because $n \geq N$), we have $a_n \leq L_N$.

Also, by Cauchy property and definition of supremum, $L_N - a_n < \frac{\epsilon}{2}$ (details omitted).

Therefore:

$$|a_n - L| \leq |a_n - L_N| + |L_N - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Therefore $a_n \to L$. ■ ■

Remark

The direction Cauchy ⇒ convergent critically uses completeness of $\mathbb{R}$.
In $\mathbb{Q}$, there exist Cauchy sequences that don't converge (to a rational).
Example: The sequence $1.4, 1.41, 1.414, 1.4142, \ldots$ (rational approximations to $\sqrt{2}$) is Cauchy in $\mathbb{Q}$ but does not converge to any rational number.
This is why we needed to construct $\mathbb{R}$—to complete $\mathbb{Q}$ by adding these missing limits!

9.7 Subsequences and Bolzano-Weierstrass

Definition 9.5 (Subsequence)**.** *Given a sequence (a_n), a **subsequence** is a sequence (a_{n_k}) where $n_1 < n_2 < n_3 < \ldots$ is a strictly increasing sequence of indices.*

Informally: Select infinitely many terms from (a_n) in order.

Example 9.6. *From $(1, 2, 3, 4, 5, 6, \ldots)$:*

- *Even terms: $(2, 4, 6, 8, \ldots)$ is the subsequence (a_{2k})*
- *Odd terms: $(1, 3, 5, 7, \ldots)$ is the subsequence (a_{2k-1})*
- *Powers of 2: $(2, 4, 8, 16, \ldots)$ is the subsequence (a_{2^k})*

Theorem 9.6 (Subsequences of Convergent Sequences)**.** *If $a_n \to L$, then every subsequence (a_{n_k}) also converges to L.*

Proof. Let $\epsilon > 0$ be given.

Since $a_n \to L$, there exists N such that for all $n \geq N$: $|a_n - L| < \epsilon$.

Since n_k is strictly increasing and $n_k \geq k$ for all k, we have: for $k \geq N$, $n_k \geq k \geq N$.

Therefore $|a_{n_k} - L| < \epsilon$ for all $k \geq N$.

Thus $a_{n_k} \to L$. ■ ■

Theorem 9.7 (Bolzano-Weierstrass Theorem)**.** *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof Sketch. Let (a_n) be bounded: $|a_n| \leq M$ for all n.

The sequence lives in the interval $[-M, M]$.

Idea: Repeatedly bisect intervals to find a convergent subsequence.

Step 1: Divide $[-M, M]$ into $[-M, 0]$ and $[0, M]$. At least one half contains infinitely many terms of (a_n). Choose such a half and call it I_1.

Step 2: Divide I_1 in half. Again, one half contains infinitely many terms. Choose such a half and call it I_2.

Continue: Obtain nested intervals $I_1 \supseteq I_2 \supseteq I_3 \supseteq \dots$ with $\text{length}(I_k) = \frac{2M}{2^k} \to 0$.

Choose $a_{n_1} \in I_1$, $a_{n_2} \in I_2$ with $n_2 > n_1$, $a_{n_3} \in I_3$ with $n_3 > n_2$, etc.

The subsequence (a_{n_k}) is Cauchy (since terms are in intervals of shrinking length).

By Cauchy criterion, (a_{n_k}) converges. ■ ■

Key Idea

Bolzano-Weierstrass is **powerful**:
Even if a bounded sequence doesn't converge (like $(-1)^n$), we can always extract a convergent subsequence.
This theorem is crucial in proving existence results in analysis.

9.8 Series: Infinite Sums

Intuition

A **series** is an "infinite sum":

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots$$

But what does this mean rigorously? We can't literally "add infinitely many things."
Solution: Define convergence via partial sums.

Definition 9.6 (Series and Partial Sums)**.** *Given a sequence (a_n), the **n-th partial sum** is:*

$$S_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n$$

*The **series** $\sum_{n=1}^{\infty} a_n$ **converges** if the sequence of partial sums (S_n) converges.*
If $S_n \to S$, we write:

$$\sum_{n=1}^{\infty} a_n = S$$

*and call S the **sum of the series**.*

Example 9.7 (Geometric Series)**.** *Consider* $\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots$ *for* $|r| < 1$.

The partial sum is:

$$S_n = \sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r} \quad \textit{(geometric sum formula)}$$

As $n \to \infty$*:*

$$S_n = \frac{1 - r^{n+1}}{1 - r} \to \frac{1}{1 - r} \quad \textit{(since } r^{n+1} \to 0 \textit{ when } |r| < 1\textit{)}$$

Therefore:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1 - r} \quad \textit{for } |r| < 1$$

If $|r| \geq 1$*, the series diverges.*

Example 9.8 (Harmonic Series)**.** *The harmonic series is:*

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

Claim*: This series diverges (even though* $\frac{1}{n} \to 0$*).*

Proof*: Group terms:*

$$\begin{aligned} S_n &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots \to \infty \end{aligned}$$

Therefore $S_n \to \infty$*, and the series diverges.* ■

Theorem 9.8 (Necessary Condition for Convergence)**.** *If* $\sum_{n=1}^{\infty} a_n$ *converges, then* $a_n \to 0$.

Proof. Suppose $S_n = \sum_{k=1}^{n} a_k \to S$.

Then:

$$a_n = S_n - S_{n-1} \to S - S = 0$$

Therefore $a_n \to 0$. ■ ■

Warning

The converse is **false**: $a_n \to 0$ does **not** imply $\sum a_n$ converges.
Counterexample: Harmonic series $\sum \frac{1}{n}$ has $\frac{1}{n} \to 0$ but diverges.

9.9 Looking Forward: Continuity and Calculus

Intuition

With sequences and limits mastered, we can now rigorously define:
Continuity: f is continuous at x if $f(x_n) \to f(x)$ whenever $x_n \to x$
Derivatives: $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$
Integrals: $\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$
All of calculus reduces to limits of sequences. We've built the foundation.

From Sequences to All of Analysis

Sequences $\to$ Limits $\to$ Continuity $\to$ Derivatives $\to$ Integrals

Every concept in calculus is built on the ϵ-N definition of convergence. Completeness of $\mathbb{R}$ ensures all these limiting processes work.

"In analysis, everything is a limit." — Anonymous

Chapter 10

Continuity: Functions that Preserve Closeness

10.1 From Sequences to Functions

Intuition

We've studied sequences: discrete points approaching a limit.
Now we study **continuous functions**: functions where "nearby inputs produce nearby outputs."
Informal idea: A function f is continuous if you can draw its graph without lifting your pen.
Rigorous idea: f is continuous at x if $f(x_n) \to f(x)$ whenever $x_n \to x$.
This chapter makes continuity precise and proves its fundamental properties.

Historical Context

The Evolution of Continuity
Ancient Mathematics (300 BCE - 1600 CE):

- Greeks used continuous curves geometrically (circles, conics)
- No formal definition—continuity was intuitive
- Archimedes: Method of exhaustion assumed continuity implicitly

Early Calculus (1650-1800):

- Newton, Leibniz: Used continuous functions freely
- Euler: "A continuous function is one whose equation is given by a single analytic expression"
- **Problem**: What about piecewise functions? No rigorous definition

19th Century Rigor:

- **Bolzano (1817)**: First rigorous definition using sequences

- **Cauchy (1821)**: "f is continuous if infinitely small changes in x produce infinitely small changes in $f(x)$" (still vague)
- **Weierstrass (1860s)**: The modern ϵ-δ definition
- **Key insight**: Replace vague "infinitely small" with precise quantifiers

Impact:

- Allowed rigorous proofs of Intermediate Value Theorem, Extreme Value Theorem
- Revealed surprising phenomena: continuous but nowhere differentiable functions
- Foundation for topology (continuous functions between topological spaces)

The ϵ-δ definition is one of the great achievements of 19th-century mathematics.

10.2 Continuity at a Point

Definition 10.1 (Continuity at a Point (Sequential))**.** *Let $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$, and let $c \in D$.*

*The function f is **continuous at** c if:*

For every sequence (x_n) in D with $x_n \to c$, we have $f(x_n) \to f(c)$.

***In symbols**:*

$$\forall (x_n) \subseteq D : x_n \to c \implies f(x_n) \to f(c)$$

Key Idea

Sequential continuity: If inputs approach c, outputs approach $f(c)$.
This means: The limit of f at c equals the value of f at c.
Equivalently: You can "pass the limit through the function":

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right)$$

Example 10.1 (Continuous Function)**.** *Let $f(x) = x^2$. Prove f is continuous at $c = 2$.*

***Proof**: Let (x_n) be any sequence with $x_n \to 2$.*

We need to show $f(x_n) = x_n^2 \to 4 = f(2)$.

By algebra of limits:

$$\lim_{n\to\infty} x_n^2 = \left(\lim_{n\to\infty} x_n\right)^2 = 2^2 = 4$$

Therefore $f(x_n) \to f(2)$, so f is continuous at 2. ✓

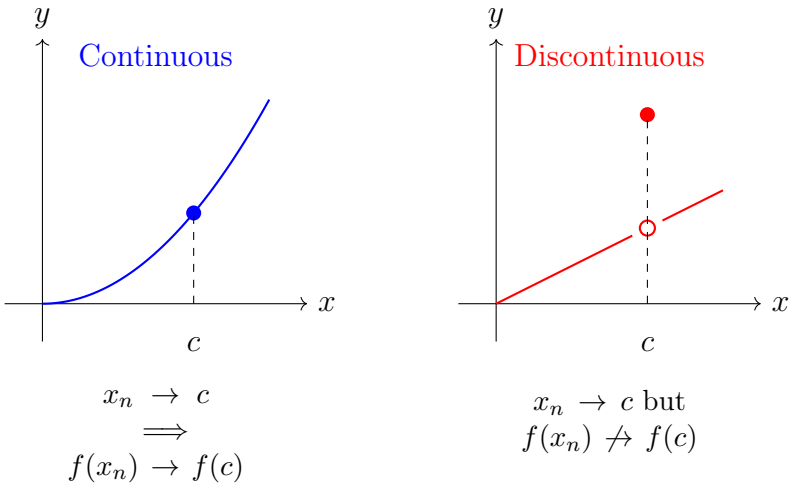
(This argument works for any c, so $f(x) = x^2$ is continuous everywhere.)

Example 10.2 (Discontinuous Function)**.** *Define* $f : \mathbb{R} \to \mathbb{R}$ *by:*

$$f(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$$

***Claim**: f is not continuous at $c = 0$.*
***Proof**: Consider the sequence $x_n = \frac{1}{n} \to 0$.*
Then $f(x_n) = 0$ for all n (since $x_n \neq 0$), so $f(x_n) \to 0$.
But $f(0) = 1 \neq 0$.
Therefore $f(x_n) \not\to f(0)$, so f is not continuous at 0. ■

Continuous vs. Discontinuous



10.3 The ϵ-δ Definition

Intuition

The sequential definition is intuitive, but sometimes we need a definition that doesn't mention sequences.
Weierstrass's ϵ-δ definition captures the same idea directly:
"For any desired closeness ϵ of outputs, there exists a required closeness δ of inputs."

Definition 10.2 (Continuity at a Point (ϵ-δ))**.** *Let $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$, and let $c \in D$.*

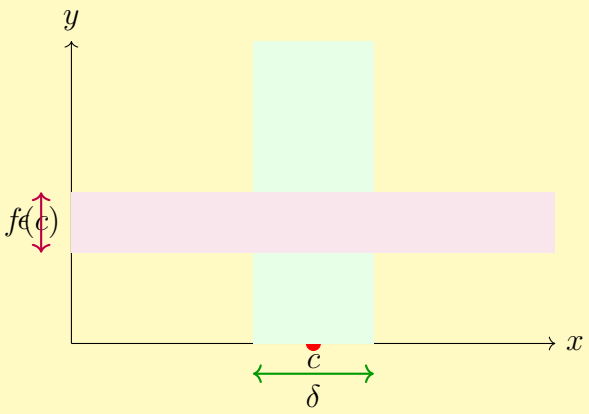
*The function f is **continuous at** c if:*

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall x \in D : |x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

***In words**: For any ϵ-neighborhood around $f(c)$, we can find a δ-neighborhood around c whose image lies entirely within the ϵ-neighborhood.*

Key Idea

The ϵ-δ game:
Challenger: Gives you $\epsilon > 0$ (tolerance for output)
You: Must find $\delta > 0$ such that whenever $|x - c| < \delta$, we have $|f(x) - f(c)| < \epsilon$
If you can always win, f is continuous at c.
Geometric interpretation:



The green region (width 2δ) maps into the purple region (height 2ϵ).

Theorem 10.1 (Equivalence of Definitions)**.** *The sequential definition and ϵ-δ definition of continuity are equivalent.*

Proof. **(ϵ-$\delta \Rightarrow$ Sequential)**:

Assume f satisfies the ϵ-δ condition at c. Let (x_n) be a sequence in D with $x_n \to c$.

We show $f(x_n) \to f(c)$.

Let $\epsilon > 0$ be given. By ϵ-δ continuity, there exists $\delta > 0$ such that:

$$|x - c| < \delta \implies |f(x) - f(c)| < \epsilon$$

Since $x_n \to c$, there exists N such that for all $n \geq N$: $|x_n - c| < \delta$.

Therefore for $n \geq N$: $|f(x_n) - f(c)| < \epsilon$.

Thus $f(x_n) \to f(c)$. ✓

(Sequential $\Rightarrow$ ϵ-δ):

Assume f satisfies the sequential condition at c. We prove ϵ-δ by contradiction.

Suppose f does not satisfy ϵ-δ. Then there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists x with:

$$|x - c| < \delta \quad \text{but} \quad |f(x) - f(c)| \geq \epsilon$$

For each $n \in \mathbb{N}$, choose $\delta = \frac{1}{n}$. Then there exists x_n such that:

$$|x_n - c| < \frac{1}{n} \quad \text{but} \quad |f(x_n) - f(c)| \geq \epsilon$$

The sequence (x_n) satisfies $x_n \to c$ (since $|x_n - c| < \frac{1}{n} \to 0$).

By sequential continuity, $f(x_n) \to f(c)$.

But $|f(x_n) - f(c)| \geq \epsilon$ for all n, contradicting $f(x_n) \to f(c)$.

Therefore the ϵ-δ condition must hold. ■ ■

Example 10.3 (Using ϵ-δ to Prove Continuity)**.** *Prove $f(x) = 3x + 2$ is continuous at $c = 1$ using ϵ-δ.*

***Proof**: Let $\epsilon > 0$ be given. We need to find $\delta > 0$ such that:*

$$|x - 1| < \delta \implies |f(x) - f(1)| < \epsilon$$

Note that $f(1) = 3(1) + 2 = 5$.

$$\begin{aligned} |f(x) - f(1)| &= |(3x + 2) - 5| \\ &= |3x - 3| \\ &= 3|x - 1| \end{aligned}$$

We want $3|x - 1| < \epsilon$, so $|x - 1| < \frac{\epsilon}{3}$.

***Choose** $\delta = \frac{\epsilon}{3}$.*

Then for $|x - 1| < \delta$:

$$|f(x) - f(1)| = 3|x - 1| < 3\delta = 3 \cdot \frac{\epsilon}{3} = \epsilon$$

Therefore f is continuous at $c = 1$. ■

(This argument works at any c, so $f(x) = 3x + 2$ is continuous everywhere.)

10.4 Continuous Functions

Definition 10.3 (Continuous on a Set)**.** *A function $f : D \to \mathbb{R}$ is **continuous on** D (or simply **continuous**) if f is continuous at every point $c \in D$.*

Theorem 10.2 (Algebra of Continuous Functions)**.** *If f and g are continuous at c, then:*

1. *$f + g$ is continuous at c*
2. *$f - g$ is continuous at c*
3. *$f \cdot g$ is continuous at c*
4. *$\frac{f}{g}$ is continuous at c (if $g(c) \neq 0$)*
5. *cf is continuous at c for any constant $c \in \mathbb{R}$*

Proof of Sum Rule. Let (x_n) be a sequence with $x_n \to c$.

Since f is continuous at c: $f(x_n) \to f(c)$.

Since g is continuous at c: $g(x_n) \to g(c)$.

By algebra of limits:

$$(f+g)(x_n) = f(x_n) + g(x_n) \to f(c) + g(c) = (f+g)(c)$$

Therefore $f+g$ is continuous at c. ■

The other rules follow similarly. ■

Example 10.4 (Polynomial Functions)**.** *Every polynomial $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is continuous on $\mathbb{R}$.*

***Proof**:*

- *Constant functions are continuous (trivial)*
- *$f(x) = x$ is continuous (easy ϵ-δ proof with $\delta = \epsilon$)*
- *x^k is continuous (by product rule, since $x^k = x \cdot x \cdots x$)*
- *$a_k x^k$ is continuous (by scalar multiplication)*
- *$p(x)$ is continuous (by sum rule)*

■

Example 10.5 (Rational Functions)**.** *Every rational function $r(x) = \frac{p(x)}{q(x)}$ is continuous on its domain (where $q(x) \neq 0$).*

***Proof**: By quotient rule, since polynomials are continuous.* ■

Theorem 10.3 (Composition of Continuous Functions)**.** *If $f : D \to \mathbb{R}$ is continuous at c and $g : E \to \mathbb{R}$ is continuous at $f(c)$ (with $f(D) \subseteq E$), then $g \circ f$ is continuous at c.*

Proof. Let (x_n) be a sequence in D with $x_n \to c$.

Since f is continuous at c: $f(x_n) \to f(c)$.

Let $y_n = f(x_n)$. Then (y_n) is a sequence in E with $y_n \to f(c)$.

Since g is continuous at $f(c)$: $g(y_n) \to g(f(c))$.

Therefore:

$$(g \circ f)(x_n) = g(f(x_n)) = g(y_n) \to g(f(c)) = (g \circ f)(c)$$

Thus $g \circ f$ is continuous at c. ■ ■

Example 10.6. *$h(x) = \sin(x^2 + 3x)$ is continuous on $\mathbb{R}$.*

***Proof**:*

- *$f(x) = x^2 + 3x$ is continuous (polynomial)*
- *$g(y) = \sin(y)$ is continuous (proven using trigonometric identities and ϵ-δ)*
- *$h = g \circ f$ is continuous (composition rule)*

■

10.5 The Intermediate Value Theorem

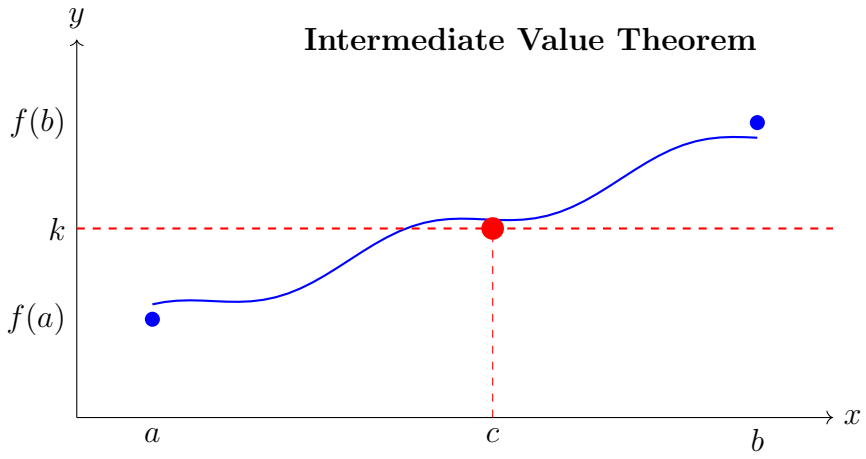
Intuition

If you drive from elevation 100m to elevation 200m, you must pass through every elevation in between.
More generally: A continuous function on an interval takes all values between any two of its values.
This seemingly obvious statement requires completeness of $\mathbb{R}$ to prove!

Theorem 10.4 (Intermediate Value Theorem (IVT)). *Let $f : [a, b] \to \mathbb{R}$ be continuous on the closed interval $[a, b]$.*

If $f(a) < k < f(b)$ (or $f(b) < k < f(a)$), then there exists $c \in (a, b)$ such that $f(c) = k$.

***In words**: A continuous function on a closed interval attains every value between its endpoints.*



If $f(a) < k < f(b)$, then $\exists c \in (a, b)$ with $f(c) = k$

Proof. Assume $f(a) < k < f(b)$ (the other case is similar).

Define:

$$S = \{x \in [a, b] : f(x) < k\}$$

Properties of S:

- $S \neq \emptyset$ (since $a \in S$, as $f(a) < k$)
- S is bounded above (by b)

By completeness, $c = \sup(S)$ exists.

We show $f(c) = k$.

Claim 1: $f(c) \leq k$.

Proof: Suppose $f(c) > k$. Let $\epsilon = f(c) - k > 0$.

By continuity, there exists $\delta > 0$ such that $|x - c| < \delta \implies |f(x) - f(c)| < \epsilon$.

For $x \in (c - \delta, c + \delta)$:

$$f(x) > f(c) - \epsilon = f(c) - (f(c) - k) = k$$

This means $f(x) > k$ for all x near c, so no element of S is close to c, contradicting $c = \sup(S)$. ✓

Claim 2: $f(c) \geq k$.

Proof: Suppose $f(c) < k$. Let $\epsilon = k - f(c) > 0$.

By continuity, there exists $\delta > 0$ such that $|x - c| < \delta \implies |f(x) - f(c)| < \epsilon$.

For $x \in (c, c + \delta)$:

$$f(x) < f(c) + \epsilon = f(c) + (k - f(c)) = k$$

This means $c + \frac{\delta}{2} \in S$, contradicting $c = \sup(S)$. ✓

Therefore $f(c) = k$. ■ ■

Key Idea

The IVT uses completeness crucially:

1. We define $S = \{x : f(x) < k\}$
2. We take $c = \sup(S)$ (this requires completeness!)
3. We show $f(c) = k$ using continuity

Without completeness (e.g., in $\mathbb{Q}$), the theorem fails.
Counterexample in $\mathbb{Q}$: $f(x) = x^2$ on $[1, 2]$ is continuous, and $f(1) = 1 < 2 < 4 = f(2)$, but there is no $c \in \mathbb{Q}$ with $f(c) = 2$ (since $\sqrt{2} \notin \mathbb{Q}$).

Example 10.7 (Root Finding). *Show that $x^3 - 3x + 1 = 0$ has a solution in $[0, 1]$.*

***Proof**: Let $f(x) = x^3 - 3x + 1$.*

f is continuous (polynomial).

$f(0) = 1 > 0$ and $f(1) = 1 - 3 + 1 = -1 < 0$.

By IVT, there exists $c \in (0, 1)$ such that $f(c) = 0$. ■

Example 10.8 (Fixed Point). *Every continuous function $f : [0, 1] \to [0, 1]$ has a fixed point (a point c where $f(c) = c$).*

***Proof**: Let $g(x) = f(x) - x$.*

g is continuous (difference of continuous functions).

$g(0) = f(0) - 0 = f(0) \geq 0$ (since $f(0) \in [0, 1]$).

$g(1) = f(1) - 1 \leq 0$ (since $f(1) \in [0, 1]$).

***Case 1**: If $g(0) = 0$, then $f(0) = 0$, so $c = 0$ is a fixed point.*

***Case 2**: If $g(1) = 0$, then $f(1) = 1$, so $c = 1$ is a fixed point.*

***Case 3**: If $g(0) > 0$ and $g(1) < 0$, then by IVT, there exists $c \in (0, 1)$ with $g(c) = 0$, so $f(c) = c$.* ■

10.6 The Extreme Value Theorem

Theorem 10.5 (Extreme Value Theorem (EVT)). *If $f : [a, b] \to \mathbb{R}$ is continuous on the closed bounded interval $[a, b]$, then f attains its maximum and minimum.*

That is, there exist $c, d \in [a, b]$ such that:

$$f(c) \leq f(x) \leq f(d) \quad \textit{for all } x \in [a, b]$$

Proof of Maximum (Minimum is Similar). **Step 1: f is bounded above.**

Suppose not. Then for each $n \in \mathbb{N}$, there exists $x_n \in [a, b]$ with $f(x_n) > n$.

The sequence (x_n) is bounded (in $[a, b]$), so by Bolzano-Weierstrass, there exists a convergent subsequence $x_{n_k} \to c$ for some $c \in [a, b]$.

By continuity, $f(x_{n_k}) \to f(c)$.

But $f(x_{n_k}) > n_k \to \infty$, contradiction.

Therefore f is bounded above. ✓

Step 2: f attains its supremum.

Let $M = \sup\{f(x) : x \in [a, b]\}$ (exists by completeness and Step 1).

For each $n \in \mathbb{N}$, by definition of supremum, there exists $x_n \in [a, b]$ with:

$$f(x_n) > M - \frac{1}{n}$$

By Bolzano-Weierstrass, there exists a subsequence $x_{n_k} \to d$ for some $d \in [a, b]$.

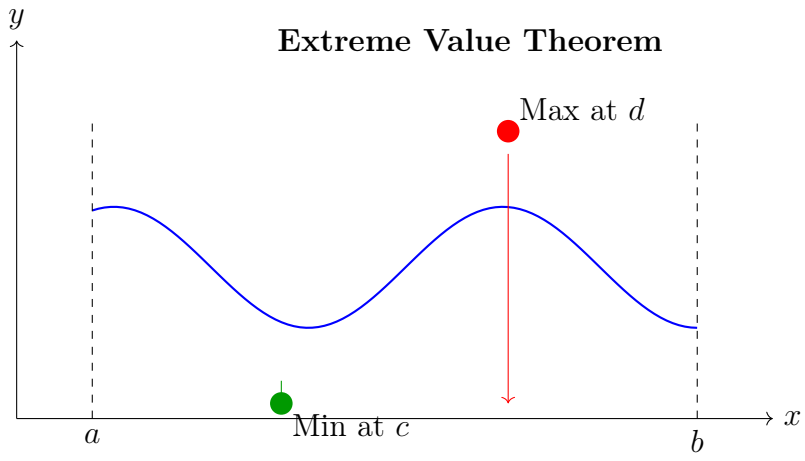
By continuity, $f(x_{n_k}) \to f(d)$.

Since $M - \frac{1}{n_k} < f(x_{n_k}) \leq M$ and $\frac{1}{n_k} \to 0$, we have:

$$f(x_{n_k}) \to M$$

By uniqueness of limits, $f(d) = M$.

Therefore f attains its maximum at d. ■ ■



Continuous on $[a, b]$ $\implies$ Max and Min exist

Warning

The EVT requires:

1. **Continuity**: Without continuity, max/min may not exist

2. **Closed interval**: On open intervals like (a, b), max/min may not exist
3. **Bounded interval**: On unbounded intervals like $[a, \infty)$, max/min may not exist

Counterexamples:

- $f(x) = x$ on $(0, 1)$: No max or min (open interval)
- $f(x) = \frac{1}{x}$ on $(0, 1]$: No max (not continuous at 0)
- $f(x) = x$ on $[0, \infty)$: No max (unbounded interval)

Key Idea

Why EVT matters:
Many optimization problems reduce to: "Find the maximum/minimum of a continuous function on a closed bounded interval."
EVT guarantees the solution exists—we just need to find it!
Typical strategy:

1. Check continuity
2. Check domain is closed and bounded
3. Find critical points (where derivative is zero or undefined)
4. Check endpoints
5. Maximum is the largest value among these

10.7 Uniform Continuity

Intuition

A function is continuous if at each point c, we can find a suitable δ for any ϵ.
But δ might depend on both ϵ *and* the point c.
Uniform continuity: The same δ works for *all* points simultaneously.

Definition 10.4 (Uniform Continuity)**.** *A function $f : D \to \mathbb{R}$ is **uniformly continuous on** D if:*

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall x, y \in D : |x - y| < \delta \implies |f(x) - f(y)| < \epsilon$$

***Key difference**: δ depends only on ϵ, not on the specific points x or y.*

Example 10.9 (Uniformly Continuous Function)**.** *$f(x) = 3x + 2$ on $\mathbb{R}$ is uniformly continuous.*

***Proof**: Let $\epsilon > 0$ be given.*

For any $x, y \in \mathbb{R}$:

$$|f(x) - f(y)| = |3x + 2 - 3y - 2| = 3|x - y|$$

We want $3|x - y| < \epsilon$, so $|x - y| < \frac{\epsilon}{3}$.
***Choose** $\delta = \frac{\epsilon}{3}$ (independent of x, y!).*
Then $|x - y| < \delta \implies |f(x) - f(y)| = 3|x - y| < 3\delta = \epsilon$.
Therefore f is uniformly continuous. ■

Example 10.10 (Continuous but Not Uniformly Continuous)**.** *$f(x) = x^2$ on $\mathbb{R}$ is continuous everywhere but not uniformly continuous.*

***Proof**: Suppose f were uniformly continuous. Then for $\epsilon = 1$, there exists $\delta > 0$ such that:*

$$|x - y| < \delta \implies |x^2 - y^2| < 1$$

Choose $x = n$ and $y = n + \frac{\delta}{2}$ for large n.
Then $|x - y| = \frac{\delta}{2} < \delta$, but:

$$|x^2 - y^2| = |n^2 - (n + \frac{\delta}{2})^2| = |n\delta - \frac{\delta^2}{4}| \approx n\delta \to \infty$$

For sufficiently large n, this exceeds 1, contradiction.
Therefore f is not uniformly continuous on $\mathbb{R}$. ■

Theorem 10.6 (Continuity on Compact Intervals)**.** *If $f : [a, b] \to \mathbb{R}$ is continuous on the closed bounded interval $[a, b]$, then f is uniformly continuous on $[a, b]$.*

Proof Sketch. Suppose f is not uniformly continuous. Then there exists $\epsilon > 0$ such that for all $\delta > 0$, there exist x, y with:

$$|x - y| < \delta \quad \text{but} \quad |f(x) - f(y)| \geq \epsilon$$

For each n, choose $\delta = \frac{1}{n}$, obtaining sequences $(x_n), (y_n)$ with:

$$|x_n - y_n| < \frac{1}{n} \quad \text{but} \quad |f(x_n) - f(y_n)| \geq \epsilon$$

By Bolzano-Weierstrass, (x_n) has a convergent subsequence $x_{n_k} \to c \in [a, b]$.
Since $|x_{n_k} - y_{n_k}| < \frac{1}{n_k} \to 0$, we also have $y_{n_k} \to c$.
By continuity, $f(x_{n_k}) \to f(c)$ and $f(y_{n_k}) \to f(c)$.
Therefore $|f(x_{n_k}) - f(y_{n_k})| \to 0$, contradicting $|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon$. ■ ■

10.8 Looking Forward: Differentiation

Intuition

Continuity means: Small changes in input produce small changes in output.
Differentiability strengthens this: Changes in output are *linearly proportional* to changes in input (locally).
The derivative $f'(x)$ is the "instantaneous rate of change"—the slope of the tangent line.
Next chapter: We'll define derivatives rigorously using limits and prove the fundamental theorems of calculus.

The Hierarchy of Smoothness

Discontinuous $\subset$ Continuous $\subset$ Differentiable $\subset$ Smooth (C^∞)

Continuity is the foundation. Differentiability adds rate-of-change.
We've now built enough machinery to define derivatives rigorously.

"Continuity protects existence; differentiability protects uniqueness."

Chapter 11

Differentiation: Instantaneous Rate of Change

11.1 From Continuity to Differentiability

Intuition

Continuity asks: "Does the function have jumps?"
Differentiability asks: "Does the function have sharp corners?"
The derivative $f'(x)$ measures the **instantaneous rate of change** of f at x:

- Geometrically: Slope of the tangent line
- Physically: Velocity (if f is position)
- Economically: Marginal cost/revenue

This chapter develops differentiation rigorously from limits.

Historical Context

The Birth of Calculus
Ancient Precursors (c. 250 BCE - 1600 CE):

- **Archimedes**: Computed tangent lines to spirals (geometric methods)
- **Fermat (1629)**: Method of "adequality" (proto-derivatives)
- **Barrow (1670)**: Geometric tangent method (Newton's teacher)

The Revolution (1665-1675):

- **Newton (1665)**: "Fluxions"—rates of change of "fluents"
- **Leibniz (1675)**: Notation $\frac{dy}{dx}$, infinitesimals dx, dy
- Both discovered: Differentiation and integration are inverse operations
- **Priority dispute**: One of history's most bitter mathematical feuds

The Rigor Gap (1700-1850):

- Euler, Lagrange, Laplace: Manipulated derivatives powerfully but non-rigorously
- **Berkeley's attack (1734)**: "What are these infinitesimals? Ghosts of departed quantities!"
- No answer: Infinitesimals weren't properly defined

19th Century Foundations:

- **Cauchy (1821)**: First limit-based definition of derivative
- **Weierstrass (1860s)**: Rigorous ϵ-δ formulation
- **Dedekind (1872)**: Completed $\mathbb{R}$, making all limits rigorous
- Result: Calculus became a branch of analysis with complete proofs

Modern View: Derivatives are limits. Infinitesimals can be made rigorous (non-standard analysis, 1960s) but are not needed for standard calculus.

11.2 The Derivative: Definition and Interpretation

Definition 11.1 (Derivative at a Point)**.** *Let $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$, and let $c \in D$ be an interior point (not an endpoint).*

*The **derivative of** f **at** c is:*

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

provided this limit exists.

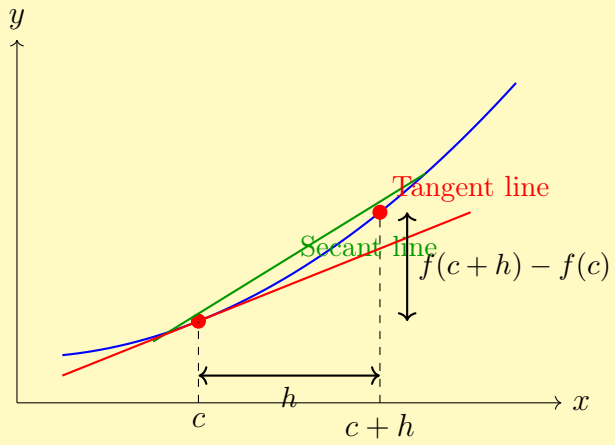
*If the limit exists, we say f is **differentiable at** c.*

***Alternative form** (using x instead of $c + h$):*

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

Key Idea

The difference quotient $\frac{f(c+h)-f(c)}{h}$ is the **average rate of change** of f over the interval $[c, c + h]$.
As $h \to 0$, this becomes the **instantaneous rate of change**.
Geometric interpretation:



As $h \to 0$, secant $\to$ tangent

The derivative $f'(c)$ is the slope of the tangent line at $x = c$.

Example 11.1 (Computing a Derivative from Definition)**.** *Find the derivative of* $f(x) = x^2$ *at* $c = 3$.

***Solution**:*

$$\begin{aligned} f'(3) &= \lim_{h\to 0} \frac{f(3+h)-f(3)}{h} \\ &= \lim_{h\to 0} \frac{(3+h)^2-9}{h} \\ &= \lim_{h\to 0} \frac{9+6h+h^2-9}{h} \\ &= \lim_{h\to 0} \frac{6h+h^2}{h} \\ &= \lim_{h\to 0} \frac{h(6+h)}{h} \\ &= \lim_{h\to 0} (6+h) \\ &= 6 \end{aligned}$$

Therefore $f'(3) = 6$. ■

(More generally, for $f(x) = x^2$*, we get* $f'(c) = 2c$ *for any* c*.)*

Example 11.2 (Function Not Differentiable)**.** *Consider* $f(x) = |x|$ *at* $c = 0$.

***Right-hand derivative**:*

$$\lim_{h\to 0^+} \frac{|h|-0}{h} = \lim_{h\to 0^+} \frac{h}{h} = 1$$

***Left-hand derivative**:*

$$\lim_{h\to 0^-} \frac{|h|-0}{h} = \lim_{h\to 0^-} \frac{-h}{h} = -1$$

Since the left and right limits disagree, $\lim_{h\to 0} \frac{|h|}{h}$ *does not exist.*

Therefore $f(x) = |x|$ *is not differentiable at* $x = 0$ *(sharp corner).* ■

Definition 11.2 (Derivative Function)**.** *If* f *is differentiable at every point in its domain, the* ***derivative function*** *is:*

$$f' : D \to \mathbb{R}, \quad f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

Notations:

$$f'(x) = \frac{df}{dx} = \frac{d}{dx} f(x) = Df(x) = D_x f$$

11.3 Differentiability Implies Continuity

Theorem 11.1. *If* f *is differentiable at* c*, then* f *is continuous at* c*.*

Proof. Assume f is differentiable at c, so $f'(c) = \lim_{h\to 0} \frac{f(c+h)-f(c)}{h}$ exists.

We need to show $\lim_{h\to 0} f(c+h) = f(c)$, i.e., $\lim_{h\to 0}[f(c+h) - f(c)] = 0$.

Note that:

$$f(c+h) - f(c) = \frac{f(c+h) - f(c)}{h} \cdot h$$

Taking limits as $h \to 0$:

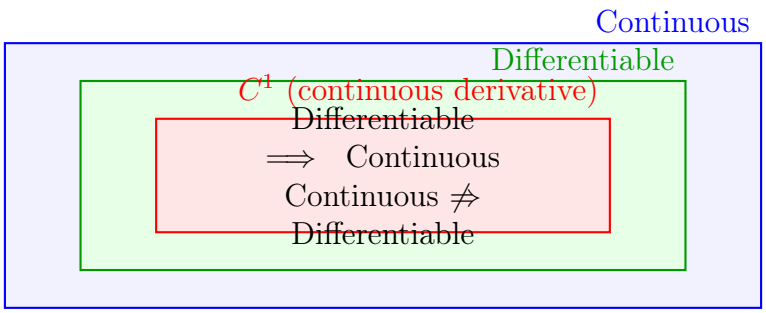
$$\begin{aligned}
\lim_{h\to 0}[f(c+h) - f(c)] &= \lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h} \cdot h\right] \\
&= \left[\lim_{h\to 0} \frac{f(c+h) - f(c)}{h}\right] \cdot \left[\lim_{h\to 0} h\right] \\
&= f'(c) \cdot 0 = 0
\end{aligned}$$

Therefore $f(c+h) \to f(c)$ as $h \to 0$, so f is continuous at c. ■ ■

Warning

The converse is **false**: Continuity does not imply differentiability.
Example: $f(x) = |x|$ is continuous at 0 but not differentiable at 0.
More extreme: Weierstrass (1872) constructed a function continuous *everywhere* but differentiable *nowhere*—a continuous but infinitely jagged curve!

Hierarchy of Function Properties



11.4 Differentiation Rules

Theorem 11.2 (Power Rule)**.** *For $f(x) = x^n$ where $n \in \mathbb{N}$:*

$$f'(x) = nx^{n-1}$$

Proof. We use the binomial theorem:

$$\begin{aligned} f'(c) &= \lim_{h\to 0} \frac{(c+h)^n - c^n}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k=0}^{n} \binom{n}{k} c^{n-k}h^k - c^n\right] \\ &= \lim_{h\to 0} \frac{1}{h}\left[c^n + nc^{n-1}h + \binom{n}{2}c^{n-2}h^2 + \cdots + h^n - c^n\right] \\ &= \lim_{h\to 0}\left[nc^{n-1} + \binom{n}{2}c^{n-2}h + \cdots + h^{n-1}\right] \\ &= nc^{n-1} \end{aligned}$$

Therefore $(x^n)' = nx^{n-1}$. ■ ■

Theorem 11.3 (Algebra of Derivatives)**.** *If f and g are differentiable at c, then:*

1. ***Constant multiple**: $(cf)' = cf'$ for any $c \in \mathbb{R}$*
2. ***Sum rule**: $(f+g)' = f' + g'$*
3. ***Product rule**: $(fg)' = f'g + fg'$*
4. ***Quotient rule**: $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ (if $g(c) \neq 0$)*

Proof of Product Rule.

$$(fg)'(c) = \lim_{h\to 0} \frac{f(c+h)g(c+h) - f(c)g(c)}{h}$$

Trick: Add and subtract $f(c+h)g(c)$:

$$\begin{aligned} &= \lim_{h\to 0} \frac{f(c+h)g(c+h) - f(c+h)g(c) + f(c+h)g(c) - f(c)g(c)}{h} \\ &= \lim_{h\to 0}\left[f(c+h)\cdot\frac{g(c+h)-g(c)}{h} + g(c)\cdot\frac{f(c+h)-f(c)}{h}\right] \\ &= \lim_{h\to 0} f(c+h)\cdot\lim_{h\to 0}\frac{g(c+h)-g(c)}{h} + g(c)\cdot\lim_{h\to 0}\frac{f(c+h)-f(c)}{h} \\ &= f(c)\cdot g'(c) + g(c)\cdot f'(c) \quad \text{(using continuity of } f\text{)} \end{aligned}$$

Therefore $(fg)' = f'g + fg'$. ■ ■

Proof of Quotient Rule. Let $h(x) = \frac{f(x)}{g(x)}$ where $g(c) \neq 0$.

$$\begin{aligned} h'(c) &= \lim_{x\to c} \frac{h(x) - h(c)}{x - c} \\ &= \lim_{x\to c} \frac{\frac{f(x)}{g(x)} - \frac{f(c)}{g(c)}}{x - c} \\ &= \lim_{x\to c} \frac{f(x)g(c) - f(c)g(x)}{(x - c)g(x)g(c)} \end{aligned}$$

Trick: Add and subtract $f(c)g(c)$ in the numerator:

$$\begin{aligned} &= \lim_{x\to c} \frac{f(x)g(c) - f(c)g(c) + f(c)g(c) - f(c)g(x)}{(x - c)g(x)g(c)} \\ &= \lim_{x\to c} \frac{[f(x) - f(c)]g(c) - f(c)[g(x) - g(c)]}{(x - c)g(x)g(c)} \\ &= \lim_{x\to c} \left[\frac{f(x) - f(c)}{x - c} \cdot \frac{g(c)}{g(x)g(c)} - \frac{f(c)}{g(c)} \cdot \frac{g(x) - g(c)}{x - c} \cdot \frac{1}{g(x)}\right] \\ &= f'(c) \cdot \frac{1}{g(c)} - \frac{f(c)}{g(c)} \cdot g'(c) \cdot \frac{1}{g(c)} \\ &= \frac{f'(c)g(c) - f(c)g'(c)}{g(c)^2} \end{aligned}$$

Therefore $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$. ■ ■

Example 11.3 (Using Differentiation Rules)**.** *Find the derivative of* $h(x) = (3x^2 + 5x)(x^3 - 2)$.

Method 1 (Product rule)*:*

$$\begin{aligned} h'(x) &= (3x^2 + 5x)' \cdot (x^3 - 2) + (3x^2 + 5x) \cdot (x^3 - 2)' \\ &= (6x + 5)(x^3 - 2) + (3x^2 + 5x)(3x^2) \\ &= 6x^4 - 12x + 5x^3 - 10 + 9x^4 + 15x^3 \\ &= 15x^4 + 20x^3 - 12x - 10 \end{aligned}$$

Method 2 (Expand first)*:*

$$h(x) = 3x^5 + 5x^4 - 6x^2 - 10x$$

$$h'(x) = 15x^4 + 20x^3 - 12x - 10$$

Both methods agree. ✓

Theorem 11.4 (Chain Rule)**.** *If g is differentiable at c and f is differentiable at $g(c)$, then $f \circ g$ is differentiable at c, and:*

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c)$$

In Leibniz notation: If $y = f(u)$ and $u = g(x)$, then:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Proof Sketch. The intuitive "proof" is:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

("canceling" du).

But $\frac{dy}{dx}$ is not a fraction—it's a limit! The rigorous proof is more subtle.

Define $\phi(h) = \frac{g(c+h)-g(c)}{h} - g'(c)$ for $h \neq 0$, and $\phi(0) = 0$.

Then $\phi(h) \to 0$ as $h \to 0$ (by definition of $g'(c)$), and:

$$g(c+h) = g(c) + [g'(c) + \phi(h)]h$$

Let $k = g(c+h) - g(c)$. Similarly:

$$f(g(c)+k) = f(g(c)) + [f'(g(c)) + \psi(k)]k$$

where $\psi(k) \to 0$ as $k \to 0$.

Substituting:

$$\begin{aligned}(f \circ g)(c+h) &= f(g(c+h)) \\ &= f(g(c)+k) \\ &= f(g(c)) + [f'(g(c)) + \psi(k)]k \\ &= f(g(c)) + [f'(g(c)) + \psi(k)][g'(c) + \phi(h)]h\end{aligned}$$

Therefore:

$$\frac{(f \circ g)(c+h) - (f \circ g)(c)}{h} = [f'(g(c)) + \psi(k)][g'(c) + \phi(h)]$$

Taking $h \to 0$ (which forces $k \to 0$ by continuity of g):

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c)$$

■ ■

Example 11.4 (Chain Rule)**.** *Find the derivative of* $h(x) = (x^2 + 3x)^5$.

Solution*: Let* $f(u) = u^5$ *and* $g(x) = x^2 + 3x$*. Then* $h = f \circ g$.

By chain rule:

$$\begin{aligned}h'(x) &= f'(g(x)) \cdot g'(x) \\ &= 5(g(x))^4 \cdot (2x+3) \\ &= 5(x^2+3x)^4 \cdot (2x+3)\end{aligned}$$

■

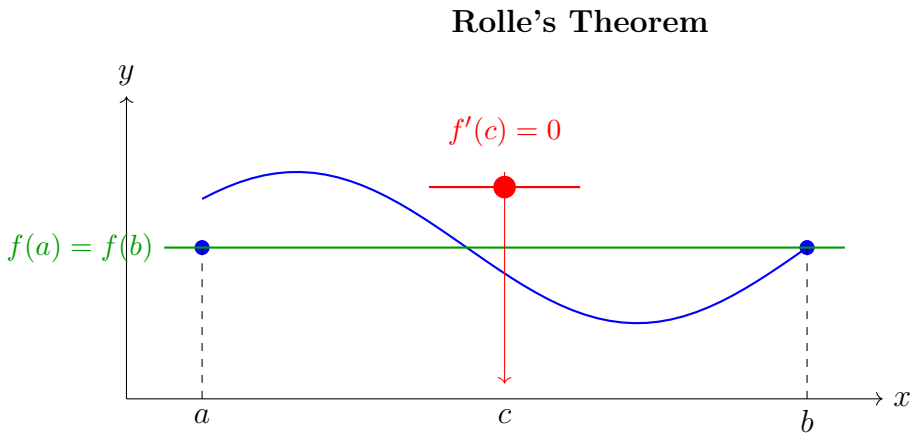
11.5 The Mean Value Theorem

Intuition

If you drive 100 miles in 2 hours, at some moment your instantaneous speed was exactly 50 mph (the average).
More generally: Between any two points on a differentiable curve, there's a point where the tangent line is parallel to the secant line.
This simple-sounding theorem has profound consequences for all of analysis.

Theorem 11.5 (Rolle's Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).*

If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.



If $f(a) = f(b)$, then $\exists c$ with horizontal tangent

Proof. Since f is continuous on $[a, b]$, by EVT, f attains its maximum and minimum.

Case 1: If f is constant, then $f'(x) = 0$ for all $x \in (a, b)$. Done. ✓

Case 2: If f is not constant, then either the maximum or minimum occurs at an interior point $c \in (a, b)$ (since $f(a) = f(b)$, they can't both be at endpoints if f is non-constant).

Without loss of generality, assume f attains its maximum at $c \in (a, b)$.

Then $f(c) \geq f(x)$ for all $x \in [a, b]$.

For $h > 0$ small:

$$\frac{f(c+h) - f(c)}{h} \leq 0 \quad (\text{since } f(c+h) \leq f(c))$$

Taking $h \to 0^+$: $f'(c) \leq 0$.

For $h < 0$ small:

$$\frac{f(c+h) - f(c)}{h} \geq 0 \quad (\text{negative divided by negative})$$

Taking $h \to 0^-$: $f'(c) \geq 0$.

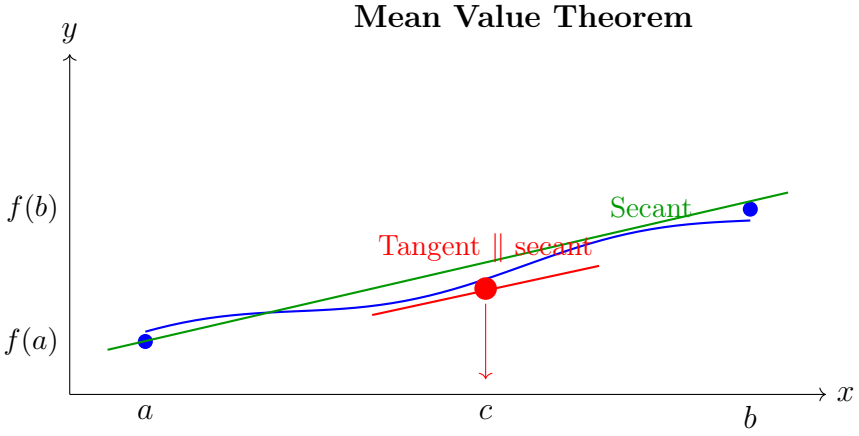
Therefore $f'(c) = 0$. ■ ■

Theorem 11.6 (Mean Value Theorem (MVT))**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).*

Then there exists $c \in (a, b)$ such that:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

***In words**: The instantaneous rate of change at some point equals the average rate of change.*



$\exists c$ where tangent slope = secant slope

Proof. Define an auxiliary function that measures the vertical distance from the secant line:

$$g(x) = f(x) - \left[f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right]$$

(The term in brackets is the secant line through $(a, f(a))$ and $(b, f(b))$.)

Properties of g:

- g is continuous on $[a, b]$ and differentiable on (a, b) (since f is)
- $g(a) = f(a) - f(a) = 0$
- $g(b) = f(b) - f(b) = 0$

By Rolle's Theorem, there exists $c \in (a, b)$ with $g'(c) = 0$.

But:

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

Therefore:

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

Rearranging:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

■ ■

Key Idea

MVT is the foundation of differential calculus. Almost every major theorem follows from it:

- $f' = 0 \implies f$ is constant
- $f' > 0 \implies f$ is increasing
- $f' = g' \implies f = g + C$
- Taylor's theorem (approximating functions by polynomials)

11.6 Consequences of the Mean Value Theorem

Theorem 11.7 (Zero Derivative Implies Constant)**.** *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on (a, b).*

Proof. Let $x_1, x_2 \in (a, b)$ with $x_1 < x_2$.

By MVT applied to $[x_1, x_2]$, there exists $c \in (x_1, x_2)$ such that:

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

Since $f'(c) = 0$:

$$f(x_2) - f(x_1) = 0 \implies f(x_2) = f(x_1)$$

Since x_1, x_2 were arbitrary, f is constant. ■ ■

Theorem 11.8 (Increasing/Decreasing Test)**.** *Let f be continuous on $[a, b]$ and differentiable on (a, b).*

1. *If $f'(x) > 0$ for all $x \in (a, b)$, then f is strictly increasing on $[a, b]$*
2. *If $f'(x) < 0$ for all $x \in (a, b)$, then f is strictly decreasing on $[a, b]$*
3. *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is increasing on $[a, b]$*

Proof of (1). Let $x_1 < x_2$ in $[a, b]$.

By MVT, there exists $c \in (x_1, x_2)$ such that:

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

Since $f'(c) > 0$ and $x_2 - x_1 > 0$:

$$f(x_2) - f(x_1) > 0 \implies f(x_2) > f(x_1)$$

Therefore f is strictly increasing. ■ ■

Example 11.5 (Finding Intervals of Increase/Decrease). *Let $f(x) = x^3 - 3x^2 + 2$. Find where f is increasing and decreasing.*

Solution:

$$f'(x) = 3x^2 - 6x = 3x(x-2)$$

Critical points: *$f'(x) = 0$ when $x = 0$ or $x = 2$.*

Sign analysis:

- *$x < 0$: $f'(x) = 3(-)(-) = (+) > 0 \implies f$ increasing*
- *$0 < x < 2$: $f'(x) = 3(+)(-) = (-) < 0 \implies f$ decreasing*
- *$x > 2$: $f'(x) = 3(+)(+) = (+) > 0 \implies f$ increasing*

Therefore: f increases on $(-\infty, 0]$, decreases on $[0, 2]$, increases on $[2, \infty)$. ■

Theorem 11.9 (Antiderivatives Differ by a Constant). *If $f'(x) = g'(x)$ for all $x \in (a, b)$, then there exists a constant C such that $f(x) = g(x) + C$ for all $x \in (a, b)$.*

Proof. Let $h(x) = f(x) - g(x)$.

Then $h'(x) = f'(x) - g'(x) = 0$ for all $x \in (a, b)$.

By the previous theorem, h is constant, say $h(x) = C$.

Therefore $f(x) - g(x) = C$, i.e., $f(x) = g(x) + C$. ■ ■

Key Idea

This theorem justifies the "$+C$" in antiderivatives:
If $F'(x) = f(x)$, then *any* antiderivative of f has the form $F(x) + C$.
This is why indefinite integrals always include $+C$!

11.7 Higher Derivatives and Concavity

Definition 11.3 (Higher Derivatives). *If f' is differentiable, we define the* ***second derivative:***

$$f''(x) = (f')'(x) = \frac{d^2 f}{dx^2}$$

Similarly: f''' (third derivative), $f^{(4)}$ (fourth), ..., $f^{(n)}$ (n-th derivative).

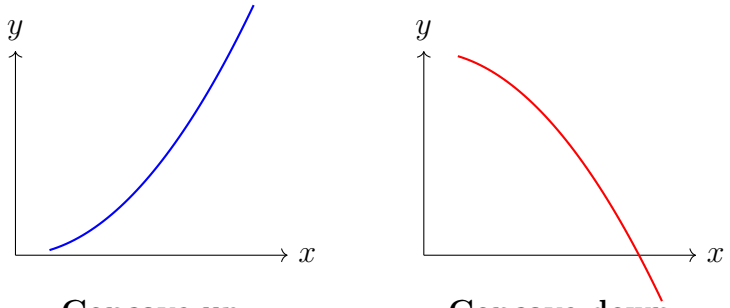
A function is C^n if $f^{(n)}$ exists and is continuous.

A function is ***smooth*** *or C^∞ if $f^{(n)}$ exists for all n.*

Definition 11.4 (Concavity). *A function f is:*

- ***Concave up*** *on (a, b) if $f''(x) > 0$ for all $x \in (a, b)$*
- ***Concave down*** *on (a, b) if $f''(x) < 0$ for all $x \in (a, b)$*

A point c where concavity changes is an ***inflection point.***



Concave up
$f'' > 0$
"Holds water"

Concave down
$f'' < 0$
"Spills water"

Theorem 11.10 (Second Derivative Test)**.** *Let f'' be continuous near c, and suppose $f'(c) = 0$.*

1. *If $f''(c) > 0$, then f has a local minimum at c*
2. *If $f''(c) < 0$, then f has a local maximum at c*
3. *If $f''(c) = 0$, the test is inconclusive*

Proof Sketch. If $f'(c) = 0$ and $f''(c) > 0$, then f' is increasing near c (since $f'' > 0$).
Therefore $f' < 0$ for $x < c$ (just left of c) and $f' > 0$ for $x > c$ (just right).
So f decreases before c and increases after c, making c a local minimum. ■ ■

11.8 L'Hôpital's Rule

Intuition

How do we compute limits like $\lim_{x\to 0} \frac{\sin x}{x}$ or $\lim_{x\to\infty} \frac{e^x}{x^2}$?
When both numerator and denominator approach 0 (or both ∞), we get indeterminate forms: $\frac{0}{0}$ or $\frac{\infty}{\infty}$.
L'Hôpital's Rule: In such cases, we can differentiate the numerator and denominator separately!

Theorem 11.11 (L'Hôpital's Rule, 1696)**.** *Suppose f and g are differentiable on an open interval I containing a (except possibly at a), and $g'(x) \neq 0$ for $x \in I \setminus \{a\}$.*

Case 1 ($\frac{0}{0}$ form): *If $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, and if $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists, then:*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

Case 2 ($\frac{\infty}{\infty}$ form): *If $\lim_{x\to a} |g(x)| = \infty$ and $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ exists, then:*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

The theorem also holds for one-sided limits and limits at $\pm\infty$.

Proof of Case 1 (Sketch). We prove the $\frac{0}{0}$ case. Assume $f(a) = g(a) = 0$ (extend by continuity if needed).

For x near a (but $x \neq a$), apply the Cauchy Mean Value Theorem (a generalization of MVT):

There exists c between a and x such that:

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}$$

Since $f(a) = g(a) = 0$:

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}$$

As $x \to a$, we have $c \to a$ (since c is between a and x).

If $\lim_{x\to a} \frac{f'(x)}{g'(x)} = L$, then:

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{c\to a} \frac{f'(c)}{g'(c)} = L$$

■ ■

Warning

Common mistakes:

1. L'Hôpital's Rule is **NOT** the Quotient Rule!

   We differentiate numerator and denominator *separately*, not as a quotient:

   $$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} \quad \text{(L'Hôpital)}$$

   NOT:

   $$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \quad \text{(Quotient Rule)}$$

2. Only use L'Hôpital when you have indeterminate form $\frac{0}{0}$ or $\frac{\infty}{\infty}$

3. If $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ doesn't exist, L'Hôpital's Rule doesn't help (but the original limit might still exist!)

Example 11.6 (Basic Application)**.** *Compute* $\lim_{x\to 0} \frac{\sin x}{x}$.

***Solution**: As $x \to 0$, both $\sin x \to 0$ and $x \to 0$, so we have $\frac{0}{0}$ form.*

Apply L'Hôpital's Rule:

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{(\sin x)'}{(x)'} = \lim_{x\to 0} \frac{\cos x}{1} = \cos 0 = 1$$

Therefore $\lim_{x\to 0} \frac{\sin x}{x} = 1$. ■

Example 11.7 (Multiple Applications)**.** *Compute* $\lim_{x\to 0} \frac{e^x-1-x}{x^2}$.

***Solution**: As $x \to 0$: numerator $\to 0$ and denominator $\to 0$, so $\frac{0}{0}$ form.*

Apply L'Hôpital's Rule:

$$\lim_{x\to 0} \frac{e^x - 1 - x}{x^2} = \lim_{x\to 0} \frac{e^x - 1}{2x}$$

Still $\frac{0}{0}$ form! Apply L'Hôpital again:

$$\lim_{x\to 0} \frac{e^x - 1}{2x} = \lim_{x\to 0} \frac{e^x}{2} = \frac{1}{2}$$

Therefore $\lim_{x\to 0} \frac{e^x-1-x}{x^2} = \frac{1}{2}$. ■

Example 11.8 ($\frac{\infty}{\infty}$ Form)**.** *Compute* $\lim_{x\to\infty} \frac{x^2}{e^x}$.

***Solution**: As $x \to \infty$: both $x^2 \to \infty$ and $e^x \to \infty$, so $\frac{\infty}{\infty}$ form.*

Apply L'Hôpital's Rule:

$$\lim_{x\to\infty} \frac{x^2}{e^x} = \lim_{x\to\infty} \frac{2x}{e^x}$$

Still $\frac{\infty}{\infty}$! Apply again:

$$\lim_{x\to\infty} \frac{2x}{e^x} = \lim_{x\to\infty} \frac{2}{e^x} = 0$$

Therefore $\lim_{x\to\infty} \frac{x^2}{e^x} = 0$.

***Interpretation**: Exponential functions grow faster than polynomials!* ■

11.9 Looking Forward: Integration

Intuition

Differentiation answers: "What is the rate of change?"
Integration (next chapter) answers the reverse question: "What function has this rate of change?"
Also: Integration computes areas, volumes, arc lengths, work, probability distributions, and more.
The **Fundamental Theorem of Calculus** connects differentiation and integration:

$$\int_a^b f'(x)\,dx = f(b) - f(a)$$

Differentiation and integration are inverse operations—this is the crowning achievement of calculus.

Differentiation: The Foundation Is Complete

Derivative as limit → Algebra of derivatives → MVT → Applications

We can now analyze rates of change, optimization, curve sketching.
Next: Integration—the inverse operation and the key to computing totals.

"The derivative measures; the integral totals."

Chapter 12

Integration: The Fundamental Theorem of Calculus

12.1 From Differentiation to Integration

Intuition

Differentiation asks: "What is the rate of change?"
Integration asks two related questions:

1. **Antiderivative problem**: What function has f as its derivative?
2. **Area problem**: What is the area under the curve $y = f(x)$?

The miracle: These two problems have the *same answer*.
The **Fundamental Theorem of Calculus** connects them:

$$\text{Area under } f \text{ from } a \text{ to } b = F(b) - F(a), \quad \text{where } F' = f$$

This chapter makes this connection rigorous.

Historical Context

Ancient Origins (300 BCE - 1600 CE)
Archimedes (c. 250 BCE):

- Computed areas using **method of exhaustion**
- Found area of parabolic segment: $\frac{4}{3} \times$ triangle
- Computed π by exhausting circle with polygons
- Method: Inscribe and circumscribe, then squeeze

The Dark Ages (500-1400 CE): Greek works preserved in Arabic translations
Early Modern (1400-1650):

- **Cavalieri (1635)**: "Indivisibles"—areas as infinite sums of lines

- **Fermat (1636)**: Found areas under $y = x^n$ by summing rectangles
- **Wallis (1656)**: Extended to rational exponents

The Breakthrough (1665-1675)
Newton (1665-1666) (unpublished until 1704):

- Discovered: Antiderivatives compute areas
- *"Integration is the inverse of differentiation"*
- Used for orbits, optics, gravitation

Leibniz (1673-1675):

- Independent discovery of Fundamental Theorem
- Invented notation: $\int$ (elongated S for "sum"), dx (infinitesimal)
- $\int f(x)\,dx$ read as "sum of $f(x)$ times infinitesimal dx"
- His notation won: we still use $\int$ and dx today

18th Century (1700-1800): Euler, Lagrange, Laplace—powerful techniques, no rigor
Rigorization (1800-1900)
Cauchy (1823):

- First rigorous definition of integral as limit of sums
- Proved Fundamental Theorem using limits

Riemann (1854):

- Generalized Cauchy's approach
- **Riemann integral**: $\int_a^b f = \lim \sum f(x_i^*)\Delta x_i$
- Characterized integrable functions (continuous except at finitely many points)

20th Century: Lebesgue (1902) invented more powerful integral for measure theory. But Riemann's integral suffices for calculus.
Modern view: Integration is the inverse of differentiation, and also computes signed areas.

12.2 The Riemann Integral: Definition

Definition 12.1 (Partition)**.** *A **partition** of $[a, b]$ is a finite sequence:*

$$P = \{x_0, x_1, \ldots, x_n\} \quad \textit{where} \quad a = x_0 < x_1 < \cdots < x_n = b$$

*The **mesh** or **norm** of P is:*

$$\|P\| = \max_{1 \le i \le n} (x_i - x_{i-1})$$

(the width of the largest subinterval).

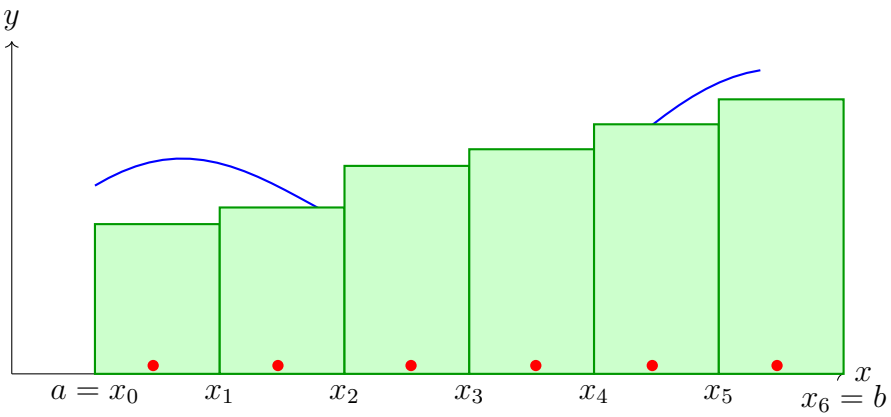
Definition 12.2 (Riemann Sum)**.** *Let $f : [a, b] \to \mathbb{R}$ be bounded, and let $P = \{x_0, \ldots, x_n\}$ be a partition.*

*Choose **sample points** $x_i^* \in [x_{i-1}, x_i]$ for each i.*

*The **Riemann sum** is:*

$$S(f, P, \{x_i^*\}) = \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}) = \sum_{i=1}^{n} f(x_i^*)\Delta x_i$$

***Geometric interpretation**: Sum of signed areas of rectangles.*



Riemann sum: $\sum f(x_i^*)\Delta x_i$
As mesh $\to$ 0, rectangles $\to$ true area

Definition 12.3 (Riemann Integrability)**.** *A function $f : [a, b] \to \mathbb{R}$ is **Riemann integrable** if there exists $L \in \mathbb{R}$ such that:*

For every $\epsilon > 0$, there exists $\delta > 0$ such that for any partition P with $\|P\| < \delta$ and any choice of sample points $\{x_i^*\}$:

$$|S(f, P, \{x_i^*\}) - L| < \epsilon$$

*We write $L = \int_a^b f(x)\,dx$ and call this the **Riemann integral** of f over $[a, b]$.*

***In words**: All Riemann sums converge to the same limit as mesh $\to 0$.*

Alternative Approach: Darboux Sums

The Riemann integral as defined above uses **tagged partitions** (partitions with chosen sample points). This is intuitive but technically cumbersome for proofs.

An equivalent definition uses **Darboux sums** (upper and lower sums):

$$U(f,P) = \sum_{i=1}^{n} \sup_{x\in[x_{i-1},x_i]} f(x)\cdot(x_i - x_{i-1}) \quad \text{(Upper sum)}$$

$$L(f,P) = \sum_{i=1}^{n} \inf_{x\in[x_{i-1},x_i]} f(x)\cdot(x_i - x_{i-1}) \quad \text{(Lower sum)}$$

A function is Riemann integrable if and only if:

$$\sup_P L(f,P) = \inf_P U(f,P)$$

Advantage: No need to consider all possible sample point choices—just supremum and infimum over each subinterval. Many theorems (especially integrability criteria) have cleaner proofs using Darboux sums.
For this text, we use the tagged partition approach for its intuitive connection to approximating areas, but readers should be aware that the Darboux formulation is often preferred for technical work.

Key Idea

Three key ideas:

1. The integral is a **limit** (like derivatives)
2. The limit must be **independent** of choice of partition and sample points
3. Not all functions are integrable (e.g., Dirichlet's function: $f(x) = 1$ if $x \in \mathbb{Q}$, $f(x) = 0$ if $x \notin \mathbb{Q}$)

Theorem 12.1 (Continuous Functions are Integrable)**.** *If $f : [a,b] \to \mathbb{R}$ is continuous, then f is Riemann integrable.*

Proof Sketch. Since f is continuous on the compact interval $[a,b]$, by uniform continuity theorem (Chapter 11), f is uniformly continuous.

Given $\epsilon > 0$, choose $\delta > 0$ such that:

$$|x-y| < \delta \implies |f(x) - f(y)| < \frac{\epsilon}{b-a}$$

For any partition P with $\|P\| < \delta$, on each subinterval $[x_{i-1}, x_i]$, the variation of f is at most $\frac{\epsilon}{b-a}$.

Therefore, for any two choices of sample points, the Riemann sums differ by at most:

$$\sum_{i=1}^{n} \frac{\epsilon}{b-a}(x_i - x_{i-1}) = \frac{\epsilon}{b-a}\cdot(b-a) = \epsilon$$

By Cauchy criterion (analogous to sequences), the Riemann sums converge. ■

(A complete proof requires more care with the Cauchy criterion for integrals.) ■

12.3 Properties of the Integral

Theorem 12.2 (Linearity of Integration)**.** *If f and g are integrable on $[a, b]$, then:*

1. $\int_a^b [f(x) + g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$
2. $\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$ *for any* $c \in \mathbb{R}$

Proof. These follow directly from linearity of limits and sums:

$$\sum [f(x_i^*) + g(x_i^*)]\Delta x_i = \sum f(x_i^*)\Delta x_i + \sum g(x_i^*)\Delta x_i$$

Taking limits as $\|P\| \to 0$ gives the result. ■ ■

Theorem 12.3 (Comparison Properties)**.** *If f and g are integrable on $[a, b]$:*

1. *If $f(x) \geq 0$ for all $x \in [a, b]$, then $\int_a^b f(x)\,dx \geq 0$*
2. *If $f(x) \leq g(x)$ for all $x \in [a, b]$, then $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$*
3. $\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx$

Proof. (1): If $f(x) \geq 0$, then every Riemann sum satisfies $S(f, P, \{x_i^*\}) \geq 0$. Taking limits preserves inequalities.

(2): Apply (1) to $g - f \geq 0$ and use linearity.

(3): Note that $-|f(x)| \leq f(x) \leq |f(x)|$. Integrate and use (2). ■ ■

Theorem 12.4 (Additivity Over Intervals)**.** *If f is integrable on $[a, c]$ and $[c, b]$ where $a < c < b$, then:*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

Proof. Consider a partition P of $[a, b]$ that includes c as a partition point.

Then P splits into partitions P_1 of $[a, c]$ and P_2 of $[c, b]$, and:

$$S(f, P, \{x_i^*\}) = S(f, P_1, \{x_i^*\}) + S(f, P_2, \{x_i^*\})$$

Taking limits as $\|P\| \to 0$ gives the result. ■ ■

Definition 12.4 (Conventions)**.** *We extend the integral notation by defining:*

1. $\int_a^a f(x)\,dx = 0$ *(zero-width interval)*
2. $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$ *(reversed limits)*

With these conventions, additivity holds for any ordering of a, c, b.

12.4 The Fundamental Theorem of Calculus

Intuition

The Fundamental Theorem comes in two parts:

- **Part 1**: Integration creates antiderivatives
- **Part 2**: Antiderivatives evaluate definite integrals

Together, they say: *Differentiation and integration are inverse operations.* This is the **central result of calculus**.

Theorem 12.5 (Fundamental Theorem of Calculus, Part 1)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous. Define:*

$$F(x) = \int_a^x f(t)\,dt$$

Then F is differentiable on (a, b) and $F'(x) = f(x)$ for all $x \in (a, b)$.
***In words**: The function $F(x) = \int_a^x f(t)\,dt$ is an antiderivative of f.*

Proof. Fix $x \in (a, b)$. We compute $F'(x)$ from the definition:

$$\begin{aligned} F'(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left[\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right] \\ &= \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,dt \quad \text{(by additivity)} \end{aligned}$$

Since f is continuous at x, for any $\epsilon > 0$, there exists $\delta > 0$ such that:

$$|t - x| < \delta \implies |f(t) - f(x)| < \epsilon$$

For $|h| < \delta$, all $t \in [x, x + h]$ (or $[x + h, x]$ if $h < 0$) satisfy $|t - x| < \delta$, so:

$$f(x) - \epsilon < f(t) < f(x) + \epsilon$$

Integrating over $[x, x + h]$ (assuming $h > 0$ for simplicity):

$$(f(x) - \epsilon)h < \int_x^{x+h} f(t)\,dt < (f(x) + \epsilon)h$$

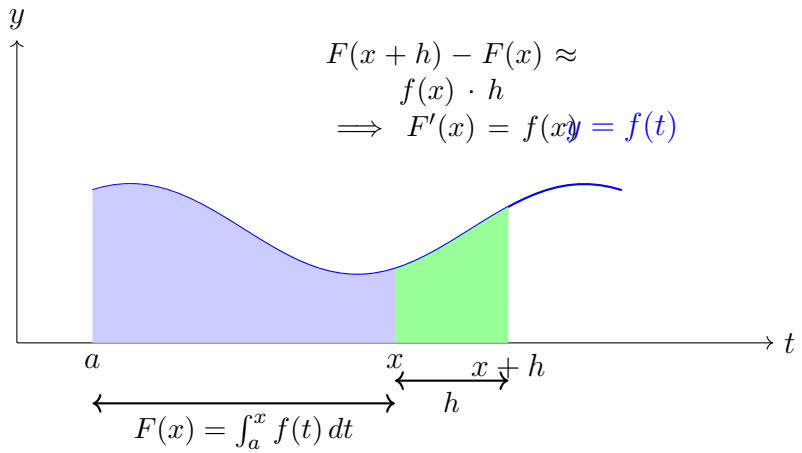
Dividing by $h > 0$:

$$f(x) - \epsilon < \frac{1}{h}\int_x^{x+h} f(t)\,dt < f(x) + \epsilon$$

Taking $h \to 0$, we squeeze:

$$F'(x) = f(x)$$

The case $h < 0$ is similar. ■ ■



Theorem 12.6 (Fundamental Theorem of Calculus, Part 2)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous, and let F be any antiderivative of f (i.e., $F'(x) = f(x)$).*

Then:

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

***Notation**: We write $F(b) - F(a) = [F(x)]_a^b$ or $F(x)\Big|_a^b$.*

Proof. By Part 1, we know that $G(x) = \int_a^x f(t)\,dt$ is an antiderivative of f.

Since F is also an antiderivative of f, we have $F'(x) = G'(x) = f(x)$ for all $x \in (a, b)$.

By the theorem from Chapter 12 (antiderivatives differ by a constant), there exists C such that:

$$F(x) = G(x) + C$$

Evaluating at $x = a$:

$$F(a) = G(a) + C = \int_a^a f(t)\,dt + C = 0 + C = C$$

Therefore $C = F(a)$, so $F(x) = G(x) + F(a)$.

Evaluating at $x = b$:

$$F(b) = G(b) + F(a) = \int_a^b f(t)\,dt + F(a)$$

Rearranging:

$$\int_a^b f(t)\,dt = F(b) - F(a)$$

■ ■

Key Idea

The Fundamental Theorem says:
To compute $\int_a^b f(x)\,dx$, you don't need to compute limits of Riemann sums!
Instead:

1. Find *any* antiderivative F of f (i.e., $F' = f$)
2. Evaluate $F(b) - F(a)$

This transforms integration into antidifferentiation—a much easier problem.

Example 12.1 (Using FTC Part 2)**.** *Compute* $\int_0^2 x^2\,dx$.

***Solution**: We need an antiderivative of* $f(x) = x^2$.

Since $\frac{d}{dx}\left(\frac{x^3}{3}\right) = x^2$, *we can take* $F(x) = \frac{x^3}{3}$.

By FTC Part 2:

$$\int_0^2 x^2\,dx = \left[\frac{x^3}{3}\right]_0^2 = \frac{8}{3} - \frac{0}{3} = \frac{8}{3}$$

■

***Geometric verification**: The area under* $y = x^2$ *from* 0 *to* 2 *is indeed* $\frac{8}{3}$ *(can be verified by Riemann sums).*

Example 12.2 (Signed Areas)**.** *Compute* $\int_{-1}^1 x\,dx$.

***Solution**:* $F(x) = \frac{x^2}{2}$ *is an antiderivative of* x.

$$\int_{-1}^1 x\,dx = \left[\frac{x^2}{2}\right]_{-1}^1 = \frac{1}{2} - \frac{1}{2} = 0$$

***Interpretation**: The area above the* x*-axis (for* $x > 0$*) exactly cancels the area below (for* $x < 0$*).*

Integrals compute ***signed area****, not total area.* ■

12.5 Integration Techniques

Theorem 12.7 (Substitution Rule)**.** *Let* $g : [a, b] \to \mathbb{R}$ *be continuously differentiable, and let* f *be continuous on the range of* g.

Then:

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du$$

***Mnemonic**: Set* $u = g(x)$, *so* $du = g'(x)\,dx$. *Then "substitute".*

Proof. Let F be an antiderivative of f, so $F' = f$.

By the chain rule:

$$\frac{d}{dx}[F(g(x))] = F'(g(x)) \cdot g'(x) = f(g(x)) \cdot g'(x)$$

Therefore $F(g(x))$ is an antiderivative of $f(g(x))g'(x)$.

By FTC Part 2:

$$\int_a^b f(g(x))g'(x)\,dx = [F(g(x))]_a^b = F(g(b)) - F(g(a))$$

But also:

$$\int_{g(a)}^{g(b)} f(u)\,du = [F(u)]_{g(a)}^{g(b)} = F(g(b)) - F(g(a))$$

Therefore the two integrals are equal. ■ ■

Example 12.3 (Substitution)**.** *Compute* $\int_0^1 2xe^{x^2}\,dx$*.*

Solution*: Let* $u = x^2$*, so* $du = 2x\,dx$*.*

When $x = 0$*:* $u = 0$*. When* $x = 1$*:* $u = 1$*.*

Therefore:

$$\int_0^1 2xe^{x^2}\,dx = \int_0^1 e^u\,du = [e^u]_0^1 = e - 1$$

■

Theorem 12.8 (Integration by Parts)**.** *If* u *and* v *are continuously differentiable on* $[a, b]$*, then:*

$$\int_a^b u(x)v'(x)\,dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)\,dx$$

Mnemonic*:* $\int u\,dv = uv - \int v\,du$*.*

Proof. By the product rule:

$$\frac{d}{dx}[u(x)v(x)] = u'(x)v(x) + u(x)v'(x)$$

Rearranging:

$$u(x)v'(x) = \frac{d}{dx}[u(x)v(x)] - u'(x)v(x)$$

Integrating both sides from a to b:

$$\int_a^b u(x)v'(x)\,dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)\,dx$$

■ ■

Example 12.4 (Integration by Parts)**.** *Compute* $\int_0^1 xe^x\,dx$*.*

Solution*: Let* $u = x$ *(so* $u' = 1$*) and* $v' = e^x$ *(so* $v = e^x$*).*

By integration by parts:

$$\begin{aligned}\int_0^1 xe^x\,dx &= [xe^x]_0^1 - \int_0^1 e^x\,dx \\ &= (1 \cdot e) - (0 \cdot 1) - [e^x]_0^1 \\ &= e - (e - 1) \\ &= 1\end{aligned}$$

■

12.6 Applications of Integration

Example 12.5 (Area Between Curves). *Find the area between $y = x^2$ and $y = x$ from $x = 0$ to $x = 1$.*

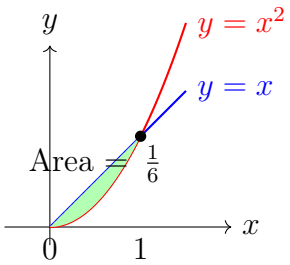
***Solution**: The curves intersect at $x = 0$ and $x = 1$.*

For $0 \le x \le 1$, we have $x \ge x^2$ (since $x - x^2 = x(1-x) \ge 0$).

Area between curves:

$$\begin{aligned} A &= \int_0^1 (x - x^2)\,dx \\ &= \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 \\ &= \frac{1}{2} - \frac{1}{3} \\ &= \frac{1}{6} \end{aligned}$$

■



Example 12.6 (Arc Length). *The **arc length** of a curve $y = f(x)$ from $x = a$ to $x = b$ is:*

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx$$

***Derivation**: An infinitesimal segment has length:*

$$ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx = \sqrt{1 + [f'(x)]^2}\,dx$$

Integrating gives total arc length.

***Example**: Arc length of $y = x^{3/2}$ from $x = 0$ to $x = 1$:*

$$L = \int_0^1 \sqrt{1 + \left(\frac{3}{2}x^{1/2}\right)^2}\,dx = \int_0^1 \sqrt{1 + \frac{9x}{4}}\,dx$$

(This can be computed using substitution $u = 1 + \frac{9x}{4}$.)

Example 12.7 (Volume of Revolution). *Rotating $y = f(x)$ around the x-axis from $x = a$ to $x = b$ creates a solid with volume:*

$$V = \pi \int_a^b [f(x)]^2\,dx$$

***Derivation**: A thin disk at position x has:*

- *Radius:* $r = f(x)$
- *Thickness:* dx
- *Volume:* $\pi r^2 dx = \pi[f(x)]^2\,dx$

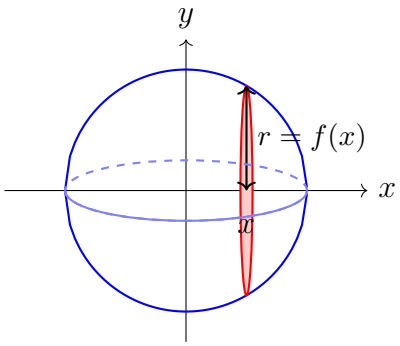
Integrating sums all disks.

***Example**: Volume of sphere of radius R:*

Rotate $y = \sqrt{R^2 - x^2}$ (upper semicircle) around x-axis from $x = -R$ to $x = R$:

$$\begin{aligned} V &= \pi \int_{-R}^{R} (R^2 - x^2)\,dx \\ &= \pi \left[R^2 x - \frac{x^3}{3} \right]_{-R}^{R} \\ &= \pi \left[\left(R^3 - \frac{R^3}{3} \right) - \left(-R^3 + \frac{R^3}{3} \right) \right] \\ &= \pi \cdot \frac{4R^3}{3} = \frac{4\pi R^3}{3} \end{aligned}$$

The classical formula! ✓



Rotating $y = \sqrt{R^2 - x^2}$
creates a sphere
Volume: $\frac{4\pi R^3}{3}$

12.7 Improper Integrals

Definition 12.5 (Improper Integrals)**.** *If f is continuous on $[a, \infty)$, we define:*

$$\int_a^{\infty} f(x)\,dx = \lim_{t \to \infty} \int_a^t f(x)\,dx$$

provided the limit exists.

Similarly, if f has a vertical asymptote at $x = b$, we define:

$$\int_a^b f(x)\,dx = \lim_{t \to b^-} \int_a^t f(x)\,dx$$

If the limit exists (and is finite), the improper integral ***converges****. Otherwise it* ***diverges****.*

Example 12.8 (Convergent Improper Integral)**.** *Compute* $\int_1^\infty \frac{1}{x^2}\,dx$*.*

Solution*:*

$$\begin{aligned}\int_1^\infty \frac{1}{x^2}\,dx &= \lim_{t\to\infty}\int_1^t \frac{1}{x^2}\,dx \\ &= \lim_{t\to\infty}\left[-\frac{1}{x}\right]_1^t \\ &= \lim_{t\to\infty}\left(-\frac{1}{t}+1\right) \\ &= 0+1=1\end{aligned}$$

Therefore $\int_1^\infty \frac{1}{x^2}\,dx = 1$ *(converges).* ■

Example 12.9 (Divergent Improper Integral)**.** *Compute* $\int_1^\infty \frac{1}{x}\,dx$*.*

Solution*:*

$$\begin{aligned}\int_1^\infty \frac{1}{x}\,dx &= \lim_{t\to\infty}\int_1^t \frac{1}{x}\,dx \\ &= \lim_{t\to\infty}[\ln x]_1^t \\ &= \lim_{t\to\infty}(\ln t - \ln 1) \\ &= \lim_{t\to\infty}\ln t = \infty\end{aligned}$$

Therefore $\int_1^\infty \frac{1}{x}\,dx$ *diverges.* ■

Moral*:* $\int_1^\infty \frac{1}{x^p}\,dx$ *converges if and only if* $p > 1$*.*

12.8 Looking Forward: Complex Numbers and Beyond

Intuition

With integration, we've completed the core of **single-variable calculus**:

- Limits and continuity
- Derivatives and rates of change
- Integrals and accumulation
- The Fundamental Theorem connecting them

Next steps in the compendium:

1. **Complex numbers**: Extending $\mathbb{R}$ to $\mathbb{C}$, solving $x^2+1=0$
2. **Abstract algebra**: Groups, rings, fields—the structure behind arithmetic
3. **Linear algebra**: Vector spaces, matrices, linear transformations

4. **Topology**: Generalizing continuity beyond $\mathbb{R}$
5. **Multivariable calculus**: Derivatives and integrals in $\mathbb{R}^n$

Each builds on the foundations we've laid.

Calculus Complete: The Inverse Operations United

Riemann integral as limit → FTC Part 1 (integration creates antiderivatives) → FTC Part 2 (antiderivatives evaluate integrals) → Applications

Differentiation and integration are inverse operations. This connection is the heart of calculus and the gateway to all of analysis.

"In mathematics, the art of asking questions is more valuable than solving problems."
— Georg Cantor

About This Book

Foundations of Mathematics presents a complete, rigorous construction of mathematics from first principles. Beginning with formal logic and axiomatic set theory, this text builds the number systems (, , ,), develops the foundations of analysis, and proves the fundamental theorems of calculus.

Written in the spirit of Bourbaki but with modern pedagogical clarity, each concept is motivated intuitively before formal definitions. The text includes:

- Complete ZFC axioms with detailed explanations
- Rigorous proofs of all major theorems (MVT, FTC, IVT, EVT, Cantor's diagonal argument)
- Color-coded exposition: intuition boxes, key ideas, warnings, and historical notes
- Comprehensive coverage from foundations through integration
- TikZ diagrams throughout for visual clarity

Suitable for: Advanced undergraduates, graduate students, and anyone seeking a complete, rigorous understanding of mathematical foundations.

The Collins Compendium
Formal Edition

Building Mathematics from the Ground Up

Typeset in LaTeX • **Collins Mwangi** • 2025