Set Theory

Muchang Bahng

Spring 2025

Contents

1 Naive Set Theory

Unlike axiomatic set theories, which are defined using formal logic, naive set theory was defined informally at the end of the 19th century by Cantor, in natural language (like English). It describes the aspects of mathematical sets using words (e.g. *satisfying, such as, ...*) and suffices for the everyday use of set theory in modern mathematics. However, as we will see, this leads to paradoxes.

Definition 1.1 (Set)

A **set** is a well-defined collection of distinct objects, called **elements**.

This definition tells us *what* a set is, but does not define *how* sets can be formed, and what operations on sets will again produce a set. The term *well-defined* cannot by itself guarantee the consistency and unambiguity of what exactly constitutes and what does not constitute a set, and therefore this is not a formal definition. Attempting to achieve this will be done in axiomatic set theory, like ZFC.

Definition 1.2 (Membership)

If x is a member of A, we write $x \in A$. For any x, it must be the case that either $x \in A$ (exclusive or) $x \notin A$.

Definition 1.3 (Equality)

Two sets A and B are defined to be equal, denoted as $A = B$, when they have precisely the same elements. That is, if $x \in A \iff x \in B$. This means that a set is completely determined by its elements, and the description is immaterial.

Definition 1.4 (Empty Set)

There exists an empty set, denoted $\emptyset$ or $\{\}$, which is a set with no members at all. Because a set is described by its elements, there can only be one empty set.

Now we show how to construct sets.

Definition 1.5 (Set-Builder Notation)

We can construct a set in two ways.

1. We list its elements between curly braces.
   (a) The set $\{1, 2\}$ denotes the set containing 1 and 2. By equality $\{1, 2\} = \{2, 1\}$.
   (b) Repetition/multiplicity is irrelevant, and so $\{1, 2, 2\} = \{1, 1, 1, 2\} = \{1, 2\}$
2. We denote
$$S = \{x | P(x)\} \tag{1}$$
where P is a property. If x satisfies this property, then $x \in S$.

Naive set theory claims that this construction *always* produces a set. Therefore, a well-defined property is enough to always produce a set of elements satisfying P.

Example 1.1 (Empty Set)

Let $S = \{x \mid x \neq x\}$. For any x, $P(x)$ is false and so S contains no elements. Therefore $S = \emptyset$.

Example 1.2 (Singleton Set)

The set $\{x \mid x = a\} = \{a\}$.

Example 1.3 (Russell Set)

Let $R = \{x \mid x \notin x\}$, i.e. the set of all sets that do not contain themselves as elements.

Theorem 1.1 (Russell's Paradox)

The Russell set exists and does not exist.

Proof.

We will determine if R is an element of itself.

1. If $R \in R$, then by it does contain itself, so it does not satisfy the property and $R \notin R$.
2. If $R \notin R$, then it satisfies the property, so $R \in R$.

Therefore, it is both the case that $x \in R$ and $x \notin R$, which contradicts the membership definition. Therefore, R is both a set from set-builder construction and not a set due to the membership definition.

Theorem 1.2 (Existence of Universe)

Let U be the set of everything, known as the **universal set**. The universal set does exist and does not exist.

Proof.

We can define $U' = \{x \mid \{\} = \{\}\}$, which defines a set. Then the property P that $\{\} = \{\}$ is always true, and U' would contain everything, and by the definition of equality $U = U'$. Now since the Russell set R is both a set and not a set from Russell's paradox, we have $R \in U$ and $R \notin U$, which means that U cannot exist. Therefore U does not exist.

So the sufficiency a well-defined property to be able to construct a set is *too powerful* in that we can construct *any* set we want. This leads us to construct the Russell set, which opens up a lot of paradoxes. Therefore, we would like to restrict the notion of well-defined in a way, which leads to axiomatic set theories.

Definition 1.6 (Subsets)

Given two sets A and B, A is a **subset** of B if every element of A is also an element of B. A subset of B that is not equal to B is called a **proper subset**.

Theorem 1.3 (Equality)

It follows from the definition of equality that

$$A \subset B \text{ and } B \subset A \iff A = B \tag{2}$$

Definition 1.7 (Power Set)

The set of all subsets of a set A is called the **power set** of A, denoted by 2^A.

We could define other things like the union, etc., but I won't bother with it when I will define them for ZFC later.

2 Zermelo-Fraenkel-Choice (ZFC) Set Theory

So with these paradoxes in mind, we would like to construct an axiomatic formulation of sets. My take is to think that sets "exist" out there somewhere in the universe, and our job is to find them. Cantor with his naive set theory believed that for every meaningful property of things there is a set whose members are exactly all the things with that property. Russell shows this this cannot be the case. Nevertheless, *some* sets exist, and we have intuitive experience thinking about finite sets. Therefore, the axioms of set theory are a limited list of *assumptions* that we hope are true about that actually existing universe of sets. As long as they are true, then whatever we conclude from them by valid reasoning steps must also be true.[1] Hence we have the following definition, which first requires the familiar property of acting like a collection of something, and then obeys the axioms we set.

Definition 2.1 (Set)

A **set** X is anything
1. that has the innate property of containing elements, and
2. obeys the axioms of ZFC.

Let's first talk about the language, where they are defined formally using the axioms in the next subsection. From first-order logic, note that we have the following symbols in our alphabet $\mathcal{L}_{\text{ZFC}}$.

1. The logical connectives $\neg$, $\vee$, $\wedge$.
2. The quantifier symbols $\exists, \forall$
3. Brackets ().

To represent sets, we also need symbols, and the membership property requires us to define a symbol for that too.

1. A countably infinite amount of variables used for representing sets.
2. The set membership symbol $\in$. In fact, when we say $x \in A$, this is a proposition formed from the predicate $P(x)$.

This is what we have to work with so far. We will construct the rest of the symbols $(=, \subset, \supset, \cup, \cap)$ from the axioms. So far we don't even know if there exists any set that obeys the following axioms! Therefore, we will assert the existence of at least one set, namely the empty set.

2.1 ZF Axioms

Now we state the axioms, which is the foundation of ZF set theory.

Axiom 2.1 (Empty Set)

The empty set containing no elements exists.

Definition 2.2 (Empty Set)

The empty set is denoted $\emptyset$.

This asserts the existence of at least 1 set, which we will build on to create more sets.

[1] This idea is called naive Platonism.

Axiom 2.2 (Axiom of Extensionality)

Two sets are equal (are the same set) if they have the same elements.

$$\forall A \forall B \big[\forall x (x \in A \iff x \in B) \iff A = B \big] \tag{3}$$

Definition 2.3 (Equality)

This axiom allows us to define the equality operator $=$, which we now add to our alphabet.

Theorem 2.1 (Sets Don't Contain Repeated Elements)

Furthermore, this axiom also implies that sets are unique up to distinct elements. That is,

$$\{1, 1, 2\} = \{1, 2\} = \{1, 1, 2, 2\} \tag{4}$$

Axiom 2.3 (Axiom of Regularity)

Every non-empty set A contains a member x such that A and x are disjoint sets.

$$\forall A \big[A \neq \emptyset \implies \exists x (x \in A \land A \cap x = \emptyset) \big] \tag{5}$$

This, along with the axioms of pairing and union, implies that no set is an element of itself and that every set has an ordinal rank.

The axiom assists us in regulating which sets are viable and which are not, preventing Russell's paradox.

Axiom 2.4 (Axiom Schema of Restricted Comprehension, or Specification)

Subsets, like in naive set theory, are constructed using set builder notation. In general, the **subset** of a set A obeying a formula $\phi(x)$ with one free variable x may be written as

$$\{x \in A \mid \phi(x)\} \tag{6}$$

The axiom schema of specification states that this subset always exists.[a]

Definition 2.4 (Subset, Superset)

The axiom of specification allows us to denote subsets. Notationally, if A is a subset of B, then we write $A \subset B$. Similarly, we say A is a **superset** of B, written $A \supset B$, if $B \subset A$.

Definition 2.5 (Intersection)

This also allows us to define intersection as

$$A \cap B \coloneqq \{x \in A \mid x \in B\} \tag{7}$$

and we can define the intersection of an arbitrary collection of sets $\mathcal{F}$ as the following. Let $A \in \mathcal{F}$.

$$\bigcap \mathcal{F} \coloneqq \{x \in A \mid \forall B (B \in \mathcal{F} \implies x \in B)\} \tag{8}$$

[a] Note that this axiom does not allow the construction of entities of the more general form $\{x \mid \phi(x)\}$. This restriction is obviously needed to avoid Russell's paradox, hence the name *restricted* comprehension.

Unfortunately, the union cannot be expressed in this specification schema, and we need a separate axiom for this.

Definition 2.6 (Set Minus)

We can however define set minus. Given sets A, B

$$A \setminus B \coloneqq \{x \in A \mid x \notin B\} \tag{9}$$

Definition 2.7 (Set Complement)

Given B and a subset $A \subset B$, the **complement** of A with respect to B is

$$A^c \coloneqq \{x \in B \mid x \notin A\} = B \setminus A \tag{10}$$

Axiom 2.5 (Axiom of Pairing)

If A, B are sets, then there exists a set which contains A and B as elements.[a]

$$\forall A \forall B \exists C((A \in C) \land (B \in C)) \tag{11}$$

This allows us to construct sets from old ones.

Theorem 2.2 (Nested Sets)

By the axiom of pairing, if we have a set X, then $\{X\}$ is also a set, since we can set $A = B = X$ which asserts the existence of $\{X, X\} = \{X\}$.

Axiom 2.6 (Axiom of Union)

For any set of sets $\mathcal{F}$, there is a set A containing every element that is a member of $\mathcal{F}$.

$$\forall \mathcal{F} \exists A \forall X \forall x \big[(x \in X \land X \in \mathcal{F}) \implies x \in A\big] \tag{12}$$

This formula doesn't directly assert the existence of $\cup \mathcal{F}$ (?).

Definition 2.8 (Union)

The set $\cup \mathcal{F}$ can be constructed from A in the above using the axiom schema of restricted comprehension.

$$\cup \mathcal{F} = \{x \in A \mid \exists X(x \in X \land X \in \mathcal{F})\} \tag{13}$$

Axiom 2.7 (Axiom of Infinity)

The axiom of infinity guarantees the existence of at least one infinite set. That is, given a set w, let $S(w) = w \cup \{w\}$ be a set.[a] Then, there exists a set X such that

1. $\emptyset \in X$, and
2. if $w \in X$, then $S(w) \in X$.

In logic terms,

$$\exists X \big[\emptyset \in X \land \forall y(y \in X \implies S(y) \in X)\big] \tag{14}$$

Since we have axiomatically claimed the two premises to be true, by propositional logic, namely

[a] For example, if $A = \{1, 2\}$ and $B = \{2, 3\}$,then $\{\{1, 2\}, \{2, 3\}\}$ exists.

modus ponens, this implies the existence of at least one set X with infinitely many members.

Definition 2.9 (Von Neumann Ordinals)

The **Von Neumann ordinals** is the minimal set X satisfying the axiom of infinity. It is the set containing

$$\begin{aligned} 0 &= \{\} = \emptyset \\ 1 &= \{0\} = \{\emptyset\} \\ 2 &= \{0,1\} = \{\emptyset, \{\emptyset\}\} \\ 3 &= \{0,1,2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \\ 4 &= \{0,1,2,3\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\} \\ \ldots &= \ldots \end{aligned}$$

This provides the foundation to construct the most basic mathematical sets: the natural numbers denoted $\mathbb{N}$.

Now that we have constructed the Von Neumann ordinals, we are allowed to do *indexing*.

Axiom 2.8 (Axiom of Power Set)

The axiom of power set states that for any set A, there is a set B that contains every subset[a] of A.

$$\forall A \exists B \forall S (S \subset A \implies S \in B) \tag{15}$$

The axiom of schema of specification is then used to define the power set as the subset of such B containing the subset of A exactly.

$$2^X = \{Y \in B \mid Y \subset X\} \tag{16}$$

Definition 2.10 (Cartesian Product)

The power set axiom allows for the definition of the **Cartesian product** of two sets X and Y. Note that if $x \in X, y \in Y$, then by the axiom of union $x, y \in X \cup Y$ and by the axiom of power set $\{x\}, \{x,y\} \in \mathcal{P}(X \cup Y)$. Therefore, using the axiom of power set again we can define

$$(x,y) \coloneqq \{\{x\}, \{x,y\}\} \in \mathcal{P}(\mathcal{P}(X \cup Y)) \tag{17}$$

and the Cartesian product is defined

$$X \times Y :== \{(x,y) \in \mathcal{P}(\mathcal{P}(X \cup Y)) \mid x \in X \wedge y \in Y\} \tag{18}$$

which is axiomatically a valid set by the axiom schema of specification.

From this we can define the Cartesian product of any finite collection of sets recursively. It is indeed the case that $(X \times Y) \times Z$ is a different set from $X \times (Y \times Z)$, but as we will see in later functions, we can define a canonical bijection between them, treating them as equivalent. Furthermore, notice that we have not defined the Cartesian product of infinite sets yet. We can define them using functions actually.

[a] Since w is a set, by the axiom of pairing $\{w\}$ is a set, and by the axiom of union $w \cup \{w\}$ is a set.
[a] Note that subset is defined by the axiom of restricted comprehension.

2.2 Functions

The definition of Cartesian products allows us to formally define **correspondences**. The most notable correspondences are *functions* and *relations*.

Definition 2.11 (Function)

Given two sets X, Y, a function $f : X \to Y$ is a subset $f \subset X \times Y$ satisfying the following
1. For all $x \in X$, there exists $y \in Y$ s.t. $(x, y) \in f$.[a]
2. For all $x \in X$ and $y, y' \in Y$, if $(x, y) \in f$ and $(x, y') \in f$, then $y = y'$.[b]

The set X is said to be the **domain** and Y the **codomain**.

$$X \xrightarrow[f]{} Y$$

Figure 1: A diagram representing the function $f : X \to Y$.

Definition 2.12 (Image, Preimage)

Given some $f : X \to Y$ and $A \subset X$, the **image** of A under f is defined

$$f(A) \coloneqq \{y \in Y \mid \exists x \in X(f(x) = y)\} \tag{19}$$

Given $B \subset Y$, the **preimage** of B under f is defined

$$f^{-1}(B) \coloneqq \{x \in X \mid f(x) \in B\} \tag{20}$$

Axiom 2.9 (Axiom Schema of Replacement)

This axiom asserts that the image of a set under any definable function will fall inside a set.

Again, how do we even know for sure that these axioms aren't contradictory? The answer is that we don't, and that is why we take them as axioms rather than provable theorems. Fortunately, from the formulation in the early 20th century up until now, no contradictions have been found, and if there is one, then it would be very bad news for us.

Definition 2.13 (Injectivity, Surjectivity, Bijectivity)

A function $f : X \to Y$ is said to be
1. **injective** if $\forall x \in X, \forall x' \in X\big(f(x) = f(x') \implies x = x'\big)$.
2. **surjective** if $\forall y \in Y \exists x \in X(y = f(x))$.
3. **bijective** if it is injective and surjective.

Definition 2.14 (Inverse Function)

If a function $f : X \to Y$ is bijective, then there exists an **inverse function** $f^{-1} : Y \to X$ satisfying

$$\forall x \in X\big[f(f^{-1}(x)) = f^{-1}(f(x)) = x\big] \tag{21}$$

[a] This says that f must be defined for all inputs in X.
[b] In other words, f must map one element to exactly one element.

Definition 2.15 (Restriction)

If $f : X \to Y$ and $X_0 \subset X$, we define the **restriction** of f to X_0 to be the function mapping to Y whose rule is

$$f|_{X_0} \coloneqq \{(x, f(x)) \in fx \in X_0\} \tag{22}$$

Definition 2.16 (Composition)

Given functions $f : X \to Y$, $g : Y \to Z$, we define the **composite**, denoted $g \circ f$ or $g(f(\cdot))$, of f and g as the subset

$$g \circ f \coloneqq \{(x, z) \in X \times Z \mid \exists y \in Y (f(x) = y \land f(y) = z)\} \tag{23}$$

Theorem 2.3 (Compositions)

A composite is a function.

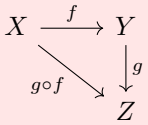

Figure 2: Commutative diagram representing a composition of functions.

Proof.

Using the definition above, we prove the two properties.

1. For all $x \in X$, there exists $y \in Y$ s.t. $(x, y) \in f$. Similarly, for all $y \in Y$, there exists $z \in Z$ s.t. $(y, z) \in g$. Therefore, for all $x \in X$, there exists a $y \in Y$, which follows that there exists also a $z \in Z$. Therefore $g \circ f$ is defined for all inputs in X.
2. For all $x \in X$ and $z, z' \in Z$, say that $(x, z), (x, z') \in g \circ f$. Then by definition of composition there exists a $y, y' \in Y$ s.t. $f(x) = y, f(y) = z$ and $f(x) = y', f(y') = z'$. Since f is a function, $y = y'$. Since g is a function, $y = y' \implies z = z'$. Therefore $g \circ f$ is a function.

For the computer science students, note that a function behaves precisely like functional dependencies in a relational database. A composition represents a natural join.

Theorem 2.4 (Associativity)

Composition is associative. That is, consider $f : Y \to Z, g : X \to Y, h : W \to X$ functions. Then

$$(f \circ g) \circ h = f \circ (g \circ h) \tag{24}$$

Therefore, we write this as

$$f \circ g \circ h \tag{25}$$

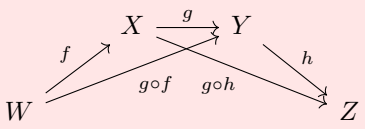

Figure 3

Proof.

Consider any $w \in W$, and let us label $x = h(w)$, $y = g(x)$, $z = f(y)$, where x, y, z must be uniquely determined by w since it is a function. Then,

$$((f \circ g) \circ h)(w) = (f \circ g)(h(w)) = (f \circ g)(x) = z \tag{26}$$
$$(f \circ (g \circ h))(w) = f((g \circ h)(w)) = f(y) = z \tag{27}$$

and they coincide for all $w \in W$.

If we are familiar with algebra, this gives the set of functions $\{f : X \to X\}$ the structure of a *monoid* under composition. We can also talk about commutativity.

Definition 2.17 (Commutativity)

Two functions $f, g : X \to X$ are said to be **commute** if

$$f \circ g = g \circ f \tag{28}$$

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g} \\ X & \xrightarrow{f} & X \end{array}$$

Figure 4: Commutative diagram representing commuting functions f, g.

2.3 Relations

Definition 2.18 (Relation)

A binary relation R on a set A is a subset of $A \times A$. We write aRb if and only if $(a, b) \in R$.[a]

But not all relations may be meaningful or interesting. Therefore we usually like to have certain properties on these relations.

1. *Reflexive.* For all $a \in A$, aRa
2. *Symmetric.* For all $a, b \in A$, if aRb then bRa
3. *Antisymmetric.* For all $a, b \in A$, if aRb and bRa then $a = b$
4. *Transitive.* For all $a, b, c \in A$, if aRb and bRc then aRc
5. *Total.* For all $a, b \in A$, either aRb or bRa

Definition 2.19 (Equivalence Relation)

An **equivalence relation** on a set A is a relation, denoted $\sim$ satisfying
1. *Reflexive.* For all $a \in A$, $a \sim a$
2. *Symmetric.* For all $a, b \in A$, if $a \sim b$ then $b \sim a$
3. *Transitive.* For all $a, b, c \in A$, if $a \sim b$ and $b \sim c$ then $a \sim c$

Given an equivalence relation, we can define an **equivalence class** as

$$[y] \coloneqq \{x \in A \mid x \sim y\} \tag{29}$$

[a] It is a way of describing precisely which two elements are related to one another.

Definition 2.20 (Partition)

A **partition** of a set X is a collection of disjoint nonempty subset of X whose union is all of A.

Theorem 2.5 ()

The set of equivalence classes of a set X with an equivalence relation $\sim$ is a partition of X.

Proof.

Assume the contrary. If X has one element, then its equivalence class is $[x]$ and this is trivially proven. If X has at least 2 elements, let us call them $x, y \in X$ with $x \neq y$. $[x], [y]$ are their equivalence classes. Clearly due to reflexivity, $x \in [x]$ and $y \in [y]$ and so they are nonempty. Since we assumed that this is not a partition, there exists some $z \in X$ in both $[x], [y]$. But $z \in [x] \implies z \sim x$ and $z \in [y] \implies y \sim z$. So by transitivity, $x \sim z$, meaning that $[x] = [y]$. Therefore they must be the same element of a partition.

Definition 2.21 (Order)

An **order** is a relation R, usually denoted $\leq$.

1. *Reflexive.* For all $a \in A$, $a \leq a$
2. *Antisymmetric.* For all $a, b \in A$, if $a \leq b$ and $b \leq a$ then $a = b$
3. *Transitive.* For all $a, b, c \in A$, if $a \leq b$ and $b \leq c$ then $a \leq c$

If it has the final property, it is known as a **total order**/**linear order**. Otherwise it is known as a **partial order**.

1. *Total.* For all $a, b \in A$, either $a \leq b$ or $b \leq a$

Given an order $\leq$, we can define the additional symbols to mean

1. $a < b \iff (a \leq b) \wedge (a \neq b)$.
2. $a \geq b \iff \neg(a < b)$.
3. $a > b \iff \neg(a \leq b)$.

For convenience of notation, we also write $a < x < b \iff (a < x) \wedge (x < b)$.

Definition 2.22 (Interval)

Given a totally ordered set X, we denote the **intervals** as

1. $(a, b) \coloneqq \{x \in X \mid a < x < b\}$
2. $[a, b) \coloneqq \{x \in X \mid a \leq x < b\}$
3. $(a, b] \coloneqq \{x \in X \mid a < x \leq b\}$
4. $[a, b] \coloneqq \{x \in X \mid a \leq x \leq b\}$

Definition 2.23 (Bounds)

Let X be a set and $Y \subset X$. Then

1. z is an **upper bound** of Y if $\forall y \in Y (y \leq z)$.
2. z is a **lower bound** of Y if $\forall y \in Y (z \leq y)$.

2.4 Axiom of Choice

The axioms up to this point are pretty much undisputed and completes ZF set theory. Now that we've defined a function, let's quickly extend the definition of a Cartesian product into an arbitrary union of sets.

Definition 2.24 (Cartesian Product)

If $\{X_\alpha\}_{\alpha \in A}$ is an indexed family of sets, then their **Cartesian product** is defined as a set of functions. That is,

$$\prod_{\alpha \in A} X_\alpha \coloneqq \left\{ f : A \to \bigcup_{\alpha \in A} X_\alpha \;\middle|\; \forall \alpha \in A, f(\alpha) \in X_\alpha \right\} \tag{30}$$

Each function f is called a **choice function**, which assigns to each X_α some element $f(\alpha) \in X_\alpha$.

Therefore, we have used the power set axiom to define a finite Cartesian product, to then define a function, to then define a general Cartesian product. But this detail is irrelevant later on. Note also that this definition of Cartesian product is not the same as that of the previous definition. The binary Cartesian product is defined as $(a, b) = \{\{a\}, \{a, b\}\}$ while this defines as a function $f : \{1, 2\} \to A, B$. But once we have overwritten the old definition (which is still necessary!) we can just forget about it and use this new definition of Cartesian product since there is a canonical bijection between them. It is a lot less annoying to think of ordered tuples as just tuples rather than as sets of sets.

However, in our definition, we just call this a "set of functions" and have never proved that it actually contains anything. But we can see obviously that if this Cartesian product is nonempty then there exists a choice function, and if there exists a choice function then the Cartesian product is nonempty. It would be ideal if we can prove one of the two conditions, but it turns out we can't, and therefore we introduce the final axiom, called the *axiom of choice*. Though controversial, it is required in the proofs of some notable theorems. If we include this axiom of choice, then we have ZFC set theory. The axiom of choice has many equivalent definitions.

Colloquially, the axiom of choice says that a Cartesian product of a collection[2] of non-empty sets is non-empty. That is, it is possible to construct a new set by choosing one element from each set, even if the collection is infinite.

Axiom 2.10 (Axiom of Choice)

Let us have an indexed family $X = \{S_i\}_{i \in I}$ of nonempty sets. Then the axiom states the following, which are all equivalent.

1. There exists an indexed set $\{x_i\}_{i \in I}$ such that $x_i \in S_i$ for every $i \in I$.
2. $\prod_{i \in I} S_i$ is nonempty.
3. There exists a choice function $f : I \to \cup_{i \in I} S_i$.

The existence of a choice function when X is finite is easily proved from the ZF axioms, and AC only matters for certain infinite sets. One may argue that if each S_i is nonempty, then choose $s_i \in S_i$ and you're done! While this is an intuition for why the axiom of choice may be true, we can't make the *choice* of all the infinitely many s_i in any "canonical" fashion. That is,k while this works for any single i at a time, this doesn't define a function $i \mapsto s_i$. Note that for any sets where you *can* make this choice (e.g. there is a total ordering on X, so choose the minimum element), AC holds as a theorem and not as an axiom.

Example 2.1 ()

Let I be the set of all nonempty subsets of $\mathbb{R}$, and $X_i = i \in I$. Then an element f in $\prod_{i \in I} X_i$ is a function which picks an element $f(T) \in T$ for every nonempty $T \subset \mathbb{R}$. How do you *define* such an f? If we have $\mathbb{N}$ instead of $\mathbb{R}$, we could take $f(T) = \min(T)$, but this doesn't work for $\mathbb{R}$. Therefore, there is no canonical choice of an element in a nonempty set of real numbers. But AC tells us that we don't have to worry about this. It gives us such a function, even if we cannot "write it down" (which means, construct it from the other ZF axioms).
If we let I be the set of all nonempty *open* subsets of $\mathbb{R}$, then there is a choice function. Choose

[2]Note that this does not have to be finite

any bijection $\tau : \mathbb{N} \to \mathbb{Q}$, and then assign to each nonempty open subset $U \subset \mathbb{R}$ the element $\tau(\min\{n \in \mathbb{N} \mid \tau(n) \in U\})$. This works since $U \cap \mathbb{Q} \neq \emptyset$.

It is characterized as nonconstructive because it asserts the existence of a choice function but says nothing about how to construct one, unlike the axiom of infinity. This choice function was used in the proof of the following, which turns out to be equivalent.

Axiom 2.11 (Axiom of Well-Ordering)

For any set X, there exists a binary relation R which *well-orders* X, i.e. is a total order and has the property that every nonempty subset of X has a least element under the order R.

$$\forall X \exists R(R \text{ well-orders } X) \tag{31}$$

We can see generally that we would like to use a choice function to select a representative element of each set in X. Then we can use these to construct an order. Finally, we state the last form of the axiom of choice.

Axiom 2.12 (Zorn's Lemma)

Let X be a partially ordered set that satisfies the two properties.
1. P is nonempty.
2. Every *chain* (a subset $A \subset P$ where A is totally ordered) has an upper bound in P.

Then P has at least one maximal element.

Zorn's lemma is required to show that every vector space has a basis.

2.5 Rules of Set Theory

Let's first talk about rules following the union, intersection, and set minus operators.

Theorem 2.6 (deMorgan's Laws)

The following hold for sets X, Y, Z.

$$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z) \tag{32}$$
$$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z) \tag{33}$$
$$X \setminus (Y \cup Z) = (X \setminus Y) \cap (X \setminus Z) \tag{34}$$
$$X \setminus (Y \cap Z) = (X \setminus Y) \cup (X \setminus Z) \tag{35}$$

Proof.

Theorem 2.7 (Symmetric Difference)

Given sets X, Y,

$$(X \setminus Y) \cap (Y \setminus X) = (X \cup Y) - (X \cap Y) \tag{36}$$

Proof.

Now let's see how these operations behavior under functions.

Theorem 2.8 (Preservation Under Mapping Back and Forth)

Given $f : A \to B$, with $A_0, A_1 \subset A$ and $B_0, B_1 \subset B$, the following hold
1. $A_0 \subset f^{-1}(f(A_0))$, with equality holding if f is injective.
2. $f(f^{-1}(B_0)) \subset B_0$, with equality holding if f is surjective.

Proof.

Listed.
1. Assume that $x \in A_0$. Then $f(x) \in f(A_0)$. The preimage is
$$f^{-1}(f(A_0)) \coloneqq \{y \in A \mid f(y) \in f(A_0)\} \tag{37}$$
and x certainly satisfies the condition that $f(x) \in f(A_0)$. Therefore $x \in f^{-1}(f(A_0))$ and so $A_0 \subset f^{-1}(f(A_0))$.
Now assume that f is injective. It suffices to prove that $f^{-1}(f(A_0)) \subset A_0$ since the other direction is proven for all functions. We prove this by proving the contrapositive, i.e. $x \notin A_0 \implies x \notin f^{-1}(f(A_0))$. Suppose $x \notin A_0 \implies f(x) \notin f(A_0) \implies f^{-1}(f(x)) \not\subset f^{-1}(f(A_0))$ by definition of the image and preimage. But note that since f is injective, $f^{-1}(f(x)) = x$.[a] and thus $x \notin f^{-1}(f(A_0))$.
2. We prove this using the contrapositive. Assume that $x \notin B_0$. Then, with abuse of notation, we have by definition of the preimage and the image $f^{-1}(x) \not\subset f^{-1}(B_0) \implies f(f^{-1}(x)) \not\subset f(f^{-1}(B_0))$. But $f(f^{-1}(x)) = \{x\}$, since we are just mapping the preimage of x back across to f. Therefore, $x \notin f(f^{-1}(B_0))$.
Now assume that f is surjective. It suffices to prove that $B_0 \subset f(f^{-1}(B_0))$. Assume $y \in B_0$. Since f is surjective, we know that $f^{-1}(y)$ is nonempty in A. We can state $f^{-1}(y) \subset f^{-1}(B_0)$[b] which then implies $f(f^{-1}(y)) \subset f(f^{-1}(B_0))$.[c] But $f(f^{-1}(y)) = y$ as mentioned previously, and so $y \in f(f^{-1}(B_0))$.

Example 2.2 ()

To see why equality does not hold in general for the two cases, look at the counterexamples below.
1. $A_0 \not\supset f^{-1}(f(A_0))$.
2. $f(f^{-1}(B_0)) \not\supset B_0$. Consider $X = Y = \{0, 1\}$ and $f : X \to Y$ defined $f(0) = f(1) = 0$. Consider $C = Y$. We have $f^{-1}(C) = f^{-1}(0) \cup f^{-1}(1) = X \cup \emptyset = X$. Then $f(f^{-1}(C)) = f(X) = \{0\} \neq C$.

Theorem 2.9 (Preservation Under Preimages)

Given $f : A \to B$, with $A_0, A_1 \subset A$ and $B_0, B_1 \subset B$, f preserves the inclusion, union, intersection, and set difference under the preimage.
1. *Inclusion.* $B_0 \subset B_1 \implies f^{-1}(B_0) \subset f^{-1}(B_1)$.
2. *Union.* $f^{-1}(B_0 \cup B_1) = f^{-1}(B_0) \cup f^{-1}(B_1)$.
3. *Intersection.* $f^{-1}(B_0 \cap B_1) = f^{-1}(B_0) \cap f^{-1}(B_1)$.
4. *Set Difference.* $f^{-1}(B_0 \setminus B_1) = f^{-1}(B_0) \setminus f^{-1}(B_1)$.

[a] More specifically, if we treat x as the singleton set, $f(x)$ is also a singleton set by definition of a function. Since f is injective, the preimage of a singleton set must be a singleton set. If it were not, then there exists x, y with $x \neq y$ that maps to the same z, which contradicts the definition of injectivity.
[b] The formal proof of this is given in Munkres 1.2.2.a.
[c] Again formal proof of this given in Munkres 1.2.2.e.

Proof.

Listed.

1. *Inclusion.* If $x \in B_0$, then $f^{-1}(x) \subset A$ maps to x by definition. But since $x \in B_0$, $f^{-1}(x)$ maps to a point in B_0, and so $f^{-1}(x) \subset f^{-1}(B_0)$. Since $B_0 \subset B_1$ by assumption, $x \in B_1$, and by the previous logic but with B_0 replaced by B_1 we have $f^{-1}(x) \subset f^{-1}(B_1)$. We have just proved that $f^{-1}(x) \in f^{-1}(B_0) \implies f^{-1}(x) \in f^{-1}(B_1)$, and so $f^{-1}(B_0) \subset f^{-1}(B_1)$.
2. *Union.* We prove bidirectionally.
   (a) $f^{-1}(B_0 \cup B_1) \subset f^{-1}(B_0) \cup f^{-1}(B_1)$. Let $x \in f^{-1}(B_0 \cup B_1)$ which by definition of the preimage means $f(x) \in B_0 \cup B_1$. Therefore $f(x) \in B_0$ or B_1. Without loss of generality, let $f(x) \in B_0$. Then we have
   $$x \in f^{-1}(f(x)) \subset f^{-1}(B_0) \tag{38}$$
   where the first inclusion comes from [Munkres 1.2.1.a] when treating $A_0 = \{x\}$, and the second subset comes from [Munkres 1.2.2.a] when treating $B_0 = \{f(x)\}, B_1 = B_1$. Therefore $x \in f^{-1}(B_0) \subset f^{-1}(B_0) \cup f^{-1}(B_1)$.
   (b) $f^{-1}(B_0) \cup f^{-1}(B_1) \subset f^{-1}(B_0 \cup B_1)$. Let $x \in f^{-1}(B_0) \cup f^{-1}(B_1)$. Without loss of generality, let $x \in f^{-1}(B_0)$ which by definition of the preimage implies $f(x) \in B_0 \subset (B_0 \cup B_1) \implies f(x) \in (B_0 \cup B_1)$. Therefore, we have
   $$x \in f^{-1}(f(x)) \subset f^{-1}(B_0 \cup B_1) \tag{39}$$
   where the inclusion claim comes from [Munkres 1.2.1.a] when treating $A_0 = \{x\}$, and the subset claim comes from [Munkres 1.2.2.a] when treating $B_0 = \{f(x)\}, B_1 = B_0 \cup B_1$. Therefore $x \in f^{-1}(B_0 \cup B_1)$.

   Therefore, $f^{-1}(B_0) \cup f^{-1}(B_1) = f^{-1}(B_0 \cup B_1)$.
3. *Intersection.* We prove bidirectionally.
   (a) $f^{-1}(B_0 \cap B_1) \subset f^{-1}(B_0) \cap f^{-1}(B_1)$. Assume $x \in f^{-1}(B_0 \cap B_1)$, which by definition of the preimage means $f(x) \in B_0 \cap B_1$. So
   $$f(x) \in B_0 \implies x \in f^{-1}(f(x)) \subset f^{-1}(B_0) \tag{40}$$
   $$f(x) \in B_1 \implies x \in f^{-1}(f(x)) \subset f^{-1}(B_1) \tag{41}$$
   where the inclusion claim comes from [Munkres 1.2.1.a] when treating $A_0 = \{x\}$, and the subset claim comes from [Munkres 1.2.2.a] when treating $f(x)$ as a singleton set. Therefore x is in both of the preimages and so $x \in f^{-1}(B_0) \cap f^{-1}(B_1)$.
   (b) $f^{-1}(B_0) \cap f^{-1}(B_1) \subset f^{-1}(B_0 \cap B_1)$. Let $x \in f^{-1}(B_0) \cap f^{-1}(B_1)$. Then by definition of intersection and preimage,
   $$x \in f^{-1}(B_0) \implies f(x) \in B_0 \tag{42}$$
   $$x \in f^{-1}(B_1) \implies f(x) \in B_1 \tag{43}$$
   and so $f(x) \in B_0 \cap B_1$ by definition of intersection. This means by definition of the preimage that $x \in f^{-1}(B_0 \cap B_1)$.
4. *Set Difference.* We prove bidirectionally.
   (a) $f^{-1}(B_0 \setminus B_1) \subset f^{-1}(B_0) \setminus f^{-1}(B_1)$. Let $x \in f^{-1}(B_0 \setminus B_1)$ which by definition of preimage means $f(x) \in B|0 \setminus B_1$. This implies two things. First,
   $$f(x) \in B_0 \implies x \in f^{-1}(f(x)) \subset f^{-1}(B_0) \tag{44}$$
   where the inclusion comes from [Munkres 1.2.1.a] when treating $A_0 = \{x\}$ as the single set, and the subset claim comes from [Munkres 1.2.2.a] stating that inclusions are preserved under the preimage operator. Secondly, we claim that
   $$f(x) \notin B_1 \implies x \notin f^{-1}(B_1) \tag{45}$$
   since if $x \in f^{-1}(B_1)$, then $f(x) \in B_1$ by definition of the preimage.

(b) $f^{-1}(B_0) \setminus f^{-1}(B_1) \subset f^{-1}(B_0 \setminus B_1)$. Let $x \in f^{-1}(B_0) \setminus f^{-1}(B_1)$. Then the following holds

$$x \in f^{-1}(B_0) \implies f(x) \in B_0 \tag{46}$$
$$x \notin f^{-1}(B_1) \implies f(x) \notin B_1 \tag{47}$$

from the definition of the preimage and the contrapositive of its implication. Therefore $f(x) \in B_0 \setminus B_1$ which by definition of the preimage $x \in f^{-1}(B_0 \setminus B_1)$.

Theorem 2.10 (Preservation Under Images)

Given $f : A \to B$, with $A_0, A_1 \subset A$ and $B_0, B_1 \subset B$, f preserves the inclusion and union under the image, but inclusion properties for the intersection and set difference hold.

1. *Inclusion.* $A_0 \subset A_1 \implies f(A_0) \subset f(A_1)$.
2. *Union.* $f(A_0 \cup A_1) = f(A_0) \cup f(A_1)$.
3. *Intersection.* $f(A_0 \cap A_1) \subset f(A_0) \cap f(A_1)$, and equality holds if f is injective.
4. *Set Difference.* $f(A_0 \setminus A_1) \supset f(A_0) \setminus f(A_1)$, and equality holds if f is injective.

Proof.

Listed.

1. *Inclusion.* Let $x \in A_0$. Then by definition of the image $f(x) \in f(A_0)$. Since $A_0 \subset A_1$, then $x \in A_1$ and it immediately follows that $f(x) \in f(A_1)$. Therefore $f(A_0) \subset f(A_1)$.
2. *Union.* We prove bidirectionally.
   (a) $f(A_0 \cup A_1) \subset f(A_0) \cup f(A_1)$. Let $y \in f(A_0 \cup A_1)$. Then by definition there exists some $x \in A_0 \cup A_1$ s.t. $f(x) = y$. WLOG let $x \in A_0$. Then by definition $y = f(x) \in f(A_0) \subset f(A_0) \cup f(A_1)$.
   (b) $f(A_0) \cup f(A_1) \subset f(A_0 \cup A_1)$. Let $y \in f(A_0) \cup f(A_1)$. WLOG $y \in f(A_0)$, and there exists some $x \in A_0$ s.t. $f(x) = y$. Since $x \in A_0$, $x \in A_0 \cup A_1$, and by definition $y = f(x) = f(A_0) \cup f(A_1)$.
3. *Intersection.* Assume that $y \in f(A_0 \cap A_1)$. Then by definition there exists some $x \in A_0 \cap A_1$ s.t. $f(x) = y$. So we have

$$x \in A_0 \implies f(x) \in f(A_0) \tag{48}$$
$$x \in A_1 \implies f(x) \in f(A_1) \tag{49}$$

   and therefore $y = f(x) \in f(A_0) \cap f(A_1)$.
   To prove equality, it suffices to show that $f(A_0) \cap f(A_1) \subset f(A_0 \cap A_1)$ if f is injective. Assume that $y \in f(A_0) \cap f(A_1)$. Then $y \in f(A_0)$, and so there exists an $x \in A_0$ s.t. $y = f(x) \in f(A_0)$. By the same logic there exists an $x' \in A_1$ s.t. $y = f(x') \in f(A_1)$. But since f is injective, this implies that $x = x'$. So $x \in A_0 \cap A_1$, and so $y = f(x) \in f(A_0 \cap A_1)$.
4. *Set Difference.* Assume that $y \in f(A_0) \setminus f(A_1)$. Since $y \in f(A_0)$, there exists some $x \in A_0$ s.t. $y = f(x)$. Since $y \notin f(A_1)$, there exists no $x' \in A_1$ s.t. $y = f(x')$. Therefore, $x \in A_0 \setminus A_1 \implies y = f(x) \in f(A_0 \setminus A_1)$.
   To prove equality, it suffices to show that $f(A_0 \setminus A_1) \subset f(A_0) \setminus f(A_1)$ if f is injective. Assume that $y \in f(A_0 \setminus A_1)$. Then there exists some $x \in A_0 \setminus A_1$ s.t. $f(x) = y$. We claim that x is unique since if there were two x, x', then $f(x) = f(x')$ with $x \neq x'$, which means f is not injective. We see that $x \in A_0 \implies y = f(x) \in f(A_0)$, and $x \notin A_1 \implies y = f(x) \notin f(A_1)$. Therefore, $x \in f(A_0) \setminus f(A_1)$.

Example 2.3 (Intersection Not Necessarily Preserved)

Note that intersection is not necessarily preserved. To see why, look at the counterexample. Consider $A = \{0, 1\}, B = \{1, 2\}$, and define

$$f(0) = f(2) = 0, f(1) = 1 \tag{50}$$

Then $f(A) = f(B) = \{0, 1\} \implies f(A) \cap f(B) = \{0, 1\}$. On the other hand, we have $A \cap B = \{1\} \implies f(A \cap B) = \{1\}$.

Theorem 2.11 (Composition)

Let $f : X \to Y$ and $g : Y \to Z$.

1. f injective and g injective $\implies$ $g \circ f$ injective.
2. f surjective and g surjective $\implies$ $g \circ f$ surjective.
3. f bijective and g bijective $\implies$ $g \circ f$ bijective.

Theorem 2.12 (Injectivity/Surjectivity)

Let $f : X \to Y$, $g : Y \to Z$, and $h = g \circ f$. The following hold:

1. h injective $\implies$ f injective.
2. h surjective $\implies$ g surjective.
3. h bijective $\implies$ f injective and g bijective.

Corollary 2.1 (Bijection Equals Existence of Inverse)

$f : X \to Y$ has a inverse function $f^{-1} : B \to A$ iff it is bijective.

Corollary 2.2 (Decomposition)

Any function $h : X \to Y$ can be decomposed to the form $h = g \circ f$, where f is injective and g is surjective.

Proof.

Given X, let us define an equivalence class where for any $x, y \in X$, $x \sim y$ iff $f(x) = f(y)$. Call this quotient space $X/\sim$. Then we can define the mappings.

1. $\iota : X \to X/\sim$ which maps each element to its equivalence class. $\iota(x) = [x]$
2. $f' : X/\sim \to Y$ which maps each class to the element of Y that it maps to. $f'([x]) = f(x)$.

ι is surjective since for every $[x] \in X/\sim$, there exists at least one element $x \in X$ that maps to it. f' is injective since have squished all the points x that map to the same y into a single class $[x]$.

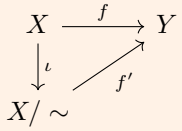

Figure 5: Decomposition of f into surjective ι and injective f'.

Theorem 2.13 (Inverse of Inverses)

If f is bijective, then $f = (f^{-1})^{-1}$.

Theorem 2.14 (Finite Set Mappings)

Suppose X and Y are finite sets, each with n elements, and $f : X \to Y$. If f is injective or bijective, then f is bijective.

Theorem 2.15 (Inverse of Compositions)

If f, g are both bijective, then

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1} \tag{51}$$

2.6 Cardinality

Definition 2.25 (Cardinality)

The **cardinality**, or the **cardinal number**, of a set X is the number of elements in X.

Definition 2.26 (Equipotence)

Two sets A and B are **equipotent**, written $A \approx B$, if there exists a bijective map $f : A \to B$. This implies that their cardinalities are the same: $|A| = |B|$. It has the following properties:

1. Reflexive: $A \approx A$
2. Symmetric: $A \approx B$ implies $B \approx A$
3. Transitive: $A \approx B$ and $B \approx C$ implies $A \approx C$

Definition 2.27 (Classes of Cardinal Numbers)

For any positive integer n, let J_n be the set whose elements are the integers $1, 2, \dots, n$. For any set A, we define

1. A is **finite** if $A \approx J_n$ for some n. The empty set is also considered to be finite.
2. A is **infinite** if it is not finite.
3. A is countable if $A \approx \mathbb{N}$.
4. A is uncountable if A is neither finite nor countable.
5. A is at most countable if A is finite or countable.

At this point, we may already be familiar with the fact that $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable. Let us formalize the statement that a countable infinity is the smallest type of infinity. We can show this by taking a countable set and showing that every infinite subset must be countable. If it was uncountable, then this would mean that a countable set contains an uncountable set.

Theorem 2.16 ()

Every infinite subset of a countable set A is countable.

Theorem 2.17 ()

An at most countable union of countable sets is countable.

Theorem 2.18 ()

A finite Cartesian product of countable sets is countable.

Now, how do we prove that a set is uncountable? We can't really use the contrapositive of Theorem 2.6, since to prove that an arbitrary set A is uncountable, then we must find an infinite subset that is not countable. But now we must prove that this subset itself is not countable, too! Therefore, we can use this theorem.

Theorem 2.19 ()

Given an arbitrary set A, if every countable subset B is a proper subset of A, then A is uncountable.

Proof.

Assume that A is countable. Then A itself is a countable subset of A, but by the assumption, A should be a proper subset of A, which is absurd. Therefore, A is uncountable.

2.7 Exercises

Exercise 2.1 (Math 531 Spring 2025 PS1.2)

Exercise 2.2 (Shifrin Abstract Algebra Appendix 2.3)

Let $f : X \to Y$. Let $A, B \subset X$ and $C, D \subset Y$. Prove or give a counterexample (if possible, provide sufficient hypotheses for each statement to be valid):

1. $f(A) \cup f(B) = f(A \cup B)$
2. $f(A) \cap f(B) = f(A \cap B)$
3. $f(A - B) = f(A) - f(B)$
4. $f^{-1}(C) \cup f^{-1}(D) = f^{-1}(C \cup D)$
5. $f^{-1}(C) \cap f^{-1}(D) = f^{-1}(C \cap D)$
6. $f^{-1}(C - D) = f^{-1}(C) - f^{-1}(D)$
7. $f(f^{-1}(C)) = C$
8. $f^{-1}(f(A)) = A$

Solution 2.1

Listed.

1.

Exercise 2.3 (Munkres Topology 1.5.5)

Which of the following subset of $\mathbb{R}^\omega$ can be expressed as the Cartesian product of subsets of $\mathbb{R}$?[a]

Solution 2.2

Listed. We will denote the sets in question as A.

1. We claim that
$$A = \mathbb{Z} \times \mathbb{Z} \times \dots \tag{52}$$
2. Let us denote $\mathbb{R}_{\geq i}$ be the set of reals greater than or equal to i. This is clearly a subset of $\mathbb{R}$.

[a] Note that the existence of these sets depend on the axiom of choice.

Then

$$A = \prod_{i=1}^{\infty} \mathbb{R}_{\geq i} \tag{53}$$

3. We claim

$$A = \left(\prod_{i=1}^{100} \mathbb{R} \right) \times \left(\prod_{j=1}^{\infty} \mathbb{Z} \right) \tag{54}$$

4. This is not possible

3 Numbers Systems

3.1 Natural Numbers

Note that the axiom of infinity allows us to construct the ordinal numbers. This was based off of two things. First the assertion that the empty set is contained and that if a set w in the ordinal numbers, then $w \cup \{w\}$ is also contained. The second property has a name.

Definition 3.1 (Inductive Set)

Let X be a set. Then X is said to be **inductive** if every element has a **successor**, i.e. a construction of a different element y from x.

$$x \in X \implies f(x) \in X \tag{55}$$

A set $X \subset \mathbb{R}$ is inductive if for each number $x \in X$, it also contains $x + 1$.

From this, with a few more structures we can define the naturals.

Definition 3.2 (Natural Numbers)

The natural numbers $\mathbb{N}$ is the set of von Neumann ordinals, with each element represented by a numerical symbol (e.g. $1, 2, \ldots$).

$$\begin{aligned} 0 &= \{\} = \emptyset \\ 1 &= \{0\} = \{\emptyset\} \\ 2 &= \{0, 1\} = \{\emptyset, \{\emptyset\}\} \\ 3 &= \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \\ 4 &= \{0, 1, 2, 3\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\} \\ \ldots &= \ldots \end{aligned}$$

The successor function $S(x)$ is rewritten in different notation as $S(x) = x + 1$. It also has the relation $\leq$ defined as the set

$$\{(a, b) \in \mathbb{N} \times \mathbb{N} \mid \exists n (S^n(a) = b)\} \tag{56}$$

where S^n is the successor operation composed n times. Addition is defined as

$$a + b \coloneqq S^b(a) = S^a(b) \tag{57}$$

and multiplication is defined recursively as

1. $m \times 0 \coloneqq 0$.
2. $m \times (n + 1) \coloneqq m \times n + m$

which is familiar to the process of adding m to itself n times.

Lemma 3.1 (Well Ordering Principle of Naturals)

Every nonempty subset of $\mathbb{N}$ has a minimal element.

Proof.

Take a subset $A \subset \mathbb{N}$.
1. If $0 \in A$, the minimum is 0.
2. Else if $1 \in A$, the minimum is 1.
3. ...

We can use this inductive property of natural numbers to prove properties of them. Note that this can only be used to prove for finite (yet unbounded) numbers!

Lemma 3.2 (Induction Principle)

Given $P(n)$, a property depending on a natural number $n \in \mathbb{N}$,
1. if $P(n_0)$ is true for some $n_0 \in \mathbb{N}$, and
2. if for every $k \geq n_0$, $P(k)$ true implies $P(k+1)$ true,

then $P(n)$ is true for all $n \geq n_0$.

Lemma 3.3 (Strong Induction Principle)

Given $P(n)$, a property depending on a positive integer n,
1. if $P(n_0), P(n_0+1), \dots, P(n_0+m)$ are true for some positive integer n_0, and nonnegative integer m, and
2. if for every $k > n_0 + m, P(j)$ is true for all $n_0 \leq j \leq k$ implies $P(k)$ is true,

then $P(n)$ is true for all $n \geq n_0$.

Theorem 3.1 (Equivalence of 3 Principles)

The well-ordering principle, induction principle, and the strong induction principle are all equivalent.

Proof.

We prove the steps.
1. *Well Ordering* $\implies$ *Strong Induction.*
2. *Strong Induction* $\implies$ *Induction.*
3. *Induction* $\implies$ *Well-Ordering.*

The idea behind the strong induction principle leads to the proof using infinite descent. Infinite descent combines strong induction with the fact that every subset of the positive integers has a smallest element, i.e. there is no strictly decreasing infinite sequence of positive integers.

Theorem 3.2 (Infinite Descent)

Given $P(n)$, a property depending on positive integer, assume that $P(n)$ is false for a set of integers $\mathcal{S}$. Let the smallest element of $\mathcal{S}$ be n_0. If $P(n_0)$ false implies $P(k)$ false, where $k < n_0$, then by contradiction $P(n)$ is true for all n.

3.2 Integers

Theorem 3.3 (Countability)

$\mathbb{Z}$ is countable.

3.3 Rational Numbers

Theorem 3.4 (Rational Numbers)

$\mathbb{Q}$ is countable.

Proof.

Since $\mathbb{N} \approx \mathbb{Z}$, it suffices to prove that $\mathbb{N} \times \mathbb{N}$ is countable. We wish to find the bijection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. We claim that

$$f(x,y) = \frac{1}{2}\{(x+y-1)^2 - (x+y-1) + 2\} + x - 1 \tag{58}$$

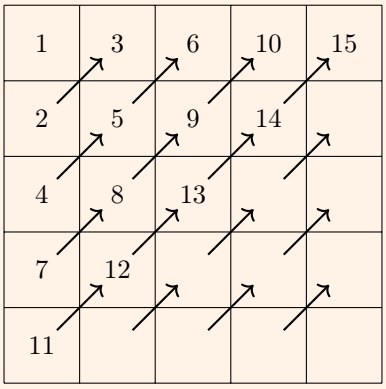

Figure 6: You can see that given (x,y) it is on the $(x+y+1)$th diagonal, which starts from the $\frac{1}{2}\big((x+y+1)^2 - (x+y+1) + 2\big)$th number and increments by $x-1$. Therefore, we have the formula above.

3.4 Real Numbers

The construction of the real numbers is done in my real analysis notes. We will prove the uncountability of the reals assuming you know the construction.

Theorem 3.5 (Uncountability of Infinite Binary Numbers)

Let A be the set of all sequences whose elements are the digits 0 and 1. Then, A is uncountable.

We also know that $\mathbb{R}$ is equipotent to A, and so the corollary follows.

Corollary 3.1 (Uncountability of Reals)

$\mathbb{R}$ is uncountable.

3.5 Exercises

Exercise 3.1 (Math 531 Spring 2025, PS1.1)

Find a formula for the sum of the first n odd numbers and prove that it is correct.

Solution 3.1

I claim that $f(n) = n^2$. I prove using induction. For $n = 1$, $f(1) = n^2 = 1^2 = 1$. Now assume f holds for some $k \in \mathbb{N}$. Then, the kth off number is $2k - 1$. Therefore

$$f(k+1) = f(k) + (2k+1) = k^2 + 2k + 1 = (k+1)^2 \tag{59}$$

and the formula holds for $k = 1$. By the principle of induction, $f(n) = n^2$ is true for all $n \in \mathbb{N}$.

Exercise 3.2 (Shifrin Abstract Algebra 1.1.4.C)

We check for $n = 1$ denoting our formula as f. Indeed, we have

$$f(1) = \frac{1 \cdot 2 \cdot 3}{6} = 1 = 1^2 \tag{60}$$

For the induction step, assume that $f(k)$ is true for some $k \in \mathbb{N}$. Then,

$$\begin{aligned} f(k+1) &= f(k) + (k+1)^2 && (61)\\ &= \frac{k(k+1)(2k+1)}{6} + \frac{6(k+1)^2}{6} && (62)\\ &= \frac{(k+1)\{k(2k+1) + 6(k+1)\}}{6} && (63)\\ &= \frac{(k+1)(2k^2+7k+6)}{6} && (64)\\ &= \frac{(k+1)(k+2)(2(k+1)+1)}{6} && (65)\\ &= f(k+1) && (66) \end{aligned}$$

Therefore f holds for all $n \in \mathbb{N}$.

Exercise 3.3 (Shifrin Abstract Algebra 1.1.4.G)

We prove the base cases for $n = 1, 2, 3$.

1. $n = 1$. $n + 2 = 3$ is divisible by 3.
2. $n = 2$. $n + 4 = 6$ is divisible by 3.
3. $n = 3$. $n + 2 = 3$ is divisible by 3.

For our inductive step, assume that for some $n = k \in \mathbb{N}$, one of the elements in $S_k = \{k, k+2, k+4\}$ is divisible by 3. Let us denote this element a. We wish to show that this claim is true for $n = k+3$ on the set $S_{k+3} = \{k+3, k+5, k+7\}$. Since $a \in S_k$, this means that $a+3 \in S_{k+3}$, and $3|a \implies 3|(a+3)$. So we can always identify the element $a+3$. Since we proved the base cases for $n = 1, 2, 3$, and proved the recursive step, we have essentially proved the claim for all naturals of the form $3k+1, 3k+3, 3k+3$ ($k \in \mathbb{N}_0$), which is precisely the natural numbers.

Exercise 3.4 (Shifrin Abstract Algebra 1.1.4.J)

Let $n = 1$. Then $1 + x \geq 1 + x$ trivially. For the induction step, assume that this inequality holds for some $n \in \mathbb{N}$. Then, we have

$$\begin{aligned} 1 + (n+1)x &= 1 + nx + x && (67)\\ &\leq (1+x)^n + x && (68)\\ &\leq (1+x)^n + x(1+x)^n && (69)\\ &= (1+x)^{n+1} && (70) \end{aligned}$$

where the prove the penultimate step by applying the ordered field axioms to the 2 cases:

1. If $x \geq 0$, then addition preserves order so $1+x \geq 0+1 = 1$. Since $1+x, 1 > 0$, order is preserved under multiplication by a positive element, so $(1+x)^2 \geq 1 + x \geq 1$. Using induction, we can show that for all $n \in \mathbb{N}$, $(1+x)^n \geq 1$, and again by preservation of order under multiplication by a positive element, this implies $x(1+x)^n \geq x$ for all $n \in \mathbb{N}$.
2. If $0 > x > -1$, we have $0 < 1 + x < 1$ and by the same induction proof, we can bound $0 < (1+x)^n < 1$ for all n. Finally by reversal of order under multiplication by a negative element, we have $x(1+x)^n > x$.

Therefore, we take the less restrictive of the 2 bounds: $x(1+x)^n \geq x$.

Exercise 3.5 (Shifrin Abstract Algebra 1.1.7)

Let us denote

$$x = \frac{1+\sqrt{5}}{2}, \quad y = \frac{1-\sqrt{5}}{2} \tag{71}$$

Note the identities

$$x^2 = \left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{3+2\sqrt{5}}{2} = 1 + x \tag{72}$$

$$y^2 = \left(\frac{1-\sqrt{5}}{2}\right)^2 = \frac{3-2\sqrt{5}}{2} = 1 + y \tag{73}$$

We check the base case for $n = 1$

$$a_1 = \frac{1}{\sqrt{5}}(x+y) = \frac{1}{\sqrt{5}}\frac{2\sqrt{5}}{2} = 1 \tag{74}$$

and for $n = 2$

$$a_2 = \frac{1}{\sqrt{5}}(x^2 - y^2) = \frac{1}{\sqrt{5}}((1+x) - (1+y)) = \frac{1}{\sqrt{5}}(x+y) = a_1 = 1 \tag{75}$$

For the inductive step, assume that this formula holds for some $k-1, k \in \mathbb{N}$. Then, we have

$$\begin{aligned} a_{k+1} &= a_k + a_{k-1} && (76)\\ &= \frac{1}{\sqrt{5}}(x^{k-1} - y^{k-1}) + \frac{1}{\sqrt{5}}(x^k - y^k) && (77)\\ &= \frac{1}{\sqrt{5}}\left\{x^{k-1}(1+x) - y^{k-1}(1+y)\right\} && (78)\\ &= \frac{1}{\sqrt{5}}(x^{k+1} - y^{k+1}) && (79) \end{aligned}$$

and we are done.