Warwick Mathematics Exchange

MA3H3

Set Theory

2024, April 13th

Desync, aka The Big Ree

Contents

Introduction

Set theory is, among other things, a branch of mathematics just as any other. However, it has also taken on the role of being one of the most popular choices of foundations: almost all mathematics may be embedded within set theory.

In this document, we use the symbol $\subseteq$ for the subset relation, and $\subset$ for the proper subset relation (as opposed to using $\subset$ for subset and $\subsetneq$ for proper subset). However, we occasionally use the symbol $\subsetneqq$ whenever it is important in a proof that the containment is proper, or to otherwise add emphasis when the inequality is important. This symbol is purposefully distinct from the symbol $\subsetneq$ used in the other convention.

Disclaimer: I make *absolutely no guarantee* that this document is complete nor without error. In particular, any content covered exclusively in lectures (if any) will not be recorded here. This document was written during the 2023 academic year, so any changes in the course since then may not be accurately reflected.

Notes on formatting

New terminology will be introduced in *italics* when used for the first time. Named theorems will also be introduced in *italics*. Important points will be **bold**. Common mistakes will be underlined. The latter two classifications are under my interpretation. YMMV.

Content not taught in the course will be outlined in the margins like this. Anything outlined like this is not examinable, but has been included as it may be helpful to know alternative methods to solve problems.

The table of contents above, and any inline references are all hyperlinked for your convenience.

History

First Edition: 2024-03-25*
Current Edition: 2024-04-13

Authors

This document was written by R.J. Kit L., a maths student. I am not otherwise affiliated with the university, and cannot help you with related matters.

Please send me a PM on Discord @Desync#6290, a message in the WMX server, or an email to Warwick.Mathematics.Exchange@gmail.com for any corrections. (If this document somehow manages to persist for more than a few years, these contact details might be out of date, depending on the maintainers. Please check the most recently updated version you can find.)

If you found this guide helpful and want to support me, you can buy me a coffee!

(Direct link for if hyperlinks are not supported on your device/reader: ko-fi.com/desync.)

*Storing dates in big-endian format is clearly the superior option, as sorting dates lexicographically will also sort dates chronologically, which is a property that little and middle-endian date formats do not share. See ISO-8601 for more details. This footnote was made by the computer science gang.

1 Transfinite Iteration

We recall some definitions about the topology on $\mathbb{R}$:

1. A subset $U \subseteq \mathbb{R}$ is *open* if for every point $x \in U$, there exists $\varepsilon > 0$ such that $\mathbb{B}(x,\varepsilon) = (x-\varepsilon,x+\varepsilon) \subseteq U$.
2. A set $F \subseteq \mathbb{R}$ is *closed* if its complement $\mathbb{R} \setminus F$ is open.
3. Equivalently (in metric spaces), a set $F \subseteq \mathbb{R}$ is closed if and only if it contains the limit of every convergent sequence in F. That is, if $(x_n)_{n=1}^{\infty} \subseteq F$ converges to $x \in \mathbb{R}$, then $x \in F$.
4. A point $p \in F$ is *isolated* (in F) if there exists $\varepsilon > 0$ such that $\mathbb{B}(p,\varepsilon) \cap F = \{p\}$, or equivalently, $\mathbb{B}(p,\varepsilon) \cap \big(F \setminus \{p\}\big) = \emptyset$.

Example.

- Any interval with positive measure has no isolated points.
- The entire set $\mathbb{R}$ and the empty set $\emptyset$ has no isolated points.
- The set of rationals $\mathbb{Q}$ has no isolated points.
- The middle-third Cantor set has no isolated points.
- The point contained in a singleton set is isolated.
- Every point of $\mathbb{Z}$ is isolated (take any $\varepsilon < \frac{1}{2}$).
- The point 0 in the set $[-2,-1] \cup \{0\}$ is isolated (take any $\varepsilon < 1$).

We are interested in *removing* the isolated points. In the last example above, removing the isolated points yields the interval [0,1], which has no isolated points.

For any set $F \subseteq \mathbb{R}$, denote by $D(F)$ the *derived set* obtained by removing the isolated points from F, or equivalently, the set of all limit points of F. Note that if F is closed, the derived set $D(F)$ is also closed since isolated points, as singletons, of F are always open in F.

Example.

- $D\big([0,1]\big) = [0,1]$.
- $D\big([0,1] \cup \{2\}\big) = [0,1]$.
- $D\big(\{0\}\big) = \emptyset$.
- $D\big(\{\frac{1}{n} : n \in \mathbb{Z}^+\}\big) = \emptyset$.

The question is, "if we start with a closed set and remove all the isolated points, do we always get a set without isolated points?" Or more precisely, "given a closed set F, is the derived set $D(F)$ always free of isolated points?"

It turns out that the answer is "no" – in removing the isolated points, a point that was not isolated before may then become isolated in the resulting set.

For instance, define the set $X \subset \mathbb{R}$ by

$$X = [-2,-1] \cup \{0\} \cup \left\{\frac{1}{n} : n \in \mathbb{Z}^+\right\}$$

Nothing in the interval is isolated and 0 is not isolated since there are elements arbitrarily close to it, but all the elements $\frac{1}{n}$ are isolated (take $\varepsilon = \frac{n}{4}$ for each point $\frac{1}{n}$). Removing these points yields the derived set

$$D(X) = [-2,-1] \cup \{0\}$$

which has an isolated point, 0. The derived set of this set is then the interval $D\big(D(X)\big) = [-2, -1]$, and from that point onwards, applying the derivation operator leaves the set unchanged, as the set is now free from isolated points. Using superscripts to denote the number of iterations, we have the sequence

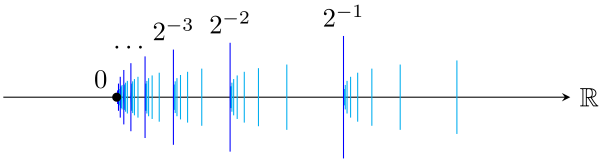
$$\begin{aligned}
D^0(X) &= X \\
D^1(X) &= [-2, -1] \cup \{0\} \\
D^2(X) &= [-2, -1] \\
D^3(X) &= [-2, -1]
\end{aligned}$$

so $D^2(X)$ is the first iteration where the set has no isolated points, after which the sequence stabilises to a fixed point.

In the previous example, X took 2 steps to stabilise because we had a sequence of isolated points that, when removed, produced a new isolated point. We can take inspiration from this to construct a set that takes another iteration to stablise by including sequences of isolated points that tend towards another sequence of isolated points. To avoid collisions, we use a geometric series rather than harmonic.

Define the set $Z \subset \mathbb{R}$ by

$$Z = \{0\} \cup \{2^{-n} : n \in \mathbb{Z}^+\} \cup \{2^{-n} + 2^{-n-m} : n,m \in \mathbb{Z}^+\}$$



The points of the form $2^{-n} + 2^{-n-m}$ are all isolated, and for each fixed n, the sequence $(2^{-n} + 2^{-n-m})$ tends to 2^{-n} as $m \to \infty$, so none of the points 2^{-n} are isolated. Also, the point 0 is not isolated, because the sequence (2^{-n}) tends towards it.

Thus, we have

$$\begin{aligned}
D^0(Z) &= \{0\} \cup \{2^{-n} : n \in \mathbb{Z}^+\} \cup \{2^{-n} + 2^{-n-m} : n,m \in \mathbb{Z}^+\} \\
D^1(Z) &= \{0\} \cup \{2^{-n} : n \in \mathbb{Z}^+\} \\
D^2(Z) &= \{0\} \\
D^3(Z) &= \emptyset \\
D^4(Z) &= \emptyset
\end{aligned}$$

So Z takes three iterations to stabilise.

By adding more and more sequences that converge to the sequences added in the previous step, this construction generalises, and it is possible to construct a set

$$E = \{0\} \cup \bigcup_{i \in \mathbb{Z}^+} \left\{ \sum_{j=1}^{i} 2^{-\sum_{k=1}^{j} n_k} : n_\alpha \in \mathbb{Z}^+ \right\}$$

such that $D^n(E)$ has isolated points for all natural n. This means that even if we remove the isolated points at all steps n, the set

$$D^\omega(E) \coloneqq \bigcap_{n \in \mathbb{N}} D^n(E)$$

still has isolated points. (Note that ω is just notation right now.)

This is just a set, so it makes sense for us to apply the derivation operator again, which we may choose to denote by

$$D^{\omega+1}(E) \coloneqq D\big(D^{\omega}(E)\big)$$

Again, this is still a set, so we may define

$$D^{\omega+2}(E) \coloneqq D\big(D^{\omega+1}(E)\big)$$

and so on, until

$$D^{\omega+\omega}(E) \coloneqq \bigcap_{\alpha=0,1,2,\ldots,\omega,\omega+1,\omega+2,\ldots} D^{\alpha}(E)$$

and we may suggestively define the notation $D^{2\cdot\omega}$ to represent $D^{\omega+\omega}$ more compactly, which suggest a further generalisation.

For any set S, we may then form the sequence of sets

$$D^0(S),D^1(S),\ldots,D^{\omega}(S),D^{\omega+1}(S),\ldots,D^{\overbrace{\omega+\omega}^{\omega\cdot 2}}(S),D^{\omega\cdot 2+1}(S),\ldots,D^{\omega\cdot 3}(S),\ldots,D^{\overbrace{\omega\cdot\omega}^{\omega^2}}(S),\ldots$$

The natural numbers can be used for two distinct purposes: to describe the size or *cardinality* of a set, or to describe the *position* of an element in a sequence, or more precisely, the *order-type* (sometimes called *ordinality*) of an ordered set.

The order-type of an ordered set is the first number not required to label the elements of the set. For instance, the set $\{a,b,c\}$ may be labelled by 0, 1, and 2, so it has ordinality 3. For finite sets, cardinality and order-type coincide.

However, the sequence above is too long for every element to be labelled by a natural number – after all, the natural numbers have all been exhausted by the time we reach $D^{\omega}(S)$ – so its order-type is greater than that of the naturals. On the other hand, the sequence is still countable, so its cardinality is the same as the naturals.

We still have not formally defined what any of these ω symbols mean – only the notation involving derivations that use them – but informally, they are the order-types of sets beyond the natural numbers.

2 The Set-Theoretic Universe

One way of specifying a (finite) set is to list out its members in curly brackets, e.g. $\{a,b,c,d\}$. The symbol $\in$ is called the *membership relation*, indicating that the object to the left is an “element of” or “member of” the set, so $a \in \{a,b,c,d\}$.

- Sets are unordered, so $\{b,a,c,d\} = \{a,b,c,d\}$;
- Elements are unique, so $\{a,a,a,b,b,c,d,d,d\} = \{a,b,c,d\}$.

The most basic set is the empty set, denoted by $\emptyset$, containing no elements. It may also be represented in the manner above, listing its elements as: $\{\}$.

The symbol $\subseteq$ is the *subset relation* between sets: $A \subseteq B$ if and only if every element in A is also an element in B.

Note that there are two ways for a set to be “in” another set: either as an element, or as a subset. To reduce ambiguity, whenever we say that A is “in” B, we mean $A \in B$, while the wording A is *contained* in B means $A \subseteq B$.

2.1 Atoms

Let A denote the collection of objects we want to talk about that are not themselves sets. For instance, the real number $\sqrt{2}$, the natural number 5, or the imaginary unit i, or anything else. Such an element is called an *atom* or *urelement*.

The goal is now to build a hierarchy of sets

$$V_0 \subseteq V_1 \subseteq V_2 \subseteq V_3 \subseteq \cdots$$

such that V_0 is the collection of all sets that can be formed from atoms. That is, an element of V_0 is a subset of A, so $V_0 = \mathcal{P}(A)$ is the power set of A.

Then, V_1 is the collection of all sets whose members are either atoms or sets in V_0. That is, an element of V_1 is a subset of $A \cup V_0$, so $V_1 = \mathcal{P}(A \cup V_0)$. A set containing only atoms is certainly a set containing only atoms *or* elements of V_1, so we also have $V_0 \subseteq V_1$.

We then recursively define

$$V_{n+1} \coloneqq \mathcal{P}(A \cup V_n)$$

and we have $V_n \subseteq V_{n+1}$ by induction.

The empty set is in V_0, so we have $\{\emptyset\} \in V_1$,$\{\{\emptyset\}\} \in V_2$,$\{\{\{\emptyset\}\}\} \in V_3, \ldots$, but the infinite set

$$\{\emptyset,\{\emptyset\},\{\{\emptyset\}\},\{\{\{\emptyset\}\}\},\ldots\}$$

is not in V_n for any natural n, since there will always be an element with at least $n+1$ nested brackets. To remedy this, we may take the infinite union

$$V_\omega \coloneqq \bigcup_{n\in\mathbb{N}} V_n$$

(we still haven't defined ω, but the notation is coming in handy), which immediately extends to

$$\begin{aligned} V_{\omega+1} &\coloneqq \mathcal{P}(A \cup V_\omega) \\ V_{\omega+2} &\coloneqq \mathcal{P}(A \cup V_{\omega+1}) \\ &\vdots \end{aligned}$$

2.2 No Atoms

It turns out that atoms aren't really necessary.

Most higher mathematical objects are defined as sets equipped with certain operations on them, so as long as we can encode relations, and functions as sets, everything else should follow from there.

The previous hierarchy is also simpler without atoms. Because there are no atoms, the only set containing atoms is the empty set, so $V_0 = \{\emptyset\}$. Then, we have $V_1 = \mathcal{P}(V_0)$, and more generally,

$$V_{n+1} \coloneqq \mathcal{P}(V_n)$$

The ωth term, V_ω is the same as before, but the following sets follow the new pattern:

$$\begin{aligned} V_\omega &\coloneqq \bigcup_{n\in\mathbb{N}} V_n \\ V_{\omega+1} &\coloneqq \mathcal{P}(V_\omega) \\ V_{\omega+2} &\coloneqq \mathcal{P}(V_{\omega+1}) \\ &\vdots \end{aligned}$$

3 Unrestricted Comprehension

Rather than explicitly listing out the elements *extensionally*, we often define sets *intensionally* by specifying a property P and collecting all objects x that satisfy that property:

$$X = \{x : P(x)\}$$

However, we need to be careful with what we allow as the property P.

3.1 Frege's Natural Numbers

One intuitive implementation of the natural numbers as pure sets was given by Gottlob Frege. The idea is to define the number 1 as the set of all sets that have exactly one element. This may appear circular (1 vs "one"), but luckily, we can define the predicate $P(S)$ = "S has one element" in first order logic without appealing to a prior notion of "one":

$$P(S) \coloneqq \exists x : \Big(x \in S \wedge \big(\forall y : (y \in S \to x = y)\big)\Big)$$

so the number 1 would be implemented as

$$[[1]] \coloneqq \big\{S : P(S)\big\}$$

Any finite number n can be similarly implemented as the set $[[n]]$ of all sets containing exactly n elements.

This construction seems reasonable. But let us now consider the set $\big\{[[1]]\big\}$. This set has exactly one element, namely $[[1]]$, so we have

$$[[1]] \in \big\{[[1]]\big\} \in [[1]]$$

There is, as yet, no reason that this circularity is incorrect, but it does seem somewhat suspect.

3.2 Universal Sets

Consider the *universal set* U that contains all sets:

$$U \coloneqq \{S : S \text{ is a set}\}$$

Since U is itself a set, it must be an element of U, so we have $U \in U$. This is again concerning.

3.3 Russell's Paradox

Consider the following set, proposed by Bertrand Russell:

$$R \coloneqq \{S : S \notin S\}$$

That is, R is the set of all sets that do not contain themselves. The problems then arise when we ask if $R \in R$ or not. If $R \in R$, then it is a set that does not contain itself, so we must have $R \notin R$. Conversely, if $R \notin R$, then by definition, $R \in R$.

3.4 Unrestricted Comprehension

In the previous examples, we obtained some questionable chains of membership, culminating with the self-contradictory membership of the Russell set. Informally, the problem is that these sets are "too large".

One method of resolving this problem is to switch to type theory. This approach has its own advantages, but the simple solution we will use is to introduce an axiom that restricts how sets are built.

So far, we have been using *unrestricted comprehension* to generate sets by collecting all objects that satisfy any given properties, but as we have seen, this is problematic. The first step is to only allow the collection of objects from an existing set. That is, we cannot form sets

$$\{x : P(x)\}$$

but only

$$\{x \in X : P(x)\}$$

where X is a known set. This already resolve some paradoxes – Russell's paradox included – but there is some ambiguity in what we mean by "property".

Let S be the set of the natural numbers that may be defined in less than 100 characters:

$$S = \{x \in \mathbb{N} : x \text{ can be defined in less than 100 characters}\}$$

So, for instance, "the third natural number", 3, is in S, as is "the 10000th prime number", 104 729, or "the numerator of the 15th coefficient of the Maclaurin series for tan", 689 005 380 505 609 448.

This set is huge. But it is finite: there are finitely many characters we can use, and as such, there are only finitely many strings with fewer than 100 characters. So, there is a smallest natural not in S.

The string,

"n is defined to be the smallest natural number that
cannot be defined in fewer than 100 characters."

that describes the number $n \notin S$ is 99 characters long. So, $n \in S$.

Despite restricting the set comprehension to only collect natural numbers, we still ran into a contradiction – we also need to restrict what qualifies as a "property". We replace "properties" with *formulae* in first order logic.

A formula may contain some or all of the following symbols:

- Logical symbols: $\wedge$, $\vee$, $\to$, $\leftrightarrow$, and $\neg$;
- Quantifiers: $\forall$:, $\exists$: (the : are optional);
- Variable symbols: x_1,x_2,x_3, . . ., or a,b,c,x,y,z,A,B,C, . . .;
- Scoping symbols: (,);
- The equality symbol: =;
- The membership symbol: $\in$.

The syntax of valid formulae is defined recursively. Given a collection of valid symbols as above, the *atomic formulae* are as follows:

- The string $x = y$ is a valid formula for any variables x,y;
- The string $x \in y$ is a valid formula for any variables x,y.

and given two formulae φ and ψ, the following are all valid formulae:

- $(\varphi \wedge \psi)$;
- $(\varphi \vee \psi)$;
- $(\varphi \to \psi)$;
- $(\varphi \leftrightarrow \psi)$;
- $\neg\varphi$;

- $\forall x : \varphi$ for any variable x;
- $\exists x : \varphi$ for any variable x.

The inclusion of brackets ensures that every formula can be parsed unambiguously.

Note that we do not yet have the symbols $\neq$ and $\notin$. However, $x \neq y$ is just an abbreviation for the formula $\neg(x = y)$, and similarly, $x \notin y$ is an abbreviation of $\neg(x \in y)$.

We have also not yet defined the symbol $\subseteq$, but we can do so as:

$$A \subseteq B \coloneqq \forall x(x \in A \to x \in B)$$

The notation of using memberships in quantifiers such as $\forall x \in X : P(x)$ or $\exists x \in X : P(x)$ can also be defined by:

$$\forall x \in X : P(x) \coloneqq \forall x\big(x \in X \to P(x)\big)$$
$$\exists x \in X : P(x) \coloneqq \exists x\big(x \in X \land P(x)\big)$$

(where $P(x)$ is some first order formula with free variable x).

Using these symbols, we can now begin to express some basic axioms for set theory.

4 The Axioms of ZF

4.1 Axiom of Extensionality

Axiom of Extensionality.
If two sets have exactly the same members, then they are equal:

$$\forall X \forall Y \big(\forall z(z \in X \leftrightarrow z \in Y) \to x = y\big)$$

Note that our theory concerns itself only with sets, so it matters not if we use upper or lowercase letters as variable symbols.

4.2 Axiom of The Empty Set

Axiom of the Empty Set.
There exists a set with no elements:

$$\exists E \forall x : x \notin E$$

We can prove a theorem with these two axioms:

Theorem 4.1. *There exists exactly one set with no members.*

Proof. By the axiom of the empty set, there exists at least one such set.

For uniqueness, suppose A and B are sets with no members. Then, for every x, the implication $x \in A \to x \in B$ holds vacuously, as does the reverse implication, so $A = B$ by the axiom of extensionality. ■

We call this set the *empty set*, with the theorem above justifying the wording "*the* empty set" over "*an* empty set".

4.3 Axiom of Pairing

Axiom of Pairing.
For any two sets u and v, there exists a set that contains exactly u and v as elements.

$$\forall u \forall v \exists X \forall x\big(x \in X \leftrightarrow (x = u \vee x = v)\big)$$

We denote the set obtained from pairing u and v by $\{u,v\}$, with uniqueness given by extensionality. If $u = v$, then we also denote this by $\{u\}$.

4.4 Axiom of Binary Union

Axiom of Binary Union.
For any two sets u and v, there is a set whose members are those sets that are members of u or of v:

$$\forall u \forall v \exists U \forall x : \big(x \in U \leftrightarrow (x \in u \vee x \in v)\big)$$

4.5 Axiom of the Power Set

Axiom of the Power Set.
For any set u, there is a set whose elements are exactly the subsets of u:

$$\forall u \exists P \forall s (s \subseteq u \leftrightarrow s \in P)$$

or omitting the abbreviation $\subseteq$,

$$\forall u \exists P \forall s\big(\forall x (x \in s \rightarrow x \in s) \leftrightarrow s \in P\big)$$

4.6 Free and Bound Variables

A *free variable*, also called a *parameter*, is a variable that is not *bound* by a preceding quantifier.

Example. In the formula

$$\forall x : x = y$$

the variable x is bound, while y is free.

Quantifiers have a certain *scope* in which they bind their variables, so more accurately, we should talk about free and bound *instances* or *occurrences* of variables.

Example. Consider the formula

$$(\forall x : x \in y) \wedge (\exists y : y \in X)$$

Clearly, x is bound and X is free. However, y is free in the first clause and bound in the second.

This is similar to variable binding in other areas of mathematics:

$$x + \int_0^1 x \, dx$$

Although the integral uses the symbol x, that instance of the symbol is bound, and doesn't really have anything to do with the free variable x outside of the integral. We could (and should) use a different label for that bound variable:

$$x + \int_0^1 t\,dt$$

Similarly, the symbol used to denote a bound variable may be changed freely in a formula of first order logic:

$$(\forall x : x \in y) \wedge (\exists z : z \in X)$$

(This is closely related to the notion of α-conversion in the lambda calculus.)

4.7 Truth Values

A formula that has no free variables is called a *sentence*. Assuming classical first order logic – specifically, the law of the excluded middle – any sentence has a truth value (though they may not be decidable in any particular theory c.f. Gödel's incompleteness theorems).

Formulae that are not sentences do not have truth values, because they contain free variables. A formula that is not a sentence may be true for all possible valuations of their parameters, such as

$$x = x$$

which is true for all x (such a formula is a *tautology*), but the formula itself does not have a truth value.

5 More Axioms

5.1 Axiom Schema of Specification

Most of the existence axioms so far have been of the form

$$\forall t_1 \forall t_2 \ldots \forall t_k \exists B \forall x \big(x \in B \leftrightarrow \varphi \big)$$

where φ is some meaningful statement about x and the sets $t_1, \ldots, t_k$. For instance, we have $k = 2$ with $t_1 = u$ and $t_2 = v$ in the pairing or axiom union, and $k = 0$ in the axiom of the empty set.

Note that for this formula to be meaningful, the statement φ cannot contain B as a variable, or else it's not a very useful definition of the set B. Also, to resolve Russell's paradox as before, we will only allow the collection of elements from some existing set. This leads to the next set of axioms:

Unlike the axioms we have seen so far, this is an axiom *schema* because it contains infinitely many axioms – one for each property φ.

Axiom Schema of Specification.
Let φ be any formula that does not contain the variable name B and has only bound variables, except for $x, t_1, \ldots, t_k$. Then, the following is an axiom:

$$\forall t_1 \forall t_2 \ldots \forall t_k \forall A \exists B \forall x \big(x \in B \leftrightarrow (x \in A \wedge \varphi) \big)$$

That is, for any property φ of x and any set A, there exists a set B that contains exactly the elements of A for which $\varphi(x)$ holds, and φ may depend on additional parameters $t_1, \ldots, t_k$.

This axiom schema is also known as *separation*, since it allows us to separate out the elements of a set that satisfy a property, or as *restricted comprehension*, since it constrains what sets can be constructed via set comprehension.

These axioms define new sets, which we denote as

$$B_{t_1, \ldots t_k, A} = \{x \in A : \varphi(x, t_1, \ldots, t_k)\}$$

or

$$B = \{x \in A : \varphi(x)\}$$

Theorem 5.1. *For any sets A and T, there is a unique set B whose elements are precisely those that are members of both A and T.*

This set is called the *intersection* of A and T, and is denoted (as usual) by $B = A \cap T$.

Proof. The axiom schema of specification contains the axiom

$$\forall t_1 \forall A \exists B \forall x \big(x \in B \leftrightarrow (x \in A \wedge x \in t_1)\big)$$

which implies the existence of at least one such set. For uniqueness, suppose X and Y are both intersections of A and T. Then, for each $x \in X$,

$$x \in X \leftrightarrow (x \in A \wedge x \in B) \leftrightarrow x \in Y$$

so $X = Y$ by extensionality. ■

We can also prove that Russell's paradox does not occur using these axioms:

Theorem 5.2. *Russell's set $R = \{x : x \notin x\}$ does not exist.*

Proof. Suppose R exists. If $R \in R$, then $R \notin R$; and if $R \notin R$, then $R \in R$. In either case, we have a contradiction. ■

Theorem 5.3. *There is no set of all sets. That is,*

$$\neg \exists U \forall x : x \in U$$

Proof. Suppose U is such a set, and let φ be the formula $x \notin x$. The formula φ does not contain U and has x free with no other bound variables, so specification yields the set

$$R = \{x \in U : x \notin x\}$$

That is,

$$x \in R \leftrightarrow (x \in U \wedge x \notin x)$$

Because U contains every set, $x \in U$ is a tautology, so

$$x \in R \leftrightarrow x \notin x$$

This implies that R is Russell's set, which does not exist by the previous theorem. ■

5.2 Axiom of Union

Using the axiom of binary union, we can form the union $A \cup B$ of two sets, and by repeating it, we can form the union of three or more sets as $(A \cup B) \cup C$. However, we cannot form arbitrary unions of infinitely many sets u_1,u_2, ...

Because we attempting to axiomatise sets, we will require that any collection of sets we are attempting to take the union of is itself a set. That is, we wish to take the union of the members of a set $A = \{u_1,u_2,\ldots\}$. Such a union would be a set B whose elements are exactly the members of the members of A.

Axiom of Union.
For any set A, there exists a set B whose members are precisely the members of the members of A:

$$\forall A \exists B \forall x \big(x \in B \leftrightarrow \exists y (y \in A \land x \in y)\big)$$

This set is unique through an argument identical to that for intersections. We denote this set by

$$B = \bigcup A$$

That is, $x \in \bigcup A$ if and only if there is a set $y \in A$ such that $x \in y$.

Note that $\bigcup$ is a unary operator, while the previous $\cup$ was binary. However, the new operator is stronger in that we can take the union of more sets.

Theorem 5.4. *The axiom of union and axiom of pairing imply the axiom of binary union.*

Proof. Pairing sets u and v gives $\{u,v\}$, and the union $\bigcup\{u,v\}$ is precisely $u \cup v$. ■

5.3 Arbitrary Intersections

We have already found the binary intersection using specification, but we would like to define the intersection of members of any set. Unlike for unions, a new axiom isn't required for this.

Theorem 5.5. *For any non-empty set A, there is a unique set B whose members are precisely those that are members of all members of A. That is,*

$$x \in B \leftrightarrow \forall y (y \in A \to x \in y)$$

Proof. Suppose A is non-empty and fix some set $w \in A$. Then, the set

$$\big\{x \in w : \forall y (y \in A \to x \in y)\big\}$$

exists by specification, and its members are precisely those that are members of every member of A. Uniqueness follows from extensionality. ■

Analogously to unions, we write $B = \bigcap A$ for this unary intersection. We can also express binary unions as the special case $x \cap y = \bigcap\{x,y\}$.

Theorem 5.6. *Let A be non-empty and let $A \subseteq B$. Then,*

$$\bigcap A \supseteq \bigcap B$$

Proof. Let $x \in \bigcap B$. By the definition of the intersection,

$$\begin{aligned} x \in \bigcap B &\leftrightarrow \forall y (y \in B \to x \in y) \\ &\to \forall y (y \in A \to x \in y) \\ &\leftrightarrow x \in \bigcap A \end{aligned}$$

so $\bigcap B \subseteq \bigcap A$ as required. ■

6 Ordered Pairs

One important basic mathematical structure is the *ordered pair*, written as (x,y) or $\langle x,y\rangle$, satisfying the characteristic property

$$\langle x,y\rangle = \langle a,b\rangle \leftrightarrow (x = a \wedge y = b)$$

The goal is to find a representation of this structure made from pure sets that satisfies the property above.

Let's first see two encodings that don't work:

1. $\langle x,y\rangle \coloneqq \{x,y\}$. For any sets x and y, we have,

$$\begin{aligned}\langle x,y\rangle &= \{x,y\}\\ &= \{y,x\}\\ &= \langle y,x\rangle\end{aligned}$$

2. $\langle x,y\rangle \coloneqq \big\{x,\{y\}\big\}$. For any sets x and y, we have,

$$\begin{aligned}\big\langle \{x\},y\big\rangle &= \big\{\{x\},\{y\}\big\}\\ &= \big\{\{y\},\{x\}\big\}\\ &= \big\langle \{y\},x\big\rangle\end{aligned}$$

In either encoding, if $x \neq y$, we have a contradiction.

The standard *Kuratowski construction* of the ordered pair is given by

$$\langle x,y\rangle \coloneqq \big\{\{x\},\{x,y\}\big\}$$

which exists by repeated applications of pairing.

Theorem 6.1. *The Kuratowski construction satisfies the characteristic property of the ordered pair. That is,*

$$\big\{\{x\},\{x,y\}\big\} = \big\{\{a\},\{a,b\}\big\} \leftrightarrow (x = a \wedge y = b)$$

Proof. Suppose that

$$\big\{\{x\},\{x,y\}\big\} = \big\{\{a\},\{a,b\}\big\}$$

We have $\{x,y\} \in \big\{\{a\},\{a,b\}\big\}$, so either

(*a*) $\{x,y\} = \{a\}$ or

(*b*) $\{x,y\} = \{a,b\}$

holds. We also have $\{x\} \in \big\{\{a\},\{a,b\}\big\}$, so either

(*c*) $\{x\} = \{a\}$ or

(*d*) $\{x\} = \{a,b\}$

holds.

If (a) holds, then $x = y = a$ and the equation reduces to

$$\big\{\{a\}\big\} = \big\{\{a\},\{a,b\}\big\}$$

so $x = y = a = b$. If (b) holds and

- if (c) also holds, we have $x = a$ and $\{x,y\} = \{x,b\}$. If $x = b$, then $x = y = a = b$. Otherwise, $y = b$.

- if (d) also holds, then $x = y = a = b$.

In all cases, $x = a$ and $y = b$.

The other direction is trivial. ■

Theorem 6.2.

(i) *There is a formula with free variables x,y,z that is satisfied if and only if $z = \langle x,y \rangle$.*

(ii) *There is a formula with free variable z that is satisfied if and only if z is an ordered pair.*

(iii) *There is a formula with free variable x,z that is satisfied if and only if z is an ordered pair with x as the first coordinate.*

(iv) *There is a formula with free variable y,z that is satisfied if and only if z is an ordered pair with y as the second coordinate.*

Proof.

(i)
$$\varphi(x,y,z) = \exists L \exists R \Big(\underbrace{(L \in z) \wedge (R \in z) \wedge \forall t \big(t \in z \to (t = L \vee t = R)\big)}_{z \text{ has precisely two elements, } L \text{ and } R} \\ \wedge \underbrace{(x \in L) \wedge \forall l (l \in L \to l = x)}_{L \text{ contains precisely } x} \wedge \underbrace{(x \in R) \wedge (y \in R) \wedge \forall r (r \in R \to r = x \vee r = y)}_{R \text{ contains precisely } x \text{ and } y} \Big)$$

The next three follow trivially as partial applications of this formula:

(ii) $\exists x \exists y : \varphi(x,y,z)$

(iii) $\exists y : \varphi(x,y,z)$

(iv) $\exists x : \varphi(x,y,z)$

■

6.1 Cartesian Product

Now we have ordered pairs, we can define cartesian products as follows:

$$A \times B \coloneqq \big\{ \langle a,b \rangle : a \in A \wedge b \in B \big\}$$

However, we have not yet proven the existence of this set. First, we identify a larger set that contains all such ordered pairs, before using specification to obtain the cartesian product.

Lemma 6.3. *If $a \in A$ and $b \in B$, then $\langle a,b \rangle \in \mathcal{P}\mathcal{P}(A \cup B)$.*

Proof.

$$\begin{aligned} a \in A \wedge b \in B \to a &\in A \cup B \wedge b \in A \cup B \\ &\leftrightarrow \{a\} \subseteq A \cup B \wedge \{a,b\} \subseteq A \cup B \\ &\leftrightarrow \{a\} \in \mathcal{P}(A \cup B) \wedge \{a,b\} \in \mathcal{P}(A \cup B) \\ &\leftrightarrow \big\{\{a\},\{a,b\}\big\} \in \mathcal{P}(A \cup B) \\ &\leftrightarrow \big\{\{a\},\{a,b\}\big\} \in \mathcal{P}\mathcal{P}(A \cup B) \\ &\leftrightarrow \langle a,b \rangle \in \mathcal{P}\mathcal{P}(A \cup B) \end{aligned}$$

■

Theorem 6.4. *The cartesian product of two sets is a set.*

Proof. Consider the formula

$$\psi(a,b,z) \equiv a \in A \wedge b \in B \wedge \varphi(a,b,z)$$

where φ is the formula from Theorem 6.2. Specification then yields the set

$$\big\{z \in \mathcal{P}\mathcal{P}(A \cup B) : \psi(a,b,z)\big\}$$

which, by the lemma above, is $A \times B$. Uniqueness follows from extensionality. ■

The cartesian product is not commutative, as the pairs are ordered, and it is also not associative, as $(A \times B) \times C$ consists of pairs of the form $\langle\langle a,b\rangle, c\rangle$, while $A \times (B \times C)$ consists of pairs of the form $\langle a,\langle b,c\rangle\rangle$. They are, however, naturally isomorphic.

To reduce the number brackets required, the convention is that the product operator binds to the left. That is, $A \times B \times C \times D \times E$ should be read as $((((A \times B) \times C) \times D) \times E)$.

7 Relations and Functions

7.1 Relations

A *relation* R is a set of ordered pairs. If $\langle x,y\rangle \in R$, then we use infix notation and write xRy.

Given a relation R, we define the *domain*, *range*, and *field* of R as

$$\begin{aligned}\operatorname{dom}(R) &\coloneqq \{x \mid \exists y : xRy\}\\ \operatorname{ran}(R) &\coloneqq \{y \mid \exists x : xRy\}\\ \operatorname{field}(R) &\coloneqq \operatorname{dom}(R) \cup \operatorname{ran}(R)\end{aligned}$$

Lemma 7.1. *If $\langle x,y\rangle \in A$, then $x,y \in \bigcup\bigcup A$.*

Proof.

$$\begin{aligned}\langle x,y\rangle &\in A\\ \big\{\langle x,y\rangle\big\} &\subset A\\ \Big\{\{\{x\},\{x,y\}\}\Big\} &\subset A\\ \bigcup\Big\{\{\{x\},\{x,y\}\}\Big\} &\subset \bigcup A\\ \{\{x\},\{x,y\}\} &\subset \bigcup A\\ \bigcup\{\{x\},\{x,y\}\} &\subset \bigcup\bigcup A\\ \{x,y\} &\subset \bigcup\bigcup A\\ x,y &\in \bigcup\bigcup A\end{aligned}$$

■

Corollary 7.1.1. *The domain, range, and field of a relation are sets.*

Proof. By specification, the following are sets:

$$\begin{aligned}\mathrm{dom}(R) &= \left\{x \in \bigcup\bigcup R \;\middle|\; \exists y : xRy\right\}\\ \mathrm{ran}(R) &= \left\{x \in \bigcup\bigcup R \;\middle|\; \exists x : xRy\right\}\end{aligned}$$

so field$(R) = \mathrm{dom}(R) \cup \mathrm{ran}(R)$ is a set by union. ■

7.2 Functions

In ordinary mathematics, we think of a function as a special kind of correspondence between pairs of object, where every given object x (the "input") is assigned exactly one corresponding object y (its "image") by the function. This means that each such object can be represented as an ordered pair $\langle x,y\rangle$.

A *function* F is a relation such that for every $x \in \mathrm{dom}(R)$ there exists a unique y with $\langle x,y\rangle \in F$. If $\langle x,y\rangle \in F$ for some function F, we write $y = F(x)$ to denote this.

The *inverse* of a relation R is the set

$$R^{-1} := \big\{\langle x,y\rangle : yRx\big\}$$

and the *composition* of two relations R and T is the set

$$R \circ T := \big\{\langle x,y\rangle : \exists z(xTz \wedge zRy)\big\}$$

Theorem 7.2. *If F and G are functions, then the composition $F \circ G$ is a function with domain*

$$\mathrm{dom}(F \circ G) = \big\{x \in \mathrm{dom}(G) : G(x) \in \mathrm{dom}(F)\big\}$$

Proof. Suppose $\langle x,y\rangle,\langle x,y'\rangle \in F \circ G$. Then, there exist t and t' such that

$$\begin{aligned}\langle x,t\rangle &\in G \qquad \langle t,y\rangle \in F\\ \langle x,t'\rangle &\in G \qquad \langle t',y'\rangle \in F\end{aligned}$$

Since G is a function, $t = t'$, so we have

$$\begin{aligned}\langle t,y\rangle &\in F\\ \langle t,y'\rangle &\in F\end{aligned}$$

and since F is a function, $y = y'$, so $\langle x,y\rangle = \langle x,y'\rangle$ and $F \circ G$ is a function. ■

7.3 Images

Given a relation R and a set X, the (R-)*image* of X under R is the set

$$R[X] := \big\{y \;\big|\; \exists x \in X : \langle x,y\rangle \in R\big\}$$

and the (R-)*preimage* is the set

$$\begin{aligned}R[X] &:= \big\{y \;\big|\; \exists x \in X : \langle x,y\rangle \in R^{-1}\big\}\\ &= \big\{y \;\big|\; \exists x \in X : \langle y,x\rangle \in R\big\}\end{aligned}$$

Note that it is not necessary to have $X \subseteq \mathrm{dom}(R)$ when taking the R-image, or $X \subseteq \mathrm{ran}(R)$ when taking the R-preimage, since $R[X] = R[X \cap \mathrm{dom}(R)]$ and $R^{-1}[X] = R^{-1}[X \cap \mathrm{ran}(R)]$.

A function F is *injective* if for every y there is at most one x such that $\langle x,y\rangle \in F$.

Theorem 7.3. *A function F is injective if and only if F^{-1} is a function.*

Otherwise F^{-1} is a relation but not a function.

Theorem 7.4. *For any functions F and G, we have*

- $(F \circ G)(x) = F\big(G(x)\big)$;
- $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$.

We write $f : A \to B$ to denote a function f with $\operatorname{dom}(f) = A$ and $\operatorname{ran}(f) \subseteq B$. A function f is *surjective* if $\operatorname{ran}(f) = B$.

Theorem 7.5.

- *For any $A \neq \emptyset$, there is no function $f : A \to \emptyset$;*
- *For any B, there is a unique function $f : \emptyset \to B$ given by $f = \emptyset$.*

7.4 Cantor's Diagonal Argument

Theorem (Cantor)**.** *For any set A, there is no surjection $f : A \to \mathcal{P}(A)$.*

Proof. Let $f : A \to \mathcal{P}(A)$ be a function, and define the set

$$S = \big\{a \in A : a \notin f(a)\big\}$$

which is a set by specification. Clearly, $S \subseteq A$, so $S \in \mathcal{P}(A)$.

Suppose $S \in \operatorname{ran}(f)$. Then, there exists $a_0 \in A$ such that $S = f(a_0)$. If $a_0 \in S$, then $a_0 \notin f(a_0) = S$, and if $a_0 \notin S$, then $a_0 \in f(a_0) = S$. This is a contradiction, so $S \notin \operatorname{ran}(f)$, and f is not surjective. ■

We define the *identity function* on a set A by

$$I_A \coloneqq \big\{\langle a,a\rangle : a \in A\big\}$$

That is, $I_A(a) = a$ for all $a \in A$.

Theorem 7.6. *If A is non-empty and there is an injective function $f : A \to B$, then there exists a surjective function $g : B \to A$ such that $g \circ f = I_A$.*

Proof. Let $a_0 \in A$. If $b \in \operatorname{ran}(f)$, then define $g(b)$ to be the unique $a \in A$ such that $f(a) = b$ (since f is injective), otherwise define $g(b) = a_0$. This function (set) exists because

$$g = f^{-1} \cup \big\{\langle b,a_0\rangle \in B \times A : b \notin \operatorname{ran}(f)\big\}$$

is the union of two sets, the latter of which exists by specification.

Clearly, g is surjective and we have $g \circ f = I_A$. ■

Theorem 7.7. *If $f : A \to B$ is surjective, then there exists an injective function $g : B \to A$ such that $f \circ g = I_B$.*

The sets $f^{-1}\big[\{b\}\big]$ for $b \in B$ are non-empty and partition A.

For each b, select $a_b \in f^{-1}\big[\{b\}\big]$. Then, the map $g : B \to A$ defined by $b \mapsto a_b$ is injective, and clearly, $f \circ g = I_B$.

However, this proof is not yet valid with the axioms we have so far – we have yet to prove that the set constructed above exists. The problem is that we do not know if collecting an element a_b from infinitely many sets $f^{-1}\big[\{b\}\big]$ yields a set.

Axiom of Choice (first form).
For any relation R, there exists a function $F \subseteq R$ such that $\operatorname{dom}(F) = \operatorname{dom}(R)$.

That is, for each element $x \in \operatorname{dom}(R)$, the function F "chooses" exactly one y with $\langle x,y\rangle \in R$. That is, exactly one element $y \in R\big[\{x\}\big]$.

Proof of Theorem 7.7. (AC) Suppose $f : A \to B$ is surjective, and consider the relation f^{-1}. We have $\operatorname{dom}(f^{-1}) = B$ and $\operatorname{ran}(f^{-1}) = A$, so by the axiom of choice, there exists a function $F \subset f^{-1}$ with domain B. For every $b \in B$, we must have $\langle F(b),b\rangle \in f$, so $f\big(F(b)\big) = b$. That is, $f \circ F = I_B$. ■

8 Constructing Numbers

8.1 Axiom of Infinity

Since we are attempting to embed all of mathematics into set theory, we should find sets that correspond to natural numbers, as well as a set that contains all natural numbers.

Zermelo proposed the following construction of the natural numbers:

- $0 = \emptyset$;
- $1 = \{\emptyset\}$;
- $2 = \{\{\emptyset\}\}$;
- $3 = \{\{\{\emptyset\}\}\}$;
- $n = \{n-1\}$.

However, the generally accepted (and in many senses better) convention given by von Neumann is as follows:

- $0 = \emptyset$;
- $1 = \{0\} = \{\emptyset\}$;
- $2 = \{0,1\} = \{\emptyset,\{\emptyset\}\}$;
- $3 = \{0,1,2\} = \{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\}$;
- $4 = \{0,1,2,3\} = \{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\},\{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\}\}$;
- $n = \{0,1,\ldots,n-1\}$.

One major advantage of this encoding is that the set representing n now has n elements (while in Zermelo's construction, every natural number n has exactly one element – $n-1$).

We also have

$$0 \in 1 \in 2 \in 3 \in \cdots$$

and

$$0 \subset 1 \subset 2 \subset 3 \subset \cdots$$

For a set x, we define its *successor* x^+ by $x^+ = x \cup \{x\}$.

Example.

- $0^+ = 0 \cup \{0\} = \{0\} = 1$;
- $1^+ = 1 \cup \{1\} = \{0,1\} = 2$;

- $2^+ = 2 \cup \{2\} = \{0,1,2\} = 3$;
- $3^+ = 3 \cup \{3\} = \{0,1,2,3\} = 4$;

so $0 = \emptyset$, $1 = \emptyset^+$, $2 = \emptyset^{++}$, $3 = \emptyset^{+++}$.

It is now intuitively clear what natural numbers are in set theory, but without a further axiom, we cannot write a formula $\varphi(x)$ that identifies if x is a natural number or not, and we also cannot form the set of all natural numbers yet.

Lemma 8.1. *For all $m,n \in \omega$.*

(*i*) $0 \neq n^+$;

(*ii*) $m \in n \to m^+ \in n^+$;

(*iii*) $m^+ = n^+ \to m = n$.

(You may recognise some of these as Peano axioms.)

A set A is *inductive* if it contains the empty set and is closed under the successor operation. That is,

$$\forall x(x \in A \to x^+ \in A)$$

Axiom of Infinity.
There is an inductive set:

$$\exists A\big(\emptyset \in A \wedge \forall x(x \in A \to x^+ \in A)\big)$$

or omitting $\emptyset$ and x^+,

$$\exists A\big(\exists e(\forall z : z \notin e) \wedge e \in A \wedge \forall x(x \in A \to x \cup \{x\} \in A)\big)$$

The axiom of infinity gives the existence of inductive sets, but does not provide uniqueness. We are only interested in the “smallest” such set.

A *natural number* is a set that is a member of every inductive set.

Theorem (Existence of Natural Numbers)**.**

(*i*) *There is a set ω whose elements are precisely the natural numbers.*

(*ii*) *The set ω is the unique set that is a subset of every inductive set.*

Proof. Let A be an inductive set given by the axiom of infinity. Then, by specification, the following is a set:

$$\begin{aligned} \omega &= \big\{a \in A : \forall S((\forall y : y \in S \to y^+ \in S) \to x \in S)\big\} \\ &= \{a \in A : x \text{ is an element of every inductive set}\} \end{aligned}$$

Clearly, its elements are precisely the natural numbers, and we also have $\omega \subseteq S$ for any inductive set S, with uniqueness given by extensionality. ■

We will use the notation $\mathbb{N} = \{0,1,2,\ldots\}$ to refer to the natural numbers when the encoding is irrelevant, and ω whenever von Neumann’s convention is required.

Theorem (Induction Principle for ω)**.** *Any inductive subset of ω coincides with ω.*

Proof. Clear from the definition of ω. ■

Theorem (Proof by Induction)**.** *Suppose φ is a property of natural numbers such that $\varphi(0)$ holds and for every natural number x, $\varphi(x) \to \varphi(x^+)$. Then $\varphi(x)$ holds for all natural numbers x.*

Proof. Let $X = \{x \in \omega : \varphi(x)\} \subseteq \omega$. By assumption, $\varphi(0)$ holds, so $0 = \emptyset \in X$, and because $\varphi(x) \to \varphi(x^+)$, X is inductive. Thus, by the previous theorem, $X = \omega$. ■

Theorem 8.2. *Every element of ω is either 0 or the successor x^+ of a unique $x \in \omega$.*

Proof. Define the set

$$A = \{n \in \omega : n = \emptyset \vee \exists m \in \omega : n = m^+\}$$

Clearly, $0 = \emptyset$ is a member of A, and A is closed under the successor operation, so A is inductive, and hence $A = \omega$. Injectivity of the successor function (Theorem 8.1) implies uniqueness. ■

8.2 Ordering of ω

We define the *less than* relation $<_\omega$ on ω by

$$<_\omega := \big\{\langle m,n\rangle \in \omega \times \omega : m \in n\big\}$$

and the *less than or equal to* relation $\leq_\omega$ on ω by

$$\leq_\omega := \big\{\langle m,n\rangle \in \omega \times \omega : m \in n \vee m = n\big\}$$

If $\langle m,n\rangle \in <_\omega$, we write $m < n$, and similarly, if $\langle m,n\rangle \in \leq_\omega$, we write $m \leq n$.

Theorem 8.3. *The relation $\in$ linearly orders ω. That is, it is:*

(*i*) *irreflexive: $\forall n \in \omega : n \notin n$;*

(*ii*) *transitive: $\forall x,y,z \in \omega : (x \in y \wedge y \in z) \to x \in z$;*

(*iii*) *linear/total: $\forall m,n \in \omega : m \in n \vee m = n \vee n \in m$.*

Proof.

(*i*) Define the set

$$A = \{n \in \omega : n \notin n\}$$

We show A is inductive via induction on n.

The empty set has no elements, so $\emptyset \notin \emptyset$ holds, and $\emptyset \in A$.

Now assume that $n \in A$ and suppose for a contradiction that $n^+ \in n^+ = n \cup \{n\}$. Then, either $n^+ \in n$ or $n^+ = n$. In the former case, $n \in n \cup \{n\} = n^+ \in n$, so by transitivity (*ii*), $n \in n$, contradicting that $n \in A$. In the latter case, $n \in n \cup \{n\} = n^+ = n$, so $n \in n$, again contradicting that $n \in A$.

It follows that $n^+ \notin n^+$, so $n^+ \in A$. So, $A \subseteq \omega$ is inductive, giving $A = \omega$, and hence $\in$ is an irreflexive relation on ω.

(*ii*) Fix $x,y \in \omega$ and define the set

$$A = \{z \in \omega : x \in y \in z \to x \in z\}$$

We show A is inductive via induction on z.

If $z = \emptyset$, then the implication holds vacuously, so $\emptyset \in A$.

Now assume that $z \in A$, and suppose that $x \in y \in z^+$. As $y \in z^+ = z \cup \{z\}$, we either have $y \in z$ or $y = z$. If $y \in z$, then by the inductive hypothesis $x \in z$, and since $z \subseteq z^+$, we have $x \in z^+$, so $z^+ \in A$.

If $y = z$, then $x \in y$ gives $x \in z$. Again, $z \subseteq z^+$, so $x \in z^+$, and $z^+ \in A$.

So, $A \subseteq \omega$ is inductive, giving $A = \omega$, and hence $\in$ is a transitive relation on ω.

(*iii*) Define the set

$$B = \{n \in \omega : \emptyset = n \vee \emptyset \in n\}$$

We show B is inductive via induction on n.

If n is empty, then $n = \emptyset$ so $n \in B$.

Now assume that $n \in B$, so either $\emptyset \in n$ or $\emptyset = n$. In the former case, $\emptyset \in n \in n \cup \{n\} = n^+$, and in the latter case $\emptyset \in \emptyset \cup \{\emptyset\} = n^+$. In either case, $\emptyset \in n^+$ by transitivity (*ii*), so $B \subseteq \omega$ is inductive and hence $B = \omega$, so $\emptyset \in n \vee \emptyset = n$ holds for all $n \in \omega$.

Now define the set

$$A = \{m \in \omega : \forall n(m \in n \vee m = n \vee n \in m)\}$$

If $m = \emptyset$, then $m \in A$ by the above result.

Now assume $m \in A$. If $n \in m$, then $n \in m^+$ by transitivity. If $n = m$, then $n^+ = m^+$. If $m \in n$, then $m^+ \in n^+$, so either $m^+ \in n$ or $m^+ = n$. In all cases, one of the conditions hold, so $m^+ \in A$. So, $A \subseteq \omega$ is inductive, giving $A = \omega$, and hence $\in$ is total on ω.

■

8.3 Recursion

Theorem (Recursion on ω)**.** *Let X be a set and $x \in X$. Let $r : X \to X$ be a function. Then, there is a unique function $f : \omega \to X$ such that*

(*i*) $f(0) = x$;

(*ii*) $f(n^+) = r\big(f(n)\big)$.

Informally, given a set X, and an element $x \in X$, every function $r : X \to X$ generates a unique sequence of elements $(x_i)_{i=1}^{\infty} \subseteq X$ such that $x_0 = x$ and $x_{n+1} = r(x_n)$. The theorem above then says that the mapping that sends n to x_n defines a function $\omega \to X$.

That is, the following diagram commutes

$$\begin{tikzcd}
& \mathbb{N} \arrow[r, "s"] \arrow[d, "f"] & \mathbb{N} \arrow[d, "f"] \\
\{*\} \arrow[ur, "0"] \arrow[dr, "x"'] & & \\
& X \arrow[r, "r"'] & X
\end{tikzcd}$$

Proof sketch. Call a function v *acceptable* if the following four properties hold:

- $\mathrm{dom}(v) \subseteq \omega$;
- $\mathrm{ran}(v) \subseteq A$;
- $0 \in \mathrm{dom}(v) \to v(0) = a$;
- $n \in \omega \wedge n^+ \in \mathrm{dom}(v) \to n \in \mathrm{dom}(v) \wedge v(n^+) = F\big(v(n)\big)$.

The last property implies that if $n \in \text{dom}(v)$, then $\{0,1,...,n\} \subseteq \text{dom}(v)$. The empty set is also an acceptable function, as is the function $v = \{\langle 0,a\rangle\}$.

Let $\mathcal{K}$ be the collection of acceptable functions. Because any acceptable function is a function with domain contained in ω and range contained in A, all of its elements are ordered pairs in $\omega \times A$. That is, every acceptable function v is a subset of $\omega \times A$, or an element in $\mathcal{P}(\omega \times A)$. Thus, $\mathcal{K} \subseteq \mathcal{P}(\omega \times A)$ constitutes a set by specification.

Define

$$f := \bigcup \mathcal{K}$$

That is, f is the relation formed from the union of all acceptable functions. This relation f satisfies $\langle n,y\rangle \in h \leftrightarrow \exists v \in \mathcal{K} : \langle n,y\rangle \in v$.

It can be proven that f is itself an acceptable function, and moreover, its domain is ω, thus satisfying the conclusions of the theorem. This function can also be shown to be unique by considering the set

$$T = \{n \in \omega : f_1(n) = f_2(n)\}$$

and proving that it is inductive. ■

8.4 Classes & Class-Functions

What happens if we try to iterate the power set operation? That is, does this theorem on recursion prove the existence of a function f on ω for which $f(0) = \emptyset$ and $f(n^+) = \mathcal{P}(f(n))$? The range of such a function would be the collection $\{\emptyset,\mathcal{P}(\emptyset),\mathcal{P}\mathcal{P}(\emptyset),\mathcal{P}\mathcal{P},\mathcal{P}(\emptyset),\ldots\}$.

With our current axioms, we cannot prove the existence of this function, nor of this set. The problem is that the power set operation is *not* a function, as it is not a set – the power set operation may be applied to any set, so its domain would be the collection of all sets, which we know is not a set.

A *class* is a collection of sets satisfying a formula φ:

$$H = \{x : \varphi(x)\}$$

This is similar to the earlier unrestricted comprehension, but now, these collections do not a priori constitute sets, being only classes unless proved otherwise.

Example. The class V defined by

$$V = \{x : x = x\}$$

is the class of all sets, called the *universe.*

A class that is not a set is called a *proper class.* For instance, the universe class is a proper set. A class that happens to be a set is sometimes called a *small class.*

We can now define more precisely what kind of operation $x \mapsto \mathcal{P}(x)$ is.

A *class-function* is a class F whose elements are ordered pairs and for every set x there is exactly one set y such that $\langle x,y\rangle \in F$. So, a class-function is an operation "definable with a formula". A class-function can thus be regarded as an operation $V \to V$, but it is not a function since it is not a set.

A class (and therefore a class-function) is defined by a formula $\varphi(x)$, which one designated free variable x, and possibly other unlisted parameters. As we have defined it, whenever $\varphi(x)$ holds for a class-function, x must be an ordered pair. This is inconvenient.

Instead, we will talk about class-functions using formulae $\varphi(x,y)$ such that for every set x, there is exactly one other set y such that $\varphi(x,y)$ holds.

Theorem 8.4. *There is a formula that is satisfied if and only if $\varphi(x,y)$ defines a class function.*

Proof.

$$\underbrace{\forall x \exists y : \varphi(x,y)}_{\text{existence}} \land \underbrace{\forall x \forall y \forall y' \Big(\big(\varphi(x,y) \land \varphi(x,y')\big) \to y = y' \Big)}_{\text{uniqueness}}$$

■

8.5 Axiom Schema of Replacement

The axiom schema of specification can be rephrased as saying that the "intersection" of any class with a set is a set. More generally, one can safely assume that every sufficiently small class is a set, and proper sets are those that contain "too many" elements.

If we take the "image" of a set under a class-function, intuitively we do not get a "bigger" thing (i.e. a class) than the set we started with. This is the motivation behind the next axiom.

Axiom Schema of Replacement.
The image of a set under a class-function is a set; if φ is any formula that does not contain B, then:

$$\forall A \Big(\underbrace{\forall x \forall y \forall y' \Big(\big(x \in A \land \varphi(x,y) \land \varphi(x,y')\big) \to y = y' \Big)}_{\varphi \text{ is a class-function on at least } A} \to \underbrace{\exists B \forall y \Big(y \in B \leftrightarrow \exists x (x \in A \land \varphi(x,y)) \Big)}_{\text{there is a set } B \text{ consisting of } \varphi\text{-images of elements of } A} \Big)$$

That is, if φ represents a definable function f, A represents its class domain, and $f(x)$ is a set for every $x \in A$, then the image of f is a subset of some set B.

Using the axiom schema of replacement, we can prove a more general recursion theorem that does not presuppose the existence of the set A.

Theorem (Recursion on ω, Class Form)**.** *Let a be any set and let $\varphi(x,y)$ define a class-function. Then, there is a set A and a unique function $f : \omega \to A$ such that $f(0) = a$ and $\varphi\big(f(n), f(n^+)\big)$ for every $n \in \omega$.*

Proof sketch. Call a function n-acceptable if it is acceptable and has domain n. First, we prove that for each $n \in \omega$, there exists an n-acceptable function, and furthermore, that any n-acceptable function and m-acceptable function agree on $n \cap m = \min(n,m)$. This implies there is a unique n-acceptable function for each $n \in \omega$.

Then, let $\varphi(x,y)$ be the formula that if $x \in \omega$, then y is x-acceptable, and otherwise $y = \emptyset$. This defines a class-function by the uniqueness proved above, and the axiom of replacement for φ implies the existence of the set $\mathcal{K}$ of all acceptable functions. Then, as before, $f = \bigcup \mathcal{K}$ satisfies the conclusions of the theorem. ■

Corollary 8.4.1. *There is a function f with domain ω for which $f(0) = \emptyset$ and $f(n^+) = \mathcal{P}\big(f(n)\big)$ for all $n \in \omega$.*

Proof. Apply the class form of the recursion theorem for the formula $\varphi(x,y) \equiv y = \mathcal{P}(x)$; and let $a = \emptyset$. ■

Corollary 8.4.2. *There is a set $\big\{\emptyset, \mathcal{P}(\emptyset), \mathcal{PP}(\emptyset), \mathcal{PPP}(\emptyset), \ldots\big\}$.*

Proof. This is the range of the function of the previous corollary. ■

8.6 Addition and Multiplication on ω

Theorem 8.5 (Parametric Recursion on ω)**.** *Let $f_0 : A \to B$ and $u : B \times A \to A$ be functions. Then, there exists a unique function $f : A \times \omega \to B$ such that*

- $f(a{,}0) = f_0(a)$ *for all* $a \in A$*;*
- $f(a{,}n^+) = u\big(a{,}f(a{,}n)\big)$ *for all* $n \in \omega$ *and* $a \in A$*.*

Corollary 8.5.1. *There is a unique function $+ : \omega \times \omega \to \omega$ such that*

- $+(m{,}0) = m$ *for all* $m \in \omega$*;*
- $+(m{,}n^+) = \big(+(m{,}n)\big)^+$ *for all* $m{,}n \in \omega$*.*

Proof. Let $A = B = \omega$, and let $f_0 : \omega \to \omega$ be the identity function, and $u : \omega \times \omega \to \omega$ be defined by $(a{,}n) \mapsto n^+$ in the previous theorem. ■

We will write this function using infix notation like with relations. Intuitively, this theorem states that the equations $m + 0 = m$ and $m + (n + 1) = (m + n) + 1$ uniquely characterise addition.

Corollary 8.5.2. *There is a unique function $\cdot : \omega \times \omega \to \omega$ such that*

- $m \cdot 0 = 0$ *for all* $m \in \omega$*;*
- $m \cdot n^+ = m + m \cdot n$ *for all* $m{,}n \in \omega$*.*

Proof. Let $A = B = \omega$, $f_0 : \omega \to \omega$ be the constant zero function, and $u : \omega \times \omega \to \omega$ be defined by $(a{,}n) \mapsto a + n$ in the previous theorem. ■

Theorem 8.6 (Basic Properties of ω)**.** *The following general properties all hold:*

(i) $\forall a{,}b \in \omega : a +_\omega b = b +_\omega a$ *(commutativity of addition);*

(ii) $\forall a{,}b{,}c \in \omega : (a +_\omega b) +_\omega c = a +_\omega (b +_\omega c)$ *(associativity of addition);*

(iii) $a +_\omega 0_\omega = a$ *(existence of additive identity);*

(iv) $a + 1 = a^+$ *(equivalence of successor and addition);*

(v) $\forall a{,}b \in \omega : a \cdot_\omega b = b \cdot_\omega a$ *(commutativity of multiplication);*

(vi) $\forall a{,}b{,}c \in \omega : (a \cdot_\omega b) \cdot_\omega c = a \cdot_\omega (b \cdot_\omega c)$ *(associativity of multiplication);*

(vii) $\forall a{,}b{,}c \in \omega : a \cdot_\omega (b +_\omega c) = a \cdot_\omega b +_\omega a \cdot_\omega c$ *(distributivity of multiplication over addition);*

(viii) $\forall a \in \omega : a \cdot_\omega 1_\omega = a$ *(existence of multiplicative identity);*

(ix) $0_\omega \neq 1_\omega$ *(non-degeneracy);*

(x) $\forall a{,}b \in \omega : a \cdot_\omega b = 0_\omega \to (a = 0_\omega \vee b = 0_\omega)$ *(zero divisors).*

Proof of (i). Fix $a \in \omega$ and define the set

$$S = \{b \in \omega : a + b = b + a\}$$

If $b = 0$, then $a + b = b + a$ holds by property (iii), so $0 \in S$.

Now assume $b \in S$. Then, by the definition of addition,

$$\begin{aligned} a + b^+ &= (a + b)^+ \\ &= (b + a)^+ \end{aligned}$$

$$= b + a^+$$

so $b^+ \in S$, and $S \subseteq \omega$ is inductive, so $S = \omega$.

The proofs for the other properties are similar. ■

Theorem 8.7. *For any natural numbers m,n,p*

$$m \in n \quad \leftrightarrow \quad m + p \in n + p$$

and if $p \neq 0$, then also

$$m \in n \quad \leftrightarrow \quad m \cdot p \in n \cdot p$$

Theorem 8.8. *For any natural numbers m,n,p*

$$m + p \in n + p \quad \rightarrow \quad m = n$$

and if $p \neq 0$, then also

$$m \cdot p = n \cdot p \quad \rightarrow \quad m = n$$

Theorem 8.9. *ω is well-ordered by $<_\omega$. That is, every non-empty subset $A \subseteq \omega$ has an $<_\omega$-minimal element.*

8.7 Equivalence Relations

A relation $R \subseteq A \times A$ is an *equivalence relation* on A if it is:

- reflexive: $\forall x \in A : xRx$;
- symmetric: $\forall x,y \in A : xRy \leftrightarrow yRx$;
- transitive: $\forall x,y,z \in A : (xRy \wedge yRz) \rightarrow xRz$.

We define the *equivalence class* $[x]_R$ of $x \in A$ under R as the set

$$[x]_R \coloneqq \{t : xRt\}$$

Note that this is indeed a set by specification, as $[x]_R \subseteq \operatorname{ran}(R) = A$.

Theorem 8.10. *Two elements are equivalent under R if and only if their equivalence classes are equal.*

Proof. Suppose xRy, and let $a \in [x]_R$. Then, by definition, xRa, so by symmetry and transitivity, aRy, so $a \in [y]_R$, and hence $[x]_R \subseteq [y]_R$. Now, let $b \in [y]_R$. Then, by definition, yRb, so by transitivity, xRb, so $b \in [x]_R$, and hence $[y]_R \subseteq [x]_R$. So, $[x]_R = [y]_R$.

For the reverse implication, suppose $[x]_R = [y]_R$. By reflexivity, $y \in [y]_R$, and since $[x]_R = [y]_R$, we also have $y \in [x]_R$, so xRy, as required. ■

Theorem 8.11. *Equivalence classes partition A. That is, the union of all equivalence classes is A, and their pairwise intersections are empty.*

Proof. By reflexivity, $x \in [x]_R$ for all x, so the union of all equivalence classes must be A.

Let $x,y \in A$ be distinct, and suppose $[x]_R \cap [y]_R$ is non-empty. Let $a \in [x]_R \cap [y]_R$, so $a \in [x]_R$ and $a \in [y]_R$. Then, by definition, xRa and yRa, so by the previous theorem, $[x]_R = [a]_R = [y]_R$. ■

8.8 Integers

We can represent natural numbers and the operations of addition and multiplication as sets. The next goal is to find an encoding of the integers, then of the rationals.

A rational number p/q may be expressed as a pair of integers, p and q – but this representation is not unique. For instance, $p/q = 2p/2q$. So, the rationals are really an equivalence relation of these representations on pairs of integers.

We take the same approach for constructing the integers: just a rational is an equivalence class of quotients of integers, an integer can be expressed as an equivalence class of differences of naturals. For instance,

$$-3 = 0 - 3 = 1 - 4 = 2 - 5 = 3 - 6 = \cdots$$

Define $\sim$ to be the equivalence relation on $\omega \times \omega$ for which $\langle a,b\rangle \sim \langle x,y\rangle$ if and only if $a + y = b + x$. The *set of integers* $\mathbb{Z}$ is defined to be the set of equivalence classes

$$\mathbb{Z} := \omega \times \omega/\sim$$

For instance, the integer $2_\mathbb{Z}$ is the equivalence class

$$2_\mathbb{Z} = \big[\langle 2,0\rangle\big] = \big\{\langle 2,0\rangle,\langle 3,1\rangle,\langle 4,2\rangle,\langle 5,3\rangle,\ldots\big\}$$

while the integer $-3_\mathbb{Z}$ is the equivalence class

$$-3_\mathbb{Z} = \big[\langle 0,3\rangle\big] = \big\{\langle 0,3\rangle,\langle 1,4\rangle,\langle 2,5\rangle,\langle 3,6\rangle,\ldots\big\}$$

Note that in this construction, ω is not a subset of $\mathbb{Z}$, and that

$$0_\omega = \emptyset \neq \big\{\langle n,n\rangle : n \in \omega\big\} = 0_\mathbb{Z}$$

How should we define addition? Informally, we have

$$(a - b) + (x - y) = (a + x) - (b + y)$$

Theorem (Addition on $\mathbb{Z}$)**.** *There is a unique function* $+_\mathbb{Z} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ *such that*

$$\big[\langle a,b\rangle\big] +_\mathbb{Z} \big[\langle x,y\rangle\big] = \big[\langle a +_\omega x, b +_\omega y\rangle\big]$$

Proof. We check that this operation is well-defined. Suppose $\langle a,b\rangle \sim \langle a',b'\rangle$ and $\langle x,y\rangle \sim \langle x',y'\rangle$, so $a + b' = b + a'$ and $x + y' = y + x'$. Adding these together, we have $a + b' + x + y' = b + a' + y + x'$. As $+_\omega$ is commutative, we have $(a + x) + (b' + y') = (b + y) + (a' + x')$, so

$$\langle a + x, b + y\rangle \sim \langle a' + x', b' + y'\rangle$$

as required. ■

For multiplication, informally, we have

$$(a - b)(x - y) = (ax + by) - (ay + bx)$$

Theorem (Multiplication on $\mathbb{Z}$)**.** *There is a unique function* $\cdot_\mathbb{Z} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ *such that*

$$\big[\langle a,b\rangle\big] \cdot_\mathbb{Z} \big[\langle x,y\rangle\big] = \big[\langle ax + by, ay + bx\rangle\big]$$

The ring unit and zero are then given by

$$\begin{aligned} 0_{\mathbb{Z}} &= \big[\langle 0,0\rangle\big] \\ 1_{\mathbb{Z}} &= \big[\langle 1,0\rangle\big] \end{aligned}$$

Theorem 8.12 (Basic Properties of $\mathbb{Z}$)**.** *Replacing $(+_\omega, \cdot_\omega, 0_\omega, 1_\omega)$ by $(+_{\mathbb{Z}}, \cdot_{\mathbb{Z}}, 0_{\mathbb{Z}}, 1_{\mathbb{Z}})$, the same results as in Theorem 8.6 for ω hold for $\mathbb{Z}$, with the addition of*

(xi) $\forall a \in \mathbb{Z} : \exists b \in \mathbb{Z} : a +_{\mathbb{Z}} b = 0_{\mathbb{Z}}$ *(existence of additive inverses).*

Although ω is not a true subset of $\mathbb{Z}$, there is a natural embedding $E : \omega \to \mathbb{Z}$ defined by $E(n) = \big[\langle n,0\rangle\big]$ such that $E(\omega)$ behaves like ω:

$$\begin{aligned} E(m +_\omega n) &= E(m) +_{\mathbb{Z}} E(n); \\ E(m \cdot_\omega n) &= E(m) \cdot_{\mathbb{Z}} E(n). \\ E(0_\omega) &= 0_{\mathbb{Z}}; \\ E(1_\omega) &= 1_{\mathbb{Z}}. \end{aligned}$$

(That is, E is a semiring homomorphism.) We also have

$$\langle m,n\rangle = E(m) - E(n)$$

8.9 Rationals

As mentioned previously, rationals can be expressed as the quotient of two integers (non-zero, in the case of the divisor). For instance,

$$\frac{1}{2} = \frac{-1}{-2} = \frac{2}{4} = \frac{-2}{-4} = \frac{3}{6} = \frac{-3}{-6} = \cdots$$

Let $\mathbb{Z}' := \mathbb{Z} \setminus \{0_{\mathbb{Z}}\}$ be the set of non-zero integers and define the equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}'$ for which $\langle a,b\rangle \sim \langle x,y\rangle$ if and only if $a \cdot y = b \cdot x$. The *set of rationals* $\mathbb{Q}$ is the set of equivalence classes

$$\mathbb{Q} := \mathbb{Z} \times \mathbb{Z}' / \sim$$

Informally, we have

$$\frac{a}{b} + \frac{x}{y} = \frac{ay + bx}{by}$$

so,

Theorem 8.13 (Addition on $\mathbb{Q}$)**.** *There is a unique function $+_{\mathbb{Q}} : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ such that*

$$\big[\langle a,b\rangle\big] +_{\mathbb{Q}} \big[\langle x,y\rangle\big] = \big[\langle ay +_{\mathbb{Z}} bx, by\rangle\big]$$

for all $a,b \in \mathbb{Z}$ and $x,y \in \mathbb{Z}'$.

Similarly, for multiplication, we should have

$$\frac{a}{b} \cdot \frac{x}{y} = \frac{ax}{by}$$

so

Theorem 8.14 (Multiplication on $\mathbb{Q}$)**.** *There is a unique function $\cdot_{\mathbb{Q}} : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ such that*

$$\big[\langle a,b\rangle\big] \cdot_{\mathbb{Q}} \big[\langle x,y\rangle\big] = \big[\langle ax, by\rangle\big]$$

for all $a,b \in \mathbb{Z}$ and $x,y \in \mathbb{Z}'$.

The unit and zero are then given by

$$\begin{aligned}0_{\mathbb{Q}} &= \big[\langle 0{,}1\rangle\big]\\ 1_{\mathbb{Q}} &= \big[\langle 1{,}1\rangle\big]\end{aligned}$$

Again, $\mathbb{Z}$ is not a true subset of $\mathbb{Q}$, but there is a natural embedding $E : \mathbb{Z} \to \mathbb{Q}$ defined by $E(a) = \big[\langle a{,}1_{\mathbb{Z}}\rangle\big]$ such that $E(\mathbb{Z})$ behaves like $\mathbb{Z}$:

$$\begin{aligned}E(a +_{\mathbb{Z}} b) &= E(a) +_{\mathbb{Q}} E(b);\\ E(a \cdot_{\mathbb{Z}} b) &= E(a) \cdot_{\mathbb{Q}} E(b);\\ E(0_{\mathbb{Z}}) &= 0_{\mathbb{Q}};\\ E(1_{\mathbb{Z}}) &= 1_{\mathbb{Q}}.\end{aligned}$$

(That is, E is a semiring homomorphism.)

Theorem 8.15 (Basic Properties of $\mathbb{Q}$)**.** *Replacing the relevant operations and constants by* ($+_{\mathbb{Q}}$, $\cdot_{\mathbb{Q}}$,$0_{\mathbb{Q}}$,$1_{\mathbb{Q}}$)*, the same results as in Theorem 8.6 and Theorem 8.12 for ω and $\mathbb{Z}$ hold for $\mathbb{Q}$, with the addition of*

(xii) $\forall a \in \mathbb{Q} : \exists b \in \mathbb{Q} : a \cdot_{\mathbb{Q}} b = 1_{\mathbb{Q}}$ *(existence of multiplicative inverses).*

8.10 Real Numbers

There are many possible approaches to construct the real numbers.

- Decimal expansions: every real number can be expressed as an integer and an infinite sequence of digits (a function $\omega \to \{0{,}...{,}9\}$).

  Disadvantages: multiplication is messy to define; also, real numbers may have two distinct decimal expansions, so we need to use equivalence classes.

- Equivalence classes of Cauchy sequences of rationals.

  Advantage: defining addition and multiplication is very easy.

- Dedekind cuts.

  Advantages: defining addition is easy, as is proving the existence of suprema; also no equivalence classes needed.

We take the third approach here.

A *Dedekind cut* is a subset $X \subset \mathbb{Q}$ such that

(i) $\emptyset \neq X \neq \mathbb{Q}$;

(ii) X is *downward closed* – that is,

$$q \in X \wedge r < q \to r \in X$$

(iii) X has no largest member – that is,

$$\neg\exists m \in X : \forall x \in X : x < m$$

(A Dedekind cut may also alternatively be defined to be the pair $\langle X{,}\mathbb{Q} \setminus X\rangle$, but X alone completely determines the pair, so no information is lost by just considering the first component.)

The *set of real numbers* $\mathbb{R}$ is the set of all Dedekind cuts.

We define the relation $<_{\mathbb{R}}$ on $\mathbb{R}$ by $x <_{\mathbb{R}} y$ if and only if $x \subset y$, and the relation $\leq_{\mathbb{R}}$ by $x \leq_{\mathbb{R}} y$ if and only if $x \subseteq y$.

Theorem 8.16. *The relation $<_{\mathbb{R}}$ is a linear ordering on $\mathbb{R}$.*

Proof. Irreflexivity and transitivity follows from that of the strict subset relation. Obviously, at most one of

$$x \subset y, \qquad x = y, \qquad y \subset x$$

can hold. To show at least one holds, suppose the first two cases fail, so $x \not\subseteq y$.

Since $x \not\subseteq y$, there exists a rational $r \in x \setminus y$. Now, let $q \in y$. If $q \geq r$, then $r \in y$, as y is downward closed, but $r \notin y$ by definition, so $q < r$. Since x is downward closed, $q \in x$. So, $y \subset x$. ■

8.10.1 Bounds

- A number $u \in \mathbb{R}$ is an *upper bound* of a set $A \subseteq \mathbb{R}$ if $a \leq_{\mathbb{R}} u$ for all $a \in A$.
- The set $A \subseteq \mathbb{R}$ is *bounded from above* if there exists an upper bound of A.
- A *least upper bound* is an upper bound less than any other upper bound.

Theorem 8.17. *Any non-empty subset of $\mathbb{R}$ that is bounded from above has a least upper bound.*

Proof. Let $A \subseteq \mathbb{R}$ be non-empty and bounded from above. We claim that $\bigcup A$ is a Dedekind cut, and is the least upper bound of A.

Since A is a collection of Dedekind cuts, which are sets of rational numbers, we have $\bigcup A \subseteq \mathbb{Q}$. Since A is non-empty, $\bigcup A \neq \emptyset$, and since A is bounded above, there exists an upper bound, say u, so $u + 1 \notin A$, and $A \neq \mathbb{Q}$.

Let $a \in \bigcup A$ and $r \in \mathbb{Q}$ such that $r < a$. Because $a \in \bigcup A$, there exists a Dedekind cut $x \in A$ such that $a \in x$. Because x is a Dedekind cut, it is downwards closed, so $r \in x$, and hence $r \in A$. So $\bigcup A$ is downward closed.

Now, let $m \in \bigcup A$, so there exists $x \in A$ such that $m \in x$. If m is a largest element of $\bigcup A$, then it would also be the largest element of x. But x is a Dedekind cut, which has no largest element.

Hence, $\bigcup A$ is a Dedekind cut. For all $x \in A$, we have $x \subseteq \bigcup A$ (that is, $x \leq_{\mathbb{R}} \bigcup A$), so $\bigcup A$ is an upper bound of A. Now, let z be any upper bound of A, so $x \leq_{\mathbb{R}} z$ ($x \subseteq z$) for every $x \in A$. Then, $\bigcup A \subseteq z$ ($\bigcup A \leq_{\mathbb{R}} z$), so $\bigcup A$ is the least upper bound. ■

For any $x,y \in \mathbb{R}$, we define the set

$$x +_{\mathbb{R}} y = \{p + q \in \mathbb{Q} : q \in x, r \in y\}$$

This coincides with our usual idea of addition, since, for example, if

$$\begin{aligned} x &= \{q \in \mathbb{Q} : q < 1\} \\ y &= \{q \in \mathbb{Q} : q < 3\} \end{aligned}$$

then

$$\begin{aligned} x +_{\mathbb{R}} y &= \{a + b \in \mathbb{Q} : a \in x, b \in y\} \\ &= \{a + b \in \mathbb{Q} : a < 1, b < 3\} \\ &= \{q \in \mathbb{Q} : q < 4\} \end{aligned}$$

Lemma 8.18. *If $x,y \in \mathbb{R}$, then $x +_{\mathbb{R}} y \in \mathbb{R}$.*

Proof. Clearly, $x +_{\mathbb{R}} y \subseteq \mathbb{Q}$. If $a,b \in \mathbb{Q}$ such that $a \notin x$ and $b \notin y$, then for every $p \in x$ and $q \in y$, $p < a$ and $q < b$, so every element $(p + q) \in x +_{\mathbb{R}} y$ is less than $a + b$, so $a + b \notin x +_{\mathbb{R}} y$, giving $x +_{\mathbb{R}} y \neq \mathbb{Q}$.

Let $a < (p + q) \in x +_{\mathbb{R}} y$. Then, $a - q < p$, so $(a - q) \in x$ by downward-closedness of x as a Dedekind cut. Then, $a = (a - q) + q \in x +_{\mathbb{R}} y$, so $x +_{\mathbb{R}} y$ is downward closed.

Suppose $p + q \in x +_{\mathbb{R}} y$ is the largest element. As x is a Dedekind cut, p is not the largest element in x, so there exists a larger element $p < p' \in x$. Then, $p + q < p' + q \in x +_{\mathbb{R}} y$, contradicting that $p + q$ was the largest.

Thus, $x +_{\mathbb{R}} y$ is a Dedekind cut. ■

Theorem 8.19 (Basic Properties of $\mathbb{R}$)**.** *Replacing the relevant operations and constants by* $(+_{\mathbb{R}}, \cdot_{\mathbb{R}}, 0_{\mathbb{R}}, 1_{\mathbb{R}})$*, the same results as in Theorem 8.6, Theorem 8.12, and Theorem 8.15 for* ω*,* $\mathbb{Z}$*, and* $\mathbb{Q}$ *hold for* $\mathbb{R}$*.*

Again, $\mathbb{Q}$ is not a true subset of $\mathbb{R}$, but there is a natural embedding $E : \mathbb{Q} \to \mathbb{R}$ defined by $E(r) = \{q \in Q : q < r\}$ such that $E(\mathbb{Q})$ behaves like $\mathbb{Q}$:

$$\begin{aligned} E(a +_{\mathbb{Q}} b) &= E(a) +_{\mathbb{R}} E(b); \\ E(a \cdot_{\mathbb{Q}} b) &= E(a) \cdot_{\mathbb{R}} E(b); \\ E(a) <_{\mathbb{R}} E(b) &\leftrightarrow a <_{\mathbb{Q}} b; \\ E(0_{\mathbb{Q}}) &= 0_{\mathbb{R}}; \\ E(1_{\mathbb{Q}}) &= 1_{\mathbb{R}}. \end{aligned}$$

(That is, E is a order semiring homomorphism.)

8.11 Complex Numbers

A complex number consists of a real part, and an imaginary part, which is just a real number scaling the imaginary unit. As such, complex numbers can easily be represented using ordered pairs: $\mathbb{C} = \mathbb{R} \times \mathbb{R}$, with addition and multiplication defined as usual:

$$\begin{aligned} \langle a,b \rangle +_{\mathbb{C}} \langle c,d \rangle &= \langle a + c, b + d \rangle \\ \langle a,b \rangle \cdot_{\mathbb{C}} \langle c,d \rangle &= \langle ac - bd, ac + bd \rangle \end{aligned}$$

Again, $\mathbb{R}$ is not a true subset of $\mathbb{C}$ – for instance, $1_{\mathbb{C}} = \langle 1_{\mathbb{R}}, 0_{\mathbb{R}} \rangle$ – but, as usual, there is a natural embedding that lifts one to the other.

9 Cardinality

Informally, the cardinality of a set means the “size” of that set, in the sense of how many elements it has.

Two sets A and B are *equinumerous* if there exists a bijection between A and B, and we denote this relation by $A \sim B$.

Equinumerosity is an “equivalence relation” since,

- For all A, $A \sim A$ via the identity on A;
- For all A,B, $A \sim B \to B \sim A$, as a bijection between A and B is also a bijection between B and A;
- For all A,B,C, $(A \sim B \wedge B \sim C) \to A \sim B$ via composition.

However, equinumerosity is not a relation in the sense that it is not a set: it is a proper class, consisting of pairs of sets that are equinumerous.

We say that two sets A and B have the *same cardinality*, written as $|A| = |B|$, if they are equinumerous.

For comparing these cardinalities, we have several options. We could say that a non-empty set A is "smaller than" (or is "at most as large as") another set B if:

(i) A is equinumerous to a subset of B. That is, $A \sim B_0 \subseteq B$;

(ii) There exists an injection $A \rightarrowtail B$;

(iii) There exists an surjection $A \twoheadrightarrow B$.

The first two are equivalent, and they also imply the third, but the third only implies the first two with the axiom of choice.

We say that the cardinality of A is at most the cardinality of B, written as $|A| \leq |B|$ if there is an injective function from A to B.

Theorem 9.1. *If $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.*

Proof. Let $f : A \to B$ and $g : B \to C$ be injective. Then, $g \circ f : A \to C$ is injective. ■

Theorem 9.2 (Cantor-Bernstein)**.** *Let A and B be sets. If there is an injection $f : A \to B$ and an injection $g : B \to A$, then there is a bijection between A and B. That is, if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Proof (König, 1906). Without loss of generality that A and B are disjoint (any elements in the intersection can be paired with their copy in the other set, and hence ignored).

For any $a \in A$, we may consider its orbit,

$$\cdots \to f^{-1}\big(g^{-1}(a)\big) \to g^{-1}(a) \to a \to f(a) \to g\big(f(a)\big) \to \cdots$$

This sequence may terminate at some point to the left, if f^{-1} or g^{-1} is not defined.

Because f and g are injective, each $a \in A$ and $b \in B$ is in exactly one such sequence, since if an element $a \in A$ occurs in two sequences, the following and preceding elements are just functions applied to that element, so the two sequences must agree. Therefore, the sequences partition the disjoint union of A and B, so it is sufficient to give a bijection between the elements of A and B in each sequence separately.

Call a sequence *A-terminating* if it terminates to the left because g^{-1} cannot be taken at a certain point. That is, the sequence begins with an element in A. Define *B-terminating* sequences similarly.

Then, for an A-terminating sequence, f is a bijection between its A-elements and its B-elements, and similarly, for a B-terminating sequence, g is a bijection between its A-elements and its B-elements. For a doubly infinite or cyclic sequence, both f and g provide bijections.

To explicitly give a bijection $A \to B$, define the set $A_0 = A \setminus g(B)$. Note that every A-terminating sequence starts with an element in A_0, or else g^{-1} would be defined for its starting element.

Then, recursively define $A_n = g(f(A_{n-1}))$, and define A' to be their union:

$$\begin{aligned} A' &= \bigcup_{n \in \mathbb{N}} A_n \\ &= A_0 \cup g(f(A_0)) \cup g(f(g(f(A_0)))) \cup \cdots \end{aligned}$$

That is, A' is the orbit of A_0 under $g \circ f$.

Then,

$$h(x) := \begin{cases} f(x) & x \in A' \\ g^{-1}(x) & x \notin A' \end{cases}$$

is a bijection $A \to B$. ■

Corollary 9.2.1. $[0,1] \sim [0,1)$. *That is, the closed unit interval* $[0,1]$ *is equinumerous with the half-open unit interval* $[0,1)$.

Proof. The functions $f : [0,1] \to [0,1)$ and $g : [0,1) \to [0,1]$ defined by $f(x) = x/2$ and $g(x) = x$ are injections. Then, we have,

$$\begin{aligned} A_0 &= [0,1] \setminus g\big([0,1)\big) \\ &= [0,1] \setminus [0,1) \\ &= \{1\} \\ A_n &= \left\{\frac{1}{2^n}\right\} \\ A' &= \bigcup_{n\in\mathbb{N}} A_n \\ &= \left\{\frac{1}{2^n} : n \in \mathbb{N}\right\} \end{aligned}$$

So

$$h(x) := \begin{cases} \frac{x}{2} & x \in A' \\ x & x \notin A' \end{cases}$$

is a bijection $[0,1] \to [0,1)$. ■

We recall Cantor's theorem and state a new corollary related to cardinality.

Theorem (Cantor)**.** *For any set A, there is no surjection* $f : A \to \mathcal{P}(A)$.

Corollary 9.2.2 (Cantor)**.** *For every set A, we have* $|A| < |\mathcal{P}(A)|$.

Proof. The function $f : A \to \mathcal{P}(A)$ defined by $a \mapsto \{a\}$ is an injection, so $|A| \leq |\mathcal{P}(A)|$, but Cantor's theorem states that there is no surjection $A \to \mathcal{P}(A)$, which implies that no bijection $A \to \mathcal{P}(A)$ may exist, so $|A| \neq |\mathcal{P}(A)|$. ■

For any sets A and B, is it true that at least one of $|A| \leq |B|$ and $|B| \leq |A|$ holds? This is surprisingly non-trivial, and is in fact equivalent to the axiom of choice. We prove this later.

9.1 Finite Sets

How do we determine if a set is finite or not? We cannot directly write a first order formula that states that a set X has finitely many elements. Fortunately, in our definition of the natural numbers, each set $n \in \omega$ is defined to be the set $\{0,1,\ldots,n-1\}$, so it "has n elements".

A set is *finite* if it is equinumerous to a natural number. This can be written as a formula for any set X as

$$\exists n \exists f \Big(\underbrace{n \in \omega \land f \subseteq X \times n}_{f \text{ is a function } X \to n}$$

$$\wedge \Big(\underbrace{\forall x \in X : \forall y \in X : \forall a \in n\big((\langle x,a\rangle \in f \wedge \langle y,a\rangle \in f) \to (\langle x,a\rangle = \langle y,a\rangle \to x = y)\big)}_{f \text{ is injective}}\Big)$$

$$\wedge \Big(\underbrace{\forall m : m \in n \to \exists x(x \in X \wedge \langle x,m\rangle \in f)}_{f \text{ is surjective}}\Big)\Big)$$

A finite set X *has cardinality* n or *has* n *elements* if there is a bijection from X to n, and we write $|X| = n$ in this case.

Theorem (Pigeonhole Principle)**.** *No natural number is equinumerous to a proper subset of itself.*

Lemma 9.3. *Let X be finite. Then, there exists a unique $n \in \omega$ such that $|X| = n$.*

Proof. By the definition of finiteness, there exists at least one such natural number. For uniqueness, suppose $|X| = n$ and $|X| = m$, so there exist bijections $f : X \to n$ and $g : X \to n$. Suppose further that $m < n$. Then $g \circ f^{-1}$ is a bijection $n \to m$, so n is equinumerous to a proper subset of itself, contradicting the pigeonhole principle. ■

Corollary 9.3.1. *No finite set is equinumerous to a proper subset of itself.*

Proof. Let X be finite, so there exists a unique $n \in \omega$ such that $|X| = n$, so there is a bijection $f : X \to n$. Let $Y \subset X$ be a proper subset, and suppose that $X \sim Y$ (X and Y are equinumerous), so there exists a bijection $g : X \to Y$. Then, $f \circ g \circ f^{-1}$ is a bijection $n \to f(Y) \subsetneq n$, contradicting the pigeonhole principle. ■

A set is *infinite* if it is not finite. Note that if X is finite and $Y \sim X$, then Y is also finite. Similarly, if X is infinite, and $Y \sim X$, then Y is also infinite.

Theorem 9.4. *ω is infinite.*

Proof. Let $s : \omega \to \omega$ be the function $s(n) = n^+$. Then, $\operatorname{ran}(s) = \omega \setminus \{0\} \subset \omega$, so ω is equinumerous to a proper subset of itself, so ω is not finite. ■

9.1.1 Dedekind Finiteness

There are other possible notions of finiteness.

A set X is *Dedekind finite* if no proper subset of X is equinumerous to X.

The pigeonhole principle implies that every finite set is Dedekind finite, but is the converse true? Yes, but this direction requires the axiom of choice.

Theorem 9.5.

(*i*) *Finite sets are Dedekind finite.*

(*ii*) *(AC) Dedekind finite sets are finite.*

Proof.

(*i*) Follows from the pigeonhole principle.

(*ii*) (*Proof sketch.*) Let A be an infinite set, and let $a_0 \in A$. Then, choose $a_1 \in A\setminus\{a_0\}$, $a_2 \in A\setminus\{a_0,a_1\}$, and so on. Since A is infinite, this process can continue forever.

Now define a function $f : A \to A$ such that $f(a_n) = a_{n+1}$, and $f(a) = a$ for any $a \notin \{a_n\}_{n\in\mathbb{N}}$. Then, f is injective, but not surjective as $a_0 \notin \operatorname{ran}(f)$, so A is equinumerous to $A \setminus \{a_0\} \subset A$.

■

9.2 Countability

A set X is *countable* if there is an injective function $f : X \to \omega$.

Clearly, finite sets are countable.

Because injections imply surjective inverses (Theorem 7.6), equivalently, a set X is countable if is empty or if there is a surjection from ω to X. We have a converse theorem (Theorem 7.7), so the previous implication is actually biconditional, but this theorem requires the axiom of choice.

However, it turns out that we can prove this result for ω without the axiom of choice.

Theorem 9.6. *A set X is countable if and only if there is a surjection $g : \omega \to X$, or if X is empty.*

Proof. Suppose X is countable, so there exists an injective function $f : X \to \omega$, so by Theorem 7.6, there exists a surjection $g : \omega \to X$, or X is empty.

If X is empty, then the unique empty function $(X = \emptyset) \to \omega$ is vacuously injective, so X is countable. Suppose otherwise that there is a surjection $g : \omega \to X$. Define $f : X \to \omega$ by

$$f(x) = \min\Big(g^{-1}\big[\{x\}\big]\Big) \subseteq \omega$$

The set $g^{-1}\big[\{x\}\big]$ is non-empty as g is surjective, and the minimum element exists as ω is well-ordered (Theorem 8.9). The preimages are also disjoint for distinct inputs (or else anything in the intersection has multiple images under g), so f is injective, as required. ■

The existence of such a surjection is often phrased as "the elements of X can be listed in a sequence", since a sequence in X is just a function $\omega \to X$.

Example. The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{N} \times \mathbb{N}$ are countable.

By the above theorem, it is enough to enumerate the elements of each set in a sequence:

- $\mathbb{N}$: 0,1,2,3,…
- $\mathbb{Z}$: 0,1,−1,2,−2,3,−3,…
- $\mathbb{Q}$: A listing of all the positive rationals was famously given by Cantor:

$$\begin{array}{cccccccccc}
\frac{1}{1} & \to & \frac{1}{2} & & \frac{1}{3} & \to & \frac{1}{4} & & \frac{1}{5} & \to \\
 & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \cdots \\
\frac{2}{1} & & \frac{2}{2} & & \frac{2}{3} & & \frac{2}{4} & & \frac{2}{5} & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & & & \cdots \\
\frac{3}{1} & & \frac{3}{2} & & \frac{3}{3} & & \frac{3}{4} & & \frac{3}{5} & \cdots \\
 & \swarrow & & \nearrow & & & & & & \cdots \\
\frac{4}{1} & & \frac{4}{2} & & \frac{4}{3} & & \frac{4}{4} & & \frac{4}{5} & \cdots \\
\downarrow & \nearrow & & & & & & & & \cdots \\
\frac{5}{1} & & \frac{5}{2} & & \frac{5}{3} & & \frac{5}{4} & & \frac{5}{5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Cantor's zig-zag argument

This lists all the positive rationals, but to include the negative rationals, we can interleave them like with the integers (and also add 0 to the beginning of the list).

- $\mathbb{N} \times \mathbb{N}$: Like the grid above, but listing (p,q) instead of $\frac{p}{q}$.

To distinguish between finite sets and infinite countable sets, we call a set *countably infinite* if it is infinite and countable.

Lemma 9.7. *Let $A \subseteq \omega$ be infinite. Then, there is a bijection $f : A \to \omega$*

Proof sketch. Intuitively, we match the smallest element of A with 0, the second smallest with 1, and so on. This does not require the axiom of choice, since ω (and hence A) is well-ordered, so no arbitrary choice functions are required, and the recursion theorem then gives the required function. ■

Theorem 9.8. *A set X is countably infinite if and only if there is a bijection between X and ω.*

Proof. Suppose there is a bijection $f : X \to \omega$, so $X \sim \omega$. Since ω is infinite, X is infinite, and since f is injective, X is countable.

Now suppose that X is countably infinite. Since X is countable, there is an injection $f : X \to \omega$. Consider $A = f(X)$. Since $f : X \to A$ is a bijection, $X \sim A$, and since X is infinite, $A \subseteq \omega$ is infinite, so there is a bijection $h : A \to \omega$ by the previous lemma. The composition $h \circ f$ is then a bijection between X and ω. ■

If X is countably infinite, then we say that its cardinality is $|X| = \aleph_0$.

Corollary 9.8.1. *If X and Y are countably infinite, then $X \sim Y$. That is, there exists a bijection XY.*

Proof. By the above theorem, $X \sim \omega$ and $Y \sim \omega$, so $X \sim Y$. ■

Theorem 9.9. *If A and B are countable, then*

(*i*) *$A \cup B$ is countable;*

(*ii*) *$A \times B$ is countable.*

Proof.

(*i*) If A or B are empty, then the union is just one of the sets, which is countable, so suppose otherwise that A and B are both non-empty.

Enumerate A as $a_1,a_2,\dots$ and B as $b_1,b_2,\dots$, possibly listing each element multiple times. Then, $A \cup B$ may be enumerated as $a_1,b_1,a_2,b_2,\dots$.

(*ii*) If A or B are empty, then the product is empty, which is countable, so suppose otherwise that A and B are both non-empty.

As A and B are countable, there exist surjections $f : \mathbb{N} \to A$ and $g : \mathbb{N} \to B$. Define $h : \mathbb{N} \times \mathbb{N} \to A \times B$ by

$$h\big(\langle a,b\rangle\big) = \big\langle f(a),g(b)\big\rangle$$

■

By induction, the union and product of finitely many countable sets is countable.

Is it true that the union of *countable many* countable sets is countable? Yes, but surprisingly, this requires the axiom of choice.

Theorem 9.10. *(AC) The union of countably many countable sets is countable. That is, if X is countable and $A \in X$ is countable for all A, then $\bigcup X$ is countable.*

Proof sketch. On a grid, enumerate each A_n on a horizontal line. Then, Cantor's zig-zag argument applies.

Slightly more formally, without loss of generality assume that $\emptyset \notin X$, and fix a surjection $f : \mathbb{N} \to X$. Each $A = f(n)$ is a non-empty countable set, so choose a surjection $g_n : \mathbb{N} \to A$, and let $h : \mathbb{N}\times\mathbb{N} \to \bigcup X$ be defined by

$$h(n,m) = g_n(m)$$

This function is a surjection as f and g_n are surjections.

Now, let $i : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be a surjection. Then $h \circ i : \mathbb{N} \to \bigcup X$ is a surjection, so $\bigcup X$ is countable. ■

In more detail, the axiom of choice is required when choosing g_n:

Define R to be the set of ordered pairs $\langle n,g\rangle$ $n \in \mathbb{N}$ such that g is a surjection $\mathbb{N} \to f(n)$. Because $f(n)$ is countable for all n, such a surjection always exists, so every n is in the domain of R, so $\operatorname{dom}(R) = \mathbb{N}$.

By the first version of the axiom of choice, there is a function $G \subseteq R$ with the same domain, $\mathbb{N}$. That is, for each n, there is a unique g_n such that $\langle n,g_n\rangle \in G$. This gives the surjection from $\mathbb{N}$ to $f(n)$, and the proof from this point is identical to the one above.

A set is *uncountable* if it is not countable, and we write $|X| > \aleph_0$.

9.3 Continuum

In set theory, *the continuum* refers to the size of $\mathbb{R}$.

Theorem 9.11 (Cantor's Diagonal Argument)**.** $\mathbb{R}$ *is not countable.*

Proof. This is Cantor's original diagonal argument.

Every real number can be uniquely expressed in base 10 as a series

$$d_0.d_1d_2d_3\ldots = n + \sum_{i=0}^{\infty} \frac{d_i}{10^i}$$

where d_0 is an integer, and d_i, $i > 0$, is a digit from 0 to 9, and the sequence (d_i) is not eventually all 9s.

Suppose that $\mathbb{R}$ is countable, so there is a list containing all real numbers

$$d_0^i.d_1^id_2^id_3^i\ldots$$

Let $r = d'_0.d'_1d'_2d'_3\ldots$ where

$$d'_i = \begin{cases} 1 & d_i = 0 \\ 0 & d_i \neq 0 \end{cases}$$

Then r is a real number not listed, as it differs from the ith real number at the ith digit. ■

Or, as a table, the list of real numbers is given by:

i	d_0	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	$\cdots$
0	0	1	2	3	4	5	6	7	8	9	$\cdots$
1	1	4	1	4	2	1	3	5	6	2	$\cdots$
2	3	1	4	1	5	9	2	6	5	3	$\cdots$
3	1	3	7	0	3	5	9	9	0	8	$\cdots$
4	1	6	1	8	0	3	3	9	8	8	$\cdots$
5	0	1	1	2	3	5	8	3	1	4	$\cdots$
6	0	1	4	2	8	5	7	1	4	2	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

and the new real number can be generated by examining the diagonal entries of the table, giving this proof its name:

i	d_0	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	$\cdots$
0	0	1	2	3	4	5	6	7	8	9	$\cdots$
1	1	4	1	4	2	1	3	5	6	2	$\cdots$
2	3	1	4	1	5	9	2	6	5	3	$\cdots$
3	1	3	7	0	3	5	9	9	0	8	$\cdots$
4	1	6	1	8	0	3	3	9	8	8	$\cdots$
5	0	1	1	2	3	5	8	3	1	4	$\cdots$
6	0	1	4	2	8	5	7	1	4	2	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
?	1	0	0	1	1	0	$\cdots$				

If $A \sim \mathbb{R}$, then we say that A has cardinality of the continuum, and we write $|A| = \mathfrak{c}$. Cantor's theorem above can then be written as $|\mathbb{R}| = \mathfrak{c} > \aleph_0$.

Theorem 9.12. $\mathbb{R} \sim (0,1)$. *That is, there is a bijection between* $\mathbb{R}$ *and the open unit interval* $(0,1)$.

Proof. The logistic function $\sigma : \mathbb{R} \to (0,1)$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

with inverse $f(x) = -\ln(\frac{1}{x-1})$ provides the required bijection. ■

Lemma 9.13. $[0,1) \sim (0,1)$. *That is, the half-open unit interval* $[0,1]$ *is equinumerous with the open unit interval* $(0,1)$.

Proof. We have previously proved that $[0,1] \sim [0,1)$ (Corollary 9.2.1). The bijection here is similar:

$$f(x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2^{n+1}} & x = \frac{1}{2^n} \\ x & \text{otherwise} \end{cases}$$

■

So, we have $[0,1] \sim [0,1) \sim (0,1) \sim \mathbb{R}$, so these sets all have the same cardinality. Combining with scaling transformations, this implies that all non-trivial intervals of $\mathbb{R}$ are equinumerous to $\mathbb{R}$ and thus have the cardinality of the continuum.

Corollary 9.13.1. *If* $S \subseteq \mathbb{R}$ *contains an open interval, then* $|S| = \mathfrak{c}$.

Proof. A suitable scaling map map injects (0,1) into the open interval in S, so there is an injection from $\mathbb{R} \sim (0{,}1)$ into S. Also, there is an injection from S into $\mathbb{R}$ given by the inclusion map. The Cantor-Bernstein theorem then implies $|S| = |\mathbb{R}| = \mathfrak{c}$. ■

Theorem 9.14. $[0{,}1] \sim [0{,}1]^2$

Corollary 9.14.1. $|\mathbb{R}^2| = \mathfrak{c}$

So far, we know the cardinalities $0{,}1{,}2,\ldots,\aleph_0,\mathfrak{c}$, and Cantor's theorem implies the existence of infinitely many infinite cardinalities given by iterated power sets:

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}\mathcal{P}(\mathbb{N})| < |\mathcal{P}\mathcal{P}\mathcal{P}(\mathbb{N})| < \ldots$$

Where does $\mathfrak{c}$ lie in this infinite chain?

Theorem 9.15. $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$

Proof. For each subset $A \subseteq \mathbb{N}$, define x_A to be the real number whose decimal expansion is

$$0.d_1d_2d_3\ldots$$

where

$$d_n = \begin{cases} 1 & n \in A \\ 0 & n \notin A \end{cases}$$

The function $A \mapsto x_A$ injects $\mathcal{P}(\mathbb{N})$ into $\mathbb{R}$, so $|\mathcal{P}(\mathbb{N})| \leq |\mathbb{R}|$.

Now, given $x \in (0{,}1) \subset \mathbb{R}$, write x in its unique binary expansion

$$0.b_0b_1b_2\ldots$$

where the sequence (b_i) is not eventually all 1s, and define the set

$$A_x := \{n \in \mathbb{N} : x_n = 1\}$$

Then, $x \mapsto A_x$ injects $(0{,}1) \sim \mathbb{R}$ into $\mathcal{P}(\mathbb{N})$, so $|\mathbb{R}| = |(0{,}1)| \leq |\mathbb{P}(\mathbb{N})|$. ■

9.3.1 Transcedental Numbers

A real number is *algebraic* if it is the root of a polynomial with integer (or rational) coefficients. A real number is *transcendental* if it is not algebraic.

It is very difficult to construct explicit examples of transcendental numbers, but surprisingly, most real numbers are transcendental:

Theorem 9.16. *The set of algebraic numbers is countable.*

Proof. A polynomial $p \in \mathbb{Z}[x]$ of degree n has the form

$$p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

with $a_n,\ldots,a_0 \in \mathbb{Z}$. Define the *height* of p to be the natural number

$$n + \sum_{i=0}^{n} |a_i|$$

Note that there are only finitely many polynomials with any given fixed height, since there are finitely many integer partitions of any natural number. Every non-zero non-constant polynomial also has at most n (i.e. finitely many) roots. So, for each height $h \in \mathbb{N}$, there are finitely many roots of polynomials of height h.

Therefore, these roots may be enumerated by enumerating the roots of polynomials of height 1, then height 2, etc. ■

Corollary 9.16.1. *There exist transcendental numbers. Furthermore, the set of transcendental numbers is not countable.*

Proof. Let A be the set of algebraic numbers and $T = \mathbb{R} \setminus A$ be the set of transcendental numbers. If T were countable, then $A \cup T = \mathbb{R}$ is countable as the union of two countable sets. But $\mathbb{R}$ is not countable. ■

9.4 Cardinal Arithmetic

If κ and λ are cardinalities, then we define $\kappa + \lambda$ to be the cardinality of any set $S = A \cup B$ where $|A| = \kappa$, $|B| = \lambda$, and $A \cap B = \emptyset$.

For natural numbers (i.e. elements of ω), this definition is equivalent to the previous definition of addition, $+_\omega$.

If κ and λ are cardinalities, then we define $\kappa \cdot \lambda$ to be the cardinality of any set $S = A \times B$ where $|A| = \kappa$, $|B| = \lambda$.

Example.

- $1 + \aleph_0 = \aleph_0$, given by $f : \{-1\} \cup \mathbb{N} \to \mathbb{N} : x \mapsto x + 1$.
- $\aleph_0 + \aleph_0 = \aleph_0$: $\mathbb{N} \sim E := \{n : \exists k \in \mathbb{N} : n = 2k\}$ and $\mathbb{N} \sim O := \{n : \exists k \in \mathbb{N} : n = 2k + 1\}$ via $x \mapsto 2x$ and $x \mapsto 2x + 1$, respectively, and $E \cup O = \mathbb{N}$.
- $2 \cdot \aleph_0 = \aleph_0$: $\mathbb{N} \times \mathbb{N}$ is countably infinite, and hence bijects to ω.
- $\mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c}$.
- $\aleph_0 \cdot \mathfrak{c} = \mathfrak{c}$.

Theorem 9.17 (Basic Properties of Cardinal Arithmetic)**.** *The following all hold:*

(i) $\forall \kappa,\lambda : \kappa + \lambda = \lambda + \kappa$ *(commutativity of cardinal addition);*

(ii) $\forall \kappa,\lambda,\mu : (\kappa + \lambda) + \mu = \kappa + (\lambda + \mu)$ *(associativity of cardinal addition);*

(iii) $\forall \kappa : \kappa + 0 = \kappa$ *(existence of additive identity);*

(iv) $\forall \kappa,\lambda : \kappa \cdot \lambda = \lambda \cdot \kappa$ *(commutativity of multiplication);*

(v) $\forall \kappa,\lambda,\mu : (\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$ *(associativity of multiplication);*

(vi) $\forall \kappa : \kappa \cdot 1 = \kappa$ *(existence of multiplicative identity);*

(vii) $\forall \kappa,\lambda,\mu : \kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$ *(distributivity of multiplication over addition);*

(viii) $\forall \kappa,\kappa',\lambda,\lambda$ *if* $\kappa \leq \kappa'$ *and* $\lambda \leq \lambda'$*, then* $\kappa + \lambda \leq \kappa' + \lambda'$ *and* $\kappa \cdot \lambda \leq \kappa' \cdot \lambda'$ *(weakly monotone);*

Proof. All trivial from basic properties of unions and products. For instance, distributivity holds because

$$A \times (B \cup C) = (A \times B) \cup (A \times C)$$

■

These cardinal operations are not, however, *strictly* monotone, as $1 + \aleph_0 = 2 + \aleph_0$, and $1 \cdot \aleph_0 = 2 \cdot \aleph_0$.

Next, we define cardinal exponentiation.

Let A and B be sets. We denote the set of functions $B \to A$ by ${}^{B}A$ or A^B. That is,

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

Note that for finite sets, A^B has $|A|^{|B|}$ many elements.

If κ and λ are cardinalities, then we define κ^λ to be the cardinality of A^B where $|A| = \kappa$ and $|B| = \lambda$.

Theorem 9.18 (Basic Properties of Cardinal Exponentiation)**.** *For all cardinalities κ,λ,μ, the following hold:*

(i) $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$;

(ii) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$;

(iii) $(\kappa^\lambda)^\mu = \kappa^{\lambda\cdot\mu}$.

Lemma 9.19. *For any cardinality κ, $2^\kappa = |\mathcal{P}(S)|$ for any set S with cardinality $|S| = \kappa$.*

Proof. Let S have cardinality κ. The set $\{0,1\}$ has cardinality 2, so $2^\kappa = |\{0,1\}^S|$ by definition. Let $f : S \to \{0,1\}$, and define the set

$$A_f := \{s \in S : f(s) = 1\}$$

Then, $f \mapsto A_f$ is a bijection from $\{0,1\}^S$ to $\mathcal{P}(S)$. ■

Corollary 9.19.1. *For any cardinality κ, $\kappa < 2^\kappa$*

Proof. Apply Cantor's theorem to the previous lemma. ■

Corollary 9.19.2. $2^{\aleph_0} = \mathfrak{c}$

Proof. By the previous lemma,

$$2^{\aleph_0} = |\{0,1\}^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})| = |\mathbb{R}| = \mathfrak{c}$$ ■

Corollary 9.19.3. $\mathfrak{c}^{\aleph_0} = \mathfrak{c}$

Proof.

$$\mathfrak{c}^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0} = \mathfrak{c}$$ ■

Corollary 9.19.4. *The set of sequences of real numbers has cardinality equal to that of the continuum. That is, $|\mathbb{R}^{\mathbb{N}}| = |\mathbb{R}|$.*

Proof. By the previous corollary,

$$|\mathbb{R}^{\mathbb{N}}| = \mathfrak{c}^{\aleph_0} = \mathfrak{c} = |\mathbb{R}|$$ ■

10 Axiom of Choice

10.1 Equivalent Formulations

We recall the first form of the axiom of choice:

Axiom of Choice (first form).
For any relation R, there exists a function $F \subseteq R$ such that $\operatorname{dom}(F) = \operatorname{dom}(R)$.

We used this to prove that the existence of a surjection $B \to A$ implies the existence of an injection $A \to B$.

We now present another more intuitive form of the axiom of choice. Informally, given a collection of non-empty sets, there is a function that chooses one element from each set.

Axiom of Choice (second form).
Let $\mathcal{S}$ be a set with $\emptyset \notin \mathcal{S}$. Then, there is a *choice function* for $\mathcal{S}$. That is, a function $\sigma : \mathcal{S} \to \bigcup \mathcal{S}$ such that $\sigma(A) \in A$ for all $A \in \mathcal{S}$.

Example. Let $\mathcal{S} \subseteq \mathcal{P}(\mathbb{N}) \setminus \{\emptyset\}$ be a collection of non-empty sets of natural numbers. A choice function is a function $\sigma : \mathcal{S} \to \bigcup \mathcal{S} = \mathbb{N}$ that sends each set $A \in \mathcal{S}$ to an element $\sigma(A) = a \in A$.

One possible choice function is given by selecting the smallest number in each set. Even if $\mathcal{S}$ contains infinitely many sets, or even all possible sets of natural numbers, it is always possible to choose the smallest element from each set to produce a new set, because $\mathbb{N}$ is well-ordered.

In this case, we don't have to invoke the axiom of choice because we can construct this function explicitly

$$\sigma = \left\{ \langle A,n \rangle \in \mathcal{P}(\omega) \times \omega : A \in \mathcal{S}, n \in A, \forall m \in \omega : m <_\omega n \to m \notin A \right\}$$

However, for some sets, a choice function is not known. For instance, the set of all non-empty subsets of the real numbers does not admit such a choice function. In this case, the axiom of choice must be invoked.

As Russell once remarked, "*The Axiom of Choice is necessary to select a set from an infinite number of pairs of socks, but not an infinite number of pairs of shoes.*"

Axiom of Choice (third form).
Let $\mathcal{A}$ be a set of pairwise disjoint non-empty sets. Then, there exists a set C who has as a member exactly one element from each member of $\mathcal{A}$. That is, for each $B \in \mathcal{A}$, $|C \cap B| = 1$.

10.1.1 Infinite Cartesian Products

Let I be an set (the *indexing set*) and let H be a function whose domain includes I. Then, for each $i \in I$, we have a set $H(i)$. We define the *indexed product* of the $H(i)$ as

$$\underset{i \in I}{\times} H(i) := \{ f : f \text{ is a function with domain } I \text{ and } f(i) \in H(i) \text{ for all } i \in I \}$$

Note that, up to natural isomorphism, this agrees with our earlier definition of an iterated cartesian product. For instance, a binary cartesian product $X_1 \times X_2$ is the set of pairs $\langle x_1,x_2 \rangle$ with $x_1 \in X_1$ and $x_2 \in X_2$, but such a pair can be naturally identified with a function $x : \{1,2\} \to X_1 \cup X_2$, with $x(1) = x_1 \in X_1$ and $x(2) = x_2 \in X_2$.

This definition of an indexed product, however, makes sense even if the indexing set is not finite, or even countable.

Clearly, if any of the sets $H(i)$ is empty, then there are no such functions, and the entire product is empty. However, is it true that the product of non-empty sets is non-empty? This is again the axiom of choice:

Axiom of Choice (fourth form).
For any indexing set I and function H with domain I, if $H(i) \neq \emptyset$ for all $i \in I$, then

$$\mathop{\times}_{i\in I} H(i) \neq \emptyset$$

Theorem 10.1. *In the presence of the other ZF axioms, the first, second, third, and fourth forms of the axiom of choice are all equivalent.*

Proof of $(1) \leftrightarrow (2)$. Assume the first form holds. Let $\mathcal{S}$ be a set with $\emptyset \notin \mathcal{S}$, and define the relation

$$R = \left\{\langle x,y\rangle \in \mathcal{S} \times \bigcup \mathcal{S} : y \in x\right\}$$

Let $F \subseteq R$ be a function such that $\operatorname{dom}(F) = \operatorname{dom}(R)$, given by the first form. Then, for every $x \in \mathcal{S}$, $y = F(x) \in x$, so F is a choice function for $\mathcal{S}$.

Now assume the second form holds. The idea is that an arbitrary relation is like a multi-valued function, and a choice function given by the second form can choose for every x in the domain exactly one corresponding y.

Let R be any relation. Then, $R \subset \operatorname{dom}(R) \times \operatorname{ran}(R)$. Let

$$\mathcal{S} = \mathcal{P}\big(\operatorname{ran}(R)\big) \setminus \{\emptyset\}$$

Let $\sigma : \mathcal{S} \to \bigcup \mathcal{S}$ be a choice function given by the second form. That is, for every $A \in \mathcal{S}$ (so $\emptyset \neq A \subseteq \operatorname{ran}(R)$), we have $\sigma(A) \in A$. For $x \in \operatorname{dom}(R)$, let

$$F(x) = \sigma\Big(\big\{y \in \operatorname{ran}(R) : \langle x,y\rangle \in R\big\}\Big)$$

This defines a function $F : \operatorname{dom}(R) \to \operatorname{ran}(R)$ with $F \subseteq R$. ■

10.2 Partial Orders

A relation $R \subseteq X \times X$ is a (*weak* or *non-strict*) *partial order* on X if it is:

(i) reflexive: $\forall x \in X : xRx$;

(ii) transitive: $\forall x,y,z \in X : (xRy \wedge yRz) \to xRz$;

(iii) antisymmetric: $\forall x,y \in X : (xRy \wedge yRx) \to x = y$.

The usual notation for a weak partial order R is $\preceq_R$ or just $\preceq$.

A relation $R \subseteq X \times X$ is a *strict partial order* on X if it is:

(i) irreflexive: $\forall x \in X : \neg xRx$;

(ii) transitive: $\forall x,y,z \in X : (xRy \wedge yRz) \to xRz$;

(iii) asymmetric: $\forall x,y \in X : xRy \to \neg yRx$;

Asymmetry is implied by irreflexivity and transitivity, so it may be omitted from this definition.

The usual notation for a strict partial order R is $\prec_R$ or just $\prec$. If $a \prec b$, then we say that *a precedes b* or that *b covers a*. The pair $(X, \prec)$ is then called a *partially ordered set* or a *poset*.

Weak and strict partial orders are essentially the same notions, differing only by the diagonal elements $\langle x,x\rangle$: given a strict partial order, adding these pairs into it yields a corresponding weak partial order, and conversely, removing these pairs from a weak partial order yields a strict partial order.

Note that not all elements in a poset may be comparable under the ordering. If every pair of elements *are* comparable, then the ordering is *total*.

Example. Consider the set of all subsets of $\mathbb{N}$ with at most three elements ordered by inclusion, $\subseteq$. The sets $\{0\}$ and $\{1\}$ are incomparable under this relation because neither $\{0\} \subseteq \{1\}$ nor $\{1\} \subseteq \{0\}$ holds.

In a poset $(X, \preceq)$, an element $x \in X$ is *maximal* if for all $y \in X$, $x \preceq y$ only if $y = x$, or equivalently, x is maximal if there does not exist any y such that $x \prec y$.

A partial ordering may have any number of maximal elements, including none. For instance, the integers have no maximal element, while the set [0,1] has one maximal element, and a set with k mutually incomparable elements has k maximal elements.

This notion is distinct from that of a *maximum* element, which is an element $x \in X$ such that $y \preceq x$ for all $y \in X$. Clearly, a maximum element is maximal, and if it exists, it is unique.

Informally, a maximal element is an element that is not less than any other element, while a maximum element is an element that is greater than every other element.

Example. Consider the set of all subsets of $\mathbb{N}$ with at most three elements ordered by inclusion, $\subseteq$. The set $\{0,1,2\}$ is maximal because it is not a subset of any other set apart from itself, but it is not a maximum, because, for example, it is not a superset of $\{3\}$.

10.2.1 Zorn's Lemma

Let $\preceq$ be a weak partial order on a set Z.

A *chain* is a set $C \subseteq Z$ such that $\preceq$ is total on C. That is, every pair of elements in c are comparable under $\preceq$:

$$\forall c_1,c_2 \in C : c_1 \preceq c_2 \lor c_2 \preceq c_1$$

Clearly, every subset of a chain is itself a chain.

An element $x \in Z$ is an *upper bound* of a chain C if $c \preceq x$ for all $c \in C$.

Zorn's Lemma.
Let $(Z, \preceq)$ be a poset, and suppose that every chain $C \subseteq Z$ has an upper bound. Then, Z has a maximal element.

Although called a lemma, we normally treat this as an axiom, for it is equivalent to the axiom of choice.

Theorem 10.2. *In the presence of the other ZF axioms, Zorn's lemma is equivalent to the axiom of choice.*

Proof, forward direction only. We prove the first form of the axiom of choice, assuming Zorn's lemma.

Let R be any relation and define

$$Z = \{f \subseteq R : f \text{ is a function}\}$$

Z is partially ordered by inclusion, $\subseteq$. (Recall that $f \subseteq g$ if and only if g extends the function f).

Let $C \subseteq Z$ be a chain. We claim that $\bigcup C \in Z$, and that $\bigcup C$ is an upper bound of C. As the union of functions in R, $\bigcup C$ is a subset of R, and since $\bigcup C$ is the union of relations, it is itself a relation.

Suppose that $\langle x,y \rangle, \langle x,y' \rangle \in \bigcup C$. Then, there are functions $f,f' \in C$ such that $\langle x,y \rangle \in f$ and $\langle x,y' \rangle \in f'$. Since C is a chain, every function is comparable. Without loss of generality, suppose $f \subseteq f'$. Then, both pairs lie within f', so $y = y'$. Thus, $\bigcup C$ is a function, so it is in Z.

Now, for any $f \in C$, we have $f \subseteq \bigcup C$ by the definition of a union, so $\bigcup C$ is an upper bound for C.

Zorn's lemma then says that $(Z, \subseteq)$ has a maximal element F.

As $F \subseteq R$, $\mathrm{dom}(F) \subseteq \mathrm{dom}(R)$. Suppose for a contradiction that $\mathrm{dom}(F) \neq \mathrm{dom}(R)$, so there exists $x_0 \in \mathrm{dom}(R) \setminus \mathrm{dom}(F)$. Let y_0 be such that $\langle x_0,y_0 \rangle \in R$, and define

$$F' = F \cup \big\{ \langle x_0,y_0 \rangle \big\}$$

Clearly, $F' \subseteq R$ is a function, so $F' \in Z$. But then, $F \subsetneq F'$, contradicting that F is maximal in Z, so $\mathrm{dom}(R) = \mathrm{dom}(F)$, as required. ■

Theorem 10.3. *(AC) Every vector space has a basis.*

Proof. Let V be a vector space over a field K, and define the set

$$Z = \{ S \subseteq V : S \text{ is linearly independent over } K \}$$

Consider the partial order on Z given by $\subseteq$. The empty set is linearly independent, and $\emptyset \in Z$, so $Z \neq \emptyset$.

Let $C \subseteq Z$ be a chain. Clearly, $\bigcup C \subseteq V$.

Suppose

$$\sum_{i=1}^{n} k_i v_i = 0$$

for some vectors $v_1, \ldots, v_n \in \bigcup C$ and scalars $k_1, \ldots, k_n \in K$. Since $v_i \in \bigcup C$, there are $S_i \in C$ with $v_i \in S_i$, and since C is a chain, the S_i also form a chain. Without loss of generality, suppose the ordering is as follows:

$$S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n$$

so $v_1, \ldots, v_n \in S_n$. Since $S_n \in Z$ is linearly independent, $k_1 = \cdots = k_n = 0$, so $\bigcup C$ is linearly independent, and hence $\bigcup C \in Z$. $\bigcup C$ is also an upper bound for C since $S \subseteq \bigcup C$ for all $S \in C$.

By Zorn's lemma, there is a maximal element $S \in Z$. We claim S is a basis for V.

Since $S \in Z$, S is linearly independent. If $S = V$, then we are done. Otherwise, let $u \in V \setminus S$. Since S is maximal in Z, $S \cup \{u\}$, is not linearly independent, so

$$k_0 u + \sum_{i=1}^{n} k_i v_i = 0$$

for some vectors $v_1, \ldots, v_n \in S$ and scalars $k_0, \ldots, k_n \in K$ not all equal to 0. If $k_0 = 0$, then $\sum_{i=1}^{n} k_i v_i = 0$ with $k_1, \ldots, k_n$ not all zero, contradicting that S is linearly independent. So, $k_0 \neq 0$, and hence

$$u = -\frac{1}{k_0}(k_1 v_1 + \cdots + k_n v_n)$$

is in the linear span of S, so S is a basis for V. ■

Corollary 10.3.1. *$\mathbb{R}$ as a vector space over $\mathbb{Q}$ has a basis. That is, there is a set $H \subset \mathbb{R}$ such that every $x \in \mathbb{R}$ can be expressed as a unique linear combination*

$$x = \sum_{i=1}^{n} q_i x_i$$

of vectors $x_1, \dots x_n \in H$ and scalars $q_1, \dots q_n \in \mathbb{Q}$.

Such a basis is called a *Hamel basis.*

10.3 Cardinal Comparability

Recall that we write $|A| \leq |B|$ if there exists an injection $A \to B$. Is this ordering total on the class of all cardinals?

Cardinal Comparability.
For any sets A and B, we have $|A| \leq |B|$ or $|B| \leq |A|$. That is, there is an injective function $A \to B$ or there is an injective function $B \to A$.
Equivalently, for any two cardinals κ and λ, we have $\kappa \leq \lambda$ or $\lambda \leq \kappa$.

It turns out that cardinal comparability is again equivalent to the axiom of choice.

10.4 Absorption Law

So far, we have seen that cardinal arithmetic for finite cardinalities agrees with arithmetic on ω. That is, if $|A| = n$ (i.e. there exists a bijection between A and $n \in \omega$) and $|B| = m$ and A and B are disjoint, then $n + m = |A \cup B| = n +_\omega m$, where the addition on the left is cardinal addition. Similarly, if $|A| = n$ and $|B| = m$, then $n \cdot m = |A \times B| = n \cdot_\omega m$.

Theorem 10.4. *(AC) For every infinite cardinality κ, $\kappa \cdot \kappa = \kappa$. That is, for every infinite set X, there is a bijection $X \to X \times X$.*

Corollary (Absorption Law). *(AC) If κ or λ is infinite, then*

$$\kappa + \lambda = \max(\kappa,\lambda)$$

If one is infinite and the other is non-zero, then

$$\kappa \cdot \lambda = \max(\kappa,\lambda)$$

Proof. Suppose without loss of generality that $\kappa \geq \lambda$. Then,

$$\kappa \leq \kappa + \lambda \leq \kappa + \kappa \leq \kappa \cdot 2 \leq \kappa \cdot \kappa = \kappa$$

so $\kappa + \lambda = \kappa = \max(\kappa,\lambda)$. If additionally $\lambda \neq 0$, then,

$$\kappa \leq \kappa \cdot \lambda \leq \kappa \cdot \kappa = \kappa$$

so $\kappa \cdot \lambda = \kappa = \max(\kappa,\lambda)$. ■

11 Well-Ordered Sets

11.1 Linearly Ordered Sets

A binary relation $<_X$ on a set X is a (*weak* or *non-strict*) *total* or *linear order* if:

- For all $a,b \in X$, exactly one of $a <_X b$, $b <_X a$, and $a = b$ holds (trichotomy);
- For all $a,b,c \in X$, if $a <_X b$ and $b <_X c$, then $a <_X c$ (transitivity).

or equivalently, if $<_X$ is a partial order in which every pair of elements is comparable or equal.

Example. The *double line* is the set $X = \mathbb{R} \times \{0,1\}$ equipped with the *lexicographical ordering* $<_{\text{lex}}$ where $\langle r,i \rangle <_X \langle s,j \rangle$ if and only if $r < s$ or $r = s$ and $i < j$.

Two ordered sets $(X, <_X)$ and $(Y, <_Y)$ are *order-isomorphic*, written as $(X, <_X) \cong (Y, <_Y)$ if there is a bijection $f : X \to Y$ that compatible with the ordering. That is, $a <_X b$ if and only if $f(a) <_Y f(b)$.

Every finite totally ordered set is order-isomorphic to a subset of the natural numbers, and similarly, every countable totally ordered set is order-isomorphic to a subset of the rationals.

11.2 Well-Ordered Sets

A total ordering $(X, <_X)$ is *well-ordered* if every non-empty subset $S \subseteq X$ has a minimal element. That is, for every $S \subseteq X$, there exists $s \in S$ such that $s \leq_X t$ for all $t \in S$.

Example.

- $\mathbb{N}$ is well-ordered.
- $\mathbb{Z}$ with its usual numerical ordering $<_{\mathbb{Z}}$ is not well-ordered.
- $\mathbb{Z}$ with the ordering $a \prec b$ if and only if $|a| < |b|$ or if $a = |b|$,

  $$0 \prec 1 \prec -1 \prec 2 \prec -2 \prec 3 \prec -3 \prec \cdots$$

  is well-ordered. This ordering is order-isomorphic to $(\mathbb{N}, <_{\mathbb{N}})$.
- $\{-\frac{1}{n} : n \in \mathbb{Z}^+\} \cup \mathbb{N}$ is well-ordered.
- $\mathbb{Z}$ with the ordering $a \prec b$ if and only if $0 \leq a < b$, $a \geq 0 > b$ or if $0 \geq a > b$,

  $$0 \prec 1 \prec 2 \prec 3 \prec \cdots \prec -1 \prec -2 \prec -3 \prec \cdots$$

  is well-ordered. This ordering is order-isomorphic to $\{-\frac{1}{n} : n \in \mathbb{Z}^+\} \cup \mathbb{N}$.
- $\mathbb{R}$ is not well-ordered.
- $(\mathbb{N} \times \mathbb{N}, <_{\text{lex}})$ is well-ordered.

Theorem 11.1. *(AC) A total ordering $(X, <_X)$ is well-ordered if and only if there is no strictly decreasing infinite sequence $x_0 > x_1 > x_2 > \cdots$ in X.*

Proof. If such a sequence exists, take $S = \{x_0,x_1,\ldots\}$. Then, S has no minimal element. Conversely, if S is non-empty and has no minimal element, then choose some element $x_0 \in S$. Because x_0 is not minimal, there exist elements $x_1 \in S$ less than x_0. Choose one such element and add it to the sequence. This element is again not minimal, and so on. ■

Theorem 11.2. *Suppose $(X, <)$ is a well-order. Then, if $x \in X$ is not maximal, then there is a unique element $x^+ \in X$ such that $x < x^+$ and there is no element $y \in X$ with $x < y < x^+$.*

Proof. Let x^+ be the minimal element of the non-empty subset $\{y \in X : y > x\}$. ■

Theorem (Fundamental Lemma)**.** *Let $(X, <)$ be a well-ordering, and let $f : X \to X$ be order preserving. That is, $x < y \to f(x) < f(y)$ for all $x,y \in X$. Then, $f(x) \geq x$ for all $x \in X$.*

Proof. Suppose the set

$$S = \{x \in X : f(x) < x\}$$

of counterexamples is non-empty. Since X is well-ordered, $S \subseteq X$ has a minimal element x_0. By definition of S, we have $f(x_0) = x_1 < x_0$, and since f is order preserving, we also have $f(x_1) < x_1$, so $x_1 \in S$. But $x_1 < x_0$, contradicting that x_0 is minimal. ■

Corollary 11.2.1. *If $(X, <_X) \cong (Y, <_Y)$ are isomorphic well-ordered sets, then the order-isomorphism $X \to Y$ is unique.*

Proof. Suppose $f,g : X \to Y$ are order-isomorphisms. Then, $f^{-1} \circ g : X \to X$ is an order-isomorphism, and by the fundamental lemma, $(f^{-1} \circ g)(x) = f^{-1}\big(g(x)\big) \geq x$ for all x. So, $g(x) \leq f(x)$ for all x. Similarly, by considering $g^{-1} \circ f : X \to X$, we obtain that $f(x) \leq g(x)$ for all x, so $f = g$. ■

11.3 Trichotomy Theorem for Well-Ordered Sets

Let $(X, <)$ be a well-ordering, and let $a \in X$. We define the *initial segment determined by the element* a to be the set

$$X \restriction a = \{x \in X : x < a\}$$

This set is well-ordered by the ordering inherited from X.

Example.

- For $(\mathbb{N}, <_{\mathbb{N}})$, $\mathbb{N} \restriction a = \{0,1,\dots,a-1\}$ equipped with the usual ordering $0 < 1 < \cdots < a-1$.
- For $\mathbb{Z}$ in the unusual ordering $0,1,2,3,\dots,-1,-2,-3,-4,\dots$, the initial segment $\mathbb{Z} \restriction -1$ is simply $\mathbb{N}$ in its usual ordering.

Theorem 11.3. *If X is well-ordered, and $x_0 \in X$, then $X \ncong X \restriction x_0$.*

Proof. Suppose $f : X \to X \restriction x_0$ is an order-isomorphism. Then, the same map is equivalently a order homomorphism $f : X \to X$, but with $f(x_0) < x_0$, contradicting the fundamental lemma. ■

Theorem (Trichotomy)**.** *Let $(X, <_X)$ and $(Y, <_Y)$ be well-ordered sets. Then, exactly one of the following holds:*

- *$X \cong Y$;*
- *There exists $x_0 \in X$ such that $X \restriction x_0 \cong Y$;*
- *There exists $y_0 \in Y$ such that $Y \restriction x_0 \cong X$;*

11.4 Well-Ordering Principle

Well-Ordering Principle (Cantor)**.**
Every set is well-orderable. That is, given any set X, there is a relation $<$ on X such that $(X, <)$ is well-ordered.

The well-ordering principle is equivalent to the axiom of choice.

Recall that every finite totally ordered set is well-ordered. We also have that $\omega = \mathbb{N}$ with natural ordering is well-ordered, but $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, with their natural ordering, are not well-ordered.

We can prove that $\mathbb{Z}$ and $\mathbb{Q}$ can be well-ordered without invoking the axiom of choice: both sets are countable, so take any bijection $f : \mathbb{Z} \to \omega$ or $f : \mathbb{Q} \to \omega$, and define the ordering $<$ on $\mathbb{Z}$ or $\mathbb{Q}$ by

$$a < b \quad \leftrightarrow \quad f(a) <_\omega f(b)$$

Since $(\omega, <_\omega)$ is well-ordered, $\mathbb{Z}$ or $\mathbb{Q}$ equipped with this ordering $<$ will also be well-ordered. However, this method does not work on $\mathbb{R}$ since it is uncountable. Here, the well-ordering principle must be used.

11.5 Order-types

Informally, the cardinality of a set describes the "size" of that set. The *order-type* instead describes the "length" of a well-ordered set.

Let $(X, <_X)$ and $(Y, <_Y)$ be well-orderings. We say that

$$\text{type}(X, <_X) = \text{type}(Y, <_Y)$$

if $(X, <_X) \cong (Y, <_Y)$.

We seem to have very good notation for the order-type of many well-ordered sets:

- $\omega = \text{type}(\mathbb{N})$;
- $n = \text{type}\big(\{0,1,\ldots,n-1\}\big)$;
- $0 = \text{type}(\emptyset)$.

Let α and β be order-types of well-ordered sets. Fix well-ordered sets $(X, <_X)$ and $(Y, <_Y)$ with order-type α and β, respectively.

- We write $\alpha = \beta$ if $X \cong Y$;
- We write $\alpha > \beta$ if there exists $x_0 \in X$ such that $X \restriction x_0 \cong Y$;
- We write $\alpha < \beta$ if there exists $y_0 \in Y$ such that $X \cong Y \restriction y_0 \cong$.

Recall that exactly one of these three options holds by the trichotomy theorem.

11.6 Ordinal Arithmetic

We have seen that cardinalities admit a sensible notion of arithmetic. What about order-types?

Let $\alpha = \text{type}(X, <_X)$ and $\beta = \text{type}(Y, <_Y)$.

- Suppose that X and Y are disjoint. The union $X \cup Y$ is well-ordered by the relation $<$ where $a < b$ if $(a \in X$ and $b \in Y)$ or $(a,b \in X$ and $a <_X b)$ or $(a,b \in Y$ and $a <_Y b)$.
- The product $X \times Y$ can also be well-ordered by the *anti-lexicographic ordering* with $\langle x,y\rangle <_{\text{anti-lex}} \langle x',y'\rangle$ if and only if $y <_Y y'$, or if $y = y'$ and $x <_X x'$.

Then, we define $\alpha + \beta = \text{type}(X \cup Y, <)$ and $\alpha \times \beta = \text{type}(X \times Y, <_{\text{anti-lex}})$.

Recall the labels used in Cantor's transfinite iteration from § 1:

$$0,1,\ldots,\omega,\omega+1,\ldots,\underbrace{\omega+\omega}_{\omega\cdot 2},\ldots,\underbrace{\omega+\omega+\omega}_{\omega\cdot 3},\ldots,\underbrace{\omega\cdot\omega}_{\omega^2},\ldots$$

Formally, this notation lists out the order-types in increasing order.

12 Transfinite Induction

Induction works on $\mathbb{N}$ precisely because $\mathbb{N}$ is well-ordered. It turns out that there is nothing special about $\mathbb{N}$, and induction works equally as well on *any* well-ordered set.

12.1 Induction on $\mathbb{N}$

To prove that a property φ holds for all natural numbers, the usual way is to show that $\varphi(0)$ holds, then to prove that for every $n \in \mathbb{N}$ if $\varphi(n)$ holds, then $\varphi(n+1)$ holds.

Theorem (Induction on $\mathbb{N}$). *Let φ be a property of natural numbers. If $\varphi(0)$ holds, and for every $n \in \mathbb{N}$, $\varphi(n)$ implies $\varphi(n+1)$, then for every $n \in \mathbb{N}$, $\varphi(n)$ holds:*

$$\Big(\varphi(0) \wedge \forall n \in N\big(\varphi(n) \to \varphi(n+1)\big)\Big) \to \forall n \in \mathbb{N} : \varphi(n)$$

This version of induction does not generalise easily to arbitrary well-ordered sets, so first we rephrase this.

Theorem (Strong Induction on $\mathbb{N}$). *Let φ be a property of natural numbers, and assume that φ satisfies the following property:*

If for every $n \in \mathbb{N}$, if $\varphi(m)$ holds for all $m < n$, then $\varphi(n)$ holds.

Then, for every $n \in \mathbb{N}$, $\varphi(n)$ holds:

$$\Big(\forall n \in \mathbb{N} : \big(\forall m \in \mathbb{N} : m < n \to \varphi(m)\big) \to \varphi(n)\Big) \to \forall n \in \mathbb{N} : \varphi(n)$$

It may seem that we are missing the base case $\varphi(0)$ in our assumptions, but for $n = 0$, the assumption becomes the following: if $\varphi(m)$ holds for all naturals $m < 0$, then $\varphi(0)$ holds. The conditional clause is vacuously true, since there are no naturals $m < 0$.

Proof. Let $\psi(n)$ be the statement "if $m < n$, then $\varphi(m)$ holds". Then, $\psi(0)$ holds as there is no $m < 0$.

Now assume $\psi(n)$ holds, so $\varphi(m)$ holds for all $m < n$ and therefore $\varphi(n)$ holds by the hypothesis of the theorem. Thus, $\varphi(m)$ holds for all $m < n+1$, which is the statement of $\psi(n+1)$. Then, standard induction for ψ gives that $\psi(n)$ holds for all $n \in \mathbb{N}$, and therefore $\varphi(n)$ holds for all $n \in \mathbb{N}$. ■

12.2 Transfinite Induction on Well-Ordered Sets

Strong induction on $\mathbb{N}$ generalises nicely to arbitrary well-ordered sets.

Theorem (Transfinite Induction). *Let $(X, <)$ be a well-ordered set, and let φ be a property of elements of X. That is, for each $x \in X$, $\varphi(x)$ is either true or false.*

Assume that for every $x \in X$, the following holds:

If $\varphi(y)$ holds for every $y < x$ then $\varphi(x)$ holds.

Then, for every $x \in X$, $\varphi(x)$ holds:

$$\Bigg(\forall x \in X\Big(\forall y \in X\big(y < x \to \varphi(y)\big)\Big)\Bigg) \to \forall x \in X : \varphi(x)$$

Proof. Define the set

$$S = \big\{x \in X : \neg\varphi(x)\big\}$$

Suppose for a contradiction that S is non-empty. Then, since X is well-ordered, $S \subseteq X$ has a minimal element x_0. Then, for every $y < x_0$, $y \notin S$ so $\varphi(y)$ holds. By the assumption, this implies $\varphi(x_0)$ holds, so $x_0 \notin S$. ■

13 Ordinals as Sets

13.1 Mapping Order-Types to Sets

We wish to define the *ordinals*: sets equipped with a natural well-ordering such that each order-type of well-ordered sets corresponds to exactly one ordinal.

We can think of this as a "map" that assigns the abstract concept of order-types to well-ordered sets. However, the order-types are just classes of well-ordered sets, so we are really looking for a class-function that sends well-ordered sets to representing well-ordered sets, such that order-isomorphic well-ordered sets have the same image.

Consider the three-element set $X = \{a,b,c\}$ with ordering $a < b < c$, and let E be a function with domain X that satisfies the equation

$$E(x) = \big\{E(y) : y <_X x\big\}$$

for all $x \in X$.

Then,

$$\begin{aligned}
E(x_0) &= \big\{E(y) : y <_X a\big\} = \emptyset = 0 \in \omega \\
E(x_1) &= \big\{E(y) : y <_X b\big\} = \big\{E(a)\big\} = \{\emptyset\} = 1 \in \omega \\
E(x_2) &= \big\{E(y) : y <_X c\big\} = \big\{E(a),E(b)\big\} = \big\{\emptyset,\{\emptyset\}\big\} = 2 \in \omega
\end{aligned}$$

More generally, for any finite well-ordered set X, the function E as defined above maps the smallest element of X to $\emptyset = 0$, the second smallest to $\{0\} = 1$, the next smallest to $\{0,1\}$, and so on; and furthermore, E gives an order-isomorphism between X and the set $n = \{0,1,2,\ldots,n-1\}$.

Note that if $x <_X x'$ for some $x,x' \in X$, then $E(x) \in E(x')$. This recursive property allows us to define E for all well-ordered sets.

Lemma (Epsilon-image of a Well-Ordered Set). *Let $(X, <_X)$ be a well-ordered set. Then, there is a unique function E with domain X such that for all $x \in X$,*

$$\begin{aligned}
E(x) &= \big\{E(y) : y \leq_X x\big\} \\
&= E\big[\{y \in X : y <_X x\}\big] \\
&= E[X \restriction x]
\end{aligned}$$

This map E has some very useful properties.

Lemma (Epsilon-images II). *The function E defined in the previous lemma satisfies the following properties:*

(i) *Whenever $x <_X x'$, we have $E(x) \in E(x')$, and also $E(x) \subsetneq E(x')$;*

(ii) *E is injective;*

(iii) *E is an order-isomorphism between $(X, <_X)$ and $(E[X], \in)$.*

(iii) *$E[X]$ is well-ordered by the $\in$ relation.*

Proof.

(i) Suppose that $x,x' \in X$, and $x <_X x'$. Then by definition of E, we have

$$E(x') = \big\{E(y) : y <_X x'\big\}$$

Since $x <_X x'$, we have $E(x) \in E(x')$.

Then, for any $E(y) \in E(x)$, we have $y < x$ by definition, so $y < x < x'$ gives $E(y) \in E(x')$, and $E(x) \subseteq E(x')$. We also have $x' \not< x$, so $E(x') \notin E(x)$, so $E(x) \subsetneq E(x')$.

(ii) Suppose for a contradiction that $E(x) = E(x')$ for some $x < x'$. Then, $E(x) \in E(x') = E(x)$. Let x_0 be the $<_X$-minimal element of X such that $E(x_0) \in E(x_0)$. Unfolding the definition of E, we have

$$E(x_0) = \{E(y) : y < x_0\} \in E(x_0)$$

Because $E(x_0) \in E(x_0)$, it is non-empty, so there exists $y < x_0$ such that $E(x_0) = E(y) \in E(x_0)$. But since $E(y) = E(x_0)$, we have $E(y) \in E(y)$, and $y < x_0$, contradicting the minimality of x_0.

(Injectivity also implies $E(x) \subsetneq E(x')$ in the first part.)

(iii) Since E is injective, it is a bijection from X to $E[X]$. Property (i) then implies that E is order preserving if $E[X]$ is equipped with the ordering given by $\in$.

(iv) X is well-ordered (by assumption), and $(X, <_X) \cong (E[X], \in)$, so $(E[X], \in)$ is well-ordered.

■

Lemma 13.1. *Let $(X, <_X) \cong (Y, <_Y)$ be order-isomorphic well-ordered sets. Then, their epsilon-images $E_X[X]$ and $E_Y[Y]$ coincide.*

Proof. This is a proof by transfinite induction. Let $f : X \to Y$ be an order-isomorphism, and define the set

$$S := \{x \in X : E_X(x) \neq E_Y\big(f(x)\big)\}$$

Suppose for a contradiction that S is non-empty. Let x_0 be the smallest element of S ($S \subseteq X$, so S is well-ordered and x_0 exists). By definition of S, we have $E_X(x) = E_Y\big(f(x)\big)$ for all $x < x_0$. Then,

$$\begin{aligned} E_X(x_0) &= \{E_X(x) : x <_X x_0\} \\ &= \{E_Y\big(f(x)\big) : x <_X x_0\} \\ &= \{E_Y\big(y\big) : y <_X f(x_0)\} \\ &= E_Y\big(f(x_0)\big) \end{aligned}$$

contradicting that $x_0 \in S$. ■

13.2 Ordinals

A set α is an *ordinal* if

- it is well-ordered by the $\in$ relation;
- whenever $y \in x \in \alpha$, we have $y \in \alpha$.

The second property is equivalent to $x \in \alpha \to x \subseteq \alpha$ (α is *transitive*).

Note that by iterating the second property, we also have that if $x_3 \in x_2 \in x_1 \in \alpha$, then $x_3 \in x_2 \in \alpha$, and $x_3 \in \alpha$. Informally, α contains as elements "all the things it references".

Theorem 13.2.

(i) *The epsilon-image of well-ordered sets are ordinals.*

(ii) *Every ordinal is the epsilon-image of a well-ordered set. Namely, each ordinal is the epsilon-image of itself.*

(iii) *Ordinals represent order-types of well-ordered sets.*

Proof.

(i) This follows from Theorem 13.1. Let $(X, <_X)$ be a well-ordering. Suppose $a \in b \in E[X]$, so $a = E(x)$ and $b = E(x')$ for some $x < x'$. Then $E(x) \in E[X]$.

(ii) Let α be an ordinal. Then, by definition,

$$E(x) = \big\{E(y) : y \in x\big\}$$

We claim that $E(x) = x$ for every $x \in \alpha$. Suppose otherwise, and let x_0 be the $\in$-minimal element in α for which $E(x_0) \neq x_0$. Then,

$$\begin{aligned} E(x_0) &= \big\{E(y) : y \in x_0\big\} \\ &= \big\{y : y \in x_0\big\} \\ &= x_0 \end{aligned}$$

This is a contradiction, so $E(x) = x$ for all $x \in \alpha$, and $E[\alpha] = \alpha$.

(iii) The previous two statements combined with Theorem 13.1 is precisely this result.

■

Theorem (Properties of Ordinals)**.**

(i) *Any element of an ordinal is itself an ordinal.*

(ii) *For any ordinals α,β,γ, if $\alpha \in \beta \in \gamma$, then $\alpha \in \gamma$.*

(iii) *For any ordinal α, we have $\alpha \notin \alpha$.*

(iv) *For any two ordinals α,β, exactly one of the $\alpha \in \beta$, $\alpha = \beta$, and $\beta \in \alpha$ holds (trichotomy).*

(v) *Any non-empty set S of ordinals has a least element. That is, there exists an ordinal $\delta \in S$ such that $\delta \in \alpha$ for all $\alpha \in S$ distinct from δ.*

(vi) *Every ordinal is the set of ordinals smaller than it: $\alpha = \{\beta : \beta < \alpha\}$, where $\beta < \alpha \equiv \beta \in \alpha$.*

(vii) *For each ordinal α, the set*

$$\alpha + 1 \coloneqq \alpha^+ \coloneqq \alpha \cup \{\alpha\}$$

is also an ordinal, and is the smallest ordinal larger than α:

$$\alpha^+ = \min\{\beta : \beta > \alpha\}$$

(viii) *If A is any set of ordinals, then $\bigcup A$ is also an ordinal, and is the least upper bound of A.*

An infinite ordinal is called a *limit ordinal* if it is not the successor of any ordinal.

Example. ω is a limit ordinal, while ω^+ and ω^{++} are successor ordinals.

Example. The following sets are ordinals:

- $0 = \emptyset$;
- $1 = \{\emptyset\}$;
- $2 = \big\{\emptyset,\{\emptyset\}\big\}$;
- $n = \{0,1,2,\ldots,n-1\}$ for every $n \in \omega$;
- $\omega = \{0,1,2,\ldots\}$;
- $\omega + 1 = \omega^+ = \omega \cup \{\omega\} = \{0,1,2,\ldots,\omega\}$;
- $\omega + 2 = \omega^{++} = \omega^+ \cup \{\omega^+\} = \{0,1,2,\ldots,\omega,\omega+1\}$;

- $\omega + n = \{0,1,2,\ldots,\omega,\omega+1,\ldots,\omega+(n-1)\}$;
- $\omega \cdot 2 = \omega + \omega = \{0,1,2,\ldots,\omega,\omega+1,\omega+2,\ldots\}$;

An ordinal α is *countable* if α is a countable set. Every ordinal in the example above is countable.

However, what is the order-type of an uncountable well-ordered set? Such sets exist, by the well-ordering principle, so it must have an order type.

We define the ordinal

$$\omega_1 := \text{the smallest uncountable ordinal}$$

Since any subset of the ordinals has a minimal element, such a smallest ordinal exists.

So, α is a countable ordinal if and only if $\alpha < \omega_1$. That is, if $\alpha \in \omega_1$. Since ω_1 is uncountable, there are uncountably many countable ordinals. This also means that not every countable ordinal can be explicitly expressed like those above.

Theorem (Burali-Forti)**.** *There is no set to which every ordinal number belongs.*

Proof. The class of ordinals is well-ordered by $\in$, and every element of an ordinal is an ordinal, so if the class of ordinals was a set, it would be an ordinal itself. But then it would be a member of itself, and no ordinal has this property. ■

13.3 Cardinals

Let A be a set. Then, its *cardinal* $\kappa = |A|$ is the smallest ordinal that is equinumerous to A.

Note that cardinals are always limit ordinals, because for any infinite ordinal α, α and $\alpha + 1 = \alpha \cup \{\alpha\}$ are equinumerous.

Example. The smallest countably infinite ordinal is ω, so $\omega = \aleph_0 = |\mathbb{N}| = |\mathbb{Z}|$.

Note that it is still sensible to retain the differing notations ω and $\aleph_0$ despite them being the same set, because cardinal and ordinal arithmetic are distinct. That is, $\omega + 1 \neq \aleph_0 + 1$. Also, $\omega = \aleph_0 = |\omega + 1|$

Let $\aleph_1$ denote the cardinality of ω_1. Actually, $\aleph_1 = \omega_1$ by the definition above. Then, $\aleph_1$ is the smallest uncountable cardinality. We similarly define $\aleph_2$ to be the smallest cardinal larger than $\aleph_1$, and so on. Note that we did not know until now that there is a smallest uncountable cardinality.

14 Applications

14.1 Transfinite Recursion

Theorem 14.1. *It is possible to draw disjoint letters* T *in $\mathbb{R}^2$ above every rational point of the x-axis.*

(A letter T above a real number x with height $h > 0$ and width $w > 0$ is a union of two line segments: the vertical line segment connecting $(x,0)$ to (x,h), and the horizontal line segment connecting $(x - \frac{w}{2},h)$ to $(x + \frac{w}{2},h)$.)

Constructive proof. For the rational number $\frac{p}{q}$ with $p \neq 0$, $q > 0$, and p,q coprime, draw a letter T with width $\frac{1}{2q^2}$ and height $\frac{1}{q}$.

Let $\frac{p}{q}$ and $\frac{p'}{q'}$ be two rationals in simplest form. If $q' = q$, then the horizontal distance of these rational numbers is at least $\frac{1}{q}$, so the letters T are clearly disjoint. Otherise, suppose without loss of generality that $q' < q$. Then,

$$\left|\frac{p}{q} - \frac{p'}{q'}\right| = \left|\frac{pq' - p'q}{qq'}\right| \geq \frac{1}{qq'} \geq \frac{1}{q^2}$$

so the horizontal segment of the letter `T` for $\frac{p}{q}$ does not reach the vertical segment of the letter for $\frac{p'}{q'}$, and since the height of the letter `T` for $\frac{p'}{q'}$ is greater than the height of the letter `T` for $\frac{p}{q}$, the letters are indeed disjoint. ■

Proof by recursion. Enumerate the rationals as x_0,x_1,x_2,.... First draw a letter `T` above x_0. Then draw a letter `T` x_1, disjoint from the one above x_0. Then draw a letter `T` above x_2, disjoint from the previously drawn letters. And so on.

At each step, there are always finitely many letters `T` drawn, so we can always draw the next one disjoint from all previous ones. ■

Theorem 14.2. *(AC)* $\mathbb{R}^3$ *is a union of disjoint circles of unit radius.*

Proof. Let $\{p_\alpha : \alpha < \mathfrak{c}\}$ enumerate the points of $\mathbb{R}^3$. That is, well-order $\mathbb{R}^3$ so that its order type is the smallest possible ordinal with cardinality of the continuum; or equivalently, fix a bijection from the cardinal/ordinal $\mathfrak{c}$ to $\mathbb{R}^3$.

For each ordinal $\alpha < \mathfrak{c}$, choose a circle C_α to cover the point p_α unless p_α was already covered by previous circles $\{C_\beta : \beta < \alpha\}$, in which case, set $C_\alpha = \emptyset$.

With transfinite recursion, define sets C_α such that for each $\alpha < \mathfrak{c}$,

- $\bigcup_{\beta\leq\alpha} C_\beta$ contains the point p_α,
- C_α is disjoint from the set $\bigcup_{\beta<\alpha} C_\beta$.

Once this is done, the non-empty sets C_α are pairwise disjoint unit circles whose union is $\mathbb{R}^3$. ■

14.2 Continuum Hypothesis

Is it true that a set $A \subset \mathbb{R}$ is either countable or there is a bijection $A \to \mathbb{R}$? That is, is it true that there is no cardinal κ with $\aleph_0 < \kappa < 2^{\aleph_0}$? Or equivalently, is it true that $\mathfrak{c} = \omega_1$?

This last statement is the *continuum hypothesis*, and it has been shown to be independent from the axioms of ZFC set theory. That is, there are models of set theory in which CH holds, and models in which it fails.

15 Axiom of Regularity

We have proved that for every natural number n, we have $n \notin n$. However, there is an additional axiom that implies that no set is an element of itself.

Axiom of Regularity.
Every non-empty set X has an element x such that $x \cap X = \emptyset$. That is:

$$\forall X\big(X \neq \emptyset \to \exists x : x \in X \wedge x \cap X = \emptyset\big)$$

That is, x does not contain any element of X. This axiom is often stated as, "*every non-empty set has an* $\in$*-minimal element.*"

Theorem 15.1 (Corollaries of Regularity)**.**

(*i*) *For every set* x, $x \notin x$.

(*ii*) *There are no sets* a *and* b *such that* $a \in b$ *and* $b \in a$.

Proof.

(*i*) Suppose $x \in x$. Let $X = \{x\}$ (this is a set by the axiom of pairing). Since x is the only element of X, we must have $x \cap X = \emptyset$ by the axiom of regularity. However, $x \in x$ and $x \in X$, so $x \in x \cap X \neq \emptyset$.

(*ii*) Suppose $a \in b$ and $b \in a$. Let $X = \{a,b\}$. By regularity, either $a \cap X = \emptyset$ or $b \cap X = \emptyset$. However, neither hold, since $a \in b \cap X$ and $b \in a \cap X$.

■

The axiom of regularity is equivalent to the statement that there is no infinite sequence of sets $x_0,x_1,x_2,\ldots$ such that

$$x_0 \ni x_1 \ni x_2 \ni \cdots$$

Since a sequence is just a function from ω, more precisely, there is no function f with domain ω such that

$$f(0) \ni f(1) \ni f(2) \ni \cdots$$

Theorem 15.2. *The axiom of regularity is equivalent to the statement that there are no functions f with domain ω such that $f(0) \ni f(1) \ni f(2) \ni \cdots$.*

Proof. Let $X = \operatorname{ran}(f)$. By regularity, $x \cap X = \emptyset$ for some $x \in X$. But $x = f(n)$ for some $n \in \omega$, so $f(n+1) \in x \cap X \neq \emptyset$.

The other direction requires the axiom of choice. Let X be a non-empty set, and let $x_0 \in X$. If $x_0 \cap X = \emptyset$, then take $x = x_0$. Otherwise, choose some $x_1 \in x_0 \cap X$ and check if $x_1 \cap X = \emptyset$. If so, take $x = x_1$. Otherwise, choose some $x_2 \in x_1 \cap X$. And so on. ■

15.1 Cumulative Hierarchy

Recall the hierarchy of sets described in § 2.1. We did not end up needing atoms, so we restate the atomless version of the hierarchy of sets more formally here.

Informally, we defined V_0 to be a certain set, then recursively defined $V_1 = \mathcal{P}(V_0)$, $V_2 = \mathcal{P}(V_1)$, and so on. We formalise the "and so on" part using ordinals.

Let

$$V_0 = \emptyset$$

(This is slightly different from the introduction, where we instead had $V_0 = \{\emptyset\}$.) Then, for each ordinal α, define

$$V_{\alpha+1} = \mathcal{P}(V_\alpha)$$

For limit ordinals α, define

$$V_\alpha = \bigcup_{\beta<\alpha} V_\beta$$

(Proving, by a version of transfinite recursion, that this is well-defined is quite some work.)

Theorem 15.3. *The axiom of regularity is equivalent to the statement that every set appears in the Cumulative Hierarchy. That is, for every set A, there is an ordinal α such that $A \in V_\alpha$.*

16 Condensed List of ZFC Axioms

Axiom of Extensionality.
If two sets have exactly the same members, then they are equal:

$$\forall X \forall Y \big(\forall z (z \in X \leftrightarrow z \in Y) \to x = y\big)$$

Axiom of the Empty Set.
There exists a set with no elements:

$$\exists E \forall x : x \notin E$$

Axiom of Pairing.
For any two sets u and v, there exists a set that contains exactly u and v as elements.

$$\forall u \forall v \exists X \forall x \big(x \in X \leftrightarrow (x = u \vee x = v)\big)$$

Axiom of the Power Set.
For any set u, there is a set whose elements are exactly the subsets of u:

$$\forall u \exists P \forall s (s \subseteq u \leftrightarrow s \in P)$$

or omitting the abbreviation $\subseteq$,

$$\forall u \exists P \forall s \big(\forall x (x \in s \to x \in s) \leftrightarrow s \in P\big)$$

Axiom Schema of Specification.
Let φ be any formula that does not contain the variable name B and has only bound variables, except for x,$t_1, \ldots$,t_k. Then, the following is an axiom:

$$\forall t_1 \forall t_2 \ldots \forall t_k \forall A \exists B \forall x \big(x \in B \leftrightarrow (x \in A \wedge \varphi)\big)$$

That is, for any property φ of x and any set A, there exists a set B that contains exactly the elements of A for which $\varphi(x)$ holds, and φ may depend on additional parameters $t_1, \ldots$,t_k.

Axiom of Union.
For any set A, there exists a set B whose members are precisely the members of the members of A:

$$\forall A \exists B \forall x \big(x \in B \leftrightarrow \exists y (y \in A \wedge x \in y)\big)$$

Axiom of Infinity.
There is an inductive set:

$$\exists A \big(\emptyset \in A \wedge \forall x (x \in A \to x^+ \in A)\big)$$

or omitting $\emptyset$ and x^+,

$$\exists A \big(\exists e (\forall z : z \notin e) \wedge e \in A \wedge \forall x (x \in A \to x \cup \{x\} \in A)\big)$$

Axiom Schema of Replacement.
The image of a set under a class-function is a set; if φ is any formula that does not contain B, then:

$$\forall A\Big(\underbrace{\forall x\forall y\forall y'\Big(\big(x \in A \wedge \varphi(x,y) \wedge \varphi(x,y')\big) \to y = y'\Big)}_{\varphi \text{ is a class-function on at least } A} \to \underbrace{\exists B\forall y\Big(y \in B \leftrightarrow \exists x(x \in A \wedge \varphi(x,y))\Big)}_{\text{there is a set } B \text{ consisting of } \varphi\text{-images of elements of } A}\Big)$$

Axiom of Regularity.
Every non-empty set X has an element x such that $x \cap X = \emptyset$. That is:

$$\forall X\big(X \neq \emptyset \to \exists x : x \in X \wedge x \cap X = \emptyset\big)$$

Axiom of Choice (first form).
For any relation R, there exists a function $F \subseteq R$ such that $\operatorname{dom}(F) = \operatorname{dom}(R)$.

Axiom of Choice (second form).
Let $\mathcal{S}$ be a set with $\emptyset \notin \mathcal{S}$. Then, there is a *choice function* for $\mathcal{S}$. That is, a function $\sigma : \mathcal{S} \to \bigcup \mathcal{S}$ such that $\sigma(A) \in A$ for all $A \in \mathcal{S}$.

Axiom of Choice (third form).
Let $\mathcal{A}$ be a set of pairwise disjoint non-empty sets. Then, there exists a set C who has as a member exactly one element from each member of $\mathcal{A}$. That is, for each $B \in \mathcal{A}$, $|C \cap B| = 1$.

Axiom of Choice (fourth form).
For any indexing set I and function H with domain I, if $H(i) \neq \emptyset$ for all $i \in I$, then

$$\mathop{\times}_{i \in I} H(i) \neq \emptyset$$

Well-Ordering Principle (Cantor)**.** (Equivalent to Choice)
Every set is well-orderable. That is, given any set X, there is a relation $<$ on X such that $(X, <)$ is well-ordered.

Cardinal Comparability. (Equivalent to Choice)
For any sets A and B, we have $|A| \leq |B|$ or $|B| \leq |A|$. That is, there is an injective function $A \to B$ or there is an injective function $B \to A$.
Equivalently, for any two cardinals κ and λ, we have $\kappa \leq \lambda$ or $\lambda \leq \kappa$.

Zorn's Lemma. (Equivalent to Choice)
Let $(Z, \preceq)$ be a poset, and suppose that every chain $C \subseteq Z$ has an upper bound. Then, Z has a maximal element.