A QUICK INTRODUCTION TO BASIC SET THEORY

ANUSH TSERUNYAN

This note is intended to quickly introduce basic notions in set theory to (undergraduate) students in mathematics not necessarily specializing in logic, thus requiring no background. It uses the sections on well-orderings, ordinals and cardinal arithmetic of [Kun83], as well as some parts from [Mos06].

Contents

Date: June 7, 2022.

1. Zermelo–Fraenkel set theory

The language of ZFC. ZF stands for *Zermelo–Fraenkel set theory* and ZFC stands for *Zermelo–Fraenkel set theory with Choice* (the latter being an extra axiom added to ZF).

To describe the *axioms* of ZFC we need to fix a language (formally speaking, a first order logic language). The *language* of ZF (same for ZFC) has only one special symbol, namely the binary relation symbol $\in$, in addition to the standard symbols that every other first order language has:

$$\text{the symbols } =, \neg, \wedge, \vee, \to, \forall, \exists, (,) \text{ and variables } x_0, x_1, x_2, \ldots.$$

To keep things informal and easy to read, we also use other letters such as x, y, z or even $A, B, \mathcal{C}, \mathcal{F}$ for variables, as well as extra symbols like $\neq, \notin, \leftrightarrow$, etc.

Using these symbols and variables we form statements about sets such as $\forall x(x = x)$, $\exists x(x \neq x)$, $\exists x(x \in y \wedge y \notin z)$. The variables are interpreted as sets and the symbols are interpreted the expected way: $=$ as "equality", $\in$ as "belongs to" (set membership), $\neg$ as "negation", $\wedge$ as "conjunction" (and), $\vee$ as "disjunction" (or), $\to$ as "implication", $\forall$ as "for all", and $\exists$ as "there exists". For example, $\forall x((x \in y \wedge y \in z) \to (x \in z))$ is interpreted as "for all sets x, if the set x is a member of the set y and the set y is a member of the set z, then the set x is a member of the set z". In mathematical logic, we refer to these kinds of statements as formulas (in the language of ZF). Here is their formal definition, although a nonpedantic reader can safely skip it.

Definition 1.1. Recursively define the notion of a *formula* in the language of ZF as follows:

(i) For variables x, y, $x = y$ is a formula;
(ii) For variables x, y, $x \in y$ is a formula;
(iii) If φ is a formula, then so is $\neg(\varphi)$;
(iv) If φ and ψ are formulas, then so are $(\varphi) \wedge (\psi)$, $(\varphi) \vee (\psi)$ and $(\varphi) \to (\psi)$;
(v) If φ is a formula and x a variable, then $\forall x(\varphi)$ and $\exists x(\varphi)$ are formulas.

When writing a formula, one doesn't have to religiously follow the rules about parentheses described in the definition above as long as there is no ambiguity in reading the formula; for example, $\neg x = y$ is technically not a formula, but it's ok to write it this way since there is no ambiguity: it is clearly the same as $\neg(x = y)$. Also, when we want to emphasize that a formula φ says something about a set (variable) x, we write $\varphi(x)$, although φ may also contain other variables; for example, $\varphi(x)$ may be $x \in y$. Lastly, we use abbreviations

$$\begin{aligned} \forall x \in X\, \varphi(x) &:\Longleftrightarrow \forall x\, (x \in X \to \varphi(x)) \\ \exists x \in X\, \varphi(x) &:\Longleftrightarrow \exists x\, (x \in X \wedge \varphi(x)). \end{aligned}$$

Axioms of ZFC. We now list all of the axioms of ZFC for reference, although we discuss only some of them here and leave the detailed discussion of the rest for later. Axioms 0–7 form the system ZF and ZFC is just ZF together with Axiom 9 (Choice).

Axiom 0. **Set existence**:

$$\exists x\, (x = x).$$

Since $x = x$ vacuously holds, this axiom just says that there is a set (one has to start with something).

Axiom 1. **Extensionality**:

$$\begin{aligned}&\forall x\,\forall y\,(\\&\quad \forall z\,(z \in x \leftrightarrow z \in y)\\&\quad \rightarrow\\&\quad x = y\\&\quad).\end{aligned}$$

This says that if two sets have the same members then they are equal; in other words, sets are uniquely determined by the members they contain.

Axiom 2. **Pairing**:

$$\begin{aligned}&\forall u\,\forall v\,\exists p\,\forall w\,(\\&\quad w \in p\\&\quad \leftrightarrow\\&\quad (w = u \vee w = v)\\&\quad).\end{aligned}$$

This says that for any sets u, v, there is a set p that contains both u, v as its members and no other members. We denote this set p by $\{u, v\}$ and call it the (unordered) *pairing* of u, v.

Axiom 3. **Union**:

$$\begin{aligned}&\forall \mathcal{C}\,\exists U\,\forall x\,(\\&\quad x \in U\\&\quad \leftrightarrow\\&\quad \exists S\, \in \mathcal{C}(x \in S)\\&\quad).\end{aligned}$$

This says that for each set $\mathcal{C}$ (think of $\mathcal{C}$ as a collection of sets), there is a set U whose members are exactly the elements that belong to some set S in this collection $\mathcal{C}$. We denote this set U by $\bigcup \mathcal{C}$ and call it the *union of the sets* in $\mathcal{C}$.

Axiom 4. **Powerset**:

$$\begin{aligned}&\forall X\,\exists \mathcal{P}\,\forall Y\,(\\&\quad Y \in \mathcal{P}\\&\quad \leftrightarrow\\&\quad Y \subseteq X\\&\quad),\end{aligned}$$

where $Y \subseteq X$ is an abbreviation for $\forall z(z \in Y \rightarrow z \in X)$. This says that for any set X there is a set $\mathcal{P}$ (think of it as a collection of sets) whose elements are exactly the subsets of X. We denote this set $\mathcal{P}$ by $\mathscr{P}(X)$ and call it the *powerset* of X.

Axiom schema 5. **Comprehension**: for every formula φ, the following is an axiom:

$$\begin{aligned}
&\forall x\, \exists y\, \forall z\, (\\
&\qquad z \in y\\
&\qquad \leftrightarrow\\
&\qquad (z \in x \wedge \varphi(z))\\
&\qquad).
\end{aligned}$$

This says that for each set X there is a set Y whose element are exactly those elements of X that satisfy φ. We denote this set Y by $\{z \in X : \varphi(z)\}$ and call it the *subset of X defined* by φ.

Note that this is not one axiom! This is an infinite collection of axioms, one for each φ. Furthermore, note that $\varphi(z)$ may also contain variables other than z, X, Y and we think of them as parameters.

Axiom 6. **Infinity**:

$$\begin{aligned}
&\exists X\, (\\
&\quad \emptyset \in X\\
&\quad \wedge\\
&\quad \forall x \in X\, (S(x) \in X)\\
&\quad),
\end{aligned}$$

where $\emptyset$ and $S(x)$ are defined as follows. It can be deduced from the above axioms that there is a set with no elements, which we denote by $\emptyset$ and call the *emptyset*. Furthermore, it can be deduced that for every set x, there is a set y whose members are exactly the elements of x and x itself, i.e., $y := x \cup \{x\}$. We denote this set y by $S(x)$ and call it the *successor* of x.

Call a set X *inductive* if $\emptyset \in X$ and for each $x \in X$, its successor $S(x)$ is also an element of X. Infinity axiom states that there exists an inductive set.

Axiom 7. **Foundation**:

$$\begin{aligned}
&\forall X\, (\\
&\quad X \neq \emptyset\\
&\quad \rightarrow\\
&\quad \exists y \in X\, \neg \exists z \in X\, (z \in y)\\
&\quad),
\end{aligned}$$

where we again use some easy-to-unravel abbreviations.

We call an element $y \in X$ an $\in$-*minimal* (or *membership-minimal*) element of X if there is no element $z \in X$ such that $z \in y$ (in particular, $y \notin y$). Foundation axiom states that the binary relation $\in$ is *well-founded*, i.e., every set X has an $\in$-minimal element. It is rarely used in set theory and not at all in other areas of mathematics. Thus, as an exercise, we try to circumvent its uses below in the text and the reader is urged to do so as well.

Axiom schema 8. **Replacement**: for every formula φ, the following is an axiom:

$$
\begin{aligned}
&\forall X\,[\\
&\quad \forall x \in X\, \exists! y\, \varphi(x,y)\\
&\quad \rightarrow\\
&\quad \exists Y\, \forall x \in X\, \exists y \in Y\, \varphi(x,y)\\
&\quad],
\end{aligned}
$$

where $\exists! y$ abbreviates "there exists a unique y".

We say that a formula $\varphi(x,y)$ *defines a function* on a set X if for each $x \in X$, there is exactly one set y such that $\varphi(x,y)$ holds. In this case, we say that a set Y *contains the φ-image* of X if for all $x \in X$, all y with $\varphi(x,y)$ are elements of Y.

Replacement axiom for φ says that for every set X, if $\varphi(x,y)$ defines a function on X, then there is a set Y containing the φ-image of X.

Axiom 9. **Choice**:

$$
\begin{aligned}
&\forall \mathcal{C}\,[\\
&\quad \emptyset \notin \mathcal{C}\\
&\quad \rightarrow\\
&\quad \exists f : \mathcal{C} \to \bigcup \mathcal{C}\, \forall S \in \mathcal{C}\, (f(S) \in S)\\
&\quad],
\end{aligned}
$$

where for sets X, Y, the notation $f : X \to Y$ means that f is a *function* from X to Y. The notion of a function, as well as that of an *ordered pair* (x,y) and *Cartesian product* $X \times Y$, can be defined using the axioms above, and we leave it as an exercise (see Exercise 1).

Choice axiom (more commonly referred to as Axiom of Choice, abbreviated, AC) states that for any set $\mathcal{C}$ with $\emptyset \notin \mathcal{C}$ (think of $\mathcal{C}$ as a collection of nonempty sets), there is a function f that assigns to each set $S \in \mathcal{C}$ an element of S.

Notation 1.2. For a set X and sets $x_0, x_1, \ldots, x_n$, we write $X = \{x_0, x_1, \ldots x_n\}$ to mean that for all sets x,

$$x \in X \leftrightarrow ((x = x_0) \vee (x = x_1) \vee \cdots \vee (x = x_n)).$$

Furthermore, for a formula φ, we write $X = \{x : \varphi(x)\}$ to mean that for all sets x,

$$x \in X \leftrightarrow \varphi(x).$$

Caution 1.3. For a given formula φ, we don't know a priori if there exists a set X such that $X = \{x : \varphi(x)\}$. In fact, it doesn't exist for some φ (see the discussion of Russel's paradox in Section 11). It is tempting to say that the existence of such a set X follows from Comprehension axiom, but the latter only gives the existence of sets of the form $\{x \in Y : \varphi(x)\}$ and not $\{x : \varphi(x)\}$. To apply Comprehension, one has to first find (prove existence of) an appropriate set Y.

Below we abandon using the awkward formal (symbolic) language of ZFC and start writing our definitions and theorems in ordinary (mathematical) English, with the understanding that if we had to, we could translate everything to the formal language.

2. Well-orderings

Definition 2.1. A binary relation $<$ on a set A is called an *ordering* (or a *strict ordering*) if for all $x, y, z \in A$,

(i) (Irreflexivity) $x \not< x$,

(ii) (Transitivity) $x < y < z \Rightarrow x < z$.

We will refer to the pair $(A, <)$ as an *ordering* or an *ordered set* if $<$ is an ordering of A.

Observation 2.2. *Conditions (i) and (ii) imply that orderings are asymmetric, i.e.*

$$x < y \Rightarrow y \not< x.$$

Remark 2.3. The term *ordering* is also often used for a binary relation $\leqslant$ on a set A that is reflexive, anti-symmetric, and transitive. This is why, the notion in Definition 2.1 is often referred to as a *strict ordering*. The two different notions are easily obtained from one another: indeed, given a nonstrict ordering $\leqslant$, one obtains its strict version by defining $a < b :\Leftrightarrow a \leqslant b$ and $a \neq b$; conversely, given a strict ordering, its nonstrict version is defined by $a \leqslant b :\Leftrightarrow a < b$ or $a = b$. In this note, we will only use strict orderings and hence we simply call it *ordering* throughout.

For an ordering $<$, we write $x \leqslant y$ to mean "$x < y$ or $x = y$".

Definition 2.4. An ordering $<$ on a set A is called *linear* (or *total*) if in addition we have:

(iii) (Totality) For all $x, y \in A$, either $x = y$ or $x < y$ or $y < x$.

A linear ordering $<$ is called a *well-ordering* if

(iv) (Well-foundedness) every subset of A has a least element x, i.e. for all $y \in A$, $x \leqslant y$. Note that such x is necessarily unique.

Definition 2.5. Let $(A, <)$ be an ordering. For $a \in A$, let $\mathrm{pred}(a, A, <)$ denote the set of predecessors of a, i.e.

$$\mathrm{pred}(a, A, <) := \{b \in A : b < a\}.$$

We simply write $\mathrm{pred}(a)$, when $(A, <)$ is clear from the context. A subset $B \subseteq A$ is called an *initial segment* if it is closed downward, i.e. for all $x \in B$, $\mathrm{pred}(x, A, <) \subseteq B$. Call B a *proper initial segment* if $B \neq A$.

Lemma 2.6. *If $(A, <)$ is a well-ordering, then any proper initial segment B is equal to* $\mathrm{pred}(a)$ *for some $a \in A$; in other words, B has a supremum.*

Proof. Take a to be the least element of $A \setminus B$. □

Definition 2.7. For ordered sets $(A, <_A), (B, <_B)$, call a function $f : A \to B$ an *isomorphism* of $(A, <_A)$ with $(B, <_B)$ (or just an *order-isomorphism*) if it is a bijection and for all $a_0, a_1 \in A$,

$$a_0 <_A a_1 \iff f(a_0) <_B f(a_1).$$

Say that $(A, <_A)$ and $(B, <_B)$ are *isomorphic*, and denote $(A, <_A) \simeq (B, <_B)$, if there is an isomorphism of $(A, <_A)$ with $(B, <_B)$.

Notation. If $(A, <)$ is a well-ordering and $B \subseteq A$, then $(B, <|_B)$ is also a well-ordering, where $<|_B := < \cap B \times B$. We will abuse the notation and write $(B, <)$ below.

For well-orderings $(A, <_A)$ and $(B, <_B)$, write $(A, <_A) \prec (B, <_B)$ if

$$(A, <_A) \simeq (\mathrm{pred}(b), <_B),$$

for some $b \in B$. Write $(A, <_A) \preceq (B, <_B)$ if $(A, <_A) \prec (B, <_B)$ or $(A, <_A) \simeq (B, <_B)$.

Observation 2.8. $(A, <_A) \preceq (B, <_B)$ *is equivalent to the existence of an isomorphism of* $(A, <_A)$ *with an initial segment of* $(B, <_B)$.

Lemma 2.9. *For well-orderings* $(A, <_A), (B, <_B)$*, there is at most one isomorphism of* $(A, <_A)$ *with an initial segment of* $(B, <_B)$.

Proof. Let B', B'' be initial segments of B (possibly equal) and suppose towards a contradiction that there are order-isomorphisms $f : A \to B', g : A \to B''$ such that $f \neq g$. Let a be the $<_A$-least element in A for which $f(a) \neq g(a)$. Without loss of generality, assume $f(a) <_B g(a)$. Then $f(a) \in B''$, so we can apply g^{-1} to the last inequality and get $a' := g^{-1}(f(a)) <_A a$. Hence, by the choice of a, $f(a') = g(a') = f(a)$, contradicting the injectivity of f. □

Corollary 2.10. $(A, <) \not\simeq (A, <)$ *for any well-ordering* $(A, <)$.

Proof. The identity map $a \mapsto a$ is already an isomorphism of $(A, <_A)$ with its initial segment, so there cannot be any other, by Lemma 2.9. □

Observation 2.11. *For any well-orderings* $(A, <_A), (B, <_B), (C, <_C)$,

(a) $(A, <_A) \prec (B, <_B) \preceq (C, <_C) \implies (A, <_A) \prec (C, <_C)$;

(b) $(A, <_A) \preceq (B, <_B) \prec (C, <_C) \implies (A, <_A) \prec (C, <_C)$.

Remark 2.12. Recall that functions are simply sets of pairs satisfying a certain condition. More precisely, a *function* $f : A \to B$ is a subset of $A \times B$ satisfying the following: for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$. (We commonly write $f(a) = b$ instead of awkward $(a, b) \in f$.) Thus, for example, the notation $f \subseteq g$ makes sense for functions f, g.

Theorem 2.13. *For well-orderings* $(A, <_A)$ *and* $(B, <_B)$*, exactly one of the following holds:*

(i) $(A, <_A) \simeq (B, <_B)$;

(ii) $(A, <_A) \prec (B, <_B)$;

(iii) $(B, <_B) \prec (A, <_A)$.

Proof. That these statements are mutually exclusive follows from Lemma 2.9 via Observation 2.11. To show that one of them holds, let

$$f := \{(a, b) \in A \times B : (\mathrm{pred}(a), <_A) \simeq (\mathrm{pred}(b), <_B)\}.$$

It is straightforward to check that f is a function from some initial segment A' of A to an initial segment B' of B. In fact, f is an isomorphism of $(A', <_A)$ with $(B', <_B)$, so it is enough to show that $A' = A$ or $B' = B$. But A' and B' cannot both be proper since otherwise $A' = \mathrm{pred}(a)$ and $B' = \mathrm{pred}(b)$ for some $a \in A \setminus A', b \in B \setminus B'$, whence $(a, b) \in f$ by the definition of f, so $a \in A'$, a contradiction. □

3. Ordinals

Simply put, ordinals are a generalization of natural numbers, suitable for listing (counting) infinite sets. They are certain well-ordered sets that serve as canonical representatives of isomorphism classes of well-ordered sets, meaning that every well-ordered set is isomorphic to exactly one ordinal (see Theorem 3.12). Besides, ordinals are very convenient to work with: for example, the relation $\prec$ between ordinals coincides with $\in$.

In ordinary mathematics, it is customary to denote sets by capital letters $A, B, \ldots$ and elements by small letters $a, b, \ldots$. In set theory however, sets themselves are often elements of other sets under consideration, so one has to abandon this tradition, as we do below.

Definition 3.1. A set x is called *transitive* if all of its elements are also its subsets, i.e. for all $y \in x$, $y \subseteq x$.

For example, the sets $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$ are transitive (check!), but the set $\{\{\emptyset\}\}$ isn't (why?).

Definition 3.2. A set α is called an *ordinal* if it is transitive and well-ordered by $\in$.

The transitive sets $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$ mentioned above are all ordinals; however, the set $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$ isn't an ordinal since $\in$ is not an ordering on it[1]. We usually denote ordinals by Greek letters α, β, γ, etc.

Lemma 3.3. *Let α be an ordinal.*

(a) $\alpha \notin \alpha$.
(b) *For any $y \in \alpha$, $y = \text{pred}(y, \alpha, \in)$.*
(c) *The least element of α is $\emptyset$ and we denote $0 := \emptyset$.*
(d) *Every $y \in \alpha$ is itself an ordinal.*

Proof. For part (a), recall that since $\in$ is an ordering on α, it must be irreflexive. Thus, for all $y \in \alpha$, $y \notin y$. Hence, if $\alpha \in \alpha$, taking $y = \alpha$ would give a contradiction. Part (b) follows from the transitivity of α and Axiom 1 (Extensionality). Part (c) follows directly from (b) since if y is the least element of α, then $\text{pred}(y, \alpha, \in) = \emptyset$. Part (d) is left as an exercise. □

For ordinals α, we will write α instead of $(\alpha, \in)$ even when we mean to view α as a well-ordered set.

Lemma 3.4. *Let α, β be ordinals.*

(a) *If $\alpha \simeq \beta$, then $\alpha = \beta$.*
(b) *$\in$ is a total order on ordinals, i.e. exactly one of the following holds: either $\alpha = \beta$, or $\alpha \in \beta$, or $\beta \in \alpha$.*
(c) *If $\alpha \subsetneq \beta$, then $\alpha \in \beta$.*

Proof. Part (a) is left as an exercise. For part (b), apply Theorem 2.13 and use (b) of Lemma 3.3 and (a) of the current lemma. Finally, (c) follows from (b) since $\alpha \subsetneq \beta$ and $\beta \in \alpha$ cannot both hold: otherwise, $\alpha \subsetneq \beta \subseteq \alpha$ and hence $\alpha \subsetneq \alpha$, a contradiction. □

In light of the above lemma, we often write $\alpha < \beta$ instead of $\alpha \in \beta$, for ordinals α, β. We also use the notation $\alpha \leqslant \beta$ to mean $\alpha = \beta$ or $\alpha < \beta$.

Lemma 3.5.

(a) *Every nonempty set of ordinals C has an $\in$-least element, i.e. there is $\alpha \in C$ such that for all $\beta \in C$ not equal to α, we have $\alpha \in \beta$. We denote this α by $\min C$.*
(b) *For every mathematical statement (formula) $\varphi(x)$, if there is an ordinal α for which $\varphi(\alpha)$ holds, then there is a least such ordinal.*
(c) *Every transitive set of ordinals is itself an ordinal.*

Proof. For (a), fix an ordinal $\gamma \in C$. If $\gamma \cap C = \emptyset$ then γ is the $\in$-least element in C and we are done. Otherwise, since $\gamma \cap C$ is a subset of γ and γ is well-ordered by $\in$, there is an $\in$-least element α in $\gamma \cap C$. Clearly, α would also be the $\in$-least element in C.

The proof of (b) is identical to that of (a); in fact, the latter follows from the former by letting $\varphi(x)$ be $x \in C$.

Part (c) follows from (b) of Lemma 3.4 and (a) of the current lemma. □

[1] The author thanks Minh Tran for this quick example.

Remark 3.6. Part (b) above allows us to freely use definitions like "Let κ be the least ordinal such that $\varphi(\kappa)$", without having to spot a set (the set C in part (a)) that would contain this κ as we would have to do if we used (a) instead.

We now define a successor operation for ordinals.

Definition 3.7. For a set x, we define the *successor* $S(x)$ of x to be the set $x \cup \{x\}$.

Note that $x \in S(x)$ and $x \subseteq S(x)$.

Lemma 3.8. *For an ordinal α, $S(\alpha)$ is the least ordinal greater than α.*

Proof. First note that $S(\alpha)$ is a transitive set of ordinals and thus is an ordinal by (c) of Lemma 3.5. By definition, $\alpha < S(\alpha)$. Now, assuming $\beta > \alpha$, we show that $S(\alpha) \leqslant \beta$. Indeed, by the transitivity of β, we have $\alpha \subseteq \beta$ and thus $S(\alpha) \subseteq \beta$. Therefore, either $S(\alpha) = \beta$ or $S(\alpha) < \beta$ by (c) of Lemma 3.4. □

Definition 3.9. An ordinal $\alpha \neq 0$ is called a *successor* if there is an ordinal β such that $\alpha = S(\beta)$; otherwise, we would say that α is a *limit* ordinal.

Thus, there are three types of ordinals: 0, successor ordinals, and limit ordinals.

Definition 3.10. For a set of ordinals C, define $\sup C = \bigcup C := \bigcup_{\alpha \in C} \alpha$.

Lemma 3.11. *Let C be a set of ordinals.*

(a) $\sup C$ *is the least ordinal β such that for all $\alpha \in C$, $\alpha \leqslant \beta$.*

(b) *If C is transitive (and hence an ordinal), then* $\sup C \notin C$ *if and only if C is a limit ordinal or $\emptyset$.*

(c) $\min C = \bigcap C := \bigcap_{\alpha \in C} \alpha$.

Proof. Left as an exercise. □

Theorem 3.12. *Any well-ordering $(A, <_A)$ is isomorphic to exactly one ordinal.*

Proof. Uniqueness follows from (a) of Lemma 3.4. For existence, let

$$A' = \{a \in A : \text{there is an ordinal } \alpha \text{ such that } \operatorname{pred}(a, A, <_A) \simeq \alpha\}.$$

By uniqueness, the ordinal α in the definition of A' is unique for each $a \in A$. Thus, by Axiom 8 (Replacement), there is a set C and a function $f : A' \to C$ such that for all $a \in A'$, $f(a)$ is an ordinal and $\operatorname{pred}(a, A, <_A) \simeq f(a)$. By shrinking C, we may assume that f is surjective. Note that A' is an initial segment of A and hence C is transitive. Thus, by (c) of Lemma 3.5, C is an ordinal. It is also easy to see that f is an isomorphism between $(A', <_A)$ and C. Now if $A' \neq A$, then $A' = \operatorname{pred}(a, A, <_A)$ for some $a \in A \setminus A'$. But then $\operatorname{pred}(a, A, <_A) \simeq C$ and hence $a \in A'$, a contradiction. Thus, $A' = A$ and we are done. □

Definition 3.13. For a well-ordering $(A, <_A)$ the unique ordinal α with $(A, <_A) \simeq \alpha$ is called the *order type* of $(A, <_A)$ and denoted by $\operatorname{tp}(A, <_A)$.

4. Transfinite induction

The usual proof of the induction theorem for $\mathbb{N}$ given in an undergraduate course uses the fact that the usual ordering on $\mathbb{N}$ is a well-ordering. The same proof works for any well-ordered set:

Theorem 4.1 (Transfinite induction)**.** *Let $(A, <)$ be a well-ordered set and let $P \subseteq A$ be a set such that for any $a \in A$,* $\operatorname{pred}(a, A, <) \subseteq P$ *implies $a \in P$. Then $P = A$.*

Proof. Assume for contradiction that $P \neq A$ and let a be the least element of $A \setminus P$. By our choice, $\operatorname{pred}(a, A, <) \subseteq P$, and hence $a \in P$, a contradiction. □

Remark 4.2. In the proofs using transfinite induction on ordinals, the cases of 0, successor ordinals and limit ordinals are usually considered separately. An example is the proof of Theorem 4.4 below.

It is common in mathematics to define sequences (indexed by $\mathbb{N}$) inductively.[2] Such an example is the Fibonacci sequence. In set theory however, we often want to inductively define sequences that are indexed by a larger well-ordered set (typically an ordinal). The next theorem shows that this indeed can be done. For notational simplicity, we state it for ordinals, but the same proof would work for arbitrary well-ordered sets.

First, note (realize?) that a sequence indexed by an ordinal κ, is nothing but a function defined on κ. The idea of inductive definition of such function is that in order to define its value at a point $\alpha \in \kappa$, we use its values at the predecessors of α. To make this precise, we need the following:

Definition 4.3. For sets A, B, a *partial function* from A to B is a usual function from a subset of A to B. We denote this by $f : A \rightharpoonup B$, and we denote the domain of f by $\mathrm{dom}(f)$. We denote the set of partial functions from A to B by $\mathrm{Partial}(A, B)$.

Theorem 4.4 (Transfinite recursion)**.** *For every ordinal κ, set A, and map*[3] *$F : \mathrm{Partial}(\kappa, A) \times \kappa \to A$, there is a unique function $f : \kappa \to A$ such that for every $\alpha \in \kappa$,*

$$f(\alpha) = F(f|_\alpha, \alpha).$$

Proof. We prove by transfinite induction on $S(\kappa)$ that for all $\gamma \leqslant \kappa$, there is a unique function $f_\gamma : \gamma \to A$ satisfying

$$f_\gamma(\alpha) = F(f_\gamma|_\alpha, \alpha), \tag{$*$}$$

for every $\alpha < \gamma$. Granted this, we would take $f = f_\kappa$ and be done.

For $\gamma = 0$, put $f_\gamma = \emptyset$ and note that this is the only function defined on $\emptyset$ in general. For γ a limit ordinal, it follows from uniqueness that for all $\alpha < \beta < \gamma$, $f_\alpha \subseteq f_\beta$ (see Remark 2.12). Thus $f_\gamma = \bigcup_{\beta<\gamma} f_\beta$ is a function from γ to A and it clearly satisfies $(*)$ as so do the functions $f_\beta, \beta < \gamma$. The uniqueness of f_γ also follows from the uniqueness of f_β for each $\beta < \gamma$.

For the successor case, let $\gamma = S(\beta)$. For $\alpha \in \gamma$, put

$$f_\gamma(\alpha) = \begin{cases} F(f_\beta, \beta) & \text{if } \alpha = \beta \\ f_\beta(\alpha) & \text{otherwise} \end{cases}.$$

It is clear that f_γ satisfies $(*)$ and that it is the unique such function. □

Remark 4.5. The proof of Theorem 4.4 shows that its conclusion holds even if the map F is only defined at points $(f, \alpha) \in \mathrm{Partial}(\kappa, A) \times \kappa$ with $\alpha = \mathrm{dom}(f)$, which is what usually happens in practice.

5. The set of natural numbers

Having defined 0 and the successor operation, we can attempt to define natural numbers as follows:

$$0 := \emptyset, 1 := S(0), 2 := S(1), ..., n + 1 := S(n), ...$$

But wait, I just used induction on the set of natural numbers while trying to define natural numbers!!! Not good.

It is still ok to define $1 := S(\emptyset)$, $2 := S(S(\emptyset))$, even $7 := S(S(S(S(S(S(S(\emptyset)))))))$, but doing so, we won't be able to define all of the natural numbers at once. We can try a top-to-bottom approach instead; that is, isolate which ordinals we would like to be called natural numbers.

[2] The correct word here is "recursively".

[3] One can relax the hypothesis by replacing sets with classes, see Section 11.

Definition 5.1. An ordinal α is called a *natural number*, if all $\beta \leqslant \alpha$ are successor ordinals or 0.

Now we want to define the set of natural numbers. It is tempting to say "Put $\mathbb{N} := \{\alpha : \alpha \text{ is a natural number}\}$", but the problem is that this may not be a set! In other words, this type of definitions is not allowed in general as it may lead to contradictions (see section Classes below). What is allowed is postulated by Axiom 5 (Comprehension) and in order to use this axiom schema, we need to know a priori that there exists a set y that contains all of the natural numbers. Using $\emptyset$ and set-theoretic operations like pairing, union, powerset, etc., we won't be able to construct such a set. This is why the existence of such a set is outright postulated in ZF and is called Axiom of Infinity (Axiom 6). It says the following:

There is a set y such that $\emptyset \in y$ and for every $x \in y$, $S(x) \in y$.

Definition 5.2. Call a set y *inductive*, if it satisfies $\emptyset \in y$ and for every $x \in y$, $S(x) \in y$.

Lemma 5.3. *Every inductive y contains all of the natural numbers.*

Proof. Assume for contradiction that there is a natural number n (in the sense of Definition 5.1) that is not in y, and let m be the least ordinal in $S(n)$ that does not belong to y.[4] m cannot be $\emptyset$ as $\emptyset \in y$, so m must be a successor ordinal. Thus $m = S(k)$ for some natural number k. By the choice of m, $k \in y$ and hence $m \in y$, a contradiction. □

Using this lemma and Axiom 6 (Infinity), we can apply Axiom 5 (Comprehension) and get

$$\omega := \{\alpha \in y : \alpha \text{ is a natural number}\}.$$

We will also use $\mathbb{N}$ for ω when we don't necessarily want to view it as an ordinal.

Proposition 5.4.

(a) *ω is a limit ordinal and is the least such.*
(b) *ω is an inductive set.*
(c) *ω is the $\subseteq$-least inductive set.*

Proof. Left as an exercise. □

Remark 5.5. Part (c) of this proposition is the statement of the (usual) induction theorem for natural numbers. It says in particular that if $P \subseteq \omega$ has the property that $0 \in P$ and $n \in P \Rightarrow S(n) \in P$, for all $n \in \omega$, then $P = \omega$.

6. Equinumerosity

There are two ways to figure out whether the number of students is equal to the number of chairs in the classroom:

(1) Count the students and count the chairs, and see if those numbers are equal.
(2) Ask the students to each pick a chair and sit on it. If there are standing students or empty chairs, then the answer is no; otherwise, it is yes.

In set theory we take the second approach because that's the one that works for infinite sets.

Definition 6.1. Two sets A, B are called *equinumerous*, noted $A \equiv B$, if there is a bijection between them. We write $A \sqsubseteq B$ if A is equinumerous with a subset of B, i.e. A injects into B. Finally, we write $A \sqsubset B$ if $A \sqsubseteq B$ but $A \not\equiv B$.

[4] We used Axiom 5 (Comprehension) here.

We leave it as an exercise[5] to prove that

$$\mathbb{N} \equiv \mathbb{Z} \equiv \mathbb{Q}$$

and

$$\mathbb{R} \equiv (0,1) \equiv [0,1] \equiv 2^{\mathbb{N}} \equiv \mathscr{P}(\mathbb{N}).$$

Notation: For sets A, B, a function $f : A \to B$ and a subset $C \subseteq A$, let $f''C$ or $f''(C)$ denote the image of C under f, i.e.

$$f''C = \{b \in B : \exists a \in C(b = f(a))\}.$$

Here is a very easy but relevant lemma:

Lemma 6.2. *Let A, B be nonempty sets. Any injection $f : A \to B$ has a surjective left inverse $g : B \to A$, i.e. $g \circ f = \mathrm{id}_A$.*

Proof. Fix $a \in A$ and define $g : B \to A$ by setting, for $b \in B$,

$$g(b) = \begin{cases} f^{-1}(b) & \text{if } b \in f''A \\ a & \text{otherwise.} \end{cases}$$

□

The following can be safely considered as the first theorem of set theory.

Theorem 6.3 (Cantor)**.** *Let A be a set.*

(a) *There is no surjection $f : A \twoheadrightarrow \mathscr{P}(A)$.*
(b) $\mathscr{P}(A) \not\sqsubseteq A$.

Proof. Part (b) follows from (a) by Lemma 6.2. For (a), assume that there is such surjection f and let $B = \{a \in A : a \notin f(a)\}$. Since f is surjective, there is $b \in A$ such that $f(b) = B$. Now either $b \in f(b)$ or $b \notin f(b)$, and both cases yield a contradiction. □

Definition 6.4. A set A is said to be *countable* if $A \sqsubseteq \mathbb{N}$.

Cantor's theorem in particular implies that $\mathscr{P}(\mathbb{N})$ (and hence also $\mathbb{R}$) is uncountable.

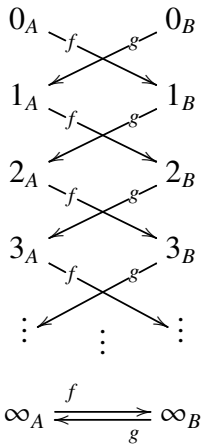
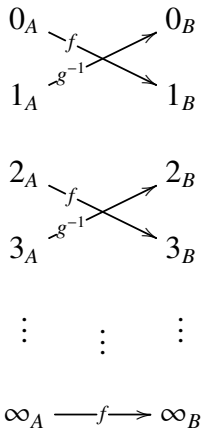
The following theorem helps when proving equinumerosity of sets.

Theorem 6.5 (Cantor–Schröder–Bernstein)**.** *For sets A, B, if $A \sqsubseteq B$ and $B \sqsubseteq A$, then $A \equiv B$.*

Proof. Prototypical case: Let $A = B = \mathbb{N} \cup \{\infty\}$, $f(n) := g(n) := n + 1$, for $n \in \mathbb{N}$, and $f(\infty) := g(\infty) := \infty$. To distinguish the two copies of $\mathbb{N} \cup \{\infty\}$, denote the elements n of A and B by n_A and n_B, respectively. We define a bijection $h : A \to B$ as follows:

$$h(a) := \begin{cases} f(a) & \text{if } a = 2n \text{ or } a = \infty \\ g^{-1}(a) & \text{if } a = 2n + 1. \end{cases}$$

[5] Read Theorem 6.5 first.

Schematically, from  we obtain 

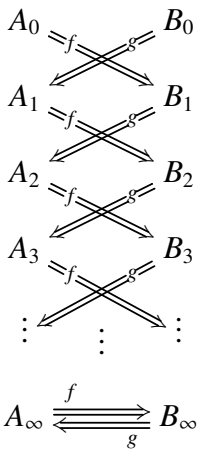
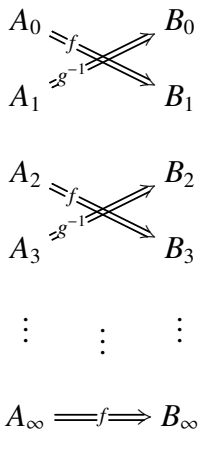
General case: Let $f : A \to B$ and $g : B \to A$ be injections. We prove by reducing this to the prototypical case as follows: we will obtain partitions

$$A = \bigsqcup_{n\in\mathbb{N}} A_n \sqcup A_\infty \text{ and } B = \bigsqcup_{n\in\mathbb{N}} B_n \sqcup B_\infty$$

such that $f'' A_n = B_{n+1}$ and $g'' B_n = A_{n+1}$ for each $n \in \mathbb{N}$, as well as $f'' A_\infty = B_\infty$, so we define a desired bijection $h : A \to B$ just like in the prototypical example, namely, for $a \in A$,

$$h(a) := \begin{cases} f(a) & \text{if } a \in A_{2n} \text{ or } a \in A_\infty \\ g^{-1}(a) & \text{if } a \in A_{2n+1} \end{cases}$$

and have basically the same picture:

from  obtain 

To carve out such partitions, we define recursively decreasing sequences $(\bar{A}_n)_{n\in\mathbb{N}}$ and $(\bar{B}_n)_{n\in\mathbb{N}}$ by $\bar{A}_0 := A$, $\bar{B}_0 := B$,

$$\bar{B}_{n+1} := f'' \bar{A}_n, \text{ and } \bar{A}_{n+1} := g'' \bar{B}_n,$$

and let $A_n := \bar{A}_n \setminus \bar{A}_{n+1}$, $B_n := \bar{B}_n \setminus \bar{B}_{n+1}$, as well as

$$A_\infty := \bigcap_{n\in\mathbb{N}} \bar{A}_n \text{ and } B_\infty := \bigcap_{n\in\mathbb{N}} \bar{B}_n.$$

It remains to verify that the partitions $A = \bigsqcup_{n\in\mathbb{N}} A_n \sqcup A_\infty$ and $B = \bigsqcup_{n\in\mathbb{N}} B_n \sqcup B_\infty$ are as desired. The injectivity of f implies that f'' commutes with set-subtraction and intersections, so we have

$f''A_n = f''(\bar{A}_n \setminus \bar{A}_{n+1}) = f''\bar{A}_n \setminus f''\bar{A}_{n+1} = \bar{B}_{n+1} \setminus \bar{B}_{n+2} = B_{n+1}$ and

$$f''A_\infty = f''\left(\bigcap_{n\in\mathbb{N}} \bar{A}_n\right) = \bigcap_{n\in\mathbb{N}} f''\bar{A}_n = \bigcap_{n\in\mathbb{N}} \bar{B}_{n+1} = \bigcap_{n\in\mathbb{N}} \bar{B}_n = B_\infty.$$

Similarly, we have the analogous statements for g, which finishes the proof. □

7. Finite/infinite sets and Axiom of Choice

Which sets should be called finite? Intuitively, a set is finite if when we count its elements, we get a natural number. Here is a formal version of this definition:

Definition 7.1. A set A is said to be *finite* if it is equinumerous with some natural number. Otherwise, we say that it is *infinite*.

There are a few other intuitive definitions of *infinite* and below we explore the relations between them.

Definition 7.2 (Dedekind)**.** A set A is called *Dedekind infinite* if it is equinumerous with a proper subset of itself, i.e. there is an injection $f : A \to A$ with $f''A \subsetneq A$. Otherwise, A is said to be *Dedekind finite*.

Proposition 7.3. *Finite sets are Dedekind finite.*

Proof. It is enough to prove that natural numbers are Dedekind finite, and we do this by induction on ω. It is trivial that $\emptyset$ is Dedekind finite since it does not have proper subsets. Assume n is Dedekind finite and show that so is $n + 1$. It is enough to show that every injection from $n + 1$ to $n + 1$ is surjective. Let $f : n + 1 \to n + 1$ be an injection. We first claim that without loss of generality we may assume that $f''n \subseteq n$: otherwise, $f(n) < n$ and $m := f^{-1}(n) < n$, and by swapping the values of $f(m)$ and $f(n)$ we will achieve $f''n \subseteq n$. Since $f|_n : n \to n$ is injective, it must also be surjective by the induction hypothesis. Thus $f(n)$ must be equal to n, and hence f is surjective as well. □

One could also define infinite sets to be exactly those sets A, for which $\omega \sqsubseteq A$. The following proposition shows that this notion coincides with Dedekind infinite.

Proposition 7.4. *A set A is Dedekind infinite if and only if $\omega \sqsubseteq A$. In particular, ω is Dedekind infinite.*

Proof. $\Rightarrow$: Let $f : A \to A$ be an injection such that $f''A \subsetneq A$. For each $n \in \mathbb{N}$, put $B_n := (f^n)''(A \setminus f''A)$. Because f is injective, it respects the setminus operation and thus, $f''B_n = (f^n)''(A) \setminus (f^{n+1})''(A)$, so, the sets B_n are pairwise disjoint. Fix $x \in B_0$ and define $g : \omega \to A$ by $n \mapsto f^n(x)$. This g is injective because $g(n) \in B_n$.
$\Leftarrow$: Instead of a proof, I will tell you the story of Hilbert's hotel. It has ω-many rooms and all of them are occupied. However, a new guest comes and demands a room. To resolve the situation, the hotel manager simply asks all of the tenants to move to the room next to their current room with one higher number, thus making room number 0 available for the new guest. □

There is yet another notion of infinite for a set A, namely, the existence of a surjection $A \twoheadrightarrow \omega$. By Lemma 6.2, $\omega \sqsubseteq A$ implies $A \twoheadrightarrow \omega$.

To summarize, we have mentioned the following notions of infinite for a set A:

(1) A is not equinumerous with a natural number.
(2) A is Dedekind infinite.
(3) $\omega \sqsubseteq A$.
(4) $A \twoheadrightarrow \omega$.

So far, we have shown that (2)⇒(1), (2)⇔(3), and (3)⇒(4). Our intuition (at least mine) tells us that all of these notions should be equivalent. Let's attempt to prove for example that (1)⇒(3).

Attempt to prove (1)⇒(3). Suppose that A is not equinumerous with a natural number, and try to recursively define an embedding $f : \omega \to A$ as follows: fix $a_0 \in A$, and for $n \in \omega$, put

$$f(n) := \begin{cases} \text{some element from } A \setminus ((f|_n)''n) & \text{if } A \setminus ((f|_n)''n) \neq \emptyset \\ a_0 & \text{otherwise.} \end{cases}$$

The first condition must hold since otherwise $f|_n$ would be a bijection from n to A, contradicting our hypothesis. Thus f is injective and we are done.

However, something is wrong with the above recursive definition. Namely, the right side is not a (predefined) function. The expression "some element from $A \setminus ((f|_n)''n)$" does not define a correspondence $n \mapsto a_n \in A \setminus ((f|_n)''n)$ for all $n \in \omega$ *at once*. What would fix our problem[6] is if we had a "choice" function $c : \mathscr{P}(A) \to A$ such that for each nonempty $B \subseteq A$, $c(B) \in B$. Then we would define our f by

$$f(n) = c(A \setminus (f|_n)''n),$$

for $n \in \omega$, and this f would clearly work.

Similarly, if we had such a choice function $c : \mathscr{P}(A) \to A$, we could prove that (4)⇒(3) as follows: if $f : A \to \omega$ is a surjection, then define $g : \omega \to A$ by $a \mapsto c(f^{-1}(a))$. Clearly g is injective.

Thus, the existence of a choice function $c : \mathscr{P}(A) \to A$ would prove the equivalence of (1)-(4). Does such a function always exist? This is not postulated by any of the axioms of ZF, and in fact, the existence of such a choice function cannot be proven from ZF[7]. Hence, we need an extra axiom (Axiom 9), called Axiom of Choice (AC), that states the following:

If A is a set whose elements are nonempty sets, then there exists a choice function; that is, a function $f : A \to \bigcup A$ such that for every $x \in A$, $f(x) \in x$.

The proof system ZF+AC is called the Zermelo–Fraenkel set theory with Choice and denoted by ZFC. Below, all of the statements or exercises that use AC are marked with (AC).

Using the above attempts to prove (1)⇒(3) and (4)⇒(3), we now get

Proposition 7.5 (AC)**.** *For a set A, the following are equivalent:*

(1) *A is not equinumerous with a natural number.*
(2) *A is Dedekind infinite.*
(3) *$\omega \sqsubseteq A$.*
(4) *$A \twoheadrightarrow \omega$.*

For this and many other reasons, most mathematicians work in ZFC rather than in ZF.

Remark 7.6. To prove (1)⇒(3) and (4)⇒(3), one does not need the full power of AC: it is enough to assume a weaker axiom, called Dependant Choice (DC), which (roughly speaking) gives a function that makes a choice depending on the previously made choices. This is exactly what we used in the attempt to prove (1)⇒(3).

[6] I'm not saying this is absolutely necessary; one could get away with something a bit weaker.

[7] This is a highly nontrivial theorem.

8. Cardinals and cardinality

Definition 8.1. An ordinal α is called a *cardinal* if there is no ordinal $\beta < \alpha$ such that $\beta \equiv \alpha$.

Proposition 7.3 implies the following:

Corollary 8.2.
(a) *Every natural number is a cardinal.*
(b) *ω is a cardinal.*

Definition 8.3. Let A be a set and κ a cardinal. A set A is said to have *cardinality* κ if $A \equiv \kappa$. We denote this by $|A| = \kappa$. We say that A *has cardinality* if it has cardinality λ for some cardinal λ.

Does every set have cardinality? The following lemma gives a reformulation of this question:

Lemma 8.4. *A set A has cardinality if and only if it can be well-ordered.*

Proof. $\Rightarrow$: If there is a bijection $f : A \to \kappa$ for some cardinal κ, just copy the well-ordering of κ to A via f.
$\Leftarrow$: If there is a well-ordering $<_A$ on A, then by Theorem 3.12, there is an ordinal α such that $(A, <_A) \simeq \alpha$. Put

$$\kappa = \min \{\beta \leqslant \alpha : A \equiv \beta\}.$$

It is clear that κ is a cardinal and $|A| = \kappa$. □

Thus, the real question is whether every set can be well-ordered or not. Think of $\mathbb{R}$ for example: the natural ordering of $\mathbb{R}$ is a linear ordering but not a well-ordering. In fact, it cannot be proved in ZF that $\mathbb{R}$ can be well-ordered. More generally, it turns out that the statement "every set can be well-ordered" is equivalent to AC and is known as Zermelo's theorem. We state this in Theorem 8.6 together with an equivalent formulation called Zorn's lemma. First, we need the following:

Definition 8.5. Let A be a set and $<$ be an ordering on it. An element $a \in A$ is called an *upper bound* for a subset $B \subseteq A$ if for any $b \in B$, $b \leqslant a$.[8] B is called a *chain* if $<|_B$ is a linear ordering on B. An element $a \in A$ is called *maximal* if there is no element $b \in A$ with $a < b$.

Theorem 8.6. *Each of the following statements is equivalent to* AC*:*
1. **Zorn's Lemma:** *For every ordered set $(A, <)$ whose every chain has an upper bound, there is a maximal element in A.*
2. **Zermelo's Theorem:** *Every set can be well-ordered.*

Proof. The proof is outlined in Exercises 9 to 11. □

Corollary 8.7 (AC)**.** *Every set has cardinality.*

9. Uncountable cardinals and the Continuum Hypothesis

We now consider the question of whether there exists uncountable cardinals; more generally, given a cardinal κ, is there a cardinal greater than κ? By Cantor's theorem, we have $\kappa \sqsubset \mathscr{P}(\kappa)$. By AC, $\mathscr{P}(\kappa)$ has cardinality and thus $\lambda := |\mathscr{P}(\kappa)| > \kappa$. This shows that the answer to the question is positive in ZFC.

Can we prove this in ZF? The answer is still yes, and it follows from the following theorem:

Theorem 9.1 (Hartogs)**.** *For every set A, there is a cardinal κ such that $\kappa \not\sqsubseteq A$. We denote the least such κ by $\aleph(A)$.*

[8] As usual, $\leqslant$ means $<$ or $=$.

Proof. Put

$$\mathrm{WO}(A) := \left\{(B, <_B) \in \mathscr{P}(A) \times \mathscr{P}(A^2) : B \subseteq A \text{ and } <_B \text{ is a well-ordering on } B\right\},$$

and, recalling Definition 3.13, put

$$\chi(A) = \{\mathrm{tp}(B, <_B) : (B, <_B) \in \mathrm{WO}(A)\};$$

this set exists by Axiom 8 (Replacement). Clearly $\chi(A)$ is a transitive set of ordinals, and hence an ordinal itself, by (c) of Lemma 3.5. To show that $\chi(A) \not\subseteq A$, assume for contradiction that there is an injection $f : \chi(A) \to A$ and put $B = f''\chi(A)$. Let $<_B$ be the push-forward of $\in|_{\chi(A)}$ to a relation on B, i.e. for $b_1, b_2 \in B$,

$$b_1 <_B b_2 \iff f^{-1}(b_1) \in f^{-1}(b_2).$$

By definition, f is an isomorphism from $(\chi(A), \in)$ to $(B, <_B)$, and hence $<_B$ is a well-ordering on B. Therefore, $(B, <_B) \in \mathrm{WO}(A)$ and thus $\mathrm{tp}(B, <_B) \in \chi(A)$. But by uniqueness, $\mathrm{tp}(B, <_B) = \chi(A)$, so $\chi(A) \in \chi(A)$, contradicting part (a) of Lemma 3.3.

A similar argument shows that $\chi(A)$ is the least ordinal with $\chi(A) \not\subseteq A$, and hence a cardinal. □

Remark 9.2. There is a version of Hartogs' theorem that doesn't require Axiom 8 (Replacement). It is stated in Exercise 13.

Notation 9.3. For a cardinal κ, put $\kappa^+ := \chi(\kappa)$. Thus κ^+ is the least cardinal greater than κ. In general, for every ordinal α, we can define the α^{th} infinite cardinal $\aleph_\alpha$ by recursion on α as follows:

$$\aleph_\alpha = \begin{cases} \omega & \text{if } \alpha = 0 \\ \aleph_\beta^+ & \text{if } \alpha = S(\beta) \\ \sup_{\beta<\alpha} \aleph_\beta & \text{if } \alpha \text{ is a limit} \end{cases}.$$

For small cardinals such as $\aleph_0, \aleph_1, \aleph_2$, it is also common to use the notation $\omega_0, \omega_1, \omega_2$, instead.

Continuum hypothesis. One of the most notorious questions in set theory (or in mathematics in general) is whether $\omega_1 = |\mathscr{P}(\mathbb{N})|$; equivalently, whether $\omega_1 = |\mathbb{R}|$. The positive answer to this is known as the Continuum Hypothesis (CH). In 1940, Gödel showed that it is consistent with ZFC that CH holds. On the other hand, Cohen proved in 1963 that the failure of CH is also consistent with ZFC, thus showing the independence of CH from ZFC. To do this, Cohen introduced his famous method of forcing, which, together with its numerous variations, has since been the main method of proving independence results.

10. Cofinality and Kőnig's theorem

Definition 10.1. For κ a cardinal and β an ordinal, a sequence $(x_\alpha)_{\alpha<\beta}$ of ordinals $x_\alpha \in \kappa$ is called *cofinal* in κ if for every $\gamma < \kappa$, there is an index $\alpha < \beta$ with $x_\alpha \geqslant \gamma$. We refer to β as the *length* of this sequence and define the cofinality of κ to be the length of the shortest cofinal sequence in κ, i.e. $\mathrm{cof}(\kappa) :=$ the least ordinal β such that there is a cofinal sequence $(x_\alpha)_{\alpha<\beta}$ in κ.

For example, $\mathrm{cof}(\omega) = \omega$, $\mathrm{cof}(\aleph_\omega) = \omega$, but $\mathrm{cof}(\omega_1) = \omega_1$ (the proof of this is outlined in the exercises below).

Definition 10.2. A cardinal κ is called *regular* if $\mathrm{cof}(\kappa) = \kappa$; otherwise, it is called *singular*.

In the examples above, w and w_1 are regular. In general, for an infinite cardinal κ, κ^+ is always regular (this is also outlined in the exercises below). However, $\aleph_\omega$ is singular.

Cantor's theorem implies that $\kappa \sqsubset 2^\kappa$ and thus $\kappa \sqsubset \kappa^\kappa$ for an infinite cardinal κ. What about $\kappa^{\mathrm{cof}(\kappa)}$? Do we still have $\kappa \sqsubset \kappa^{\mathrm{cof}(\kappa)}$? The following theorem generalizes Cantor's theorem and gives an affirmative answer to the latter question.

Notation 10.3. Let I be a set and $(A_i)_{i\in I}$ be a sequence of sets. We denote by $\bigsqcup_{i\in I} A_i$ the disjoint union of A_i, i.e. $\bigsqcup_{i\in I} A_i := \bigcup_{i\in I} \{i\} \times A_i$.

Theorem 10.4 (Kőnig)**.** *Let I be a set and $(A_i)_{i\in I}, (B_i)_{i\in I}$ be sequences. If $A_i \sqsubset B_i$ for all $i \in I$, then $\bigsqcup_{i\in I} A_i \sqsubset \prod_{i\in I} B_i$.*

We prove this theorem below, but first note that this is indeed a generalization of Cantor's theorem since applying it to $I = \omega$, $A_i = 1$ and $B_i = 2$ (for all $i \in \omega$), gives $\omega \sqsubset 2^\omega$. Moreover, we also get:

Corollary 10.5. *$\kappa \sqsubset \kappa^{\mathrm{cof}(\kappa)}$, for every infinite cardinal κ. In particular, $\aleph_\omega \sqsubset \aleph_\omega^\omega$.*

Proof. Put $I = \mathrm{cof}(\kappa)$ and let $(A_i)_{i\in I}$ be a cofinal sequence in κ, and thus, $\bigcup_{i\in I} A_i = \kappa$. Putting $B_i = \kappa$, it is clear that $A_i \sqsubset B_i$. Hence, Kőnig's theorem implies $\kappa = \bigcup_{i\in I} A_i \sqsubseteq \bigsqcup_{i\in I} A_i \sqsubset \prod_{i\in I} B_i = \kappa^{\mathrm{cof}(\kappa)}$. □

Proof of Theorem 10.4. First, we show that $\bigsqcup_{i\in I} A_i \sqsubseteq \prod_{i\in I} B_i$. Using AC, we get a sequence $(f_i)_{i\in I}$ of injections $f_i : A_i \hookrightarrow B_i$. Note that because $B_i \not\subseteq A_i$, none of these f_i is surjective, so $B_i \setminus f_i'' A_i \neq \emptyset$. Applying AC again, we get a sequence $x = (x_i)_{i\in I} \in \prod_{i\in I} B_i$ such that $x_i \in B_i \setminus f_i'' A_i$ for each $i \in I$. Finally, we define $f : \bigsqcup_{i\in I} A_i \to \prod_{i\in I} B_i$ by

$$f(i,a)(j) := \begin{cases} f_i(a) & \text{if } j = i \\ x_j & \text{otherwise.} \end{cases}$$

It is easy to check that f is injective and thus $\bigsqcup_{i\in I} A_i \sqsubseteq \prod_{i\in I} B_i$.

Now suppose towards a contradiction that there is a surjection $g : \bigsqcup_{i\in I} A_i \twoheadrightarrow \prod_{i\in I} B_i$. We will define $x \in \prod_{i\in I} B_i$ that is not in the image of g.

The main point is that the map $g_i : A_i \to B_i$ defined by $a \mapsto g(i,a)(i)$ cannot be surjective by the hypothesis, so $B_i \setminus g_i'' A_i \neq \emptyset$ and we can choose a point from it. To this end, fix a choice function $\pi : \mathscr{P}(\bigsqcup_{i\in I} B_i) \setminus \{\emptyset\} \to \bigsqcup_{i\in I} B_i$. Now for each $i \in I$, we define the value $x(i)$ such that it ensures x is not in the g-image of $\{i\} \times A_i$, namely: $x(i) := \pi\left(B_i \setminus g_i'' A_i\right)$. It is clear that x is not in the image of g, for if $x = g(i,a)$ for some $i \in I$ and $a \in A_i$, then $x(i) \neq g_i(a) = g(i,a)(i)$. □

11. Classes

This section can be safely skipped by an impatient reader. It was mainly written to satisfy the curiosity of those readers (in particular, the author's significant other), who have heard the phrase "It's a proper class, not a set." and have always wondered what it means.

Given a mathematical statement (formula) $\varphi(x)$, one may wonder if there exists a set y such that

$$y = \{x : \varphi(x)\},$$

i.e. y contains all of the sets (in the world) that satisfy $\varphi(x)$. The answer of course depends on φ. For example, if $\varphi(x)$ is the statement "$x \in z$", for some set z, then such a set y exists, namely $y = z$. Another such example is $\varphi(x) \equiv$ "$\forall y \in x (y \neq y)$" as $\{x : \varphi(x)\} = \emptyset$. On the other hand, if $\varphi(x)$ is the statement "$x \notin x$", then such y doesn't exist since if it did, then either $y \in y$ or $y \notin y$, and both yield a contradiction. (This is the famous Russell's paradox.)

However, when writing proofs or theorems, it is often convenient to refer to $\{x : \varphi(x)\}$ as if it was a set (even when it isn't) and call it a *class*. This is just abuse of language and is not part of formal mathematics. The word class stands for the totality of all sets satisfying the formula φ, and it can be eliminated from any statement that contains it, although this usually makes statements longer. For example, for a class $C = \{x : \varphi(x)\}$, the statement "$x \in C$" really means "x satisfies φ". A perhaps more convincing example is the statement "The class ON of all ordinals is well-ordered by $\in$." as it is equivalent to the statement of (b) of Lemma 3.4 together with (b) of Lemma 3.5.

The classes that are not equal to sets (like the class defined by "$x \notin x$") are called proper classes. The intuition is that proper classes are "too big" to be sets. The existence of such classes comes from the fact that it is not postulated by ZF (or ZFC) that for a given formula $\varphi(x)$, there must be a set y such that $y = \{x : \varphi(x)\}$. What is postulated is Axiom Schema of Comprehension (Axiom 5), which we recall here:

For every set y, there is a set z such that $z = \{x \in y : \varphi(x)\}$.

Thus, to avoid ending up with a proper class when trying to define a set, we make sure that all of the elements we want to put in our set a priori belong to some (perhaps larger) set. In rare cases when this is not possible, we may need to use Axiom 8 (Replacement) instead, as we did in the proof of Theorem 3.12.

Proposition 11.1. *The following classes are proper:*

(a) $R = \{x : x \notin x\}$;

(b) $V = \{x : x = x\}$;

(c) $\mathrm{ON} = \{x : x \text{ is an ordinal}\}$.

Proof. Part (a) is left as an exercise. For (b), note that if V was a set, then by Axiom 5 (Comprehension), $R = \{x \in V : x \notin x\}$ would also be a set, which is a contradiction. Part (c) is left as an exercise. □

Exercises

1. Let x, y, A, B be sets.

(a) Show that $\{x\}$ is a set.

(b) Show that $(x, y) := \{\{x\}, \{x, y\}\}$ is a set and write a formula $\varphi(z)$ that holds if and only if z is an ordered pair. Moreover, write formulas $\varphi_0(z, x)$ and $\varphi_1(z, y)$ such that $\varphi_0(z, x)$ and $\varphi_1(z, y)$ hold if and only if $z = (x, y)$. In other words, φ_0 defines the function $z \mapsto x$ and φ_1 defines the function $z \mapsto y$.

(c) Show that $A \times B := \{(x, y) : x \in A \wedge y \in B\}$ is a set.

(d) Define the notion of a function $f : A \to B$ as a certain subset of $A \times B$, i.e. write down which sets are called functions from A to B.

2. Show that the powerset of a transitive set is transitive.

3. Prove part (d) of Lemma 3.3 and part (a) of Lemma 3.4. Do not use any of the later parts (b)-(c) of Lemma 3.4 in your proofs.

4. Prove that there does NOT exist a set that contains all of the ordinals.

5. Prove Lemma 3.11.

6. Prove Proposition 5.4.

7. Prove that $\mathbb{N} \equiv \mathbb{Z} \equiv \mathbb{Q}$.

8. Prove that $\mathbb{R} \equiv (0, 1) \equiv [0, 1] \equiv 2^{\mathbb{N}} \equiv \mathscr{P}(\mathbb{N})$.

9. Prove that AC implies Zorn's Lemma.

Outline: Let $(A, <)$ be as in the statement of Zorn's Lemma and assume for contradiction that no element is maximal. Then for every chain $C \subseteq A$, the set $U(C) = \{a \in A : \forall b \in C(b < a)\}$

of strict upper bounds is nonempty. Hence, denoting the set of all chains in A (including the empty set) by $\mathrm{Chains}(A)$, AC provides a function $f : \mathrm{Chains}(A) \to A$ mapping each chain C to an element in $U(C)$. Using f, obtain a contradiction by defining an injection of $\mathcal{X}(A)$ into A by transfinite induction.

10. Prove that Zorn's Lemma implies Zermelo's Theorem.

HINT: For a set A, consider $(\mathrm{WO}(A), \prec)$, where $\mathrm{WO}(A)$ is as in the proof of Theorem 9.1.

11. Prove that Zermelo's Theorem implies AC.

CAUTION: It is easy to accidentally use AC in your proof. Make sure you don't.

12. (a) Prove without using AC that for a cardinal $\kappa \geqslant \omega$, $|\kappa \times \kappa| = \kappa$.

HINT: Define a well-ordering $<_2$ of $\kappa \times \kappa$ (using the well-ordering of κ) such that the cardinality of every proper initial segment of $(\kappa \times \kappa, <_2)$ is less than κ (think of how you would do it for $\kappa = \omega$). Conclude that $\mathrm{tp}(\kappa \times \kappa, <_2) \leqslant \kappa$.

(b) (AC) Conclude that if A_α are sets of cardinality at most κ, for $\alpha < \kappa$, then $|\bigcup_{\alpha<\kappa} A_\alpha| \leqslant \kappa$. Pinpoint exactly where you use AC.

(c) (AC) Show that for any infinite cardinal κ, κ^+ is regular. In particular, ω_1 is regular.

13. Prove the following version of Hartogs' theorem that doesn't use AXIOM 8 (Replacement):

Theorem 11.2 (Hartogs)**.** *For every set A, there is a well-ordered set $(H(A), <_{H(A)})$ such that $H(A) \not\preccurlyeq A$ and $(H(A), <_{H(A)})$ is $\preceq$-least such well-ordering, i.e. if $(B, <_B)$ is another well-ordering with the property that $B \not\preccurlyeq A$, then $(H(A), <_{H(A)}) \preceq (B, <_B)$.*

OUTLINE: Let $\mathrm{WO}(A)$ be as in Theorem 9.1 and denote by $H(A)$ the quotient of $\mathrm{WO}(A)$ by the equivalence relation $\simeq$, i.e. $H(A) = \mathrm{WO}(A)/\simeq$. For $(B, <_B) \in \mathrm{WO}(A)$, let $[(B, <_B)]$ denote the $\simeq$-equivalence class of $(B, <_B)$ in $H(A)$. Define an ordering $<_{H(A)}$ on $H(A)$ as follows: for $[(B, <_B)], [(C, <_C)] \in H(A)$,

$$[(B, <_B)] <_{H(A)} [(C, <_C)] \iff (B, <_B) \prec (C, <_C).$$

Show that $<_{H(A)}$ is actually a well-ordering and verify that $(H(A), <_{H(A)})$ satisfies the conclusion of the theorem.

14. Show that $R := \{x : x \notin x\}$ is a proper class.

15. An *open interval* in ω_1 is a set of the form $(\alpha, \beta) := \{\gamma \in \omega_1 : \alpha < \gamma < \beta\}$ or $[0, \alpha) := \alpha$, for some $\alpha < \beta < \omega_1$. The topology generated by open intervals is naturally called the *open interval topology*.

(AC) Prove that the open interval topology on ω_1 is sequentially compact (i.e. every sequence has a convergent subsequence), but not compact (in the sense of open covers).

16. Let X be a second-countable topological space.

(a) Show that X has at most continuum many open subsets.

(b) Let α, β, γ denote ordinals. A sequence of sets $(A_\alpha)_{\alpha<\gamma}$ is called *monotone* if it is either increasing (i.e. $\alpha < \beta \Rightarrow A_\alpha \subseteq A_\beta$, for all $\alpha, \beta < \gamma$) or decreasing (i.e. $\alpha < \beta \Rightarrow A_\alpha \supseteq A_\beta$, for all $\alpha, \beta < \gamma$); call it *strictly monotone*, if all of the inclusions are strict.

Prove that any strictly monotone sequence $(U_\alpha)_{\alpha<\gamma}$ of open subsets of X has countable length, i.e. γ is countable.

HINT: Use the same idea as in the proof of (a).

(c) Show that every monotone sequence $(U_\alpha)_{\alpha<\omega_1}$ open subsets of X eventually stabilizes, i.e. there is $\gamma < \omega_1$ such that for all $\alpha < \omega_1$ with $\alpha \geqslant \gamma$, we have $U_\alpha = U_\gamma$.

HINT: Use the regularity of ω_1.

(d) Conclude that parts (a), (b) and (c) are also true for closed sets.

REFERENCES

[Kun83] K. Kunen, *Set Theory: An Introduction to Independence Proofs*, Studies in Logic and the Foundations of Mathematics, vol. 102, Elsevier, 1983.

[Mos06] Y. N. Moschovakis, *Notes on Set Theory*, 2nd ed., Undergraduate Texts in Mathematics, Springer, 2006.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, IL, 61801, USA
Email address: anush@illinois.edu