MATH 434 Set Theory Lecture Notes 2018

1 Introduction

1.1 Why Set Theory?

Set theory can be used as a foundation for mathematics. What this means is that we can represent natural numbers, real numbers, functions and other mathematical objects by sets. We will see some examples of this, however we will not focus too heavily on this aspect in this course. This foundational nature means that questions about the existence of certain mathematical objects can be turned into questions about the existence of sets. It can be easier to analyse these questions in the setting of set theory where all extraneous details have been removed.

Related to set existence questions, is the question of consistency. We want any mathematical axioms we use to be consistent, i.e. we would like to know that there is no contradiction that can be derived from them. By Gödel's second incompleteness theorem, no sufficiently strong system can prove its own consistency. However, we can still say interesting things about consistency. An important result along these lines is also one of Gödel's; if the axioms of Zermelo-Fraenkel set theory are consistent then so are the axioms of Zermelo-Fraenkel along with the axiom of choice.

Set theory is well-suited to first-order logic. Because everything is a set, quantifiers in first order logic range over sets as well as sets of sets (and sets of sets of sets etc.). This allows more complicated logics, like second-order logic to be expressed in terms of first-order logic using set theory. By using first-order logic as a basis, we also have access to the compactness and completeness theorems.

1.2 Russell's Paradox

Typically when defining an object in mathematics, a mathematician will just write down a description of the object completely confident that the existence of such an object is logically consistent. In much of mathematics, this is rarely a problem for there are no 'known' inconsistencies that will arrive. But consider the following argument. Let U be the set of all sets. Now let S be the following set

$$S = \{X \in U \mid X \notin X\}.$$

Ask yourself, is $S \in S$? Thinking about this for a while you will see that $S \in S$ implies $S \notin S$ and vice versa. This is known as Russell's paradox. Where is the mistake? Is it in the existence of U, or in the creation of S from U? As we will see, set theory provides a way of dealing with this.

1.3 Axiomatic Set Theory

Russell's Paradox leads us to an axiomatic approach to set theory. We will write down as axioms rules for which sets exist. We will also add an axiom on when two sets are equal. In our arguments, we can only use sets whose existence follows from the axioms. Let us start with a simple example which we will gradually develop into the axioms of Zermelo-Fraenkel set theory. We begin with the following four axioms:

1. Extensionality – Two sets are the same if and only if they have the same members.
2. Empty set – There is a set with no members.
3. Pairing – For any sets x and y, there is a set containing exactly x and y. (We denote this by $\{x, y\}$ and call this the unordered pair.)
4. Union – For any set x there is a set y which is the union of all elements of x i.e. $z \in y$ if and only if for some set w we have $z \in w \in x$ or alternatively
$$y = \bigcup_{w \in x} w.$$

As an example of the union axiom, if A is the set

$$\{\{1, 2, 3\}, \{2, 6, \{3, 4\}\}, \{5\}\}$$

then $\bigcup_{w \in A} w$ is $\{1, 2, 3, 5, 6, \{3, 4\}\}$. The notation $\bigcup A$ is also used the same set.

We can express all these axioms using first-order logic. In order to do this, we need to determine what constant, relation and function symbols we will use. This is called the language of set theory. The language of set theory is very simple. It consists of a single binary relation ϵ; there are no function symbols or constants. Hence the atomic formulas are $x \in y$ and $x = y$. Formulas are built up from these atomic formulas using $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$, $\neg$ along with the quantifiers $\forall x$ and $\exists x$ in the usual way.

Relations like $\subseteq$ are not in the language of set theory. Instead we regard $x \subseteq y$ as short-hand for the formula $\forall z(z \in x \rightarrow z \in y)$. Other short-hand forms we will use are $a \notin b$ for $\neg(a \in b)$; $(\forall x \in y)\varphi(x)$ for $(\forall x)(x \in y \rightarrow \varphi(x))$; and $(\exists x \in y)\varphi(x)$ for $(\exists x)(x \in y \wedge \varphi(x))$. Here is how we will express the above axioms in first-order logic.

1. Extensionality – $(\forall x)(\forall y)(x = y \leftrightarrow (\forall z)(z \in x \leftrightarrow z \in y))$.
2. Empty set – $(\exists x)(\forall y)(y \notin x)$.
3. Pairing – $(\forall x)(\forall y)(\exists z)(x \in z \wedge y \in z \wedge (\forall w \in z)(w = x \vee w = y))$.

4. Union – $(\forall x)(\exists y)[(z \in y \leftrightarrow (\exists w \in x)(z \in w)]$.

We can do quite a bit with just these axioms.

Lemma 1.1. *For any two sets A and B, $A \cup B$ is a set.*

Proof. Apply pairing to form the set $\{A, B\}$. Now apply union to obtain $\bigcup\{A, B\}$. Clearly $x \in \bigcup\{A, B\}$ if and only if $x \in A$ or $x \in B$. □

1.4 The Natural Numbers I

Let us show how we can represent the natural numbers using sets. We will identify each natural number with a specific set. Let $0 = \emptyset$. By pairing, taking $x = y = \emptyset$, there is a set $\{\emptyset\}$. Let 1 be this set. Note that $1 = \{0\}$. Now by pairing we also have $\{1\}$ and then by pairing again we have $\{1, \{1\}\}$. Now apply union to this set we obtain the set $\{0, 1\}$. In general we let

$$n + 1 = \bigcup\{n, \{n\}\}.$$

Observe by induction that $n + 1 = \{0, 1, 2, \ldots, n\}$. One benefit about this way to define the natural numbers is we now have that $n < m$ if and only if $n \in m$.

This gives us all the natural numbers as sets, but not necessarily the set of all natural numbers. Observe the distinction between every natural number is a set and the set of all natural numbers exists. In order to get the set of all natural numbers we need new axioms. The following axiom is one way of expressing that there exists an infinite set. This particular way is useful because the set it gives us has some nice properties.

4. Infinity – $(\exists x)(\emptyset \in x \wedge (\forall y \in x)(y \cup \{y\} \in x))$.

Note $y \cup \{y\} \in x$ can be expressed as $(\exists z \in x)(z = y \cup \{y\})$ and $z = y \cup \{y\}$ can be expressed as $y \in z \wedge y \subseteq z \wedge (\forall w \in z)(w \in y \wedge w = y)$.

Let I be the set guaranteed by the axiom of infinity. By induction, I must contain all natural numbers we informally defined. However, I could potentially be much larger. We need a means of refining it down. As of yet we have no way of doing this. We need an axiom like the following. For any set x and any property P, the collection of elements of x satisfying property P is a set.

In fact, as we are working in first-order logic, we need to add more than one axiom. We need an axiom for every possible description.

5. Axiom Schema of Separation – For every n and every formula $\varphi(x_0, x_1, \ldots, x_n)$ with $n + 1$ free variables in the language of set theory, we have the following axiom:

$$(\forall a)(\forall p_1)(\forall p_1) \ldots (\forall p_n)(\exists b)(x \in b \leftrightarrow (x \in a \wedge \varphi(a, p_1, p_2, \ldots, p_n))).$$

In words, for any set a, any sets $p_1, p_2, \ldots p_n$, and any formula φ in the language of set theory with $n+1$ free variables, there is a set b such that $x \in b$ if and only if $\varphi(x, p_1, p_2, \ldots, p_n)$ and $x \in a$.

The sets $p_1, \ldots, p_n$ are known as parameters. The use of parameters is illustrated in the following lemma whose proof is left as an exercise.

Lemma 1.2.

(i) If A and B are sets then so is $A \cap B$.

(ii) If $\mathcal{A}$ is a set so is $\bigcap_{A \in \mathcal{A}} A$.

1.5 The Natural Numbers II

We are now in a position to define the set of all natural numbers. We call a set I an *inductive set* if $\emptyset \in I$ and for all $x \in I$ we have $x \cup \{x\} \in I$. The axiom of infinity guarantees the existence of an inductive set.

Lemma 1.3. *Let $\mathcal{I}$ be a set such that every $I \in \mathcal{I}$ is inductive. Show that $\bigcap \mathcal{I}$ is also an inductive set.*

Proof. For all $I \in \mathcal{I}$, we have $\emptyset \in I$ because I is inductive, hence $\emptyset \in \bigcap \mathcal{I}$. If $x \in \bigcap \mathcal{I}$ then for any $I \in \mathcal{I}$ we have $x \in I$. As I is inductive we know $x \cup \{x\} \in I$. Thus $x \cup \{x\} \in \bigcap \mathcal{I}$ and hence $\bigcap \mathcal{I}$ is an inductive set. □

Fix an inductive set M, as guaranteed by the axiom of infinity. Now define N as follows:

$$N \coloneqq \{n \in M \colon \forall X (\text{ if } X \text{ is inductive then } n \in X)\}.$$

Note that N is the intersection of *all* inductive sets and hence by the previous lemma it is also an inductive set. As N is the intersection of all inductive sets, the definition of N is independent of M. Clearly N is the smallest inductive set as it is contained in all other inductive sets. Now we will give a definition of $\mathbb{N}$ that we will use for this course. Remember this definition and use it when you need to prove facts about the natural numbers.

Definition 1.4.

(i) The set $\mathbb{N}$ is the smallest inductive set.

(ii) An element of $\mathbb{N}$ is called a *natural number*.

Lemma 1.5. $\mathbb{N} = \{m \in \mathbb{N} \colon m = \emptyset \vee (\exists n \in \mathbb{N})(m = n \cup \{n\})\}$

Proof. Let $X = \{m \in \mathbb{N} \colon m = \emptyset \vee (\exists n \in \mathbb{N})(m = n \cup \{n\})\}$. First $\emptyset \in X$ by definition. Now if $n \in X$, then $n \in \mathbb{N}$. Hence $n \cup \{n\}$ is also in X. Thus X is an inductive set. As $\mathbb{N}$ is the smallest inductive set it follows that $X = \mathbb{N}$. □

A philosophical question is the following, 'is $\mathbb{N}$ really the set of natural numbers?' We won't dwell on this however, note that it is has the following essential property of the natural numbers, the proof of which is left as an exercise.

Lemma 1.6. *If $X \subseteq \mathbb{N}$ and $X \neq \emptyset$, then X contains a least element i.e. for some element $n \in X$, we have that for all $m \in X$, $m \not< n$.*

1.6 Functions as Sets

Given two sets a and b, we used the pairing axiom to define the unordered pair $\{a, b\}$. We can also use the pairing axiom to produce the set $\{\{a\}, \{a, b\}\}$. This set is called an ordered pair, see exercises, and it is denoted (a, b).

We can use the axioms we have so far, along with the powerset axiom to form the cross-product $A \times B$.

6. Powerset Axiom – $(\forall a)(\exists b)(\forall x)(x \in b \leftrightarrow x \subseteq a)$. (For any set a, there is a set b containing all the subsets of a.)

We will denote the powerset of a by $\mathcal{P}(a)$.

Lemma 1.7. *There is a formula $\varphi(a, b, c)$ in the language of set theory such that $\varphi(a, b, c)$ holds if and only if $c = (a, b)$.*

Proof. As $(a, b) = \{\{a\}, \{a, b\}\}$ we can take

$$\exists i \exists j (i \in c \wedge j \in c \wedge a \in i \wedge a \in j \wedge b \in j) \\ \wedge \forall i (i \in c \to (i = \{a, b\} \vee i = \{a\}))$$

Showing that we can replace $i = \{a, b\}$ and $i = \{a\}$ by a formula in the language of set theory is an exercise. □

Lemma 1.8. *For all sets A, B the cross-product is a set.*

Proof. We have established that $A \cup B$ is a set. Now apply the power set axiom twice to obtain $E = \mathcal{P}(\mathcal{P}(A \cup B))$. Observe that for any $a \in A$ and $b \in B$, $\{a\}, \{a, b\} \in \mathcal{P}(A \cup B)$. Hence $\{\{a\}, \{a, b\}\} \in E$. Now, using φ from the previous lemma, apply separation

$$C = \{c \in E \mid (\exists a \in A)(\exists b \in B)\varphi(a, b, c)\}.$$ □

From cross-products it is possible to use separation to produce functions. A partial function f from A to B is regarded as a subset of $A \times B$ such that for all x, y, z if $(x, y), (x, z) \in f$ then $y = z$. The range and domains of a function are sets (see exercises) and so in this case, f would be total if its domain is equal to A.

Lemma 1.9. *The successor function $s : \mathbb{N} \to \mathbb{N}$ defined by $s(n) = n + 1$ is a set.*

Proof. We want s to be the set $\{(m, n) \in \mathbb{N} \times \mathbb{N} \mid n = m \cup \{m\}\}$. This can be done using separation as follows

$$\{x \in \mathbb{N} \times \mathbb{N} \colon (\exists m \in \mathbb{N})(\exists n \in \mathbb{N})(\varphi(m, n, x) \wedge m \in n \wedge m \subseteq n \wedge (\forall z \in n)(z \in m \vee z = m)\}. \quad \square$$

For the following proof we will make use of Lemma 1.6.

Theorem 1.10. *The function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by $f(n, m) = n + m$ is a set.*

Proof. Recall that we can define addition recursively as follows. For all $n \in \mathbb{N}$, we define $n+0 = n$. For all n and all m we define $n+s(m) = s(n+m)$. We can adapt this definition to the context of set theory. We will prove a significant generalisation of this method later on.

We start by defining A to be the set of all partial functions from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, with the property that they agree with the above recursive definition where defined. Think of A as the set of approximations to addition. Let

$$\begin{aligned} A = \{g \in \mathcal{P}((\mathbb{N} \times \mathbb{N}) \times \mathbb{N}) \colon & (\forall n)((n, 0) \in \mathrm{dom}(g) \to g(n, 0) = n) \\ & \wedge (\forall n)(\forall m)((n, s(m)) \in \mathrm{dom}(g) \to \\ & \quad ((n, m) \in \mathrm{dom}(g) \wedge g(n, s(m)) = s(g(n, m))))\}. \end{aligned}$$

An example of an element in A is the following set (whose existence follows easily from pairing and union)

$$\{((0, 0), 0), ((0, 1), 1), ((2, 0), 2), ((2, 1), 3), ((2, 2), 4)\}.$$

Now define $f = \bigcup A$.

Claim 1.11. The set f is a partial function.

Proof. Assume not. Then for some n, and some least m for this n, there is a $p \neq r$ such that $((n, m), p), ((n, m), r) \in f$. By definition of f, there are g_1 and g_2 in A with $((n, m), p) \in g_1$ and $((n, m), r) \in g_2$. Clearly $m \neq 0$ as this would imply that $r = p = n$ because it must be that $((n, 0), n) \in g_1 \cap g_2$. Now if $m \neq 0$, then by Lemma 1.5 there is some $l \in \mathbb{N}$ such that $m = l \cup \{l\}$. But then as m is least and clearly $l \in m$ and so $l < m$. Hence we have for some $q \in \mathbb{N}$ that$((n, l), q) \in g_1 \cap g_2$. This gives us the contradiction that $p = s(q) = r$. $\square$

Claim 1.12. The domain of f is $\mathbb{N} \times \mathbb{N}$.

Proof. Fix n. Assume for some least m, for this n, we have (n, m) not in the domain of f. First the set $\{((n, 0), n)\}$ exists by repeated applications of pairing. This is an element of A and so $f(n, 0) = n$. Hence $m \neq 0$. Thus for some $l \in \mathbb{N}$ we have $m = l \cup \{l\}$. Now (n, l) is in the domain of f and so for some $g \in A$ and some p, we have $((n, l), p) \in g$. But $g \cup \{(n, m), s(p))\} \in A$ and so (n, m) is in the domain of f. □

Claim 1.13. f agrees with the recursive definition of addition.

Proof. This follows because if $f(n, m) = p$ then $g(n, m) = p$ for some $g \in A$ and g agrees with the recursive definition. □

□

1.7 Axioms of ZFC

Here is a list of the standard axioms of ZFC. The existence of the empty set is not a standard axiom (this follows immediately from the axiom of infinity and the axiom of separation) so we omit it. We include the axiom schema of replacement, the axiom of foundation and the axiom of choice for completeness. We will discuss these in detail later.

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

$$(\forall x)(\forall y)[(\forall z)(z \in x \leftrightarrow z \in y) \rightarrow x = y]$$

Axiom of Pairing.

$$(\forall x)(\forall y)(\exists z)[x \in z \wedge y \in z \wedge (\forall w \in z)(w = x \vee w = y)]$$

Denote by $\{x, y\}$.

Axiom of Infinity. There is an inductive set.

$$(\exists x)[\emptyset \in x \wedge (\forall y \in x)(\{y\} \cup y \in x)]$$

Axiom of Union. Given a set x, there is a set y that is the union of all elements of x.

$$(\forall x)(\exists y)(\forall z)[z \in y \leftrightarrow (\exists w \in x)(z \in w)]$$

Denote by $\bigcup x$.

Axiom of Powerset. Given any set x, there is a set containing all subsets of x.

$$(\forall x)(\exists y)(\forall z)[z \in y \leftrightarrow z \subseteq x]$$

Denote by $\mathcal{P}(x)$.

Axiom Schema of Separation. For any first order formula $\varphi(x_0, x_1, \ldots x_n)$ in the language of set theory we have the following axiom.

$$(\forall x)(\forall p_1)\ldots(\forall p_n)(\exists y)(\forall z)[z \in y \leftrightarrow (z \in x \wedge \varphi(z, p_1, \ldots, p_n))]$$

Note that if we accept separation, then Russell's paradox means that there can be no universal set.

Axiom Schema of Replacement. If a first-order formula $\varphi(x, y, p_1, \ldots p_n)$ in the language of set theory defines a function with parameters $p_1, \ldots, p_n$ (i.e. for all x there is a unique y such that $\varphi(x, y, p_1, \ldots, p_n)$ holds) then the range of any set under this function is a set.

$$(\forall p_1)\ldots(\forall p_n)[((\forall x)(\forall y)(\forall z)(\varphi(x, y, p_1, \ldots p_n) \wedge \varphi(x, z, p_1, \ldots p_n) \rightarrow y = z)) \rightarrow ((\forall x)(\exists y)(\forall z)(z \in y \leftrightarrow \exists(w \in x)\varphi(w, z, p_1, \ldots p_n)))]$$

Note that the function itself is not necessarily a set.

Axiom of Foundation Any non-empty set has an ϵ-minimal element.

$$(\forall x)[(\exists y)(y \in x) \rightarrow (\exists z)(z \in x \wedge (\forall w \in x)(z = w \vee w \notin z))]$$

Axiom of Choice. For any set x, there is a choice function on the non-empty subsets of x.

$$(\forall x)(\exists f)(\forall z)[(z \subseteq x \wedge z \neq \emptyset) \rightarrow (f(z) \in z)]$$

2 Well-orders and Ordinals

What is the most important property of $\mathbb{N}$? Arguably it is that $\mathbb{N}$ has the least number property. This is the basis of inductive arguments. We can generalise inductive arguments to another mathematical object known as a well-order.

Definition 2.1. Let R be a set and let $<_R$ be a binary relation on R. We call $<_R$ a *linear order* if for all $x, y, z \in R$:

(i) $x \not<_R x$.

(ii) $x <_R y$ or $y <_R x$ or $x = y$.

(iii) $x <_R y$ and $y <_R z$ implies $x <_R z$.

Definition 2.2. A linear order $<_R$ on R is a *well-order* if for all non-empty $E \subseteq R$, there exists $x \in E$ such that for all $y \in E$, if $y \neq x$ then $x <_R y$.

Any finite linear order is a well-order. An example of an infinite well-order is $(\mathbb{N}, <)$. This well-order is denoted by ω. There are many more well-orders e.g. consider the set $\mathbb{N} \cup \{\star\}$ with the order $\prec$ defined by for all $n \in \mathbb{N}$, $n \prec \star$ and if $m, n \in \mathbb{N}$ then $m \prec n$ if and only if $m < n$. This well-order is denoted by $\omega + 1$. Later on we will see many more well-orders.

Examples of linear orders that are not well-orders include $\mathbb{Z}$, $\mathbb{R}$ and ω^*. The order ω^* is ω backwards, i.e. the order $<$ on the negative integers. An other example is $[0, 1] \cap \mathbb{Q}$ with the standard $<$ order. Note that this linear order has 0 as its least element. However, $(0, 1] \cap \mathbb{Q}$ is an example of a subset of $[0, 1] \cap \mathbb{Q}$ that does not have a least element.

Two well-orders are isomorphic if there exists an order-preserving bijection between them. Given a linear order $(R, <)$ and $a \in R$, denote by $R\restriction_a = \{x \in R \mid x < a\}$.

Question 1. Why can any strict initial segment of a well-order $(R, <)$ be written as $R\restriction_a$ for some $a \in R$? (This is not true for linear orders.)

Lemma 2.3. *Let $(R, <)$ be a well-order. Then $(R, <)$ is not isomorphic to any strict initial segment of itself.*

Proof. Let e be an element of R. Assume $f : R \to R\restriction_e$ is an isomorphism. Let a be least such that $f(a) \neq a$ (such an a exists because $f(e) \neq e$). If $f(a) < a$, then because f is an isomorphism we have that $f(f(a)) < f(a)$ contradicting the minimality of a. If $f(a) > a$ then a must be in the range of f. Hence there is some $b \neq a$ such that $f(b) = a < f(a)$ and so $b < a$, which again contradicts the minimality of a. □

Theorem 2.4. *Given any two well-orders, either they are isomorphic or one is isomorphic to a strict initial segment of the other.*

Proof. Let $(R, <)$ and $(S, \prec)$ be two well-orders. Define a partial function $f : R \to S$ by $f(a) = b$ if $(R{\restriction_a}, <)$ is isomorphic to $(S{\restriction_b}, \prec)$. This definition does define a partial function because by lemma, if $b, c \in S$ with $b \neq c$, then $(R{\restriction_a}, <)$ cannot be isomorphic to both $(S{\restriction_b}, \prec)$ and $(S{\restriction_c}, \prec)$. Further the domain of f is an initial segment of R, and the range is an initial segment of S.

I claim that f is an isomorphism between its domain and range. This holds because if $a_0 < a_1$ are in the domain of f, with $f(a_1) = b$, then $a_0 \in R{\restriction_{a_1}}$, and $(R{\restriction_{a_1}}, <)$ is isomorphic to $(S{\restriction_b}, \prec)$. If we restrict this isomorphism to $R{\restriction_{a_0}}$ then this establishes that a_0 maps under f to some $c < b$.

Now for the inductive step. If there is some $a \in R \setminus \mathrm{dom}(f)$ and $b \in S \setminus \mathrm{rng}(f)$, then take a and b least with this property for $<$ and $\prec$ respectively. Now f is an isomorphism between $R{\restriction_a}$ and $S{\restriction_b}$. But, by the definition of f, this would imply that $f(a) = b$, a contradiction.

If every $a \in R$ is in the domain of f, and every $b \in S$ is in the range of f, then f establishes that R and S are isomorphic. If R is the domain of f, but for some least $b \in S$, b is not in the range of the f. Then f defines an isomorphism between R and $S{\restriction_b}$. Similarly if S is the range of f. □

2.1 Ordinals

There is a special type of well-orders called ordinals.

Definition 2.5. A set x is called *transitive* if for all $y \in x$, $y \subseteq x$.

Equivalently x is a transitive if and only if for all z and y, $z \in y \in x$ implies $z \in x$.

Definition 2.6. A set is an *ordinal* if it is transitive and well-ordered by ϵ.

Note if α is an ordinal and $x \in \alpha$, then by the definition of a linear order we know that $x \notin x$ (this also follows from the axiom of foundation).

Definition 2.7.

(i) Given an ordinal α we define the *successor* of α to be $S(\alpha) = \alpha \cup \{\alpha\}$. This is also denoted $\alpha + 1$.

(ii) An ordinal α is called a *successor ordinal* if for some ordinal γ, $\alpha = S(\gamma)$.

(iii) An ordinal α is called a *limit ordinal* if $\alpha \neq \emptyset$ and α is not a successor ordinal.

Lemma 2.8.

(i) $\emptyset$ is an ordinal.

(ii) If α is an ordinal then $S(\alpha)$ is an ordinal.

(iii) If α is an ordinal and $\beta \in \alpha$, then β is an ordinal.

(iv) If α, β are ordinals and $\alpha \subsetneq \beta$ then $\alpha \in \beta$.

(v) If α, β are ordinals then $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$.

Proof. (i). Trivial.
(ii). If $x \in S(\alpha)$ then $x \in \alpha$ or $x = \alpha$. If $x \in \alpha$ then $x \subseteq \alpha$ as α is an ordinal and so $x \subseteq S(\alpha)$. If $x = \alpha$ then $x \subseteq S(\alpha)$ by definition so $S(\alpha)$ is transitive.

You should check that ϵ linearly orders $S(\alpha)$ with α as the maximal element. Let $E \subseteq S(\alpha)$ be non-empty. If $E = \{\alpha\}$ then E has a least element. Otherwise $E \setminus \{\alpha\} \subseteq \alpha$ is not empty and so has a least element β. If $\alpha \in E$ then as $\beta \in \alpha$ we also have that β is the least element of E.
(iii). By the transitivity of α, $\beta \subseteq \alpha$ and so β is well-ordered by ϵ. Now take any $x \in y \in \beta$. Again by the transitivity of α, $x, y \in \alpha$. Hence either $x \in \beta$ or $\beta \in x$ (because ϵ linearly orders the elements of α). If $\beta \in x$, then α is not an ordinal because $x \in y \in \beta \in x$ shows that α is not linearly ordered by ϵ. Hence $x \in \beta$.
(iv). Let δ be least such that $\delta \in \beta \setminus \alpha$. Take any $x \in \alpha$. As β is an ordinal, and x and δ are both in β, we know that either $\delta \in x$ or $x \in \delta$. If $\delta \in x$, then by transitivity of α we have $\delta \in \alpha$. But this is impossible as we chose $\delta \in \beta \setminus \alpha$. Hence $x \in \delta$ and so we can conclude that $\alpha \subseteq \delta$. Now if $\alpha \neq \delta$, then there is some $\gamma \in \delta \setminus \alpha$, but then $\gamma \in \beta \setminus \alpha$ and $\gamma \in \delta$ contradicting the minimality of δ.
(v). First $\delta = \alpha \cap \beta$ is an ordinal. It is well-ordered by ϵ because it is a subset of an ordinal, and it is transitive because α and β are transitive. Assume $\delta \neq \alpha$ and $\delta \neq \beta$. Then $\delta \subsetneq \alpha$ and so by (iii) $\delta \in \alpha$. Similarly $\delta \in \beta$. Hence $\delta \in \alpha \cap \beta = \delta$, contradicting the fact that α is linearly ordered by ϵ. □

The proof of the following lemma is left as an exercise.

Lemma 2.9.

(i) The union of a set of ordinals is an ordinal.

(ii) $\mathbb{N}$ is an ordinal.

When considering $\mathbb{N}$ as an ordinal we denote it by ω. The following lemmas give another characterisation of ω.

Lemma 2.10. *If α is a successor ordinal such that every element of α is either the empty set or a successor ordinal, then $\alpha \in \omega$.*

Proof. For this proof we want to find a least counter-example and then derive a contradiction. However, we need to do this inside some well-ordered set.

Say an ordinal α has property $(\star)$ if firstly $\alpha \notin \omega$, secondly α is a successor ordinal and thirdly every element of α is either the empty set, or a successor ordinal. Assume the lemma does not hold. Then there is some ordinal α with property $(\star)$.

Consider the set $S(\alpha)$. This is the well-ordered set we will work with. Let β be the ϵ-least element of $S(\alpha)$ with property $(\star)$. Now β exists as $\alpha \in S(\alpha)$ has property $(\star)$. As $\beta \notin \omega$, we know that $\beta \neq \emptyset$. Further either $\beta = \alpha$ or $\beta \in \alpha$. This implies that β is a successor ordinal and so $\beta = \gamma \cup \{\gamma\}$ for some ordinal γ. But $\gamma \notin \omega$ because this would imply $\beta \in \omega$. Now if $x \in \gamma$ then by the transitivity of β we have know $x \in \beta$. Hence either x is the empty set or x is a successor ordinal. Finally $\gamma \in \beta$ so γ is a successor ordinal Thus γ has property $(\star)$ contradicting the minimality of β. □

The following is a generalisation of Lemma 1.5.

Lemma 2.11. *If $n \in \omega$ and $n \neq \emptyset$, then n is a a successor ordinal such that every element of n is either the empty set or a successor ordinal.*

Proof. Again the set $\{n \in \omega \colon n = \emptyset$ or n is a successor ordinal such that every element of n is either the empty set or a successor ordinal$\}$ is an inductive set. □

In the following lemma we will make use of the axiom schema of replacement for the first time. The axiom of replacement works as follows. Assume that we have a formula in the language of set theory with two free variables $\varphi(x, y)$. Now assume further that for all x, y, z if $\varphi(x, y)$ and $\varphi(x, z)$ both hold, then $y = z$. In this case we can think of φ as defining a function (though this function may not be a set in the sense of a set of ordered pairs). The axiom of replacement says that the image of a set under this function is also a set. For example, there is a a formula $\varphi(x, y)$ that holds if $y = \mathcal{P}(x)$. So if w is a set, then by the axiom of replacement the set $\{\mathcal{P}(x) \colon x \in w\}$ is a set (note that this could be established without replacement). We will examine these functions defined by formulas further in Section 3.

Lemma 2.12. *Every well-ordered set is isomorphic to a unique ordinal.*

Proof. Let $(W, <)$ be a well-ordering. We can define a formula $\varphi(x, \alpha)$ which holds if $x \in W$, α is an ordinal, and $(W\restriction_x, <) \cong (\alpha, \epsilon)$. (This last part can be achieve by saying there exists a function f, $\mathrm{dom}(f) = W\restriction_x$, the range of f is α, and if $a, b \in W\restriction_x$, then $a < b$ if and only if $f(a) \in f(b)$.)

Let $\widehat{W} = \{x \in W \mid (\exists \alpha)\varphi(x, \alpha)\}$. Now $\widehat{W}$ exists by separation and $\widehat{W}$ forms an initial segment of W. Hence either $\widehat{W} = W$ or $\widehat{W} = W\restriction_b$ for the least $b \in W \setminus \widehat{W}$. Apply replacement to obtain a set X such that for all $x \in \widehat{W}$, the unique α such that $\varphi(x, \alpha)$ is in X. Let

$$g = \{(x, \alpha) \in \widehat{W} \times X \mid \varphi(x, \alpha)\}.$$

Let β be the range of g. It is left as an exercise to show that β is an ordinal and $g : \widehat{W} \cong \beta$. Now if $\widehat{W} = W\restriction_b$ for some $b \in W$, then we have just constructed an isomorphism between $W\restriction_b$ and an ordinal β. This contradicts the fact that $b \notin \widehat{W}$. Hence $\widehat{W} = W$ and $(W, <)$ is isomorphic to (β, ϵ). □

Now, as we will see, the axiom of choice implies that every set can be well-ordered. Hence this lemma, in conjunction with the axiom of choice, gives us *lots* of ordinals. For example because the reals can be well-ordered, there is an uncountable ordinal.

Given a well-ordering $(W, <)$, we denote the unique ordinal α such that $\alpha \cong (W, <)$ by $ot(W, <)$. (The notation ot is short for order-type.)

2.2 Operations on ordinals

Definition 2.13. Given ordinals α, β, we define:

(i) $\alpha + \beta = ot((\{0\} \times \alpha) \cup (\{1\} \times \beta), <_{lex})$ where $<_{lex}$ is the lexicographical order i.e. $(i, \gamma) <_{lex} (j, \delta)$ if $i < j$ or, $i = j$ and $\gamma \in \delta$.

(ii) $\alpha \cdot \beta = ot(\alpha \times \beta, <_{rlex})$ where $(\gamma, \delta) <_{rlex} (\iota, \pi)$, if $\delta \in \pi$ or, $\delta = \pi$ and $\gamma \in \iota$.

Lemma 2.14. *If α, β are ordinals, then $\alpha + \beta$ is also an ordinal.*

Proof. Let E be a non-empty subset of $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$. Let $E_0 = \{(i, \xi) \in E \mid i = 0\}$. If E_0 is not empty, then it contains a least element because α is well-ordered. A least element of E_0 is a least element of E. If E_0 is empty, then E is a subset of $\{1\} \times \beta$. The fact that β is well-ordered implies that E has a least element. □

Comments:

1. Observe that the operation + defined on the ordinals is not commutative because $ot(1 + \omega) = \omega \neq S(\omega) = ot(\omega + 1)$.

2. As $S(\alpha) = \alpha + 1$, this is often denoted by $\alpha + 1$.

3. Let α, β, and γ be ordinals. Consider the set $(\{0\} \times \alpha) \cup (\{1\} \times \beta) \cup (\{2\} \times \gamma)$ ordered lexicographically. It is not difficult to see that this has order type of both $(\alpha+\beta)+\gamma$ and $\alpha+(\beta+\gamma)$. Hence this operation is associative on the ordinals.

4. An alternative way of thinking about ordinal multiplication is the following: $\alpha \cdot \beta$ is the order-type obtained by replacing each element of β with an order of type α.

2.3 Limits of ordinal sequences

Let $\langle \delta_i \mid i \in \omega \rangle$ be a sequence of ordinals. Assume that for all $i \in \omega$, $\delta_i < \delta_{i+1}$. We say $\lim_i \delta_i = \gamma$ if the sequence never exceeds γ and for all $\zeta < \gamma$ there is some i such that $\delta_i > \zeta$. In fact this definition can be adapted to cope with sequences of any ordinal length, that are not necessarily increasing.

Definition 2.15. Let α be an ordinal and let $\langle \delta_\beta \mid \beta < \alpha \rangle$ be a sequence of ordinals of length α. We say that

$$\gamma = \lim_{\beta<\alpha} \delta_\beta$$

if

$$(\forall \zeta < \gamma)(\exists \beta < \alpha)(\forall \xi)[(\beta \leq \xi < \alpha) \to (\zeta < \delta_\xi \leq \gamma)]$$

Lemma 2.16. *For α an ordinal and λ a limit ordinal,*

$$\alpha + \lambda = \lim_{\beta<\lambda} \alpha + \beta.$$

Proof. (Sketch) If $\beta < \lambda$ then $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$ under the lexicographical order is an initial segment of $(\{0\} \times \alpha) \cup (\{1\} \times \lambda)$ via the inclusion mapping. Hence $\lim_{\beta<\lambda} \alpha + \beta \leq \alpha + \lambda$.

Fix some $\gamma < \alpha + \lambda$. Now if $\gamma < \alpha$, we have that $\alpha + \delta > \gamma$ for all $\delta \in \lambda$. If $\gamma > \alpha$, then γ is equal to $\alpha + \beta$ for some unique β (exercise). Further $\beta < \gamma$ and so for all δ such that $\beta \leq \delta < \lambda$ we have that $\alpha + \delta > \gamma$. □

An alternative way to define ordinal addition is by transfinite recursion. This definition extends the definition for the natural numbers. For an ordinal α, define $\alpha + 1$ to be the successor of α.

Definition 2.17. Fix α. Define

(i) $\alpha + 0 = \alpha$.

(ii) $\alpha + (\beta + 1) = (\alpha + \beta) + 1$.

(iii) $\alpha + \lambda = \lim_{\delta<\lambda} \alpha + \delta$, for λ a limit ordinal.

However, does this give us a definition of $\alpha + \beta$ for all ordinals β? We will need to do some work to prove this. There is a further issue with Definition 2.17. Let **ORD** be all the ordinals. Consider the definition for addition. This definition implicitly defines a mapping from **ORD** $\times$ **ORD** to **ORD** by $(\alpha, \beta) \mapsto \alpha + \beta$. However, we have not shown that this function exists as a set using the axioms of ZFC. This is serious point. If we want to prove, using the axioms of ZFC, that ordinal addition is associative, we need some way to refer to it. Hence we would like to show that this function exists as a set. The problem is, **it doesn't!**

We will resolve these problems in the following section. In a similar fashion we can define ordinal multiplication and ordinal exponentiation. For ordinal multiplication, we fix α and define:

Definition 2.18.

(i) $\alpha \cdot 0 = 0$.

(ii) $\alpha \cdot (\beta + 1) = (\alpha \cdot \beta) + \alpha$.

(iii) $\alpha \cdot \lambda = \lim_{\delta < \lambda} \alpha \cdot \delta$, for λ a limit ordinal.

For ordinal exponentiation, we fix α and define:

Definition 2.19.

(i) $\alpha^0 = 1$.

(ii) $\alpha^{(\beta+1)} = (\alpha^\beta) \cdot \alpha$.

(iii) $\alpha^\lambda = \lim_{\delta < \lambda} \alpha^\delta$, for λ a limit ordinal.

3 Classes and Transfinite Recursion

Inside our universe of sets we can talk about, or identify, a collection of sets e.g. all ordinals. But being able to identify a collection of sets informally does not make this collection a set e.g. the collection of all sets. Often we can describe a collection of sets as being those sets for which a certain formula in the language of set theory holds e.g. $\{x \mid x = x\}$ or $\{x \mid x \text{ is an ordinal}\}$ (note that despite the use of curly brackets neither of these are in fact sets).

We call a collection of sets defined by such a formula a class. Any formula in the language of set theory with one free variable defines a class (we could also use parameters to define classes). A class is proper if it is not a set.

Theorem 3.1. *The collection of all ordinals is a proper class.*

Proof. It is a class because the statement 'α is transitive and wellordered by ϵ' can be expressed as a first-order formula. Now if the collection of all ordinals is a set x, then consider $\bigcup x$. By the exercises $\bigcup x$ is an ordinal and hence $\bigcup x \in x$. Now if $z \in \bigcup x$, then z is an ordinal and so $z \in x$. Further if $z \in x$ then z is an ordinal and so $z \cup \{z\} \in x$. This means that $z \in \bigcup x$. Hence $\bigcup x = x$. But $\bigcup x \in \bigcup x$ contradicts the fact that $\bigcup x$ is an ordinal. □

We will denote classes using boldface font. The class of all sets is denoted $\mathbf{V}$. The class of all ordinals is denoted $\mathbf{ORD}$.

The key thing to remember about proper classes is that because they are not sets, we cannot build sets from them by applying axioms like separation, power set or choice. There is also the fact that we cannot prove theorems about all classes in ZFC because each class is defined by a formula and we cannot simply quantify over formulas. Theorems about classes are often called theorem schemas.

Corollary 3.2. *Ordinal addition is not a set function.*

Proof. The domain of a set function is a set and the ordinals form a proper class. □

Where does this leave us with respect to Definitions 2.17, 2.18, and 2.19? Fortunately, not all is lost. While we cannot have a set function for these operations, we can obtain a class function. A class function $\mathbf{F}$ is a class such that the only members of $\mathbf{F}$ are ordered pairs and for all x, y, z if (x, y) and (x, z) are members of $\mathbf{F}$ then $y = z$. In other words, a class function is a formula with two free variables such that for all x there is at most one y such that $\varphi(x, y)$ holds. For example the formula $x = y$ defines a class function, the identity function. The formula $(\forall z)(z \in y \leftrightarrow z \subseteq x)$ defines the mapping from a set to its power set. The mapping taking an ordinal to its successor is another example.

Class functions are useful because we can still prove theorems about them. For example, assume that $\varphi(x, y, z)$ defines ordinal addition i.e. if x, y are ordinals then $\varphi(x, y, z)$ holds if and only if $x + y = z$ (how this formula behaves if x or y is not an ordinal is unimportant). Now let $\psi(x)$ hold if and only if x is an ordinal. Now the following is a theorem of ZFC.

Theorem 3.3. *If α, β, γ are ordinals i.e. $\psi(\alpha)$, $\psi(\beta)$, and $\psi(\gamma)$ all hold. Then for any ordinal δ, $(\alpha+\beta)+\gamma = \delta$ if and only if $\alpha+(\beta+\gamma) = \delta$. Note that this last statement can be expressed in first-order logic as*

$$[\exists\xi(\varphi(\alpha, \beta, \xi) \wedge \varphi(\xi, \gamma, \delta))] \leftrightarrow [\exists\xi(\varphi(\beta, \gamma, \xi) \wedge \varphi(\alpha, \xi, \delta))].$$

Or more succinctly, ordinal addition is associative.

One important way to build class functions is by transfinite recursion. Before proving this theorem we will establish an essential lemma.

Lemma 3.4. *Any non-empty class of ordinals has a least element.*

Proof. Let $\mathbf{C}$ be a class of ordinals x such that $\varphi(x, p)$ holds. If $\mathbf{C}$ is not empty, let α be a member of $\mathbf{C}$. Consider the set $\{x \in \alpha \mid \varphi(x, p)\}$. Note that this is the set of members of $\mathbf{C}$ that are strictly less than α. If this set is empty, then α is the least element of $\mathbf{C}$. Otherwise because α is an ordinal, this subset of α has a least element β. Now β be must a least element of $\mathbf{C}$ because if $\gamma \in \beta$ then by transitivity $\gamma \in \alpha$ and $\varphi(\gamma, p)$ cannot hold. □

Theorem 3.5. *Let $\mathbf{F} : \mathbf{V} \to \mathbf{V}$ be a class function. There is a unique class function $\mathbf{G} : \mathbf{ORD} \to \mathbf{V}$ such that for all ordinals α,*

(i) $\mathbf{G}{\restriction_\alpha}$ *is a set.*

(ii) $\mathbf{G}(\alpha) = \mathbf{F}(\mathbf{G}{\restriction_\alpha})$.

Proof. Let δ be an ordinal. Call a function h (where h is a set) a δ-approximation if $\delta = \mathrm{dom}(h)$ and for all $\alpha \in \delta$ we have that $h(\alpha) = \mathbf{F}(h{\restriction_\alpha})$.

Claim 3.6. If h_0 and h_1 are a δ_0-approximation and a δ_1-approximation respectively, then $h_0{\restriction_{\delta_0\cap\delta_1}} = h_1{\restriction_{\delta_0\cap\delta_1}}$.

Proof. Fix any ordinal $\alpha \in \delta_0 \cap \delta_1$ and assume that $h_0{\restriction_\alpha} = h_1{\restriction_\alpha}$. Then

$$h_0(\alpha) = \mathbf{F}(h_0{\restriction_\alpha}) = \mathbf{F}(h_1{\restriction_\alpha}) = h_1(\alpha).$$

Hence there can be no least ordinal in $\delta_0 \cap \delta_1$ where h_0 and h_1 differ. □

Claim 3.7. For every $\delta \in \mathbf{ORD}$, there is a δ-approximation.

Proof. Those ordinals δ for which there is no δ-approximation form a class. If it is not empty, it must have a least element. If $\delta = 0$, the empty function is a 0-approximation. If h is a β-approximation and $\delta = \beta + 1$, then a δ-approximation is $\widehat{h} = h \cup \{(\beta, \mathbf{F}(h))\}$. Finally, if δ is a limit, then use replacement to obtain a set of β-approximations for all $\beta < \delta$ and then take their union. This is a δ-approximation. Hence there cannot be a least element without a δ-approximation. □

The functions that are δ-approximations for some δ form a class. Now let $\mathbf{G}$ be a class of ordered pairs such that (x, y) is in $\mathbf{G}$ if there exists an ordinal δ and a δ-approximation h, such that $h(x) = y$. By the phrase '(x, y) is in $\mathbf{G}$', I mean that the formula defining $\mathbf{G}$ holds for (x, y).

Note that $\mathbf{G}$ is a class because it can be defined for all ordinals by a formula in the language of set theory. Now by the two claims we have that $\mathbf{G}$ is a class function i.e. for every ordinal δ, there is a set y with (δ, y) in $\mathbf{G}$; and if (x, y) and (x, z) are both in $\mathbf{G}$ then $y = z$. Thus by the axiom of replacement, for any ordinal α the image of α under $\mathbf{G}$ is a set. It then follows that $\mathbf{G}{\restriction_\alpha}$ is a set.

Finally, assume for all ordinals $\beta < \alpha$ we have that $\mathbf{G}(\beta) = \mathbf{F}(\mathbf{G}{\restriction_\beta})$. Let h be an $\alpha + 1$ approximation. Now

$$\mathbf{G}(\alpha) = h(\alpha) = \mathbf{F}(h{\restriction_\alpha}) = \mathbf{F}(\mathbf{G}{\restriction_\alpha}).$$ □

Now let us fix α and consider $\mathbf{H} : \mathbf{ORD} \to \mathbf{ORD}$ defined recursively in Definition 2.17 i.e. $\mathbf{H}(\gamma) = \alpha + \beta$. We will show that $\mathbf{H}$ is a class function by applying Theorem 3.5. When using Theorem 3.5, we need to begin with a class function $\mathbf{F} : \mathbf{V} \to \mathbf{V}$. Theorem 3.5 gives a class function $\mathbf{G} : \mathbf{ORD} \to \mathbf{V}$. We want to choose $\mathbf{F}$ so that $\mathbf{G}$ produced by the theorem is equal to $\mathbf{H}$.

How do we define $\mathbf{F}$? We know that $\mathbf{H}(0) = \alpha + 0 = \alpha$ and

$$\mathbf{G}(0) = \mathbf{F}(\mathbf{G}{\restriction_0}) = \mathbf{F}(\emptyset).$$

The last equality holds because there is nothing in the domain of $\mathbf{G}$ less than 0 so $\mathbf{G}{\restriction_0}$ is the empty function. Thus if we define $\mathbf{F}(\emptyset) = \alpha$, then $\mathbf{H}(0) = \mathbf{G}(0)$.

Now $\mathbf{H}(1) = \alpha + 1$, and $\mathbf{G}(1) = \mathbf{F}(\mathbf{G}{\restriction_1}) = \mathbf{F}(\{(0, \alpha)\})$. This means we need to set $\mathbf{F}(\{(0, \alpha)\}) = \alpha + 1$. Iterating this idea further we see we need to have

(i) $\mathbf{F}(\{(0, \alpha), (1, \alpha + 1)\}) = \alpha + 2$.

(ii) $\mathbf{F}(\{(0, \alpha), (1, \alpha + 1), (2, \alpha + 2)\}) = \alpha + 3$.

This suggests the following definition of $\mathbf{F}$.

$$\mathbf{F}(x) = \begin{cases} \emptyset & \text{if } x \text{ is not a mapping} \\ & \text{from ordinals to ordinals,} \\ \alpha & \text{if } x \text{ is the empty function,} \\ \sup\{\gamma + 1 \mid \gamma \in \operatorname{range}(x)\} & \text{otherwise.} \end{cases}$$

If we apply Theorem 3.5 to this $\mathbf{F}$, we obtain a class function $\mathbf{G}$.

Lemma 3.8. $\mathbf{G}$ *is a strictly increasing function.*

Proof. Let $\beta > \alpha$. Then

$$\mathbf{G}(\beta) = \mathbf{F}(\mathbf{G}{\restriction}_\beta) = \sup\{\gamma + 1 \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_\beta)\} > \mathbf{G}{\restriction}_\beta(\alpha) = \mathbf{G}(\alpha). \quad \square$$

Lemma 3.9. $\mathbf{G}(\beta + 1) = \mathbf{G}(\beta) + 1$.

Proof. Note that the largest ordinal in the domain of $\mathbf{G}{\restriction}_{\beta+1}$ is β. By the previous lemma,

$$\sup\{\gamma + 1 \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_{\beta+1})\} = \mathbf{G}{\restriction}_{\beta+1}(\beta) + 1 = \mathbf{G}(\beta) + 1.$$

Hence

$$\mathbf{G}(\beta + 1) = \mathbf{F}(\mathbf{G}{\restriction}_{\beta+1}) = \sup\{\gamma + 1 \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_{\beta+1})\} = \mathbf{G}(\beta) + 1. \quad \square$$

Lemma 3.10. *If* λ *is a limit ordinal, then* $\mathbf{G}(\lambda) = \lim_{\delta<\lambda} \alpha + \delta$.

Proof. The key to this proof is noticing that

$$\sup\{\gamma + 1 \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_\lambda)\} = \sup\{\gamma \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_\lambda)\}$$

This holds because λ has no maximal element and $\mathbf{G}$ is strictly increasing. Hence if $\xi \in \lambda$, then $\xi + 1 \in \lambda$ and $\mathbf{G}{\restriction}_\lambda(\xi + 1) > \mathbf{G}{\restriction}_\lambda(\xi)$. Hence

$$\mathbf{G}(\lambda) = \mathbf{F}(\mathbf{G}{\restriction}_\lambda) = \sup\{\gamma \mid \gamma \in \operatorname{range}(\mathbf{G}{\restriction}_\lambda)\} = \lim_{\delta<\lambda} \mathbf{G}(\delta). \quad \square$$

Hence $\mathbf{G}$ has all the properties of Definition 2.17, and so $\mathbf{G}$ is equal to $\mathbf{H}$.

4 Cardinals

4.1 Basic concepts

Theorem 4.1. *No set can be mapped onto its powerset.*

Proof. Fix a set X. Let $f : X \to \mathcal{P}(X)$. Define $D = \{a \in X \mid a \notin f(a)\}$. If for some $a \in X$, $f(a) = D$, then $a \in D$ if and only if $a \notin f(a) = D$, a contradiction. Hence D is not in the range of f. □

As there is a bijection between $\mathcal{P}(\mathbb{N})$ and $\mathbb{R}$ we obtain the following corollary.

Corollary 4.2. *The set $\mathbb{N}$ cannot be mapped onto the set $\mathbb{R}$.*

Theorem 4.3 (Schröder-Bernstein)**.** *If $f : X \to Y$ and $g : Y \to X$ are injections, then there exists a bijection $h : X \to Y$.*

Proof. Let $X_0 = X$ any $Y_0 = Y$. Inductively define $Y_{n+1} = f(X_n)$, and $X_{n+1} = X_n \setminus g(Y \setminus Y_{n+1})$. Now let $X_\omega = \bigcap_{n<\omega} X_n$ and $Y_\omega = \bigcap_{n<\omega} Y_n$. Define

$$h(x) = \begin{cases} f(x) & x \in X_\omega \\ g^{-1}(x) & \text{otherwise.} \end{cases}$$

Clearly $X_{n+1} \subseteq X_n$ which implies that $Y_{n+1} \subseteq Y_n$. Now if $x \notin X_\omega$, then x is in the range of g and so, as g is injective, h is well defined. Now h is onto because $y \in Y_\omega$ implies $y \in f(X_\omega)$ and $y \notin Y_\omega$ implies that $g(y) \notin X_\omega$ and so $h(g(y)) = y$. The function h is injective because $x_0 \in X_\omega$ implies that $f(x_0) \in Y(\omega)$ and so $g(f(x_0)) \in X_\omega$. Hence if $x_1 \notin X_\omega$ we have that $h(x_0) \neq h(x_1)$. □

Theorem 4.4. *For any set E, there is an ordinal α, and a bijection $f : \alpha \to E$.*

Proof. Short Version: Let $g : \mathcal{P}(E) \setminus \emptyset \to E$ be a choice function. Define a function $\mathbf{G} : \mathbf{ORD} \to E$ by transfinite recursion.

$$\mathbf{G}(\alpha) = \begin{cases} g(E \setminus \text{range}(\mathbf{G}\restriction_\alpha)) & \text{if } E \setminus \text{range}(\mathbf{G}\restriction_\alpha) \neq \emptyset \\ \star & \text{otherwise.} \end{cases}$$

There is some least α such that $\mathbf{G}(\alpha) = \star$ otherwise by replacement $\mathbf{ORD}$ is a set. Hence $\mathbf{G}\restriction_\alpha$ is a set bijection $h : \alpha \to E$.
Long version: We still start with the choice function $g : \mathcal{P}(E) \setminus \emptyset \to E$. Define a class function $\mathbf{F} : \mathbf{V} \to \mathbf{V}$ by

$$\mathbf{F}(x) = \begin{cases} g(E) & \text{if } x = \emptyset \\ g(E \setminus \text{range}(x)) & \text{if } x \text{ is a function and } E \setminus \text{range}(x) \neq \emptyset. \\ \star & \text{if } x \text{ is a function and } E \setminus \text{range}(x) = \emptyset \\ \emptyset & \text{Otherwise.} \end{cases}$$

Now apply Theorem 3.5, to obtain a class function $\mathbf{G} : \mathbf{ORD} \to \mathbf{V}$.

Claim 4.5. If $e \in E$, then there exists at most one ordinal δ such that $\mathbf{G}(\delta) = e$.

Proof. If δ is least such that $\mathbf{G}(\delta) = e$, and $\beta > \delta$, then $e \in \text{range}(\mathbf{G}\restriction_\beta)$. So $\mathbf{G}(\beta) = \mathbf{F}(\mathbf{G}\restriction_\beta) = g(E \setminus \text{range}(\mathbf{G}\restriction_\beta)) \neq e$. □

Claim 4.6. There is some ordinal α such that $\mathbf{G}(\alpha) = \star$.

Proof. The range of $\mathbf{G}$ is $E \cup \{\star\}$. If $\star$ is not in the range, then $\mathbf{G}$ defines an injective mapping from **ORD** to a subset of E, but this means there is a class mapping from this subset of E onto **ORD**. But this would imply by replacement that **ORD** is a set. Hence $\star$ is in the range of $\mathbf{G}$. □

By an application of Lemma 3, we can take α to be the least ordinal such that $\mathbf{G}(\alpha) = \star$. But as $\mathbf{G}(\alpha) = \mathbf{F}(\mathbf{G}\restriction_\alpha)$, this implies that $\text{range}(\mathbf{G}\restriction_\alpha)$ is equal to E. □

Corollary 4.7. *Any set can be wellordered.*

Proof. Let x be a set. Take a bijection $f : \alpha \to x$. Define a wellordering on x by for all $y, z \in x$ define $y < z$ if $f^{-1}(y) < f^{-1}(z)$. □

Definition 4.8.

(i) For any set X, let $|X|$ be the least ordinal α such that there is a bijection $f : X \to \alpha$.

(ii) An ordinal α is a *cardinal* if $|\alpha| = \alpha$.

Lemma 4.9. *Let X and Y be sets. Then $|X| \leq |Y|$ if and only if there is an injection from X to Y.*

Proof. Assume that $|X| \leq |Y|$. By definition, this means that there are cardinals α, β with $\alpha \leq \beta$ and bijections $f : X \to \alpha$ and $g : Y \to \beta$. Now as $\alpha \subseteq \beta$ we have that $g^{-1} \circ f$ is an injection from X to Y.

Now assume that there is an injection from X to Y. If $|Y| \leq |X|$, then by the first part of this proof there is an injection from Y to X and so by Theorem 4.3 we have a bijection between X and Y and so $|X| = |Y|$. □

Lemma 4.10.

(i) Any natural number is a cardinal.

(ii) ω is cardinal.

(iii) Every infinite cardinal is a limit ordinal.

(iv) If $f : X \to Y$ is onto, then $|Y| \leq |X|$.

(v) If $|Y| \leq |X|$ and $Y \neq \emptyset$, then there is an onto function $f : X \to Y$.

(vi) There is no largest cardinal.

Proof. (i) The empty set meets the definition to be a cardinal. Now assume that n meets the definition to be a cardinal where $n < \omega$. We will show that $n + 1$ is also a cardinal. If $n + 1$ is not a cardinal then there is a bijection $f : n + 1 \to m + 1$ for some $m < n$. (We can use $m + 1$ instead of m as clearly there is no bijection between $n + 1$ and the empty set.) Hence f contains an ordered pair (n, x) for some $x \leq m$. If $x = m$ then $f \setminus \{(n, x)\}$ is a bijection between m and n contradicting our assumption that n is a cardinal. If $x \neq m$, then for some $y < n$ the pair $(y, m) \in f$. In this case $(f \cup \{(y, x)\} \setminus \{(n, x), (y, m)\}$ is a bijection between m and n again contradicting our assumption that n is a cardinal.
(ii) If $f : \omega \to n$ is a bijection, then the mapping $g : \omega \to n + 1$ defined by $g(0) = n$ and $g(x + 1) = f(x)$ is a bijection between ω and $n + 1$. This gives a bijection between n and $n + 1$ contradicting (1).
(iii) If α is an infinite ordinal, then $f : \alpha \to S(\alpha)$ defined by $f(0) = \alpha$, $f(n + 1) = n$ for $n < \omega$ and $f(\beta) = \beta$ otherwise, is a bijection.
(iv) Define an injection $g : Y \to X$ by $g(y)$ is the least element of $f^{-1}(y)$.
(v) If $|Y| \leq |X|$ then there is an injection $f : Y \to X$. Now as $Y \neq \emptyset$, we can take some $y \in Y$. Define $g : X \to Y$ by $g(x) = y$ if $x \notin \text{range}(f)$ and $g(x) = f^{-1}(x)$ otherwise.
(vi) Assume that κ is the largest cardinal. Then $|\mathcal{P}(\kappa)| \leq |\kappa|$ and so $|\kappa|$ can be mapped onto its powerset contradicting Theorem 4.1.

□

You should convince yourself that the argument given in the proof of (i) *does not* imply that the successor ordinal of any cardinal is a cardinal.

For any cardinal κ, we define κ^+ to be least cardinal strictly greater than κ.

4.2 Cardinal arithmetic

We define addition and multiplication on infinite cardinals by

$$\kappa + \lambda = |\kappa \times \{0\} \cup \lambda \times \{1\}| \quad \text{and} \quad \kappa \cdot \lambda = |\kappa \times \lambda|.$$

Note that cardinal arithmetic and ordinal arithmetic are *not the same!* For example, cardinal addition and multiplication are both commutative.

Theorem 4.11. *If κ is an infinite cardinal, then $|\kappa \times \kappa| = \kappa$.*

Proof. Using any standard pairing function, we can show that this theorem holds for $\kappa = \omega$ (e.g. the function $(x, y) \mapsto ((x + y)(x + y + 1))/2 + x$ is a bijection between $\omega \times \omega$ and ω).

For $\kappa > \omega$, we define an ordering on $\kappa \times \kappa$ as follows.

$$(\alpha, \beta) <_p (\gamma, \delta) \text{ if } \begin{cases} \max(\alpha, \beta) < \max(\gamma, \delta) \\ (\alpha, \beta) <_{lex} (\gamma, \delta) & \text{otherwise.} \end{cases}$$

Let $W = (\kappa \times \kappa, <_p)$ be this ordering.

Claim 4.12. W is a wellorder.

Proof. If $(\alpha_1, \beta_1) >_p (\alpha_2, \beta_2) >_p \ldots$ is an infinite decreasing sequence in W, then the sequence $\max(\alpha_1, \beta_1), \max(\alpha_2, \beta_2), \ldots$ is a non-increasing sequence of ordinals and so there must be some ordinal δ and some n for all $m \geq n$, $\max(\alpha_m, \beta_m) = \delta$. But then $(\alpha_n, \beta_n) >_{lex} (\alpha_{n+1}, \beta_{n+1}) >_{lex} \ldots$ is an infinite descending sequence and we already know that the lexicographical order is wellfounded. □

If $ot(W) > \kappa$, then for some $(\alpha, \beta) \in \kappa \times \kappa$ we have that $W\restriction_{(\alpha,\beta)}$ is isomorphic to κ. Let $\delta = \max(\alpha, \beta) + 1$ so $\delta < \kappa$ as κ is a limit ordinal. Now we have that $ot(W\restriction_{(\delta,\delta)}) > \kappa$ but then arguing in terms of cardinalities we have

$$\kappa \leq |W\restriction_{(\delta,\delta)}| = |\delta \times \delta| = ||\delta| \times |\delta|| = |\delta|.$$

The last equality follows from the induction hypothesis. But $|\delta| < \kappa$ and we have a contradiction. Hence $ot(W) = \kappa$. □

Note in order to visualize this well-ordering, it is instructive to consider the case with $\kappa = \omega$.

Corollary 4.13. *If κ and λ are infinite cardinals, then* $\max(\kappa, \lambda) = \kappa + \lambda = \kappa \cdot \lambda$.

Proof. If $\kappa \geq \lambda$, then $\kappa \leq \kappa + \lambda \leq \kappa \cdot 2 \leq \kappa \cdot \lambda \leq \kappa \cdot \kappa = \kappa$. □

This theorem is very useful and particularly its corollary that given κ many sets each of size at most κ, the union of these sets has size at most κ. However, it does mean that cardinal addition and multiplication do not give us any more cardinals. There is one operation on cardinals that does however, cardinal exponentiation.

Definition 4.14.

(a) $X^Y = \{f : f : Y \to X\}$.

(b) $\kappa^\lambda = |\kappa^\lambda|$.

In many respects, cardinal exponentiation behaves as expected. Note that in the lemma below, cardinal addition and cardinal multiplication are used.

Lemma 4.15.

(a) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$.

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

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

(d) If $\kappa \leq \lambda$ *then* $\kappa^\mu \leq \lambda^\mu$.

(e) $\kappa^0 = 1$, $1^\kappa = 1$ *and* $0^\kappa = 0$ *if* $\kappa > 0$.

Proof. (a): To prove this equality we need to find a set L of cardinality $(\kappa \cdot \lambda)^\mu$ and a set R of cardinality $\kappa^\mu \cdot \lambda^\mu$ and then construct a bijection between them (alternatively we could construct two injections and use Theorem 4.3). We take L to be the set of functions f with domain μ and codomain $\kappa \times \lambda$. We take R to be $R_1 \times R_2$ where R_1 is the set of functions with domain μ and codomain κ, and R_2 is the set of functions with domain μ and codomain λ. We can obtain bijection $T : R \to L$ by $T((g, h)) = f$ where for all $\alpha \in \mu$ $f(\alpha) = (g(\alpha), h(\alpha))$.
(d): Because $\lambda \supseteq \kappa$, the set $\{f \mid \text{dom}(f) = \mu \text{ and codomain}(f) = \kappa\}$ can be injected into the set $\{f \mid \text{dom}(f) = \mu \text{ and codomain}(f) = \lambda\}$ via the inclusion mapping.
(e): The empty function is the unique function from $\emptyset$ to κ. The function $f(\alpha) = 0$ for all $\alpha \in \kappa$ is the unique function from κ to 1. If κ is not empty, then there are no functions from κ to $\emptyset$. □

Lemma 4.16. *If* $\lambda \geq \omega$, *and* $2 \leq \kappa \leq \lambda$ *then* $\kappa^\lambda = 2^\lambda = |\mathcal{P}(\lambda)|$.

Proof.

$$2^\lambda \leq \kappa^\lambda \leq \lambda^\lambda \leq |\mathcal{P}(\lambda \times \lambda)| = |\mathcal{P}(\lambda)| = 2^\lambda.$$

□

Using cardinals we can give a satisfactory definition of finite.

Definition 4.17. A set x is *finite* if $|x| < \omega$. A set x is *countable* if $|x| \leq \omega$.

Definition 4.18. Define by transfinite recursion,

$$\begin{aligned} &\aleph_0 = \omega \\ &\aleph_{\alpha+1} = \aleph_\alpha^+ \\ &\aleph_\beta = \sup\{\aleph_\alpha \mid \alpha < \beta\} \quad \text{for } \beta \text{ limit.} \end{aligned}$$

We originally discussed Cantor's continuum hypothesis in terms of sets of real numbers. However, an equivalent definition is the following.

Definition 4.19. The *continuum hypothesis* (CH) is the statement that $2^\omega = \aleph_1$. The *generalized continuum hypothesis* (GCH) is the statement that for all ordinals α, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$.

The notation ω_α is also used to refer to the cardinal $\aleph_\alpha$ particularly when referring to the order type of the cardinal.

4.3 Inaccessible cardinals

Definition 4.20. Let α be an ordinal. The *cofinality* of α, denoted $cf(\alpha)$ is the least ordinal β such that there exists an unbounded function $f : \beta \to \alpha$. (A function f is unbounded if for all $\delta \in \alpha$ there is some $\gamma \in \beta$ such that $f(\gamma) > \delta$.) Using the inclusion mapping we know that $cf(\alpha) \leq \alpha$.

Lemma 4.21. *If the mapping $f : \beta \to \alpha$ is cofinal, then there is a nondecreasing cofinal mapping $g : \widehat{\beta} \to \alpha$ for some $\widehat{\beta} \leq \beta$.*

Proof. Let $\widehat{\beta} \leq \beta$ be least such that $\sup\{f(\gamma) \mid \gamma < \widehat{\beta}\} = \alpha$. Define $g : \widehat{\beta} \to \alpha$ by $g(\delta) = \sup\{f(\gamma)) \mid \gamma \leq \delta\}$. □

Lemma 4.22.

(i) The cofinality of an ordinal is a cardinal.

(ii) For all α, $cf(cf(\alpha)) = cf(\alpha)$.

Proof. (i) Assume there is a cofinal mapping $f : \beta \to \alpha$. Let $h : |\beta| \to \beta$ be a bijection. Now $|\beta| \leq \beta$ (as an ordinal) and $f \circ h : |\beta| \to \alpha$ is a cofinal mapping.
(ii) Let $f : \beta \to \alpha$ be cofinal and nondecreasing. Let $g : \gamma \to \beta$ be cofinal and nondecreasing. Then $f \circ g : \gamma \to \alpha$ is cofinal and nondecreasing. □

Definition 4.23. An infinite cardinal $\aleph_\alpha$ is *regular* if $cf(\omega_\alpha) = \omega_\alpha$, otherwise it is *singular*.

Lemma 4.24. *Every successor cardinal is regular (i.e. for any cardinal κ, $cf(\kappa^+) = \kappa^+$).*

Proof. If not there is an cofinal function $f : \mu \to \kappa^+$ where $\mu \leq \kappa$ is a cardinal. Now for each $\alpha \in \mu$ use the axiom of choice to pick an onto mapping $g_\alpha : \kappa \to f(\alpha)$ (here we are regarding $f(\alpha)$ as a set of ordinals forming a strict initial segment of κ^+). Now define $g : \mu \times \kappa \to \kappa^+$ by $g(\alpha, \beta) = g_\alpha(\beta)$. Now g is a mapping onto κ^+ because if $\delta \in \kappa^+$ there is some $\alpha \in \mu$ such that $f(\alpha) > \delta$. Hence $\delta \in f(\alpha)$ and so for some $\gamma < \kappa$ we have $g_\alpha(\gamma) = \delta$. Thus $g(\alpha, \gamma) = \delta$. Hence $\kappa = |\mu \times \kappa| \geq \kappa^+$ a contradiction. □

So we know that every successor cardinal is regular, how about limit cardinals. For example, $\aleph_\omega$ is singular because the function $f : \omega \to \aleph_\omega$ defined by $f(n) = \aleph_n$ is cofinal. The cardinal $\aleph_0$ is regular because any mapping $n \mapsto \omega$ is bounded (where $n \in \omega$).

Definition 4.25.

(i) A *weakly inaccessible cardinal* is an uncountable regular limit cardinal.

(ii) A *strongly inaccessible cardinal* is an uncountable regular limit cardinal κ such that for all $\lambda < \kappa$, $2^\lambda < \kappa$.

The existence of weakly inaccessible cardinals in not provable in ZFC. It is even consistent that 2^ω is weakly inaccessible.

5 Zorn's Lemma

Definition 5.1. Let $\mathbb{P} = (P, <)$ be a partial order. A *chain* in $\mathbb{P}$ is a set $X \subseteq P$ such that for all distinct $x, y \in X$, either $x < y$ or $y < x$. A chain X in $\mathbb{P}$ has an *upper bound* if for some $b \in P$ for all $x \in X$ we have that $x \leq b$.

Note that chains can be infinite, even uncountable.

Definition 5.2. Let $\mathbb{P} = (P, <)$ be a partial order. An element $m \in P$ is maximal, if for all $x \in P$ it is not the case that $x > m$.

Theorem 5.3 (Zorn's Lemma)**.** *Let $\mathbb{P} = (P, <)$ be a partial order such that every chain has an upper bound. Let $a \in P$. Then there is a maximal element $b \in P$ such that $b \geq a$.*

Before proving Zorn's Lemma, we will apply it to establish the following theorem.

Theorem 5.4. *Any vector space has a basis.*

Recall that if V is a vector space, then $B \subseteq V$ is a basis for V if:

(a) Any element of V is a linear combination of a finite number of elements of B.

(b) Any finite subset of B is linearly independent.

In fact we will say that $X \subseteq V$ is linearly independent if all finite subsets of X are linearly independent.

For example, consider the space of vectors of length ω taking real values. What would a basis for this vector space look like?

Proof of Theorem 5.4. Let V be a vector space. Let $\mathbb{P} = (P, <)$ where P is the set of all linearly independent sets of vectors. If I_1 and I_2 are both linearly independent sets of vectors then $I_1 < I_2$ if $I_1 \subseteq I_2$.

Claim 5.5. If $X \subseteq P$ is a chain, then $\bigcup X \in P$.

Proof. Take any finite subset $\{v_1, v_2, \ldots, v_k\} \in \bigcup X$. Now for each $i \in \{1, \ldots, k\}$ take $x_i \in X$ such that $v_i \in x_i$. Observe that $<$ is a linear order on $\{x_1, \ldots, x_k\}$ and as any finite linear order has a maximum element, there is some $x \in \{x_1, \ldots, x_k\}$ such that $\{v_1, v_2, \ldots, v_k\} \subseteq x$. Hence as $x \in P$ we have that $\{v_1, v_2, \ldots, v_k\}$ is linearly independent. □

It follows that any chain $X \subseteq P$ has an upper-bound in P (specifically $\bigcup X$). Hence from Zorn's lemma we can let m be a maximal element in $\mathbb{P}$.

This means that any element of V can be written as linear combination of elements of m (if not take $v \in V$ such that v is not a linear combination of elements of m, then $m \cup \{v\}$ is a linearly independent set strictly greater than m). Hence m is a basis for V. □

Proof of Zorn's Lemma. Let $a \in P$. Let α be an ordinal such that there is a bijection $f : \alpha \to P$. Note that f exists by Theorem 4.4 which uses the axiom of choice. Define by transfinite induction for every $\beta < \alpha$.

$x_0 = a$.

$$x_{\beta+1} = \begin{cases} f(\beta) & \text{if } f(\beta) > x_\beta \\ x_\beta & \text{otherwise.} \end{cases}$$

For λ limit, set x_λ equal to an upper bound for $\{x_\gamma \mid \gamma < x\}$ (e.g. use the axiom of choice to choose one out of the set of upper bounds).

Let b be an upper bound for $\{x_\beta \mid \beta < \alpha\}$. In fact, as $f(\beta) = b$ for some β, it follows that for all $\gamma > \beta$, $x_\gamma = b$. Now b must be a maximal element of P, because if $c > b$, then $c > x_\beta$ for all $\beta < \alpha$. So if $f(\gamma) = c$, then $x_{\gamma+1} = c$ a contradiction. Further $b \geq a$ because for all $\beta < \alpha$ we have that $x_\beta \geq a$. □

In fact Zorn's lemma is equivalent to axiom of choice over the axioms of ZF.

Theorem 5.6. *Zorn's Lemma implies the axiom of choice.*

Proof. Fix a set X. Let $\mathbb{P} = (P, <)$ where P is the set of all *possibly partial* functions $f : \mathcal{P}(X) \setminus \emptyset \to X$ such that $f(z) \in z$ for all z in the domain of f. We say $f < g$ if g extends f as a function i.e. g agrees with f on all elements in the domain of f.

Note that a maximal element in this partial order is a total function (otherwise the function could be extended by adding one new subset to the domain). Hence we only need to prove that any chain has an upper bound. If $C \subseteq P$ is a chain, then again $\bigcup C$ is an upper bound (in this case $\bigcup C$ is the function that maps z to y, if for some $g \in C$, $g(z) = y$). □

Along with Theorem 4.4 we have nearly established the following.

Theorem 5.7. *Over ZF, the following are equivalent:*

(a) Axiom of Choice

(b) Any set X can we well-ordered (i.e. there is an ordinal α and a bijection $\alpha \mapsto X$).

(c) Zorn's Lemma.

Proof. $(a) + (b) \implies (c)$: This is the proof of Zorn's Lemma, we also used the axiom of choice implicitly to find upper bounds for chains.
$(a) \implies (b)$: Theorem 4.4.
$(c) \implies (a)$: Previous theorem.
Finally we need to show that $(b) \implies (a)$. Let X be a set, let $f : \alpha \to X$ be a bijection for some ordinal α. Now define a choice function $g : \mathcal{P}(X) \setminus \emptyset \to X$ by $g(S) = f(\beta)$ for the least β such that $f(\beta) \in S$. □

Definition 5.8. Let X be a set. We call $\mathcal{F}$ a *filter* on X if $\mathcal{F} \subseteq \mathcal{P}(X)$ and $\mathcal{F}$ has the following properties:

(i) $\emptyset \notin \mathcal{F}$.

(ii) $A \in \mathcal{F}$ and $A \subseteq B$ implies that $B \in \mathcal{F}$.

(iii) $A \in \mathcal{F}$ and $B \in \mathcal{F}$ implies that $A \cap B \in \mathcal{F}$.

You should think of the sets in the filter as being "large".

Example 1. Let $\mathcal{C} = \{X \subseteq \omega \mid X \text{ is cofinite}\}$. It is not difficult to verify that $\mathcal{C}$ is a filter on ω.

Definition 5.9. A filter $\mathcal{F}$ on X is an ultrafilter, if for all $Y \subseteq X$ either $Y \in \mathcal{F}$ of $X \setminus Y \in \mathcal{F}$.

Example 2. Let $\mathcal{U} = \{X \subseteq \omega | 17 \in X\}$. Then $\mathcal{U}$ is an ultrafilter. In this case a set is large if and only if it contains 17.

In general, given any non-empty set X and any $x \in X$, we can define and ultrafilter on X by letting $\mathcal{U} = \{Y \subseteq X \mid x \in Y\}$. An ultrafilter of this form is called principal. The following lemma is not difficult to prove.

Lemma 5.10. *If X is a finite set, and $\mathcal{U}$ is an ultrafilter on X, then $\mathcal{U}$ is a principal ultrafilter.*

Theorem 5.11. *There is a non-principal ultrafilter on ω.*

Proof. Let $P = \{\mathcal{F} \mid \mathcal{F} \text{ is a filter on } \omega\}$. Let $\mathbb{P} = (P, <)$ were $\mathcal{F}_1 < \mathcal{F}_2$ if $\mathcal{F}_1 \subsetneq \mathcal{F}_2$.

Claim 5.12. A maximal element of $\mathbb{P}$ is an ultrafilter

Proof of claim. Take $\mathcal{F} \in P$ such that $\mathcal{F}$ is not an ultrafilter. Then for some $x \subseteq \omega$ we have that $x \notin \mathcal{F}$ and $\omega \setminus x \notin \mathcal{F}$. Now assume that there are $z_1, z_2 \in \mathcal{F}$ such that $z_1 \cap x = \emptyset$ and $z_2 \cap (\omega \setminus x) = \emptyset$. This implies that $z_1 \cap z_2 = \emptyset$ which is impossible as $\mathcal{F}$ is a filter. Hence assume without loss of generality that for all $z \in \mathcal{F}$ we have that $z \cap x \neq \emptyset$. Now define $\mathcal{G}$ such that $\mathcal{G} = \{y \subseteq \omega \mid (\exists z \in \mathcal{F})(x \cap z \subseteq y\}$.

Let's check that $\mathcal{G}$ is a filter. We have ensured that $z \cap x \neq \emptyset$ for any $z \in \mathcal{F}$, and also that $y \in \mathcal{G}$ and $w \supseteq y$ implies that $w \in \mathcal{G}$. Finally if $y_1, y_2 \in \mathcal{G}$ then for some $z_1, z_2 \in \mathcal{F}$ we have $y_i \supseteq x \cap z_i$. But as $z_1 \cap z_2 \in \mathcal{F}$, it follows that $y_1 \cap y_2 \supseteq x \cap (z_1 \cap z_2)$ and so $y_1 \cap y_2 \in \mathcal{G}$. Observe that $x \in \mathcal{G}$ and $\mathcal{F} \subseteq \mathcal{G}$. Hence $\mathcal{G} > \mathcal{F}$ and so $\mathcal{F}$ is not a maximal element of $\mathbb{P}$. □

Let $X \subseteq P$ be a chain. I'll show that $\bigcup X$ is a filter in P. Clearly $\emptyset \notin \bigcup X$ and $y \supseteq x \in \bigcup X$ implies that $y \in \bigcup X$. If $x, y \in \bigcup X$ then for some $\mathcal{F}_1, \mathcal{F}_2 \in X$, $x \in \mathcal{F}_1$ and $y \in \mathcal{F}_2$. Without loss of generality, $\mathcal{F}_2 \geq \mathcal{F}_1$

and hence $x, y \in \mathcal{F}_2$ and and so $x \cap y \in \mathcal{F}_2 \subseteq \bigcup X$. As $\bigcup X$ is an upper bound for X in $\mathbb{P}$ it follows that any chain in $\mathbb{P}$ has an upper bound. Hence by Zorn's lemma for any $\mathcal{F} \in P$, there is an maximal element $\mathcal{U} \geq \mathcal{F}$ and we have already established that $\mathcal{U}$ must be an ultrafilter. Let $\mathcal{U}$ be a maximal element such that $\mathcal{U} > \mathcal{C}$ where $\mathcal{C}$ is the cofinite filter. Now if $\mathcal{U}$ was a principal ultrafilter then for some $n \in \omega$ we have that $\{n\} \in \mathcal{U}$ but this is impossible as $\omega \setminus \{n\} \in \mathcal{C} \subseteq \mathcal{U}$. Thus $\mathcal{U}$ is a non-principal ultrafilter. $\square$

6 Ramsey's Theorem

Let x be a set. We denote by $[x]^2$ all two-element subsets of x. We denote by $[x]^n$ all n-element subsets of x.

Theorem 6.1 (Ramsey's Theorem – infinite version)**.** *Let n, k be natural numbers. Let $c : [\omega]^n \to k$ be a function. Then there exists an infinite subset $h \subseteq \omega$ such that c is constant on $[h]^n$ (i.e. for all $p, q \in [h]^n$ we have that $c(p) = c(q)$).*

Note that we call c a *colouring*, and we call h a *homogeneous* set for the colouring.

Proof. We will prove the theorem for the case $n = k = 2$. Let $\mathcal{U}$ be a non-principal ultrafilter on ω. We will use $\exists^{\mathcal{U}} a \varphi(a)$ to mean

$$\{a \in \omega \mid \varphi(a)\} \in \mathcal{U}.$$

You can read $\exists^{\mathcal{U}} a \varphi(a)$ as "there are ultrafilter many a such that $\varphi(a)$". Now for each $a \in \omega$, either

(i) $\exists^{\mathcal{U}} b(c(\{a, b\}) = 1)$ or

(ii) $\exists^{\mathcal{U}} b(c(\{a, b\}) = 0)$.

Hence either $\exists^{\mathcal{U}} a \exists^{\mathcal{U}} b(c(\{a, b\}) = 0)$, or $\exists^{\mathcal{U}} a \exists^{\mathcal{U}} b(c(\{a, b\}) = 1)$. Assume without loss of generality that $\exists^{\mathcal{U}} a \exists^{\mathcal{U}} b(c(\{a, b\}) = 0)$. Let $U = \{a \in \omega \mid \exists^{\mathcal{U}} b(c(\{a, b\}) = 0)\}$. For each $a \in U$, let $U_a = \{b \in \omega \mid (c(\{a, b\}) = 0)\}$. Note that $U \in \mathcal{U}$ and for all $a \in U$, we have $U_a \in \mathcal{U}$.

Let $a_0 = \min U$. Once $a_0, a_1, \ldots, a_n$ have been defined let

$$a_{n+1} = \min(U \cap U_{a_0} \cap U_{a_1} \cap \ldots \cap U_{a_n} \cap \{x \mid x > a_n\}).$$

Now because $\mathcal{U}$ is a non principal ultrafilter, $\mathcal{U} \supset \mathcal{C}$ (homework exercise). Hence $\{x \mid x > a_n\} \in \mathcal{U}$ and so $(U \cap U_{a_0} \cap U_{a_1} \cap \ldots \cap U_{a_n} \cap \{x \mid x > a_n\}) \in \mathcal{U}$ and in particular it implies that this set is not empty. Hence a_{n+1} is well defined. Let $h = \{a_i \mid i \in \omega\}$. The set h is infinite, because $a_{n+1} > a_n$ as $a_{n+1} \in \{x \mid x > a_n\}$. If $i < j$ then $a_j \in U_{a_i}$ and hence $c(\{a_i, a_j\}) = 0$. Thus c is constant on $[h]^2$. □

Let us see one application of Ramsey's theorem. There is another application in the assignment.

Theorem 6.2. *Let L be an infinite set and $<_L$ be a linear order on L. Then there is a subset $\{l_0, l_1, \ldots\} \subseteq L$ such that either*

(a) For all $i \in \omega$, $l_i <_L l_{i+1}$; or

(b) For all $i \in \omega$, $l_i >_L l_{i+1}$.

In other words, any infinite linear order contains an infinite ascending sequence or contains an infinite descending sequence.

Proof. Take an injection $f : \omega \to L$. Define a colouring $c : [\omega]^2 \to \{0, 1\}$ as follows. For $a < b$, define

$$c(\{a, b\}) = \begin{cases} 0 & \text{if } f(a) <_L f(b) \\ 1 & \text{if } f(a) >_L f(b). \end{cases}$$

Let h be an infinite homogeneous set. If h is homogeneous for colour 0, then for any $a, b \in h$ we have that $a < b$ implies $f(a) <_L f(b)$. Hence $(L, <_L)$ contains an infinite ascending sequence (if $e_1 < e_2 < \ldots$ are all elements of h, then $f(e_1) < f(e_2) < \ldots$ is an ascending sequence of elements in L).

Similarly, if h is homogeneous for colour 1, then $(L, <_L)$ contains an infinite descending sequence. □

Theorem 6.3 (Ramsey's Theorem - finite version)**.** *For any $k \in \omega$, there is an $n \in \omega$ such that for any function $c : [n]^2 \to \{0, 1\}$ there is an $h \subseteq n$ such that c is constant on $[h]^2$ and $|h| = k$.*

An equivalent statement is the following. For any $k \in \omega$, there is an $n \in \omega$ such that for any graph G on n vertices, either G contains a clique of size k or G contains and anti-clique of size k.

Proof. Assume that the finite version of Ramsey's theorem fails to hold for some value k. Let $\mathcal{L}$ be the language with countably many constant symbols $\{c_i\}_{i \in \omega}$ and a single binary relation R. Let S be the all sentences of the following form for all $i, j \in \omega$

(i) If $i \neq j$, we have $c_i \neq c_j$.

(ii) $\neg c_i R c_i$.

(iii) $c_i R c_j \implies c_j R c_i$.

These sentences say that the constant symbols form a graph with an edge between c_i and c_j if $c_i R c_j$. We also add sentences to say that the graph contains no cliques or anticliques of size k. For all subsets $\{i_1, \ldots, i_k\} \subseteq \omega$ of size k we include in S the sentences

(i) $\neg(c_1 R c_2 \wedge c_1 R c_3 \wedge \ldots \wedge c_2 R c_3 \wedge \ldots \wedge c_{k-1} R c_k)$.

(ii) $c_1 R c_2 \vee c_1 R c_3 \vee \ldots \vee c_2 R c_3 \vee \ldots c_{k-1} R c_k$.

Now S is finitely satisfiable because any finite subset of S can only mention finitely many constants and we know for any n there is a graph of size n that has no cliques or anti-cliques of size k. Hence by the compactness theorem for first-order logic, there is a model $\mathcal{M}$ of S. But consider the colouring $c : [\omega]^2 \to \{0, 1\}$ where $c(\{i, j\}) = 1$ if and only if $\mathcal{M} \models c_i R c_j$. This colouring has no homogeneous set of size k, let alone an infinite homogeneous set. This contradicts the infinite version of Ramsey's theorem. □

Definition 6.4. A non-principal ultrafilter $\mathcal{U}$ is called *Ramsey* if for any partition $C_0, C_1, \ldots$ of ω, then either

(i) For some i, $C_i \in \mathcal{U}$; or

(ii) There is $E \in \mathcal{U}$ such that $|E \cap C_i| = 1$ for all i.

Lemma 6.5. *An ultrafilter is Ramsey if and only if for any function $c : [\omega]^2 \to 2$, there is a set $H \in \mathcal{U}$ such that H is homogeneous for c.*

Proof. Let $\mathcal{U}$ be an ultrafilter such that for any function $c : [\omega]^2 \to 2$, there is a set $H \in \mathcal{U}$ such that H is homogeneous for c. Fix n, define $c(\{i, j\}) = 0$ if and only if $n \in \{i, j\}$. Any infinite homogeneous set for c excludes n. Thus $\mathcal{U}$ is non-principal.

Assume that for any $c : [\omega]^2 \to 2$, there is a set $H \in \mathcal{U}$ such that H is homogeneous for c. Now let $C_0, C_1, \ldots$ be a partition of ω. Define a colouring c as follows

$$c(\{x, y\}) \coloneqq \begin{cases} 0 & \text{if } x \text{ and } y \text{ belong to the same partition} \\ 1 & \text{otherwise.} \end{cases}$$

Let $H \in \mathcal{U}$ be a homogenous set for c. Now if H is homogeneous for colour 0, every element of H must be in the same partition. Hence this partition must be in $\mathcal{U}$. If H is homogeneous for colour 1, then no two elements of H can be in the same partition. Hence $|H \cap C_i| \leq 1$ for all i. Now it easy to extend H to obtain $E \in \mathcal{U}$ such that $|E \cap C_i| = 1$ for all i. Thus $\mathcal{U}$ is Ramsey.

For the other direction, assume $\mathcal{U}$ is Ramsey and fix a function $c : [\omega^2] \to 2$. For all i define

$$A_i \coloneqq \begin{cases} \{x \colon c(\{i, x\}) = 0\} & \text{if this set is } \textit{not} \text{ in the ultrafilter} \\ \{x \colon c(\{i, x\}) = 1\} & \text{otherwise.} \end{cases}$$

Now turn these sets into a partition of ω by defining $D_0 \coloneqq A_0 \cup \{0\}$ and $D_{n+1} \coloneqq (A_n \cup \{n\}) \setminus \bigcup_{l<n} D_l$. As no D_i is in $\mathcal{U}$, there must be some $E \in \mathcal{U}$ such that $|E \cap D_i| = 1$ for all i.

Now if $i \in E$, then $A_i \cap E$ is finite so either for almost all $j \in E$, $c(\{i, j\}) = 0$ or for almost all $j \in E$ $c(\{i, j\}) = 1$. List E in increasing order $e_0, e_1, \ldots$. Define an increasing subsequence as follows $k_0 \coloneqq e_0$ and

$$k_{n+1} \coloneqq \min\{e_m \colon e_m > k_n \text{ and } \forall e_j \leq k_n (\exists i \in \{0, 1\} (\forall e_l \geq e_m (c(e_j, e_l) = 1)))\}.$$

This means that if $e_i \leq k_n$, then e_i gets the same colour with any e_j greater than k_{n+1}. Consider the following partition of ω:

$$\omega \setminus E, [0, k_0], (k_0, k_1], (k_1, k_2], \ldots.$$

None of these subsets are in $\mathcal{U}$ and so there must be an element of $\mathcal{U}$ that intersects each one exactly once. By ignoring the element of the first partition, we can take $B \in \mathcal{U}$ such that $B = \{b_0, b_1, \ldots\}$ where $b_0 \in [0, k_0]$ and for $n > 0$, $b_n \in (k_{n+1}, k_n]$. Now we have the property that each element $b_k \in B$ gets the same colour with every element above except possibly b_{k+1}. Let $B_0 \coloneqq \{b_i \colon i \text{ is even}\}$ and let $B_1 \coloneqq \{b_i \colon i \text{ is odd}\}$. One of these sets is in the ultrafilter and this set has the property any element gets the same colour with every element above it. Let this set be C. Now we partition C into C_0 and C_1 where

$$C_i \coloneqq \{x \in C \colon \text{ for all } y \in C, \text{ if } y > x \text{ then } c(\{x, j\}) = i\}.$$

Either C_0 or C_1 is in $\mathcal{U}$ and both are homogeneous sets for c. □

Note that the continuum hypothesis implies the existence of Ramsey ultrafilters but this cannot be done in ZFC.

7 Loś's Theorem

Let $\mathcal{L}$ be a countable language. Let $S = \{s_0, s_1, \ldots\}$ be a set of $\mathcal{L}$ sentences. Assume S is finitely satisfiable. For all $i \in \omega$, let M_i be a model of $\{s_0, \ldots, s_i\}$. Define

$$\mathcal{M} = \prod_{i \in \omega} M_i.$$

Now if $a \in \mathcal{M}$, then really a is a function from ω such that for all i, $a(i) \in M_i$.

Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathcal{M}$. For $a, b \in \mathcal{M}$, define $a \sim b$ if $\{i \mid a(i) = b(i)\} \in \mathcal{U}$. We will now define an $\mathcal{L}$-structure which we will denote by $\mathfrak{U}$. This is called the ultraproduct of $\mathcal{M}$ over $\mathcal{U}$. The domain of $\mathfrak{U}$:

$$\{[a] \mid a \in \mathcal{M}\}.$$

For c a constant symbol in $\mathcal{L}$ define $c^{\mathfrak{U}}$ to be $[i \mapsto c^{M_i}]$.

For R an n-ary relation define $R^{\mathfrak{U}}([a_1], \ldots, [a_n])$ to hold if

$$\{i \mid R^{M_i}(a_1(i), \ldots, a_n(i))\} \in \mathcal{U}.$$

(Check well defined).

The main result that we plan to show is that $\mathfrak{U}$ is a model of S. This will follow immediately from show that for any formula φ and $[a_1], \ldots, [a_n] \in \mathfrak{U}$,

$$\mathfrak{U} \models \varphi(a_1, \ldots, a_n) \text{ if and only if } \{i \mid M_i \models \varphi(a_1(i), \ldots, a_n(i))\} \in \mathcal{U}$$

8 Cumulative Hierarchy

We will now define the class $\mathbf{V}$ by transfinite recursion.

$$\begin{aligned} V_0 &= \emptyset \\ V_{\alpha+1} &= \mathcal{P}(V_\alpha) \\ V_\beta &= \bigcup_{\alpha<\beta} V_\alpha \text{ for } \beta \text{ limit.} \end{aligned}$$

This transfinite recursion provides a class mapping from the ordinals to the universe of sets $\alpha \mapsto V_\alpha$. Let $\mathbf{V} = \bigcup_{\alpha\in\mathbf{ORD}} V_\alpha$. To see that $\mathbf{V}$ is indeed a class, consider the formula that holds for a set x if there exists an ordinal α such that $x \in V_\alpha$.

Lemma 8.1. *The power set of a transitive set is transitive.*

Proof. Let x be a transitive set. If $z \in y \in \mathcal{P}(x)$, then $y \subseteq x$ so $z \in x$. By transitivity $z \subseteq x$ and so $z \in \mathcal{P}(x)$. □

Lemma 8.2. *For each α, V_α is transitive.*

Proof. $V_0 = \emptyset$ is transitive. If V_α is transitive then by previous lemma $V_{\alpha+1}$ is transitive. Finally the union of transitive sets is clearly transitive so if λ is a limit ordinal then V_λ is transitive. □

Lemma 8.3. *If $\alpha \leq \beta$, then $V_\alpha \subseteq V_\beta$.*

Proof. $V_\alpha \subseteq V_\alpha$. If $V_\alpha \subseteq V_\beta$ then $V_\alpha \in V_{\beta+1}$ and so $V_\alpha \subseteq V_{\beta+1}$ by transitivity. If λ is a limit and $V_\alpha \subseteq V_\beta$ for some $\beta < \lambda$, then $V_\alpha \subseteq V_\lambda$. □

Our goal is to prove the following theorem.

Theorem 8.4. *If x is a set, then $x \in \mathbf{V}$.*

In order to prove this, we need to introduce a new concept.

Definition 8.5. Given a set x, we define the *transitive closure* of x, $TC(x)$, as follows.

$$x_0 = x, \quad x_{n+1} = \bigcup x_n, \quad TC(x) \bigcup_{i\in\omega} x_i.$$

Exercise 1. Determine the transitive closure of $\{\{3\}, \{0, \{2, 3, \{4\}\}, 6\}\}$?

The following lemma is very similar to Lemma 3.

Lemma 8.6. *Every nonempty class has an ϵ-minimal element.*

Proof. Let $\mathbf{C}$ be a class and take $x \in \mathbf{C}$. If $\{y \in x \mid y \in \mathbf{C}\}$ is empty then x is an ϵ-minimal element of $\mathbf{C}$. Otherwise let y be ϵ-minimal in $TC(x) \cap \mathbf{C}$. Note that here we are using the axiom of foundation. Now if $z \in y$, then $z \in TC(x)$ so by minimality of y, $z \notin \mathbf{C}$. Hence y is an ϵ-minimal element of $\mathbf{C}$. □

Proof of Theorem 8.4. The complement of a class is a class, so if there is some x not in $\mathbf{V}$ we can take an ϵ-least such element. Hence for all $y \in x$ we have that $y \in \mathbf{V}$. Define a class mapping taking y to the least ordinal α such that $y \in V_\alpha$. By replacement the range of x under this mapping is a set of ordinals. Hence there is some ordinal β such that for all $y \in x$ we have $y \in V_\beta$. But then $x \subseteq V_\beta$ and so $x \in V_{\beta+1}$. □

One immensely useful thing that Theorem 8.4 gives us is a rank function. We define $\text{rank}(x)$ to be the least ordinal α such that $x \in V_{\alpha+1}$.

Lemma 8.7.

(i) If $x \in y$, then $\text{rank}(x) < \text{rank}(y)$.

(ii) If α is an ordinal then $\text{rank}(\alpha) = \alpha$.

Proof. (i): Assume $x \in y$. Let $\alpha = \text{rank}(y)$ so $y \in V_{\alpha+1}$. This means $y \subseteq V_\alpha$ and so $x \in V_\alpha$. If $\alpha = \gamma + 1$, then $\text{rank}(x) \leq \gamma < \alpha$. If α is a limit then as V_α is defined to be $\bigcup_{\gamma<\alpha} V_\gamma$, for some $\gamma < \alpha$, $x \in V_{\gamma+1}$.
(ii): First we show that $\text{rank}(\alpha) \leq \alpha$. We prove this by induction. $0 \in \mathcal{P}(\emptyset) = V_1 = V_{0+1}$. If $\alpha \in V_{\alpha+1}$, then $\alpha + 1 = \{\alpha\} \cup \alpha \subseteq V_{\alpha+1}$ by transitivity of $V_{\alpha+1}$ and so $\alpha + 1 \in V_{\alpha+2}$. If α is a limit ordinal then if $\gamma < \alpha$, $\gamma \in V_{\gamma+1}$ and so $\gamma \in V_\alpha$. Thus $\alpha \subseteq V_\alpha$ and so $\alpha \in V_{\alpha+1}$. This shows that $\text{rank}(\alpha) \leq \alpha$. Now if for some α, $\text{rank}(\alpha) < \alpha$, let α be least with this property. Let $\beta = \text{rank}(\alpha) < \alpha$. Now $\beta \in \alpha$ so by (i), $\text{rank}(\beta) < \text{rank}(\alpha) = \beta$, contradicting the minimality of α. □

9 Relativization

Let $\mathbf{C}$ be a class and $\mathbf{E}$ a binary class relation on $\mathbf{C}$. Let φ be a formula in the language of set theory with n free variables. Take $x_1, \ldots, x_n$ in $\mathbf{C}$. We will inductively define what it means for $(\varphi(x_1, \ldots, x_n))^{\mathbf{C},\mathbf{E}}$ to hold.

(i) $(x_1 \in x_2)^{\mathbf{C},\mathbf{E}}$ holds if $x_1 \mathbf{E} x_2$.

(ii) $(x_1 = x_2)^{\mathbf{C},\mathbf{E}}$ holds if $x_1 = x_2$.

(iii) $(\varphi \wedge \psi)^{\mathbf{C},\mathbf{E}}$ holds if both $\varphi^{\mathbf{C},\mathbf{E}}$ and $\psi^{\mathbf{C},\mathbf{E}}$ hold.

(iv) $(\neg\varphi)^{\mathbf{C},\mathbf{E}}$ holds if $\varphi^{\mathbf{C},\mathbf{E}}$ does not hold.

(v) $(\exists z(\varphi(z)))^{\mathbf{C},\mathbf{E}}$ holds if for some $y \in \mathbf{C}$, $(\varphi(y))^{\mathbf{C},\mathbf{E}}$ holds.

If $(\varphi(x_1, \ldots, x_n))^{\mathbf{C},\mathbf{E}}$ holds we will write $(\mathbf{C}, \mathbf{E}) \models \varphi(x_1, \ldots, x_n)$. Mostly, we will be interested in the case where $\mathbf{E}$ is the real ϵ relation. In this case we will write $(\mathbf{C}, \epsilon) \models \varphi(x_1, \ldots, x_n)$ or just $\mathbf{C} \models \varphi(x_1, \ldots, x_n)$.

Example 3. Let $M = \{0, 1, \{0, 1\}, \{0, 1, 2\}\}$ and let $N = \{0, 1, 2, \{0, 1\}, \{0, 1, 2\}\}$. Observe

(i) $M \subseteq N$.

(ii) $M \models \{0, 1, 2\} \subseteq \{0, 1\}$.

(iii) $N \models \{0, 1, 2\} \not\subseteq \{0, 1\}$.

Recall that the symbol $\subseteq$ is not in the language of set theory. The formula $x \subseteq y$ is short for $\forall z(z \in x \to z \in y)$. Item (iii) is true because because $\forall z \in M(z \in \{0, 1, 2\} \to z \in \{0, 1\})$. The point is that $2 \notin M$ so M cannot "know" that this element is in $\{0, 1, 2\}$ and not in $\{0, 1\}$.

Definition 9.1. Let $\mathbf{M} \subseteq \mathbf{N}$. We call a formula φ with n free variables *absolute between* $\mathbf{M}$ *and* $\mathbf{N}$ if for all $x_1, \ldots, x_n \in \mathbf{M}$ we have that

$$\mathbf{M} \models \varphi(x_1, \ldots, x_n) \Leftrightarrow \mathbf{N} \models \varphi(x_1, \ldots, x_n).$$

We call φ *absolute for* $\mathbf{M}$ if φ is absolute between $\mathbf{M}$ and $\mathbf{V}$.

If $\mathbf{M}$ is a transitive class, then $(\mathbf{M}, \epsilon)$ is called a transitive model (we will usually just call $\mathbf{M}$ a transitive model). If $\mathbf{M}$ and $\mathbf{N}$ are both transitive models with $\mathbf{M} \subseteq \mathbf{N}$, then many formulas are absolute between them.

Lemma 9.2. *If* $\mathbf{M}$ *is a transitive model, then the formula* $x \subseteq y$ *is absolute for* $\mathbf{M}$.

Proof. Fix x, y in $\mathbf{M}$. If $\mathbf{M} \models x \not\subseteq y$, then $\exists z \in \mathbf{M}(z \in x \wedge z \notin y)$. Now if $z \in \mathbf{M}$ then $z \in \mathbf{V}$ as all sets are in $\mathbf{V}$. Hence $\mathbf{V} \models x \not\subseteq y$.

Conversely, if $\mathbf{V} \models x \not\subseteq y$, then $\exists z(z \in x \wedge z \notin y)$. Now because $z \in x$ and $x \in \mathbf{M}$ we have by transitivity of $\mathbf{M}$ that $z \in \mathbf{M}$. Thus $\exists z \in \mathbf{M}(z \in x \wedge z \notin y)$ and so $\mathbf{M} \models x \not\subseteq y$. □

The argument used to prove Lemma 9.2 can be generalised to a large collection of formulas.

Definition 9.3. A formula φ is called a Δ_0 formula if:

(i) φ is an atomic formula.

(ii) φ is equal to $\neg\psi$ where ψ is a Δ_0 formula.

(iii) φ is equal to $\psi \wedge \theta$, $\psi \to \theta$ or $\psi \vee \theta$ where ψ and θ are Δ_0 formulas.

(iv) φ is equal to $(\exists x \in y)\psi$ or $(\forall x \in y)\psi$ where ψ is a Δ_0 formula.

Lemma 9.4. *If* $\mathbf{M}$ *is a transitive model and* φ *is a* Δ_0 *formula, then* φ *is absolute for* $\mathbf{M}$*.*

Proof. For a transitive model, $\mathbf{M}$ is $(\mathbf{M}, \epsilon)$ so $(x_1 \in x_2)^{\mathbf{M},\epsilon}$ holds if only if $x_1 \in x_2$, and $(x_1 = x_2)^{\mathbf{M},\epsilon}$ holds if and only if $x_1 = x_2$. Hence any atomic formula is absolute. Now assume φ and ψ are absolute.

(i) $\mathbf{M} \models \neg\varphi$ if and only if $\mathbf{M} \not\models \varphi$ if and only if $\mathbf{V} \not\models \varphi$ if and only if $\mathbf{V} \models \neg\varphi$.

(ii) $\mathbf{M} \models \varphi \wedge \psi$ if and only if $\mathbf{M} \models \varphi$ and $\mathbf{M} \models \psi$ if and only if $\mathbf{V} \models \varphi$ and $\mathbf{V} \models \psi$ if and only if $\mathbf{V} \models \varphi \wedge \psi$.

Hence the absolute formulas are closed under $\neg$ and $\wedge$. Similarly they are closed under $\vee$ and $\to$. The only thing that needs proof is to show that if $\varphi(x_1, \ldots, x_n, z)$ is absolute then so is $(\exists z \in y)\varphi(x_1, \ldots, x_n, z)$. Take $y, x_1, \ldots, x_n \in \mathbf{M}$.

If $\mathbf{M} \models (\exists z \in y)\varphi(x_1, \ldots, x_n, z)$, then for some $w \in y \cap \mathbf{M}$, $\mathbf{M} \models \varphi(x_1, \ldots, x_n, w)$. Because $w \in \mathbf{M} \subseteq \mathbf{V}$, and φ is absolute, we have that $\mathbf{V} \models \varphi(x_1, \ldots, x_n, w)$ and so $\mathbf{V} \models (\exists z \in y)\varphi(x_1, \ldots, x_n, z)$. The argument reverses because if $w \in y \cap \mathbf{V}$ we have that $w \in y \in \mathbf{M}$ and so by transitivity of $\mathbf{M}$, $w \in \mathbf{M}$. □

Lemma 9.5. *If* $\vdash (\forall x)(\varphi(x) \leftrightarrow \psi(x))$*, and* $\psi(x)$ *is absolute for* $\mathbf{M}$ *then* $\varphi(x)$ *is absolute for* $\mathbf{M}$*.*

Proof.

$$\mathbf{M} \models \varphi(x) \Leftrightarrow \mathbf{M} \models \psi(x) \Leftrightarrow \mathbf{V} \models \psi(x) \Leftrightarrow \mathbf{V} \models \varphi(x).$$

□

Theorem 9.6. *If $\alpha > \omega$ is a limit ordinal, then V_α models all axioms of ZFC except possibly replacement.*

Proof. Extensionality: The formula $\forall z(z \in x \leftrightarrow z \in y) \leftrightarrow y = z$ is equivalent to the Δ_0 formula $(\forall z \in x(z \in y) \wedge \forall z \in y(z \in x)) \leftrightarrow y = z$. This is absolute as V_α is a transitive model.
Pairing: We want to show that

$$V_\alpha \models (\forall x)(\forall y)(\exists z)[x \in z \wedge y \in z \wedge (\forall w \in z)(w = x \vee w = y)]$$

Take any $x, y \in V_\alpha$, we have that $z = \{x, y\} \subseteq V_\gamma$ for $\gamma < \alpha$, hence $z \in V_{\gamma+1} \subseteq V_\alpha$. Note that the formula $x \in z \wedge y \in z \wedge (\forall w \in z)(w = x \vee w = y)$ is Δ_0. Hence $V_\alpha \models x \in z \wedge y \in z \wedge (\forall w \in z)(w = x \vee w = y)$. Because x and y were arbitrary, the axiom of pairing holds.
Separation: For separation we need to be a little careful. Take $x \in V_\alpha$. Now $x \in V_\beta$ for some $\beta < \alpha$. Let $y = \{z \in x \mid V_\alpha \models \varphi(z)\}$. To prove that the axiom of separation holds, we need to show that y is in V_α. If we can ensure that y is in $\mathbf{V}$, then it follows that $y \in \mathcal{P}(x)$ and so $y \in V_{\lambda+2}$. But is y a set inside $\mathbf{V}$? It may be that $y \neq \{z \in x \mid \mathbf{V} \models \varphi(z)\}$ because φ is interpreted differently in $\mathbf{V}$ and V_α. However, by replacing any unbounded quantifiers $\exists x$ and $\forall y$ in φ with $\exists x \in V_\alpha$ and $\forall x \in V_\alpha$ respectively we can obtain a new formula $\widehat{\varphi}$. In this case, $y = \{z \in x \mid \mathbf{V} \models \widehat{\varphi}(z)\}$, and so $y \in \mathbf{V}$ (by separation in $\mathbf{V}$)which implies that $y \in V_{\beta+2} \subseteq V_\alpha$.
Union: Exercise.
Powerset: Recall this is $(\forall x)(\exists y)(\forall z)[z \in y \leftrightarrow z \subseteq x]$. If $x \in V_\alpha$, then let $y = \mathcal{P}(x)$. Now $x \subseteq V_\gamma$ for some $\gamma < \alpha$ and so $y \in V_\alpha$. Now $(\forall z \in y)(z \subseteq x)$ is Δ_0 so absolute. Hence $V_\alpha \models (\forall z \in y)(z \subseteq x)$. Now consider $(\forall z)(z \subseteq x \rightarrow z \in y)$. If $z \in V_\alpha$ and $z \subseteq x$, then $z \in y$ because y contains all subsets of x. Hence $V_\alpha \models \forall z(z \in y \leftrightarrow z \subseteq x)$. As x was arbitrary, the axiom of powerset holds.
Choice: Take $x \in V_\alpha$. Let $f : \mathcal{P}(x) \setminus \emptyset \rightarrow x$ be a choice function in $\mathbf{V}$. Now if f is in V_α, then f is a choice function for x in V_α because

(i) $\mathcal{P}(x)$ in $\mathbf{V}$ is the same as $\mathcal{P}(x)$ in V_α.

(ii) The only property that f needs to be a choice function is that $f(y) \in y$ for all $y \in \mathrm{dom} f$. The formula $a \in b$ is absolute so if $f(y) \in y$, then $(f(y) \in y)^{V_\alpha}$.

Now $f \subseteq \mathcal{P}(x) \times x$. We know that $x \in V_\gamma$ for some $\gamma < \alpha$. Now $x \subseteq V_\gamma$ so if $y \subseteq x$, then $y \in V_{\gamma+1}$. Hence $\mathcal{P}(x) \in V_{\gamma+2}$. If $a \in \mathcal{P}(x)$ and $b \in x$, then $(a, b) \in V_{\gamma+4}$. Hence f is in $V_{\gamma+5} \subseteq V_\alpha$.
Foundation: Take $x \in V_\alpha$ such that $x \neq \emptyset$. Let $y \in x$ be rank-minimal. Now $y \in V_\alpha$ by transitivity and clearly if $z \in y$ then the $\mathrm{rank}(z) < \mathrm{rank}(y)$ so $z \notin x$. Hence y is an ϵ-minimal element of x in V_α.

Infinity: Here we use that fact that $\alpha > \omega$ and so $\omega \in V_\alpha$. First, $\emptyset \in \omega$ and for all $x \in \omega$ we have that $x \cup \{x\} \in \omega$. Note that $y = \emptyset$ and $y = x \cup \{x\}$ are both expressible by Δ_0 formulas $(\forall z \in y)(z \neq z)$ and $(\forall z \in y)(z = x \vee z \in x) \wedge x \in y \wedge x \subseteq y$ respectively. As Δ_0 formulas are absolute and $\omega \in V_\alpha$ we have that

$$V_\alpha \models \emptyset \in \omega \wedge (\forall y \in \omega)(\{y\} \cup y \in \omega).$$

Hence the axiom of infinity holds in V_α. □

Theorem 9.7. *If κ is a strongly inaccessible cardinal, then V_κ is a model of ZFC.*

The proof of the following lemma is an exercise.

Lemma 9.8. *If κ is a strongly inaccessible cardinal, then for all $\alpha < \kappa$, $|V_\alpha| < \kappa$.*

Proof of Theorem 9.7. We only need to show that V_κ models replacement. Let us remind ourselves exactly what this means. For any formula $\varphi(x, y)$, if

$$V_\kappa \models \forall x \forall y \forall z(\varphi(x, y) \wedge \varphi(x, z) \to y = z) \tag{1}$$

then we need to show that

$$V_\kappa \models \forall a \exists b \forall y(y \in b \leftrightarrow (\exists x \in a)\varphi(x, y)). \tag{2}$$

Now assume that $\varphi(x, y)$ is a formula such that (1) holds. (This means that $\varphi(x, y)$ defines a class function inside of V_κ but we will not use this directly). Let a be any set in V_κ. Hence $a \in V_\beta$ for some $\beta < \kappa$ and it follows from Lemma 9.8 that $|a| < |V_\beta| < \kappa$. Define a mapping $f : a \to \kappa$ as follows

$$f(x) = \begin{cases} 0 & \text{if } V_\kappa \models \forall y \neg\varphi(x, y) \\ \alpha & \text{least such for some } y \in V_\alpha, V_\kappa \models \varphi(x, y). \end{cases}$$

The function f cannot be cofinal in κ because κ is regular. Hence there is some $\lambda < \kappa$, such that for all $x \in a$, if for some y, $V_\kappa \models \varphi(x, y)$ then $y \in V_\lambda$. If we define

$$b = \{y \in V_\lambda \mid (\exists x \in a)V_\kappa \models \varphi(x, y)\}$$

then $b \in V_{\lambda+1} \subseteq V_\kappa$. We have now done enough to establish (2). This means that

$$V_\kappa \models (1) \to (2)$$

and so regularity holds in V_κ. □

Corollary 9.9. *ZFC cannot prove the existence of strongly inaccessible cardinals.*

Proof. $Con(ZFC)$ is essentially the statement that the axioms of ZFC are consistent. This can be expressed in first-order logic by formalising the statement 'there is no proof of $0 = 1$ from the axioms of ZFC'. Gödel's incompleteness theorem established that *if the axioms of ZFC are consistent*, then there is no proof of $Con(ZFC)$ from these axioms. Now we have just shown that from ZFC plus there exists a strongly inaccessible cardinal, there is model of ZFC (i.e. V_κ where κ is strongly inaccessible). This gives a proof of $Con(ZFC)$ from the axioms of ZFC plus there is a strongly inaccessible cardinal. Hence we can conclude that ZFC cannot prove that there exists a strongly inaccessible cardinals because if so then ZFC could prove $Con(ZFC)$. □

10 Measurable Cardinals

Recall the definition of an ultrafilter. It follows from a simple induction that if $\mathcal{U}$ is an ultrafilter on a set X and $\{A_0, \ldots, A_n\} \subseteq \mathcal{U}$, then $\bigcap i \leq n A_i$ is an element of $\mathcal{U}$. We can ask that an ultrafilter has a stronger

Definition 10.1. Let κ be a cardinal. We call an ultrafilter $\mathcal{U}$ on set X κ-complete if any $\mathcal{A} \subseteq \mathcal{U}$ such that $|\mathcal{A}| < \kappa$ we have that $\bigcap \mathcal{A} \in \mathcal{U}$.

Hence a normal ultrafilter is ω-complete. We will now consider the question of whether a κ-complete ultrafilter exists for some $\kappa > \omega$.

Question 2. Can you put a κ-complete ultrafilter on $(2^\kappa)^+$?

If $\mathcal{U}$ be κ-complete cardinal on λ, then

- $cf(\lambda) > \kappa$.
- $2^\kappa < \lambda$.

Let κ be a cardinal. We call κ a *measurable cardinal* if there is a κ-complete non-principal ultrafilter on κ.

Proof. Let $\mathcal{U}$ be a κ-complete non-principal ultrafilter on κ. It should be clear that if $A \in \mathcal{U}$, then $|A| = \kappa$ (non-singleton is in the ultrafilter and the hence the union of less than κ many singletons is not in the ultrafilter). If $f : \lambda \to \kappa$ is cofinal in κ then $\bigcup_{\alpha<\lambda} f(\alpha)$ is not the ultrafilter but this set is equal to κ. (Recall $f(\alpha)$ is the set of ordinals less than $f(\alpha)$).

Assume that there is an injection $f : \kappa \to 2^\lambda$. For each $\alpha < \lambda$ there is a set $A_\alpha \subseteq \kappa$ such that $f(\alpha)$ is constant on κ. Now consider $\bigcap_{\alpha<\lambda} A_\alpha$. Elements in this intersection map to the same function hence this his size 1 contradicting the fact that the ultrafilter is non-principal. □

11 Perfect Set Property of Closed Sets

Recall that Cantor's continuum hypothesis is the following:

$$\text{If } E \subseteq \mathbb{R} \text{ is uncountable, then } |E| = |\mathbb{R}|$$
$$\text{(i.e. there is a bijection between } E \text{ and } \mathbb{R}).$$

The word continuum is used to refer to $|\mathbb{R}|$. We will prove is that the above statement is true if we add the condition that E is closed and so closed sets cannot form a counter-example to the continuum hypothesis. But before this, let us consider open sets. If $E \subseteq \mathbb{R}$ is an open set, then $E = \emptyset$ or $|E| = |\mathbb{R}|$. This is true because if E is non-empty, it must contain an open interval (a, b) with $a < b$. Clearly this interval has the same cardinality as $(-\pi/2, \pi/2)$. This later interval has the same cardinality as $\mathbb{R}$ via the bijection $x \mapsto \arctan(x)$.

Definition 11.1. We call $P \subseteq \mathbb{R}$ a perfect set if P is closed, non-empty, and P contains no isolated points.

Theorem 11.2. *Any perfect set has size continuum.*

Proof. Recall that 2^ω is the set of all functions from ω to 2. By a previous exercise, we have $|2^\omega|$ is equal to the continuum. The idea of this proof is to construct an injection from 2^ω to the perfect set. The set 2^ω can be thought of as the set of paths through an infinite binary branching tree. We will associate a real number in the perfect set with each node on this tree. These nodes are denoted by finite binary strings and we use λ to denote the empty string.

Let $P \subseteq \mathbb{R}$ be perfect. Take x_λ to be any element of P. Define $\epsilon_0 = 1$. In general, once x_σ has been defined for all strings of length n, let ϵ_{n+1} be less than $1/4$ the minimum distance between any distinct x_σs with $|\sigma| \leq n$ defined so far, and also less than $\epsilon_n/2$. We inductively define x_τ for all strings τ as follows. First assume that x_σ has been defined.

Let $x_{\sigma 0} = x_\sigma$. Let

$$x_{\sigma 1} = \begin{cases} \sup\{z \in P \mid x_\sigma < z < x_\sigma + \epsilon_{n+1}\} & \text{if it exists} \\ \inf\{z \in P \mid x_\sigma - \epsilon_{n+1} < z < x_\sigma\} & \text{otherwise.} \end{cases}$$

Because P has no isolated points, it follows that $x_{\sigma 1}$ exists. The real $x_{\sigma 1}$ is in P because P is closed.

We define a function $f : 2^\omega \to P$. For any $a \in 2^\omega$, let $f(a) = \lim_{n\to\infty} x_{a\restriction_n}$ (where $a\restriction_n$ is the finite binary sequence comprised of the first n bits of a). Consider the sequence $x_{a\restriction_0}, x_{a\restriction_1}, x_{a\restriction_2}, \ldots$, note that $d(x_{a\restriction_n}, x_{a\restriction_{n+1}}) \leq \epsilon_n$. Hence for all n, for all $m > n$, $d(x_{a\restriction_n}, x_{a\restriction_m}) \leq 2\epsilon_n$. (Here we use the fact that $\epsilon_{n+1} \leq \epsilon_n/2$.) Thus the sequence is Cauchy and so converges. Because P is closed we have that $f(a) \in P$.

Take any $a, b \in 2^\omega$ with $a \neq b$. Let σ be the longest initial segment of a such than σ is an initial segment of b. Let $n = |\sigma 0|$. Without loss of generality assume that $\sigma 0$ is an initial segment of a and that $\sigma 1$ is an initial segment of b. Consider $x_{\sigma 0}$. It must be that $f(a) \in \overline{B}(x_{\sigma 0}; 2\epsilon_{n+1})$ (this is the closed ball of radius $2\epsilon_{n+1}$). Also it must be that $f(b) \in \overline{B}(x_{\sigma 1}; 2\epsilon_{n+1})$. However ϵ_{n+1} was chosen to be less that $d(a, b)/4$. Hence these two balls do not intersect. Thus $f(a) \neq f(b)$. This means that f is injective and so $|\mathbb{R}| = |2^\omega| \leq |P| \leq |\mathbb{R}|$. Hence P has cardinality equal to the continuum. □

The Cantor-Bendixson Derivative

Given a set $E \subseteq \mathbb{R}$, let $Ac(E)$ be the set of all accumulation points of E i.e. non-isolated points. Fix a closed $C \subseteq \mathbb{R}$. Recall that ω_1 is used to denote $\aleph_1$ when considering $\aleph_1$ as an ordinal. We will now define a process known as the *Cantor-Bendixson derivative* of C. This is a mapping $f : \omega_1 \to \mathcal{P}(\mathbb{R})$ defined by transfinite recursion as follows:

$$\begin{aligned} f(0) &= C \\ f(\alpha + 1) &= Ac(f(\alpha)) \\ f(\lambda) &= \bigcap_{\beta < \lambda} f(\beta) \quad \lambda \text{ is a limit} \end{aligned}$$

Claim 11.3. There exists some $\alpha \in \omega_1$ such that $f(\alpha) = f(\alpha + 1)$.

Proof. Assume not, then for each α, $f(\alpha) \neq f(\alpha + 1)$ As $f(\alpha) \neq f(\alpha + 1)$, then there is some isolated point in $f(\alpha)$. Hence there is some open ball $B(q; r)$ with q, r rational such that $|f(\alpha) \cap B(q, r)| = 1$ and $f(\alpha + 1) \cap B(q, r) = \emptyset$. Define a mapping $\alpha \mapsto (q, r)$ for the least pair (q, r) such that this is true. This gives us an injection from ω_1 to $\omega \times \omega$ which is a contradiction. □

This α is called the *Cantor–Bendixson rank* of C.

Theorem 11.4. *Any uncountable closed set is the union of a perfect set and a countable set.*

Proof. Fixed a closed set C, and define the mapping f as above. Let α be least such that $f(\alpha) = f(\alpha+1)$. Observe that $C \setminus f(\alpha)$ is countable. This is true because for each point x removed for C at some point in the definition of f, there is a unique rational q, r such that $x \in B(q, r)$ i.e. the ball that isolates x. There are only countably many such balls.

If $f(\alpha) = \emptyset$, then C is countable. If $f(\alpha) \neq \emptyset$, then $f(\alpha)$ is a perfect set and $C = f(\alpha) \cup (C \setminus f(\alpha))$ is the union of a perfect set and a countable set. □

Hence no closed set can be a witness to the failure of the continuum hypothesis.

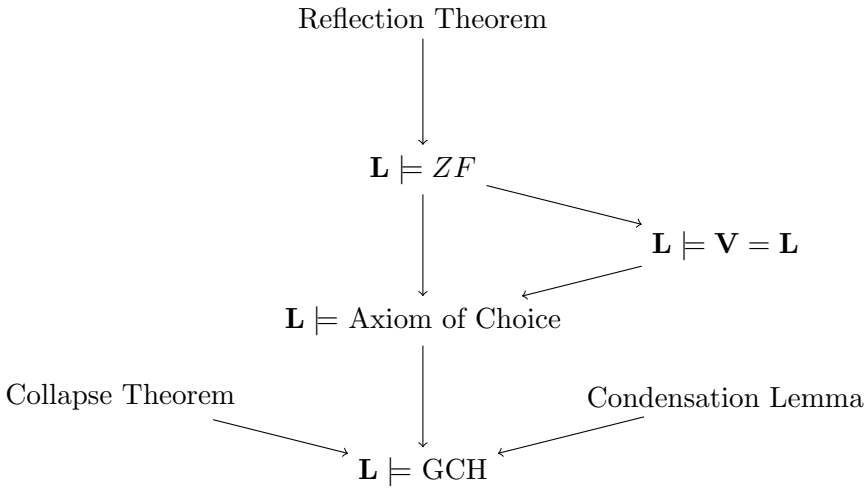

Figure 1: Outline of approach.

12 Gödel's Constructible Universe

Our objective is to prove that the continuum hypothesis is consistent with ZFC. We will do this by building a model of ZFC in which the continuum hypothesis holds. This model is called Gödel's constructible universe which we will denote by $\mathbf{L}$.

Recall that the continuum hypothesis states that $|2^\omega| = \aleph_1$. We know that $|2^\omega| = |\mathcal{P}(\omega)|$, so this indicates that we need to find a model of ZFC where the power set of ω is "small". What does this mean? If we have a model $\mathcal{M}$ of ZFC, then for some $x \in \mathcal{M}$, what $\mathcal{M}$ thinks of as the power set of x does not need to be the "real" power set of x i.e. we could have that $\mathcal{M} \models y = \mathcal{P}(x)$, but $\mathbf{V} \models y \neq \mathcal{P}(x)$. For example, if ZFC is consistent, then by the Lowenhiem-Skolem theorem, there is a *countable* set model $\mathcal{M}$ of ZFC. So from outside the model, any set inside the model (including whatever $\mathcal{M}$ thinks is 2^ω) is countable.

There is a lot of work involved in proving this result. This is not surprising as it is one of the major achievements of twentieth century logic. To obtain this result, we will need to prove a number of theorems on the way. The general plan is outlined in Figure 1.

One key idea in building Gödel's $\mathbf{L}$, is the definable power set operation.

Definition 12.1.

(i) Given $y \subseteq x$, we call y a *definable subset of* x if for some formula φ and $p_1, \ldots, p_n \in x$, we have that $z \in y$ if and only if $x \models \varphi(z, p_1, \ldots, p_n)$.

(ii) $\mathcal{P}_{\text{def}}(x) = \{y \subseteq x \mid y \text{ is a definable subset of } x\}$.

Lemma 12.2. *If* $|a| = \kappa \geq \omega$, *the* $|\mathcal{P}_{\text{def}}(a)| = \kappa$.

Proof. There are countably many formulas in the language of set theory. There are κ many possible parameters as they are κ many finite subsets of a. Hence there are at most $\omega \cdot \kappa = \kappa$ definable subsets of a. There are at least κ many definable subsets as for each $b \in a$, the singleton $\{b\}$ is a definable subset (defined by the formula $\forall y \in x(y = b)$). □

It is important to note that if $x \subseteq y$, then $\mathcal{P}_{\text{def}}(x)$ may not be equal to $\mathcal{P}_{\text{def}}(y) \cap \mathcal{P}(x)$. For example, let $z = \mathcal{P}_{\text{def}}(\omega)$. Now z is countable by the lemma above. Hence there is a surjection $f : \omega \to z$. Consider $\omega \cup \{f\}$. The set $d = \{n \mid n \notin f(n)\}$ is an element of $\mathcal{P}_{\text{def}}(\omega \cup \{f\})$. But $d \notin z$ by a diagonalization argument.

Theorem 12.3. *There is a* Δ_0 *formula* $\theta(x, y, z)$ *such that for all sets* a *and* b *such that* a *is transitive, we have that* $b = \mathcal{P}_{\text{def}}(a)$ *if and only if* $\exists z \theta(a, b, z)$.

This theorem is essential to the arguments that follow. We will not prove this theorem but only make the following remarks about it. We define a series of functions called the *basic Gödel operations*. These are:

(a) $G_1(x, y) = \{x, y\}$.

(b) $G_2(x, y) = x \times y$.

(c) $G_3(x, y) = \{(u, v) \in x \times y \mid u \in v\}$.

(d) $G_4(x, y) = x \setminus y$.

(e) $G_5(x, y) = x \cap y$.

(f) $G_6(x) = \bigcup x$.

(g) $G_7(x) = \text{dom}(x)$.

(h) $G_8(x) = \{(u, v) \mid (v, u) \in x\}$.

(i) $G_9(x) = \{(u, v, w) \mid (u, w, v) \in x\}$.

(j) $G_{10}(x) = \{(u, v, w) \mid (v, w, u) \in x\}$.

A *Gödel operation* is a composition of basic Gödel operations. Now we can define a function $g : \mathbf{V} \to \mathbf{V}$ by $g(x)$ is the closure of x under all Gödel operations. Theorem 12.3 is proved by showing that for a transitive set M:

$$\mathcal{P}_{\text{def}}(M) = \{y \subseteq M \mid y \in g(M \cup \{M\})\}. \tag{3}$$

The equality in (3) will become important when we show that the axiom of choice holds in $\mathbf{L}$.

Lemma 12.4. *The definable power set of a transitive set is transitive.*

Proof. Let x be a transitive set. If $z \in y \in \mathcal{P}_{\text{def}}(x)$, then $y \subseteq x$ so $z \in x$. By transitivity $z \subseteq x$ and so $z \in \mathcal{P}_{\text{def}}(x)$ because $z = \{w \in x \mid x \models w \in z\}$. □

Because of Theorem 12.3 and Lemma 12.4, we can make the following definition by transfinite recursion.

$$\begin{aligned} L_0 &= \emptyset \\ L_{\alpha+1} &= \mathcal{P}_{\text{def}}(L_\alpha) \\ L_\beta &= \bigcup_{\alpha<\beta} L_\alpha \text{ for } \beta \text{ limit.} \end{aligned}$$

This transfinite recursion provides a class mapping from the ordinals to the universe of sets $\alpha \mapsto L_\alpha$. Let $\mathbf{L} = \bigcup_{\alpha \in \mathbf{ORD}} L_\alpha$. Note that this definition is exactly the same as the definition of $\mathbf{V}$ except that we have replaced $\mathcal{P}$ by $\mathcal{P}_{\text{def}}$.

By the same argument we used for $\mathbf{V}$, we obtain that for each α, L_α is transitive (and so $\mathbf{L}$ is a transitive class) and additionally, if $\alpha \leq \beta$, then $L_\alpha \subseteq L_\beta$.

Remark 1. We will soon show that $\omega \in L_{\omega+1}$. However, $\mathcal{P}(\omega) \cap \mathbf{L} \neq \mathcal{P}(\omega) \cap L_{\omega+2}$. New subsets of the natural numbers (i.e. new real numbers) first occur in $\mathbf{L}$ at different ordinals. We will show that $\mathcal{P}(\omega) \cap \mathbf{L} = \mathcal{P}(\omega) \cap L_{\omega_1}$. This will be the key step in showing that the continuum hypothesis holds in $\mathbf{L}$.

Being a well-order is not absolute for transitive models. The problem lies in expressing that any subset of a well-order has a least element. However, we know by foundation that any set has an $\in$-least element. Hence to know if $(A, \in)$ is a well-order, we only need to know that $(A, \in)$ is a linear order. This is representable by a Δ_0 formula.

$$(\forall x \in A)(\forall y \in A)(\forall z \in A)((x \in y \vee y \in x \vee x = y) \wedge ((x \in y \wedge y \in z) \rightarrow x \in z)).$$

Hence being an ordinal is absolute for transitive models.

Lemma 12.5. *If M is a transitive set, then $y = \{\alpha \in M \mid \alpha$ is an ordinal $\}$ is in $\mathcal{P}_{\text{def}}(M)$.*

Proof. By the above argument there is a Δ_0-formula φ such that for all $x \in M$, $M \models \varphi(x)$ if and only if $\mathbf{V} \models \varphi(x)$ if and only if x is an ordinal. Hence $y = \{x \in M \mid M \models \varphi(x)\}$ is a definable subset of M containing all ordinals in M. □

Lemma 12.6. *For any ordinal α, $\alpha \in L_{\alpha+1}$.*

Proof. First, $0 = \emptyset = \{x \mid L_0 \models x \neq x\} \in L_1$. Now if for all $\beta < \alpha$, $\beta \in L_\alpha$, then $\alpha = \{y \mid L_\alpha \models y$ is an ordinal $\}$. This implies that $\alpha \in L_{\alpha+1}$. □

Theorem 12.7. **L** *is a model of ZF.*

Proof (the easy axioms).
Extensionality: (same argument as V_α).
Pairing:. Take $a, b \in \mathbf{L}$ so for some λ, $a, b \in L_\lambda$. Now $\{a, b\} \in \mathcal{P}_{\text{def}}(L_\lambda)$ (consider the formula $z = a \vee z = b$) and so $\{a, b\} \in \mathbf{L}$.
Union: Take $a \in \mathbf{L}$ so $a \in L_\alpha$ for some α. Let $u = \bigcup a$. We need to show that u is in **L**. Let b be the set of z such that $L_\alpha \models (\exists y)(z \in y \wedge y \in a)$. So $b \in \mathcal{P}_{\text{def}}(L_\alpha)$. Clearly $b \subseteq u$. Now take any $c \in u$. It must be that there is some d such that $c \in d \in a$. But as L_α is transitive we have that $c, d \in L_\alpha$. Hence $c \in b$ and so $u = b \in L_{\alpha+1}$. Finally $\mathbf{L} \models c \in u \leftrightarrow (\exists b(c \in b \in a))$ by absoluteness for Δ_0 formulas. Hence **L** models the axiom of union.
Foundation: (same argument as V_α).
Infinity: By Lemma 12.6, all ordinals are in **L**. So in particular $\omega \in \mathbf{L}$. This means that $\mathbf{L} \models \emptyset \in \omega \wedge \forall x \in \omega(x \cup \{x\} \in \omega)$ as this is a Δ_0 formula. Hence the axiom of infinity holds in **L**.
Power set: Let $x \in \mathbf{L}$. Let $M = \mathcal{P}(x) \cap \mathbf{L}$. Using the axiom of replacement in **V**, there is some λ such that $M \subseteq L_\lambda$. Now $M \in \mathcal{P}_{\text{def}}(L_\lambda)$ because M is equal to $\{z \in L_\lambda \mid z \subseteq x\}$. Hence $M \in \mathbf{L}$. Finally $\mathbf{L} \models (\forall z)(z \in M \leftrightarrow z \subseteq x)$, because this can be expressed by a Δ_0 formula. □

This leaves us with the axioms of separation and replacement. Replacement will follow easily once we have separation. To prove separation we need the following remarkable theorem.

Theorem 12.8 (Reflection Theorem). *For any formula φ and any ordinal α, there is an ordinal $\beta > \alpha$ such that φ is absolute between* **L** *and L_β.*

Using the Reflection Theorem we can complete the proof of Theorem 12.7.

Proof of Theorem 12.7 continued. Separation: Let φ be a first order formula, and let $x, p_1, \ldots, p_n \in \mathbf{L}$. Let $y \subseteq x$ be such that for all $z \in y$, $\mathbf{L} \models \varphi(z, p_1, \ldots, p_n)$. To prove the axiom of separation, we need to show that y is in **L**. Take some α such that $x, p_1, \ldots, p_n \in L_\alpha$. Now using the reflection theorem, find some $\beta > \alpha$ such that φ is absolute between L_β and **L**. I claim $y \in \mathcal{P}_{\text{def}}(L_\beta)$ and hence $y \in \mathbf{L}$. This claim holds because

$$z \in y \text{ if and only if } \mathbf{L} \models z \in x \wedge \varphi(z, p_1, \ldots, p_n) \text{ if and only if } L_\beta \models z \in x \wedge \varphi(z, p_1, \ldots, p_n).$$

Replacement: Now that we have separation, replacement is easy to prove. Assume $\mathbf{L} \models \forall x \forall y \forall z((\varphi(x, y) \wedge \varphi(x, z)) \rightarrow y = z)$. Now let $a \in \mathbf{L}$. By replacement in **V**, there is an ordinal α such that for all $x \in a$ if $\mathbf{L} \models (\exists y)\varphi(x, y)$ then this unique y is in L_α. Hence let $b = \{y \in L_\alpha \mid (\exists x \in a)\varphi(x, y)\}$. By separation this b is a set in **L** and hence

$$\mathbf{L} \models \forall a \exists b \forall y(y \in b \leftrightarrow (\exists x \in a)\varphi(x, y)).$$

□

A subformula of a formula φ, is formula contained in φ e.g. $x \in y \vee \forall z(z \in x)$ is a subformula of $(\forall x)(x \in y \vee \forall z(z \in x))$.

Proof of Theorem 12.8. Let $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$ list all the subformulas of φ (including φ itself). Define a class function $f : \mathbf{ORD} \to \mathbf{ORD}$ by $f(\delta)$ equals the least β such that for all $x_1, \ldots, x_k \in L_\delta$, if $\mathbf{L} \models (\exists z)\varphi_i(x_1, \ldots, x_n, z)$ then $\mathbf{L} \models (\exists z \in L_\beta)\varphi_i(x_1, \ldots, x_n, z)$.

Now define $\alpha_0 = \alpha$ and $\alpha_{n+1} = f(\alpha_n)$. Then let $\beta = \lim_{n\in\omega} \alpha_n$. Claim for all i, φ_i is absolute between L_β and $\mathbf{L}$. We prove this by induction. We already know that the atomic formulas are absolute and clearly the formulas absolute between L_β and $\mathbf{L}$ are closed under $\wedge$ and $\neg$ (see the proof of Lemma 9.4). Now assume that $\varphi_i(x_1, \ldots, x_n)$ is absolute between L_β and $\mathbf{L}$. Consider the formula $(\exists w)\varphi_i(w, x_2, \ldots, x_n)$. If for $x_2, \ldots, x_n \in L_\beta$, $L_\beta \models (\exists w)\varphi_i(w, x_2, \ldots, x_n)$, then $L_\beta \models \varphi_i(x_1, x_2, \ldots, x_n)$ for some $x_1 \in L_\beta$. Hence by absoluteness of φ_i, $\mathbf{L} \models \varphi_i(x_1, x_2, \ldots, x_n)$ and so $\mathbf{L} \models (\exists w)\varphi_i(w, x_2, \ldots, x_n)$.

Now assume that $\mathbf{L} \models (\exists w)\varphi_i(w, x_2, \ldots, x_n)$. There exists some k such that $x_2, \ldots, x_n \in L_{\alpha_k}$. Hence there is some $x_1 \in L_{\alpha_{k+1}}$ such that $\mathbf{L} \models \varphi_i(x_1, x_2, \ldots, x_n)$ but $x_1 \in L_\beta$ and so by absoluteness $L_\beta \models \varphi_i(x_1, x_2, \ldots, x_n)$ and hence $L_\beta \models (\exists w)\varphi_i(w, x_2, \ldots, x_n)$ □

The Reflection Theorem has significant generalisations that we will not cover in this course.

12.1 Choice and V = L

We will prove that the Axiom of Choice holds in $\mathbf{L}$, by showing that for all α, there is a well-ordering of L_α in $\mathbf{L}$. By Theorem 5.7, this is enough to show that choice holds in $\mathbf{L}$. We will denote this well-ordering by $<_\alpha$. In order to prove this result inductively, we will want the well-orderings that we define to work together. So we will ensure that for all ordinals α, β the following hold:

(i) For all $x \in L_\alpha \setminus L_\beta$ and for all $y \in L_\beta$, $y <_\alpha x$.

(ii) $<_{\alpha+1}$ extends L_β as a relation.

Clearly the empty relation well-orders $L_0 = \emptyset$. Now assume that we have defined $<_\alpha$ on L_α. For any n we can let $<_\alpha^n$ be the lexicographical ordering on n-tuples of elements of L_α using $<_\alpha$ i.e. if $a_1, \ldots, a_n, b_1, \ldots, b_n \in L_\alpha$ then $(a_1, \ldots, a_n) <_\alpha^n (b_1, \ldots, b_n)$ if $a_i <_\alpha b_i$ where i is the first place that the sequences differ.

At this point we make use of (3). We know that any element $x \in L_{\alpha+1}$ is equal to $g_k(p_1, \ldots, p_n)$ where g_k is a Gödel operation and $p_1, \ldots, p_n \in L_\alpha$.

The idea is to define $<_{\alpha+1}$ by considering the least Gödel operation and the least parameters that can be used to define x.

Let $\langle g_i \mid i \in \omega \rangle$ be an enumeration of all Gödel operations. Now define $<_{\alpha+1}$ as follows:

- If $x, y \in L_\alpha$, then $x <_{\alpha+1} y$ if and only if $x <_\alpha y$.
- If $x \in L_\alpha$ and $y \notin L_\alpha$, then $x <_{\alpha+1} y$.
- If $x, y \in L_{\alpha+1}$. Let (i, n) be least such that $x = g_i(z_1, \ldots, z_n)$ for some sets $z_1, \ldots, z_n \in L_\alpha$. Let $(a_1, \ldots, a_n)$ be least in $(L_\alpha)^n$ such that $x = g_i(a_1, \ldots, a_n)$. Let (j, m) be least such that $y = g_j(z_1, \ldots, z_m)$ for some sets $z_1, \ldots, z_m \in L_\alpha$. Let $(b_1, \ldots, b_m)$ be least in $(L_\alpha)^m$ such that $y = g_j(b_1, \ldots, b_m)$. Set $x <_{\alpha+1} y$ if $(i, n) < (j, m)$ or if $(i, n) = (j, m)$ and $(a_1, \ldots, a_n) <^n_\alpha (b_1, \ldots, b_n)$. Otherwise set $y <_{\alpha+1} x$.

If λ is a limit ordinal, then we simply define $x <_\lambda y$ if for some $\alpha < \lambda$, $x <_\alpha y$. This is clearly a total ordering and it is not too difficult to check that it is a well-ordering using the fact that $<_\alpha$ is a well-order.

Now we have shown that there is a well-ordering of L_α for each ordinal α. But we haven't shown that this well-ordering is in $\mathbf{L}$. We have only shown that $<_\alpha \in \mathbf{V}$. We do know that $\mathbf{L}$ is a model of ZF, and we did not use the axiom of choice in our construction of $\mathbf{L}$. Hence for any ordinal α we can let $(L_\alpha)^{\mathbf{L}}$ be the α^{th} level of the $\mathbf{L}$ hierarchy *constructed inside* $\mathbf{L}$ *itself!* We know that there is a well-ordering of $(L_\alpha)^{\mathbf{L}}$ inside $\mathbf{L}$ and so if $(L_\alpha)^{\mathbf{L}} = L_\alpha$ then we would be done. Clearly $(L_0)^{\mathbf{L}} = \emptyset = L_0$ and if λ is a limit ordinal, and for all $\alpha < \lambda$, $(L_\alpha)^{\mathbf{L}} = L_\alpha$ then $(L_\lambda)^{\mathbf{L}} = L_\lambda$. Hence we only need the following lemma.

Lemma 12.9. *If* $(L_\alpha)^{\mathbf{L}} = L_\alpha$, *then* $(L_{\alpha+1})^{\mathbf{L}} = L_{\alpha+1}$.

Proof. Let $a = L_\alpha$, let $b = (L_\alpha)^{\mathbf{L}}$. Now $\mathbf{L} \models b = \mathcal{P}_{\text{def}}(a)$. Hence by Theorem 12.3, $\mathbf{L} \models (\exists z)\theta(a, b, z)$. By the upwards absoluteness of Σ_1 formulas (exercise) we have that $\mathbf{V} \models (\exists z)\theta(a, b, z)$. Hence $\mathbf{V} \models b = \mathcal{P}_{\text{def}}(a)$ and so $b = L_{\alpha+1}$. □

The axiom of constructibility is typically written as the statement $\mathbf{V} = \mathbf{L}$. It states that every set in the universe occurs inside $\mathbf{L}$ i.e. $(\forall x)(\exists \alpha)(x \in L_\alpha)$. We have just established the following theorems.

Theorem 12.10.

$$\mathbf{L} \models \mathbf{V} = \mathbf{L}.$$

Proof. This holds by the discussion above and the fact that every ordinal is in $\mathbf{L}$. □

Theorem 12.11.

$$\mathbf{L} \models \textit{Axiom of Choice.}$$

Proof. We have shown that every set X in $\mathbf{L}$ can be well-ordered. This is enough to show that the Axiom of Choice holds in $\mathbf{L}$. (See Theorem 5.7 for details). □

Corollary 12.12. *If ZF is consistent then so is ZFC.*

Proof. Assume that ZF is consistent. Then there is a model M of ZF. We have shown that $(L)^M$ is a model of ZFC. Hence ZFC must be consistent. □

The following theorem lies at the heart of Gödel's Condensation Lemma.

Theorem 12.13. *There is a sentence σ such that for any transitive set M, $M \models \sigma$ if and only if $M = L_\delta$ for some limit ordinal δ.*

Proof Sketch. We would like to find a sentence σ such that if $M \models \sigma$, and M is transitive, then

(i) There is no largest ordinal in M.

(ii) $\delta \in \mathbf{ORD} \cap M \implies L_\delta \subseteq M$.

(iii) $x \in M \implies (\exists \delta \in \mathbf{ORD} \cap M)(x \in L_\delta)$.

These three facts would be enough to show that

$$M = \bigcup_{\delta \in \mathbf{ORD} \cap M} L_\delta.$$

The first statement is simple we just need to include $(\forall x)(\exists y)(y = x \cup \{x\})$. This holds in L_δ provided δ is a limit ordinal.

For the remaining statements we need to ensure that $(\forall x)(\exists y)(\exists z)\theta(x, y, z)$ holds in M. This ensures not only that the definable power set of any set exists, but that it can be correctly identified. It is not obvious that this holds in L_δ for limit δ but this can be shown by checking the definability of z.

Now we add a statement that says for all ordinals α, there is a function f such that

(i) $\mathrm{dom}(f) = \alpha$.

(ii) $f(0) = \emptyset$.

(iii) For all limit ordinals $\lambda \in \alpha$, $f(\lambda) = \bigcup_{\beta<\lambda} f(\beta)$.

(iv) If $\beta + 1 < \alpha$, then for some z, $\theta(f(\beta), f(\beta + 1), z)$ i.e. $f(\beta + 1)$ is the definable power set of $f(\beta)$.

If M models this statement, then for all $\beta < \alpha$, $f(\beta) = L_\beta$. By transitivity, $f(\beta) \subseteq M$. Finally we can adapt the above argument to produce a statement that says for any x, there is an ordinal α such that for the function f above, $x \in \bigcup \text{range}(f)$. □

The sentence σ can be used to prove the following theorem.

Theorem 12.14 (Gödel's Condensation Lemma). *If M is a transitive set and M is elementary equivalent to L_δ for some limit ordinal δ, then $M = L_\gamma$ for some limit ordinal γ.*

Proof. By Theorem 12.13, $L_\delta \models \sigma$ and so $M \models \sigma$ because M is elementary equivalent to L_δ. Again by Theorem 12.13, along with the transitivity of M, we have that $M = L_\gamma$ for limit ordinal γ. □

12.2 Generalized Continuum Hypothesis

Because we have shown that the axiom of choice holds in **L**, we know that all results on cardinal arithmetic hold. In this section, we will show that the generalized continuum hypothesis holds in **L**.

We have looked at a few absoluteness results that hold for transitive models of set theory. We will now consider non-transitive sets. Given a set M, we will define the collapse function $c : M \to \mathbf{V}$ by

$$c(x) = \{c(y) \mid y \in M \cap x\}.$$

This is an inductive definition on the rank of elements of M. Definitions such as this are used regularly in set theory. To see that c is defined for any set in M, assume that c is defined for all sets in M of rank less than α. Take any set $x \in M$ of rank α. Now any $y \in M \cap x$ has rank strictly less than α and so $c(y)$ is defined. Thus $c(x)$ is defined as well.

The following lemma is an immediate consequence of the definition.

Lemma 12.15.

(i) If $x, y \in M$ and $y \in x$, then $c(y) \in c(x)$.

(ii) If $x \in M$, and $a \in c(x)$, then for some $y \in x \cap M$, $c(y) = a$.

Note that this lemma implies that if c is a bijection, then it is an isomorphism.

Lemma 12.16. *The range of c is transitive.*

Proof. Let $N = c(M)$. If $b \in a \in N$, then there is some $x \in M$ such that $c(x) = a$. Now by definition $c(x) = \{c(y) \mid y \in M \cap x\}$ hence for some $y \in x$, $c(y) = b$ and therefore $b \in N$. □

The *transitive collapse* of M is the range of c. Note that the collapse function is not necessarily injective e.g. if $M = \{\{0,2\},\{0,3\}\}$ then $c(\{0,2\}) = c(\{0,3\}) = \emptyset$.

Example 4. Let $M = \{0,1,2,\{1,\{0,2\}\},\{2,\{0,2\}\}\}$. The transitive collapse of of M is the set $N = \{0,1,2,\{1\},\{2\}\}$ and in this case, the collapse function is an isomorphism between M and N.

Theorem 12.17 (Weak form of Mostowski collapse theorem). *If M models the axiom of extensionality then the collapse function is an isomorphism between M and the transitive collapse of M.*

Proof. We only need show that c is an injection (which implies it is a bijection with the range). Assume not, then there is some a of least rank such that there is a $b \neq a$ with $c(b) = c(a)$. Now as M models the axiom of extensionality, there is some $d \in M$ such that $d \in a \setminus b$ or $d \in b \setminus a$. Assume $d \in a \setminus b$. But as $c(d) \in c(a) = c(b)$, this means that there is some $e \in b$ such that $c(d) = c(e)$. However, $d \neq e$ which contradicts the minimality of a.

Similarly if $d \in b \setminus a$, then $c(d) \in c(b) = c(a)$, and so there is some $e \in a$ such that $c(d) = c(e)$. Again, $d \neq e$ and this time the existence of e contradicts the minimality of a. By Lemma 12.15, c is an isomorphism. □

Theorem 12.18. *The generalized continuum hypothesis holds in* $\mathbf{L}$.

Proof. Fix a cardinal ω_α. Take any $x \in \mathbf{L}$, such that $x \subseteq \omega_\alpha$. There is some limit ordinal δ such that $x \in L_\delta$. Let M be an elementary submodel of L_δ such that

(i) $x \in M$.

(ii) $\omega_\alpha \subseteq M$.

(iii) $|M| = |\omega_\alpha|$.

Let N be the transitive collapse of M. Let c be the collapse function. Observe that for all ordinals $\beta \in \omega_\alpha$, $c(\beta) = \beta$. This implies that $c(x) = x$. Note that N is elementarily equivalent to L_δ. Hence as N is transitive, by Theorem 12.14, for some limit ordinal γ, $N = L_\gamma$. But $|N| = |\omega_\alpha|$ and so $\gamma < |\omega_\alpha|^+$. Thus $x \in L_\gamma \subseteq L_{|\omega_\alpha|^+}$. Hence $\mathcal{P}(\omega_\alpha) \cap \mathbf{L} \subseteq L_{|\omega_\alpha|^+}$. But (by exercise) $|L_{|\omega_\alpha|^+}| \leq |\omega_\alpha|^+$ and so

$$\mathbf{L} \models |\mathcal{P}(\omega_\alpha)| = |\omega_\alpha|^+.$$ □

Corollary 12.19. *ZFC cannot prove the existence of weakly inaccessible cardinals.*

Proof. If α is weakly inaccessible, then α is strongly inaccessible in L. Hence $(V_\alpha)^{\mathbf{L}}$ is a set model of ZFC but this would prove the consistency of ZFC. □

In fact the assumption that $\mathbf{V} = \mathbf{L}$ resolves many set theoretic questions. However, most set theorists believe that $\mathbf{V}$ and $\mathbf{L}$ are quite different as the following quotation indicates.

> Maybe the following analogy will explain my attitude; we use the standard American ethnic prejudice and status system, as it is generally familiar. So a typical universe of set theory is the parallel of Mr. John Smith, the typical American; my typical universe is quite interesting (even pluralistic), it has long intervals where GCH holds, but others in which it is violated badly, many 's such that Souslin trees exist and many 's for which every Aronszajn is special, and it may have lots of measurables, with a huge cardinal being a marginal case but certainly no supercompact. This seems not less justifiable than stating that Mr. John Smith grew up in upstate New York, got his higher education in California, dropped out from college in his third year, lived in suburbia in the Midwest, is largely of anglo-saxon stock with some Irish or Italian grandfather and a shade of hispanic or black blood, with a wife living separately and 2.4 children. "Come on," I hear, "how can you treat having no or even CH? - you cannot say somewhere yes somewhere no!" True, but neither could Mr. Smith have 2.4 children, and still the mythical "normal" American citizen is in a suitable sense a very real one. In this light, $\mathbf{L}$ looks like the head of a gay chapter of the Klu Klux Klan - a case worthy of study, but probably not representative.
>
> – Saharon Shelah

13 Banach-Tarski Paradox

13.1 Rotations in $\mathbb{R}^3$

I will start by introducing $SO(3)$, the group of all rotations of the sphere. We can analyse rotations in $\mathbb{R}^3$ using a little linear algebra. Recall that a square matrix A is called *orthogonal* if $A^TA = I$ (which implies that $AA^T = I$). Observe that the determinant of an orthogonal matrix is either 1 or -1.

One key property of orthogonal matrices is that they preserves dot products i.e. if A is orthogonal then $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$. This means that the linear transformation induced by an orthogonal matrix preserves length and preserves orthogonality. This in turn implies that the columns of any orthogonal matrix form an orthonormal basis for $\mathbb{R}^3$. It is easy to verify that the converse holds i.e. if A is a matrix whose columns are an orthonormal basis for $\mathbb{R}^3$, then A is orthogonal.

If T is a rotation linear transformation in $\mathbb{R}^3$, then the matrix associated with T is orthogonal. To see this note that the unit vectors $\{e_1, e_2, e_3\}$ must be mapped under T to an orthonormal basis for $\mathbb{R}^3$. The converse does not quite hold. Once a rotation T has determined $T(e_1)$ and $T(e_2)$, then $T(e_3)$ has been uniquely determined, however there are two possible ways of extending $\{T(e_1), T(e_2)\}$ to an orthonormal basis, we could add $T(e_3)$, or $-T(e_3)$. Hence, in order to obtain the group $SO(3)$ as a subgroup of $GL_3(\mathbb{R})$ we need to restrict ourselves to transformations that are orientation preserving. This gives us that $SO(3)$ is the group of all orthogonal matrices in $GL_3(\mathbb{R})$ with determinant equal to 1.

Because a rotation in $SO(3)$ is length preserving, it maps the unit sphere S^2 to itself. So we can consider how these rotations affect on S^2. For any rotation $g \in SO(3)$ we have an induced automorphism of $SO(3)$ by the mapping $\vec{x} \mapsto g\vec{x}$. This is an example of a group action on a space because

(i) The identity element of $SO(3)$ induces the identity automorphism.

(ii) For any $g, h \in SO(3)$ and $\vec{x} \in S^2$, we that that $(gh)\vec{x} = g(h(\vec{x}))$.

The Banach-Tarski paradox says that we can partition S^2 into finitely many pieces $A_1, \ldots, A_n, B_1, \ldots B_m$ such that for some rotations $g_1, \ldots, g_n$, $h_1, \ldots h_m$ we have that:

(i) $S^2 = g_1A_1 \sqcup \ldots \sqcup g_nA_n$.

(ii) $S^2 = h_1B_1 \sqcup \ldots \sqcup h_mB_m$.

A key step in the proof of the Banach-Tarski paradox is showing that $SO(3)$ contains a subgroup isomorphic to the free group on two generators. We will investigate this subgroup but first let us prove some facts about rotations.

Lemma 13.1. *If a rotation fixes more than two points on S^2 it must be the identity.*

Proof. Let A be a rotation that fixes more than two distinct points. It follows that A fixes two points $\vec{u}$, $\vec{v}$ such that $\vec{u} \neq -\vec{v}$. Because it is a linear transformation, the rotation fixes all points in the two dimensional subspace generated by $\vec{u}$, $\vec{v}$. Let $\mathcal{B} = \{\vec{x}, \vec{y}, \vec{z}\}$ be an orthonormal basis for $\mathbb{R}^3$ such that $\vec{x}, \vec{y}$ are in the span of $\{\vec{u}, \vec{v}\}$. Take a_1, a_2, and a_3 such that $A\vec{z} = a_1\vec{x} + a_2\vec{y} + a_3\vec{z}$. This means we can write the rotation A with respect to the basis $\mathcal{B}$ as

$$A_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & a_1 \\ 0 & 1 & a_2 \\ 0 & 0 & a_3 \end{bmatrix}$$

Let P be a matrix that changes coordinates from the standard basis to $\mathcal{B}$. This means that

$$A = P^{-1} A_{\mathcal{B}} P.$$

Now as $\det(A) = 1$, it follows that $\det(A_{\mathcal{B}}) = 1$ and so $a_3 = 1$. But because $\mathcal{B}$ is an orthonormal basis for $\mathbb{R}^3$ we have that $\sqrt{(a_1)^2 + (a_2)^2 + (a_3)^2}$ is equal to the length of $A\vec{z}$ which is 1. Hence $\vec{a} = \vec{e_3}$. Hence A is the identity transformation. □

Lemma 13.2. *If a rotation is not the identity it fixes exactly two points which are antipodal.*

Proof. As A is an orthonormal matrix we have that

$$\det(A - I) = \det(A - I)^T = \det(A^T - I) = \det(A)\det(A^T - I) = \det(AA^T - A) = \det(I - A) = -\det(A - I).$$

Hence $\det(A - I) = 0$ and so 1 is a eigenvalue for A. Hence A has a eigenvector $\vec{x}$ of length 1. Thus $\vec{x}$ and $-\vec{x}$ are antipodal points on S^2 that are fixed by A. There are exactly two points by the previous lemma. □

Note that this last lemma is a theorem of Euler.

13.2 The free group

We denote by $\{a, b, a^{-1}, b^{-1}\}^{<\omega}$ the set of all finite words that can formed using the alphabet $\{a, b, a^{-1}, b^{-1}\}$. Let $|w|$ denote the length of the word w. We call a word w *reduced* if w does not contain any of the following sub-words: aa^{-1}, $a^{-1}a$, bb^{-1}, or $b^{-1}b$. The operation $^\frown$ is used to denote concatenation.

We call u a sub-word of w if $w = s^\frown u^\frown t$ for some words s and t. Clearly if w is a reduced word and u is a sub-word of w then u is also reduced. Say $w \lhd u$ if $w = v_1{}^\frown v_2$ and $u = v_1{}^\frown s^\frown v_2$ where $s \in \{aa^{-1}, a^{-1}a, bb^{-1}, b^{-1}b\}$.

Lemma 13.3. *For any word w, there is a unique reduced word u such that either $u = w$ or $u = u_0 \lhd u_1 \lhd \ldots \lhd u_n \lhd w$.*

Proof (Not assessed). As words have finite length the existence of a u follows because if $s \lhd t$ then $|s| < |t|$ so this operation can only be applied finitely often.

Assume that for some word w and some distinct reduced words u_0 and v_0 we have $u_0 \lhd u_1 \lhd \ldots \lhd u_n \lhd w$ and $v_0 \lhd v_1 \lhd \ldots \lhd v_m \lhd w$. Without loss of generality we can assume that u_0 is the empty word and w has minimal length. Clearly w does not have length 0, 1, or 2.

As w has minimal length we know that $v_m \neq u_n$. Take s, t such that $v_m = s^\frown t$ and $w = s^\frown c^\frown c^{-1}{}^\frown t$. In the reduction to the empty string via the sequence $u_0 \lhd u_1 \lhd \ldots \lhd u_n \lhd w$ we know that these c and c^{-1} are removed at some point. If they are removed together, then $s^\frown t$ also has a reduction to the empty string by applying all the other reductions in the sequence $u_0 \lhd u_1 \lhd \ldots \lhd u_n \lhd w$. This contradicts the minimality of w.

Otherwise, this c is removed with some $d = c^{-1}$ at some point and c^{-1} is removed with some $e = c$ at some point. So we can write w in one of the following forms where s_1, s_2, t_1, and t_2 are strings.

(i) $t_1{}^\frown d^\frown s_1{}^\frown c^\frown c^{-1}{}^\frown s_2{}^\frown e^\frown t_2$.

(ii) $t_1{}^\frown e^\frown s_1{}^\frown d^\frown s_2{}^\frown c^\frown c^{-1}{}^\frown t_2$.

(iii) $t_1{}^\frown c^\frown c^{-1}{}^\frown s_1{}^\frown e^\frown s_2{}^\frown d^\frown t_2$.

But now note that using the reduction of w to the empty string, it is possible to reduce all of s_1, s_2 and $t_1{}^\frown t_2$ to the empty string. Hence there is reduction of v_n to the empty string contradicting the minimality of w. For example, consider the second case above. This means that $v_n = t_1{}^\frown e^\frown s_1{}^\frown d^\frown s_2{}^\frown t_2$. Start by reducing s_1 to the empty string, then remove the pair ed, then reduce s_2 to the empty string, and finally reduce $t_1{}^\frown t_2$ to the empty string. This again contradicts the minimality of w. □

Let F_2 be the set of all reduced words i.e.

$$F_2 = \{w \in \{a, b, a^{-1}, b^{-1}\}^{<\omega} \mid w \text{ is reduced}\}.$$

We define an operation on reduced words as follows. Let $u \star w$ be equal to the unique reduced word of $u^\frown w$. It is not difficult to see that $(F_2, \star)$ forms a group. The operation $\star$ is associative because $(u \star w) \star v$ is the unique reduced word of $u^\frown w^\frown v$ as is $u \star (w \star v)$. The empty word is the identity and for inverses just reverse words and change a, b, a^{-1}, b^{-1} to a^{-1}, b^{-1}, a, b respectively, e.g. the inverse of $aba^{-1}b$ is $b^{-1}ab^{-1}a^{-1}$.

Lemma 13.4. *There is a decomposition of F_2 into four pieces A_1, A_2, A_3 and A_4 such that $F_2 = A_1 \sqcup b \star A_2 = A_3 \sqcup a \star A_4$.*

Proof. We partition F_2 as follows.

- Let $A_1 = \{w \in F_2 \mid w \text{ starts with } b\}$.
- Let $A_2 = \{w \in F_2 \mid w \text{ starts with } b^{-1}\}$.
- Let $A_3 = \{w \in F_2 \mid w \text{ starts with } a^{-1}\} \cup \{a^{-n} \mid n \in \mathbb{N}\}$.
- Let $A_4 = F_2 \setminus (A_1 \cup A_2 \cup A_3)$.

It is not difficult to check that this partition has the desired properties. □

Lemma 13.5. *There is a subgroup of $SO(3)$ isomorphic to the free group with two generators.*

Proof. Refer to the exposition of the Banach-Tarski paradox by Terry Tao handed out in the lecture. □

Consider what this lemma means. There are two rotations, let's call them a and b such that if u and w is sequences of rotations a, a^{-1}, b and b^{-1} then u and w are only identical if they have the same reduced word. For example, this means that $aba^{-1}b^{-1}$ is not the identity.

Let us denote this subgroup by H. For any $x, y \in S^2$, say $x \sim y$ if for some $g \in H$, we have that $gx = y$. This is an equivalence relation. Let $[x] = \{y \in S^2 \mid x \sim y\}$. The set $[x]$ is called the orbit of x.

There are two possible cases for the orbit of x that we need to consider. First assume that for all $g, h \in H$ such that $g \neq h$ we have that $gx \neq hx$. In this case we say that H *acts freely* on x. This implies that the orbit of x looks like the Cayley graph of the free group.

Lemma 13.6. *The set of points on which H does not act freely is countable.*

Proof. If H does not act freely on x, then take h, g such that $h \neq g$ and $hx \neq gx$. This means that $x = h^{-1}gx$ and so x is fixed by the rotation $h^{-1}g$ and $h^{-1}g \neq e$. But any rotation in H that is not the identity fixes exactly two points. As there are countably many elements of H, there are only countably many points on which H does not act freely. □

Let C be the countable set of points in S^2 on which H does not act freely. For now we are going to ignore these points. Observe that C is a collection of equivalence classes.

Here comes the set theory. For each equivalency class in $S^2 \setminus C$ choose an element. Let E be the set of representatives picked. Define

(i) $A_1 = \{y \in S^2 \mid (\exists x \in E)(\exists w \in H)(y = wx \text{ and } w \text{ starts with } b)\}$.

(ii) $A_2 = \{y \in S^2 \mid (\exists x \in E)(\exists w \in H)(y = wx \text{ and } w \text{ starts with } b^{-1})\}$.

(iii) $A_2 = \{y \in S^2 \mid (\exists x \in E)(\exists w \in H)(y = wx \text{ and } w \text{ starts with } a^{-1})\} \cup \{y \in S^2 \mid (\exists x \in E)(\exists n \in \mathbb{N})(y = a^{-n}x)\}$.

(iv) $A_4 = S^2 \setminus (A_1 \cup A_2 \cup A_3)$.

This decomposition is essentially that given in Lemma 13.4, but now we decompose each equivalence class. It follows that $S^2 \setminus C = A_1 \sqcup b \star A_2 = A_3 \sqcup a \star A_4$.

We have almost completed our proof of the Banach-Tarski paradox. We just need to deal with C, the remaining countable set. There are only countably many points in C, so we can choose a rotation $c \in H$ such that for all $x, y \in C$ and all $n \in \mathbb{N} \setminus \{0\}$ we have that $c^n x \neq y$. We can do this is follows. Choose an axis of rotation that avoids any of the points in C (possible as we have uncountably many choices). Now take an angle rotation around this axis. There are also uncountably many choices, and for each pair $(x, y) \in C \times C$ there at most countably many c such that $c^n x = y$ for some non-zero n. Hence there are only countably many angles of rotation to avoid. Let $\widehat{C} = \bigcup_{i=0}^{\infty} c^i C$.

Here is a process using the rotations a, b and c, to create two spheres from one.

(i) Starting with S^2, apply the rotation c to $\widehat{C}$. This gives $(S^2 \setminus \widehat{C}) \sqcup c\widehat{C}$ which is equal to $S^2 \setminus C$.

(ii) Now divide $S^2 \setminus C$ into the four pieces A_1, A_2, A_3, and A_4 as described above. Create two copies of $S^2 \setminus C$.

(iii) For each of these copies of $S^2 \setminus C$, apply the rotation c^{-1} to $\widehat{C} \cap (S \setminus C)$. This gives two copies of $S \setminus \widehat{C} \sqcup c^{-1}(\widehat{C} \setminus C) = S^2$.

Because we used 2 pieces in the first step, 4 pieces in the second step and 2 pieces in the third step, we could have instead used 16 pieces in the first step. This proves the Banach-Tarski paradox.