MATH 145A - SET THEORY I, FALL 2019
COURSE NOTES

SEBASTIEN VASEY

Contents

Date: November 18, 2019.

1. Introduction: What are sets, and what are they good for?

Intuitively, a set is just a collection of mathematical objects. For example, the collection of all real numbers, the collection containing only the number 1, or the collection of all triangles in the Euclidean plane, are sets. As usual in mathematics, we restrict ourselves to collections that can be described precisely, so the collection of all unicorns, of all birds, or of all random numbers, are not sets.

One way to describe a set is by listing its elements between curly brackets: the second set in our list was the set $\{1\}$. An interesting set is the empty set, denoted by $\emptyset$: it is the unique set containing no elements. Crucially, a set can itself contain other sets. Consider for example the set containing the empty set, $\{\emptyset\}$ (this is different from the empty set itself, for the same reason that receiving an empty box as a gift is more pleasant than receiving no gift at all!).

So far so good, but a playful mathematician may start asking: can a set contain itself? For example, the collection V of all sets should itself be a set, and by definition be contained in itself. As Bertrand Russell puts it, the collection of all ideas is itself an idea. On the other hand it seems most sets do not contain themselves (the empty set does not, and in fact none of the other examples we have seen do). It becomes tempting to look at the set of all sets that do not contain themselves. This is similar to writing the ultimate catalog of all non self-referential books, or asking whether the (male) barber that shaves every men not shaving themselves should shave himself, or taking me seriously when I say I am lying to you. Try to think about it: either way gives a contradiction. The non-existence of the set of all sets not containing themselves is called Russell's paradox (a more precise statement will be given later, see Theorem 2.8).

The bottom line is that just defining a set as a collection of objects, while often good-enough to get an intuition, is too vague to precisely study the mathematics of sets. In fact, "defining" a set to be a collection of objects makes the situation worse: it replaces one undefined term (set) by two undefined terms (collection and objects).

A strategy often followed in mathematics when one does not know what something is the *axiomatic method*: we keep the term undefined, but simply gives a list of properties that it should satisfy. This should already be somewhat familiar. For example, in axiomatic geometry one never defines "points" and "lines", but one has an axiom saying for example that any two points lie on a line ("lie" is also an undefined term). Modern set theory is developed in a similar way: we will have an undefined (or primitive) term "set", as well as a relation $\in$ (being a member of), and will have a number of axioms describing how sets ought to behave with respect to the relation $\in$. For example, for any sets x and y, there should be a set z so that $x \in z$, $y \in z$, and no other element is in z. We will write $\{x, y\}$ for this set. One will have an axiom saying certain sets exist (for example the empty set, or the natural numbers), as well as axioms saying how to build new sets from old ones (by taking union, intersection, power sets, etc.). The problem of what "objects" a set should contain will be solved by a radical decision: a set can only contain sets, nothing else. Numbers, shapes, etc. will all be defined in terms of sets. For example, we will define the number 0 to be the empty set, the number 1 to be the set $\{\emptyset\} = \{0\}$, the number 2 to be the set $\{0, 1\}$, and generally a natural number will be the set of its predecessors. Other mathematical objects such as real numbers, functions, relations, groups, topological spaces, etc. can all be defined in terms of sets.

This approach to set theory makes it into a rigorous foundation for mathematics, as we will outline during the first few classes. Afterward, the focus will shift to the ideas and techniques from set theory that are beautiful on their own and can also be used as mathematical tools. Let's survey some of these here.

1.1. **The sizes of infinity.** Set theory often focuses on *infinite sets*. At the end of the 19th century, Georg Cantor realized that there was a way to make sense of "how big" infinite sets were by looking at the functions one can define between them. Specifically, recall that a function f from a set X to a set Y is a *bijection* if distinct members of X are sent to distinct members of Y (i.e. f is an injection), and for any member y of Y, there is x in X so that $f(x) = y$ (i.e. f is a surjection). A bijection can be thought of as a way to rename one set into another. For example, the map $f : \{1, 2, 3\} \to \{4, 5, 6\}$ given by $f(x) = x + 3$ is a bijection. The map $f : \mathbb{R} \to \mathbb{R}_{>0}$ given by $f(x) = e^x$ is also a bijection. There is a bijection between from a finite set to another if and only if they have the same number of elements. Thus it makes sense to define two (possibly infinite) sets X and Y to *have the same cardinality* if there is a bijection from X to Y. We have seen already that $\mathbb{R}$ has the same cardinality as $\mathbb{R}_{\geq 0}$, and you may have seen before that $\mathbb{Z}$ and $\mathbb{N}$ also have the same cardinality (if not, try to prove it!). Using the now famous "diagonal argument" (actually closely related to Russell's paradox) Cantor proved that $\mathbb{N}$ and $\mathbb{R}$ did *not* have the same cardinality. Thus in a sense the (infinite) size of $\mathbb{R}$ is strictly bigger than the (infinite) size of $\mathbb{N}$. Cantor showed more generally that for any set X, $\mathcal{P}(X)$ has strictly bigger size than X.

Is there a size between that of $\mathbb{N}$ and $\mathbb{R}$? That is, is there a set of reals that is in bijection with neither the reals nor the natural numbers? Cantor conjectured that there was no such set, and the question was baptized *the continuum hypothesis*. Legend has it that Cantor went mad trying to prove it. This was for good reasons, as much later Kurt Gödel and Paul Cohen established that there is no way to prove the continuum hypothesis... Or to disprove it. The axioms of set theory do not give us enough information to solve it (in this class, we will see one direction of this: that it is impossible to disprove the continuum hypothesis).

Still, there is a beautiful theory of infinite cardinals. Cardinals are linearly ordered (for any two sets, one is always bigger than or equal to the other), one can define an arithmetic on cardinals, and prove that for infinite cardinals λ and μ, $\lambda + \mu = \lambda \cdot \mu = \max(\lambda, \mu)$ (in fact a common theme is that infinite sizes are easier to understand than finite sizes). One can generalize questions around the continuum hypothesis by studying the infinite cardinal exponential function $\lambda \mapsto 2^\lambda$.

1.2. Ordinals and transfinite induction.

Closely related to cardinals are ordinals. Ordinals indicate a relative position rather than a size: a box may contain *two* sweets (cardinal), and one may decide on an order on them, leading to a *first* sweet and a *second* sweet. Similarly to cardinals, there is a theory of infinite ordinals. It is familiar to anyone that has ever thought about how far one can count: $0, 1, 2, 3, 4, \ldots, \infty, \infty + 1, \infty + 2, \ldots \infty + \infty = \infty \cdot 2, \ldots \infty \cdot \infty, \ldots$ are only the first few ordinal numbers. Since the ∞ symbol is overused, one calls the first infinite ordinal ω instead.

Ordinals are very useful in understanding constructions with "infinitely many steps". Consider for example nested sets such as $\{\emptyset\}$ or $\{\{\{\emptyset\}, \emptyset\}\}$. One can try to measure the complexity of a set by counting the maximal number of "nesting" that appears. For example, $\emptyset$ has complexity zero, $\{\emptyset\}$ has complexity one, $\{\emptyset, \{\emptyset\}\}$ has complexity two, etc. Things are clear-enough as long as one considers only finite levels of nesting, but consider the following problem. Let $X_0 = \emptyset$, $X_1 = \{\emptyset\}$, $X_2 = \{\{\emptyset\}\}$, and generally $X_{n+1} = \{X_n\}$ for n a natural number. The complexity of X_n is n. Let $X = \bigcup_n X_n$. What should its complexity be? It cannot be any finite n, so it should be infinite. What if we want to be more precise? Its complexity should clearly be as low as possible among the infinite complexity. Let us call this complexity ω. Now consider $Y = X \cup \{X\}$. What is the complexity of Y? Clearly, Y has more nestings than X. In fact, it is tempting to say the complexity of Y should be that of X plus one, $\omega + 1$. The theory of ordinals allow us to make this precise. How high can the complexity of a set go? The answer is "as high as possible", i.e. any ordinal can be the complexity of a set, as we will see in this class.

While nesting empty sets is fun, one may reasonable ask for an example in another part of mathematics. The following definition occurs in analysis: a set of real numbers is *Borel* if it can be obtained from the following operations:

(1) Any open interval is Borel.
(2) The complement of a Borel set is Borel.
(3) If $(A_n)_{n \in \mathbb{N}}$ is a sequence of Borel set, then the union $\bigcup_{n \in \mathbb{N}} A_n$ is Borel.

(One way to make this more precise is to let the collection of Borel sets be the intersection of all collection of sets of reals that contain the open intervals and are closed under complements and taking countable unions – this does not however allow us to precisely measure the Borel complexity of sets)

This is reminiscent of a definition by induction. Open intervals are the base case, and complement and union are the inductive step. However problem occurs when one asks how many steps it took to build a given Borel set A. Say A comes from a union of Borel sets: $A = \bigcup_n A_n$. Maybe it took n steps to build A_n. How many steps then did it take to build A? Infinitely many? But then how many steps does it take to build A^c? The more precise answers are ω and $\omega + 1$: infinite ordinals can be used to describe inductions that take more than finitely-many steps. These are called *transfinite* inductions. In fact one contribution of set theory is a very general theory of induction beyond the natural numbers.

Another application of ordinals is given by Cantor's proof (to be seen in this class) that the continuum hypothesis is true for closed sets! Any closed subset of the real numbers is either countable or of the same cardinality as the reals. The key concept is that of the derivative of a closed set: given a closed set $X \subseteq \mathbb{R}$, X' is the set of limit points of X: the set of $x \in \mathbb{R}$ so that for all $\epsilon > 0$, $(x-\epsilon, x+\epsilon) \cap X \neq \{x\}$ (that is, the set of all points of X that are "surrounded by other things"). One can readily check that X' is also a closed set. For example, if $X = \{0\} \cup \{\frac{1}{n} \mid n = 1, 2, 3, \ldots\}$, then $X' = \{0\}$, and $X'' = \emptyset$. On the other hand, if $X = [0, 1]$, then $X' = X = [0, 1]$. Sets X so that $X' = X$ are called *perfect*. One can prove that non-empty perfect sets have the same cardinality as the reals.

Thus it is natural to ask what happens if we iterate this derivative operation. Define $X^{(0)} := X$, and $X^{(n+1)} = \left(X^{(n)}\right)'$. Let $X^{(\omega)} := \bigcap_{n \in \mathbb{N}} X^{(n)}$. Do we have that $\left(X^{(\omega)}\right)' = X^{(\omega)}$? Turns out the answer is "not necessarily" (finding an example is challenging but try to think about it — first find a closed X so that $X'' \neq X'''$). Still, at each iteration of the derivative, we get a subset, and so (by counting the reals) the iteration must stabilize at some point. However, finite steps are not enough: one must continue deriving "into the transfinite" to get the desired set which is its own derivative. In fact, we will see there is a countable ordinal α so that $X^{(\alpha+1)} = X^{(\alpha)}$. The least such ordinal gives us a measure of the complexity of the set X. In fact a carefuly analysis of this process shows that any closed set is either countable or contains a non-empty perfect set, hence has the cardinality of the reals.

1.3. **Infinite games.** The following very general framework for games is productively studied in set theory. We consider two-player games. The players alternate playing natural numbers (any number is allowed). The rules of the games are given as a set X of sequences of natural numbers, giving all possible plays that are winning for player I. Call this game $G(X)$.

For example, let X be the set of all sequences of natural numbers of the form $(a_0, a_1, a_2, \ldots)$, where $\sum_{k=0}^{2n} a_n$ is even for all natural numbers n. For example $(2, 2, 2, 2, 2, 2, ...)$ is in X, and so is $(4, 1, 5, 0, 0, 0, ...)$, but $(0, 1, 0, 1, 0, 1, ...)$ is not in X. During the game, player I plays a_0, player II replies with a_1, player II replies with a_2, etc. At the end, player I wins if and only if $(a_0, a_1, \ldots)$ is in X. Now suppose you are player I: can you figure out a strategy to win $G(X)$?

This may seem like an artificial framework, but whatever the "real" game, natural numbers can often be used to code the player's moves; for example a move in chess can be described as a sequence of two coordinates, like E2-E4. There are 64 possible coordinates, so a chess move can naturally be seen as a sequence of two numbers between 1 and 64 (which can easily be encoded into a single number between 1 and 64^2). Some of these numbers do not code valid chess moves, but we can encode this into the rules of the game, by saying that if an illegal move was made, the player who first made an illegal move loses (this is actually a rule in some competitive forms of chess). As opposed to chess, we do not allow a draw: one of the player must win at the end (if we are playing chess, we could declare that a draw counts as a win for player II).

For another point of view, think of the moves as specifying with more and more precision a real number (say between 0 and 1 you can think of a_n as the nth decimal digit of the number – with perhaps $a_n \geq 10$ being forbidden by the rules, or you can think of a_n as coding a rational approximation to the real number). The real number that is obtained in the end must be in the set X for player I to win. Formally, define the game $G^*(X)$, for $X \subseteq [0, 1]$ as follows: player I and II alternate playing natural numbers a_n between 0 and 9, and player I wins exactly when $0.a_0a_1\ldots$ is in X. As for chess, this can be coded as $G(X^*)$, for some X^*.

Suppose for example that $X = [0, 0.5]$. Then player I instantly wins by specifying $a_0 = 3$. No matter what happens after, the number will be of the form $0.3a_1a_2\ldots$ which is between 0 and 0.5. On the other hand, if $X = [0, 0.05]$, then player II will win: player I has to play $a_0 = 0$, and then II plays $a_1 = 6$. These two examples are somewhat silly, because who wins depend on who plays first.

For a more interesting game, allow player I to play as many digits as she wants (but still a finite number), then player II replies with one digit, and the game continues as before. Call this variation $G^{**}(X)$ (which again can be coded as $G(X^{**})$ for some X^{**}) In this variation, player I wins both games: she wins for $[0, 0.5]$ as before, and wins for $[0, 0.05]$ by playing 00 as the first move. Which closed sets are winning for II in that variation? Well, a silly example is the empty set. A more interesting example is $X = \{0, 1\}$: no matter what player I does, player II can by his next move make sure that there will be no way of ending inside X. What about a more complicated example, like $\{0\} \cup \{\frac{1}{n} \mid n = 1, 2, 3, \ldots\}$? It turns out player II can win this! Can you see how? In general, the following are true (but not so obvious). For a set $X \subseteq [0, 1]$:

(1) X is countable if and only if player II wins $G^{**}(X)$.
(2) X contains a non-empty perfect set (recall this means a set which is its own derivative) if and only if player I wins $G^{**}(X)$.

What do we mean here by "player I wins"? We mean that player I has a *winnning strategy*: a way to play that guarantees a win no matter how II plays. Call a set X *determined* if one of the two players has a winning strategy in the game for X. It is *not* obvious that any set is determined. In fact, this is false! Intuitively, even if player I can make sure "not to lose", this does not always mean she will win. However this *will* be true for closed sets (because, intuitively, closed sets are closed under taking limits so if one can avoid losing then the limit of the finite approximations will land in the set, hence one will win), and hence closed sets are determined. Putting all of this together, we have obtained a sketch[1] of a proof that either a closed set is countable or contains a perfect set. In fact, the above result tells us that any set which is determined has this property, hence satisfies the continuum hypothesis. In general, determined sets can be thought of as "simple" or "nice" and this makes it possible to prove they satisfy a number of "regularity" properties using games.

For example, the game above encodes the so-called "perfect set property" of being either countable or containing a perfect set. There are also games that

[1] Of course, this is only a very rough sketch of the main ideas. We will go go back to it in much more details later in class.

encode measurability or having the property of Baire, two important properties in analysis.

A hard result of Martin is that every Borel set is determined, and the topic of what sets beyond Borel are determined (up to making it an axiom, the *axiom of determinacy* contradicting the axiom of choice, that *all* sets are determined) is a deep topic of current research.

2. Sets and classes

We now start giving the axiom and language of set theory[2]. As explained already, the collection of all sets cannot be a set. However, it is still convenient to be able to talk about it. Thus we will really introduce two types of objects: classes and sets. Classes will be (intuitively) any collections of sets, and sets will be certain classes, intuitively those that are not too big. Note that we avoid Russell's paradox here by specifying that a class only contains sets, not arbitrary classes. Thus we will be unable to consider the class of all classes, though we will still be able to look at the class of all sets.

More precisely, the objects we will consider are called *classes*. For any two classes A and B, we can ask whether $A \in B$ (read A *is a member of* B, or A *is in* B, or A *is an element of* B). A *set* is, by definition, a class X that is a member of some other class. A *proper class* is a class that is not a set. We often use lowercase letter for sets, and uppercase letter for classes. For classes A and B, we say A is a *subclass* of B, written $A \subseteq B$, if for any set a, if $a \in A$ then $a \in B$. When in addition A is a set we say that A is a *subset* of B.

The following are the two most basic axioms of set (or really class) theory.

Axiom 2.1 (Axioms for classes)**.**

- (Extensionality) If A and B are classes, $A \subseteq B$, and $B \subseteq A$, then $A = B$.
- (Specification) If $P(x)$ is any property of sets, there is a class A whose members are exactly the sets a such that $P(a)$ is true.

The extensionality axiom says that classes are determined by their elements (for example, classes don't have colors or other attributes — their elements tell us everything we need to know).

The specification axiom (also called the *comprehension* axiom) requires some explanation. What is a property? We are leaving the exact meaning vague here — it could be made precise by using the language of formal logic. Intuitively, and for the purpose of this class, it is any precise statement that a set may or may not satisfy. For example, "x is not the empty set", "any class that is not the class of all sets contains x", or "x is a set which has exactly one member" are properties. A little bit more precisely, a property P involves a "variable" set x, and can be formed using:

- The relations $\in$ and $=$.
- Known and previously defined sets.
- Logical connectives, such as or, not, and, implies.

[2] More specifically, we will present what is known as *Morse-Kelley set theory*. It is very close to the (more standard but harder to work with) *Zermelo Frankel set theory with choice* (ZFC). Arguably, it seems there is not much difference between the two in practice: any practical mathematical statement that can be proven with one can be proven with the other.

- Quantifier over sets or over classes: "For all classes Y, ...", "For all sets y, ...".

Other properties that we will be able to make sense of once numbers have been defined are "x is a real number" or "x is a prime number". Thus the axiom of specification allows us to make any collection of sets that we can precisely describe into a class (that class may *not* be a set, however). Let us explore this a little bit further by introducing some standard notation:

Notation 2.2. If $P(x)$ is a property of sets, we write $\{a \mid P(a)\}$, read *the set of all a such that* $P(a)$, for the class described by the axiom of specification. Two alternate forms will also be used often:

- $\{a \in A \mid P(a)\}$ stands for $\{a \mid P(a) \text{ and } a \in A\}$.
- $\{E(a) \mid P(a)\}$ stands for $\{b \mid b = E(a) \text{ for some } a \text{ such that } P(a)\}$, here $E(a)$ is some expression involving a.

The following examples are informal, in the sense that we have not yet defined the sets of numbers involved. Nevertheless, they should be familiar from previous classes.

Example 2.3.

(1) $\{n \in \mathbb{Z} \mid n = 2m \text{ for some m } \in \mathbb{Z}\}$ is the set of integers.
(2) $\{n^2 \mid n \in \mathbb{Z}\}$ is the set of square integers.

Continuing the abstract development of set theory, the axiom of specification allows us to define several important classes, as well as ways to build old classes from new ones.

Definition 2.4.

- The *empty class* is the class $\{x \mid x \neq x\}$. It is denoted by $\emptyset$ (or sometimes, but rarely, by $\{\}$). The *class of all sets* is the class $\{x \mid x = x\}$. It is denoted by SET.
- For a class A, the *complement of* A, denoted A^c, is the class $\{x \mid x \notin A\}$.
- For two classes A and B, $A \cup B$, read A *union* B is the class $\{x \mid x \in A \text{ or } x \in B\}$.
- For a class C, $\bigcup C$, read *the union of all members of* C, is the class $\{x \mid \text{there exists } y \in C \text{ such that } x \in y\}$.
- For two classes A and B, $A \cap B$, read A *intersection* B is the class $\{x \mid x \in A \text{ and } x \in B\}$.
- For a class C, $\bigcap C$, read *the intersection of all members of* C is the class $\{x \mid \text{for all } y \in C,\ x \in y\}$.
- For two classes A and B, the *difference* $A - B$ (or $A \backslash B$) is defined to be $A \cap B^c$. The *symmetric difference* $A \boldsymbol{\Delta} B$ is defined to be $(A - B) \cup (B - A)$.

Notation 2.5. Given a set a, we denote by $\{a\}$ (read *singleton* a) the class $\{x \mid x = a\}$. Given two sets a and b, define the *unordered pair* $\{a, b\}$ to be the class $\{a\} \cup \{b\}$. Similarly for more than two sets.

Several easy properties (that you should be able to prove) follow from these definitions. For example, SET $= \emptyset^c$, $(A \cup B)^c = A^c \cap B^c$, $\emptyset \cup A = A$ for any class A, $\bigcup \emptyset = \emptyset$ and $\bigcap \emptyset =$ SET, $\bigcup\{a, b\} = a \cup b$ for two sets a and b, $\{a, a\} = \{a\}$, etc.

Note that $\emptyset$ is the unique class with no members, whereas SET contains all sets. Usually, $\emptyset$ is called the empty *set*. However from the two axioms given so far, it is not clear that $\emptyset$ is a set (i.e. that there is a class containing it). In fact, the axioms do not guarantee that there are any sets around. It could be that SET $= \emptyset$. We now introduce more axioms allowing us to build sets.

Axiom 2.6 (Basic set-building axioms)**.**

- (Empty set) $\emptyset$ is a set.
- (Subset) If b is a set and $a \subseteq b$, then a is a set.
- (Pairing) If a and b are sets, then $\{a, b\}$ is a set.
- (Union) If a is a set, then $\bigcup a$ is a set.
- (Power set) If b is a set, then the class $\{a \mid a \subseteq b\}$ is a set. We call it the *power set of* b, written $\mathcal{P}(b)$.

With these axioms, we can prove that more objects are set. For example:

Lemma 2.7.

(1) $\{\emptyset, \{\emptyset\}\}$ is a set.
(2) If a and b are sets, then $a \cup b$ and $a \cap b$ are sets.
(3) If a, b, and c are sets, then $\{a, b, c\}$ is a set.

Proof.

(1) By repeatedly applying the pairing and empty set axiom. Note that $\{\emptyset\} = \{\emptyset, \emptyset\}$.
(2) By pairing, $\{a, b\}$ is a set. By union, $a \cup b = \bigcup\{a, b\}$ is a set. For intersection, observe that $a \cap b \subseteq a$, so use the subset axiom.
(3) By pairing, $\{a\} = \{a, a\}$ is a set, and similarly $\{b\}$, $\{c\}$ are sets. Now by the previous part applied twice $\{a, b, c\} = \{a\} \cup \{b\} \cup \{c\}$ is a set.

□

On the other hand, some classes are too big to be sets:

Theorem 2.8 (Russell's paradox)**.** SET is a proper class (i.e. not a set).

Proof. Suppose for a contradiction that SET is a set. Note that $A \subseteq$ SET for any class A so by the subset axiom, any class is a set. Now consider the class $C = \{x \mid x \notin x\}$ of all sets that are not members of themselves. This class exists by the axiom of specification. By the previous observation, it is a set. Let us ask whether $C \in C$. If $C \in C$, then since C is a set, it must satisfy the defining property of C, so $C \notin C$, contradiction. If $C \notin C$, then again because it is a set, it must be a member of C, so $C \in C$, contradiction again. □

In practice, any class that can be defined explicitly from known sets using operations such as union, power set, subset, and pairing will be a set. It is only the "very big" and "nonconstructive" classes that are not sets. We will see more examples later. Unless asked explicitly, you do not have to justify why a given object is a set in this class — it is usually clear from its definition.

From now on, we will drop our convention and use uppercase letters also for sets.

2.1. **The natural numbers and the axiom of infinity.**

The axioms given so far allow us to build sets by starting with the empty set and iterating operations such as pairing, power sets, and unions. For example, we can build the sets $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$, $\{\{\{\emptyset\}\}\}$, etc. The class of all sets will contain all of those, but the axioms do not yet allow us to form a *set* containing them. In fact, the axioms do not even allow us to conclude the existence of any infinite set. Thus the next step is to add an axiom allowing us to build the natural numbers. It is natural to *define* the number 0 to be the empty set. More generally, a natural number will be identified with the set of its predecessors. For example, $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, and so on. To make sense of this "and so on", we define for *any* set x its successor, denoted $\mathrm{S}(x)$ (or $\mathrm{S}\,x$), by $\mathrm{S}(x) := x \cup \{x\}$ (note that this is a set). Thus $S(x)$ is the "number" whose predecessors are all the predecessors of x, plus x itself. You can check that $S(0) = 1$ and $S(1) = 2$. The natural numbers will then be built by collecting the sets obtained by iterating the successor operation from 0. More precisely it will have the following property:

Definition 2.9. A class A is *closed under successors* if for any $x \in A$, $S(x) \in A$.

For example, SET is closed under successors. Also, $\emptyset$ is closed under successors vacuously. We want to define $\mathbb{N}$ to be the smallest set closed under successors and containing 0. This can be done by taking the intersections of all such sets. However the axioms so far do not guarantee us that there is *any* such *set*. This is the content of the axiom of infinity:

Axiom 2.10 (Axiom of infinity)**.** There is a *set* that contains $\emptyset$ and is closed under successors.

Definition 2.11. Define the set of natural numbers $\mathbb{N}$ by:

$$\mathbb{N} := \bigcap \{X \mid \emptyset \in X \text{ and } X \text{ is closed under successors}\}$$

Note that $\mathbb{N}$ is indeed a set: by the axiom of infinity, there is X a set that contains $\emptyset$ and is closed under successors. By definition, $\mathbb{N} \subseteq X$, and by the subset axiom any subset of a set is a set.

The set $\mathbb{N}$ so-defined satisfies the principle of induction, in the following sense:

Theorem 2.12 (Induction for $\mathbb{N}$)**.** Assume that A is a subset of $\mathbb{N}$ satisfying the following two properties:

(1) $0 \in A$.
(2) If $n \in A$, then $S(n) \in A$.

Then $A = \mathbb{N}$.

Proof. The second condition is just saying that A is closed under successors. The first says that A contains 0. By definition, $\mathbb{N}$ is the intersection of all such sets, so in particular $\mathbb{N} \subseteq A$. By extensionality, $\mathbb{N} = A$. □

Later, we will study induction in a more general setup. The natural numbers $\mathbb{N}$ so-defined satisfy the *Peano axioms*:

- For all natural numbers n, m, if $\mathrm{S}\,n = \mathrm{S}\,m$, then $n = m$.
- There is no natural number n such that $\mathrm{S}\,n = 0$.
- Induction for $\mathbb{N}$ (as in Theorem 2.12): if A is a subset of natural numbers containing 0 and closed under successors, then $A = \mathbb{N}$.

We have already proved the third. To see the second, observe that $n \in \mathrm{S}\,n$ for any natural number n, and 0 is the empty set, so it has no member. For the first, assume $\mathrm{S}\,n = \mathrm{S}\,m$. Then $n \cup \{n\} = m \cup \{m\}$. Assume for a contradiction that $n \neq m$. Then the only way for the previous equality to be true is that $n \in m$ and $m \in n$. By problem 5b in assignment 1, $n \subseteq m$ and $m \subseteq n$, so $m = n$, contradiction.

We can also define the usual ordering on $\mathbb{N}$: indeed, $n < m$ in the usual sense exactly when n is a predecessor of m, i.e. $n \in m$. You will explore some of the properties of this relation in assignment 1.

Using just the Peano axioms, it is also possible to recursively define addition on $\mathbb{N}$: $n + 0 = 0$, and $\mathrm{S}(n + m) = n + \mathrm{S}\, m$. Making sense of this definition precisely requires a tool called the *recursion theorem*, to be proven later, but here we accept (as you did in previous math classes) that such definitions are permissible. We can then define the order in a different (but equivalent) way by $n < m$ if and only if there exists a nonzero k such that $n + k = m$. Unless explicitly mentioned, you can assume that addition, multiplication, and the ordering on $\mathbb{N}$ are defined and have the usual properties.

To study such definitions in more generality, it is convenient to study abstractly functions and relations.

2.2. Ordered pairs, cartesian product, relations, and functions.

So far, we have only considered *unordered pairs* $\{a, b\}$. We now define a notion of *ordered* pair. We do so not only for sets a and b but more generally for classes.

Definition 2.13. For classes A and B, the *ordered pair* (A, B) is defined to be the following class:

$$(A, B) := \{\{\{a\}, 0\} \mid a \in A\} \cup \{\{\{b\}, 1\} \mid b \in B\}$$

We prove two crucial results: A and B can indeed be recovered from the ordered pair (A, B), and moreover if A and B are sets, then the ordered pair is also a set.

Lemma 2.14.

(1) Assume A, B, C, and D are classes. If $(A, B) = (C, D)$, then $A = C$ and $B = D$.
(2) If a and b are sets, then (a, b) is a set.

Proof.

(1) Assume $(A, B) = (C, D)$. We show $A = C$ and the proof that $B = D$ is similar. We first show $A \subseteq C$. Assume $a \in A$. Then $\{\{a\}, 0\} \in (A, B) = (C, D)$. Thus either $\{\{a\}, 0\} = \{\{c\}, 0\}$ for some $c \in C$, or $\{\{a\}, 0\} = \{\{d\}, 1\}$ for some $d \in D$. In the second case, we know that $0 \neq 1$ so $\{d\} = 0 = \emptyset$, which is impossible (as $d \in \{d\}$). Thus we must be in the first case: $\{\{a\}, 0\} = \{\{c\}, 0\}$. Again, $\{c\} \neq 0$, so $\{a\} = \{c\}$, so $a = c$, so $a \in C$. Similarly, $C \subseteq A$, so $A = C$, as desired.
(2) The members of (a, b) are nested sets, but at the "bottom" of these nested sets is either a member of a, a member of b, 0, or 1. Thus let us set $X := a \cup b \cup \{0, 1\}$. X is a set by the union axiom. Now for $a_0 \in a$, $\{a_0\}$ is a subset of X, so a member of $\mathcal{P}(X)$. Similarly for $\{b_0\}$, where $b_0 \in b$. On the other hand, 0 and 1 are members of X, so we can let $Y = \mathcal{P}(X) \cup X$. Moving to the next level of nesting, for $a_0 \in a$, $\{\{a_0\}, 0\}$ is a subset of Y, hence a member of $\mathcal{P}(Y)$, and similarly for $\{\{b_0\}, 1\}$, $b_0 \in b$. In the end, (a, b) is therefore a subset of $\mathcal{P}(Y)$, so by the subset axiom a set.

□

Remark 2.15. From the pair $p = (A, C)$, we can recover the class A as follows: A is the class of sets x such that for *some* classes A', C', $x \in A'$ and $(A', C') = p$. We can proceed similarly to recover B.

Definition 2.16. For classes A, B, C, define the *ordered triple* (A, B, C) to be $(A, (B, C))$. Similarly define ordered 4-tuple and 5-tuples (we will see later how to precisely define n-tuples for any natural number n).

Definition 2.17. For classes A and B, we denote by $A \times B$ (the *cartesian product* of A and B) the class of all ordered pairs with left component from A and right component from B. Formally:

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

We may write $A \times B \times C$ instead of $A \times (B \times C)$.

Lemma 2.18. If A and B are sets, then $A \times B$ is a set.

Proof. Assignment 1. □

We can now define relations and functions. A possible minor difference compared to what you saw in a previous class is that we make the domain and codomain part of the function/relation. We also define function and relation for classes, although this will not be used too often.

Definition 2.19. A (binary) *class relation* is a triple $R = (A, B, \Gamma)$, where:

- A is a class, called the *domain* of R. We may write $\mathrm{dom}(R)$ for A.
- B is a class, called the *codomain* of R. We may write $\mathrm{cod}(R)$ for B.
- Γ is a subclass of $A \times B$, called the *graph* of R. We may write $\Gamma(R)$ for Γ.

When A and B are both sets, we call R a *(set) relation.* We say that R is *from A to B*. If $A = B$, we say that R is *on* A. For $a \in A$, $b \in B$, we write aRb for $(a, b) \in \Gamma$.

Definition 2.20. A *(class) function* is a (class) relation F such that for any $a \in \mathrm{dom}(F)$ there exists a unique $b \in \mathrm{cod}(F)$ so that aFb. We write $f(a)$ for this unique b. We write $F : A \to B$ to indicate that F is a (class) function from A to B. We also may call a class function a *map.*

Example 2.21.

1. The relation R on $\mathbb{N}$ (so the domain and codomain is $\mathbb{N}$) defined by nRm if and only if $n \in m$ is the usual ordering on $\mathbb{N}$, typically written $R =<_{\mathbb{N}}$, or just "$<$ on $\mathbb{N}$". More formally, $\Gamma(R) = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid n \in m\}$.
2. More generally, for any class A, membership ($\in$) induces a relation $\in_A$ on A defined by $x \in_A y$ if and only if $x \in y$.
3. Define a function $f : \mathbb{N} \to \mathbb{N} - \{0\}$ by $f(n) = \mathrm{S}(\mathrm{S}(n))$. This is the usual "$n + 2$" function. The definition is shorthand for: $\mathrm{dom}(f) = \mathbb{N}$, $\mathrm{cod}(f) = \mathbb{N} - \{0\}$, and $\Gamma(f) = \{(n, \mathrm{S}(\mathrm{S}(n))) \mid n \in \mathbb{N}\}$.
4. For any class A, there is an empty relation R on A which has A as domain and codomain, and the empty graph. There is also a full relation R' which has also A as domain and codomain, and graph $A \times A$. If $A = \emptyset$, they coincide. Neither are functions, unless $A = \emptyset$ or (for R') if $A = \{x\}$ for some set x.
5. On any class A, there is a class function $\mathrm{id}_A : A \to A$ defined by $\mathrm{id}_A(a) = a$. It is called the *identity function on A.* This can also be thought of as the *equality relation on A*, where two elements of A are related exactly when they are equal.

(6) More generally, for any classes $A \subseteq B$, we can define a class function $i_{A,B} : A \to B$ define by $i_{A,B}(a) = a$. It is called the *inclusion of A into B.* Unless $A = B$, we do *not* consider it the same as id_A (while their graphs and domains are the same, they have different codomains).

(7) Taking unions gives a class function from SET to SET. More precisely one can define $F : \mathrm{SET} \to \mathrm{SET}$ by $F(A) = \bigcup A$. Similarly for complement, intersection, and power set. We can also consider functions of two variables. For example, binary union can be thought of as a function $G : \mathrm{SET} \times \mathrm{SET} \to \mathrm{SET}$ defined by $G((A, B)) = A \cup B$. We usually write $G(A, B)$ instead of $G((A, B))$.

The following are operations on relations that you may be familiar with:

Definition 2.22. Let R be a class relation from A to B and S be a class relation from B to C.

- The *composition* $S \circ R$ is the relation T from A to C defined by aTc if and only if there exists $b \in B$ so that aRb and bSc.
- The *inverse* R^{-1} of R is the relation from B to A defined by $bR^{-1}a$ if and only if aRb.
- The *complement* $\not R$ of R is the relation from A to B defined by $a \not R\, b$ if and only if it is not true that aRb.
- For any class $A_0 \subseteq A$, the *image* of A_0 under R is defined by $R[A_0] := \{b \in B \mid aRb \text{ for some } a \in A_0\}$. The *range* of R, $\mathrm{ran}(R)$ is the image $R[A]$ of A under R.
- For any class A_0 (but usually $A_0 \subseteq A$), the *restriction* $R \upharpoonright A_0$ of R to A_0 is the relation with domain A_0, codomain B, and graph $\Gamma(R) \cap (A_0 \times B)$.
- For any class B' (but usually $B \subseteq B'$), the *corestriction* $R \upharpoonleft B'$ of R to B' is the relation with domain A, codomain B', and graph $\Gamma(R) \cap (A \times B')$.

Remark 2.23. If $f : A \to B$ is a class function and $A_0 \subseteq A$, then $f \upharpoonright A_0$ is also a function. Similarly, if $B \subseteq B'$, then $f \upharpoonleft B'$ is also a function.

The next definitions are properties of relations that you also have probably seen before:

Definition 2.24. Let R be a class relation on A.

- R is *reflexive* if aRa for any $a \in A$.
- R is *irreflexive* (or *antireflexive*) if $a \not R\, a$ for any $a \in A$.
- R is *symmetric* if aRb implies bRa for any $a, b \in A$.
- R is *antisymmetric* if aRb and bRa imply that $a = b$, for any $a, b \in A$.
- R is *transitive* if aRb and bRc imply aRc whenever $a, b, c \in A$.
- R is *total* if for any $a, b \in A$, either aRb or bRa.
- R is *trichotomous* if for any $a, b \in A$, either aRb, bRa, or $a = b$.
- R is an *equivalence relation* if it is reflexive, symmetric and transitive.

Definition 2.25.

- For a (class) equivalence relation E on A and $a \in A$, the *equivalence class* of a, denoted $[a]_E$ is the class of all $b \in A$ so that aEb.
- A *partition* P of a set A is a set of sets such that:
  (1) $\emptyset \notin P$.
  (2) $A, B \in P$, $A \neq B$ implies $A \cap B = \emptyset$.

(3) $\bigcup P = A$.

You will show the following standard result about equivalence relations in the assignments:

Theorem 2.26. If E is an equivalence relation on a set A, then $A/E := \{[a]_E \mid a \in A\}$ is a partition of A. Conversely, if P is a partition of the set A, then there is an equivalence relation E_P on A so that $A/E_P = P$.

Finally, recall the following basic properties of functions:

Definition 2.27. Let $F : A \to B$ be a class function.

- F is an *injection* (or *injective*) if for any $a_1, a_2 \in A$, $F(a_1) = F(a_2)$ implies $a_1 = a_2$.
- F is a *surjection* (or *surjective*, or *onto*) if for any $b \in B$, there exists $a \in A$ so that $F(a) = b$. In other words, the image $F[A]$ of A under F is equal to B.
- F is a *bijection* (or *bijective*) if it is both an injection and a surjection.

Most of these concepts will be reviewed in the exercises, and used substantially later.

2.3. **Sequences, indexed unions, and the axiom of replacement.** In the introduction, we mentioned sets of the form $\bigcup_{n\in\mathbb{N}} S_n$. How exactly do we make sense of this? We first need a way to describe sequences of sets. A simple way is to think of a sequence of sets as a function $F : \mathbb{N} \to \text{SET}$ (so $S_0 = F(0)$, $S_1 = F(1)$, etc.). This works fine, but what about sequence of classes? Remember that there is no class of all classes, so there would be no possible codomain for such a function. The following slightly more general definition does the trick:

Definition 2.28. A *class sequence* is a pair $S = (I, \Gamma)$, where:

- I is a class, called the *domain* or *index* of the sequence.
- Γ is a subclass of $I \times \text{SET}$.

Given S and $i \in I$, the *ith element* S_i of the sequence is defined to be the class $\{x \mid (i, x) \in \Gamma\}$. We usually suppress explicit mention of Γ, and just write $(S_i)_{i\in I}$ for a class sequence. When I is a set and each S_i is a set, we just call S a *sequence* (or a *set sequence*, or *sequence of sets*, for emphasis).

Note that sequences are very close to functions, but the difference is that they have no codomain. We adopt a couple of conventions to simplify our lives:

Notation 2.29 (Conventions regarding sequences)**.**

(1) When $(S_i)_{i\in I}$ is a sequence, we may not always precisely distinguish it from the corresponding class function $F : I \to \text{SET}$ given by $F(i) = S_i$. If context tells us that all the S_i's are part of a common set (for example $S_i \in \mathbb{N}$ for all $I \in I$, i.e. we have a sequence of natural numbers), then we may also identify the sequence with the corresponding function $F : I \to \mathbb{N}$.
(2) We may also not always distinguish between the ordered pair (a, b) and the sequence $(a_i)_{i\in\{0,1\}}$ given by $a_0 = a$, $a_1 = b$. Similarly for longer tuples.

We now define indexed unions and intersections:

Definition 2.30. If $(S_i)_{i\in I}$ is a class sequence, $\bigcup_{i\in I} S_i$ is the class $\{x \mid x \in S_i \text{ for some } i \in I\}$. Similarly, $\bigcap_{i\in I} S_i = \{x \mid x \in S_i \text{ for all } i \in I\}$.

Remark 2.31. If each S_i is a set (but I is not necessarily a set), then $\bigcup_{i \in I} S_i = \bigcup\{S_i \mid i \in I\}$. Going the other way, we can always describe a union of the form $\bigcup X$ using the indexed union notation, because $\bigcup X = \bigcup_{a \in X} a$ (the class sequence involved here has domain X and graph $\Gamma = \{(a, a) \mid a \in X\}$).

As a concrete example, consider the $\mathbb{N}$-indexed sequence $S_0 = \mathbb{N}$, $S_1 = \mathcal{P}(\mathbb{N})$, $S_2 = \mathcal{P}(\mathcal{P}(\mathbb{N}))$, and so on (we will learn how to make this "and so on" precise later). Let $X := \bigcup_{n \in \mathbb{N}} S_n$. The class X is very big: it contains all subsets of $\mathbb{N}$, all functions from $\mathbb{N}$ to $\mathbb{N}$ (which we can use to encode real numbers), all sets of functions from $\mathbb{N}$ to $\mathbb{N}$, etc. In fact, most familiar objects of mathematics are already inside X. We would like to know that X is a set, but none of the axioms introduced so far suffice for this. One may want to say that this follows from the union axiom because $X = \bigcup\{S_n \mid n \in \mathbb{N}\}$, but what tells us that $\{S_n \mid n \in \mathbb{N}\}$ is a set? This will be a consequence of the axiom of replacement:

Axiom 2.32 (Replacement)**.** If F is a class function whose domain is a set, then the range of F is a set.

Lemma 2.33. If $S = (S_i)_{i \in I}$ is a sequence of sets, then $\bigcup_{i \in I} S_i$ is a set. In fact, S itself is a set.

Proof. Consider the class function $F : I \to \text{SET}$ given by $F(i) = S_i$. The domain of F is I which by assumption is a set, and the range of F is exactly the class $\{S_i \mid i \in I\}$. By the axiom of replacement this is a set. By the union axiom, $\bigcup_{i \in I} S_i = \bigcup\{S_i \mid i \in I\}$ is a set. To see that S itself is a set, recall that $S = (I, \Gamma)$, where I is (by assumption) a set, so it suffices to show that Γ is a set. This is because Γ is a subclass of $I \times (\bigcup_{i \in I} S_i)$, and we have just argued both components of this cartesian product are sets. □

2.4. **Infinite products and the axiom of choice.** We now generalize the definition of the cartesian product:

Definition 2.34. Let $(S_i)_{i \in I}$ be a class sequence, with I a set. The *product* $\prod_{i \in I} S_i$, is the class:

$$\{x \mid x \text{ is a sequence with domain } I \text{ and } x_i \in S_i \text{ for all } i \in I\}$$

Note that the cartesian product $A \times B$ is (essentially) the product $\prod_{i \in \{0,1\}} A_i$, where $A_0 = A$, $A_1 = B$. There are two trivial cases: if $I = \{x\}$, then $\prod_{i \in I} S_i$ is essentially just S_x itself, while if $I = \emptyset$, $\prod_{i \in I} S_i$ contains exactly one element, the empty sequence (formally described by the pair $(\emptyset, = \mathit{emptyset}$: both its domain and its graph are the empty set).

The product is closely related to the following simpler class:

Definition 2.35. For sets B and A, let ${}^{B}A$ denote the class of all functions from B to A.

Lemma 2.36. For any sets B and A, ${}^{B}A$ is a set.

Proof. Assignment. □

Lemma 2.37. Let $(S_i)_{i \in I}$ be a sequence of sets. Let $A := \bigcup_{i \in I} S_i$. There is a class bijection between $\{f \in {}^{I}A \mid f(i) \in S_i \text{ for all } i \in I\}$ and $\prod_{i \in I} S_i$.

Proof. Given a function $f : I \to A$ such that $f(i) \in S_i$ for all $i \in I$, the sequence $(f(i))_{i \in I}$ is an element of $\prod_{i \in I} A_i$. Going the other direction, an element x of $\prod_{i \in I} A_i$ gives rise to a function $f : I \to A$ given by $f(i) = x_i$. Check that this induces a bijection. □

Given this lemma and the naturality of the bijection, we will often not strictly distinguish between $\prod_{i \in I} A_i$ and the set of functions f from I to $\bigcup_{i \in I} A_i$ such that $f(i) \in A_i$ for all $i \in I$. In particular, if $A_i = A$ for all $i \in I$, we will often not distinguish between ${}^I A$ and $\prod_{i \in I} A_i$.

We now deduce that products of sets are always sets (and so the map of Lemma 2.37 is really a set function):

Lemma 2.38. If $(S_i)_{i \in I}$ is a sequence of sets, then $\prod_{i \in I} S_i$ is a set.

Proof. First, let $A := \bigcup_{i \in I} S_i$. This is a set (Lemma 2.33). Therefore ${}^I A$ is a set (Lemma 2.36). Let F be the class bijection given by Lemma 2.37. Its domain is a subclass of ${}^I A$, hence a set by the subset axiom. By the axiom of replacement, its range $\prod_{i \in I} S_i$ must be a set. □

There is one annoying question left to discuss: it is easy to check that if $S_i = \emptyset$ for some $i \in I$, then $\prod_{i \in I} S_i = \emptyset$. Conversely, if $S_i \neq \emptyset$ for all $i \in I$, then we would like to conclude that $\prod_{i \in I} S_i \neq \emptyset$. This seems obvious: choose $x_i \in S_i$ for each $i \in I$ (possible as S_i is not empty), and then the corresponding sequence x is in $\prod_{i \in I} S_i$. However it turns out there is no way to formalize this infinite choice argument. It is always possible to choose an element from a fixed non-empty class S, and if we have a second class S', we can always choose an element from it as well. However when we have infinitely-many classes problems start appearing. Thus we make it into an axiom that the required choice can indeed be performed. In fact, we will require that there is a function from SET to SET picking an element from any non-empty set:

Axiom 2.39 (Choice)**.** There exists a class function $F : \text{SET} \to \text{SET}$ such that $F(a) \in a$ for all non-empty sets a.

We call an F as in the statement of the axiom of choice a *global choice function.* In fact, the axiom of choice as stated here is sometimes called the *axiom of global choice*, and the term *axiom of choice* is reserved for what we may call the *axiom of local choice*: for any non-empty set X there is a function $f : \mathcal{P}(X) \to X$ such that $F(a) \in a$ for all non-empty subsets a of X.

Lemma 2.40. If $(S_i)_{i \in I}$ is a sequence of sets and $S_i \neq \emptyset$ for all $i \in I$, then $\prod_{i \in I} S_i \neq \emptyset$.

Proof. Fix a global choice function $F : \text{SET} \to \text{SET}$. Define a sequence $x = (x_i)_{i \in I}$ by $x_i = F(S_i)$. Then $x \in \prod_{i \in I} S_i$. □

We will see that the axiom of choice has several important fundamental consequences, including that every set (in fact every class) can be well-ordered. It also implies several apparent "paradoxes": not all sets are determined, not all subsets of reals are measurable, and even the Banach-Tarski paradox (there is a decomposition of the unit ball into finitely-many pieces so that moving and rotating these pieces around produces two disjoint unit balls). Nevertheless, it can also be used to prove strong theorems (e.g. every vector spaces has a basis). It is now well established as an axiom for mathematics, and we will use it freely.

For completeness, we state the remaining axiom for set theory, although we will only discuss it a bit later. There is no need to understand it yet.

Axiom 2.41 (Foundation)**.** For any non-empty class A, there is $a \in A$ such that $a \cap A = \emptyset$.

For convenience, we have collected all the axioms for set theory in Appendix A.

The axiom of choice in action: the prisonners puzzle. Just for fun, and as an example of a simple somewhat paradoxical result involving the axiom of choice, we give the following puzzle.

Infinitely-many prisonners, indexed by natural numbers, will be put inside a room. Each of the prisonner will have a hat, either red or blue. Prisonners will be able to see the color of other hats but they won't see their own hat. The prisonners will be asked to all guess the color of their own hat at the same time. While in the room, they will not be allowed to communicate in any way. However they are allowed to discuss a strategy before entering the room.

We show that there is a way that all but finitely many prisonners can guess correctly.

For this, think of the prisonners as members of the set $\mathbb{N}$. An assignment of hat can be thought of as a function assigning a color to each member of $\mathbb{N}$. Thinking of blue as 0 and red as 1, this means each assignment is a function $f : \mathbb{N} \to 2$, i.e. $f \in {}^{\mathbb{N}}2$. Consider the binary relation E on ${}^{\mathbb{N}}2$ defined as follows: fEg if only if $\{n \in \mathbb{N} \mid f(n) \neq g(n)\}$ is finite. In other words, fEg if and only if f and g agree everywhere except for finitely-many places. Observe that E is an equivalence relation (exercise). Let F be the class function given by the axiom of choice. This function allows us in particular to pick a representative out of each E-equivalence class. The prisonners agree on F before entering the room.

Assume now the prisonners enter the room, and let f be the assignment of hats. Prisonner number n will know $f(m)$ for any $m \neq n$. Therefore he or she will know $[f]_E$ (but not f). Thus $[f]_E$ will be known by all prisonners. Now let $g := F([f]_E)$, and have prisonner n guess the hat color $g(n)$. Since gEf, all but finitely-many prisonners will be correct.

3. Defining, and counting, sets of numbers

Using induction (or more precisely recursion, to be justified formally later), one can define the usual structure on the set of natural numbers $\mathbb{N}$: addition ($m+0 = m$, $m + \mathrm{S}\, n = \mathrm{S}(m + n)$), multiplication ($m \cdot 0 = 0$, $m \cdot \mathrm{S}\, n = m \cdot n + m$), and the ordering ($n < m$ if and only if $n + k = m$ for some nonzero k). Another approach to defining the ordering is to set $n < m$ if and only if $n \in m$, and this turns out to be an equivalent definition. The following are key properties of the relation $<$ on $\mathbb{N}$ (which we take as a given, but you should be able to prove them using induction):

Fact 3.1. The ordering $<$ on $\mathbb{N}$ is a strict linear order: an irreflexive, transitive, and trichotomous relation. Moreover, if A is a non-empty subset of $\mathbb{N}$, then A has a minimal element: an element $a \in A$ so that $a \leq a'$ for all $a' \in A$.

Formally, addition is the unique function $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ that satisfies the properties set out in its recursive definition (but we write $m + n$ instead of the awkward $+(m,n)$). We will take for granted the usual properties of addition (they can all be proven by induction — you will look at commutativity in assignment 1). For example:

Fact 3.2. Addition and multiplication have the following properties, for all $n, m, k \in \mathbb{N}$:

- (Associativity) $n + (m + k) = (n + m) + k$, and similarly for multiplication.
- (Commutativity) $n + m = m + n$, and similarly for multiplication.
- (Cancellation) $n+m = k+m$ implies $n = k$, and similarly for multiplication (if $m \neq 0$).
- (Preservation of order) If $n \leq k$, then $n + m \leq k + m$, and similarly for multiplication.
- (Distributivity) $(n + m)k = nk + mk$.

We emphasize that these properties are not *hard* to prove, but they are just long and tedious exercises in the induction axiom. See the book by Hrbacek and Jech, or Edmund Landau's *Foundation of analysis* (which goes all the way to the complex numbers)[3].

3.1. **The integers.** Let us now construct the integers. For this we look at pairs of natural numbers (a, b), supposed to represent the "number" $a - b$. We have to identify two pairs (a, b), (c, d) if $a - b = c - d$. We do not yet have subtraction available, but we know this equation should be equivalent to $a + d = b + c$. Thus we define:

Definition 3.3. Define a relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by $(a, b) \sim (c, d)$ if and only if $a + d = b + c$.

Lemma 3.4. $\sim$ is an equivalence relation.

Proof. Reflexivity and symmetry are clear. We prove transitivity. Assume $(a, b) \sim (c, d) \sim (e, f)$. Then $a + d = b + c$ and $c + f = d + e$. Adding f to both sides of the first equation, $a + d + f = b + c + f$. By the second equation, $b + c + f = b + d + e$. Thus $a + f + d = b + e + d$. By cancellation, $a + f = b + e$, so $(a, b) \sim (e, f)$, as desired. □

Definition 3.5.

- We define the set $\mathbb{Z}^*$ by $\mathbb{Z}^* := (\mathbb{N} \times \mathbb{N})/\sim$.
- We define an ordering $\leq$ on $\mathbb{Z}^*$ as follows: $[(a, b)] \leq [(c, d)]$ if and only if $a + d \leq b + c$.
- We define addition on $\mathbb{Z}^*$ as follows: $[(a, b)] + [(c, d)] = [(a + c, b + d)]$.
- We define multiplication on $\mathbb{Z}^*$ as follows: $[(a, b)] \cdot [(c, d)] = [(ac+bd, ad+bc)]$.
- We define the *negative* of a member $[(a, b)]$ of $\mathbb{Z}^*$ to be $-[(a, b)] = [(b, a)]$. We then define $a - b$ to be $a + (-b)$.

Again, it has to be checked that these definitions make sense (i.e. that they do not depend on the choice of representative of the equivalence class), and that the operations behave as expected. This can be done. For example, it is easy to check

[3] A famous passage in the "Preface to the student" reads: "The multiplication table is not to be found in this book, not even the theorem $2 \cdot 2 = 4$; but I would recommend, as an exercise [...] that you make the following definitions: $2 = 1 + 1$, $4 = (1 + (1 + (1 + 1)))$ and then prove the theorem."

that $a + (-a) = [(0,0)]$ for any $a \in \mathbb{Z}^*$. Note also that if we define $\mathbb{N}^*$ to be the set $\{[(a,0)] \mid a \in \mathbb{N}\}$, then $\mathbb{N}^*$ "behaves like $\mathbb{N}$" inside $\mathbb{Z}^*$. More precisely, define a function $f : \mathbb{N} \to \mathbb{Z}^*$ by $f(n) = [(n,0)]$. Then f has the following properties:

(1) It is an injection (and its image is $\mathbb{N}^*$).
(2) If $n < m$, then $f(n) < f(m)$.
(3) $f(n+m) = f(n) + f(m)$.
(4) $f(nm) = f(n)f(m)$.

This means that f preserves the structure of $\mathbb{N}$, and *embeds* it into $\mathbb{Z}^*$. The "copy" of $\mathbb{N}$ inside $\mathbb{Z}^*$ is $\mathbb{N}^*$, and we say that f is an isomorphism from $\mathbb{N}$ into $\mathbb{N}^*$. Thus if we had started out by defining the natural numbers to be $\mathbb{N}^*$, we would not have lost anything, since we can always use f to translate between $\mathbb{N}$ and $\mathbb{N}^*$. Thinking about it differently, we can now define the integers to be the set $\mathbb{Z} := \mathbb{N} \cup (\mathbb{Z}^* - \mathbb{N}^*)$, with the operations appropriately translated (this is just to have that $\mathbb{N}$ is a subset of $\mathbb{Z}$, but usually one forgets this detail and just identifies $\mathbb{N}$ and $\mathbb{N}^*$).

In fact, there is no real need to remember the exact definition of $\mathbb{Z}$ in terms of equivalence classes: all one needs to remember is that $\mathbb{Z}$ is a set with some operations on it satisfying some properties. In fact, we will later characterize abstractly the ordering of $\mathbb{Z}$.

3.2. **The rationals.** Similarly, to define the rationals $\mathbb{Q}$, we look at the ordered pairs (a,b) of integers with $b \neq 0$, meant to represent the number $\frac{a}{b}$. We identify pairs that represent the same fraction: $(a,b) \sim^* (c,d)$ if and only if $ad = bc$. We then can proceed in a similar fashion as before. Again, for the details see Landau's book (reference on the course website).

3.3. **The reals.** There are several ways to define the reals: one approach is via Cauchy sequences. Another goes via Dedekind cuts. A construction can be found in Landau or in many analysis textbooks (e.g. Abbott's). In any case, from now on we assume the existence of the set of real numbers $\mathbb{R}$, as well as an ordering on $\mathbb{R}$, operations like addition, multiplication, subtraction, division, etc. A brief review of some of what we will assume known about $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ is in the appendix.

3.4. **Countable and uncountable sets.** We finish with a brief review of countability. A more powerful theory of infinite sizes will be developped later, here we just want to prove some basic facts that will help us study the integers and the rationals. Intuitively, countable sets are those that have size "at most" that of the natural numbers, and finite sets are those that have size "strictly less" than that of the natural numbers.

Definition 3.6. A set X is *infinite* if there is an injection from $\mathbb{N}$ into X. It is called finite otherwise.

Definition 3.7. A set X is *countable* if it is either empty or there is a surjection from $\mathbb{N}$ onto X. A set that is not countable is called *uncountable*. A set that is infinite and countable is called *countably infinite*.

Note that some finite sets like $\emptyset$ or $\{\emptyset\}$ are countable by definition (with a little bit of work, we can show that all finite sets are countable, but we don't need this for now). The following are very basic properties of countable sets that we will use without comments:

Lemma 3.8.

(1) $\mathbb{N}$ is countable.
(2) Any subset of a countable set is countable.
(3) If A is countable, B is a set, and there is a surjection from A onto B, then B is countable.

Proof.

(1) Because the identity function $\mathrm{id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ is a surjection (in fact a bijection).
(2) Assume B is countable, as witnessed by a surjection $f : \mathbb{N} \to B$. Let $A \subseteq B$. If A is empty, then it is countable by definition so assume that A is not empty. We define a surjection $g : \mathbb{N} \to A$ as follows: let $S := f^{-1}[A]$ be the inverse image of A (that is, $S = \{n \in \mathbb{N} \mid f(n) \in A\}$). Since A is not empty, one can pick $a \in A$. Define $g : \mathbb{N} \to A$ as follows: $g(n) = f(n)$ if $n \in S$, and $g(n) = a$ otherwise. Then g is a surjection, as $g[\mathbb{N}] \supseteq g[S] = f[S] = A$.
(3) Assume A is countable and let $g : A \to B$ be a surjection. If $A = \emptyset$, then since g is a surjection the only possibility is that $B = \emptyset$, so B is countable. Assume now that $A \neq \emptyset$. Since A is countable, there is a surjection $f : \mathbb{N} \to A$. Then $g \circ f : \mathbb{N} \to B$ is a surjection, so B is countable.

□

To prove that more sets are countable, recall that for $n \in \mathbb{N}$ and X a set ${}^{n}X$ is the set of all functions from n into X (recall we are thinking of n as the set $\{0, 1, 2, \ldots, n-1\}$). For example, if $n = 2$ then ${}^{2}X = {}^{\{0,1\}}X$. A function f from $\{0, 1\}$ to X is essentially just a pair $(f(0), f(1))$. Similarly, a member of ${}^{n}X$ is essentially an n-tuple. In particular, if $n = 1$, ${}^{1}X$ is essentially just X. It may be fun to think of the case $n = 0$.

We also define ${}^{<\mathbb{N}}X$, called the set of *finite sequences* of elements of X, to be $\bigcup_{n \in \mathbb{N}} {}^{n}X$. We will prove that if X is countable this set is countable. We start with a lemma:

Lemma 3.9. $\mathbb{N} \times \mathbb{N}$ is countable.

Proof. Recall that any nonzero natural number n factors as $n = 2^k m$, for m an odd number and $k \in \mathbb{N}$. Moreover k and m are unique. This is a consequence (which is easier to prove) of the fundamental theorem of arithmetic. Define a function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ as follows: $f(0) = (0, 0)$, and if n is positive, $n = 2^k m$ with m odd, $f(n) = (k, \frac{m-1}{2})$. It is easy to see that f is a surjection. Indeed, given a pair (a, b) in $\mathbb{N} \times \mathbb{N}$, we have that $f(n) = (a, b)$, where $n = 2^a(2b + 1)$. □

A similar proof, using prime factorization. Define a surjection $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ as follows: $f(0) = 0$ (or any other value), and if $n > 0$, let a be the maximal natural number so that 2^a divides n, and let b be the maximal natural number so that 3^b divides n. Set $f(n) = (a, b)$. This is a surjection: if $(a, b) \in \mathbb{N} \times \mathbb{N}$, then $f(2^a 3^b) = (a, b)$. □

Theorem 3.10.

(1) If $(A_n)_{n \in \mathbb{N}}$ is a sequence of countable sets, then $\bigcup_{n \in \mathbb{N}} A_n$ is countable.
(2) If A and B are countable, then $A \times B$, $A \cup B$, ${}^{n}A$ for all $n \in \mathbb{N}$, and ${}^{<\mathbb{N}}A$ are all countable.
(3) $\mathbb{Z}$ and $\mathbb{Q}$ are countable.

Proof.

(1) Let $A := \bigcup_{n \in \mathbb{N}} A_n$. For simplicity, we assume none of the A_n's are empty (otherwise just remove them from the sequence). For each $n \in \mathbb{N}$, fix a surjection $f_n : \mathbb{N} \to A$ (we are using the axiom of choice here, can you see how?). Define a function $f : \mathbb{N} \times \mathbb{N} \to A$ by $f(n, m) = f_n(m)$. This is a surjection: if $a \in A$, then $a \in A_n$ for some $n \in \mathbb{N}$, and so $a = f_n(m)$ for some $m \in \mathbb{N}$. Therefore $a = f(n, m)$, so a is in the range of f. Since we have shown earlier that $\mathbb{N} \times \mathbb{N}$ is countable, this establishes countability of A.

(2) Again, for simplicity we assume that A and B are not empty (otherwise the results are much easier to prove). Let $f : \mathbb{N} \to A$, $g : \mathbb{N} \to B$ be surjections. Then there is a surjection from $\mathbb{N} \times \mathbb{N}$ onto $A \times B$ given by sending (n, m) to $(f(n), g(m))$. Since $\mathbb{N} \times \mathbb{N}$ is countable, $A \times B$ is countable. Now $A \cup B$ is countable: this is just a special case of the previous part. We can establish that ${}^nA$ is countable by induction on n. If $n = 0$, this is obvious. Assume now that ${}^nA$ is countable and we want to establish that ${}^{n+1}A$ is countable. There is a bijection from $({}^nA \times A)$ to ${}^{n+1}A$ (exercise). The two component of the cartesian product are countable, so ${}^nA \times A$, and hence ${}^{n+1}A$, must also be countable. Finally, ${}^{<\omega}A = \bigcup_{n \in \mathbb{N}} {}^nA$ so it is countable by previous parts.

(3) The map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ defined by $f(n, m) = n - m$ is a surjection, so $\mathbb{Z}$ is countable. This in turn means that $\mathbb{Z} \times (\mathbb{Z} - \{0\})$ is countable. The map $f : \mathbb{Z} \times (\mathbb{Z} - \{0\}) \to \mathbb{Q}$ defined by $f(n, m) = \frac{n}{m}$ is a surjection, so $\mathbb{Q}$ is also countable.

□

So far, we have only shown that certain sets are countable. Are there uncountable sets? The answer is yes: this is Cantor's famous diagonal argument. We give a simple version here, showing that $\mathcal{P}(\mathbb{N})$ is uncountable, and more generally that $\mathcal{P}(X)$ is always "strictly bigger" than X. With more work (that we will do eventually), one can show that $\mathcal{P}(\mathbb{N})$ is in bijection with $\mathbb{R}$, and hence that $\mathbb{R}$ is uncountable. We will give another, completely different proof of uncountability of $\mathbb{R}$ in the next section.

Theorem 3.11 (Cantor's theorem)**.** For any set A, there is no surjection from A onto $\mathcal{P}(A)$. In particular, $\mathcal{P}(\mathbb{N})$ is uncountable.

Proof. Let $F : A \to \mathcal{P}(A)$ be any function. Let $B := \{a \in A \mid a \notin F(a)\}$. We show that B is not in the range of F, hence that F is not surjective. Indeed, suppose for a contradiction that $F(a) = B$ for some $a \in A$. There are two possibilities: either $a \in B$ or $a \notin B$. If $a \in B$, then by definition of B, $a \notin F(a) = B$, so this is impossible. On the other hand, if $a \notin B$, then $a \notin B = F(a)$, so by definition of B, $a \in B$, contradiction again. □

Where is the "diagonal" in this argument? You may want to try identifying sets with their characteristic functions, listing line by line the values of each characteristic functions, and think about how the set B in the proof above compares to the values in the diagonal. Details in class!

You should compare this proof to Russell's paradox (Theorem 2.8). In fact, Cantor's theorem gives another proof that SET is not a set (in fact, it's really the same proof but using Cantor's theorem as a lemma instead of proving a special case of it):

Corollary 3.12. SET is a proper class.

Proof. Assume for a contradiction that SET is a set. By the power set axiom, $\mathcal{P}(\text{SET})$ is then also a set. In fact, if you think about it, $\mathcal{P}(\text{SET}) = \text{SET}$ (the members of any set are sets themselves). Thus the identiy function is a surjection of SET onto $\mathcal{P}(\text{SET})$, contradicting Cantor's theorem. □

4. Linear orderings

In this section, we start studying a certain type of relation, the *(partial) orderings*. One goal is to abstractly characterize the ordering on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$.

Definition 4.1. A relation is an *ordering* (or a *partial order*) if it is transitive, antisymmetric, and reflexive.

Note that we can alternatively describe an ordering by its *strict* ordering $<$. We make this precise now:

Definition 4.2. A relation is a *strict ordering* (or *strict partial order*) if it is transitive and[4] *irreflexive*.

If R is an ordering on A, we can get a strict ordering S by setting aSb if and only if aRb and $a \neq b$. Conversely if S is a strict ordering on A, we get an ordering R by setting aRb if and only if $a = b$ or aSb. Thus the strict and non-strict ways of describing the order are interchangeable, and we will take advantage of this. We will often denote an ordering by $(A, \leq)$ (or $(A, <)$ if it is a strict ordering), where A is the set and $\leq$ the relation. We may also sometimes call $(A, \leq)$ a *partially ordered set*, or *poset*.

Definition 4.3. Two elements a, b of an ordering $(A, \leq)$ are *comparable* if $a \leq b$ or $b \leq a$. They are *incomparable* if they are not comparable. An ordering where any two elements are comparable is called a *total (or linear) order*. When the ordering is strict, we may talk about a strict linear order (alternatively, a *strict linear order* is a relation that is transitive, antisymmetric, irreflexive, and trichotomous). A subset B of A in which any two elements are comparable is called a *chain*.

Example 4.4.

(1) $(\mathbb{N}, <)$ is a (strict) linear order, where $n < m$ if and only if $n \in m$ (or equivalently if and only if there is $k \in \mathbb{N}$ so that $n + k = m$). You should be able to prove this, but can take it as granted. Similarly, $(\mathbb{Z}, <)$, $(\mathbb{Q}, <)$, and $(\mathbb{R}, <)$ are strict linear orders (where $<$ each time means the expected order).

(2) For any set A (typically, $A = \mathcal{P}(B)$ for a set B), $(A, \subseteq)$ is a partial order, where $\subseteq$ is the usual subset relation (of course restricted to A: more precisely the partial order is the relation R on the class A whose graph is $\{(x, y) \in A \times A \mid x \subseteq y\}$. The order may not be linear (consider for example $\mathcal{P}(\{0, 1\}, \subseteq)$: $\{1\}$ and $\{0\}$ are *incomparable* in that ordering.

[4] Antisymmetry follows from transitivity and irreflexivity: if $aRbRa$, then aRa by transitivity, which is impossible by irreflexivity.

(3) For a set A the relation $\in_A$ on A (usually written $(A, \in)$) defined by $a \in_A b$ if and only if $a \in b$ may or may not be a strict partial order. For example, if $A = \mathbb{N}$ then it is a strict partial order, but if $A = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$, then (A, ϵ) is not transitive.

(4) $(\mathbb{N}, |)$ is a partial order, where $|$ is the divisibility relation: n divides m if there exists a natural number k so that $kn = m$. In that order, 2 and 3 are incomparable, but 2 is below 4. Any number divides 0, and 1 divides any number. The set $\{1, 2, 4, 8, 16, ...\}$ of powers of two is a chain in that partial order.

Fundamental is what it means to embed a partially ordered set into another (we will focus on linear orders for now).

Definition 4.5. An *order embedding* between two posets $(A, \leq_A)$ and $(B, \leq_B)$ is an injection $f : A \to B$ such that for any $a_1, a_2 \in A$, $a_1 \leq_A a_2$ if and only if $f(a_1) \leq_A f(a_2)$. If in addition f is a bijection, then f is called an (order) *isomorphism*, the two posets are said to be *isomorphic*, and we write $(A, \leq_A) \cong (B, \leq_B)$.

For linear order, the definitions simplify as follows:

Lemma 4.6. Let $(A, \leq_A)$ and $(B, \leq_B)$ be linear orders and $f : A \to B$ be a function. If f is *order-preserving*, in the sense that $a <_A a'$ in A imply $f(a) <_A f(a')$, then f is an order embedding. If in addition f is a surjection, then it is an isomorphism.

Proof. Exercise. □

We will often suppress mention of the order if they are clear from context. For example, we may say "the linear order $\mathbb{N}$", where it is understood that we mean the usual order on $\mathbb{N}$.

Intuitively, two partially ordered sets that are isomorphic "look the same up to renaming". On the other hand if $(A, \leq_A)$ embeds into $(B, \leq_B)$, then there is a part of B that looks like A.

Example 4.7.

(1) The function $f : \{1, 2\} \to \{3, 4, 5\}$ given by $f(1) = 3$, $f(2) = 5$ is an order-embedding between $\{1, 2\}$ and $\{3, 4, 5\}$, with the usual orderings. It is not an isomorphism as it is not bijective. On the other hand the corestriction of f to codomain $\{3, 5\}$ *is* an isomorphism.

(2) The function $f : \mathbb{N} \to \mathbb{N} - \{0\}$ given by $f(n) = n + 1$ is an isomorphism.

(3) The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = e^x$ is an order embedding, and its corestriction to the positive reals give an isomorphism from $(\mathbb{R}, <)$ to $((0, \infty), <)$.

(4) The function $f : (-\pi/2, \pi/2) \to \mathbb{R}$ given by $f(x) = \tan(x)$ is an isomorphism (!). Draw a picture!

(5) Using the previous two examples, we can deduce that for any $-\infty \leq a < b \leq \infty$, the open interval of reals (a, b) is isomorphic to $\mathbb{R}$ (exercise). We will see another proof of this shortly.

(6) The inclusion $f : \mathbb{N} \to \mathbb{Z}$ is an order embedding. We say that $(\mathbb{N}, \leq)$ is a *suborder* of $(\mathbb{Z}, \leq)$.

(7) $\mathbb{N}$ and $\mathbb{Z}$ are *not* isomorphic. Indeed, the former has a minimal element, 0, and the latter does not have a minimal element. In more details, suppose

for a contradiction there is an isomorphism $f : \mathbb{N} \cong \mathbb{Z}$. Let $m := f(0) - 1$. Since f is a bijection, there is $n \in \mathbb{N}$ so that $f(n) = m$. Note however that $f(n) = m < f(0)$, so $n < 0$, which is impossible.

(8) Similarly, $\mathbb{Q}$ is not isomorphic to either $\mathbb{N}$ or $\mathbb{Z}$ (exercise).

(9) *Any* partial order $(A, \leq)$ embeds into the partial order $(\mathcal{P}(A), \subseteq)$: define $f : A \to \mathcal{P}(A)$ by $f(a) = \{b \in A \mid b \leq a\}$.

Let us investigate the property that really sets appart $\mathbb{Q}$ from, say, $\mathbb{Z}$ (considered as orderings). The key is that any member x of $\mathbb{Z}$ has a successor (an element y so that $x < y$ but no element x' satisfies $x < x' < y$). This is not true for $\mathbb{Q}$: if q is a rational, then there is no rational immediately after q. In fact:

Lemma 4.8. If $r < q$ are rational numbers, then there is a rational number r' with $r < r' < q$.

Proof. Take $r' = \frac{r+q}{2}$. □

Orderings having that property are given a name:

Definition 4.9. A linear order $(A, \leq)$ is *dense* if for any $a < b$ in A, there exists $c \in A$ so that $a < c < b$. More generally, a subset A_0 of A is *dense in* $(A, \leq)$ if for any $a < b$ in A, there is $c \in A_0$ with $a < c < b$.

Thus a linear order $(A, \leq)$ is dense if A is dense in $(A, \leq)$. For example, $\mathbb{N}$ and $\mathbb{Z}$ are not dense (there are no integers between 0 and 1), while we have just shown that $\mathbb{Q}$ is dense. The same proof applies to show that $\mathbb{R}$ is dense, but in fact we have the following stronger fundamental property of $\mathbb{R}$ (see for example Abbott's book for a proof):

Fact 4.10. $\mathbb{Q}$ is dense in $\mathbb{R}$. That is, if $x < y$ are real numbers, then there is a *rational* $r \in \mathbb{Q}$ so that $x < r < y$.

The opposite of dense is the concept of a *discrete* linear order:

Definition 4.11. A linear order $(A, \leq)$ is *discrete* if for any $a \in A$:

- Either a is minimal or it has an *immediate predecessor* (an element $b < a$ so that there is no c with $b < c < a$).
- Either a is maximal or it has an *immediate successor* (an element $b > a$ so that there is no c with $b > c > a$).

For example, any finite linear order is discrete, and $\mathbb{N}$ and $\mathbb{Z}$ are discrete. What sets them appart is that $\mathbb{N}$ has a minimum and $\mathbb{Z}$ does not (it has *no endpoints*). There are however discrete orders that are countable, are not empty, have no endpoints, and are not isomorphic to $\mathbb{Z}$ (exercise). However the following characterization is true:

Theorem 4.12. Let $(A, \leq)$ be a discrete linear order such that for any $x \leq y$, $\{z \in A \mid x \leq z \leq y\}$ is finite.

(1) If A has a minimum but no maximum, then it is isomorphic to $\mathbb{N}$.

(2) If A has no endpoints, then it is isomorphic to $\mathbb{Z}$.

Proof. Exercise. □

Let us now work toward characterizing $\mathbb{Q}$. We have seen that $\mathbb{Q}$ is dense, and we also know that it has no endpoints. These properties are also satisfied by $\mathbb{R}$ and we know that $\mathbb{Q}$ is not isomorphic to $\mathbb{R}$ (because $\mathbb{Q}$ is countable but $\mathbb{R}$ is not). We will give another proof that $\mathbb{Q}$ is not isomorphic to $\mathbb{R}$ soon, but in any case we better restrict ourselves to *countable* dense linear orders without endpoints. A trivial case is left: the empty order. After we eliminate it, we obtain a characterization of $\mathbb{Q}$. The proof is due to Cantor and proceeds by what is called a back and forth argument.

Theorem 4.13 (The back and forth theorem)**.** Any two non-empty countable dense linear orders without endpoints are isomorphic. In particular, any such ordering is isomorphic to $\mathbb{Q}$.

Proof. Let $(A, \leq)$ and $(B, \leq)$ be two such orders. Since A is countable and not empty, we can fix a surjection $f : \mathbb{N} \to A$. Let $(a_n)_{n \in \mathbb{N}}$ be defined by $a_n = f(n)$. Similarly, let $(b_n)_{n \in \mathbb{N}}$ enumerate B. We will define recursively a sequence of functions $(f_n)_{n \in \mathbb{N}}$ such that for all $n \in \mathbb{N}$:

(1) There is $m \in \mathbb{N}$ so that $\text{dom}(f_n) \subseteq \{a_k \mid k < m\}$, $\text{cod}(f_n) \subseteq \{b_k \mid k < m\}$.
(2) $a_n \in \text{dom}(f_{n+1})$ and $b_n \in \text{cod}(f_{n+1})$.
(3) f_n is a bijection, and if $a < a'$ are in $\text{dom}(f_n)$, then $f_n(a) < f_n(a')$ (that is, f_n is an isomorphism between the corresponding restrictions of the order on A and B).
(4) For $k \leq n$, $\text{dom}(f_k) \subseteq \text{dom}(f_n)$ and $f_k = f_n \upharpoonright \text{dom}(f_k)$. That is, functions defined later extend the earlier functions.

Suppose for a moment that we have managed to build such a sequence. Then we can define $f : A \to B$ as follows: given $a \in A$, there is $n \in \mathbb{N}$ such that $a = a_n$. Pick the least such n. By assumption, $a = a_n \in \text{dom}(f_{n+1})$, so define $f(a)$ to be $f_{n+1}(a)$. To check that f is an isomorphism, we apply Lemma 4.6 and check that it is order-preserving and surjective. First, f is order-preserving: let $a < a'$ be in A. Pick $n, m \in \mathbb{N}$ least such that $a_n = a$, $a_m = a'$. Pick $k \in \mathbb{N}$ such that $n, m < k$. Then $a, a' \in \text{dom}(f_k)$, and $f(a) = f_{n+1}(a) = f_k(a) < f_k(a') = f_{m+1}(a') = f(a')$. Finally, f is surjective because for any $b \in B$, there is $n \in \mathbb{N}$ such that $b = b_n$ and so $b \in \text{cod}(f_{n+1})$. Thus $b = f_{n+1}(a)$ for some $a \in A$. Taking k minimal so that $a = a_k$, we have that $f(a) = f_{k+1}(a)$. Taking $m > k, n$, we get that $f(a) = f_m(a) = f_{n+1}(a) = b$, so b is in the range of f.

It remains to show how to build the sequence $(f_n)_{n \in \mathbb{N}}$. For the base case, we let f_0 be the unique function with empty domain and codomain. Assume now that f_n has been defined, for a fixed $n \in \mathbb{N}$. We define f_{n+1}. We first define a function g_n with domain $\text{dom}(f_n) \cup \{a_n\}$ so that $g \upharpoonright \text{dom}(f_n) = f_n$. If $a_n \in \text{dom}(f_n)$, there is nothing left to do: set $g_n = f_n$. Otherwise for $a \in \text{dom}(f_n)$, let $g_n(a) = f_n(a)$, and we are left to define $g_n(a_n)$. There are three cases:

- If $a_n > a$ for all $a \in \text{dom}(f_n)$, then since B has no endpoints we can pick $b \in B$ so that $b > b'$ for all $b' \in \text{cod}(f_n)$. We let $g_n(a) = b$.
- If $a_n < a$ for all $a \in \text{dom}(f_n)$, then similarly since B has no endpoints we can pick $b \in B$ so that $b < b'$ for all $b' \in \text{cod}(f_n)$. We then let $g_n(a) = b$.
- Otherwise, we can pick a maximal among the members of $\text{dom}(f_n)$ so that $a < a_n$ and a' minimal among the members of $\text{dom}(f_n)$ so that $a_n < a'$. Since B is dense, there exists $b \in B$ so that $f_n(a) < b < f_n(a')$. It is easy to check that setting $g_n(a_n) = b$ preserves the order.

This completes what is called the "forth" step. We now do the "back" step: if $b_n \in \text{cod}(g_n)$, there is nothing left do do: set $f_{n+1} = g_n$. Otherwise, for $a \in \text{dom}(g_n)$

set $f_{n+1}(a) = g_n(a)$, and we are left to find $a \in A$ to add to the domain of f_{n+1} (i.e. so we can let $\text{dom}(f_{n+1}) = \text{dom}(g_n) \cup \{a\}$ and define $f_{n+1}(a) = b_n$ — another way to see this is we want to find $f_{n+1}^{-1}(b_n)$). There are again three cases:

- If $b_n > b$ for all $b \in \text{cod}(g_n)$, then since A has no endpoints we can pick $a \in A$ so that $a > a'$ for all $a' \in \text{dom}(g_n)$. We let $f_{n+1}(a) = b_n$.
- If $b_n < b$ for all $b \in \text{cod}(f_n)$, then similarly we can choose a.
- Otherwise, pick b maximal among the members of $\text{cod}(g_n)$ so that $b < b_n$ and b' minimal among the members of $\text{cod}(g_n)$ so that $b_n < b'$. Since A is dense, there exists $a \in A$ so that $g_n^{-1}(b) < a < g_n^{-1}(b')$. We let $f_{n+1}(a) = b_n$.

□

It follows, for example, that $(0, 1) \cap \mathbb{Q}$ is isomorphic to $\mathbb{Q}$! Proving this directly is not so easy... For example using tangent (as in Example 4.7(4)) does not work because it is not true that the tangent of a rational number is a rational number.

To understand better what sets $\mathbb{Q}$ appart from $\mathbb{R}$, let us now talk a little bit about minimums in linearly ordered sets. Given $(B, \leq)$ linearly ordered and $A \subseteq B$, a *minimum* (or *least element*) of A is an element $a \in A$ such that $a \leq a'$ for all $a' \in A$. Dually define the concept of a *maximum* (or *greatest element*). More generally, a *lower bound* of A is any $b \in B$ such that $b \leq a'$ for all $a' \in A$. It is easy to check that:

- A minimum is unique, if it exists.
- A minimum of A is exactly a lower bound that is also a member in A.

There are cases in which no minimum exist, but still there is something very close: consider the open interval $(0, 1)$ in $\mathbb{R}$. It has no minimum, since 0 is not in the set, but 0 is a greatest element of the set of lower bounds. In general, we say that $b \in B$ is an *infimum of A* if b is a greatest element of the set of lower bounds of A. Dually define the concept of an *upper bound* and of a *supremum.* Again, one can prove:

- An infimum is unique, if it exists.
- Any minimum of A is an infimum, and an infimum of A is a minimum if and only if it is a member of A.

The following definition is what separates $\mathbb{R}$ from $\mathbb{Q}$.

Definition 4.14. A linearly ordered set $(B, \leq)$ is *complete* if for any non-empty set $A \subseteq B$, if A has a lower bound, then it has an infimum, and if A has an upper bound, then it has a supremum.

Example 4.15.

(1) Any finite linear order is complete, because any non-empty set there has a minimum and a maximum (proving this requires working by induction on cardinality, and we haven't defined cardinality yet, but we soon will).
(2) $\mathbb{N}$ is complete: any non-empty subset A has a minimum, if A is bounded above, it will also have a maximum (exercise).
(3) $\mathbb{Z}$ is also complete (exercise).
(4) $\mathbb{Q}$ is *not* complete: consider the set A of all rational numbers q such that $q < \sqrt{2}$. The set A is bounded above (by 2), but we know that $\sqrt{2}$ is not rational, so it follows that A has no supremum in $\mathbb{Q}$: if r was such a supremum, we would have $\sqrt{2} < r$, and it is a basic property of the

real numbers that between any two distinct reals there is a rational. In particular we can find r' a rational so that $\sqrt{2} < r' < r$, contradicting that r was minimal among all the upper bounds.

(5) A fundamental property of $\mathbb{R}$ is that it *is* complete.

We can now give a proof that $\mathbb{R}$ is uncountable. This was historically Cantor's first proof:

Corollary 4.16. $\mathbb{R}$ is uncountable.

Proof. If not, $\mathbb{R}$ is countable. We know also that $(\mathbb{R}, \leq)$ is not empty, dense, and without endpoints. By Theorem 4.13, $(\mathbb{R}, \leq)$ is isomorphic to $(\mathbb{Q}, \leq)$. This is a contradiction because $(\mathbb{Q}, \leq)$ is not complete, but $(\mathbb{R}, \leq)$ is. □

We finish by characterizing the ordering on $\mathbb{R}$:

Definition 4.17. A linear ordering $(A, \leq)$ is *separable* if there is a non-empty countable set $A_0 \subseteq A$ such that A_0 is dense in A.

We have seen that $\mathbb{R}$ is separable, as witnessed by $\mathbb{Q}$.

Theorem 4.18. Any linear ordering without endpoints that is complete and separable is isomorphic to $\mathbb{R}$.

Proof. Exercise. □

Corollary 4.19. Any open interval (like $(0, 1)$) is isomorphic to $\mathbb{R}$.

Proof. Immediate from the theorem, since any such interval is complete, separable (take the intersection of the rationals with it), and without endpoints. □

5. Wellfounded relations, induction, and recursion

In this section, we define a very general framework for induction arguments, and construction by recursion. Recall that we have already rigorously established (Theorem 2.12) how to do induction on the natural numbers. However you are probably familiar with the fact that one can also induct on $\mathbb{N} \cup \{-1\}$, or on the set of pairs of natural numbers. However one cannot run an induction argument on $\mathbb{Z}$, since $\mathbb{Z}$ has no minimal element to start the induction on. In fact, it is this feature — every set has a minimal element — that makes induction work. We could look at minimal elements defined with respect to some (partial) order, but we take an even more general approach and look at *any* class relation:

Definition 5.1. Let R be a class relation on a class B.

- Let $a, b \in B$. We say that a is an *R-predecessor* of b if aRb. We let $\text{pred}_R(b)$ denote the class of R-predecessors of b.
- Let A be a subclass of B An element $a \in B$ is *R-minimal in A* if $a \in A$ and any R-predecessor of a is not in A. When R is clear from context, we omit it and say that a is *minimal in A*, or *a minimal element of A*.
- R is *wellfounded* if any non-empty subclass A of B has an R-minimal element.
- If B is a set and R is wellfounded strict linear order, we call R a *wellordering.*

Example 5.2.

(1) The usual strict ordering on $\mathbb{N}$ is a wellordering. In fact, for any inger n, the usual ordering on $\{m \in \mathbb{Z} \mid m \geq n\}$ is a wellordering. On the other hand, $(\mathbb{Z}, <)$ is not a wellordering (because $\mathbb{Z}$ itself does not have a minimal element). Note however that $(\mathbb{N}, \leq)$ is technically *not* wellfounded, just because 0 is related to 0, so is not $\leq$-minimal in the sense given above.
(2) The set $\mathbb{Q}_{\geq 0}$ of nonnegative rationals with the usual order is not a wellordering, because $\{q \in \mathbb{Q}_{\geq 0} \mid q > 0\}$ has no minimal element.
(3) The strict divisibility partial order on $\mathbb{N}$ is wellfounded: suppose A is a non-empty subset of natural numbers. Take a minimal element a of A according to the usual ordering on $\mathbb{N}$. Then a is also a minimal element according to divisibility, because if a' divides a, and $a' \neq a$, then $a' < a$ so $a' \notin A$. Note however that A may not have a *unique* minimal element. For example, the set $\{2, 3\}$ has both 2 and 3 as minimums.
(4) The strict partial order $(\mathcal{P}(\mathbb{N}), \subsetneq)$ is not wellfounded: for $k \in \mathbb{N}$, let A_k be the set of natural numbers divisible by 2^k. Then $\{A_k \mid k \in \mathbb{N}\}$ has no minimum in $\mathcal{P}(\mathbb{N})$ (because A_{k+1} is a strict subset of A_k for each $k \in \mathbb{N}$.
(5) Consider the following partial ordering on $\mathbb{N} \times \mathbb{N}$: $(a, b) \leq (c, d)$ if and only if $a \leq c$ and $b \leq d$ (where $\leq$ is the usual ordering on $\mathbb{N}$). Then the strict version $<$ is wellfounded: try to prove this as an exercise.

An interesting question to ask is whether the relation $\in$ (say on the class SET) is wellfounded. For example, do there exists weird sequences $(a_n)_{n \in \mathbb{N}}$ of sets so that $a_{n+1} \in a_n$ for all $n \in \mathbb{N}$? More simply, does there exist a set x such that $x \in x$? Or even $\{x\} \in x$? Even if they potentially could exist, such sets do not seem very useful. Thus we explicitly rule them out with an axiom:

Axiom 5.3 (Foundation)**.** For any non-empty class A, there is $a \in A$ such that $A \cap a = \emptyset$.

Lemma 5.4. The axiom of foundation is equivalent (modulo the other axioms) to the statement that the ϵ relation on SET is wellfounded.

Proof. Assume the axiom of foundation. Let A be any non-empty class. Pick $a \in A$ as given by the axiom of foundation. Since $A \cap a = \emptyset$, we have in particular that whenever $a' \in a$, $a' \notin A$. Thus a is ϵ-minimal in A, as desired. This shows that the ϵ relation on SET is wellfounded. The converse is completely similar. □

From now on, we assume the axiom of foundation (although we will not use it for a long time). We now state the principle of induction for any wellfounded relation.

Theorem 5.5 (The induction theorem)**.** Let R be a wellfounded class relation on B and let A be a subclass of B. Assume that for any $b \in B$, if all the R-predecessors of b are in A, then b is also in A. Then we can conclude that $A = B$.

Proof. Assume for a contradiction that $A \neq B$ and let $B_0 := B - A$. By wellfoundedness, let b be R-minimal in B_0. For any R-predecessor b' of B, we know that $b' \notin B_0$ (because b was minimal), so $b' \in A$. Thus all the R-predecessors of b are in A, hence $b \in A$ by assumption. This is a contradiction because $A \cap B_0 = \emptyset$. □

To get more intuition for the statement of Theorem 5.5 let us apply it to the relation $<$ on $B = \mathbb{N}$. Let $A \subseteq \mathbb{N}$. Suppose we know that for any natural number n, if all the predecessors of n are in A, then n is also in A. The conclusion of Theorem 5.5 tells us that $A = \mathbb{N}$. This is essentially the statement of the strong induction principle for natural numbers: if we can prove that whenever a statement is true for the predecessors of n then it is true for n itself, then the statement is true of all natural numbers. Note that the base case is hidden here: vacuously, all the predecessors of 0 are in A. We will later apply Theorem 5.5 to do induction on certain wellorderings called ordinals.

For now, we examine a related topic: construction by recursion. We want to (finally) formalize constructions such as the following: define a function $f : \mathbb{N} \to \mathbb{N}$ recursively as follows: $f(0) = 0$, $f(1) = 1$, and $f(n+2) = f(n)+f(n+1)$ (giving the Fibonacci sequence). What is happening is that we are given a function F taking as input the "previous cases" and giving as output the result of putting these previous cases together. This is formalized as follows (we restrict ourselves to relations that are "not too big", called set-like):

Definition 5.6. Let R be a class relation on B.

- For any $b \in B$, recall that $\text{pred}_R(b)$ denote the class of all R-predecessors of b.
- A subclass A of B is an *R-initial segment* if it is closed under predecessors: for any $a \in A$, $\text{pred}_R(a) \subseteq A$. When R is clear from context, we omit it and just say that A is an initial segment.
- We say that R is *set-like* if $\text{pred}_R(b)$ is a set for any $b \in B$.

Example 5.7.

- $\{0, 1, 2\}$ is an initial segment of $(\mathbb{N}, <)$, but the set A of even numbers is not (because $1 < 2$, $2 \in A$, but $1 \notin A$).
- Any relation (on a set) is set-like. There are however other set-like relations. For example, the membership relation $\in$ on SET is set-like (exercise).

Theorem 5.8 (The recursion theorem)**.** Let R be a set-like wellfounded class relation on a class B. Let D be the class of pairs (s, b), where $b \in B$ and s is a sequence with domain $\text{pred}_R(b)$. Let $F : D \to \text{SET}$ be a class function. There exists a unique class function $f : B \to \text{SET}$ such that for any $b \in B$, $f(b) = F((f(a))_{a \in \text{pred}_R(b)}, b)$.

Before proving the recursion theorem, let us see how we could use it to build the Fibonacci sequence described above. We take R to be the ordering $<$ on $B = \mathbb{N}$. This is clearly set-like because B is a set. The function $F : D \to \text{SET}$ is defined as follows: given a sequence $s = (a_m)_{m \in \text{pred}_<(n)}$ and $n \in \mathbb{N}$, if the a_m's are not natural numbers, then we define $F(s, n)$ arbitrarily, e.g. $F(s, n) = 0$. Otherwise if $n = 0$, define $F(s, n) = 0$. Similarly, if $n = 1$, define $F(s, n) = 1$. If $n \geq 2$, define $F(s, n) = a_{n-2} + a_{n-1}$. Usually, these details are tedious to write down and we omit them, but it is important to understand how the recursion theorem makes all of this work.

Proof of the recursion theorem. We are again going to proceed by induction, intuitively by looking at the least place where f is not defined. Formally, call a class function $f : B_0 \to \text{SET}$ a *partial solution* if B_0 is an R-initial segment of B and for all $b \in B_0$, $f(b) = F((f(a))_{a \in \text{pred}_R(b)}, b)$. We are looking for a partial solution with domain B. We proceed in several steps:

(1) If f is a partial solution and $B_0 \subseteq \text{dom}(f)$ is an initial segment of B, then $f \restriction B_0$ is a partial solution: this is immediate from the definition.

(2) If $h_0 : B_0 \to \text{SET}$ and $h_1 : B_1 \to \text{SET}$ are partial solutions, then h_0 and h_1 agree on the intersection $B_0 \cap B_1$ of their domain. Indeed, note first that $h_0 \upharpoonright (B_0 \cap B_1)$ and $h_1 \upharpoonright (B_0 \cap B_1)$ are both still partial solutions (being an initial segment is preserved by intersections). Thus it suffices to show that any two partial solutions with the same domain (say B_0) are equal. So suppose $f, g : B_0 \to \text{SET}$ are partial solutions. We are going to prove that $f(b) = g(b)$ for all $b \in B_0$ by induction on R. That is, we are going to use Theorem 5.5 with the relation R restricted to B_0, and the set $A := \{b \in B_0 \mid f(b) = g(b)\}$. So fix $b \in B_0$ and assume that for all R-predecessors a of b, we know that $f(a) = g(a)$. We have to see that $f(b) = g(b)$. This is immediate from the following computation:

$$f(b) = F((f(a))_{a \in \text{pred}_R(b)}, b) = F((g(a))_{a \in \text{pred}_R(b)}, b) = g(b)$$

Note this in particular proves the uniqueness part of the recursion theorem (setting $B_0 = B$). It remains to prove existence.

(3) For all $b \in B$, there is a partial solution f_b whose domain contains b. We prove this by induction on b. That is, we use Theorem 5.5 with the set $A = \{b \in B \mid \text{ there is a partial solution with domain containing } b\}$. Fix $b \in B$ and suppose that for all R-predecessors a of b, there is a partial solution f_a whose domain contains a. Define $f : \{b\} \cup \bigcup_{a \in \text{pred}_R(b)} \text{dom}(f_a) \to \text{SET}$ as follows: First, $f(b') = f_a(b')$ if $b' \in \text{dom}(f_a)$ for some $a \in \text{pred}_R(b)$. Second, $f(b) = F((f_a(a))_{a \in \text{pred}_R(b)}, b)$. Observe that f is well-defined by the uniqueness of partial solutions. Let us check that f is indeed a partial solution. Let $b' \in \text{dom}(f)$. If $b' = b$, this is just the definition of $f(b)$, and otherwise this means that $b' \in \text{dom}(f_a)$ for some $a \in \text{pred}_R(b)$. Then:

$$f(b') = f_a(b') = F((f_a(a_0))_{a_0 \in \text{pred}_R(a)}, b') = F((f_{a_0}(a_0))_{a_0 \in \text{pred}_R(a)}, b') = F((f(a_0))_{a_0 \in \text{pred}_R(a)}, b')$$

Where the first equality is by definition of f, the second is because f_a is a partial solution, the third is because the partial solutions f_a and f_{a_0} must agree at the intersections of their domains, in particular at a_0. The fourth equality is by definition of f again.

(4) There is a partial solution with domain B. Indeed, define $f : B \to \text{SET}$ by $f(b) = f_b(b)$ (where f_b is the partial solution defined in the previous part). As before, we get that f is a partial solution.

□

The following is another property of relations:

Definition 5.9. A class relation R on B is *extensional* if for any $b_1, b_2 \in B$, if $\text{pred}_R(b_1) = \text{pred}_R(b_2)$, then $b_1 = b_2$.

Example 5.10.

- The membership relation $\in$ on SET is extensional. This is a restatement of the extensionality axiom. However membership is not always extensional when restricted to a different class. Can you think of an example?
- Any strict linear order $(A, <)$ is extensional: say $\text{pred}_R(b_1) = \text{pred}_R(b_2)$ and $b_1 \neq b_2$. Then without loss of generality $b_1 < b_2$, hence $b_1 \in \text{pred}_R(b_2) - \text{pred}_R(b_1)$, contradiction.

- The strict divisibility relation on $\mathbb{N}$ is not extensional: 2 and 3 have the same proper divisors (namely 1) but are not equal.

We conclude this section by proving that, in fact, *any* wellfounded and set-like class relation that is extensional is isomorphic to one that is of the form membership on some class. This gives a "concrete" way to think of such relations. You can think of this as a variation on the fact that any poset set embeds into a poset of the form $(\mathcal{P}(X), \subseteq)$ (Example 4.7(9)).

Definition 5.11. We say an arbitrary class X is a *set-initial segment* (also sometimes called *transitive*) if it is an ϵ-initial segment, where ϵ is seen as a binary relation on SET. Explicitly, X is a set-initial segment if for any $x \in X$ and any $y \in x$, we have that $y \in X$. In other words, $x \in X$ implies $x \subseteq X$.

Note that $\mathbb{N}$ is a set-initial segment. Indeed, from the construction we have that if $m \in n$ and $n \in \mathbb{N}$, then $m \in \mathbb{N}$. Similarly, for any $n \in \mathbb{N}$, the set $\{m \in \mathbb{N} \mid m < n\}$ (which happens to be equal to n itself!) is a set-initial segment.

Definition 5.12. Let R be a class relation on A and S be a class relation on B. An *isomorphism* from R to S is a class function $f : A \to B$ such that f is a bijection and for any $a_1, a_2 \in A$, $a_1 R a_2$ if and only if $f(a_1) R f(a_2)$.

For a class X, let ϵ_X denote the membership relation on X.

Theorem 5.13 (Mostowski collapsing theorem)**.** Let R be a set-like, extensional, wellfounded relation on a class A. Then there exists a unique set-initial segment X and a unique isomorphism f from R to ϵ_X.

Proof. The idea is as follows: suppose R is just the usual ordering on $A = \{1, 2\}$. We send 1 to $\emptyset$, and then will send 2 to the set containing its predecessors, namely $\{\emptyset\}$. In general, define a class function $g : A \to \text{SET}$ by recursion on R as follows:

$$g(b) = \{g(a) \mid a \in \text{pred}_R(b)\}$$

More precisely, we apply the recursion theorem to the class function F defined by $F((x_a)_{a \in \text{pred}_R(b)}, b) = \{x_a \mid a \in \text{pred}_R(b)\}$.

Let X be the range of g, and let f be the corestriction of g to codomain X. First, X is a set-initial segment. Indeed, if $x \in X$, then x is of the form $x = \{g(a) \mid a \in \text{pred}_R(b)\}$ for some b. Any $y \in x$ is of the form $g(a)$ for some a, i.e. it is also in X.

We now check that f is an isomorphism. By construction, f is surjective. To check that f is injective. Suppose not. Pick an R-minimal $b \in B$ such that there is $b' \neq b$ with $f(b) = f(b')$. Thus $f(b) = \{f(a) \mid a \in \text{pred}_R(b)\} = \{f(a') \mid a' \in \text{pred}_R(b')\} = f(b')$. If $\text{pred}_R(b) = \text{pred}_R(b')$, extensionality of R implies that $b = b'$, so $\text{pred}_R(b) \neq \text{pred}_R(b')$. If there exists $a \in \text{pred}_R(b) - \text{pred}_R(b')$, we then have that $f(a) = f(a')$ for some $a' \in \text{pred}_R(b')$ with $a \neq a'$. Since aRb, this contradicts the R-minimality of b. If there exists $a' \in \text{pred}_R(b') - \text{pred}_R(b)$, the argument is symmetric.

Thus f is a bijection. Finally, if aRb then by definition $f(a) \in f(b) = \{f(a') \mid a' \in \text{pred}_R(b)\}$. Conversely, if $a, b \in B$ and $f(a) \in f(b)$, then $f(a) = f(a')$ for some $a' \in \text{pred}_R(b)$, and by injectivity $a = a'$, so aRb. This concludes the proof that f is an isomorphism.

It remains to see that f and X are unique. Suppose that f' is another isomorphism from R to $\epsilon_{X'}$ for X' another set-initial segment. Let $h := f' \circ f^{-1}$. This is an isomorphism from ϵ_X to $\epsilon_{X'}$. We prove that h is the identity on X, which will imply that $f = f'$. We show by induction on $x \in X$ that $h(x) = x$. Assume inductively that $h(x_0) = x_0$ for all x_0 with $x_0 \in x$. Let $x_0 \in x$. Then $h(x_0) \in h(x)$ (as h is an isomorphism), and by assumption $h(x_0) = x_0$. Thus $x \subseteq h(x)$. Conversely, if $y \in h(x)$, then as X' is a set-initial segment, $y \in X'$, and as h is an isomorphism there exists $x_0 \in X$ such that $h(x_0) = y$. Thus $h(x_0) \in h(x)$, so $x_0 \in x$, so $h(x_0) = x_0 = y$, so $y \in x$. This proves that $h(x) \subseteq x$, hence that $h(x) = x$. □

Definition 5.14. We call X as above the *collapse* of R and f the *collapsing map.*

Example 5.15.

(1) If R is the usual strict ordering on $\mathbb{N} - \{0\}$, then the collapse X will be $\mathbb{N}$: one can show 1 is sent to 0, 2 to 1, etc. A quick way to prove this is via uniqueness of the collapse: we know that $\mathbb{N} - \{0\}$ and $\mathbb{N}$ are isomorphic, and $\mathbb{N}$ is a set-initial segment, so $\mathbb{N}$ must be the collapse.

(2) If X is a set-initial segment, then $\mathcal{P}(X)$ is a set-initial segment, so it is its own collapse. The simplest case is when $X = \emptyset$, so $\mathcal{P}(\emptyset) = \{\emptyset\}$. A slightly harder case is $\mathcal{P}(\mathcal{P}(\emptyset)) = \{\emptyset, \{\emptyset\}\}$.

6. Well-orderings and ordinals

Recall:

Definition 6.1. A *well-ordering* is a wellfounded strict linear order. When the domain of the relation is a class rather than a set, we talk of a *class well-ordering.*

Example 6.2. $(\{0, 1\}, <)$ and $(\mathbb{N}, <)$ are well-orderings, but $(\mathbb{Z}, <)$ is not. Also, strict divisibility on $\mathbb{N}$ is not a well-ordering, since we require the ordering to be total.

Well-orderings are, in particular, wellfounded relations, so one can do induction on them. The collapsing theorem also tells us that each well-ordering is (uniquely) isomorphic to the membership relation on a set-initial segment. Thus to avoid renamings, it makes sense to study only the well-orderings given by the membership relation on a set-initial segment. We give those a name:

Definition 6.3. An *ordinal* is a set α such that:

(1) α is a set-initial segment.
(2) $(\alpha, \in)$ is a well-ordering.

Recall that for α to be a set-initial segment means that if $y \in \alpha$ and $x \in y$, then $x \in \alpha$, see Definition 5.11. Said another way, $x \in \alpha$ implies that $x \subseteq \alpha$.

The other part of the definition means that the membership relation $\in_\alpha$ on α (where $x \in_\alpha y$ if and only if $x \in y$) is a well-ordering.

Definition 6.3 may seem strange, so let us look at some examples:

Example 6.4.

- $\emptyset$ is an ordinal. This is easy to check: it is vacuously a set-initial segment, and the membership relation on $\emptyset$ is a well-ordering. Following our earlier convention, we will call this ordinal 0.

- $\{\emptyset\}$ is also an ordinal. Indeed, if $x \in \{\emptyset\}$, then $x = \emptyset$, so $x \subseteq \emptyset$. Thus $\{\emptyset\}$ is a set-initial segment. It is also straightforward to check that $(\{\emptyset\}, \in)$ is a well-ordering. Following our earlier convention, we will call this ordinal 1. Observe that $1 = \{0\}$.
- $\{\emptyset, \{\emptyset\}\}$ is also an ordinal. Observe that this is $\{0, 1\}$. We will call this ordinal 2.
- $x = \{\{\emptyset\}\}$ is not an ordinal because it is not a set-initial segment: $\{\emptyset\} \in x$, but $\{\emptyset\} \not\subseteq x$, as $\emptyset \notin x$.
- The set $\mathbb{N}$ of natural numbers is also an ordinal. Indeed, you showed in assignment 1 that if $n \in \mathbb{N}$ then $n \subseteq \mathbb{N}$ so $\mathbb{N}$ is a set-initial segment. That $(\mathbb{N}, \in)$ is a well-ordering is of course a fundamental property of $\mathbb{N}$. We use the letter ω to denote $\mathbb{N}$, when we think of it as an ordinal.
- Any natural number n is an ordinal. You showed in assignment 1 that $m \in n$ implies $m \subseteq n$, and of course the ordering $(n, \in)$ is simply the usual ordering on the set $\{0, 1, \dots, n-1\}$. Thus ordinals generalize the natural numbers: all natural numbers are ordinals, but not conversely: there are more ordinals. We have already seen an ordinal that is not a natural number, ω. It is the first *infinite* ordinal. The natural numbers are called finite ordinals
- What about another infinite ordinal? Consider $\omega \cup \{\omega\}$. You should be able to check that it is a set-initial segment and also well-ordered by membership. It is the successor of ω. We will call it $\mathrm{S}\,\omega$, or $\omega + 1$.

You can think of ordinals as numbers denoting position, or rank. 0 is the zeroth ordinal, 1 is the first ordinal, ..., ω is the ωth ordinal, $\omega + 1$ is the ordinal right after ω, etc. They are related to, but different from, *cardinals*, which we will consider later. For example, notice that the function $f : \omega \to \omega \cup \{\omega\}$ given by $f(0) = \omega$, $f(n) = n+1$ for $n \in \omega$ is a bijection. Thus ω and $\omega \cup \{\omega\}$ have the same cardinality, but they will be considered different as ordinals: ω has only finite ordinals before it. It corresponds to the first infinite step in an inductive process. On the other hand $\omega \cup \{\omega\}$ corresponds to the step right after ω.

In passing, let us restate that *any* well-ordering will be isomorphic to a unique ordinal:

Theorem 6.5. If $(A, <)$ is a well-ordering, then there is a unique ordinal α and a unique isomorphism of $(A, <)$ into $(\alpha, \in)$.

Proof. This is a particular case of the collapsing theorem. □

Definition 6.6. If $(A, <)$ is a well-ordering, we call the unique ordinal α isomorphic to $(A, <)$ the *order type* of $(A, <)$.

We can think of the ordinals as being themselves ordered by the membership relation. In this way, each ordinal is really the set of its predecessors (just like for the natural numbers). Thus we can think of $\omega \cup \{\omega\}$ as the ordinal whose predecessors are the predecessors of ω and ω itself.

In general, we can make the following definition:

Definition 6.7. For an ordinal α, define the *successor of* α, $\mathrm{S}\,\alpha$, by $\mathrm{S}\,\alpha = \alpha \cup \{\alpha\}$.

Lemma 6.8. If α is an ordinal, then $\mathrm{S}\,\alpha$ is an ordinal.

Proof. Exercise. □

Note that for natural numbers (i.e. finite ordinals), this coincide with the previous definition. However we can apply successors to *any* ordinal. For example the successor of ω is $\mathrm{S}\,\omega = \omega \cup \{\omega\}$ and the successor of $\mathrm{S}\,\omega$ is $\omega \cup \{\omega\} \cup \{\omega \cup \{\omega\}\}$, etc. Let us now make the ordering on the ordinals precise:

Definition 6.9. Let OR denote the class of all ordinals. We define a (class) relation $<$ on it by $\alpha < \beta$ if and only if $\alpha \in \beta$.

As usual, we will write $\alpha \leq \beta$ to mean that $\alpha < \beta$ or $\alpha = \beta$. We work toward showing that $(\mathrm{OR}, <)$ is a class well-ordering. The proof is a generalization of the fact, seen in assignments, that $(\mathbb{N}, \in)$ is a well-ordering.

Lemma 6.10.

(1) If α is an ordinal and $\beta \in \alpha$, then β is an ordinal. That is, OR is a set-initial segment.
(2) $(\mathrm{OR}, <)$ is set-like.
(3) $(\mathrm{OR}, <)$ is wellfounded.
(4) $(\mathrm{OR}, <)$ is a strict partial order.
(5) If α and β are ordinals, then $\alpha \leq \beta$ if and only if $\alpha \subseteq \beta$.
(6) $(\mathrm{OR}, <)$ is trichotomous: if α and β are ordinals, either $\alpha < \beta$, $\beta < \alpha$, or $\alpha = \beta$.

Proof.

(1) Let $\beta \in \alpha$. We know that α is an ordinal, so a set-initial segment, hence $\beta \subseteq \alpha$. Because $(\alpha, \in)$ is a well-ordering, this immediately implies that $(\beta, \in)$ is a well-ordering. Let us check that β is a set-initial segment: let $a \in b \in \beta$. We have to see that $a \in \beta$. We know that $b \in \alpha$ as α is set-initial, and therefore also $a \in \alpha$. Thus we have $a, b, \beta \in \alpha$ and $a \in b \in \beta$, so by transitivity of the ordering $(\alpha, \in)$, $a \in \beta$.

(2) Let $\alpha \in \mathrm{OR}$. Then any ϵ-predecessor of α in OR must (by definition of membership) be a member of α, so the set of predecessors of α is a subclass of α, hence a set.

(3) Let $A \subseteq \mathrm{OR}$ be a non-empty class. Pick $\alpha \in A$. If α is minimal in A, we are done already. If not, $A_0 := A \cap \alpha$ is a non-empty subset of α, and α is a well-ordering, so pick $\alpha_0 \in A_0$ ϵ-minimal in A_0 (with respect to $(\alpha, \in)$). We claim that α_0 is also ϵ-minimal in A with respect to $(\mathrm{OR}, \in)$. Indeed, if $a \in \alpha_0$ and $a \in A$, then as α is a set-initial segment, $a \in \alpha$, so $a \in A_0$, contradiction.

(4)
- OR is irreflexive: if $\alpha \in \alpha$, then this means that the set $\{\alpha\}$ has no minimal element in α, contradiction (we could also use the axiom of foundation).
- OR is transitive: if $\alpha \in \beta \in \gamma$, then as γ is a set-initial segment, $\alpha \in \gamma$.
- OR is antisymmetric: if $\alpha \in \beta$ and $\beta \in \alpha$, then by transitivity, $\alpha \in \alpha$, contradicting irreflexivity.

(5) Let α and β be ordinals. If $\alpha \leq \beta$, then either $\alpha = \beta$, or $\alpha \in \beta$ and in the latter case as β is a set-initial segment, $\alpha \subseteq \beta$. Let us now prove the converse. Assume that $\alpha \subseteq \beta$. If $\alpha = \beta$, we are done so assume that $\alpha \subsetneq \beta$. Let $A := \beta - \alpha$.

<u>Claim</u>: For $x \in \beta$, $x \in A$ if and only if $\alpha \leq x$.

<u>Proof</u>: If $x \in A$, then $x \in \beta$ and $x \notin \alpha$ by definition. As β is a total ordering, $\alpha \leq x$. Conversely, if $\alpha \leq x$, then $x \notin \alpha$, so $x \in A$. $\dagger_{\text{Claim}}$

Now β is a well-ordering so let γ be the minimal element of A. We prove that $\gamma = \alpha$: on the one hand, if $x \in \gamma$ then x cannot be in A as it would contradict the minimality of γ, so $x \in \alpha$. Thus $\gamma \subseteq \alpha$. Conversely, if $x \in \alpha$, then since $\alpha \subseteq \beta$, $x \in \beta$. If $x \notin \gamma$, then as β is total, $\gamma \leq x$. By the claim, $\alpha \leq \gamma$, so $\alpha \leq x$, contradicting the starting assumption that $x \in \alpha$.

Since $\alpha = \gamma$ and $\gamma \in \beta$, we have that $\alpha \in \beta$, as desired.

(6) Let α, β be ordinals and assume that $\beta \not\leq \alpha$. We will show that $\alpha < \beta$. By the previous part, $\beta \not\leq \alpha$ implies that $\beta \not\subseteq \alpha$. Let γ be a minimal element of $A := \beta - \alpha$. Then for any $x \in \gamma$, $x \in \alpha$, so $\gamma \subseteq \alpha$, so $\gamma \leq \alpha$. However by definition of A, $\gamma \notin \alpha$, so $\gamma \not< \alpha$, so the only possibility is that $\gamma = \alpha$. As $\gamma \in \beta$, $\alpha < \beta$.

□

We summarize the results of the lemma in one theorem:

Theorem 6.11. $(\mathrm{OR}, <)$ is a set-like set-initial segment and a class well-ordering.

Proof. By the previous lemma. □

The importance of Theorem 6.11 is that we can do induction and recursion on the ordinals. We note also in passing that it establishes yet another proof that SET is not a set:

Corollary 6.12 (Burali-Forti paradox)**.** OR is a proper class. Therefore SET is also a proper class.

Proof. Since $\mathrm{OR} \subseteq \mathrm{SET}$, the last sentence follows from the first and the contrapositive of the subset axiom. Now assume for a contradiction that OR is a set. Then by Theorem 6.11, it would be an ordinal itself. Hence $\mathrm{OR} \in \mathrm{OR}$, so the set $\{\mathrm{OR}\}$ does not have an ϵ-minimal element, a contradiction to wellfoundedness. □

What are other ways to build ordinals than taking successors? Well, we have a chain of finite successors of ω: $\omega < \mathrm{S}\,\omega < \mathrm{S}\,\mathrm{S}\,\omega < \ldots$, can we take the limit of that chain? The answer is yes! It is given by the union:

Lemma 6.13. If X is a set of ordinals, then $\bigcup X$ is an ordinal.

Proof. Let $Y := \bigcup X$. Let us first check that Y is a set-initial segment. Let $y \in Y$ and let $a \in y$. Since $y \in Y$, $y \in \alpha$ for some $\alpha \in X$ by definition of the union. Since α is an ordinal, it is a set-initial segment, so $a \in \alpha \subseteq \bigcup X = Y$, so $a \in Y$, as desired.

To finish, note that $Y \subseteq \mathrm{OR}$, and we have proven that $(\mathrm{OR}, \in)$ is a well-ordering, so $(Y, <)$ is itself a well-ordering on general grounds (any restriction of a well-ordering is a well-ordering). □

The union of a set of ordinals will give the minimal ordinal α so that all the members of X are predecessors of α. Thus we define:

Definition 6.14. For X a set of ordinals, we write $\sup(X)$ instead of $\bigcup X$. If $(\alpha_i)_{i \in I}$ is a sequence of ordinals, we write $\sup_{i \in I} \alpha_i$ for $\sup\{\alpha_i \mid i \in I\}$.

In the exercise, you will prove that this is really a supremum in the usual sense.

A consequence of the previous discussion is that we can let $X = \{\omega, \mathrm{S}\,\omega, \mathrm{S}\,\mathrm{S}\,\omega, \ldots\}$, and consider $\alpha := \sup(X)$. This is an ordinal, usually written $\omega + \omega$. Note that, just like ω, it is not the successor of any ordinal. More generally, we make the following definition:

Definition 6.15. An ordinal α is *successor* if there exists an ordinal β such that $\alpha = \beta + 1$. We say that α is *limit* if it is not zero and not a successor.

Thus ω and $\omega + \omega$ are limits, while $1, 2, \mathrm{S}\,\omega, \mathrm{S}\,\mathrm{S}\,\omega$ are successors. The ordinal 0 is the only ordinal that is neither limit nor successor. The following characterization of limit ordinals will be useful.

Lemma 6.16. Let α be a strictly positive ordinal. The following are equivalent:

1. α is limit.
2. $\alpha = \sup_{\beta < \alpha} \beta$.
3. There exists a set $X \subseteq \alpha$ such that for any $\beta < \alpha$, there exists $\gamma \in X$ with $\beta < \gamma$.

Proof. Exercise. □

Consider what happens when using the induction theorem on OR. We are given a class P of ordinals (a certain property), and want to prove that $P = \mathrm{OR}$. To do this, we have to prove that for any ordinal α, if all the predecessors of α are in P, then α is in P as well. It is often convenient (but by no means necessary) to split this verification into three cases: α zero, α successor, or α limit. Thus we obtain the following induction theorem:

Theorem 6.17 (Induction theorem for ordinals)**.** Let $A \subseteq \mathrm{OR}$. Suppose that the following three conditions holds:

- (Base case) $0 \in A$.
- (Successor step) If $\alpha \in A$, then $\mathrm{S}\,\alpha \in A$.
- (Limit step) If α is a limit ordinal and $\beta \in A$ for all $\beta < \alpha$, then $\alpha \in A$.

Then $A = \mathrm{OR}$.

Proof. This is a special case of Theorem 5.5, where we have split into three cases (α zero, successor, or limit) the verification that for all α, if the predecessors of α are in A then α is in A. □

This result may look familiar, but note that compared to the induction theorem for $\mathbb{N}$ (Theorem 2.12), there is an additional limit case to check! In the case of induction for the natural numbers, this limit case is not relevant, as all the ordinals below ω are either zero or successors. However we will see plenty of examples in this course where the limit case matters.

There is also a corresponding recursion theorem for ordinals. A precise statement is left to you, but the basic idea is that to define a function by recursion on the ordinals, one should specify a base case, a successor step, *and* a limit step.

6.1. **Ordinal arithmetic.** Let us for example see how to define *ordinal addition.*

Definition 6.18. For each fixed ordinal α, we define by induction on β an ordinal $\alpha + \beta$ as follows:

- (Base case) $\alpha + \beta = \alpha$ if $\beta = 0$.
- (Inductive step) $\alpha + \beta = \mathrm{S}(\alpha + \gamma)$ if $\beta = \mathrm{S}\,\gamma$ for some ordinal γ.
- (Limit step) $\alpha + \beta = \sup_{\gamma < \beta}(\alpha + \gamma)$ if β is limit.

Note again that the definition is very similar to that of addition on the natural numbers. In fact, it coincides on the finite ordinal. However here there is a limit step that handles addition of bigger ordinals.

Formally we really have, for each fixed ordinal α, defined by recursion a certain class function $f_\alpha : \mathrm{OR} \to \mathrm{OR}$ that sends β to α if $\beta = 0$, to $\mathrm{S}\, f_\alpha(\gamma)$ if $\beta = \mathrm{S}\,\gamma$, and to $\sup_{\gamma<\beta} \alpha+\gamma$ if β is limit. We then define $+ : \mathrm{OR} \times \mathrm{OR} \to \mathrm{OR}$ by $+(\alpha, \beta) := f_\alpha(\beta)$, and write $\alpha + \beta$ instead of $+(\alpha, \beta)$. Typically of course we won't write all these tedious steps and will adopt the style above. We may not even mention β explicitly in the clauses of the later definitions by induction. For example, let us define now *ordinal multiplication*:

Definition 6.19. For a fixed ordinal α, we define by induction on β an ordinal $\alpha \cdot \beta$ as follows:

- (Base case) $\alpha \cdot 0 = 0$.
- (Inductive step) $\alpha \cdot (\mathrm{S}\,\gamma) = (\alpha \cdot \gamma) + \alpha$.
- (Limit step) $\alpha \cdot \delta = \sup_{\gamma<\delta} \alpha \cdot \gamma$ if δ is limit.

Finally, and for more practice, let us even define *ordinal exponentiation*:

Definition 6.20. For a fixed ordinal α, we define by induction on β an ordinal α^β as follows:

- (Base case) $\alpha^0 = 1$.
- (Inductive step) $\alpha^{\mathrm{S}\,\gamma} = \alpha^\gamma \cdot \alpha$.
- (Limit step) $\alpha^\delta = \sup_{\gamma<\delta} \alpha^\gamma$ if δ is limit.

Example 6.21.

(1) Addition, multiplication, and exponentiation all coincide with their familiar definition on the natural numbers (i.e. for $n, m < \omega$).

(2) For any ordinal α, $\alpha + 1 = \alpha + \mathrm{S}\,0 = \mathrm{S}(\alpha + 0) = \mathrm{S}\,\alpha$, so adding one to an ordinal is the same as taking its successor. We will use this without comments from now on.

(3) *WARNING*: ordinal addition is *not* commutative. To see this, note that $\omega + 1 = \mathrm{S}\,\omega$ by the above. However $1 + \omega = \sup_{n<\omega}(1 + n) = \omega \neq \mathrm{S}\,\omega$.

(4) Ordinal addition *is* fortunately associative, and satisfies some other properties of addition on the natural numbers, such as:
   - (Right cancellativity) $\alpha + \gamma = \alpha + \beta$ implies $\gamma = \beta$. You will prove this in the assignments.
   - (Monotonicity)If $\alpha \le \alpha'$ and $\beta \le \beta'$, then $\alpha+\beta \le \alpha'+\beta'$. This should be easy from the definitions. What about strict monotonicity?
   - (Right continuity) For any non-empty set X of ordinals, $\alpha+\sup(X) = \sup_{\beta\in X}(\alpha + \beta)$. You will prove this in the assignments. What about left continuity?

(5) For any ordinal α, $\alpha \cdot 1 = \alpha \cdot 0 + \alpha = 0 + \alpha$. We claim that $0 + \alpha = \alpha$ (this requires a proof, as in general addition is not commutative). We prove this by induction on α: if $\alpha = 0$, then $0 + 0 = 0$. If $\alpha = \beta + 1$ is successor, $0 + (\beta + 1) = (0 + \beta) + 1 = \beta + 1$. The first equality here is by the successor step in the definition of addition (recall that $\beta + 1$ is exactly the same as $\mathrm{S}\,\beta$). The second is by the induction hypothesis. Finally, if α is limit, $0 + \alpha = \sup_{\gamma<\alpha}(0 + \gamma) = \sup_{\gamma<\alpha} \gamma = \alpha$, where the second equality is by the induction hypothesis, and the third equality is by Lemma 6.16.

(6) *WARNING*: ordinal multiplication is *not* commutative either: $\omega \cdot 2 = \omega + \omega = \sup_{n<\omega} \omega + n$. In particular, $\omega \cdot 2$ it is strictly greater than ω. On the other hand, $2 \cdot \omega = \sup_{n<\omega} 2 \cdot n = \omega$.
(7) Note that $\alpha^1 = \alpha$ and $\alpha^2 = \alpha \cdot \alpha$, for any ordinal α. On the other hand, $2^\omega = \sup_{n<\omega} 2^n = \omega$.

For practice with transfinite induction, let us prove another property of ordinal addition:

Lemma 6.22. If $\alpha \leq \beta$ are ordinals, there exists a unique ordinal γ such that $\beta = \alpha + \gamma$.

Proof. Uniqueness is left as an exercise for the assignments. Let's prove existence: we fix α and proceed by transfinite induction on β. Precisely, we fix an arbitrary ordinal β and assume that the statement is true for all ordinals $\beta_0 < \beta$. Here the cases where $\beta < \alpha$ are vacuous (or alternatively, one could think of our induction as being on the restriction of $<$ to the class of ordinals that are at least α). Thus our "base case" becomes $\alpha = \beta$. In this case, $\gamma = 0$ works. Assume now that $\beta > \alpha$. If β is a successor, $\beta = \beta_0 + 1$, for $\beta_0 \geq \alpha$. By the induction hypothesis, there is γ_0 such that $\alpha + \gamma_0 = \beta_0$. Then $\alpha + (\gamma_0 + 1) = (\alpha + \gamma_0) + 1 = \beta_0 + 1 = \beta$, so $\gamma := \gamma_0 + 1$ works.

Finally, if β is a limit ordinal, for each ordinal β_0 with $\alpha \leq \beta_0 < \beta$, we have by induction an ordinal γ_{β_0} such that $\alpha + \gamma_{\beta_0} = \beta_0$. Let $\gamma := \sup_{\alpha \leq \beta_0 < \beta} \gamma_{\beta_0}$. Note that $\beta = \sup_{\alpha \leq \beta_0 < \beta} \beta_0 = \sup_{\alpha \leq \beta_0 < \beta} (\alpha + \gamma_{\beta_0}) = \alpha + \sup_{\alpha \leq \beta_0 < \beta} \gamma_{\beta_0} = \alpha + \gamma$. All the equalities should be clear, except perhaps the one before last: this is given by the right continuity property of ordinal addition, to be shown in the assignments. □

To visualize ordinal addition and multiplication, let us introduce an equivalent definition in terms of linear orderings. Recall that we can think of an ordinal α as a well-ordering $(\alpha, \in)$. The ordinal $\alpha + \beta$ is obtained by "concatenating" the corresponding two linear orderings. More formally, for any two strict linear orderings I and J, define their *concatenation*, $I \oplus J$ to be the relation R on $(I \times \{0\}) \cup (J \times \{1\})$ defined by $(i, \ell) R (j, \ell')$ if and only if $\ell < \ell'$, or $\ell = \ell'$ and $i < j$. Intuitively, $I \oplus J$ is the disjoint union of I and J, ordered by putting the elements of I first and then those of J, with the ordering of I and J inherited for the components.

Relatedly, also define $I \times J$ to be the relation R defined on the set $I \times J$ by $(i_1, j_1) R (i_2, j_2)$ if and only if $i_1 < i_2$ or $i_1 = i_2$ and $j_1 < j_2$ (this is the lexicographic, or dictionary order). One can show:

Lemma 6.23. If I and J are well-orderings, then $I \oplus J$ and $I \times J$ are well-orderings.

Proof. Exercise. □

Any well-ordering is isomorphic to a unique ordinal (Theorem 6.5). Thus one could have defined $\alpha + \beta$ to be the unique ordinal isomorphic to the well-ordering $\alpha \oplus \beta$, and $\alpha \cdot \beta$ to be the unique ordinal isomorphic to the well-ordering $\beta \times \alpha$ (we are reversing the order of factor here just as a convention, to avoid having addition right continuous but multiplication left continuous). As an exercise, try to convince yourself this would have yielded an equivalent definition (we will use this freely):

Theorem 6.24. $\alpha + \beta$ is the unique ordinal isomorphic to $\alpha \oplus \beta$, and $\alpha \cdot \beta$ is the unique ordinal isomorphic to the lexicographic ordering on $\beta \times \alpha$.

Proof. Exercise. □

We will not need a similar characterization of ordinal exponentiation, but you may still try to figure it out, as a challenge!

Some properties of addition and multiplication are easier to prove using the characterization in terms of orderings, while some are easier to prove directly from the definitions using transfinite induction. You should try to familiarize yourself with both points of view.

6.2. **An application: Cantor normal form and Goodstein sequences.** Any natural number k as a sum of the form $k = 2^{n-1}c_{n-1} + 2^{n-2}c_{n-2} + \dots 2^0 c_0$, with $c_i \in \{0, 1\}$ for all $i < n$, and $c_{n-1} \neq 0$. Further, such a representation is unique. This is the well-known base two representation. For example, $11 = 2^3 \cdot 1 + 2^2 \cdot 0 + 2^1 \cdot 1 + 2^0 \cdot 1$ (said another way, 11 in base ten is 1011 in binary). More generally, any natural number has a base b representation for any natural number $b \geq 2$. Can one generalize such representations to ordinals? The answer is yes. In the assignments you will prove the following:

Theorem 6.25 (Cantor's normal form theorem)**.** Let $\gamma \geq 2$ be an ordinal. For any ordinal α, there exists $n < \omega$, ordinals $\alpha_0 > \alpha_1 > \dots > \alpha_{n-1}$, and $c_0, c_1, \dots, c_{n-1}$ such that $0 < c_i < \gamma$ for all $i < n$ and:

$$\alpha = \gamma^{\alpha_0} c_0 + \gamma^{\alpha_1} c_1 + \dots + \gamma^{\alpha_{n-1}} c_{n-1}$$

(with the convention that the empty sum is zero). Moreover this representation is unique.

The representation above is called the *base γ representation of α.*

Example 6.26.

(1) If $\gamma = 2$, then the base 2 representation of ω is $\omega = 2^\omega \cdot 1$.
(2) If $\alpha = 3 + \omega^2 + \omega + 5$, then the base ω representation of α is $\alpha = \omega^2 + \omega + 5 = \omega^2 \cdot 1 + \omega \cdot 1 + \omega^0 \cdot 5$.

Let us use the base ω representation to understand a finite object, the *Goodstein sequences.* The definition of a weak Goodstein sequence may be easier to understand:

Definition 6.27. For a natural number m, the *weak Goodstein sequence* of m is a sequence $(a_n)_{n \in \mathbb{N}}$ of natural numbers defined inductively as follows:

- $a_0 = m$.
- Given a_k, if $a_k = 0$ then $a_{k+1} = 0$. Otherwise, if $a_k > 0$, write it in base $k + 2$: $a_k = (k+2)^{n-1}c_{n-1} + (k+2)^{n-2}c_{n-2} + \dots + (k+2)^0 c_0$, with $0 \leq c_i < k+2$. Define a_{k+1} to be $(k+3)^{n-1}c_{n-1} + \dots + (k+3)^0 c_0 - 1$. That is, a_{k+1} is obtained by subtracting one from the number whose base $k+3$ representation has the same coefficient as the base $k+2$ representation of a_k.

For example, let us consider the weak Goodstein sequence of $m = 21$. We have that $a_0 = 21 = 2^4 + 2^2 + 2^0$. Thus $a_1 = 3^4 + 3^2 + 2^0 - 1 = 3^4 + 3^2 = 90$. Then $a_2 = 4^4 + 4^2 - 1 = 271$. In base 5, 271 is $271 = 4^4 + 4^1 \cdot 3 + 4^0 \cdot 3$. Thus $a_3 = 5^4 + 5^1 \cdot 3 + 5^0 \cdot 3 - 1 = 642$, $a_4 = 1315$, $a_5 = 2422$, $a_6 = 4119$, etc. It seems it goes on forever. However, for smaller number the weak Goodstein sequence stabilizes at

0: for $m = 4 = 2^2$, $a_1 = 3^2 - 1 = 8 = 3^1 \cdot 2 + 3^0 \cdot 2$, $a_2 = 4^1 \cdot 2 + 4^0 \cdot 2 - 1 = 9$, $a_3 = 5^1 \cdot 2 + 5^0 \cdot 1 - 1 = 5^1 \cdot 2 = 10$, $a_4 = 6^1 \cdot 2 - 1 = 11 = 6^1 \cdot 1 + 6^0 \cdot 5$, $a_5 = 7 \cdot 1 + 6^0 \cdot 5 - 1 = 11$, ..., $a_9 = 11$, $a_{10} = 10$, $a_{11} = 9$, ..., $a_{20} = 0$. The fact this takes so long may make us wonder about the eventual behavior of weak Goodstein sequences starting at higher numbers. This suspicion turns out to be warranted:

Theorem 6.28. For any natural number m, the weak Goodstein sequence of m is eventually zero.

Proof. Assume without loss of generality that $m > 0$ and let $a_0, a_1, \ldots$ be the weak Goodstein sequence of m. Fix a natural number k. Write a_k as $(k+2)^{\alpha_0} c_0 + \ldots + (k+2)^{\alpha_{n-1}} c_{n-1}$, with $\alpha_0 > \alpha_1 > \ldots > \alpha_{n-1}$ and $0 < c_i < k+2$ for any $i < n$. Consider the ordinal $\beta_k := \omega^{\alpha_0} c_0 + \ldots + \omega^{\alpha_{n-1}} c_{n-1}$ obtained by keeping the same coefficient but changing the base to ω. Of course, $a_k \leq \beta_k$. Assume for a contradiction that $a_k \neq 0$ for any k. Then $\beta_k > 0$, and moreover you should be able to convince yourself that $\beta_k > \beta_{k+1}$ (if $n > 1$, then all the coefficients remain the same, except for c_{n-1} which is decreased by 1; if $n = 1$, then c_0 is decreased by one, and we add a term that is strictly less than ω^{α_0}). Thus $\{\beta_k \mid k \in \mathbb{N}\}$ is a set of ordinals that does not have a minimum, contradiction. □

A careful look at the proof shows that one can give an elementary proof of the theorem by looking at a certain lexicographic ordering on the sequence of coefficients, and proving that it is a well-order. Ordinals streamline the analysis, however. A harder problem can be obtained by looking at the *Goodstein sequence* of a number m, obtained as before but also writing the exponents in base $(k+2)$, as well as the exponents of the exponents, etc. For example, $21 = 2^4 + 2^2 + 1 = 2^{2^2} + 2^2 + 1$, so a_1 will be $3^{3^3} + 3^3$, which is approximately $7.6 \cdot 10^{12}$. The next term is approximately $1.3 \cdot 10^{154}$, etc. Still one can show that (after a very long, but still finite time), the Goodstein sequence of any number reaches zero.

7. The axiom of foundation and the cumulative hierarchy

For more practice with ordinals, let us give another example of a definition by transfinite recursion. We will use it to analyze the universe of sets.

Definition 7.1 (The cumulative hierarchy)**.** Define a class sequence $(V_\alpha)_{\alpha \in \mathrm{OR}}$ by recursion on α as follows:

- $V_0 = \emptyset$.
- $V_{\alpha+1} = \mathcal{P}(V_\alpha)$.
- $V_\delta = \bigcup_{\alpha < \delta} V_\alpha$ if δ is limit.

We let $V := \bigcup_{\alpha \in \mathrm{OR}} V$.

The letter V stands for Von Neumann, who introduced this hierarchy. It can also stands for a picture of the hierarchy itself: at the bottom, there are few sets, and then it grows wider and wider.

Lemma 7.2 (Basic properties of the cumulative hierarchy)**.**

(1) For any ordinal α, V_α is a set-initial segment.
(2) For any ordinals $\alpha \leq \beta$, $V_\alpha \subseteq V_\beta$.
(3) For any ordinal α, $\alpha \subseteq V_\alpha$.
(4) Any set is in V_α for some α. In other words, $\mathrm{SET} = V$.

Proof. We leave the first three parts as exercises. For the fourth, suppose this is not true. By the axiom of foundation, $(\mathrm{SET}, \in)$ is wellfounded, so we can take an ϵ-minimal set x so that $x \notin V$. By minimality, for each $y \in x$ there exists $\alpha = \alpha_y$ so that $y \in V_\alpha$. Let $\beta := \sup_{y \in x} \alpha_y$. We have that $y \in V_\beta$ for all $y \in x$, so $x \subseteq V_\beta$, so $x \in V_{\beta+1}$, a contradiction. □

Note that even if the axiom of foundation is false, we could just ignore all the bad sets that are not in V: all the sets we have constructed so far are definitely in V, and V is closed under all the constructions we have. This is a strong argument for the axiom of foundation.

Definition 7.3. For any set x, define its *foundation rank* to be the minimal α such that $x \in V_{\alpha+1}$.

Example 7.4. The foundation rank of $\emptyset$ is 0, while that of $\{\{\emptyset\}\}$ is 2, that of ω is ω, and that of $\{43, \omega, \{\omega\}\}$ is $\omega + 2$. In general the foundation rank of an ordinal α is exactly α, and the foundation rank of V_α itself is also α (exercise!). The foundation rank of $\mathbb{R}$ is of the form $\omega + k$ for some positive natural number k depending on the exact definition adopted. For many set-theoretical purposes, we can identify $\mathbb{R}$ with $\mathcal{P}(\mathbb{N})$, whose foundation rank is $\omega + 1$.

The V_α hierarchy grows extremely quickly. For finite $n < \omega$, the cardinality of V_n is a tower of 2's of height n. V_ω is countable, but $V_{\omega+1}$ is not. Sometimes, V_ω is called the universe of arithmetic, since anything there is finite, and studying it is analogous to studying $(\mathbb{N}, +, \cdot)$ without being able to quantify over the subsets of $\mathbb{N}$ (in a sense, they are "proper classes" if we cut off the universe of sets at V_ω). $V_{\omega+1}$ is the universe of *second-order* arithmetic, since studying it is analogeous to studying $(\mathcal{P}(\mathbb{N}), \mathbb{N}, +, \cdot)$ (if we identify $\mathbb{R}$ with $\mathcal{P}(\mathbb{N})$, we are doing "elementary analysis", i.e. we do not have access to fancy subsets of reals, just to simple ones that we can explicitly define, e.g. the set of all positive reals, all intervals of reals, or more generally any Borel set). At $V_{\omega+\omega}$, we have reached the universe of "ordinary mathematics": all the reals, functions from reals to reals, function spaces of such functions, functions between function spaces, etc. are all in $V_{\omega+\omega}$. Note we are still only at a low countable stage: we have barely begun! For example, no uncountable ordinal shows up in $V_{\omega+\omega}$ (and indeed in any V_α for α countable).

In fact, it is a recurrent theme in set theory that one can use (and often needs) higher level to study the lower ones. For example, we used ordinals around ω^ω to study weak Goodstein sequences, very finite objects. For the regular Goodstein sequences, one needs ordinals around $\epsilon_0 = \omega^{\omega^{\omega^{\cdots}}}$.

8. Cardinals

The next goal is to build a theory of sizes of sets. It is convenient to us ordinals for that purpose. For example, the finite ordinals (i.e. the natural numbers) each correspond to a distinct number of elements. This is not so for the infinite ordinals. For example, $\omega + 1$ is in bijection with ω (send ω to 0, any natural number n to $n + 1$). In fact, there are a great many distinct countable ordinals (uncountably many!): $\omega + 2, \omega + \omega$, $\omega \cdot \omega$, ω^ω, are all examples (can you see why? Go back to the definitions of the operations and the property of countable sets if you are stuck). Nevertheless, some ordinals are uncountable, and we can look at the least one as defining the first uncountable size. More generally:

Definition 8.1. An ordinal α is a *cardinal* if for any $\beta < \alpha$, there is no bijection from β onto α.

Thus any natural number n is a cardinal (prove it by induction!) and ω itself is a cardinal. However $\omega + 1$ is not, as observed.

We will most of the time use letters such as μ, κ, θ, λ when we want to think of ordinals as cardinals, and $\alpha, \beta, \gamma, \delta$ for ordinals.

We have seen that any well-ordering is (uniquely) isomorphic to an ordinal. However given a set (like $\mathbb{Z}$), we can put many different well-orderings on it. For example, we could list it as $0, 1, 2, \ldots, -1, -2, -3, \ldots$. This gives a well-ordering that collapses to $\omega+\omega$. We could on the other hand list it as $0, -1, 1, -2, 2, -3, 3, \ldots$, which collapses to ω. When looking at cardinals, we are interested only about the "number of elements" in the set, and not about how exactly the set is ordered. Thus it makes sense to look at the least possible order type (recall that the *order type* of a well-ordering is defined to be the ordinal isomorphic to it):

Definition 8.2. The *cardinality* of a set X, written $|X|$, is the minimal ordinal α such that there is a well-ordering $<$ on X so that $(X, <) \cong (\alpha, \in)$.

We have just seen that $|\mathbb{Z}| \leq \omega$, and it is clear that $|\mathbb{Z}| > n$ for any natural number n, hence $|\mathbb{Z}| = \omega$. We will deal with the question of whether there even *is* a way of ordering an arbitrary set X soon. For now we observe that the cardinality is indeed a cardinal. It relies on the following easy result:

Lemma 8.3. If $f : \alpha \to X$ is a bijection, then there is a well-ordering on X isomorphic to α.

Proof. Set $x < y$ if and only if $f^{-1}(x) \in f^{-1}(y)$. □

Lemma 8.4. For any set X, $|X|$ is a cardinal.

Proof. Let α be an ordinal that is not a cardinal. Let $f : \beta \to \alpha$ be a bijection with $\beta < \alpha$. Let $(X, <)$ be a well-ordering isomorphic to $(\alpha, \in)$ via the isomorphism $\pi : X \to \alpha$. Then $f \circ \pi^{-1}$ is a bijection, hence by Lemma 8.3 there is a well-ordering of X of type β. □

To finish making sense of the definition of cardinality, we have to show that for any set X there is a well-ordering on X. This is not so easy: how do you well-order $\mathbb{R}$? The intuition is that you can build your well-order inductively, picking element after element. This "picking" of course uses the axiom of choice.

Theorem 8.5 (The well-ordering theorem)**.** For any set X, there is a well-ordering on X. In particular, $|X|$ is well-defined.

Proof. Fix a global choice function $F : \text{SET} \to \text{SET}$. We define a class sequence $(a_\alpha)_{\alpha \in \text{OR}}$ by induction on α as follows: given $(a_\beta)_{\beta < \alpha}$, let $a_\alpha := F(X - \{a_\beta \mid \beta < \alpha\})$.

We claim that there is an ordinal α such that $X - \{a_\beta \mid \beta < \alpha\} = \emptyset$. If not, this means that $a_\alpha \in X$ for all α, and moreover $a_\alpha \neq a_\beta$ for all $\alpha \neq \beta$. Thus the class function $f : \text{OR} \to X$ given by $f(\alpha) = a_\alpha$ is an injection, so (by an assignment problem) there must be a surjection $g : X \to \text{OR}$. Since X is a set, the axiom of replacement implies that OR is a set, contradicting Corollary 6.12.

Let γ be minimal such that $X - \{a_\beta \mid \beta < \gamma\} = \emptyset$. This means, by construction, that $X = \{a_\alpha \mid \alpha < \gamma\}$. Moreover, the construction ensures also the a_α's are all distinct. Thus the map $\alpha \mapsto a_\alpha$ is a bijection from γ onto X, and hence induces a well-ordering (Lemma 8.3). □

You will see in the assignments that in fact the axiom of choice is equivalent to a global form of the well-ordering theorem:

Theorem 8.6. There is a class well-ordering on SET. In fact, there is a bijection from OR onto SET.

Proof. An assignment problem. □

You will also explore another equivalent statement, Zorn's lemma. Let us now explore properties of cardinals, that we will use without comments.

Definition 8.7. Let CARD be the class of all cardinals. We order it with the restriction of the ordering on OR.

Lemma 8.8 (Basic properties of cardinals)**.**

(1) $(\text{CARD}, <)$ is a class well-ordering.
(2) $|\alpha| \leq \alpha$, for any ordinal α.
(3) $||X|| = |X|$ for any set X.
(4) For sets X and Y, the following are equivalent:
  (a) $|X| \leq |Y|$.
  (b) There is an injection from X to Y.
  (c) $X = \emptyset$ or there is a surjection from Y to X.
(5) $|X| = |Y|$ if and only if there is a bijection from X to Y.
(6) CARD is unbounded: for any ordinal α there is a cardinal $\lambda > \alpha$. In particular, CARD is a proper class.

Proof.

(1) Because $\text{CARD} \subseteq \text{OR}$, and $(\text{OR}, <)$ is a class well-ordering.
(2) Immediate from the definition: $(\alpha, \in)$ is a well-ordering of α of type α.
(3) Also immediate from the definition of a cardinal.
(4) An assignment problem.
(5) First observe that if $\lambda = |X| = |Y|$, then by definition there is a bijection f from X onto λ, and a bijection g from Y onto λ. Thus $g^{-1}f$ is a bijection from X onto Y. Conversely, if there is a bijection from X to Y, then there is an injection from X to Y and a surjection from X to Y. By the previous part, $|X| \leq |Y|$ and $|Y| \leq |X|$, so $|X| = |Y|$.
(6) The "in particular" part follows from assignment 3. Fix an ordinal α, and consider $X := \mathcal{P}(\alpha)$. By Cantor's theorem, there is no surjection of α onto X, hence there cannot be a surjection of α onto $\lambda := |X|$. By the previous part (and as $X \neq \emptyset$), this shows that $\alpha < \lambda$.

□

Let us now revisit Section 3. We defined there a set X to be *infinite* if there is an injection from $\mathbb{N}$ into X, and *finite* otherwise. We also defined X to be *countable* if $X = \emptyset$ or there is a surjection from $\mathbb{N}$ onto X. We can now prove the expected equivalences:

Corollary 8.9.

(1) A set X is infinite if and only if $|X| \geq \omega$, and finite if and only if $|X| < \omega$ (i.e. X is in bijection with a natural number)

(2) A set X is countable if and only if $|X| \leq \omega$.

Proof.

(1) Note that $|\mathbb{N}| = \omega$. Thus for a set X there is an injection from $\mathbb{N}$ into X if and only if $\omega \leq |X|$ (by the basic properties). The characterization of finiteness immediately follows.
(2) Also immediate from the basic properties.

□

We can also compute the cardinality of the reals. We make key use of the fact that to show two sets have the same cardinality it is enough to exhibit an injection and a surjection separately, rather than build a single bijection.

Example 8.10. $|[0,1]| = |\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

Proof. First observe that $|\mathcal{P}(\mathbb{N})| = |{}^{\mathbb{N}}2|$, via the bijection sending each set to its characteristic function. Next, the function $s \mapsto \sum_{k=0}^{\infty} s(k)2^{-k-1}$ gives a surjection of ${}^{\mathbb{N}}2$ onto $[0,1]$, so $|\mathcal{P}(\mathbb{N})| \geq |[0,1]|$ (this map is not a bijection, as the sequence $1,0,0,\ldots$ is sent to the same place as $0,1,1,1,\ldots$). On the other hand, the function $s \mapsto \sum_{k=0}^{\infty} s(k)2^{-2k-2}$ is an injection of ${}^{\mathbb{N}}2$ into $[0,1]$, so $|\mathcal{P}(\mathbb{N})| \leq [0,1]$. This shows that $|[0,1]| = |\mathcal{P}(\mathbb{N})|$. We have seen previously that the open interval $(0,1)$ is in bijection with (in fact order-isomorphic to) $\mathbb{R}$, so $|(0,1)| = |\mathbb{R}|$. Since trivially $|(0,1)| \leq |[0,1]| \leq |\mathbb{R}|$, we conclude that $|[0,1]| = |\mathbb{R}|$. We will see another proof of this using cardinal arithmetic later (Example 8.22). □

It is a basic fact, used implicitly several times already, that any class of ordinals can be enumerated in increasing order. More precisely:

Lemma 8.11. Let A be a class of ordinals.

- If A is a proper class, then there exists a unique class sequence $(a_\alpha)_{\alpha \in \mathrm{OR}}$ such that $A = \{a_\alpha \mid \alpha \in \mathrm{OR}\}$ and $\alpha < \beta$ implies $a_\alpha < a_\beta$.
- If A is a set, there exists a unique ordinal δ (called the *order type of* A) and a unique sequence $(a_\alpha)_{\alpha < \delta}$ such that $A = \{a_\alpha \mid \alpha \in \delta\}$ and $\alpha < \beta < \delta$ implies $a_\alpha < a_\beta$.

Proof. This is a straightforward consequence of the Mostowski collapsing theorem. Alternatively, one can construct the sequence by hand by taking a_0 to be the minimal element of A, a_1 to be the minimal element of $A - \{a_0\}$, and so on (by induction). □

The lemma above applies, in particular, to the class CARD of all cardinals. Explicitly, we can define the αth infinite cardinal. First, the following gives the successor operation for cardinals:

Definition 8.12. For a cardinal λ, let λ^+ be the minimal cardinal μ such that $\mu > \lambda$. We call λ the *(cardinal) successor of* λ.

Definition 8.13. By induction on α, define[5] $\aleph_\alpha$, the *αth infinite cardinal*, as follows:

(1) $\aleph_0 = \omega$.
(2) $\aleph_{\gamma+1} = (\aleph_\gamma)^+$.

[5] The letter $\aleph$ is pronounced "aleph", it is the first letter of the Hebrew alphabet.

(3) $\aleph_\delta = \sup_{\gamma < \delta} \aleph_\gamma$.

We have implicitly used:

Lemma 8.14. If X is a set of cardinals, then $\sup(X)$ is also a cardinal.

Proof. Exercise (suppose not...). □

Note that each cardinal is also an ordinal. When we want to think of $\aleph_\alpha$ as an ordinal, we will write ω_α instead of $\aleph_\alpha$ (one exception: we always write ω instead of ω_0). Thus the cardinals look like:

$$0, 1, 2, \ldots, \aleph_0, \aleph_1, \ldots, \aleph_\omega, \ldots, \aleph_{\omega+\omega}, \ldots, \aleph_{\omega_1}, \ldots, \aleph_{\omega_2}, \ldots \aleph_{\omega_\omega}, \ldots$$

You should be able to argue that any infinite cardinal λ is of the form $\aleph_\alpha$ for some α (here is one way: argue that $\lambda \leq \aleph_\lambda$ and take α least so that $\lambda \leq \aleph_\alpha$).

8.1. **Cardinal arithmetic.** We now define addition, multiplication, and exponentiation *on cardinals.* These operations are different from the ones defined on ordinals (remember that we only care about sizes). Thus we expect to have $\aleph_0 + 1 = \aleph_0$, since adding one element to a countable set keeps the set countable. In fact, we will more generally define infinite sums and infinite products, and for this the following definition will be convenient:

Definition 8.15. Let $(A_i)_{i \in I}$ be any class sequence. Then define $\bigsqcup_{i \in I} A_i$ (the *disjoint union* of the A_i's) by:

$$\bigsqcup_{i \in I} A_i = \bigcup_{i \in I} (\{i\} \times A_i).$$

Also define $A \sqcup B := \bigcup_{i \in \{0,1\}} A_i$, where $A_0 = A$, $A_1 = B$.

Definition 8.16 (Cardinal addition, multiplication, and exponentiation)**.**

- If $(\lambda_i)_{i \in I}$ is a sequence of cardinals, define $\sum_{i \in I} \lambda_i := |\bigsqcup_{i \in I} \lambda_i|$, and $\prod_{i \in I} \lambda_i := |\prod_{i \in I} \lambda_i|$.
- If λ and μ are cardinals, define $\lambda + \mu$ to be $\sum_{i \in \{0,1\}} \lambda_i$, where $\lambda_0 = \lambda$, $\lambda_1 = \mu$, and similarly for $\lambda \cdot \mu$.
- If λ and μ are cardinals, let λ^μ (read "λ to the power μ") be the cardinality of ${}^\mu\lambda$ (the set of functions from μ to λ).

Note that we have abused notation in the definition of the product: the symbol $\prod$ on the left and right hand side of the equation do not stand for the same thing. No confusion should result, but it is good to be aware of the difference.

It is also important to remember that we have overloaded the operators: for example (as we will see) $\aleph_0 + \aleph_0 = \aleph_0$, whereas for the $+$ of *ordinal* arithmetic, $\omega + \omega \neq \omega$. Thus it is important to know, in a given context, whether $+$ denotes the addition of ordinals or of cardinals. It is usually clear from context which is meant (based on the use of different letters for cardinals and ordinals). Interestingly, however, ordinal and cardinal arithmetic coincide for finite numbers.

Note that the definition, for example of a sum, is not sensitive to the exact choice of sets with the given cardinality: $\lambda + \mu$ is just the cardinality of $A \cup B$ where A and B are any two disjoint sets of size λ and μ respectively (play around with bijections to see why). In more details, we will use the following without comments:

Remark 8.17.

- If $(X_i)_{i\in I}$ is a sequence of disjoint sets, and $|X_i| = \lambda_i$, then $\sum_{i\in I}\lambda_i = |\bigcup_{i\in I} X_i|$. Similarly, if $(Y_i)_{i\in I}$ is any sequence of sets and $|Y_i| = \mu_i$, then $\prod_{i\in I}\mu_i = |\prod_{i\in I} Y_i|$.
- If $|X| = \lambda$ and $|Y| = \mu$, then $\lambda^\mu = |{}^Y X|$.

Lemma 8.18 (Basic properties of cardinal arithmetic)**.**

(1) Addition and multiplication of cardinals are associative and commutative (in fact any reindexing of sums or products give the same result).
(2) (Relationship between sum and product) $\sum_{i\in I}\lambda = |I|\cdot\lambda$.
(3) (Relationship between product and exponentiation) $\prod_{i\in I}\lambda = \lambda^{|I|}$.
(4) $(\lambda^\mu)^\theta = \lambda^{\mu\cdot\theta}$.
(5) $\lambda + 0 = \lambda$, $\lambda^+ = \lambda + 1$ if λ is finite, $\lambda\cdot 0 = 0$, $\lambda\cdot 1 = \lambda$, $\lambda^0 = 1$, $\lambda^1 = \lambda$, $\sum_{i\in\emptyset}\lambda_i = 0$, $\prod_{i\in\emptyset}\lambda_i = 1$.
(6) (Monotonicity) If $\mu_i \le \lambda_i$ for all $i \in I$, then $\sum_{i\in I}\mu_i \le \sum_{i\in I}\lambda_i$, $\prod_{i\in I}\mu_i \le \prod_{i\in I}\lambda_i$. If in addition $\mu_i \ge 2$ for all $i \in I$, then $\sum_{i\in I}\mu_i \le \prod_{i\in I}\mu_i$.
(7) $|\mathcal{P}(\lambda)| = 2^\lambda$. Thus $\lambda < \lambda^+ \le 2^\lambda$.

Proof. All are very easy. The equality $|\mathcal{P}(\lambda)| = 2^\lambda$ follows from identifying subsets of λ with their characteristic function, and the fact that $\lambda < 2^\lambda$ is then immediate from Cantor's theorem. By definition, λ^+ is the minimal cardinal strictly above λ, so $\lambda < \lambda^+ \le 2^\lambda$ is clear.

We also prove that $(\lambda^\mu)^\theta = \lambda^{\mu\cdot\theta}$, since we will use that often. For this, we give a bijection from ${}^\theta({}^\mu\lambda)$ onto ${}^{\mu\times\theta}\lambda$. Given as input a function $f : \theta \to {}^\mu\lambda$, send it to the function $g_f : \mu\times\theta \to \lambda$ given by $g_f(\alpha,\beta) = f(\alpha)(\beta)$. It's a straightforward unpacking of the definitions to see that $f \mapsto g_f$ gives a bijection (called "currying" by programming language experts). □

Infinite cardinal arithmetic is, in a sense, much easier than finite cardinal arithmetic (note that the result below give in particular another proof that $\aleph_0\cdot\aleph_0 = \aleph_0$, i.e. that $\mathbb{N}\times\mathbb{N}$ is countable):

Theorem 8.19 (Binary addition and multiplication are simple)**.** For any infinite cardinals λ and μ, $\lambda + \mu = \lambda\cdot\mu = \max(\lambda,\mu)$.

Proof. Assume without loss of generality that $\lambda \le \mu$. It is clear that $\mu \le \lambda + \mu \le \lambda\cdot\mu \le \mu\cdot\mu$, so we will be done once we have established that $\mu\cdot\mu \le \mu$.

We prove more precisely by induction on a (possibly finite) cardinal μ that $\mu\cdot\mu < \max(\mu^+,\aleph_0)$. If μ is finite, then $\mu\cdot\mu$ and μ^+ are also finite, hence strictly less than $\aleph_0$. Assume now that μ is infinite and the result is true for all cardinals $\theta < \mu$. We need to find a well-order of $\mu\times\mu$ of type at most μ. First, let $<_{\text{lex}}$ denote the lexicographic ordering on $\mu\times\mu$. We use it to define another ordering $\prec$ on $\mu\times\mu$ as follows: $(\alpha,\beta) \prec (\alpha',\beta')$ if and only if either $\max(\alpha,\beta) < \max(\alpha',\beta')$ or $\max(\alpha,\beta) = \max(\alpha',\beta')$ and $(\alpha,\beta) <_{\text{lex}} (\alpha',\beta')$. It is easy to check that $\prec$ is a strict linear ordering. It is also a well-ordering: if A is a non-empty subset of $\mu\times\mu$, first find $\gamma := \min\{\max(\alpha,\beta) \mid (\alpha,\beta)\in A\}$, then find the least element of $\{(\alpha,\beta)\in A \mid \max(\alpha,\beta) = \gamma\}$ according to the lexicographic ordering (which you have shown in assignments to be a well-ordering).

We claim that $(\mu\times\mu,\prec)$ is isomorphic to $(\mu,\in)$. Suppose not. Then $(\mu\times\mu,\prec) \cong (\rho,\in)$ for some $\rho > \mu$. Let π be the isomorphism and let $(\alpha,\beta) := \pi^{-1}(\mu)$ (we use here that $\rho > \mu$). Observe that, by definition of the ordering, any pair (α',β')

which is $\prec$-below (α, β) must satisfy $\max(\alpha', \beta') \leq \gamma_0$, where $\gamma_0 := \max(\alpha, \beta)$. Let $\gamma := \gamma_0 + 1$. If γ_0 is finite, then $\gamma + 1$ is finite too so $\gamma < \mu$. If γ_0 is infinite, then by the induction hypothesis $|\gamma| = |\gamma_0 + 1| = |\gamma_0| + 1 \leq |\gamma_0||\gamma_0| = |\gamma_0| \leq \gamma_0 < \mu$.

By definition of γ, we have that $\mathrm{pred}_{\prec}((\alpha, \beta)) \subseteq \gamma \times \gamma$. As $\gamma < \mu$, we must have that $\mathrm{pred}_{\prec}((\alpha, \beta))| \leq |\gamma \times \gamma| = |\gamma| \cdot |\gamma|$. Since $\gamma < \mu$, $|\gamma| < \mu$, hence by the induction hypothesis, $|\gamma||\gamma| < \max(|\gamma|^+, \aleph_0) \leq \mu$. Thus (α, β) has strictly less than μ-many predecessors. This is a contradiction, since π was an isomorphism and μ has exactly μ-many predecessors in the ordinals. □

Note that this implies we cannot reach a cardinal from below using *ordinal* addition and multiplication. We will use this without comments:

Corollary 8.20. Let λ be an infinite cardinal. Then:

(1) If $\alpha, \beta < \lambda$, then $\alpha + \beta, \alpha \cdot \beta < \lambda$. In particular, λ is a limit ordinal.
(2) If $\alpha < \lambda$, $\{\beta < \lambda \mid \beta \geq \alpha\}$ has order type (and in particular cardinality) λ.

Proof.

(1) Note that, from the characterization of ordinal addition and multiplication in terms of concatenation and cartesian product of linear orders, $|\alpha + \beta| = |\alpha| + |\beta|$ and $|\alpha\beta| = |\alpha||\beta|$ (the operations on the left hand side of the equations are the ordinal ones, but on the right hand side they are the cardinal ones). If $\lambda = \aleph_0$, then α and β are finite, so $|\alpha| + |\beta|$, $|\alpha||\beta|$ are also finite. Otherwise, we may assume by taking α, β bigger if needed that they are both infinite and then we have that $|\alpha| + |\beta| = |\alpha||\beta| = \max(|\alpha|, |\beta|)$. Since $|\alpha|, |\beta| < \lambda$, we are done. The "in particular" part follows by taking $\beta = 1$.
(2) Let $S := \{\beta < \lambda \mid \beta \geq \alpha\}$. Check that the map $\gamma \mapsto \gamma + \alpha$ is an order isomorphism from $(\lambda, \in)$ onto $(S, \in)$.

□

Corollary 8.21 (Infinite sums are simple too)**.** If $(\lambda_i)_{i \in I}$ is a sequence of infinite cardinals, then:

$$\sum_{i \in I} \lambda_i = \max(|I|, \sup_{i \in I} \lambda_i)$$

Proof. Exercise. □

Example 8.22. $|\mathbb{R}| = |[0, 1]| = 2^{\aleph_0}$.

Proof. We have seen already (Example 8.10) that $|[0, 1]| = |\mathcal{P}(\mathbb{N})| = 2^{\aleph_0}$. Therefore $2^{\aleph_0} \leq |\mathbb{R}|$. It suffices to see that $|\mathbb{R}| \leq 2^{\aleph_0}$. Note that $|[n, n + 1]| = |[0, 1]|$ for any integer n (take the bijection $x \mapsto x + n$). Moreover, $\mathbb{R} = \bigcup_{n \in \mathbb{Z}} [n, n + 1]$. Therefore we have:

$$|\mathbb{R}| = \left| \bigcup_{n \in \mathbb{Z}} [n, n + 1] \right| \leq \sum_{n \in \mathbb{Z}} |[n, n + 1]| = \sum_{n \in \mathbb{Z}} |[0, 1]| = |\mathbb{Z}||[0, 1]| = \aleph_0 \cdot 2^{\aleph_0} = 2^{\aleph_0}$$

□

Example 8.23 (Some fun with exponentiation)**.** $2^{\aleph_0} = \aleph_0^{\aleph_0} = \aleph_1^{\aleph_0}$.

Proof. By Cantor's theorem, $\aleph_1 = (\aleph_0)^+ \leq 2^{\aleph_0}$. Thus:

$$2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq \aleph_1^{\aleph_0} \leq \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0}$$

The left and right hand sides are the same, so all the inequalities must be equalities. □

It is worth pointing out that not all questions in cardinal arithmetic have easy answer, and in fact cardinal arithmetic is an active area of research in set theory. For example one can ask:

Question 8.24. Is $2^{\aleph_0} = \aleph_1$?

Since we have just seen that $|\mathbb{R}| = 2^{\aleph_0}$, this is equivalent to the *continuum hypothesis*: any subset of real is either countable or of the same cardinality as that of the reals. Indeed, if $|\mathbb{R}| > \aleph_1$, it would be easy to find $X \subseteq \mathbb{R}$ so that $|X| = \aleph_1$ (for example take the first ω_1-many elements of a well-ordering of the reals). Note that the question of evaluating $2^{\aleph_0}$ is the same as that of evaluating the infinite product $2 \cdot 2 \ldots = \prod_{n \in \mathbb{N}} 2$. In particular, infinite products can be complicated to evaluate. However, they reduce to exponentiation. Just for fun, let us first look at a particular case:

Example 8.25. For any cardinal λ, define $\lambda!$ to be the product $\prod_{2 \leq \mu \leq \lambda} \mu$. That is, $\lambda! = 2 \cdot 3 \cdot \ldots \cdot \lambda$. Assume λ is infinite. We have a trivial lower bound: $\lambda! \geq \prod_{2 \leq \mu \leq \lambda} 2 = 2^\lambda$. We also have a trivial upper bound: $\lambda! \leq \prod_{2 \leq \mu \leq \lambda} \lambda = \lambda^\lambda$. If λ were finite, similar bounds would hold (we have to replace λ by $\lambda - 1$) but they are often too simple-minded and in application we may need better bounds (the ultimate ones being derived from what is called Stirling's formula). However if λ is infinite the upper and lower bounds coincide! $2^\lambda \leq \lambda^\lambda \leq \left(2^\lambda\right)^\lambda = 2^{\lambda \cdot \lambda} = 2^\lambda$. Thus $\lambda! = 2^\lambda$ and we are done. Infinite arithmetic is easier than finite arithmetic!

Theorem 8.26 (Infinite products reduce to exponentiation)**.** Let μ be an infinite cardinal and let $(\lambda_i)_{i<\mu}$ be an increasing sequence of nonzero cardinals (that is, $0 < \lambda_i \leq \lambda_j$ for any $i \leq j < \mu$). Then:

$$\prod_{i<\mu} \lambda_i = \left(\sup_{i<\mu} \lambda_i\right)^\mu$$

Proof. Write $I := \mu$. Let $\lambda := \sup_{i \in I} \lambda_i$. Clearly, $\prod_{i \in I} \lambda_i \leq \prod_{i \in I} \lambda = \lambda^{|I|} = \lambda^\mu$. It remains to see the reverse inequality. If $\lambda_i = 1$ for all $i < \mu$, then $\lambda = 1$ and the result is immediate. Thus we can assume $\lambda_i \geq 2$ for some $i < \mu$, and hence that $\lambda_j \geq 2$ for all $j \geq i$. We can always reindex by letting $\theta_j := \lambda_{i+j}$ for $j < \mu$, so assume without loss of generality that $\lambda_j \geq 2$ for all $j < \mu$.

As $\mu \cdot \mu = \mu$, there is a bijection f from $\mu \times \mu$ onto I, so we can find $(I_\alpha)_{\alpha<\mu}$ pairwise disjoint subsets of I so that $I = \bigcup_{\alpha<\mu} I_\alpha$ (take $I_\alpha = f[\{\alpha\} \times \mu]$). Note that for any α, I_α is unbounded in μ (otherwise it would not have cardinality μ). In particular, $\sup_{i \in I_\alpha} \lambda_i = \lambda$. Thus we have:

$$\prod_{i \in I} \lambda_i = \prod_{\alpha<\mu} \prod_{i \in I_\alpha} \lambda_i \geq \prod_{\alpha<\mu} \sum_{i \in I_\alpha} \lambda_i = \prod_{\alpha<\mu} \max(|I_\alpha|, \lambda) \geq \prod_{\alpha<\mu} \lambda = \lambda^\mu$$

which is as desired. □

Exercise 8.27. Assume you are given an arbitrary (not necessarily increasing) sequence $(\lambda_i)_{i\in I}$ of cardinals. Explain how to compute $\prod_{i\in I}\lambda_i$.

We end with the following aesthetically pleasing statement, which has strong consequences:

Theorem 8.28 (König's theorem)**.** Assume $(\lambda_i)_{i\in I}$, $(\mu_i)_{i\in I}$ are sequences of cardinals such that for all $i \in I$, $\lambda_i < \mu_i$. Then:

$$\sum_{i\in I}\lambda_i < \prod_{i\in I}\mu_i$$

Proof. It is easy to see that $\sum_{i\in I}\lambda_i \le \prod_{i\in I}\mu_i$ (for the corner case where $\lambda_i = 1$ for all i, we use that $\lambda_i < \mu_i$, hence $\mu_i \ge 2$). It remains to see the inequality is strict. For convenience, let $(A_i)_{i\in I}$ be a sequence of disjoint sets with $|A_i| = \lambda_i$ for all $i \in I$ and let $(B_i)_{i\in I}$ be a sequence of sets with $|B_i| = \mu_i$ for all $i \in I$. If $\sum_{i\in I}\lambda_i = \prod_{i\in I}\mu_i$, we would have a surjection $F : \bigcup_{i\in I} A_i \to \prod_{i\in I} B_i$. We do a kind of generalized diagonalization. For each $i \in I$, let $B'_i := \{b \in B_i \mid b = F(a)_i \text{ for some } a \in A_i\}$. Note that $B'_i \neq B_i$ as $|F[A_i]| \le |A_i| = \mu_i < \lambda_i = |B_i|$. Pick $b_i \in B_i - B'_i$. As F is surjective, there must exists $a \in \bigcup_{i\in I} A_i$ so that $(F(a))_i = b_i$. However, $a \in A_i$ for some $i \in I$, and so $b_i \in B'_i$, contradicting the definition of b_i. □

As a corollary, we recover Cantor's theorem:

Corollary 8.29. For any cardinal λ, $\lambda < 2^\lambda$.

Proof. $\lambda = \sum_{i<\lambda} 1$, while $2^\lambda = \prod_{i<\lambda} 2$. □

9. Regular and singular cardinals

We move toward a finer understanding of cardinals: in a sense, the cardinal $\aleph_\omega$ is "easier to reach" than $\aleph_1$: the former is the supremum of the countable set $\{\aleph_n \mid n < \omega\}$. We make that idea precise.

Definition 9.1 (Cofinality)**.** Let $P = (A, \le)$ be a partial ordering.

(1) A set $S \subseteq A$ is *cofinal* (in P) if for all $a \in A$ there exists $b \in S$ so that $a \le b$.
(2) The *cofinality* of P, written $\mathrm{cf}(P)$, is the minimal cardinal λ such that there exists a cofinal subset of A of cardinality λ.
(3) For an ordinal α, the *cofinality* of α, written $\mathrm{cf}(\alpha)$, is the cofinality of $(\alpha, \in)$.

Example 9.2. Let $P = (A, \le)$ be a partial ordering.

(1) A is a cofinal set in P. Thus $\mathrm{cf}(P) \le |A|$. In particular, if α is an ordinal $\mathrm{cf}(\alpha) \le |\alpha|$.
(2) If for every $a \in A$ there exists a maximal element $a' \in A$ with $a \le a'$, then $\mathrm{cf}(P)$ is the cardinality of the set of all maximal elements of P. Indeed, it is easy to check that the set of maximal elements of P is a cofinal set, and conversely if S is a cofinal set and $a \in A$ is maximal, there must exist $a' \in S$ so that $a \le a'$, and as a is maximal $a = a'$.
(3) In particular, if P has a single maximal element (that is, a greatest element) then $\mathrm{cf}(P) = 1$. This is the case for example, if $P = (\mathcal{P}(\mathbb{N}), \subseteq)$ or if P is any linear order with a maximum (like as a successor ordinal).

(4) As another example, consider the ordering $P = (\mathcal{P}(\mathbb{N}) - \{\mathbb{N}\}, \subseteq)$. Its set of maximal elements is $\{\mathbb{N} - \{n\} \mid n \in \mathbb{N}\}$, thus $\text{cf}(P) = \aleph_0$.
(5) If P has a finite cofinal set S, then for any $a \in A$ there exists $b \in S$ so that $a \leq b$. As S is finite, there exists $b' \in S$ maximal so that $b \leq b'$. Thus in this case every element of P is bounded by a maxima element.
(6) The usual orderings on $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ do not have maximal elements, so the previous observations do not apply, hence their cofinality is at least $\aleph_0$. On the other hand, the set $\mathbb{N}$ is a cofinal subset of $\mathbb{R}$ (hence also of $\mathbb{Z}$ and $\mathbb{Q}$), and it follows that the cofinality of all these orderings is $|\mathbb{N}| = \aleph_0$.
(7) Let α be an ordinal. We have seen already that if α is a successor ordinal then $\text{cf}(\alpha) = 1$. Of course, $\text{cf}(0) = 0$. If α is a limit ordinal, then it does not have a maximum, so $\aleph_0 \leq \text{cf}(\alpha) \leq |\alpha|$. For example, $\text{cf}(\omega + \omega) = \omega$ (even though the cofinality is always a cardinal, we may often think of it as an ordinal, so we will write either $\text{cf}(\omega + \omega) = \omega$ or $\text{cf}(\omega + \omega) = \aleph_0$). A cofinal set is given by $\omega + \omega$ itself, or more interestingly by the set $\{\omega + n \mid n < \omega\}$. We will see however that $\text{cf}(\aleph_1) = \omega_1$. On the other hand, $\text{cf}(\aleph_\omega) = \omega$, as witnessed by the countable set $\{\aleph_n \mid n < \omega\}$.

We can classify cardinals based on their cofinality:

Definition 9.3 (Regular and singular cardinals)**.** A cardinal λ is *regular* if $\text{cf}(\lambda) = \lambda$, and *singular* otherwise.

Remark 9.4. For any ordinal α, $\text{cf}(\text{cf}(\alpha)) = \text{cf}(\alpha)$. In particular, $\text{cf}(\alpha)$ is always a regular cardinal.

Note that $\text{cf}(0) = 0$ and $\text{cf}(n) = 1$ for n a positive natural number. Thus 0 and 1 are technically regular cardinals while bigger natural numbers are singulars (although these notions are really interesting only for infinite cardinals). On the other hand, $\text{cf}(\aleph_0) = \aleph_0$, so $\aleph_0$ is regular. What about $\aleph_1$? It turns out it is regular too. Since $\aleph_1 = \aleph_0^+$, this is a special case of:

Lemma 9.5. If λ is an infinite cardinal, then λ^+ is regular.

Proof. Assume for a contradiction that $A \subseteq \lambda^+$ is cofinal and has cardinality at most λ. This means that for any $\alpha \in \lambda^+$ there exists $\beta \in A$ so that $\alpha + 1 \leq \beta$, so $\alpha < \beta$. Remember however that this means $\alpha \in \beta$. Thus any element of λ^+ is contained in a member of A, so $\lambda^+ \subseteq \bigcup A$ (it is in fact easy to see equality holds). However, $|A| \leq \lambda$, and for all $\beta \in A$, $|\beta| < \lambda^+$, so $|\beta| \leq \lambda$. Thus $\lambda^+ \leq |\bigcup A| = |\bigcup_{\beta \in A} \beta| \leq \sum_{\beta \in A} |\beta| \leq |A|\lambda = \lambda \cdot \lambda = \lambda$, a contradiction. □

Let us restate the lemma in words: just like for ordinals, there are two types of nonzero cardinals: those that come from the successor operations and those that do not:

Definition 9.6. A cardinal λ is a *successor cardinal* if $\lambda = \mu^+$ for some cardinal μ. We say that λ is a *limit cardinal* if it is not zero and not a successor cardinal.

Exercise 9.7. Show that $\aleph_\alpha$ is a successor cardinal if and only if α is a successor ordinal and a limit cardinal if and only if α is a limit ordinal or zero.

Thus we have just seen that infinite successor cardinals are regular. In particular, $\aleph_{\alpha+1}$ is regular for any ordinal α, so $\aleph_1, \aleph_2, \aleph_3$ are all regulars. On the other hand we have also seen that $\aleph_\omega$ is singular: $\text{cf}(\aleph_\omega) = \aleph_0$. More generally:

Lemma 9.8.

(1) If δ is a limit ordinal, then $\text{cf}(\aleph_\delta) = \text{cf}(\delta)$.
(2) If λ is a limit cardinal of cofinality θ, then there exists a sequence $(\lambda_i)_{i<\theta}$ that is strictly increasing ($i < j < \theta$ implies $\lambda_i < \lambda_j$) and so that $\{\lambda_i \mid i < \theta\}$ is cofinal in λ. In particular, $\sup_{i<\theta} \lambda_i = \lambda$.

Proof. Assignment 5. □

Remark 9.9. In general, if $I = (A, \leq)$ is any linear order and θ is its cofinality, we can always find a strictly increasing sequence $(x_i)_{i<\theta}$ of members of A so that $\{x_i \mid i < \theta\}$ is cofinal (you will prove this in assignment 5). This implies in particular that the cofinality of any linear order is a regular cardinal (but the cofinality may be singular for partial orders! Can you think of an example?)

Does that mean that all limit cardinals are singulars? Not quite since $\aleph_0$ is regular. Are there other regular limits? Let us consider what a regular *uncountable* limit cardinal λ should look like. It must be of the form $\lambda = \aleph_\delta$ for a limit ordinal δ, and by regularity we must have that $\lambda = \text{cf}(\lambda) = \text{cf}(\delta)$. Since clearly $\text{cf}(\delta) \leq \delta \leq \lambda$, the only possibility is that $\delta = \lambda$. In particular, $\lambda = \aleph_\lambda$ (it is an $\aleph$ *fixed point*). Note one can construct *singular* cardinals μ so that $\mu = \aleph_\mu$ (exercise), but here we are looking for a cardinal like this that is also regular. Such cardinals are called *large cardinals* and their existence cannot be proven with our axioms (though it is generally believed that assuming they exist does not yield to a contradiction either).

In a sense, regular cardinals are nicer to deal with than singular cardinals: they are closed under more operations (if you have a regular uncountable cardinal λ, you can take a the supremum of a countable set of ordinals below it and still be below λ). As we will see, singular cardinals also exhibit interesting combinatorial behavior, however.

9.1. **More cardinal arithmetic.** We have seen how to compute the cofinality of $\aleph_\alpha$ for any α. Given that we do not know for which α we have $2^{\aleph_0} = \aleph_\alpha$, it makes sense to ask whether we can still compute the cofinality of $2^{\aleph_0}$. We will give a partial answer: this cofinality has to be uncountable. This turns out to be (provably) the best that can be done from our axioms. We first prove a statement that is essentially equivalent to König's theorem.

Theorem 9.10 (König's theorem, second version)**.** For any infinite cardinal λ, $\lambda^{\text{cf}(\lambda)} > \lambda$.

Proof. Let $\theta := \text{cf}(\lambda)$. If λ is a successor cardinal, then λ is regular (Lemma 9.5) so $\theta = \lambda$, and so we just want to prove that $\lambda^\lambda > \lambda$. This holds by Cantor's theorem. Assume now that λ is a limit cardinal. From Lemma 9.8, we get a strictly increasing sequence $(\lambda_i)_{i<\theta}$ of cardinals below λ such that $\sup_{i<\theta} \lambda_i = \lambda$. By the first version of König's theorem (Theorem 8.28):

$$\sum_{i<\theta} \lambda_i < \prod_{i<\theta} \lambda$$

The product on the right hand side is equal to $\lambda^\theta = \lambda^{\text{cf}(\lambda)}$. The sum on the left hand side is, by Corollary 8.21, equal to $\max(\theta, \sup_{i<\theta} \lambda_i) = \lambda$. Thus we obtain the desired result. □

Corollary 9.11. For any infinite cardinal λ, $\text{cf}(2^\lambda) > \lambda$. In particular, $2^{\aleph_0}$ has uncountable cofinality.

Proof. Let $\mu := 2^\lambda$. If $\text{cf}(\mu) \leq \lambda$, then we would get that $\mu < \mu^{\text{cf}(\mu)} \leq \mu^\lambda = \left(2^\lambda\right)^\lambda = 2^{\lambda \cdot \lambda} = 2^\lambda = \mu$, a contradiction. □

Note that, as $\text{cf}(2^\lambda) \leq 2^\lambda$, we have obtained yet another proof of Cantor's theorem (at least for infinite cardinals). In conclusion, we have the following restrictions on the class function $\lambda \mapsto 2^\lambda$:

- Monotonicity: if $\lambda_1 \leq \lambda_2$, then $2^{\lambda_1} \leq 2^{\lambda_2}$.
- König's theorem: $\text{cf}(2^\lambda) > \lambda$.

It turns out that, *if λ is regular*, these are the only restrictions one can prove. In particular, $2^{\aleph_0}$ can consistently be any cardinal of uncountable cofinality.

This does not mean that the subject of cardinal arithmetic is exhausted: there remains the very interesting case of singular cardinals, and furthermore one can also ask more generally about the function λ^θ for arbitrary λ and θ. We have seen that computing any infinite product reduces to this function.

For now, let us prove a simple result about exponentiation of singular cardinals. To state it easily, we will use the following notion:

Definition 9.12. For X any set and α an ordinal, let ${}^{<\alpha}X := \bigcup_{\beta < \alpha} {}^{\beta}X$. For λ and μ cardinals, define $\lambda^{<\mu}$ (read "λ to the weak power of μ") to be $|{}^{<\mu}\lambda|$.

Exercise 9.13 (Basic properties of the weak power)**.** Let λ and θ be cardinals.

(1) $\lambda^{<0} = 0$.
(2) If $\theta > 0$, then $0^{<\theta} = 1$.
(3) $1^{<\theta} = \theta$. In particular, $\lambda^{<\theta} \geq \theta$ if $\lambda \geq 1$.
(4) If $\lambda \geq 2$, then $\lambda^{<\theta} \leq \lambda^\theta$.
(5) If θ is infinite, $\lambda^{<\theta^+} = \lambda^\theta$
(6) $\lambda^{<\theta} = \sum_{\kappa < \theta} \lambda^\kappa$.
(7) $\lambda^{<\aleph_0} = \lambda + \aleph_0$.

Theorem 9.14. For any cardinal $\lambda \geq 2$ and any infinite cardinal θ:

$$\lambda^\theta = \left(\lambda^{<\theta}\right)^{\text{cf}(\theta)}$$

Proof. One inequality is easy: $\left(\lambda^{<\theta}\right)^{\text{cf}(\theta)} \leq \left(\lambda^\theta\right)^\theta = \lambda^{\theta \cdot \theta} = \lambda^\theta$. It remains to see the other inequality. Let $\delta := \text{cf}(\theta)$ and let $\{\alpha_i \mid i < \delta\}$ be a cofinal set, enumerated in increasing order.

The map $f \mapsto (f \upharpoonright \alpha_i)_{i<\delta}$ gives an injection from ${}^\theta\lambda$ into $\prod_{i<\delta} {}^{\alpha_i}\lambda$. The latter product clearly has cardinality at most $\left(\lambda^{<\theta}\right)^\delta$, so we are done. □

Note that the equation above is true but useless if θ is regular. However if θ is singular we deduce:

Corollary 9.15. For any cardinal $\lambda \geq 2$ and any infinite singular cardinal θ, if there is a cardinal μ such that $\lambda^\kappa = \mu$ for unboundedly-many $\kappa < \theta$, then we also have that $\lambda^\theta = \mu$.

Proof. Let $\delta := \text{cf}(\theta)$. As θ is singular, $\delta < \theta$. Using Theorem 9.14, $\lambda^\theta = \left(\lambda^{<\theta}\right)^\delta$. Now $\lambda^\kappa \geq \kappa$, so $\lambda^{<\theta} \geq \theta$, hence by assumption $\lambda^{<\theta} = \sum_{\kappa<\theta} \lambda^\kappa = \theta \cdot \mu = \mu$. We know that for some $\kappa > \delta$, $\lambda^\kappa = \mu$ (we are using here that $\delta < \theta$), so $\lambda^\theta = \mu^\delta = (\lambda^\kappa)^\delta = \lambda^{\kappa \cdot \delta} = \lambda^\kappa = \mu$. □

Example 9.16. If $2^{\aleph_n} = \aleph_{\omega+1}$ for any $n < \omega$, then $2^{\aleph_\omega} = \aleph_{\omega+1}$. To see this, set $\lambda := 2$, $\theta := \aleph_\omega$, and $\mu := \aleph_{\omega+1}$ in the above.

10. The Borel hierarchy

We leave cardinal arithmetic for now and do real analysis instead. Since we are doing set theory, we will study *sets* of real numbers. The continuum hypothesis is essentially about understanding arbitrary sets of reals, so let us start with the simple ones:

Definition 10.1. A set X of real numbers is *open* if for every $x \in X$ there exists $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subseteq X$. Equivalently, X is a union of open intervals. A set is *closed* if it is the complement of an open set.

Observe that the empty set and the set of all reals are open and closed. Arbitrary unions and finite intersections of open sets are open. Dually, arbitrary intersections and finite unions of closed sets are closed. It is easy to check that the point $\{x\}$ is closed for any real number x, so any set can be built from an arbitrary union of closed sets. What if, however, we look at sets that can be built from countable union of closed sets? We obtain what is called an F_σ set. We can then take countable intersections of those, and continue building a transfinite hierarchy of sets. Sets obtained from open sets using complement, countable intersections and countable unions are called Borel sets:

Definition 10.2. A *σ-algebra on a set* X is a set $\mathcal{A}$ of subsets of X such that:

- $X \in \mathcal{A}$.
- If $S \in \mathcal{A}$, then $X - S \in \mathcal{A}$ (that is, $\mathcal{A}$ is *closed under complement*).
- If $(S_n)_{n \in \mathbb{N}}$ is a countable sequence of members of $\mathcal{A}$,then $\bigcup_{n \in \mathbb{N}} S_n \in \mathcal{A}$ (that is, $\mathcal{A}$ is *closed under countable unions*).

Note that a σ-algebra is closed under countable intersections as well (by De Morgan's law). Of course, $\mathcal{P}(\mathbb{R})$ is a σ-algebra on $\mathbb{R}$. Moreover, it is easy to check that an arbitrary intersection of σ-algebra is a σ-algebra. Thus we define:

Definition 10.3. The *Borel σ-algebra* on $\mathbb{R}$ is the intersection of all σ-algebras on $\mathbb{R}$ that contain the open sets. A set of reals is *Borel* if it belongs to the Borel σ-algebra. We denote the set of all Borel sets by $\mathcal{B}$.

In other words, the Borel σ-algebra is the smallest σ-algebra containing the open sets, and a set is Borel if it can be generated from an open set using complement and countable unions (maybe repeated infinitely-many times).

The Borel sets have some interesting characterizations. Lebesgue and Baire were interested in studying "constructive" definitions of functions, and so they looked at the smallest class of functions from $\mathbb{R}$ to $\mathbb{R}$ containing the continuous functions and closed under composition and taking pointwise limits. The Borel sets are exactly the sets that can be inverse images of open intervals using such functions.

To analyze exactly how $\mathcal{B}$ is generated, we define the following hierarchy:

Definition 10.4 (The Borel hierarchy)**.** By induction on α, we define two class sequences $\left(\mathbf{\Sigma}^0_\alpha, \mathbf{\Pi}^0_\alpha\right)_{\alpha \in \mathrm{OR} - \{0\}}$ as follows:

- $\boldsymbol{\Sigma}^0_1$ is the set of all open sets of reals.
- Given $\boldsymbol{\Sigma}^0_\alpha$, $\boldsymbol{\Pi}^0_\alpha$ is the set $\{\mathbb{R} - X \mid X \in \boldsymbol{\Sigma}^0_\alpha\}$.
- For $\alpha \geq 2$ and given $\left(\boldsymbol{\Sigma}^0_\beta, \boldsymbol{\Pi}^0_\beta\right)_{\beta<\alpha}$, a set S is in $\boldsymbol{\Sigma}^0_\alpha$ if there exists sets $(S_n)_{n\in\mathbb{N}}$ such that $S = \bigcup_{n\in\mathbb{N}} S_n$ and $S_n \in \bigcup_{\beta<\alpha}(\boldsymbol{\Sigma}^0_\beta \cup \boldsymbol{\Pi}^0_\beta)$ for all $n < \omega$.

We also define $\boldsymbol{\Delta}^0_\alpha := \boldsymbol{\Sigma}_\alpha \cap \boldsymbol{\Pi}_\alpha$ and $\mathcal{B}_\alpha := \boldsymbol{\Sigma}^0_\alpha \cup \boldsymbol{\Pi}^0_\alpha$.

The reasons the hierarchy does not start at zero are tied to certain logical characterizations of the sets, and not very important. The superscript 0 suggests there are other hierarchies beyond the Borel sets (we will see one later). The letter $\boldsymbol{\Sigma}$ stands for sum (union), while $\boldsymbol{\Pi}$ is for product (intersections). Thus $\boldsymbol{\Sigma}^0_1$ is the set of all closed sets, while $\boldsymbol{\Delta}^0_1$ is the set of all sets that are both closed and open – the clopen sets: just $\emptyset$ and $\mathbb{R}$. In general, $\boldsymbol{\Delta}^0_\alpha$ is closed under complements.

Note that in the definition of Σ^0_α for $\alpha \geq 2$, we could take $S_n \in \bigcup_{\beta<\alpha} \Pi^0_\beta$ only (try to prove this; if you get stuck, first prove that open sets are countable unions of closed sets).

Of course, $\boldsymbol{\Delta}^0_\alpha \subseteq \boldsymbol{\Sigma}^0_\alpha \subseteq \mathcal{B}^0_\alpha$ and $\boldsymbol{\Delta}^0_\alpha \subseteq \boldsymbol{\Pi}^0_\alpha \subseteq \mathcal{B}^0_\alpha$. More interestingly, for $\alpha < \beta$ we have that $\mathcal{B}^0_\alpha \subseteq \boldsymbol{\Delta}^0_\beta$.

Since there are only $2^{2^{\aleph_0}}$-many sets of reals, we expect there should be an α so that $\mathcal{B}_\alpha = \mathcal{B}_\beta$ for all $\beta \geq \alpha$. In fact we have:

Lemma 10.5. $\mathcal{B}_{\omega_1}$ is a σ-algebra on $\mathbb{R}$. In particular, $\mathcal{B}_{\omega_1} = \mathcal{B} = \boldsymbol{\Delta}^0_\alpha = \boldsymbol{\Sigma}^0_\alpha = \boldsymbol{\Pi}^0_\alpha = \mathcal{B}^0_\alpha$ for all $\alpha \geq \omega_1$.

Proof. If $X \in \Sigma^0_{\omega_1}$, then there exists $(S_n)_{n<\omega}$ such that $X = \bigcup_{n<\omega} S_n$ and for all $n < \omega$, $S_n \in \mathcal{B}_{\alpha_n}$ for some $\alpha_n < \omega_1$. Let $\alpha := \sup_{n\in\mathbb{N}} \alpha_n$. Since ω_1 is regular, $\alpha < \omega_1$, so $S_n \in \mathcal{B}_\alpha \subseteq \boldsymbol{\Delta}^0_{\alpha+1}$ for all $n < \omega$, so $X \in \boldsymbol{\Sigma}^0_{\alpha+2} \subseteq \mathcal{B}_{\alpha+2}$. Thus $\boldsymbol{\Sigma}^0_{\omega_1} = \bigcup_{\alpha<\omega_1} \boldsymbol{\Sigma}^0_\alpha$. Similarly, $\boldsymbol{\Pi}^0_{\omega_1} = \bigcup_{\alpha<\omega_1} \boldsymbol{\Pi}^0_\alpha$. Thus $\mathcal{B}_{\omega_1} = \bigcup_{\alpha<\omega_1} \mathcal{B}_\alpha$. A similar argument using regularity of ω_1 shows that $\mathcal{B}_{\omega_1}$ is closed under countable union and complement. This proves that $\mathcal{B}_{\omega_1}$ is a σ-algebra. The definition shows it has to contain any σ-algebra containing the open sets, so $\mathcal{B}_{\omega_1} = \mathcal{B}$. It is also easy to see that the hierarchy must collapse after $\mathcal{B}_{\omega_1}$. □

Remark 10.6. It can be shown that $\boldsymbol{\Delta}_\alpha \subsetneq \boldsymbol{\Sigma}_\alpha \subsetneq \boldsymbol{\Delta}_{\alpha+1}$ for any $\alpha < \omega_1$. Thus ω_1 really is the least ordinal α such that $\mathcal{B}_{\alpha+1} = \mathcal{B}_\alpha$: the hierarchy really takes ω_1-many stages to grow. We will analyze this phenomenon more in depth later when we talk about trees.

Definition 10.7. The *Borel rank* of a Borel set X is the minimal ordinal α such that $X \in \mathcal{B}_{\alpha+1}$.

Thus open and closed sets have Borel rank 0, countable union of open sets have Borel rank 1, countable union of sets of Borel rank 1 have rank at most 2, etc. If we want to understand sets of reals, it may make sense to start with the simple one, at the bottom of the Borel hierarchy, and then walk our way up. Note however that we cannot hope to reach *all* sets this way: there are sets of reals that are not Borel (exercise). For example:

Lemma 10.8 (The continuum hypothesis for open sets)**.** Any non-empty open set has cardinality $2^{\aleph_0}$.

Proof. A non-empty open set X must contain an interval of the form $(x - \epsilon, x + \epsilon)$ for $\epsilon > 0$ and $x \in X$. We can pick a nonzero natural number n so that $\frac{1}{n} < \epsilon$.

In particular, X contains an interval of the form $[x, x + \frac{1}{n}]$. This is in bijection with $[0, \frac{1}{n}]$, hence (using the bijection induced by multiplication by n) with $[0, 1]$. We saw already (Example 8.22) that $|[0, 1]| = |\mathbb{R}| = 2^{\aleph_0}$, hence it must follow that $2^{\aleph_0} = |[0, \frac{1}{n}]| \leq |X| \leq |\mathbb{R}| = 2^{\aleph_0}$, so $|X| = 2^{\aleph_0}$. □

10.1. **The continuum hypothesis for closed sets.** We now work toward establishing the continuum hypothesis for closed sets. The following notions will play a key role:

Definition 10.9. Let X be a set of real numbers.

(1) An element $x \in \mathbb{R}$ is a *limit point of* X if there is a sequence of members of $X - \{x\}$ that converges to x. Equivalently, every open set that contains x intersects $X - \{x\}$. If x is not a limit point but $x \in X$, we say it is an *isolated point of* X. Equivalently, x is an isolated point if there is $\epsilon > 0$ so that $(x - \epsilon, x + \epsilon) \cap X = \{x\}$.
(2) Let X' be the set of all limit points of X.
(3) X is called *perfect* if it is closed, not empty, and does not contain any isolated point.

Exercise 10.10. A set X of real numbers is closed if and only if $X' \subseteq X$. Thus a non-empty set X is perfect if and only $X' = X$.

Exercise 10.11. X' is closed for any set of reals X.

Example 10.12.

(1) Consider the set $X := 0 \cup \{\frac{1}{n} \mid n \in \mathbb{N} - \{0\}\}$. X is closed, 0 is a limit point, and the other points are isolated. Thus $X' = \{0\}$. In X', 0 is now isolated, so $X'' = \emptyset$.
(2) Any closed interval $[a, b]$ with $a < b$ is perfect.

The operation $X \mapsto X'$ is sometimes called the *Cantor-Bendixon derivative.* We will study what happens when we iterate it many times. For now, we explain what this has to do with the continuum hypothesis:

Theorem 10.13. Any perfect set has cardinality $2^{\aleph_0}$.

Proof. Let P be a perfect set. We will build an injection of ${}^{\mathbb{N}}2$ into P. We first build $(I_s)_{s \in {}^{<\mathbb{N}}2}$ such that, for each $n \in \mathbb{N}$ and each $s \in {}^n2$:

- I_s is an open interval of length at most $\frac{1}{n+1}$.
- $I_s \subseteq I_{s \restriction k}$ for any $k < n$.
- $\overline{I_{s \frown 0}} \cap \overline{I_{s \frown 1}} = \emptyset$. Here, $s \frown \ell$ denotes the sequence of length $n + 1$ obtained by adding ℓ to the end of s. On the other hand, $\overline{I}$ denotes the closure of I: the smallest closed set containing I (in this case, this is just the closed interval with the same endpoints).
- $I_s \cap P$ is not empty.

The construction is by induction on n. For $n = 0$, we pick any $x \in P$ and let $I_{\langle\rangle}$ be any open interval of length at most one containing x. Given I_s, pick $x \in I_s \cap P$. As x is not isolated, there exists $y \in I_s \cap P$ with $x \neq y$. Without loss of generality, $x < y$. Pick z with $x < z < y$ and pick $I_{s \frown 0} \subseteq I_s$ any open interval of length at most $\frac{1}{n+2}$ containing x but not z. Similarly define $I_{s \frown 1} \subseteq I_s$ to be any open interval of length at most $\frac{1}{n+2}$ containing y but not z.

Now for each sequence $f \in {}^{\mathbb{N}}2$ and each $n \in \mathbb{N}$, we can pick a real $r_{f,n} \in I_{f \upharpoonright n} \cap P$, and let r_f be the limit of the sequence of $r_{f,n}$. Note that because P is closed, $r_f \in P \cap \bigcap_{n<\omega} \overline{I_{f \upharpoonright n}}$, and there is a unique element in the latter set (any other element would have to be within distance $\frac{1}{n+1}$ of r_f for any n). In particular, if $f \neq g$ then $r_f \neq r_g$. Thus the map $f \mapsto r_f$ is an injection of ${}^{\mathbb{N}}2$ into P. □

Note that if U is a non-empty open set, then it contains an open interval (a, b), and hence a perfect set of the form $[a - \epsilon, b - \epsilon]$. Thus Theorem 10.13 gives another proof of the continuum hypothesis for open sets. Generally, a strategy to prove that a certain set of reals has cardinality $2^{\aleph_0}$ is to show it contains a perfect set, so we introduce the following definition:

Definition 10.14. A set of reals X has the *perfect set property* if it is either countable or contains a perfect set.

As just discussed, any open set has the perfect set property, and moreover:

Corollary 10.15. If X is a set of reals with the perfect set property, then it is either countable or has cardinality $2^{\aleph_0}$.

Proof. Clearly, $|X| \leq |\mathbb{R}| = 2^{\aleph_0}$, so the result follows from Theorem 10.13. □

Remark 10.16. Any finite set is closed, and any countable set is $\mathbf{\Sigma}^0_2$ (exercise!), so sets higher in the Borel hierarchy are all uncountable.

We work toward establishing that closed sets have the perfect set property. For this, we will simply iterate the Cantor-Bendixon derivative:

Definition 10.17. For a set X of reals and an ordinal α, define $X^{(\alpha)}$, the *αth derivative of* X, by induction on α as follows:

- $X^{(0)} = X$.
- $X^{(\alpha+1)} = \left(X^{(\alpha)}\right)'$, the set of limit points of $X^{(\alpha)}$.
- $X^{(\delta)} = \bigcap_{\alpha<\delta} X^{(\alpha)}$ if δ is limit.

If X is closed, it is clear that $X^{(\beta)} \subseteq X^{(\alpha)}$ for $\alpha \leq \beta$. Since there are only so-many reals, the process must terminate at some point.

Definition 10.18. The *Cantor-Bendixon rank* of a set of reals X is the minimal ordinal α such that $X^{(\alpha)} = X^{(\alpha+1)}$.

Lemma 10.19. For any closed set X of real numbers and any ordinal α, $X - X^{(\alpha)}$ is countable. In particular, the Cantor-Bendixon rank of X is a countable ordinal.

Proof. The rational numbers are countable, hence so is the set $\mathbb{Q} \times \mathbb{Q}$, so let $f : \mathbb{N} \to \mathbb{Q} \times \mathbb{Q}$ be a bijection and for each $k \in \mathbb{N}$, write $f(k) = (a_k, b_k)$, and let I_k be the open interval whose endpoints are a_k and b_k.

Let $a \in X - X^{(\alpha)}$. There is a unique ordinal $\gamma = \gamma_a < \alpha$ such that a is an isolated point of $X^{(\gamma)}$ (past γ, it is removed from the set; before γ it is a limit point). Therefore there exists a small open interval containing a but no other point of $X^{(\gamma)}$. We can make sure this interval has rational endpoints, so there exists a natural number k such that $I_k \cap X^{(\gamma)} = \{a\}$. Let $k(a)$ be the minimal such k.

We claim that the map sending a to $f(a) = k(a)$ is an injection from $X - X^{(\alpha)}$ into $\mathbb{N}$. Indeed, suppose that $a, b \in X - X^{(\alpha)}$ and $k = k(a) = k(b)$. By definition,

$I_k \cap X^{(\gamma_a)} = \{a\}$ and $I_k \cap X^{(\gamma_b)} = \{b\}$. Without loss of generality, $\gamma_a \leq \gamma_b$, so $X^{(\gamma_b)} \subseteq X^{(\gamma_a)}$, hence $I_k \cap X^{(\gamma_b)} \subseteq I_k \cap X^{(\gamma_a)} = \{a\}$, so $\{a\} = \{b\}$, so $a = b$.

To see the "in particular" part, suppose for a contradiction that $X^{(\alpha)} \neq X^{(\alpha+1)}$ for all $\alpha < \omega_1$. Then for each $\alpha < \omega_1$ we can pick $a_\alpha \in X^{(\alpha+1)} - X^{(\alpha)}$, and from the definition $\alpha \neq \beta$ implies $a_\alpha \neq a_\beta$. This witnesses that $X^{(\omega_1)} - X$ is uncountable, contradiction. □

Theorem 10.20 (Cantor-Bendixon theorem)**.** Any uncountable closed set of reals is the union of a countable set and a perfect set. In particular, any closed set has the perfect set property.

Proof. Let X be an uncountable closed set of reals. Let α be its Cantor-Bendixon rank. Let $A := X - X^{(\alpha)}$ and let $P := X^{(\alpha)}$. By construction, $P = P'$ and $X = A \cup P$. Moreover, A is countable by Lemma 10.19. As X is uncountable, P must be uncountable and so in particular non-empty. This shows that P is perfect. □

Corollary 10.21 (The continuum hypothesis for closed sets)**.** Any closed set of reals is either countable or has cardinality $2^{\aleph_0}$.

Proof. By Theorem 10.20 and Corollary 10.15. □

We could try to prove that every Borel set has the perfect set property by showing using induction on α that every set in $\mathcal{B}_\alpha$ has the perfect set property. This was Cantor's dream, perhaps (with a more complicated hierarchy), yielding to a proof of the continuum hypothesis. We have just done the case $\alpha = 1$, and we can go a little bit further: any countable union of closed sets (i.e. any $\mathbf{\Sigma}^0_2$ set) has the perfect set property. Indeed, the perfect set property is preserved by countable unions. It is not so obvious how to proceed for sets in $\mathbf{\Pi}^0_2$ (countable intersections of open sets). In fact, it is *not* true that a countable intersections of sets with the perfect set property always has the perfect set property. It is also not true that the perfect set property is preserved by taking complements.

One approach is to prove a more general version of the Cantor-Bendixon theorem for any appropriate topological space (any Polish space). Then one shows that given a Borel set, one can change the topology to make this Borel set clopen, and then apply the Cantor-Bendixon argument!

In the later sections, we discuss another, more powerful, approach using infinite games.

11. Infinite games: a first introduction

Definition 11.1. Fix a non-empty set X and a set $A \subseteq {}^\omega X$, consider the following game, called $G_X(A)$. There are two players, player I and II. They alternate playing elements from X: player I plays $a_0 \in X$, player II plays $a_1 \in X$, player I plays $a_2 \in X$, and so on. At the end, we obtain a sequence $a = (a_n)_{n<\omega}$, called a *run* (or *play*) of the game. Player I wins this run of the game if $a \in A$, and player II wins otherwise. The set A is called the *payoff set.*

We will focus on the case $X = \omega$, in which case we just write $G(A)$.

There are many games that fit this scheme. A game like chess does not last infinitely-many moves, but we could simply ignore the last moves after one of the player has won. We do not however allow a draw (so if we wanted to study chess, we would need some convention, like letting the second player win in case of a draw).

As another example, fix a set $A \subseteq \mathbb{R}$, and let:

$$A^* := \{a = (a_n)_{n<\omega} \in {}^{\omega}2 \mid \sum_{n=0}^{\infty} a_n 2^{-n} \in A\}$$

Then the game $G_2(A^*)$ has player I and II alternate playing zero and ones, and player I wins if and only if the corresponding run $a_0 a_1 \ldots$ is such that $\sum_{n=0}^{\infty} a_n \in A$. We want to use such games to encode properties like the existence of a perfect set.

We are interested in who will win the game $G(A)$, for a particular payoff set A. For example, if $A = {}^{\omega}\omega$ it is clear that player I always wins no matter what she does. On the other hand if $A = \emptyset$ player II always wins. If A contains a single element, say the all zero sequence, it is theoretically possible for player I to win, but one would really need player II to collaborate. Player II has what we call a winning strategy: a way to answer the moves of player I so that the final sequence always avoids A. In this case, the winning strategy is very simple: just play any nonzero number a_1, then do anything. Let us more generally define what a winning strategy is.

Definition 11.2. Let $A \subseteq {}^{\omega}\omega$.

(1) A *strategy* in $G(A)$ is any function $\sigma : {}^{<\omega}\omega \to \omega$.
(2) If σ is a strategy and $b \in {}^{\omega}\omega$, then we let $\sigma * b$ denote the run that results if I plays according to σ and II plays b. Specifically, $\sigma * b$ is defined inductively: for any natural number n, $(\sigma * b)_{2n+1} = b_n$, and $(\sigma * b)_{2n} = \sigma((\sigma * b) \upharpoonright 2n)$. We similarly define $b * \sigma$, the run that results if I plays b and II plays according to σ.
(3) A strategy σ is *winning for player I in* $G(A)$ if $\sigma * b \in A$ for any $b \in {}^{\omega}\omega$. That is, σ ensures a win for player I no matter what player II does. Similarly, σ is *winning for player II in* $G(A)$ if $b * \sigma \notin A$ for all $b \in {}^{\omega}\omega$.
(4) We say that A (or the game $G(A)$) is *determined* if either I or II has a winning strategy in $G(A)$.

Note that it is not possible for both players to have winning strategy: otherwise we could use the strategies against each other. Thus at most one player has a winning strategy.

Example 11.3.

(1) ${}^{\omega}\omega$ and $\emptyset$ are determined (I and II win the game with any strategy). Any singleton set is also determined, as we have seen. In fact, any finite set is determined (why?). In fact, even countable sets are determined (exercise).
(2) The game[6] $G(A)$ is *finite* if there exists a natural number n so that for any $a, b \in {}^{\omega}\omega$, if $a \upharpoonright n = b \upharpoonright n$, then $a \in A$ if and only if $b \in A$. In other words, after n-many moves we know the winner already. For example, chess and tic tac toe are finite games. Note however that if A is a singleton, like the all zero sequence, then $G(A)$ is not finite (even though A is itself a finite set).

[6] This is really a property of A, but it may be more helpful to think of it as a property of the game.

Let us argue that any finite game is determined: let us say for illustration that $n = 4$. In order for II to have a winning strategy, we must have:

$$(\forall a_0 \exists a_1 \forall a_2 \exists a_3) a_0 a_1 a_2 a_3 00 \ldots \in A$$

(where $\forall$ means "for all", $\exists$ means "there exists")
On the other hand I wins exactly when:

$$(\exists a_0 \forall a_1 \exists a_2 \forall a_3) a_0 a_1 a_2 a_3 00 \ldots \notin A$$

These two formulas are the negations of each other, so exactly one must hold, i.e. the game is determined. This illustrates an interesting reason to consider games: they allow us to easily think about properties that are complicated, in the sense that they are expressed with many quantifier alternations.

It is easy to believe that any set is determined. Indeed, any run is either won by player I or II, so it seems hard to imagine situations when one would not have a strategy. Such a situation can however happen:

Theorem 11.4. There exists sets that are *not* determined.

Proof. A rough attempt is to count the number of strategies: ${}^{<\omega}\omega$ is countable, hence there are only $2^{\aleph_0}$-many strategies, while there are $2^{2^{\aleph_0}}$-many subsets of ${}^{\omega}\omega$. This does not immediately imply that there are sets with no strategies, since a strategy can work for more than one game (the strategy that always plays 0 wins for I when $A = {}^{\omega}\omega$ and when $A = {}^{\omega}\omega - \{11111\ldots\}$).

We do a more sophisticated diagonalization argument: let $(\sigma_\alpha)_{\alpha < 2^{\aleph_0}}$ list all the strategies. We define $(a_\alpha, b_\alpha)_{\alpha < 2^{\aleph_0}}$ by induction on α as follows. Assume that $(a_\beta)_{\beta < \alpha}$ have been defined. Think of a_α as "αth possible play" of player I, and as b_α as a corresponding "αth possible play" for player II.

Note that $b \mapsto b * \sigma$ and $b \mapsto \sigma * b$ are injections (why?), so $b \mapsto \sigma * b$ is an injection for any strategy σ. Thus $\{\sigma * b \mid b \in {}^{\omega}\omega\}$ has cardinality $2^{\aleph_0}$. On the other hand, $\{\sigma_i * b_j \mid i, j < \alpha\} \cup \{a_j * \sigma_i \mid i, j < \alpha\}$ only has cardinality at most $\aleph_0 |\alpha| |\alpha| < 2^{\aleph_0}$, so there must be b_α so that in particular $\sigma_\alpha * b_\alpha \neq a_i * \sigma_i$ for $i < \alpha$. Similarly, there is a_α so that $a_\alpha * \sigma_\alpha \neq \sigma_i * b_i$ for $i \leq \alpha$.

Let $A := \{a_\alpha * \sigma_\alpha \mid \alpha < 2^{\aleph_0}\}$, $B := \{\sigma_\alpha * b_\alpha \mid \alpha < 2^{\aleph_0}\}$. We claim that A is not determined (and in fact neither is B). The key fact is that $A \cap B = \emptyset$: if $a_\alpha * \sigma_\alpha = b_\beta * \sigma_\beta$, then without loss of generality $\alpha \leq \beta$, and we get a contradiction to the construction of a_α and b_β.

Let σ be a strategy. Say $\sigma = \sigma_\alpha$. σ is not winning for player I in $G(A)$, because $\sigma_\alpha * b_\alpha$ is in B, hence not in A. Similarly, σ is not winning for player II in $G(A)$ because $a_\alpha * \sigma_\alpha$ is in A. □

The set constructed in the proof of Theorem 11.4 was artificial. The more important results we want to prove is that "simple sets" are determined. We have already seen, for example, that finite games are determined. We will prove more eventually, but it may be helpful to study one more game first.

Definition 11.5 (The perfect set game)**.** The *perfect set game* on a set $A \subseteq \mathbb{R}$ is defined as follows: player I plays a move of the form (U_0^0, U_0^1), where U_0^0, U_0^1 are open intervals with rational endpoints, length strictly less than 2^{-0}, and $U_0^0 \cap U_0^1 = \emptyset$. Player II replies a move of the form $a_0 \in \{0, 1\}$ (giving a choice of set). We then require that player I plays (U_1^0, U_1^1), open intervals with rational endpoints, length strictly less than 2^{-1}, $U_1^0 \cap U_1^1 = \emptyset$, but also $\overline{U_1^0} \cup \overline{U_1^1} \subseteq U_0^{a_0}$. Player II then continues and plays $a_1 \in \{0, 1\}$ and so on.

Note that any run of the perfect set game yields a unique real number $x \in \bigcap_{n<\omega} U_n^{a_n}$. Player I wins the perfect set game on A if $x \in A$, and player II wins otherwise.

Note that the perfect set game can be described as a game of the form $G_X(A^*)$, where X is the set whose elements are $0, 1$, and all open intervals with rational endpoints. Since X is a countable set, we could also encode the perfect set game as a game of the form $G_\omega(A^{**})$ However it is easier to describe it colloquially as above.

Theorem 11.6. Let $A \subseteq \mathbb{R}$.

(1) If player I has a winning strategy in the perfect set game on A, then A contains a perfect set.
(2) If player II has a winning strategy in the perfect set game on A, then A is countable.

Proof.

(1) If player I has a winning strategy σ in the perfect set game on A, let P be the set of all real numbers x produced by playing the strategy σ against all possible binary sequences. We have that $\emptyset \neq P \subseteq A$ as σ is winning for I. P contains no isolated points: if x is isolated, then we can find a small interval U around it that does not intersect A. Say U has length ϵ, and fix n so that $2^{-n} < \epsilon$. Let a be the play of II that yields x when played against σ. Then II can change a_n and the real number produced by the run will land in $U - \{a\}$, so outside of A, contradiction.

It remains to see that P is closed. Let x be a limit point of P. We describe a run of the game that will produce x. Suppose I plays (U_0^0, U_1^1). We claim that x is in U_0^0 or in U_1^1. Otherwise, there is a small neighborhood U of x disjoint from both U_0^0 and U_1^1, hence disjoint from $P - \{x\}$. This contradicts the fact that x is a limit point of P. If $x \in U_0^\ell$, we play $a_0 = \ell$. Continue like this inductively: given U_n^0, U_n^1, we know for the same reason as before that $x \in U_n^0$ or $x \in U_n^1$ (otherwise we could find a small neighborhood around x that the strategy for I avoids).

(2) Suppose player II has a winning strategy σ. Let x be a real number. Call a partial run $s = (U_0^0, U_0^1), a_0, (U_1^0, U_1^1), \ldots, (U_n^0, U_n^1), a_n$ *x-compatible* if it satisfies the rules of the game, is played according to σ, and $x \in U_n^{a_n}$. We claim that for every real number $x \in A$, there is an x-compatible run $s = s_x$ that cannot be extended to a longer one. Indeed if not, any x-compatible run can be further extended, yielding eventually to a run producing x. This is impossible since $x \in A$ and σ is assumed to be winning for II.

We now show that the map $x \mapsto s_x$ is an injection from A into the set of partial runs. Since there are only countably-many such runs, this will show that A is countable. Assume that $x \neq y$ but $s_x = s_y$. Pick disjoint open intervals U_{n+1}^0, U_{n+1}^1 small-enough, satisfying the rules, and so that $x \in U_{n+1}^0$, $y \in U_{n+1}^1$. Now if σ chooses U_{n+1}^0, this means that s_x was not maximal, and if σ chooses U_{n+1}^1 this means that s_y was not maximal.

□

In conclusion, we get that if a set of reals A is determined, in the sense that either player I or player II has a winning strategy in the perfect set game on A, then it has the perfect set property! Thus the perfect set property holds for "simple" sets of reals, where "simple" now means determined. In particular we obtain another proof that closed sets have the perfect set property:

Theorem 11.7. If $A \subseteq \mathbb{R}$ is a closed set, then either I or II has a winning strategy in the perfect set game for A. In particular, A has the perfect set property

Proof. The last sentence follows from Theorem 11.6. Assume that II does *not* have a winning strategy in the perfect set game for A. We describe a winning strategy σ for player I. In short, the strategy is to "not lose".

In details, given a partial run $s = (U_0^0, U_0^1), a_0, (U_1^0, U_1^1), a_1, \dots, (U_{n-1}^0, U_{n-1}^1), a_{n-1}$, we say that s is *okay* if II does *not* have a winning strategy from the position given by s. By hypothesis, we know that the empty partial run is okay. Given an okay partial run s as above, we also know there is a pair of open intervals (U_n^0, U_n^1) satisfying the rules so that for any $a_n \in \{0, 1\}$, $s \frown (U_n^0, U_n^1), a_n$ is okay. Indeed, otherwise that would just mean that player II can counter any move in such a way that he now has a winning strategy, and so would give a winning strategy from the position given by s, contradicting that s is okay. Define $\sigma(s) = (U_n^0, U_n^1)$. For a non-okay run s, we can define $\sigma(s)$ arbitrarily (such runs will never come up – they correspond to situations where player I made a mistake).

We have to see that σ is a winning strategy. Let $(U_0^0, U_0^1), a_0, (U_1^0, U_1^1), a_1 \dots$ be a run played according to σ. Let x be the real number produced by that run. Note that $U_n^{a_n} \cap A \neq \emptyset$ for any $n < \omega$: otherwise the partial run would not have been okay. Pick $x_n \in U_n^{a_n} \cap A$. Then it is easy to see that x_n converges to x. Since A is closed, $x \in A$, as desired. □

12. Trees and topology

We have seen that any finite game is determined. It may help to restate this result a little bit differently:

Definition 12.1. For any fixed set X, $n < \omega$ and $s \in {}^nX$, let $N_s := \{x \in {}^\omega X \mid x \restriction n = s\}$. We call any set of the form N_s a *basic open set, or a basic open neighborhood, in* ${}^\omega X$. A subset A of ${}^\omega X$ is called *open* if it is the union of a collection of basic open sets. We say that A is *closed* if ${}^\omega X - A$ is open.

If you know some topology, you can check that a subset of ${}^\omega X$ is open if and only if it is open in the product topology, where X is given the discrete topology.

If you do not know any topology this is fine too. We will only use the following definition:

Definition 12.2. A *topological space* is a pair (A, τ), where A is a set, and τ is a collection of subsets of A. The members of τ are called *open sets.* We also say that τ is a *topology on* A. We require:

- $\emptyset, A \in \tau$.
- τ is closed under unions: any union of open sets is open.
- τ is closed under finite intersections: any finite intersections of open sets is open.

In any topology, one defines a set to be closed if it is the complement of an open set. One example of a topology is given by $\mathbb{R}$ with its usual open sets. For practice, let us check:

Lemma 12.3. For any set X, the open sets introduced in Definition 12.1 give a topology on X.

Proof. Note that the empty set is open (it is the empty union), and also ${}^{\omega}X$ is open (it is the union of all the basic open sets, or alternatively it is equal to $N_{\langle\rangle}$). The open sets are closed under union by definition. Why are they closed under finite intersections? It is of course enough to see that if U and V are open, then $U \cap V$ is open. For this, it suffices (by distributivity of union and intersection) to see that $N_s \cap N_t$ is open for any finite sequence $s \in {}^{m}X$ and $t \in {}^{n}X$ of elements of X. Without loss of generality, $m \leq n$. There is $k \leq m$ largest such that $s \upharpoonright k = t \upharpoonright k$. If $k = m$, this means that s is an initial subsequence of t, so $N_s \cap N_t = N_t$. Otherwise $k < m$ and so $s(k) \neq t(k)$. In this case, $N_s \cap N_t = \emptyset$, which is also open. □

You should convince yourself that $G(A)$ is a finite game if and only if for some $n < \omega$ and $S \subseteq {}^{n}\omega$, $A = \bigcup_{s \in S} N_s$.

Exercise 12.4. Prove that for each $A \subseteq {}^{\omega}\omega$ so that $G(A)$ is finite, A is clopen. Is the converse true?

This suggests that one could generalize the determinacy of finite games by proving the determinacy of clopen sets. We will in fact prove even more: any open *or* closed set is determined. However it makes sense to delay the proof a little bit to develop some tools to handle these new spaces and think about their relationship with the reals. We will focus on two especially important spaces:

Definition 12.5. The *Baire space* is the set ${}^{\omega}\omega$ with the topology from Definition 12.1. The *Cantor space* is the set ${}^{\omega}2$, again with the topology from Definition 12.1.

Thinking of a real number as a sequence of digits (say in a certain base), it seems clear the Cantor and Baire space are closely related to the real line. They are however not exactly the same topologically. For example the basic open sets are clopen, but in $\mathbb{R}$ the only clopen sets are $\mathbb{R}$ and the empty set. We can realize the Cantor and Baire spaces as specific subspaces of $\mathbb{R}$:

- Let $f : {}^{\omega}2 \to \mathbb{R}$ be defined by $f(s) = \sum_{n=0}^{\infty} 2s(n)3^{-n-1}$. Let C be the range of f. C is often called the *Cantor set*: it contains all the numbers in $[0, 1]$ whose base 3 representation does not contain a 1. Come to class to see a picture of the Cantor set! We can think of C as a topological space, with the topology *induced from* $\mathbb{R}$: the open sets are the intersections of open sets of $\mathbb{R}$ with C (exercise: prove it is indeed a topological space).

  The map $f \upharpoonright C$ is a bijection and plays will with the topology of C: the inverse image of open sets from C under f is an open set in the Cantor space. The image of open sets from the Cantor space are also open in C. Such a map is called a *homeomorphism*. This means that C is topologically indistinguishable from the Cantor space. However it is often easier to think of strings of zero and ones than base 3 representation of real numbers, so we work with ${}^{\omega}2$ most of the time. The essence of the space though, is that it is a complete binary splitting tree of height ω. We will make this precise later.

- Let $f : {}^{\omega}\omega \to \mathbb{R}$ be given by:

$$f(s) = s(0) + 1 + \cfrac{1}{s(1) + 1 + \frac{1}{s(2)+1+\frac{1}{s(3)+1+\ldots}}}$$

  The expression on the right hand side is called a *continued fraction.* It can be shown that every real number has a continued fraction representation. The continued fraction of a rational number terminates, while that of an irrational does not. Thus the range of f is the set I of irrational numbers. Again, it can be shown that $f \upharpoonleft I$ is a homeomorphism, so ${}^{\omega}\omega$ is essentially the same as the irrational numbers. Still the essence of it is just that it is an infinitely-branching tree.

We prefer to work with the Cantor or Baire space rather than with the real numbers for several reasons that will become apparent as we go along. In essence, we care about the combinatorics of these spaces more than about topological properties such as connectedness, etc. Thus thinking of them as trees is much more convenient. Since they are closely related to the reals, very often proofs involving them have a translation to the reals, or we can use the maps above or others to directly get conclusions about the reals without needing to translate the proofs.

Spaces such as $\mathbb{R}$, ${}^{\omega}2$, and ${}^{\omega}\omega$ are *Polish spaces*: separable complete and metrizable topological spaces (do not worry if you do not know what these terms mean). There is a general theory of Polish spaces, and again many of the arguments we will do just for the Baire or Cantor spaces generalize to arbitrary Polish spaces.

What do these spaces have to do with trees, and what is a tree, really? We will use the following definition:

Definition 12.6. A *tree* on a set X is a set $T \subseteq {}^{<\omega}X$ such that for all $m < n < \omega$, if $s \in T$ and $s \in {}^{n}X$, then $s \upharpoonright m \in T$. That is, T is closed under taking initial subsequences.

We will adopt the following notation:

Definition 12.7. For $s \in {}^{\alpha}X$, $t \in {}^{\beta}X$ with α, β ordinals, we write $s \subseteq t$ (s is an *initial subsequence of* t) if $\alpha \leq \beta$ and $t \upharpoonright \alpha = \beta$.

The *length* of a sequence s is just its domain. For $s \in {}^{m}\omega$, $t \in {}^{n}\omega$, we write $s \frown t$ for the *concatenation* of the two sequences s and t: it is the unique sequence u of length $m+n$ such that $s \subseteq u$ and $u(m+k) = t(k)$ for $k < n$. We adopt this notation for infinite sequences too, and may also write things like $s \frown 0$ or $0 \frown 1 \frown s$.

We will focus on trees with $X = 2$ or $X = \omega$. We think of the empty sequence $\langle\rangle$ as the root (it is part of any non-empty tree, since it is the initial subsequence of length zero), and then grow it with sequences of length one, etc. Come to class for pictures again!

A tree may fail to be very high: for example all elements may have length one or zero. It could however happen that a sequence goes "all the way":

Definition 12.8. A *branch* of a tree T on X is an infinite sequence $x \in {}^{\omega}X$ such that $x \upharpoonright n \in T$ for all $n < \omega$. We denote by $[T]$ the set of all branches of T.

Example 12.9.

(1) $[{}^{<\omega}X] = {}^{\omega}X$.

(2) If $s \in {}^{<\omega}\omega$, then we can form the tree T_s of all sequences $t \in {}^{<\omega}\omega$ so that $s \subseteq t$ or $t \subseteq s$. Its set of branches $[T_s]$ is just the basic open set N_s.
(3) For any fixed $x \in {}^{\omega}\omega$, we have that $\{x\} = [T]$, where $T = \{x \restriction n \mid n < \omega\}$ (the tree is just a line).
(4) For each $n < \omega$, let $s_n : n \to \omega$ be the constantly n sequence: $s_n(i) = n$ for any $i < n$. Let $T := \{s_n \restriction m \mid m \leq n < \omega\}$. By construction, T is a tree on ω but $[T] = \emptyset$. However T still has arbitrarily long sequences.

In all the examples we had, $[T]$ was a closed set. This is not a coincidence:

Lemma 12.10. Let $C \subseteq {}^{\omega}X$. Then C is closed if and only if $C = [T]$ for some tree T on X.

Proof. Exercise. □

Intuitively, one can think of a tree as a set of rational numbers. The branches are those irrationals numbers that can be arbitrarily well approximated by elements of the tree. Since the rationals are dense, it makes sense that branches should be closed under limits and that any closed sets of reals should be approximable by a tree.

12.1. **More on infinite games.** Armed with the vocabulary of trees, we can adapt the proof of Theorem 11.7 to show:

Theorem 12.11 (Gale-Stewart theorem)**.** For any non-empty set X, if $A \subseteq {}^{\omega}X$ is open or closed, then A is determined.

Proof. We prove the result if A is open. If A is closed, you should be able to adapt the proof of Theorem 11.7 to show what is required.

Suppose that I does not have a winning strategy in $G_X(A)$. We show that II has a winning strategy. Let $C := {}^{\omega}X - A$. Since A is open, C is closed. By Lemma 12.10, $C = [T]$ for some tree T on X. For any fixed sequence $s \in {}^{<\omega}X$, let $A_s := \{t \in {}^{\omega}X \mid s \frown t \in A\}$ and call s of even length an *okay sequence* if player I does *not* have a winning strategy in the game $G_X(A_s)$. Note that $A_{\langle\rangle} = A$, so by assumption the empty sequence is okay.

We define a strategy σ for player II as follows: given an okay sequence s of even length, we know that no matter what number $\ell \in X$ player I plays, there must be $x_\ell \in X$ so that $s \frown \ell \frown x_\ell$ is still okay (otherwise, s itself would not be okay). Let $\sigma(s \frown \ell) := x_\ell$. On other inputs, define σ arbitrarily. We claim that σ is winning for player II. Let $x \in {}^{\omega}X$ and let $y := x * \sigma$. We have to see that $y \notin A$. Note that for every even $n < \omega$, $N_{y\restriction n} \cap C \neq \emptyset$ (otherwise $y \restriction n$ would not be an okay sequence). In particular, $y \restriction n \in T$, so y is a branch of T: $y \in [T]$, so $y \notin A$. □

The perfect set game takes a particularly simple form in the Cantor space:

Definition 12.12. The *perfect set game* on a set $A \subseteq {}^{\omega}2$ is played as follows: player I plays a sequence in $s_0{}^{<\omega}2$ (possibly empty), and player II plays $a_0 \in \{0, 1\}$. Player I then plays another sequence $s_1 \in {}^{<\omega}2$, player II replies with $a_1 \in \{0, 1\}$, and so on. Player I wins the game if $s_0 \frown i_0 \frown s_1 \ldots$ is in A, and player II wins otherwise.

Note that this version of the perfect set game is a game of the form $G_X(A^*)$, where $X = {}^{<\omega}2 \cup \{0, 1\}$. It is not too hard to see that the complexity of A^* is closely related to that of A. For example, convince yourself that if A is closed (in

${}^{\omega}2$) then A^* is also closed (in ${}^{\omega}X$). Similarly if A is Borel then A^* is Borel, and in fact they are at the same levels of the hierarchy.

We can now redo the proof of Theorem 11.6 to get:

Exercise 12.13. If player I has a winning strategy in the perfect set game on $A \subseteq {}^{\omega}2$, then A contains a perfect set. If player II has a winning strategy in the perfect set game on A, then A is countable.

Here, the notion of a perfect set in ${}^{\omega}2$ is defined as expected: it is a set P that is not empty, closed, and has no isolated point. A point x of P is isolated if there exists a basic open set N_s so that $P \cap N_s = \{x\}$. It is instructive to reprove in this case that a perfect set must have cardinality $2^{\aleph_0}$.

Using the Gale-Stewart theorem, we deduce that any closed set of the Cantor space has the perfect set property.

Exercise 12.14. Adapt the proof of the Cantor-Bendixon theorem to show that any closed subset of the Cantor space has the perfect set property.

Can we go further than closed determinacy? Are Borel sets determined? The answer is yes, although we will only prove a related result.

Fact 12.15 (Martin's theorem)**.** Any Borel set is determined.

Interestingly, the proof of Martin's theorem uses very big cardinals (though still existing within our axioms), at the level of $|V_{\omega_1}|$ (recall that $\mathbb{R}$ and all objects of mainstream mathematics live in $V_{\omega+\omega}$). It can be shown that such big cardinals are needed.

Corollary 12.16. Any Borel set has the perfect set property.

12.2. **Illfounded and wellfounded trees.** We saw in Example 12.9 that there could be trees with arbitrarily long sequences yet no branches. This cannot happen in an important particular case (the theorem is by a different König than the one on cardinal arithmetic[7]):

Definition 12.17. A tree T on a set X is *finitely splitting* (or *finitely branching*) if for any $s \in T$, $\{x \in X \mid s \frown x \in T\}$ is finite.

For example, if X is finite then any tree on X is finitely branching.

Theorem 12.18 (Kőnig's tree infinity lemma)**.** An infinite finitely splitting tree has a branch.

Proof. Let T be an infinite finitely splitting tree. For $s \in T$, we define T_s to be the tree of sequences $t \in T$ such that $s \subseteq t$ or $t \subseteq s$. We define a sequence $(s_n)_{n<\omega}$ of members of T inductively such that $s_n \in {}^{n}X$, $s_n \subseteq s_m$ whenever $n < m$, and T_{s_n} is infinite for any natural number n.

We pick s_0 to be the empty sequence. Note that $T_{\langle\rangle} = T$, which is infinite by assumption. Given s_n, let $t_1, \dots, t_k$ be the sequences in $T \cap {}^{n+1}X$ that extend s_n. There are only finitely-many since T is finitely splitting. Moreover, $T_{s_n} = T_{t_1} \cup T_{t_2} \cup \ldots \cup T_{t_k}$ and T_{s_n} is infinite so there exists $i \leq k$ so that T_{t_i} is also infinite. Set $s_{n+1} = t_i$.

Now that the construction is done, define $x \in {}^{\omega}X$ by $x(n) = s_{n+1}(n)$. Then $x \restriction n = s_n$, so x is a branch of T. □

[7] in fact the accent on the o is not of the same type.

Remark 12.19 (If you know some topology)**.** In general topology, Tychonoff's theorem tells us that a product of compact spaces is compact. In particular, a product of finite spaces is compact. One can deduce Kőnig's lemma from this: can you see how?

While very simple, Kőnig's lemma has a few nice applications. Our first application is in graph theory. Recall that a graph G is a pair (V, E), where V is a set (called the set of *vertices* of G) and E is a subset of $[V]^2$, the set of 2-element subsets of V. The members of E are called *edges* of G. For λ a cardinal, a λ*-coloring* of $G = (V, E)$ is a function $f : V \to \lambda$ such that $\{x, y\} \in E$ implies $f(x) \neq f(y)$ (in words, adjacent vertices are given different colors). A graph is λ*-colorable* if there is a λ-coloring of G. The *chromatic number* of G is the least cardinal λ such that G is λ-colorable. A *subgraph* of a graph $G = (V, E)$ is a graph $G_0 = (V_0, E_0)$ so that $V_0 \subseteq V$ and $E_0 \subseteq E$.

Theorem 12.20 (De Bruijn-Erdős)**.** If $G = (V, E)$ is a countable[8] graph and k is a natural number, G is k-colorable if and only if all its finite subgraphs are k-colorable.

Proof. If G is k-colorable, then clearly all its subgraphs are as well. Assume now that all finite subgraphs of G are countable. Of course, we might as well assume that G is infinite. Write $V = \{v_n \mid n < \omega\}$. We define a tree T on k as follows: $s \in T \cap {}^n k$ if the map $v_i \mapsto s(i)$ gives a k-coloring of $G \restriction \{v_0, \ldots, v_{n-1}\}$. Here, we write $G \restriction V_0$ for the graph $(V_0, E \cap [V_0]^2)$.

Intuitively, T is the tree of all k-coloring of finite restrictions of G. T is a tree (restrictions of k-colorings are k-colorings). Also, T is finitely splitting, since k is finite. Moreover, T is infinite because every n-vertices subgraph of G is k-colorable, and this k-coloring gives a member of $T \cap {}^n k$.

By Kőnig's lemma, there is a branch x through T. Define $f : V \to k$ by $f(v_n) := x(n)$. This is the desired coloring. □

Another fun application of Kőnig's lemma is a quick proof of the finite Ramsey theorem. We start with a simple puzzle: in a party with six students, three of them all know each other, or three of them all do not know each other. In general, the content of the finite Ramsey theorem is that for each natural number k there exists a natural number n so that in any party with n students there are k of them that all know or all do not know each other. More formally, we can think of coloring the edges of a complete graph on n vertices in two colors (depending on whether the two students know or do not know each other). From such a coloring, Ramsey's theorem tells us we can extract a homogeneous set: a set of vertices whose corresponding induced subgraph is monochromatic. The graph terminology does not help much here, so we will simply think of an edge coloring as a function $F : [X]^2 \to k$ for some set X (recall again that $[X]^2$ is the set of 2-element subsets of X). Rather than deal with finite sets, we will prove that in any infinite party there is a group of infinitely-many friends or infinitely-many strangers.

Theorem 12.21 (Infinite Ramsey theorem)**.** For any infinite set X, any natural number k, and any $F : [X]^2 \to k$, there exists an infinite $H \subseteq X$ such that $F \restriction [H]^2$ is constant.

[8] The result is true even if the graph is not countable, but the proof is harder.

Proof. We build inductively a sequence of elements $(x_n)_{n<\omega}$ from X, and a sequence of sets $(S_n)_{n<\omega}$ such that for all $n < \omega$:

1. $S_n \subseteq X$ is infinite.
2. $x_n \in S_n$.
3. $S_{n+1} \subseteq S_n$.
4. $F \restriction \{\{x_n, y\} \mid y \in S_{n+1}\}$ is constant.

This is possible: take $S_0 = X$, x_0 any element in X. Given S_n, x_n, for each $c < k$ let $P_c := \{y \in S_n - \{x_n\} \mid F(\{x, y\}) = c\}$. Note that $S_n - \{x\}$ is infinite, and $P_0, P_1, \dots, P_{k-1}$ is a finite partition of it, so there must exist $c < k$ so that P_c is infinite. Let $S_{n+1} := P_c$, and pick any $x_{n+1} \in S_{n+1}$.

Now that the construction is done, for each $n < \omega$, let c_n denote the constant value of $F \restriction \{\{x_n, y\} \mid y \in S_{n+1}\}$. There must be $c < k$ such that $c_n = c$ for infinitely-many $n < \omega$. Let $I := \{n < \omega \mid c_n = c\}$, and let $H := \{x_n \mid n \in I\}$. This is as desired: suppose $n < m$ are both in H. Then $x_m \in S_m \subseteq S_{n+1}$, so $F(\{x_n, x_m\}) = c_n = c$. □

The H in the statement is called a *homogeneous set.* We show why the finite Ramsey theorem follows:

Theorem 12.22 (Finite Ramsey theorem)**.** For any $k < \omega$ there exists $n < \omega$ such that if $F : [n]^2 \to k$, there is $H \subseteq n$ so that $|H| = k$ and $F \restriction [H]^2$ is constant.

Proof. Suppose not. Fix $k < \omega$ such that for all $n < \omega$ there is a "bad" coloring $F_n : [n]^2 \to k$ such that for no $H \subseteq n$ of size k is $F \restriction [H]^2$ constant.

Fix a bijection $n \mapsto u_n$ from ω to $[\omega]^2$. Define a tree T on k as follows: $s \in {}^m k$ is in T if and only if the coloring $c : \{u_\ell \mid \ell < m\} \to k$ given by $c(u_\ell) = s(\ell)$ is such that there is no $H \subseteq \omega$ so that $|H| = k$, $[H]^2 \subseteq \operatorname{dom}(c)$, and $c \restriction [H]^2$ is constant.

Intuitively, T is the tree of attempts to build a bad coloring of $[\omega]^2$. The bad colorings F_n, $n < \omega$ witness that T is infinite. Since k is finite, T is also finitely splitting. By Kőnig's lemma, there is a branch x through T. Define $c : [\omega]^2 \to k$ by $c(u_\ell) := x(\ell)$. By the infinite Ramsey theorem, there is $H \subseteq \omega$ such that $|H| = k$ and $c \restriction [H]^2$ is constant (in fact the infinite Ramsey theorem even guarantees an infinite such H). Pick $n < \omega$ big-enough that $[H]^2 \subseteq \{u_\ell \mid \ell < n\}$. Then $x \restriction n \notin T$, contradiction. □

We now want to understand when a tree has no infinite branches. If we picture the tree as growing downward, and we think of this as depicting a partial order, then not having a branch means being wellfounded. Thus we define:

Definition 12.23. A tree T on X is *wellfounded* if $[T] = \emptyset$ and *illfounded* otherwise.

Exercise 12.24. Prove that a tree T on X is wellfounded if and only if the relation R on X given by sRt if and only if $t \subsetneq s$ is wellfounded.

To develop the theory of wellfounded trees, it makes sense to come back to wellfounded relations more generally.

Definition 12.25. Let R be a relation on a set A. A *rank for R* is a function $\rho : A \to \alpha$, where α is an ordinal and aRb implies $\rho(a) < \rho(b)$ for any $a, b \in A$.

Theorem 12.26. A relation R on a set A is wellfounded if and only if there is a rank function for R.

Proof. If R is *not* wellfounded, then by an exercise in the assignment there is a sequence $(a_n)_{n<\omega}$ in A so that so that $a_{n+1} R a_n$ for all $n < \omega$. If $\rho : A \to \alpha$ was a rank function, we would have $\rho(a_{n+1}) < \rho(a_n)$ for all $n < \omega$, showing that α itself is not wellfounded, contradiction.

Conversely, assume that R is wellfounded. Since R is wellfounded and set-like (A is a set), we can proceed by recursion to define $\rho : A \to \alpha$, where $\alpha = (|A| + \aleph_0)^+$. Fix $a \in A$ and assume inductively that ρ is defined on the R-predecessors of a. Let $\rho(a) := \sup\{\rho(b) + 1 \mid bRa\}$. In particular if a is minimal in A, $\rho(a) = 0$. From the definition of ρ, it is clear that it is a rank function. □

Definition 12.27. For any wellfounded relation R on A, we let ρ_R be the rank function defined in the proof above. The *rank of R* is defined to be the supremum of $\{\rho_R(a) \mid a \in A\}$. The *rank* of a wellfounded tree T is the rank of the relation $\supsetneq$ on T.

More explicitly, given a wellfounded tree T on X, define inductively $\rho_T : T \to (|T| + \aleph_0)^+$ by $\rho_T(s) = \sup\{\rho_T(t) + 1 \mid s \subsetneq t, t \in T\}$. Note that this coincide with the definition above. Moreover, if X is countable then the codomain of the rank function is ω_1, so the rank of every tree is a countable ordinal. Note also that the rank of a tree is just the rank of the root, the empty sequence.

12.3. **The projective sets.** A subset of the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is *open* if it is a union of open rectangles (sets of the form $(a, b) \times (c, d)$). Starting from this definition, we can then define the Borel hierarchy on $\mathbb{R}^2$. The following question was the source of many mistakes: given a Borel subset A of $\mathbb{R}^2$, is the projection $\text{proj}(A)$ of A Borel? Here, $\text{proj}(A)$ it the set of real numbers x so that $(x, y) \in A$ for some real number y.

The answer turns out to be no. It is connected to questions such as whether a Borel subset X of $\mathbb{R}^2$ (say with $\text{proj}(X) = \mathbb{R}$) can be *uniformized* by a Borel function: a function $f : \mathbb{R} \to \mathbb{R}$ so that the inverse images of Borel sets are Borel and $(x, f(x)) \in X$ whenever $x \in \mathbb{R}$. Such problems come up often: for example, finding a nice function "implicitly defined" by the equation $y = x^2$ is the same as finding a good uniformization of the corresponding graph.

Unfortunately it is not true that Borel sets have Borel uniformizations. However it turns out that X can be uniformized by a function whose graph will be the complement of the projection of a Borel set.

Really, we are starting to study a new hierarchy, that of *projective sets.* We will define them for the Baire space ${}^\omega\omega$. Indeed, the Baire space has the convenient property of being isomorphic to its product! In details, we will often identify ${}^\omega\omega \times {}^\omega\omega$ with ${}^\omega(\omega \times \omega)$, using the map π given by $(f, g) \mapsto (n \mapsto (f(n), g(n)))$. It is easy to see that this map preserves all topological properties: for example if X is an open subset of ${}^\omega\omega \times {}^\omega\omega$ (in the sense that it is union of sets of the form $U \times V$, with U, V open in the Baire space), then $\pi[X]$ is open as well (the converse is also true). In fact, when we say a certain set $X \subseteq {}^\omega\omega \times {}^\omega\omega$ is (say) $\boldsymbol{\Sigma}^0_\alpha$, we *mean* that $\pi[X]$ is $\boldsymbol{\Sigma}^0_\alpha$, where we define the hierarchy $\boldsymbol{\Sigma}^0_\alpha$ in the obvious way, starting with letting Σ^0_1 be the open sets of the Baire space. We are abusing notation here: we should really write $\Sigma^0_\alpha(\mathbb{R})$ and $\Sigma^0_\alpha({}^\omega\omega)$, but this is a bit too heavy.

Note also that $\omega \times \omega$ is in bijection with ω, so we can similarly identify ${}^\omega\omega$ and ${}^\omega(\omega \times \omega)$ and may do so silently again.

Definition 12.28 (The projective hierarchy)**.** Define by induction on $n < \omega$, $n \geq 1$ classes of subsets of ${}^{\omega}\omega$ called $\mathbf{\Sigma}^1_n, \mathbf{\Pi}^1_n$, as follows:

(1) $\mathbf{\Sigma}^1_1$ is the set of all sets of the form $\text{proj}(X)$, where X is a Borel subset of ${}^{\omega}\omega \times {}^{\omega}\omega$.
(2) $\mathbf{\Pi}^1_n$ is the set $\{{}^{\omega}\omega - X \mid X \in \mathbf{\Sigma}^1_n\}$.
(3) For $n \geq 2$, $\mathbf{\Sigma}^1_n$ is the set of all sets of the form $\text{proj}(X)$, where $X \in \mathbf{\Sigma}^1_m \cup \mathbf{\Pi}^1_m$ for $m < n$.

We also define $\mathbf{\Delta}^1_n := \mathbf{\Sigma}^1_n \cap \mathbf{\Pi}^1_n$. A *projective set* is a set that is a member of $\mathbf{\Sigma}^1_n$ for some $n \geq 1$.

In words, projective sets are those that can be obtained from Borel sets by taking projections and complements. As for the Borel sets, we have that $\mathbf{\Sigma}^1_n, \mathbf{\Pi}^1_n \subsetneq \mathbf{\Delta}^1_{n+1}$ for each n. It is easy to see that every Borel set is in $\mathbf{\Delta}^1_1$. However it is a nontrivial theorem that there are no other sets: $\mathbf{\Delta}^1_1$ is the set of all Borel subsets of ${}^{\omega}\omega$.

The following notion is related:

Definition 12.29. A subset A of ${}^{\omega}\omega$ is called *analytic* if there exists a closed set C such that $A = \text{proj}(C)$.

Since any closed set is Borel, any analytic set is $\mathbf{\Sigma}^1_1$. We will show that the converse is also true. In particular, any Borel set is analytic.

Lemma 12.30. The projection of an analytic set is analytic.

Proof. Immediate: do some coding. □

Lemma 12.31. Analytic sets are closed under countable unions and intersections.

Proof sketch. Let us show that analytic sets are closed under countable intersections. The proof that they are closed under countable unions is similar and easier. The intuition is that taking a projection corresponds to a quantifier of the form "there exists $x \in {}^{\omega}\omega$", while taking a countable union corresponds to a less powerful quantifier of the form "there exists $n \in \omega$". On the other hand a countable intersection is a quantifier of the form "forall $n \in \omega$, which is harder to handle.

Assume $(A_n)_{n<\omega}$ is a sequence of analytic sets. Fix closed sets C_n so that $A_n = \text{proj}(C_n)$. Let $A := \bigcap_{n<\omega} A_n$. Let us also define $C := \bigcup_{n<\omega} C_n \times \{c_n\}$, where c_n is the constantly n sequence. It is easy to check that C is also a closed set.

We then have that $x \in A$ if and only if for all $n < \omega$, $x \in A_n = \text{proj}(C_n)$ if and only if for all $n < \omega$, there exists $y = y_n \in C_n$ such that $(x, y) \in C_n$ if and only if for all $n < \omega$ there exists $y = y_n \in {}^{\omega}\omega$ such that $(x, y, c_n) \in C$.

This last statement is equivalent to: there exists $z \in {}^{\omega}\omega \times \omega$ so that for all $n < \omega$, $(x, z^n, c_n) \in C$. Here, z^n denotes the map $m \mapsto z(m, n)$. What is happening is that we are coding the y_n's into a single sequence z. Now for a fixed natural number n, let C'_n be the set of all pairs (x, z) such that $(x, z^n, c_n) \in C$. One can check that C'_n is closed, hence $C' := \bigcap_{n<\omega} C'_n$ is closed. We therefore get that $x \in A$ if and only if there exists z such that $(x, z) \in C'$, so $A = \text{proj}(C')$. □

Lemma 12.32. Any Borel set is analytic.

Proof. Trivially, any closed set is analytic. Any open set is a countable union of closed sets (exercise), so any open set is analytic. The result then follows from an easy induction. □

Lemma 12.33. The analytic sets are exactly the $\mathbf{\Sigma}_1^1$-sets.

Proof. Immediate from the fact that a Borel set is analytic, and the projection of an analytic set is analytic. □

Lemma 12.34. For any natural number $k \geq 1$, $\mathbf{\Sigma}_k^1$, $\mathbf{\Pi}_k^1$ are closed under countable unions and countable intersections.

Proof. This follows from induction and the fact that analytic sets are closed under these operations (really we have to imitate the proof of Lemma 12.31). □

Note that it does *not* follow that if A_n is a set in Σ_n^1, then $\bigcup_{n<\omega} A_n$ is a projective set! Thus we could further iterate and look at $\mathbf{\Sigma}_\alpha^1$ for an ordinal α (and then project those!). Projective sets are complicated-enough, so we will not go further.

Whether projective sets are determined depends on large cardinal axioms. We may see later a proof that analytic sets are determined using what is called a *measurable cardinal*, a very large cardinal whose existence cannot be proven from our axioms. Even larger cardinal axioms are needed for all projective sets. We give the statement of the result for information only:

Theorem 12.35 (Martin-Steel theorem)**.** If there are infinitely-many Woodin cardinals, then all projective sets are determined.

12.4. **Trees of uncountable height.** We now generalize our definition of a tree.

Definition 12.36. A *(long) tree* on a set X is a subset T of ${}^{<\alpha}X$, for some ordinal α. We require that if $s \in T$, then $s \upharpoonright \beta \in T$ for any $\beta \in \mathrm{dom}(s)$. As before, we write $s \subseteq t$ if $t \in {}^{\beta}X$, $s \in {}^{\gamma}X$, $\gamma \leq \beta$, and $t \upharpoonright \gamma = s$.

We will not always mention the "long", and will often simply call a long tree a tree. The following concepts are important:

Definition 12.37. Given a (long) tree T on X and an ordinal α, the *αth level* $\mathrm{Lev}_\alpha(T)$ of T is the set $T \cap {}^{\alpha}X$. The *height of* T, written $\mathrm{ht}(T)$, is the minimal ordinal α such that $\mathrm{Lev}_\alpha(T) = \emptyset$.

For example, the height of any tree $T \subseteq {}^{<\omega}X$ is at most ω. Even for such trees, one should not confuse rank with height: the height can be at most ω and is always defined. The rank is defined only for wellfounded trees, and can be any countable ordinal (as you will see in the assignments). Consider for example the tree T consisting of the constant sequences considered in Example 12.9. It is easy to see that its rank is ω. Consider however the tree T' of sequences of the form $0 \frown s$, for $s \in T$. Here, $\frown$ denotes concatenation of sequences. The rank of T' is $\omega + 1$: indeed the rank of the sequence 0 in T' is ω, the same as the rank of the empty sequence in T.

For long trees, the height can be any ordinal. For example the tree ${}^{<\omega_1}\omega$ (on ω) has height ω_1. The notion of a branch is defined just like before:

Definition 12.38. A *branch* of a long tree T on X is a sequence $x \in {}^{\mathrm{ht}(T)}X$ such that $x \upharpoonright \alpha \in T$ for all $\alpha < \mathrm{ht}(T)$.

We will focus on trees whose height is a cardinal and whose levels are "small":

Definition 12.39. For κ an infinite cardinal, a *κ-tree* is a (long) tree T such that:

- $\mathrm{ht}(T) = \kappa$.

- $|\operatorname{Lev}_\alpha(T)| < \kappa$ for all $\alpha < \kappa$.

Note that a tree of height ω is an $\aleph_0$-tree if and only if it is finitely splitting. Thus:

Theorem 12.40 (Kőnig's lemma, second version)**.** Any $\aleph_0$-tree has a branch.

Does this generalize to other cardinals? If T is a κ-tree, does it necessarily have a branch? It turns out that the answer is no. Counterexamples are given a name:

Definition 12.41. A κ-tree is *Aronszajn* if it does not have a branch.

Thus we may restate Kőnig's lemma once again:

Theorem 12.42 (Kőnig's lemma, third version)**.** There are no Aronszajn $\aleph_0$-trees.

The case $\kappa = \aleph_1$ is drastically different. Intuitively, $\aleph_1$ is "not as compact" as $\aleph_0$:

Theorem 12.43. There *is* an Aronszajn $\aleph_1$-tree on ω.

Proof. We start with $T_0 := \{f : \alpha \to \omega \mid f \text{ is an injection}\}$. Note that T_0 is a tree on ω. Moreover, T_0 has height ω_1 since for any countable ordinal α there is an injection of α into ω. Further, T_0 has no branch since any such branch would have to be an injection of ω_1 into ω, which cannot exist (by definition of ω_1). However T_0 is not quite the tree we are looking for since the levels are too big: for each infinite countable ordinal α, $|\operatorname{Lev}_\alpha(T_0)| = 2^{\aleph_0}$ (exercise – not needed for the proof). We will "thin out" T_0 to get our Aronszajn tree. To do this, we will attempt to build a branch through T_0, but allow ourselves finitely-many changes at each step. The resulting tree of attempts will be what we want.

Let's get to work! First define a relation $\sim$ on T_0 by $f \sim g$ if $\beta := \operatorname{dom}(f) = \operatorname{dom}(g)$, and $\{\alpha < \beta \mid f(\alpha) \neq g(\alpha)\}$ is finite. You should check that $\sim$ is an equivalence relation on T_0 and that each equivalence class is countable.

We build $(f_\alpha)_{\alpha<\omega_1}$ a sequence of elements in T_0 such that for all $\alpha < \omega_1$:

(1) $\operatorname{dom}(f_\alpha) = \alpha$.
(2) $\alpha < \beta < \omega_1$ implies $f_\beta \restriction \alpha \sim f_\alpha$.
(3) $\omega - \operatorname{ran}(f_\alpha)$ is infinite.

If we can do it, then this will be enough: let $T := \bigcup_{\alpha<\omega_1} [f_\alpha]_\sim$. T is an $\aleph_1$-tree, which cannot have a branch (as $T \subseteq T_0$, and T_0 itself does not have a branch).

The definition proceeds by induction on α. If $\alpha = 0$, let f_α be the empty function with codomain ω. If $\alpha = \beta + 1$ and we are given f_β, pick some $n \in \omega - \operatorname{ran}(f_\beta)$, and define $f_\alpha : \alpha \to \omega$ by $f_\alpha(\beta) = n$, $f_\alpha(\gamma) = f_\beta(\gamma)$ for $\gamma < \beta$. The limit case remains, and it is the hard one. Assume δ is a limit ordinal and $(f_\alpha)_{\alpha<\delta}$ are given (it may be instructive to pause and figure out the case $\alpha = \omega$ first). We know that $\operatorname{cf}(\delta) = \omega$, so fix $(\alpha_n)_{n<\omega}$ cofinal in δ.

Claim: There exists $f : \delta \to \omega$ such that $f \in T_0$ and $f \restriction \alpha_n \sim f_{\alpha_n}$ for all $n < \omega$.

Proof of Claim: We define $(g_n)_{n<\omega}$ by induction on n such that for all $n < m < \omega$, $g_n \in \operatorname{Lev}_{\alpha_n}(T_0)$, $g_n \sim f_{\alpha_n}$ and $g_m \restriction \alpha_n = g_n$.

Take $g_0 := f_{\alpha_0}$. Given g_n, let S be the set of $\beta < \alpha_{n+1}$ with $\alpha_n \leq \beta$ and $f_{\alpha_{n+1}}(\beta) = g_{\alpha_n}(\alpha)$ for some $\alpha < \alpha_n$. Intuitively, S is the set of problematic points β, where we cannot just set $g_{n+1}(\beta) = f_{\alpha_{n+1}}(\beta)$. This is not an issue because S is finite! Indeed, if $S' := \{\alpha < \alpha_n \mid f_{\alpha_n}(\alpha) \neq g_n(\alpha)\} \cup \{\alpha < \alpha_n \mid f_{\alpha_{n+1}}(\alpha) \neq f_{\alpha_n}(\alpha)\}$, then S' is finite, and if $\beta \in S$, then there must be a unique $\alpha = \alpha_\beta \in S'$ so that

$f_{\alpha_{n+1}}(\beta) = g_n(\alpha)$. Because $f_{\alpha_{n+1}}$ is an injection, the map $\beta \mapsto \alpha_\beta$ is an injection of S into S', so as S' is finite, S is finite too.

Now write $S = \{\beta_k \mid k < m\}$, and pick distinct $a_0, a_1, \ldots, a_{m-1}$ in $(\omega - \operatorname{ran}(g_n)) \cap (\omega - \operatorname{ran}(f_{\alpha_{n+1}}))$ (note that this set is infinite, as g_n and $f_{\alpha_{n+1}} \restriction \alpha_n$ disagree on only finitely-many points). Define $g_{n+1} : \alpha_{n+1} \to \omega$ by:

- $g_{n+1}(\alpha) = f_{\alpha_{n+1}}(\alpha)$ if $\alpha \notin S$, $\alpha \geq \alpha_n$.
- $g_{n+1}(\alpha) = g_n(\alpha)$ if $\alpha < \alpha_n$.
- $g_{n+1}(\beta_k) = a_k$.

This completes the construction. Now we can simply define $f : \delta \to \omega$ by $f(\alpha) = g_n(\alpha)$, where n is minimal such that $\alpha < \alpha_n$. $\dagger_{\text{Claim}}$

This almost works: f is an injection with domain δ that differs from each f_α in at most finitely-many points. The only problem is that $\omega - \operatorname{ran}(f)$ may not be infinite (in turn, it will yield to problems at the next successor stage!). Thus we will change f so that its range avoids $f(\alpha_n)$, for n odd. Let $f_\delta : \delta \to \omega$ be defined by $f_\delta(\alpha) = f(\alpha)$ if $\alpha \neq \alpha_n$ for any $n < \omega$, and $f_\delta(\alpha) = f(\alpha_{2n})$ if $\alpha = \alpha_n$ for $n < \omega$.

Let us check that f_δ is as desired:

(1) f_δ has domain δ and is an injection (easy to see), so $f_\delta \in T_0$.
(2) $\omega - \operatorname{ran}(f_\delta) \supseteq \{f(\alpha_{2n+1}) \mid n < \omega\}$, so is infinite.
(3) For any $n < \omega$, $f \restriction \alpha_n \sim f_{\alpha_n}$. Since $f_\delta \restriction \alpha_n$ differs from $f \restriction \alpha_n$ on a subset of $\{\alpha_m \mid m < n\}$, we have that $f_\delta \restriction \alpha_n \sim f \restriction \alpha_n \sim f_{\alpha_n}$.

In general, if $\alpha < \delta$, we can fix $n < \omega$ such that $\alpha < \alpha_n$ and we have $f_\delta \restriction \alpha_n \sim f_{\alpha_n}$, so $f_\delta \restriction \alpha = (f_\delta \restriction \alpha_n) \restriction \alpha \sim f_{\alpha_n} \restriction \alpha \sim f_\alpha$, so $f_\delta \restriction \alpha \sim f_\alpha$.

□

The reason that the tree constructed above is Aronszajn is *not* that we get stuck at successors: we get stuck at limits! If we have a sequence $s \in {}^\delta X$ such that for all $\alpha < \delta$, $s \restriction \alpha \in T$ and $s \restriction \alpha$ has extensions to all levels (in particular to δ), this does *not* mean that $s \in T$. In fact it may be interesting to note that the tree constructed in the above proof has the following property: any element can always be extended further. In general, any $\aleph_1$-tree can be "prunned" to have that property:

Lemma 12.44. If T is an $\aleph_1$-tree, then there exists an $\aleph_1$-tree $T_0 \subseteq T$ such that for any $s \in T_0$ and any $\alpha < \omega_1$, there exists $t \in \operatorname{Lev}_\alpha(T_0)$ so that either $t \subseteq s$ or $s \subseteq t$.

Proof. For $s \in T$, let $N_s := \{t \in T \mid s \subseteq t\}$. Also write $\operatorname{Lev}_{\leq\alpha}(T)$ for $\bigcup_{\beta \leq \alpha} \operatorname{Lev}_\beta(T)$. Note that if $\alpha < \omega_1$, then $\operatorname{Lev}_{\leq\alpha}(T)$ is a countable union of countable sets, hence countable.

Let T_0 be the set of members of T that have uncountably-many extensions. That is, T_0 is the set of $s \in T$ such that N_s is uncountable.

Note that T_0 is a tree, and all its levels are countable (as it is a subtree of T). We now show that every element of T_0 has extensions to all countable levels (this will in particular show that the empty sequence has extensions to all countable levels, hence that T_0 is an $\aleph_1$-tree). Let $s \in T_0$. Say $s \in \operatorname{Lev}_\alpha(T_0)$. Let $\beta \geq \alpha$. Let $I := \operatorname{Lev}_\beta(T) \cap N_s$. We have that $N_s \subseteq \operatorname{Lev}_{\leq\beta}(T) \cup \bigcup_{t \in I} N_t$. The first set in this union is countable. The second is a countable union, and N_s is uncountable, thus there must be some $t \in I$ so that N_t is uncountable. We then have that $t \in T_0$, as desired. □

12.5. **Aronszajn trees and the basis problem for linear orders.** To understand a given linear order L, it helps to know which "simple" linear orders embed into L. The following is an application of the infinite Ramsey's theorem (exercise):

Theorem 12.45. For any infinite linear order L, either ω or ω^R embed into L.

Here, for a linear order I, I^R is the "reverse" of I: $a <_{I^R} b$ if and only if $a >_I b$.

What about uncountable linear orders? Is there a simple collection C of uncountable linear orders so that any uncountable linear order contains a copy of a member of C? This is called the *basis problem for linear order* (and is a good final project topic!). What are the simple uncountable linear orders we know? ω_1, ω_1^R, and the reals (or more generally any uncountable suborder of the reals). Note that you have seen in an assignment that ω_1 and ω_1^R do not embed into the reals. The list is, however, not complete:

Definition 12.46. An *Aronszajn line* is an uncountable linear order L so that neither ω_1, nor ω_1^R, nor any uncountable $S \subseteq \mathbb{R}$ (with the usual order) embed into L.

Theorem 12.47. There is an Aronszajn line.

Proof. Let $T \subseteq {}^{<\omega_1}\omega$ be an Aronszajn $\aleph_1$-tree on ω (Theorem 12.43). Let T' be the set of members of T that have extensions to all levels (T' is the set of $s \in T$ so that for all $\alpha < \omega_1$ there is s_α so that $s \subseteq s_\alpha$ or $s_\alpha \subseteq s$). Note that T' is also an Aronszajn $\aleph_1$-tree: since T is uncountable and each level is countable, at each level there must be an element that is extended by uncountably-many other elements. These elements cannot be all below a certain level β (since again $\bigcup_{\alpha<\leq\beta} \text{Lev}_\alpha(T)$ is a union of countable sets, hence countable).

By Lemma 12.44, we may assume without loss of generality that every element of T has an extension to every higher level.

Order T lexicographically: let $L = (T, \leq)$, where $a \leq b$ if one of the following conditions hold:

- $a \subseteq b$.
- $a \not\subseteq b$, $b \not\subseteq a$, and if $\alpha < \omega_1$ is least so that $a \restriction (\alpha + 1) \neq b \restriction (\alpha + 1)$, then $a(\alpha) < b(\alpha)$.

You should check that L is a linear ordering. Further, it is uncountable since T is uncountable (remember that T has height ω_1, which means in particular it has at least one element per level).

Also observe that any element of L has a successor: since any $a \in T$ has an extension to a higher level, we can let $n < \omega$ be least so that $a \frown n \in T$. Then $a \frown n$ is the $<$-successor of a. This implies that no uncountable $S \subseteq \mathbb{R}$ embeds into L: it is not true that every element in such an S has a successor (otherwise one could pick a rational between each element and its successor, getting an injection of S into the rationals).

We now show that ω_1 does not embed into L. Suppose for a contradiction it does: let $f : \omega_1 \to L$ be an order embedding. Write $a_\alpha := f(\alpha)$. We build $(b_\alpha)_{\alpha<\omega_1}$ by induction on α such that for all $\alpha < \omega_1$:

(1) $b_\alpha \in \text{Lev}_\alpha(T)$.
(2) $b_\beta \restriction \alpha = b_\alpha$ whenever $\alpha < \beta < \omega_1$.
(3) There is $\beta < \omega_1$ such that for all $\gamma \geq \beta$, $b_\alpha \subseteq a_\gamma$.

If this can be done, then we would get a branch $b : \omega_1 \to \omega$ given by $b(\alpha) = b_{\alpha+1}(\alpha)$, contradicting that T is Aronszajn. For the construction, set $b_0 := \langle\rangle$, and if δ is a limit ordinal, let $b_\delta : \delta \to \omega$ be given by $b_\delta(\alpha) = b_{\alpha+1}(\alpha)$. Note that since for each $\alpha < \delta$ there is β_α so that $b_\alpha \subseteq a_\gamma$ for all $\gamma \geq \beta_\alpha$, we can let $\beta := \sup_{\alpha<\delta} \beta_\alpha$ and get that $b_\alpha \subseteq a_\gamma$ for all $\gamma \geq \beta$, hence $b_\delta \subseteq a_\gamma$ for all $\gamma \geq \beta$ (we are using here that ω_1 is regular to see that $\beta < \omega_1$).

The successor case requires some thoughts. Fix $\alpha < \omega_1$ and assume that b_α is given. Let $S := \{n < \omega \mid b_\alpha \frown n \in T\}$. Note that S is not empty since every element has extensions to all levels in T. For $n \in S$, let $A_n := \{\gamma < \omega_1 \mid b_\alpha \frown n \subseteq a_\gamma\}$. By assumption, for some $\beta < \omega_1$ and all $\gamma \geq \beta$, $b_\alpha \subseteq a_\gamma$. Without loss of generality, β is big-enough so that $a_\gamma \neq b_\alpha$ for any $\gamma \geq \beta$. Thus $\bigcup_{n \in S} A_n = [\beta, \omega_1)$. Moreover the A_n's are disjoint, so for each $\gamma \in [\beta, \omega_1)$, there is a unique $n = n_\gamma < \omega$ such that $\gamma \in A_n$. By definition of the ordering L, $\gamma \leq \gamma'$ implies $n_\gamma \leq n_{\gamma'}$. It follows there must be a maximal $n \in S$ such that $n = n_\gamma$ for some $\gamma \in [\beta, \omega_1)$ (otherwise, for each $n \in S$ there is γ_n so that $n_{\gamma_n} > n$, and $\gamma := \sup_{n<\omega} \gamma_n$ cannot be in any of the A_n's). It also follows that for this maximal $n \in S$ and $\gamma \in [\beta, \omega_1)$ such that $n_\gamma = n$, we have that $n_{\gamma'} = n$ for all $\gamma' \geq n$. Therefore we can let $b_{\alpha+1} := b_\alpha \frown n$.

The proof that ω_1^R does not embed into L is analogous: take the minimal rather than the maximal n in the argument above. □

13. Filters and ideals

In this section, we explore abstract notions of "largeness" that are specified axiomatically.

Definition 13.1. Let S be a non-empty set. A *filter* on S is a collection F of subsets of S such that:

(1) $S \in F$, $\emptyset \notin F$.
(2) If $X \in F$ and $X \subseteq Y \subseteq S$, then $Y \in F$.
(3) If $X, Y \in F$, then $X \cap Y \in F$.

Intuitively, a filter specifies a collection of sets to be thought of as "large". The full set (S) should be large, and any extension of a large set should be large. We also require that the intersection of two large sets should itself be large (this is maybe easier to understand when looking at the dual statement: a union of two small sets should be small). The concept is named "filter" because it allows us to filter out the small sets that do not matter and keep only the large ones, who do.

The dual notion is also interesting to look at:

Definition 13.2. Let S be a non-empty set. An *ideal* on S is a collection I of subsets of S such that:

(1) $\emptyset \in I$, $S \notin I$.
(2) If $Y \in I$ and $X \subseteq Y$, then $X \in I$.
(3) If $X, Y \in I$, then $X \cup Y \in I$.

An ideal specifies a collection of sets to be thought of as small, with similar properties to those of a filter. The two concepts are directly related as follows:

Exercise 13.3. A collection F of subsets of S is a filter if and only if $\{S - X \mid X \in F\}$ is an ideal.

The collection $\{S - X \mid X \in F\}$ is called the *dual ideal* of F. Similarly, we can define the dual filter of an ideal.

Why is the dual of a filter named an ideal? This comes from algebra!

Definition 13.4. A *commutative ring* is a triple $(R, +, \cdot)$ where R is a set and $+, \cdot$ are binary operations (i.e. functions from $R \times R$ to R) satisfying, for all $a, b, c \in R$:

(1) (Associativity) $(a + b) + c = a + (b + c)$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
(2) (Commutativity) $a + b = b + a$, $a \cdot b = b \cdot a$.
(3) (Existence of additive identity) There exists an element $0 \in R$ such that $x + 0 = x$ for all $x \in R$ (such an element is unique).
(4) (Existence of multiplicative identity) There exists an element $1 \in R$ such that $x \cdot 1 = x$ for all $x \in R$ (such an element is unique)
(5) (Existence of additive inverse) There exists $(-a)$ such that $a + (-a) = 0$.
(6) (Distributivity) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

An *ideal* of $(R, +, \cdot)$ is a non-empty subset $I \subseteq R$ such that $a, b \in I$ implies $a + b \in I$, and $r \in R$, $a \in I$ implies $ra \in I$.

The classical example of a commutative ring is the ring of integers $(\mathbb{Z}, +, \cdot)$. The ideals there are the sets of the form $m\mathbb{Z} := \{mx \mid x \in \mathbb{Z}\}$: multiples of a fixed number. However rings can also arise in the context of sets:

Exercise 13.5. Let S be a non-empty set. Prove that $(\mathcal{P}(S), \Delta, \cap)$ is a commutative ring, where Δ denotes symmetric difference ($A \Delta B = (A - B) \cup (B - A)$). Prove that $I \subseteq \mathcal{P}(S)$ is an ideal in this ring (in the sense of Definition 13.4) if and only if it is an ideal on S in the sense of Definition 13.2.

In fact, the ring $(\mathcal{P}(S), \Delta, \cap)$ is an example of a *Boolean ring*: a ring where $x \cdot x = x$ for any x.

It is time to give some examples of filters:

Example 13.6.

(1) For any non-empty set S, the *trivial filter* $\{S\}$ is a filter on S. Any other filter on S contains the trivial filter.
(2) More generally, if $A \subseteq S$ is a non-empty set, the *principal filter generated by* A is $\{X \subseteq S \mid A \subseteq X\}$ (the trivial filter was the case $A = S$). Dually, one can define the *principal ideal generated by* A (when $A \neq S$) to be $\{X \subseteq S \mid X \subseteq A\}$. Note that this use of the word principal agrees with the terminology of rings (an ideal I in a commutative ring R is *principal* if it is of the form $\{rm \mid r \in R\}$, for some $m \in R$). The principal filters are quite boring for us: interesting examples of filters are nonprincipal.
(3) If S is an infinite set, a subset $X \subseteq S$ is *cofinite* if $S - X$ is finite. The set of all cofinite subsets of S forms a filter on S, called the *cofinite filter*. The dual ideal is the ideal of all finite subsets of S. The closure of this ideal under union is given by the fact that the union of two finite sets is always finite. Note that the cofinite filter is not principal: if $A \subseteq S$ is cofinite, then $A - \{a\}$ is also cofinite for any $a \in A$.
(4) The *density* of a set of natural numbers A, written $d(A)$ is $\lim_{n\to\infty} \frac{|A \cap \{0, \ldots, n\}|}{n+1}$ (if it exists – if this limit does not exist, A does not have a density). Note that $d(\mathbb{N}) = 1$, the density of any finite set is 0, the density of the even numbers (and of the odd numbers) is $\frac{1}{2}$, etc. As an exercise, try to find an infinite set with zero density, and a set with no density.

Let $I := \{A \subseteq \mathbb{N} \mid d(A) = 0\}$. This is an ideal. To prove that it is closed under union, we use that $d(A \cup B) \leq d(A) + d(B)$ (if both of these exist). Note that the inequality is an equality if A and B are disjoint. In particular, the dual filter is the set of all $A \subseteq \mathbb{N}$ such that $d(A) = 1$. This is again not a principal filter.

The following concept from analysis is closely related to that of a filter:

Definition 13.7. A (finitely additive) *measure* on a set S is a function $m : \mathcal{P}(S) \to \mathbb{R}$ such that:

1. $m(\emptyset) = 0$, $m(S) > 0$.
2. (Monotonicity) $A \subseteq B \subseteq S$ implies $m(A) \leq m(B)$.
3. (Finite additivity) If A and B are disjoint, then $m(A \cup B) = m(A) + m(B)$.

Example 13.8.

1. Density would be an example of a measure on $\mathbb{N}$, except it is not defined for all sets.
2. Similarly, Lebesgue measure in analysis is almost an example of a measure on $[0, 1]$ (satisfying $m((a, b)) = b - a$) but it is not defined on every set.
3. For S a finite set, define the *counting measure* on S by $m(X) := |X|$.
4. For S a set and $a \in S$, define $m(X)$ to be 1 if $a \in X$, 0 otherwise. This is the measure concentrating on a.

The last two examples are quite trivial. Are there nontrivial measures? Could we for example somehow extend density to be defined for all sets in such a way that we get a measure? It may help to first connect measures with filters:

Exercise 13.9. Let S be a non-empty set. If m is a measure on S, then $\{X \subseteq S \mid m(X) = m(S)\}$ is a filter.

Is there a converse? If we start with a filter F on S can we get a measure? We could try defining $m(X) = 1$ if $X \in F$, and $m(X) = 0$ if $X \notin F$, but this does not quite work: say $S = \mathbb{N}$, A is the set of even numbers, B is the set of odd numbers, and F is the cofinite filter. Then $m(A) = 0 = m(B) = 0$, but $m(A \cup B) = 1$. Thus the cofinite filter does not have enough sets to get a measure. We need the filter to be "as big as possible":

Definition 13.10. An *ultrafilter* on a non-empty set S is a filter U on S such that for any $X \subseteq S$, $X \in U$ or $S - X \in U$. A *prime ideal* on S is an ideal P on S such that for any $X \subseteq S$, $X \in P$ or $S - X \in P$.

Exercise 13.11. Show that an ideal P on S is prime (in the sense above) if and only if it is prime (in the ring sense: an ideal I is *prime* if $ab \in I$ implies $a \in I$ or $b \in I$) in $(\mathcal{P}(S), \Delta, \cap)$.

Example 13.12. If S is a non-empty set and $A \subseteq S$ is not empty, the principal filter $F = \{X \subseteq S \mid A \subseteq X\}$ generated by A is an ultrafilter if and only if $A = \{a\}$ for some $a \in S$. The "if" part is because for any subset X of S either X contains a or the complement $S - X$ of X does. The "only if" part is because if $\{a, b\} \subseteq A$, $a \neq b$, then neither $\{a\}$ nor its complement is in the filter.

It is not so easy (in fact impossible, without the axiom of choice) to find examples of nonprincipal ultrafilters. We do get the connection we wanted with measures: call a measure m on a set S *two-valued* if $m(X) \in \{0, m(S)\}$ for any $X \subseteq S$.

Exercise 13.13. Let S be a non-empty set.

(1) If $m : \mathcal{P}(S) \to \mathbb{R}$ is a measure, then the filter $\{X \subseteq S \mid m(X) = m(S)\}$ is an ultrafilter if and only if m is two-valued.
(2) If U is an ultrafilter on S, then the function $m : \mathcal{P}(S) \to \mathbb{R}$ given by $m(X) = 1$ if $X \in U$ and $m(X) = 0$ if $X \notin U$ is a two-valued measure.

To build nonprincipal ultrafilters, we have to extend existing filters as much as possible. For example, we can start with the cofinite filter, decide whether to add the odd or the evens, close the result to a filter, continue until we cannot anymore. We now make this rigorous. First, we look at the collections that can generate filters:

Definition 13.14. Let G be a collection of subsets of a non-empty set S. We say that G has the *finite intersection property* if G is not empty and whenever $X_1, X_2, \dots, X_n \in G$, we have that $X_1 \cap X_2 \dots \cap X_n \neq \emptyset$.

Note that any filter has the finite intersection property. The set of all cofinite sets together with the set of even numbers also does. Trivially, $\{A\}$ has the finite intersection property, for any non-empty $A \subseteq S$. Thus the construction of a principal filter is generalized by:

Lemma 13.15. Let G be a collection of subsets of a non-empty set S. If G has the finite intersection property, then there exists a filter F on S such that $G \subseteq F$.

Proof. Let F be the set of all $X \subseteq S$ such that there exists $X_1, \dots, X_n \in G$ so that $X_1 \cap \dots \cap X_n \subseteq X$. Clearly, $G \subseteq F$. F is a filter: $S \in F$ as G is not empty and $\emptyset \notin F$ as G has the finite intersection property. By definition, if $X \subseteq Y \subseteq S$ and $X \in F$, then $Y \in F$. Finally, if $X, Y \in F$, then $X \supseteq X_1 \cap \dots \cap X_n$, and $Y \supseteq Y_1 \cap \dots \cap Y_m$, for $X_1, \dots, X_n, Y_1, \dots, Y_m \in G$. Thus $X \cap Y \supseteq X_1 \cap \dots \cap X_n \cap Y_1 \cap \dots \cap Y_m$, so $X \cap Y \in F$. □

We call F as above a filter *generated by* G.

Lemma 13.16. Let S be a non-empty set and let G be a collection of subsets of S having the finite intersection property. If $X \subseteq S$, then either $G \cup \{X\}$ or $G \cup \{S - X\}$ has the finite intersection property.

Proof. Suppose not. Pick $X_1, \dots, X_n \in G$ such that $X_1 \cap \dots \cap X_n \cap X = \emptyset$. Also pick $Y_1, \dots, Y_m \in G$ such that $Y_1 \cap \dots Y_m \cap (S - X) = \emptyset$. We then have:

$$\begin{aligned} X_1 \cap \dots X_n \cap Y_1 \cap \dots \cap Y_m &= (X_1 \cap \dots X_n \cap Y_1 \cap \dots \cap Y_m) \cup (X \cup (S - X)) \\ &\subseteq (X_1 \cap \dots \cap X_n \cap X) \cup (Y_1 \cap \dots \cap Y_m \cap (S - X)) \\ &= \emptyset \end{aligned}$$

Contradicting the hypothesis that G had the finite intersection property. □

Definition 13.17. A filter F on a set S is *maximal* if for any filter F' on S, if $F \subseteq F'$, then $F = F'$.

Theorem 13.18. A filter F on a set S is an ultrafilter if and only if it is maximal.

Proof. First assume that F is an ultrafilter. Let F' be a filter such that $F \subseteq F'$. We show that $F' \subseteq F$. Let $X \in F'$. As F is an ultrafilter, either $X \in F$ or $S - X \in F$.

In the former case we are done. The latter case is not possible: if $S - X \in F$ then $S - X \in F'$, and so $\emptyset = X \cap (S - X) \in F'$, contradiction.

Now assume conversely that F is maximal. We show it is an ultrafilter. Let $X \subseteq S$. As F has the finite intersection property, Lemma 13.16 ensures that either $F \cup \{X\}$ or $F \cup \{S - X\}$ has the finite intersection property. If $F \cup \{X\}$ has the finite intersection property, then by Lemma 13.15 there is a filter F' containing $F \cup \{X\}$. By maximality, $F' = F$, so $X \in F$. Similarly, if $F \cup \{S - X\}$ has the finite intersection property then $S - X \in F$. □

Remark 13.19. Thus any prime ideal in the ring $(\mathcal{P}(S), \Delta, \cap)$ is maximal. This is not true for arbitrary commutative rings (take for example the ideal generated by the variable x in the ring $\mathbb{Z}[x]$ of polynomials with integer coefficient).

Corollary 13.20. *For any filter F on a set S, there exists an ultrafilter U such that $F \subseteq U$.*

The proof uses:

Theorem 13.21 (Zorn's lemma)**.** *If P is a partially ordered set where every chain has an upper bound, then P has a maximal element.*

Proof. Assignment 6. □

Proof of Corollary 13.20. Consider the partial order P of all filters extending F, ordered by $\subseteq$. Every chain C of filters has an upper bound, given by $\bigcup_{F' \in C} F'$. Thus by Zorn's lemma there is a maximal element U in P. By Theorem 13.18, U must be an ultrafilter. □

We finish by giving an application to analysis, and answering the original question about extending density to a measure.

Definition 13.22. Let F be a filter on $\mathbb{N}$. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. We say that a real number a is the *F-limit* of $(a_n)_{n \in \mathbb{N}}$ (written $a = \lim_F a_n$) if for any $\epsilon > 0$, $\{n \in \mathbb{N} \mid |a_n - a| < \epsilon\} \in F$.

In words, the F-limit of a_n is a if the set of n such that a_n is close to a is large. This has a number of good properties:

Exercise 13.23. Let F be a filter on $\mathbb{N}$.

(1) The F-limit is unique if it exists.
(2) If F' is a filter, $F \subseteq F'$, and the F-limit of a sequence $(a_n)_{n \in \mathbb{N}}$ is a, then the F'-limit of the sequence is also a.
(3) If F is the cofinite filter, then the F-limit is the usual limit.
(4) If U is an ultrafilter, then *any* bounded sequence has a U-limit.
(5) If $\lim_F a_n = a$ and $\lim_F b_n = b$, then:
    (a) $\lim_F (a_n + b_n) = a + b$.
    (b) For any $c \in \mathbb{R}$, $\lim_F ca_n = ca$.
    (c) If $a_n \leq b_n$ for all n, then $a \leq b$.

The fact that any bounded sequence has a U-limit when U is an ultrafilter is particularly convenient. For example, the sequence $0, 1, 0, 1, 0, 1, \ldots$ has a U-limit (which will be 0 or 1, depending on whether the set of even or odd numbers is in U). This choice of limit is coherent, in the sense that for example the additivity and monotonicity properties of limits are preserved. We can deduce the existence of a measure extending the density:

Theorem 13.24. There exists a measure m on $\mathbb{N}$ such that $m(A) = d(A)$ whenever $A \subseteq \mathbb{N}$ has a density.

Proof. Let U be an ultrafilter extending the cofinite filter on $\mathbb{N}$ (exists by Theorem 13.20). Define $m : \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ by $m(A) = \lim_U \frac{|A \cap \{0,1\ldots,n\}|}{n+1}$. This limit exists since the corresponding sequence is bounded (between 0 and 1). Since U extends the cofinite filter and the limit with respect to the cofinite filter is the usual one, we get that $m(A) = d(A)$ whenever $A \subseteq \mathbb{N}$ has a density. In particular, $m(\mathbb{N}) = 1$ and $m(\emptyset) = 0$. Monotonicity of m is also straightforward from the corresponding property of the U-limit. Similarly, finite additivity of m follows from the additivity property of the U-limit. □

14. Clubs and stationary sets

The examples in the previous section concentrated on $\mathbb{N}$. In this section, we study a very interesting example of a filter on regular uncountable cardinals: the club filter. We will see that this filter behaves very differently from the usual filters on the natural numbers.

Definition 14.1. Let λ be a regular uncountable cardinal (for example, $\lambda = \omega_1$).

(1) A set $C \subseteq \lambda$ is *closed* if for every limit ordinal $\delta < \lambda$ and every increasing sequence $(\alpha_i)_{i<\delta}$ with $\alpha_i \in C$ for all $i < \delta$, we have that $\sup_{i<\delta} \alpha_i \in C$.
(2) A set $X \subseteq \lambda$ is *unbounded* (in λ) if for every $\alpha < \lambda$, there exists $\beta \in X$ such that $\alpha < \beta$.
(3) A set $C \subseteq \lambda$ is *club* (in λ) if it is both closed and unbounded.

The definition of closed coincides with the notion of a closed set according to the *order topology* on λ: open sets are unions of open intervals, where an open interval is a set of the form $(\alpha, \beta) := \{\gamma < \lambda \mid \alpha < \gamma < \beta\}$. Note that this is not very interesting when $\lambda = \omega$, which is why we restricted ourselves to uncountable regular cardinals. Note also that an unbounded set is the same as a cofinal set. Thus in particular club subsets of λ are cofinal, so (by regularity of λ) must have cardinality λ. However the fact that they are closed gives us that they are in some sense the "big" cofinal sets.

Example 14.2. For definiteness, set $\lambda = \omega_1$ (although the examples generalize to any regular uncountable cardinal λ).

(1) For any ordinal $\alpha < \omega_1$, the set $[\alpha, \omega_1)$ of all $\beta < \omega_1$ so that $\alpha \leq \beta$ is club (in ω_1). In particular, $[0, \omega_1) = \omega_1$.
(2) The set of all countable limit ordinals is club. It is unbounded, and the supremum of an increasing sequence of limit ordinals must be a limit ordinal (exercise!).
(3) The set of all countable successor ordinals is *not* club. It is unbounded, but not closed: for example the sequence $1, 2, 3, \ldots$ is a sequence of successor ordinals with supremum ω, which is not a successor ordinal.
(4) It is *not* true that a superset of a club is a club. For example the set of all countable ordinals that are either limit and distinct from ω or finite is not club, even though the set of all countable limit ordinals distinct from ω is a club.

Club sets can also be characterized in terms of how they are enumerated:

Definition 14.3. Let λ be a regular uncountable cardinal and let $C \subseteq \lambda$. A *normal enumeration of* C is a sequence $(\alpha_i)_{i<\lambda}$ such that:

(1) $C = \{\alpha_i \mid i < \lambda\}$.
(2) The sequence is strictly increasing: $i < j < \lambda$ implies $\alpha_i < \alpha_j$.
(3) The sequence is *continuous*: if $\delta < \lambda$ is a limit ordinal, then $\alpha_\delta = \sup_{i<\delta} \alpha_i$.

Lemma 14.4. Let λ be a regular uncountable cardinal. A subset $C \subseteq \lambda$ is club if and only if it has a normal enumeration.

Proof. If C is club, inductively let $\alpha_i := \min(C - \{\alpha_j \mid j < i\})$. Note that as λ is regular and C is unbounded, $|C| = \lambda$, so α_i is well-defined for any $i < \lambda$, and as $C \subseteq \lambda$, $C = \{\alpha_i \mid i < \lambda\}$ (another way to see this is to look at the Mostowski collapse $\pi : C \to \lambda$, and observe that $\alpha_i = \pi^{-1}(i)$). The sequence is strictly increasing by definition. Continuity follows from the fact that C is closed.

The converse is left as an exercise. □

As the previous example shows, the club sets do *not* form a filter. They do however *generate* a filter, as they satisfy a strong form of the finite intersection property:

Theorem 14.5. Let λ be a regular uncountable cardinal. If A and B are club, then $A \cap B$ is club.

Proof. It is straigthforward to check that $A \cap B$ is closed (in general, any intersection of closed sets is closed). The proof of unboundedness is more interesting (and would not generalize to $\lambda = \omega$!). Fix $\alpha < \lambda$. We have to find $\beta \in A \cap B$ such that $\alpha < \beta$. As A is unbounded, we know there is $\beta_0 \in A$ such that $\alpha < \beta_0$. β_0 may not be in B, since B is unbounded there is $\beta_1 \in B$ with $\beta_0 < \beta_1$. Whereas β_0 was in A, β_1 may no longer be in A. So we find $\beta_2 \in A$ with $\beta_1 < \beta_2$. We continue in this way to build an increasing sequence $(\beta_n)_{n<\omega}$ of ordinals such that $\alpha < \beta_0$, $\beta_n \in A$ if n is even, and $\beta_n \in B$ if n is odd. Let $\beta := \sup_{n<\omega} \beta_n$. Note that $\beta = \sup_{k<\omega} \beta_{2k}$, so because A is closed $\beta \in A$. Similarly, $\beta = \sup_{k<\omega} \beta_{2k+1}$, so $\beta \in B$. Thus $\beta \in A \cap B$ is the ordinal we were looking for. □

Definition 14.6. For a regular uncountable cardinal λ, the *club filter on* λ is the set of all $X \subseteq \lambda$ such that there is a club C with $C \subseteq X$.

Corollary 14.7. The club filter on a regular uncountable λ is indeed a filter.

Proof. The set λ of *all* ordinals below λ is clearly club. The empty set is not club (it is closed, but not unbounded). If a set X contains a club, then any extension of it contains a club. Finally, the fact that the intersection of two clubs is a club (Theorem 14.5) shows that the clubs have the finite intersection property, so as in the proof of Lemma 13.15, the club filter is closed under intersections. □

The proof of Theorem 14.5 in fact generalizes to any intersections of size less than λ. For example, the union of countably many clubs is always a club:

Theorem 14.8. Let λ be a regular uncountable cardinal. If $\alpha < \lambda$ is a nonzero ordinal and $(C_i)_{i<\alpha}$ is a sequence of clubs, then $\bigcap_{i<\alpha} C_i$ is club.

Proof. Exercise. □

The theorem of course does not generalize to the intersection of λ or more clubs. Consider for example for $\alpha < \lambda$ the club $C_\alpha := [\alpha, \lambda)$. We have that $\bigcap_{\alpha<\lambda} C_\alpha = \emptyset$ which is not club. There is however a way to push the result of Theorem 14.8 a little bit farther:

Definition 14.9. Let λ be a regular uncountable cardinal and let $(X_\alpha)_{\alpha<\lambda}$ be a sequence of subsets of λ. The *diagonal intersection* of the X_α's, written $\triangle_{\alpha<\lambda} X_\alpha$, is the set $\{\gamma < \lambda \mid \gamma \in \bigcap_{\alpha<\gamma} X_\alpha\}$. Dually, the *diagonal union* of the X_α's, written $\triangledown_{\alpha<\lambda} X_\alpha$, is the set $\{\gamma < \lambda \mid \gamma \in \bigcup_{\alpha<\gamma} X_\alpha\}$.

In words, an ordinal γ is in the diagonal intersection if it is in all the "previous" sets: those indexed by an ordinal before γ.

Example 14.10. Let λ be a regular uncountable cardinal.

(1) $\triangle_{\alpha<\lambda}[\alpha, \lambda) = \lambda$.
(2) If $0 < \alpha < \lambda$ and $(X_\gamma)_{\gamma<\alpha}$, define $X_\gamma := \lambda$ for $\gamma \geq \alpha$. Let $X := \triangle_{\gamma<\lambda} X_\gamma$. Then $\bigcap_{\gamma<\alpha} X_\gamma \subseteq X \subseteq [0, \alpha) \cup \bigcap_{\gamma<\alpha} X_\gamma$.

Theorem 14.11. Let λ be a regular uncountable cardinal. If $(C_\alpha)_{\alpha<\lambda}$ is a sequence of clubs, then $\triangle_{\alpha<\lambda} C_\alpha$ is club.

Proof. Exercise. □

The dual ideal of the club filter is called the *nonstationary ideal.* Let us first introduce stationary sets:

Definition 14.12. Let λ be a regular uncountable cardinal.

(1) A set $S \subseteq \lambda$ is *stationary* if it intersects every club: for every club $C \subseteq \lambda$, $S \cap C \neq \emptyset$.
(2) The *nonstationary ideal on* λ is the set of all $X \subseteq \lambda$ that are *not* stationary.

Note that if X is in the nonstationary ideal, then there is a club C such that $X \cap C = \emptyset$, hence $C \subseteq \lambda - X$, so $\lambda - X$ is in the club filter. Conversely, if Y is in the club filter, as witnessed by the club $C \subseteq Y$, then $(\lambda - Y) \cap C = \emptyset$, so $\lambda - Y$ is in the nonstationary ideal. Thus the nonstationary ideal is indeed the dual of the club filter. In particular, it is really an ideal.

One analogy that may be helpful to keep in mind is the following: sets in the club filter are those of measure 1. Stationary sets are those of strictly positive measure, and nonstationary sets are those of measure 0.

Example 14.13. Let λ be a regular uncountable cardinal.

(1) Any club is stationary (because the intersections of two clubs is club, so in particular not empty).
(2) If S is stationary and $S \subseteq T$, then T is stationary.
(3) If S is stationary and C is club, then $S \cap C$ is in fact stationary (not just nonempty): indeed for any other club D, $C \cap D$ is club, so $(S \cap C) \cap D = S \cap (C \cap D)$ has to be nonempty.
(4) The empty set is not stationary. More generally, any bounded set (i.e. any subset of α for some ordinal $\alpha < \lambda$) is not stationary: the set $[\alpha, \lambda)$ is club for all $\alpha < \lambda$.
(5) The set of all successor ordinals is an unbounded set which is not stationary: it is disjoint from the club of all limit ordinals.

(6) If $\theta < \lambda$ is an infinite regular cardinal, $S_\theta := \{\delta < \lambda \mid \mathrm{cf}(\delta) = \theta\}$ is stationary. To see this, let C be a club of λ. Let $(\alpha_i)_{i<\lambda}$ be a normal enumeration of C. Then $\alpha_\theta = \sup_{i<\theta} \alpha_i$ is the desired ordinal in $C \cap S_\theta$.

(7) If $\lambda > \aleph_1$ then $S_{\aleph_0}$ and $S_{\aleph_1}$ (see above) are disjoint stationary sets. This implies, in this case:
   (a) That stationary sets do *not* form a filter.
   (b) That S_θ does *not* contain a club for any regular infinite $\theta < \lambda$: if it did then it would have to intersect both $S_{\aleph_0}$ and $S_{\aleph_1}$.
   (c) That the club filter F is not an ultrafilter. We have just argued that $S_{\aleph_0} \notin F$. However $S_{\aleph_0}$ meets every element of F, so its complement cannot be in F either.

   When $\lambda = \aleph_1$, there is only one regular infinite cardinal below $\aleph_1$, which is $\aleph_0$. Thus $S_{\aleph_0}$ is just the set of all limit ordinals, hence is club. It nevertheless remains true that some stationary subsets of ω_1 are not club, hence the first and third item remain true, although the proof is harder (we will do it later).

15. Fodor's lemma and its applications

The most important fact about stationary sets is Fodor's lemma. To introduce it, we need the notion of a regressive function.

Definition 15.1. Let λ be a regular uncountable cardinal and let $X \subseteq \lambda$. A function $f : X \to \lambda$ is *regressive* if $f(\alpha) < \alpha$ for every nonzero $\alpha \in X$.

One example of a regressive function is, for each ordinal $\alpha < \lambda$, the constantly α function $f : [\alpha + 1, \lambda) \to \lambda$ defined by $f(\beta) = \alpha$ for all $\beta \in [\alpha + 1, \lambda)$. Fodor's lemma tells us that any function that is regressive on a stationary set has to be constant on a stationary set.

Theorem 15.2 (Fodor's lemma)**.** Let λ be a regular uncountable cardinal and let S be a stationary set. If $f : S \to \lambda$ is regressive, then there exists a stationary $T \subseteq S$ such that $f \restriction T$ is constant.

Proof. For each $\alpha < \lambda$, let $S_\alpha := f^{-1}[\{\alpha\}]$. Suppose for a contradiction that S_α is not stationary for any $\alpha < \lambda$. Let $S' := \bigtriangledown_{\alpha<\lambda} S_\alpha$ be the diagonal union of the S_α's (Definition 14.9). By the dual of Theorem 14.11, S' must be nonstationary. However, $S' = S$: by definition, $S' \subseteq S$, and if $\gamma \in S$, then $\beta := f(\gamma) < \gamma$ by assumption, hence $\gamma \in f^{-1}[\{\beta\}] = S_\beta$, so as $\beta < \gamma$, $\gamma \in \bigtriangledown_{\alpha<\lambda} S_\alpha$. □

Let us now consider several applications of Fodor's lemma.

15.1. The transfinite subway.

We start with something recreational! Consider a subway going from the airport (station 0) to the Hilbert[9] hotel (station ω_1). That's right, there are $\aleph_1$-many stations, numbered with each ordinal $\alpha \leq \omega_1$. The subway goes through each station in turn. At the airport, the subway arrives empty and a countable infinity of passengers board. In the next station, exactly one passenger gets off and another countable infinity boards (as you can see, this is a very realistic

[9] The Hilbert hotel is the infinite hotel where you can always make space by shifting the rooms by one: if the rooms are numbered $0, 1, 2, \ldots$, the hotel is full, and a new guest comes along, one can simply ask the guests at room n to relocate to room $n + 1$, freeing room 0 for the new customer.

model). In general, at station α, exactly one passenger gets off (if the subway is not empty, otherwise of course nobody gets off), and then a countable infinity of passengers boards the subway. We assume that somebody who gets off the subway *cannot* board the subway again at a later station.

The question is: how many people are in the subway when it arrives at the Hilbert hotel (station ω_1)?

It may seem like there should be an uncountable infinity of passengers, just because more get in at each station than get off, and there are uncountably-many stations. However this intuition is wrong: ω_1 is way too big for such thinking.

Let us analyze what happens at station α. If the subway is empty, nobody gets off. If the subway is not empty, we know exactly one passenger gets off. This passenger had to get in at some station, say it is station $\beta < \alpha$ (there is only one such station by the assumption that the same passenger cannot get in at multiple stations). Thus let us define a function $f : \omega_1 \to \omega_1$ by $f(\alpha) = \alpha$, if the subway arrives empty at station α, or $f(\alpha) = \beta$ if the subway arrives non-empty at station α and the passenger who gets off boarded at station β.

Let $S := \{\alpha < \omega_1 \mid f(\alpha) < \alpha\}$ be the set of stations where the subway arrives non-empty. It may seem like the subway will always arrive non-empty – except at the airport – and so S should be very big but this is completely wrong: we claim that S is not even stationary. Indeed if S were stationary, then by Fodor's lemma applied to $f \restriction S$ we would get that there exists a stationary $S_0 \subseteq S$ such that $f \restriction S_0$ is constant. Stationary sets are unbounded (Example 14.13), so by regularity of ω_1, S_0 is in particular uncountable. Say $f(\alpha) = \beta$ for all $\alpha \in S_0$. Since by assumption at distinct stations distinct passengers get off, this means that uncountably-many passengers boarded at station β, a contradiction.

Since S is not stationary, this means that its complement $C := \{\alpha < \omega_1 \mid f(\alpha) = \alpha\}$ is in the club filter (in fact one can argue it is club but this is not needed). In particular, there is an unbounded set of stations $\alpha < \omega_1$ so that the subway arrives empty at station α. We conclude that the subway also arrives empty at the Hilbert hotel! Indeed, if a passenger is still on the subway at the Hilbert hotel, it must have boarded the subway at a certain station α, but then would have had to exit it by station $\min(C - (\alpha + 1))$, way before the Hilbert hotel.

15.2. **Splitting a stationary set.** If stationary sets are those of strictly positive measure, then it makes sense to ask whether given a stationary S, there could exist stationary $S_1, S_2 \subseteq S$ disjoint (analogous to splitting the interval $(0, 0.5)$ into $(0, 0.2)$ and $(0.3, 0.5)$, say). We saw that in some special cases this could be done, for example if $\lambda = S = \omega_2$ we can look at the ordinals of cofinality ω and of cofinality ω_1 and they form disjoint stationary subsets of S. When $\lambda = \omega_1$, it is already not so clear how to split S (and in this case S is club so of measure 1!). We show that this is always possible. In fact, S can be split into λ-many disjoint stationary pieces.

Fact 15.3 (Solovay splitting theorem)**.** Let λ be a regular uncountable cardinal. If S is a stationary subset of λ, then there exists λ-many pairwise disjoint stationary subsets of λ.

Corollary 15.4. The club filter on any regular uncountable cardinal is not an ultrafilter.

Proof. If λ is a regular uncountable cardinal. We can in particular split λ into two disjoint stationary sets S and T. Then S is not in the club filter, as otherwise we would have $T \cap S \neq \emptyset$. However, $\lambda - S$ is not in the club filter either since S meets every club. □

Some cases of the Solovay splitting theorem are easier (and were known before Solovay proved the general case). In fact, we showed already that the club filter is not an ultrafilter for $\lambda \geq \omega_2$ (Example 14.13). Thus for the corollary it suffices to prove the case $\lambda = \omega_1$. We focus on it first for concreteness.

Lemma 15.5. For any stationary set $S \subseteq \omega_1$, there are $\aleph_1$-many pairwise disjoint stationary subsets of S.

Proof. Without loss of generality, S consists only of limit ordinals (since the limit ordinals form a club their intersection with S is stationary). For each $\alpha \in S$, fix $(\alpha_n)_{n<\omega}$ a cofinal sequence in α (this is a slight abuse of notation: both α and n can vary!).

After fixing this "data", let us see how the cofinal sequences interact. For each $n < \omega$ and each $\gamma < \omega_1$, let $X_n^\gamma := \{\alpha \in S \mid \alpha_n = \gamma\}$. In words, this is the set of α where the nth point in a cofinal sequence of α is equal to γ.

What happens if we fix n but change γ? Well, $X_n^\gamma \cap X_n^{\gamma'} = \emptyset$ for $\gamma \neq \gamma'$. We can also say that $\bigcup_{\gamma<\omega_1} X_n^\gamma = S$: any α_n has to be some γ. Can we say more? Well, for each fixed $n < \omega$, the map $\alpha \mapsto \alpha_n$ is regressive as a map from S to ω_1. By Fodor's lemma, there exists $S' \subseteq S$ stationary and $\gamma < \omega_1$ so that $\alpha_n = \gamma$ for all $\alpha \in S'$. In other words, X_n^γ is stationary.

It would be great it there could be several γ's such that X_n^γ is stationary. After some thoughts, the α_n's are increasing, so they ought to eventually go past any fixed γ. It could be however that $\alpha_0 = 0$ for all α, but varying n we should be able to get what we want.

More precisely, for each fixed $\gamma < \omega_1$, there exists $n = n_\gamma < \omega_1$ such that $S_{n,\geq\gamma} := \{\alpha \in S \mid \alpha_n \geq \gamma\}$ is stationary. If not, there exists $\gamma < \omega_1$ so that for each $n < \omega$, $C_n := \{\alpha \in S \mid \alpha_n < \gamma\}$ is club. Thus $\bigcap_{n<\omega} C_n = \{\alpha \in S \mid \sup_{n<\omega} \alpha_n \leq \gamma\}$ is also club, but the latter is just $\alpha \in S \mid \alpha \leq \gamma\} \subseteq [0, \gamma]$, which is not club, contradiction.

Now we can do even more prunning: consider the map $\gamma \mapsto n_\gamma$ as a function from ω_1 to ω. By the pigeonhole principle (make this precise as an exercise, or use Fodor's lemma again!), there exists an unbounded $W \subseteq \gamma$ and $n < \omega$ such that $n = n_\gamma$ for all $\gamma \in W$. In other words, $|W| = \omega_1$. Now let us show that $\{\gamma < \omega_1 \mid X_n^\gamma \text{ is stationary}\}$ is unbounded, and this will give us what we want: then the X_n^γ's such that X_n^γ is stationary partition S into ω_1-many stationary sets. Fix $\gamma_0 < \omega_1$. Pick $\gamma \in W$ with $\gamma_0 < \gamma$. Consider the map $\alpha \mapsto \alpha_n$ from $S_{n,\geq\gamma}$ to ω_1. It is regressive, so by Fodor's lemma there exists ρ such that $\alpha_n = \rho$ for stationarily-many α's. By construction, $\rho \geq \gamma$ and X_n^ρ is stationary, as desired. □

The proof generalizes to split any set of ordinals of a fixed cofinality:

Definition 15.6. For $\kappa < \lambda$ both regular infinite cardinals, let $S_\kappa^\lambda := \{\delta < \lambda \mid \text{cf}(\delta) = \kappa\}$.

Lemma 15.7. If $\kappa < \lambda$ are regular infinite cardinals, then any stationary subset of S_κ^λ has λ-many pairwise disjoint stationary subsets.

Proof. Exercise. Generalize Lemma 15.5. □

Proof of Solovay's theorem if λ is successor. Let S be a stationary subset of λ. Without loss of generality, S contains only limit ordinals (intersect it with the club of all limits, if needed). Say $\lambda = \mu^+$. Let X be the set of all regular cardinals below λ. We have that $X \subseteq \mu + 1$, so $|X| \leq \mu$. Moreover $S = \bigcup_{\kappa \in X}(S^\lambda_\kappa \cap S)$. If all of the components of this union were nonstationary, then we would have a union of fewer than λ-many nonstationary sets, which has to be nonstationary by the dual of the fact that an intersection of fewer than λ-many club is club. This is impossible, as S is stationary. Thus for some $\kappa \in X$, $S^\lambda_\kappa \cap S$ is stationary, and we can apply Lemma 15.7 to it. □

In case λ is limit, the proof of Solovay's theorem is also within reach but a little bit tedious so we omit it (you can find it, for example, in Jech's graduate set theory book).

Let us simply observe that if there exists $S \subseteq \lambda$ stationary such that $S \cap S^\lambda_\kappa$ is nonstationary for any infinite regular $\kappa < \lambda$, then λ has to be quite a large cardinal. First, we have seen that λ has to be limit. An uncountable regular limit cardinal is called a *weakly inaccessible cardinal*, and the existence of a weakly inaccessible cardinal is already unprovable from the axioms of set theory. However the situation is much worse. Let $S_0 := \{\alpha \in S \mid \mathrm{cf}(\alpha) = \alpha\}$ be the set of regular cardinals in S. We have that S_0 has to be stationary! If not, then $S_1 := \{\alpha \in S \mid \mathrm{cf}(\alpha) < \alpha\}$ is stationary, hence the map $\alpha \mapsto \mathrm{cf}(\alpha)$ is regressive on S_1, so by Fodor's lemma must be constant on a stationary subset S_1' of S_1. In particular, $S_1' \subseteq S^\lambda_\kappa$ for some κ, a contradiction to the assumption that $S \cap S^\lambda_\kappa$ is not stationary.

A regular uncountable cardinal λ such that $S_0^* := \{\alpha < \lambda \mid \alpha = \mathrm{cf}(\alpha)\}$ is stationary is called a *weakly Mahlo cardinal.* Weakly Mahlo cardinals are weakly inaccessible, and therefore the set C of limit cardinals below λ is club! Thus $S_0^* \cap C$ has to be stationary as well: not only is λ weakly inaccessible, but it contains a stationary set of weakly inaccessible cardinals. In particular it cannot be the first weakly inaccessible (in fact it is the λth weakly inaccessible). Nevertheless, there are good reasons to believe that weakly Mahlo cardinals should exist.

In conclusion, we have really shown that Solovay's theorem holds for a lot of cardinals, definitely for all the cardinals that can be proven to exist from the usual axioms of set theory.

15.3. **The Δ-system lemma.** As opposed to Fodor's lemma, Solovay's splitting theorem, or even the transfinite subway puzzle, the next result could have been stated on the first day of this class.

Theorem 15.8 (Δ-system lemma)**.** If $\{A_i \mid i \in I\}$ is an uncountable collection of finite sets, then there exists $J \subseteq I$ uncountable and a set A such that $A_i \cap A_j = A$ for all $i \neq j$ both in J.

A collection of sets whose pairwise intersections are constant (as in the conclusion of the theorem) is called a *Δ-system*, or a *sunflower* (draw a Venn diagram if you want to know where the name comes from!). There is also a version of the Δ-system lemma in finite combinatorics: for any natural number n and k, there is a very big natural number M so that any collection of M-many sets each of cardinality at most k contains a Δ-system of size n. The number M can be computed from n and k ($M = k!n^{k+1}$ suffices; but it is not known whether this upper bound is sharp).

It may be fun to try proving the Δ-system lemma on your own first. Surprisingly, Fodor's lemma makes it very easy (even though there are elementary proofs that do not need it). The trick is to rename everything so that we deal with sets of ordinals.

Proof of the Δ-system lemma. Without loss of generality, $I = \omega_1$. Also without loss of generality, $A_i \subseteq \omega_1$ for all $i < \omega_1$. The map $i \mapsto |A_i|$ is regressive on a stationary set (in fact on a club set), so by Fodor must be constant on a stationary set. In fact we only need it to be constant on a set of cardinality $\aleph_1$, so we could also appeal to the pigeonhole principle. Thus without loss of generality there is a natural number n such that $n = |A_i|$ for all $i < \omega_1$.

We want to "spread out" the A_i's along ω_1. Thus let $C := \{\alpha < \omega_1 \mid \max(A_i) < \alpha \text{ for all } i < \alpha\}$. You should be able to check that C is club. For $k \leq n$, let $S_k := \{\alpha \in C \mid |A_\alpha \cap \alpha| = k\}$. We know that $S_0 \cup S_1 \cup \ldots \cup S_n = C$, so since the nonstationary ideal is closed under union, there must be $k \leq n$ so that S_k is stationary. For $\alpha \in S_k$ and any fixed $m < k$, the map $f_m : S_k \to \omega_1$ sending α to the mth element of A_α is regressive by definition of S_k. Applying Fodor's lemma to f_0, we get $S^0 \subseteq S_k$ stationary so that $f_0 \upharpoonright S^0$ is constant. Now apply Fodor's lemma to $f_1 \upharpoonright S^0$ to get $S^1 \subseteq S^0$ stationary so that $f_1 \upharpoonright S^1$ is constant. Continue in this way and get a stationary $S \subseteq S_k$ so that $f_m \upharpoonright S$ is constant for all $m < k$. Let a_m be the value that f_m takes on S. Let $A := \{a_m \mid m < k\}$. In words, for each $\alpha \in S$, a_m is the mth element of A_α. This means that $A \subseteq A_\alpha \cap A_\beta$ whenever $\alpha \neq \beta$ are in S. It still could be the intersection is bigger, but by definition of S_k, we know that the elements k, $k+1$, ..., $n-1$ of S_α are all at least α. Thus we proceed as follows: let $\alpha_0 \in S$, and let $\alpha_1 > \max(A_{\alpha_0})$ be in S. Then the elements k, $k+1$, ..., n of A_{α_1} are all at least α_1, whereas those of A_{α_0} are strictly below α_1, so $A_{\alpha_0} \cap A_{\alpha_1} = A$. We continue in this way, defining inductively $(\alpha_i)_{i<\omega_1}$ all in S such that $\alpha_i > \max(A_{\alpha_j})$ for all $j < i$. This is possible by regularity of ω_1. In the end, take $J := \{\alpha_i \mid i < \omega_1\}$ and $\{A_j \mid j \in J\}$ is the desired Δ-system. □

15.4. **Silver's theorem.** The final application of Fodor's theorem we investigate is to cardinal arithmetic! For λ a regular (infinite) cardinal, it is known that the value of 2^λ can be basically anything as long as $\text{cf}(2^\lambda) > \lambda$ (König's theorem), and $\mu < \lambda$ implies $2^\mu \leq 2^\lambda$. For example the axioms of set theory do not disprove that $2^{\aleph_0} = \aleph_{32}$ and $2^{\aleph_1} = \aleph_{\omega_{59}+43}$.

If λ is a *singular* (infinite) cardinal, the story is different. Of course we must still have that for $\mu < \lambda$, $2^\mu \leq 2^\lambda$, but it turns out that the value of 2^λ cannot "jump" too far from $\sup_{\mu<\lambda} 2^\mu$. Making this precise and finding the sharpest theorems is still an ongoing research program in set theory. The first singular infinite cardinal is $\aleph_\omega$, and Saharon Shelah proved not too long ago that if $2^{\aleph_n} < \aleph_\omega$ for all $n < \omega$, then $2^{\aleph_\omega} < \aleph_{\omega_4}$ (why 4? Is it sharp? nobody knows). This is a delicate argument because ω is small and we cannot take advantage of the powerful club filter! Thus we are going to focus instead on $\aleph_{\omega_1}$, the first singular cardinal of uncountable cofinality.

The simplest possible behavior of cardinal exponentiation is given by the generalized continuum hypothesis (GCH), which says that $2^\lambda = \lambda^+$ for any infinite cardinal λ (as small as it could be, by Cantor's theorem). For an infinite cardinal λ, we will say that *the GCH holds below* λ if $2^\mu = \mu^+$ for all infinite $\mu < \lambda$. We will prove:

Theorem 15.9 (Silver's theorem)**.** If the GCH holds below $\aleph_{\omega_1}$, then it holds at $\aleph_{\omega_1}$: $2^{\aleph_{\omega_1}} = \aleph_{\omega_1+1}$.

Note that there is nothing special about $\aleph_{\omega_1}$ here: Silver's theorem holds for any singular cardinal of uncountable cofinality. Nevertheless, we focus on $\aleph_{\omega_1}$ for notational simplicity. The proof will study the following objects:

Definition 15.10. Two functions (or sequences) f, g with domain ω_1 are *almost disjoint* if there exists $\alpha < \omega_1$ such that for all $\beta \in [\alpha, \omega_1)$, $f(\beta) \neq g(\beta)$. A collection F of functions (or sequences) is *almost disjoint* if any two distinct elements of F are almost disjoint.

The name comes from the notion of almost disjoint sets: sets whose intersection is bounded (say in ω_1). Two sets are almost disjoint if and only if their characteristic functions are almost disjoint.

To estimate $2^{\aleph_{\omega_1}}$, it suffices to estimate the size of certain almost disjoint families:

Lemma 15.11. Let μ be a cardinal such that any almost disjoint family $F \subseteq \prod_{\alpha<\omega_1} \mathcal{P}(\aleph_\alpha)$ has cardinality at most μ. Then $2^{\aleph_{\omega_1}} \leq \mu$

Proof. For each $X \subseteq \aleph_{\omega_1}$, let $f_X \in \prod_{\alpha<\omega_1} \mathcal{P}(\aleph_\alpha)$ be given by $f_X(\alpha) := X \cap \aleph_\alpha$. Let $F := \{f_X \mid X \subseteq \aleph_{\omega_1}\}$. The map $X \mapsto f_X$ is an injection from $\mathcal{P}(\aleph_{\omega_1})$ into F: if $X \neq Y$ then this is witnessed by some element in their symmetric difference which by definition of $\aleph_{\omega_1}$ must be in $\aleph_\alpha$ for some $\alpha < \omega_1$, and so $X \cap \aleph_\alpha \neq Y \cap \aleph_\alpha$. In fact, $X \cap \aleph_\beta \neq Y \cap \aleph_\beta$ for all $\beta \in [\alpha, \omega_1)$, which shows that F is an almost disjoint family. The result follows. □

We now work on estimating sizes of almost disjoint families. The key lemma below uses Fodor's lemma.

Lemma 15.12. Assume GCH below $\aleph_{\omega_1}$. Let $(A_\alpha)_{\alpha<\omega_1}$ be sets such that $|A_\alpha| \leq \aleph_\alpha$ for all $\alpha < \omega_1$. If $F \subseteq \prod_{\alpha<\omega_1} A_\alpha$ is a family of almost disjoint functions, then $|F| \leq \aleph_{\omega_1}$.

Proof. By renaming, we can assume without loss of generality that $A_\alpha \subseteq \aleph_\alpha$ for all $\alpha < \omega_1$. Fix $f \in F$. For all $\alpha < \omega_1$, we have $f(\alpha) \in \aleph_\alpha$, hence $f(\alpha) < \aleph_\alpha$. Let S be the set of all countable limit ordinals. If $\alpha \in S$, then as $\aleph_\alpha = \sup_{\beta<\alpha} \aleph_\beta$, we must have that $f(\alpha) < \aleph_\beta$ for some $\beta < \alpha$. Let $f^*(\alpha)$ be the least such β. Then $f^* : S \to \omega_1$ is a regressive function on a stationary set, so by Fodor's lemma there exists $S_f \subseteq S$ stationary and $\gamma_f < \omega_1$ such that $f^*(\alpha) = \gamma$ for all $\alpha \in S_f$. In particular, $f \restriction S_f$ has range contained in $\aleph_\gamma$. Let $F' := \{f \restriction S_f \mid f \in F\}$. The map $\phi : F \to F'$ given by $\phi(f) = f \restriction S_f$ is an injection: if $f \neq g$ are in F, then by almost disjointness for all high-enough α, $f(\alpha) \neq g(\alpha)$, so as S_f is stationary (so unbounded) there exists such an α in S_f, so (assuming $S_g = S_f$ – otherwise $\phi(f) \neq \phi(g)$ as they have different domains) $f \restriction S_f \neq g \restriction S_f$. Therefore $|F| \leq |F'|$. An element of F' is determined by its domain (a subset of ω_1) and a function from this domain to $\aleph_\gamma$, for some $\gamma < \omega_1$. Thus:

$$|F| \leq |F'| \leq 2^{\aleph_1} \sum_{\gamma<\omega_1} \aleph_\gamma^{\aleph_1} \leq \aleph_2 \aleph_{\omega_1} = \aleph_{\omega_1}$$

□

The following sharpening is given by the proof:

Lemma 15.13. Assume GCH below $\aleph_{\omega_1}$. Let $(A_\alpha)_{\alpha<\omega_1}$ be sets such that $\{\alpha < \omega_1 \mid |A_\alpha| \leq \aleph_\alpha\}$ is stationary. If $F \subseteq \prod_{\alpha<\omega_1} A_\alpha$ is a family of almost disjoint functions, then $|F| \leq \aleph_{\omega_1}$.

Proof. Same proof as Lemma 15.12: instead of starting with S we start with the set of limit ordinals $\alpha < \omega_1$ so that $|A_\alpha| \leq \aleph_\alpha$. This is the intersection of a stationary set with a club, hence a stationary set. □

The proof of Lemma 15.12 is powerful, but does not quite bring us all the way to Silver's theorem yet, since Lemma 15.11 allows the αth component of the sequence to be in $\mathcal{P}(\aleph_\alpha)$ rather than $\aleph_\alpha$. Note that $|\mathcal{P}(\aleph_\alpha)| = \aleph_{\alpha+1}$ under the GCH hypothesis. Thus we have to prove:

Lemma 15.14. Assume GCH below $\aleph_{\omega_1}$. Let $(A_\alpha)_{\alpha<\omega_1}$ be sets such that $|A_\alpha| \leq \aleph_{\alpha+1}$ for all $\alpha < \omega_1$. If $F \subseteq \prod_{\alpha<\omega_1} A_\alpha$ is a family of almost disjoint functions, then $|F| \leq \aleph_{\omega_1+1}$.

Proof. Without loss of generality, $A_\alpha \subseteq \aleph_{\alpha+1}$ for all $\alpha < \omega_1$.

Let U be an ultrafilter on ω_1 extending the club filter. Note that every $S \in U$ must be stationary (S meets every element of U, hence in particular every club!). Define a relation $<$ on F by $f < g$ if $\{\alpha < \omega_1 \mid f(\alpha) < g(\alpha)\} \in U$ (in words, $f(\alpha) < g(\alpha)$ for almost every α, where the precise meaning of "almost every" is given by U). The relation $<$ is a strict linear order on U (exercise). Toward counting F, let us first count for a fixed $f \in F$, how many elements there are before f. Specifically, let $\operatorname{pred}(f) := \{g \in F \mid g < f\}$ be the set of predecessors of f. For $g \in \operatorname{pred}(f)$, let $S = S_g := \{\alpha < \omega_1 \mid g(\alpha) < f(\alpha)\}$. We have that $S \in U$, hence S is stationary. Since for any $\alpha < \omega_1$, $|f(\alpha)| \leq \aleph_\alpha$, we are very close to being able to apply Lemma 15.13. The remaining issue is that the set S depends on g, but for each fixed stationary set T, $\operatorname{pred}(f)_T := \{g \in F \mid g(\alpha) < f(\alpha) \text{ for all } \alpha \in T\}$ has cardinality at most $\aleph_{\omega_1}$ by Lemma 15.13. Now observe that $\operatorname{pred}(f) = \bigcup_{T \text{ stationary}} \operatorname{pred}(f)_T$ and there are at most[10] $2^{\aleph_1}$-many stationary subsets of $\aleph_1$, so $|\operatorname{pred}(f)| \leq 2^{\aleph_1}\aleph_{\omega_1} = \aleph_{\omega_1}$.

We have just shown that $|\operatorname{pred}(f)| \leq \aleph_{\omega_1}$ for any $f \in F$. It follows that $|F| \leq \aleph_{\omega_1+1}$. In general, if L is any linear order where all the predecessor sets have cardinality λ, then L has cardinality at most λ^+ (exercise). □

We can now prove Silver's theorem:

Proof of Silver's theorem. By Lemma 15.11, it suffices to see that any almost disjoint family $F \subseteq \prod_{\alpha<\omega_1} \mathcal{P}(\aleph_\alpha)$ has cardinality at most $\aleph_{\omega_1+1}$. By the GCH hypothesis, $|\mathcal{P}(\aleph_\alpha)| = \aleph_{\alpha+1}$. Thus the result follows from Lemma 15.14 applied to $A_\alpha := \mathcal{P}(\aleph_\alpha)$. □

16. Universes for set theory

In this section, we build universes for set theory, with the goal of studying the compatibility of the axiom of choice or the continuum hypothesis with the other axioms.

[10] In fact there are exactly that many. Can you see why?

16.1. **Demystifying independence.** There is nothing mystical about the term "model for set theory" and the fact that the continuum hypothesis or the axiom of choice cannot be proven from the other axioms. Such phenomenons occur all the times in mathematics. Consider for example the axioms of an ordering. We could present them this way: there is an "undefined binary relation" $\leq$ and it has to satisfy three axioms: $x \leq x$ for all x, $x \leq y$ and $y \leq z$ implies $x \leq z$ for all x, y, z, and $x \leq y$, $y \leq x$ implies $x = y$.

A *model* for these axioms is intuitively a "universe" (including a concrete interpretation of the relation $\leq$) where the axioms hold. In the case of orderings, such a model is nothing but what we call a partially ordered set. For example, $(\mathbb{N}, \leq)$ and $(\mathcal{P}(\{1, 2, 3\}), \subseteq)$ are models for the axioms of orderings.

Consider now the property that $\leq$ is total: $x \leq y$ or $y \leq x$ for any x, y. It is impossible to prove this property from the axioms of orderings. It is quite hard to prove this impossibility directly (one would have to make precise what a proof is, and then somehow consider all possible proofs). An easier way is to observe that, whatever a proof is, if a property were provable from the axioms, then in any model of the axioms the property should hold too! Hence it suffices to exhibit a model for the axioms that does not satisy linearity (such as $(\mathcal{P}(\{1, 2, 3\}), \subseteq)$, and this will show linearity is impossible to prove. If we exhibit a model for linearity (such as $(\mathbb{N}, \leq)$), this will show that nonlinearity is impossible to prove, or said another way that linearity is impossible to disprove. Thus linearity is independent of the axioms of ordering. So many words for this straightforward fact!

When dealing with set theory, the situation is conceptually similar but of course much harder in practice. One difficulty is that we are studying set theory within set theory (i.e. our background axioms for mathematics are set theory). Another difficulty is that it is quite hard to build models for set theory (the axioms are much more complicated). Nevertheless, it helps to keep the simple example of linear orderings in mind and extrapolate: to prove that it is impossible to disprove the continuum hypothesis, it suffices to build a model for the axioms of set theory satisfying the continuum hypothesis, and similarly for the axiom of choice.

16.2. **Zermelo-Frankel (ZF) set theory.** So far we have been working with Morse-Kelley set theory (MK). It will be convenient to introduce slightly different axioms for sets than the ones we have been working with so far: Zermelo-Frankel (ZF) set theory. In ZF set theory, there are only sets, nothing else (no classes). Just as in MK, there is also an undefined relation $\in$, and we use the same terminology as before. For example, for A, B sets, $A \subseteq B$ means that any set a which is a member A is also a member of B. The *ZF axioms* are the following:

- Extensionality: If A and B are sets such that $A \subseteq B$ and $B \subseteq A$, then $A = B$.
- Comprehension: If A is a set and $P(x)$ is a property of sets, then there is a set $A_0 \subseteq A$ whose members are exactly the sets $a \in A$ satisfying $P(a)$. We write $\{a \in A \mid P(a)\}$ for this set.
- Set existence: There is a set (note it follows using comprehension that there is a set $\emptyset$ with no members).
- Pairing: if a and b are sets, then there is a set c such that $a \in c$ and $b \in c$. (By comprehension, we can find a subset of c containing exactly a and b. We write $\{a, b\}$ for this set.)

- Union: if A is a set, there is a set B such that $b \in B$ whenever there is $a \in A$ so that $b \in a$ (Again by comprehension, this allows us to define $\bigcup A$ with the same meaning as before; also let $A \cup B$ mean $\bigcup\{A, B\}$).
- Power set: if A is a set, there is a set B such that for any set a, if $a \subseteq A$ then $a \in B$ (By comprehension, we can then define $\mathcal{P}(A)$ as before).
- Infinity: There is a set A such that $\emptyset \in A$ and if $x \in A$ then $x \cup \{x\} \in A$.
- Replacement: If $P(x, y)$ is a property depending on two sets x and y, A is a set, and for all $a \in A$ there exists a unique set b such that $P(a, b)$, then there is a set B so that for all $a \in A$ there is $b \in B$ with $P(a, b)$.
- Foundation: For any non-empty set A, there is $a \in A$ such that $a \cap A = \emptyset$ (where $\cap$ has the usual meaning – intersections exist by comprehension).

The main difference with MK set theory, which we have been sing so far, is that there are no classes. Thus the statement of replacement is in terms of property rather than in terms of an existing class function. The axiom of comprehension is a replacement for specification: it allows us to form the set of all sets satisfying a given property, but only if we can bound this set by a given set. The other axioms are very similar to what we had before. The axiom of choice is not part of ZF set theory. *ZFC set theory (Zermelo-Frankel with choice)* is defined to be ZF with the added axiom of local choice:

- Local choice: for any set X, there is a function $F : X \to \bigcup X$ such that $F(a) \in a$ for all non-empty $a \in X$.

Here, the notion of function is defined as before. It is clear that MK is more powerful than ZFC, since all the axioms of ZFC follow from MK. On the other hand, MK is in practice not really stronger: with minor changes, all the theorems we have seen are true also in ZFC. Objects like the class of all ordinals do not exist, but we can instead work with the property "x is an ordinal". Arguably, MK is usually more convenient to work with but when we start doing metamathematics and analyze universe of set theory, we will see that having a set theory without classes, like ZFC, is handy.

In this section, unless stated explicitly, we still work in MK but study models for ZF and ZFC.

16.3. **Models of ZF.** Intuitively, a model (or universe) of ZF is a class M that satisfies all the axioms of ZF, when we relativize to that set. Relativizing an axiom to M means that the word "set" is replaced by "set in M" everywhere in the axiom.

Consider for example the class $M = \{\emptyset, \{1\}\}$. Does M satisfy the axiom of extensionality? No, because in M, $\{1\}$ has no element (because it has no elements that are in M). Thus in M, $\{1\}$ is both a subset and a superset of the empty set. This situation is similar to the reason why $(\mathbb{N}, +)$ does not have additive inverses: while it is true that $-1 + 1 = 0$, -1 is not in $\mathbb{N}$.

To avoid this issue with extensionality, it suffices to require that M is set-initial (that is, $x \in y \in M$ implies $x \in M$). In this case, we call M an *inner* model.

Definition 16.1. A *model of ZF* (or of ZFC) is a class M that satisfies all the axioms of ZF (or ZFC), in the sense that they hold when relativized to M (the word "set" is replaced by "set in M" everywhere). An *inner model* of ZF (or of ZFC) is a model that is also set-initial.

Two of the axioms are immediately implied by set-initiality:

Lemma 16.2. If M is a set-initial class, then it satisfies the extensionality and foundation axioms.

Proof. Both are straightforward. $\square$

Obviously, SET is an inner model of ZFC. Are there others? Let us look at the V_α's. They are set-initial, which is a good start. First, $V_0 = \emptyset$ is not an inner model of ZFC, because it does not satisfy set existence (and infinity; but the other axioms hold vacuously). Second, $V_{\alpha+1}$ is also not an inner model of ZFC, for any ordinal α. Indeed, $V_\alpha \in V_{\alpha+1}$ (recall that $V_{\alpha+1} = \mathcal{P}(V_\alpha)$) but the power set axiom fails in $V_{\alpha+1}$. We have to be a bit careful here: relativized to $V_{\alpha+1}$, the power set axiom says that if $A \in V_{\alpha+1}$, then there is $B \in V_{\alpha+1}$ such that for all $a \in V_{\alpha+1}$, $a \subseteq A$ implies $a \in B$. That is, we only have to collect the subsets of A that happen to be in $V_{\alpha+1}$. In this specific case however, if $a \subseteq V_\alpha$ is any subset then $a \in V_{\alpha+1}$ by definition, so the B described by the power set axiom would need to be $\mathcal{P}(V_\alpha) = V_{\alpha+1}$, which is not in $V_{\alpha+1}$.

We are left with V_δ for δ limit ordinals. These do satisfy the power set axiom. They also satisfy set existence and pairing: if $a, b \in V_\delta$ then $a \in V_\alpha$ and $b \in V_\beta$ for $\alpha, \beta < \delta$. Without loss of generality $\alpha \leq \beta$, so also $a \in V_\beta$. Now note that $\{a, b\} \subseteq V_\beta$ so is in $V_{\beta+1} \subseteq V_\alpha$. What about infinity? If $\delta = \omega$, it is not going to be satisfied: any $A \in V_\omega$ is a finite set so cannot satisfy the axiom (one can check that this is the only axiom of ZFC that fails in V_ω). However if $\delta > \omega$, then $A = V_\omega$ witnesses that the axiom of infinity is satisfied ($A = \omega$ would also work of course). Assume from now on that $\delta > \omega$. What about the union axiom? Fix $A \in V_\delta$. Then $A \in V_\alpha$ for some $\alpha < \delta$. The members of A must also be in V_α (it is set-initial), hence are subsets of V_α. Therefore $\bigcup A \subseteq V_\alpha$, so $\bigcup A \in V_{\alpha+1} \subseteq V_\delta$. Note that, as opposed to the power set axiom, we did not need to relativize the union axiom: since it depends only on sets contained in A (and sets contained in those), the fact that V_δ is set-initial gives us that "union" means the same thing in SET $= V$ and in V_δ. We say that the operation of taking unions is *absolute* between V_δ and SET.

Very good, so so far we have shown that if $\delta > \omega$ is a limit ordinal, then V_δ satisfies extensionality, set existence, pairing, union, power set, infinity, and foundation. We are left with comprehension and replacement. For both, we need to discuss how to relativize properties, and so we inevitably have to come back to the question of what a property actually is. Let's give the following still slightly imprecise description: a *property* $P(x)$ is a finite expression that can involve the following statements:

- $x \in y$, for x, y variables, or $x \in a$, for a an arbitrary fixed set (called a *parameter* of the property), $a \in x$, $a \in b$, for a, b both parameters.
- Similarly, $x = y$, $x = a$, $a = x$, $a = b$.
- Composite statements, put together using "and", "or", "implies", "if and only if", "not". For example you can say "not $x \in y$" (which really one would write as $x \notin y$).
- Quantified statement of the form "For all sets y, $Q(x, y)$" or "There exists a set y, $Q(x, y)$", where Q is another property.

Note that some of these are redundant. For example, "For all sets y, $Q(x, y)$" can also be written "not there exists a set y, not $Q(x, y)$", or in English: it is not true that there exists a set y not satisfying $Q(x, y)$. Also note that since we are dealing with ZF and not MK, we do not allow our properties to quantify over classes.

You should convince yourself that all the properties that we used can really be defined in this style. Now, *relativizing* a property to a class M means simply replacing "for all sets y" by "for all sets y in M", and similarly for there exists. We also require the parameters of the property to all come from M. Thus the comprehension axiom relativized to a class M says that for any property $P(x)$ with parameters from M and any set $A \in M$, the set $A_0' \subseteq A$ is in A, where the members of A_0' are exactly the sets $a \in A$ satisfying the property P relativized to M (i.e. with all quantifiers bounded by M). We write $P^M(x)$ for the property P relativized to M.

In passing, note that the comprehension and replacement axioms of ZF are not really single axioms but axiom *schemes*: they consist of infinitely-many axioms, one for each fixed property.

Back to V_δ for $\delta > \omega$ limit: let $A \in V_\delta$ and let $P(x)$ be a property, with parameters $a_1, \dots, a_n$ from V_δ. P has only finitely-many parameters, so there must be $\alpha < \delta$ so that $A \in V_\alpha$ and all the a_i's are in V_α. Let $A_0 := \{a \in A \mid P^{V_\delta}(a)\}$. Note that $A_0 \subseteq A \subseteq V_\alpha$, so $A_0 \in V_{\alpha+1}$. Thus the comprehension axiom holds in V_δ.

We will discuss replacement in a minute, but what about the axiom of local choice? Again, if $X \in V_\delta$, then $X \in V_\alpha$ for $\alpha < \delta$. Let $F : X \to \bigcup X$ be a choice function (in SET). Note that $F \in \mathcal{P}^{50}(X \cup \bigcup X \cup \{0, 1\})$, where $\mathcal{P}^{50}$ denotes the power set iterated 50 times (an overkill of course). Thus $F \in V_{\alpha+50} \subseteq V_\delta$, and hence local choice holds in V_δ. We have shown:

Lemma 16.3. If $\delta > \omega$ is a limit ordinal, then V_δ satisfies all the axioms of ZFC except possibly for replacement.

Replacement is trickier. Its truth depends on whether one can "reach" δ from below. For example, $V_{\omega+\omega}$ contains $\mathcal{P}(\omega)$, which has cardinality at least ω_1. However ω_1 is not in $V_{\omega+\omega}$! Thus if we fix a well-order $<$ of $\mathcal{P}(\omega)$, it will be in $V_{\omega+\omega}$, but its Mostowski collapse will not. Since the Mostowski collapse can be defined by an explicit (very long) formula, and what is described by this formula does not change when relativizing it, we get a failure of the axiom of replacement. In general, if $\delta > \omega$ is limit and there exists $A \in V_\delta$ such that $|A| \geq \delta$, then V_δ does not satisfy replacement.

What is the cardinality of members of V_δ? In fact, what is the cardinality of V_δ itself? The answer is most easily described recursively:

Definition 16.4. For a fixed cardinal λ, define by induction on the ordinal α a cardinal $\beth_\alpha(\lambda)$ as follows:

- $\beth_0(\lambda) = \lambda$.
- $\beth_{\beta+1}(\lambda) = 2^{\beth_\beta(\lambda)}$.
- $\beth_\delta(\lambda) = \sup_{\gamma<\delta} \beth_\gamma(\lambda)$ if δ is a limit ordinal.

($\beth$ is read "beth". It is the second letter of the Hebrew alphabet). We write $\beth_\alpha$ instead of $\beth_\alpha(\aleph_0)$.

Lemma 16.5.

1. For any ordinal α, $|V_\alpha| = \beth_\alpha(0)$.
2. If $\alpha \geq \omega \cdot \omega$, $\beth_\alpha(0) = \beth_\alpha$.
3. If $\delta > \omega$ is a limit ordinal and replacement holds in V_δ, then $\beth_\delta = \delta$. In particular, δ is a cardinal.

Proof.

(1) Straightforward induction.
(2) $\beth_\beta \leq \beth_{\omega+\beta}(0)$, and if $\alpha \geq \omega \cdot \omega$, $\omega + \alpha = \alpha$, so the result follows.
(3) We need to ensure that $|A| < \delta$ for any $A \in V_\delta$. In particular, $|V_\alpha| = \beth_\alpha < \delta$ for all $\alpha < \delta$, so the result follows.

□

Thus we have seen that if V_δ is a model of ZFC, then $\delta = \beth_\delta$. The converse is not true. For example, if λ is the first cardinal so that $\lambda = \beth_\lambda$ and V_λ were a model of ZFC, then V_λ would in particular satisfy the property that there exists a cardinal μ with $\mu = \beth_\mu$. All the properties involved are absolute between V_λ and SET, hence $\mu = \lambda$, so $\mu \notin V_\lambda$, a contradiction. Generally, in order for V_λ to be a model of ZFC, it must be impossible to define λ from within V_λ. Let us now study a case where this happens:

Definition 16.6. A cardinal λ is *strong limit* if it is infinite and $2^\mu < \lambda$ for all $\mu < \lambda$. We say that λ is *strongly inaccessible* (or just *inaccessible*) if it is uncountable, regular, and strong limit.

Exercise 16.7. Let λ be a regular uncountable cardinal. Prove that λ is strongly inaccessible if and only if $\lambda = \beth_\lambda$. Show similarly that λ is weakly inaccessible if and only if $\lambda = \aleph_\lambda$.

Note that strong limit implies limit, hence strongly inaccessible implies weakly inaccessible. We will see that the existence of strongly inaccessible cardinals cannot be proven from the axioms of MK. Cardinals whose existence cannot be proven from the axioms are often called *large cardinals*.

Theorem 16.8. If λ is an inaccessible cardinal, then V_λ is an inner model of ZFC.

Proof. By Lemma 16.3, it suffices to check that V_λ satisfies the replacement axiom. Fix a property $P(x, y)$ and a set $A \in V_\lambda$ such that for all $a \in A$ there is a unique $b \in V_\lambda$ such that $P(a, b)^{V_\lambda}$. Using replacement (in SET), we can find a function $f : A \to V_\lambda$ such that for all $a \in A$, $f(a)$ is the unique $b \in V_\lambda$ so that $P(a, b)^{V_\lambda}$. We have to see that the range of f is a member of V_λ. Since $A \in V_\lambda$, we know that $A \in V_\alpha$ for some $\alpha < \lambda$. In particular, $A \subseteq V_\alpha$, and $|V_\alpha| = \beth_\alpha(0) \leq \beth_\alpha < \lambda$ (Exercise 16.7). Thus $|A| < \lambda$. Let $\mu := |A|$ and write $A = \{a_i \mid i < \mu\}$. For each $i < \mu$, $f(a_i) \in V_\lambda$ so there exists $\alpha_i < \lambda$ such that $f(a_i) \in V_{\alpha_i}$. Let $\alpha := \sup_{i<\mu} \alpha_i$. Since λ is regular, $\alpha < \lambda$, and by definition the range of f is a subset of V_α, hence a member of $V_{\alpha+1} \subseteq V_\lambda$. □

Remark 16.9. By a similar proof, we also get that $V_{\lambda+1}$ is an inner model of MK (in the expected sense: we interpret "class" in the axioms by "member of $V_{\lambda+1}$"; note that then sets will be members of V_λ). The same result remains valid if we work within ZFC set theory. This gives a precise sense in which MK is not too much stronger than ZFC: from ZFC and an inaccessible cardinal, we can build a model for MK.

Let us now show that it is impossible to prove the existence of inaccessible cardinals. Intuitively (and as said already), if there is an inner model M of ZFC with no inaccessible cardinals, then any hypothetical proof of existence of inaccessible cardinals from the axioms of ZFC would in particular hold inside M and give that

M has inaccessible cardinals. Thus to show it is impossible to prove existence of inaccessible cardinals in ZFC, it suffices to build a model of ZFC with no inaccessible cardinals.

Corollary 16.10. There is an inner model M of ZFC that does not have inaccessible cardinals (more precisely, M satisfies the property "there are no inaccessible cardinals"). Thus it is impossible to prove the existence of inaccessible cardinals from the axioms of ZFC.

Proof. If there are no inaccessible cardinals, then let $M :=$ SET. Otherwise, let λ be the minimal inaccessible cardinal and let $M := V_\lambda$. By Theorem 16.8, M is an inner model of ZFC. If M satisfied the property that there were an inaccessible cardinal, then let $\mu \in M$ witness that property. The property is absolute, and hence μ would have to be inaccessible in SET as well, and since $\mu < \lambda$ this contradicts the minimality of λ. □

Remark 16.11. A similar argument, using Remark 16.9 shows that it is also impossible to prove the existence of inaccessible cardinals from the axioms of MK. In fact, while we will focus from now on on building models of ZF and ZFC, many of our arguments are generalizable to MK with more work.

Are there other models of ZFC? For example, is there a singular cardinal λ such that V_λ is a model of ZFC? The answer is yes.[11]

Definition 16.12. If M and N are classes, M is an *elementary submodel of* N, written $M \preceq N$, if $M \subseteq N$ and for any property $P(x_0, \ldots, x_{n-1})$ of sets and any $a_0, \ldots, a_{n-1}$ in M, $P^M(a_0, \ldots, a_{n-1})$ holds if and only if $P^N(a_0, \ldots, a_{n-1})$ holds.

Intuitively, an elementary submodel M of N satisfy the same properties as N, even when there are parameters from M in the description of the properties. Note that $\preceq$ is transitive, irreflexive and antisymmetric. We aim to build an elementary submodel of SET (it will in particular be a model of ZFC). The next two results realize this goal.

Lemma 16.13 (Tarski-Vaught test)**.** Let M, N be classes with $M \subseteq N$. The following are equivalent:

(1) $M \preceq N$.
(2) For any property $P(x, x_0, \ldots, x_{n-1})$ and any $a_0, \ldots, a_{n-1}$ in M, if there exists $a \in N$ so that $P(a, a_0, \ldots, a_{n-1})^N$, then there exists $a' \in M$ so that $P(a', a_0, \ldots, a_{n-1})^N$.

Proof. That the second condition follows from $M \preceq N$ is immediate from the definition. Assume now that the second condition holds. Let $P(x_0, \ldots, x_{n-1})$ be a property and let $a_0, \ldots, a_{n-1}$ be in M. We proceed by induction on the structure of P: we can think of it as being built up from *atomic* statements ($x \in y$ or $x = y$) using logical connectives ("and", "or", "implies", "if and only if", "not"), and quantifiers (for all or there exists). We prove by induction on P that $P^M(a_0, \ldots, a_{n-1})$ if and only if $P^N(a_0, \ldots, a_{n-1})$ for any choice of $a_0, \ldots, a_{n-1} \in M$. If P is atomic ($x_0 \in x_1$ or $x_0 = x_1$), then this is immediate. If P is built up from less complex

[11] But remember that we are working within MK, a stronger system of axioms. Within ZFC, the answer would be "not necessarily" because of a theorem in logic, Gödel's incompleteness theorem, saying that ZFC cannot prove its own consistency, i.e. cannot always exhibit a set model for itself.

formulas using logical connectives, then the result is also straightforward from the induction hypothesis. Suppose now that P is of the form "there exists a set y such that $Q(y, x_0, \dots, x_{n-1})$", where Q is a less complex property. If $P^M(a_0, \dots, a_{n-1})$, fix $a \in M$ such that $Q^M(a, a_0, \dots, a_{n-1})$. By the induction hypothesis, we also have $Q^N(a, a_0, \dots, a_{n-1})$, hence $P^N(a_0, \dots, a_{n-1})$. For the converse, assume that $P^N(a_0, \dots, a_{n-1})$. By the second condition in the statement of the lemma (that we are assuming), there is $a' \in M$ so that $Q^N(a, a_0, \dots, a_{n-1})$. By the induction hypothesis again, $Q^M(a, a_0, \dots, a_{n-1})$, and hence $P^M(a_0, \dots, a_{n-1})$. The case where P is given by a for all quantifier can be reduced to that of an existential quantifier (forall y, Q (y) is the same as not there exists y not Q (y) ...). □

Theorem 16.14 (Löwenheim-Skolem-Tarski theorem)**.** If N is a class and $X \subseteq N$ is a set, then there exists $M \preceq N$ such that $X \subseteq M$ and $|M| \leq |X| + \aleph_0$.

Proof. We build M satisfying the second condition of Lemma 16.13. Specifically, we say a set $M \subseteq N$ is *existentially closed over* $A \subseteq M$ if for any property $P(x, x_0, \dots, x_{n-1})$ and any $a_0, \dots, a_{n-1}$ in A, if there exists $a \in N$ so that $P(a, a_0, \dots, a_{n-1})^N$, then there exists $a' \in M$ so that $P(a', a_0, \dots, a_{n-1})^N$.

Properties are simply finite strings from a finite possible choice of letters. Thus there are countably-many such properties, which we can enumerate as $(P_n)_{n<\omega}$ (formalizing this fully logically is a little bit tricky, and we skip this part: we are using the full power of MK, including the ability to quantify over classes). We build $(M_n)_{n<\omega}$ an increasing sequence of subsets of N so that $X \subseteq M_0$, $|M_n| \leq |X| + \aleph_0$, and M_{n+1} is existentially closed over M_n for all $n < \omega$. Set $M_0 := X$, and given M_n, simply enumerate all tuples $(k, a_0, \dots, a_{r-1})$, where $k, r < \omega$ and $a_0, \dots, a_{r-1} \in M_n$. There are at most $|M_n|^{<\aleph_0} + \aleph_0 = |M_n| + \aleph_0$ many such tuples, and for each one there may correspond a property $P_k(x, a_0, \dots, a_{r-1})$ which has a solution in N. In this case, pick one such a solution and put it in M_{n+1}. In the end, set $M := \bigcup_{n<\omega} M_n$. □

Corollary 16.15.

(1) For any infinite cardinal λ, there is $M \preceq \text{SET}$ of cardinality λ. In particular, there is a model of ZFC of cardinality λ. In fact, we can even find an *inner* model of ZFC of cardinality λ.
(2) There exists a cardinal λ such that $V_\lambda \preceq \text{SET}$ (so in particular is a model of ZFC). In fact, the class of cardinals λ such that $V_\lambda \preceq \text{SET}$ is closed and unbounded (in the expected sense).

Proof.

(1) Apply the Löwenheim-Skolem-Tarski theorem with $X = \lambda$ to find $M \preceq \text{SET}$ of cardinality λ. To get an *inner* model, we apply the Mostowski collapsing lemma to $(M, \in)$: it is extensional because it satisfies the extensionality axiom, it is set-like because it is a set, and it is wellfounded because by the axiom of foundation even $(\text{SET}, \in)$ is wellfounded. We get that $(M, \in)$ is isomorphic to some $(M', \in)$, where M' is set-initial. Because it is isomorphic to $(M, \in)$, M' is still a model of ZFC.
(2) Let C be the class of cardinals λ such that $V_\lambda \preceq \text{SET}$. First prove that if δ is a limit ordinal and $(M_\alpha)_{\alpha<\delta}$ are such that $M_\alpha \preceq \text{SET}$ for all $\alpha < \delta$, then $\bigcup_{\alpha<\delta} M_\alpha \preceq \text{SET}$ (this is straightforward using the Tarski-Vaught test, for example). This shows that C is closed. To see it is unbounded fix

an ordinal β and define inductively $(\lambda_n)_{n<\omega}$, $(M_n)_{n<\omega}$ as follows: fix any cardinal $\lambda_0 > \beta$, and given λ_n for $n < \omega$, find a set $M_n \preceq \text{SET}$ such that $V_{\lambda_n} \subseteq M_n \preceq \text{SET}$, and let λ_{n+1} be big-enough so that $M_n \subseteq V_{\lambda_{n+1}}$. In the end, let $\lambda := \sup_{n<\omega} \lambda_n$, and observe that $V_\lambda = \bigcup_{n<\omega} M_n \preceq \text{SET}$, so $\lambda \in C$.

□

16.4. **Skolem's paradox.** A fun consequence of Corollary 16.15 is that there exists a countable inner model M of ZFC. Now, since M is a model of ZFC, it must satisfy the statement "there exists uncountable sets". However M is countable, and since it is set-initial any of its member is also a subset of M, so is countable. Isn't this a contradiction?

The resolution is that the property $P(x)$ given by "x is uncountable" is not absolute between SET and M. Precisely, we know there is $a \in M$ such that $P^M(a)$, and we also know that it is not true that $P(a)$, but it is not a contradiction because we do not know that $M \preceq \text{SET}$. In fact, we have just shown that there is no countable set M that is a model of ZFC and such that both M is set-initial and $M \preceq \text{SET}$.

On the other hand, we *can* take $M \preceq \text{SET}$ countable. Then if $a \in M$ is such that $P^M(a)$, we know that $a \cap M$ is countable (because M is countable), so there is a surjection $f : \omega \to a \cap M$, but this does not mean that $f \in M$, so there is no contradiction!

16.5. **The Gödel operations.** Corollary 16.15 is not directly useful to study statements such as the continuum hypothesis or the axiom of choice, since the models it produces all satisfy the same properties. Toward building more interesting models of ZFC, we now work toward characterizing when a proper class is an inner model of ZFC. An inner model of ZFC must be closed under the following functions, called the *Gödel operations*. Intuitively they are the explicit, absolute operations that can be performed on sets:

Definition 16.16. The *Gödel operations* are the following class functions from $\text{SET} \times \text{SET}$ to SET:

(1) $G_0(A, B) := A \times B$ $(= \{(a, b) \mid a \in A, b \in B\})$.
(2) $G_1(A, B) := \{(a, b) \in A \times B \mid a \in b\}$.
(3) $G_2(A, B) := \{a \in A \mid \text{ there exists } a' \in A \text{ such that } (a, a') \in B\}$.
(4) $G_3(A, B) := A - B$.
(5) $G_4(A, B) := A \cap B$.
(6) $G_5(A, B) := \bigcup A$.
(7) $G_6(A, B) := \{A, B\}$.
(8) $G_\pi(A, B) := \{(a_{\pi(0)}, a_{\pi(1)}, \dots, a_{\pi(n-1)}) \mid a_0, \dots, a_{n-1} \in A, (a_0, \dots, a_{n-1}) \in B\}$, for each $n < \omega$ and each permutation (i.e. bijection) $\pi : n \to n$.

We will study what it means to be closed under the Gödel operations:

Definition 16.17. Let $F := (F_i)_{i\in I}$ be a class sequence such that for each $i \in I$, there exists a natural number n_i such that F_i is a function from SET^{n_i} to SET (where SET^{n_i} is the cartesian product of SET with itself n_i times and we make the usual identifications). Let A be a class. We say that A is *closed under* F (or *closed under the functions from* F) if for any $i \in I$, if $a_0, \dots, a_{n_i-1}$ are all in A, then $F_i(a_0, \dots, a_{n_i-1}) \in A$. We say that A is *closed under the Gödel operations* if it is closed under $(G_i)_{i\in I}$, where $I = \{0,1,2,3,4,5,6\} \cup \{\pi \mid \pi \text{ a permutation of } n,\, n < \omega\}$.

Lemma 16.18. Assume $F := (F_i)_{i\in I}$ is a class sequence such that for each $i \in I$, F_i is a class function from SET^{n_i} to SET, for some $n_i < \omega$. Assume M is a class which is closed under F. If I is a set and $X \subseteq M$ is a set, then there exists $A \subseteq M$ such that $X \subseteq A$, A is closed under F, and $|A| \leq |I| + |X| + \aleph_0$. In particular, if M is a class which is closed under the Gödel operations and $X \subseteq M$ is a set, then there exists $A \subseteq M$ which contains M, is closed under the Gödel operations, and satisfies $|A| \leq |X| + \aleph_0$.

Proof. Exercise. □

What the lemma is saying, intuitively, is that it is always possible to close a set under a given number of operations. For a familiar example, take the set of real numbers $\mathbb{R}$ and let $A \subseteq \mathbb{R}$ be a countable set. Then you should convince yourself that a special instance of the lemma says there is a countable set $B \subseteq \mathbb{R}$ containing A that is closed under addition and multiplication (i.e. if $a, b \in B$ then $a + b, a \cdot b \in B$).

Similarly here, we see that if we start with a class closed under the Gödel operations (like SET), we can find a countable subset that is closed under the Gödel operations. Note that any elementary submodel of SET will be closed under the Gödel operations (why?), but not necessarily vice-versa. Indeed:

Exercise 16.19. For an ordinal α, V_α is closed under the Gödel operations if and only if α is zero or a limit ordinal.

Intuitively, closing a set under the Gödel operations give us everything that can be explicitly defined with parameters from that set, and nothing else. We now characterize when a class is an inner model of ZF using the Gödel operations. This also will give us a tool to quickly check the axioms of ZF.

Theorem 16.20. Let M be a set-initial proper class. The following are equivalent:

(1) M is a model of ZF.
(2) M is closed under the Gödel operations, and for any set $X \subseteq M$, there exists a set $Y \in M$ so that $X \subseteq Y$.

Proof. First, if M is a model of ZF then it is clear that it must be closed under the Gödel operations (they are all absolute between M and SET). Now fix $X \subseteq M$. There exists an ordinal α such that $X \in V_\alpha$. Let $Y := V_\alpha \cap M$. Then $X \subseteq Y$ and $Y \in M$, because the construction of the V_α's can be done in ZF, and when we relativize the definition of V_α to M, we obtain $V_\alpha \cap M$. In other words, Y is the object that M "thinks" is V_α, so $Y \in M$.

Conversely, assume that M is closed under the Gödel operations and for any set $X \subseteq M$ there is a set $Y \in M$ so that $X \subseteq Y$. We show that M is a model of ZF by verifying all the axioms.

- M satisfies extensionality and foundation, because any set-initial class satisfies them (Lemma 16.2). Set existence is also immediate: M is a proper class. Pairing and union are because M is closed under the Gödel operations.
- For the power set axiom, fix $A \in M$. We have to see that $\mathcal{P}(A) \cap M$ is contained in a member of M (remember that we are relativizing the power

set axiom to M). Note that $\mathcal{P}(A) \cap M \subseteq M$, so the $Y \in M$ so that $\mathcal{P}(A) \cap M \subseteq Y$ witnesses the power set axiom.

- For the axiom of infinity, note that $\emptyset \in M$: if we pick $a \in M$, $\emptyset = a - a = G_3(a, a)$. Similarly, if $x \in M$ then $x \cup \{x\} = \bigcup\{x, \{x\}\} \in M$ by closure under the Gödel operations. Thus $\omega \subseteq M$, and hence we can find $Y \in M$ so that $\omega \subseteq Y$, witnessing the axiom of infinity.
- For replacement, let $A \in M$ be a set and let $P(x, y)$ be a property such that for any $a \in A$ there is a unique $b = b_a \in M$ so that $P^M(a, b)$ holds. By replacement (in SET), we can find a set $B \subseteq M$ so that $b_a \in B$ for all $a \in A$. Now the $Y \in M$ so that $B \subseteq Y$ witnesses replacement.
- It remains to prove the comprehension axiom. Fix a set $A \in M$. We prove more generally that for any property $P(x_1, \ldots, x_n)$, possibly involving parameters from M, the set $A^P := \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in A, P^M(a_1, \ldots, a_n)\}$ is in M. We can assume that P does not contain the equality sign: if $x = y$ appears in P, replace it by "for all z, $z \in x$ if and only if $z \in y$". Also, we can assume P does not contain "for all x ..", since we can replace it by "not there exists x not ...". Similarly, without loss of generality P does not use "or", "implies" or "if and only if": "if and only if" can be replaced by the conjunction of two "implies", Q_1 implies Q_2 is equivalent to "not Q_1 or Q_2", and Q_1 or Q_2 is equivalent to "not (not Q_1 and not Q_2)".

  As in the proof of the Tarski-Vaught test (Lemma 16.13), we now proceed by induction on the complexity of P. If P is $x_1 \in x_2$, then $A^P = G_1(A, A)$. If P is $u \in x_1$, for u a parameter, then first let $A_0 := G_1(\{u\}, A) = G_1(G_6(u, u), A) = \{(u, a) \mid u \in a \in A\}$. Next, let $A_1 := G_\pi(A_0)$, where π permutes the first and second component. Thus $A_1 = \{(a, u) \mid u \in a \in A\}$. Finally, let $A_2 := G_2(A_1)$, and let $A_3 := G_0(A_2, A)$. We have that $A^P = A_3$, so we are done by closure under the Gödel operations. The other atomic cases (for example if P is $u \in x_2$, or $x_2 \in u$) are similar.

  Assume now that P is "Q_1 and Q_2", and A^{Q_1}, A^{Q_2} are both in M. Then $A^P = A^{Q_1} \cap A^{Q_2} = G_4(A^{Q_1}, A^{Q_2})$, which is in M by closure under the Gödel operations. Similarly, if P is "not Q", and Q has n variables, then $A^P = A^n - A^Q = G_3(A^n - A^Q)$, where A^n is obtained by taking cartesian products n times (G_0).

  Finally, for quantifiers we apply G_0, G_2, and G_π for an appropriate permutation π (we are taking advantage of the fact that we defined $(a_1, a_2, \ldots a_n)$ to be $(a_1, (a_2, \ldots, a_n))$).

□

The last part of the proof of Theorem 16.20 proved something more general, which we record for future reference:

Lemma 16.21. If N is a set-initial class closed under the Gödel operations, $M \in N$, and $P(x_1, \ldots, x_n)$ is a property of sets, possibly with parameters from M then $\{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in M, P^N(a_1, \ldots, a_n)\}$ is in N.

In the next section, we investigate an example of a proper class model of ZF which is not just SET.

17. The constructible universe

Using the Gödel operations, we now define the smallest possible proper class inner model of ZF. The idea is simple: whereas when building the V_α's we took the full power set at each step, this construction is completely understand to understand questions like the continuum hypothesis: in one step, we throw in absolutely all the subsets of natural numbers. Instead, we define a hierarchy that grows much more slowly and only includes what is absolutely needed instead of the full power set.

Definition 17.1. For a set X, let $\operatorname{cl}(X)$ denote the closure of X under the Gödel operations (that is, the intersection of all sets $Y \supseteq X$ closed under the Gödel operations – see Lemma 16.18). Let $\operatorname{Def}(X) := \operatorname{cl}(X \cup \{X\}) \cap \mathcal{P}(X)$. We call $\operatorname{Def}(X)$ the *definable power set of X.*

Remark 17.2. It can be shown (try it!) that $\operatorname{Def}(X)$ is exactly the set of all $Y \subseteq X$ such that there exists a property $P(x)$ with parameters from X with $Y = \{a \in X \mid P^X(a)\}$. In that sense, $\operatorname{Def}(X)$ is the set of all subsets of X that are definable from inside X.

The following are the basic properties of $\operatorname{Def}(X)$.

Lemma 17.3. Let X be a set-initial set.

1. $X \cup \{X\} \subseteq \operatorname{Def}(X) \subseteq \mathcal{P}(X)$.
2. $\operatorname{Def}(X)$ contains all finite subsets of X.
3. $\operatorname{Def}(X)$ is set-initial.
4. If X is infinite, $|X| = |\operatorname{Def}(X)|$.
5. $\operatorname{Def}(X) = \mathcal{P}(X)$ if and only if X is finite.
6. If $\alpha \subseteq X$ is an ordinal, then $\alpha \in \operatorname{Def}(X)$.

Proof.

1. Since X is set-initial, $X \cup \{X\} \subseteq \mathcal{P}(X)$, and $X \cup \{X\} \subseteq \operatorname{cl}(X \cup \{X\})$. The fact that $\operatorname{Def}(X) \subseteq \mathcal{P}(X)$ is immediate from the definition.
2. Let $X_0 \subseteq X$ be a finite set. We proceed by induction on $|X_0|$. If $X_0 = \emptyset$, then $X_0 = X - X = G_3(X, X)$ so by closure under the Gödel operations, $X_0 \in \operatorname{Def}(X)$. If $X_0 = X_0' \cup \{a\}$, with $a \in X$ and $X_0' \in \operatorname{Def}(X)$, then $X_0 = G_5(G_6(X_0', G_6(a, a)))$, so is in $\operatorname{Def}(X)$.
3. Let $a \in B \in \operatorname{Def}(X)$. Then $B \subseteq X$, so $a \in X$, so $\{a\} \in \operatorname{cl}(X)$, so $a = \bigcup\{a\} \in \operatorname{cl}(X)$. Since X is set-initial, $a \subseteq X$, so $a \in \operatorname{cl}(X) \cap \mathcal{P}(X) \subseteq \operatorname{Def}(X)$, as desired.
4. By Lemma 16.18, $|\operatorname{cl}(X)| \leq |X| + \aleph_0$, so because $\operatorname{Def}(X) \subseteq \operatorname{cl}(X)$, the result follows.
5. If X is infinite, then by the previous part $|\operatorname{Def}(X)| = |X|$ but by Cantor's theorem $|\mathcal{P}(X)| > |X|$, so $\operatorname{Def}(X) \neq \mathcal{P}(X)$. If X is finite, then because any finite set is in $\operatorname{Def}(X)$, we get that $\operatorname{Def}(X) = \mathcal{P}(X)$.
6. Since X is set-initial, if there exists $\beta \geq \alpha$ with $\beta \in X$, then $\alpha \in X$, so $\alpha \in \operatorname{Def}(X)$. Assume now that $\mathrm{OR} \cap X = \alpha$. Consider the property $P(x)$ given by "x is an ordinal". By Lemma 16.21, $\{a \in X \mid P^{\operatorname{cl}(X \cup \{X\})}(a)\}$ is a member of $\operatorname{cl}(X \cup \{X\})$. Since the property P is absolute between $\operatorname{cl}(X \cup \{X\})$ and SET, this set is just $X \cap \mathrm{OR} = \alpha$.

□

Using the definable power set, we give the definition of the constructible hierarchy (due to Kurt Gödel).

Definition 17.4 (The constructible hierarchy)**.** Define inductively $(L_\alpha)_{\alpha \in \mathrm{OR}}$ as follows[12]

(1) $L_0 = \emptyset$.
(2) $L_{\alpha+1} = \mathrm{Def}(L_\alpha)$.
(3) $L_\delta = \bigcup_{\alpha < \delta} L_\alpha$, for δ a limit ordinal.

The *constructible universe* is the class $L := \bigcup_{\alpha \in \mathrm{OR}} L_\alpha$. We may also write L_{OR} for L. A set is *constructible* if it is a member L.

Lemma 17.5 (Basic properties of the constructible hierarchy)**.** For all ordinals α:

(1) L_α is set-initial.
(2) If $\alpha < \beta$, then $L_\alpha \subsetneq L_\beta$.
(3) $L_\alpha \subseteq V_\alpha$, with equality if and only if $\alpha \leq \omega$.
(4) If α is infinite, $|L_\alpha| = |\alpha|$.
(5) L is closed under the Gödel operations, and more generally L_α is closed under the Gödel operations if and only if α is zero or a limit ordinal.
(6) $\alpha \subseteq L_\alpha$, and in fact $L_\alpha \cap \mathrm{OR} = \alpha$.

Proof. All are easy exercises using Lemma 17.3. □

Given the work already done, it is quite straightforward to see that the constructible universe is a model of ZF. Very interestingly, it is also a model of ZFC, and the proof does not use the axiom of choice! This shows that the axiom of choice cannot be disproven from ZF.

Theorem 17.6. *L is a proper class inner model of ZFC.*

Proof. We apply Theorem 16.20 to see that L is a model of ZF. L is set-initial because each L_α is set-initial. L is a proper class because $L_\alpha \subsetneq L_{\alpha+1}$ for each ordinal α. L is closed under the Gödel operations by Lemma 17.5. Finally, if $X \subseteq L$ is a set, then by the axiom of replacement (in SET) there must exist α such that $X \subseteq L_\alpha$. Note that $L_\alpha \in L_{\alpha+1} \subseteq L$, so setting $Y := L_\alpha$, we have that $Y \in L$ and $X \subseteq Y$.

For choice, we prove even more: we give a class well-ordering $\trianglelefteq$ of L such that its restriction $\trianglelefteq \restriction L_\alpha$ is a member of L for all α. Moreover, the definition is uniform: there is a property $P(x, y)$ without parameters such that for $a, b \in L$, $a \trianglelefteq b$ if and only if $P^L(a, b)$. Thus a variant of *global* choice holds in L.

We only sketch how to construct $\trianglelefteq$ but the idea is simply to order elements by the order in which they were constructed: if X is a set and $a \in \mathrm{cl}(X)$, then there exists $x_1, \ldots, x_n$ so that a can be obtained by the composition of a finite number of Gödel operations, starting with $x_1, \ldots, x_n$. There are countably many operations, and they can (without using the axiom of choice) each be associated a natural number n. A composition of finitely-many of them can therefore be described by a finite sequence of natural numbers. There are many explicit well-orderings of finite sequences of natural numbers (one way is to use powers of primes, another is to order them lexicographically). Thus for fixed parameters $x_1, \ldots, x_n$, we can well-order the elements of $\mathrm{cl}(\{x_1, \ldots, x_n\})$. In particular, if we are inductively given

[12] What does the letter L stand for? One theory says it stands for “laws”. I prefer the theory it stands for “Langsam”, meaning slow in German, because the hierarchy grows very slowly.

a well-ordering $\trianglelefteq_\alpha$ of L_α, then we can well-order $L_{\alpha+1}$ as follows: order all the elements of L_α as with $\trianglelefteq_\alpha$, and put the elements of $L_{\alpha+1} - L_\alpha$ above. The first element will be L_α. If $a \in L_{\alpha+1} - (L_\alpha \cup \{L_\alpha\})$, let $n = n_a < \omega$ be least so that $a \in \mathrm{cl}(x_1, \dots, x_n)$ for some $x_1, \dots, x_n \in L_\alpha \cup \{L_\alpha\}$. Then pick $(x_1^a, \dots, x_n^a)$ to be the lexicographically least such tuple (among tuples of the same size, where the ordering of the components is $\trianglelefteq$).

We now know that a is obtained from $x_1^a, \dots, x_n^a$ via a composition of finitely-many Gödel operations. Let m be a natural number encoding in a canonical way exactly which operations are needed and in which order to compose them. This must be defined in such a way that from m and $x_1^a, \dots, x_n^a$ we can uniquely recover a. There may be several such codes m (corresponding to several different ways to compose Gödel operations to obtain a), so we let m^a be the least one.

We then define $a \trianglelefteq_{\alpha+1} b$, for $a, b \in L_{\alpha+1} - (L_\alpha \cup \{L_\alpha\})$, to hold if either $n_a < n_b$ or if $n := n_a = n_b$ and $(x_1^a, \dots, x_n^a, m^a)$ is lexicographically least than $(x_1^b, \dots, x_n^b, m^b)$ (where $\trianglelefteq_\alpha$ is used to well-order the first n components, and the usual ordering on ω is used for the last one).

Finally, if δ is a limit ordinal, we just order L_δ by $a \trianglelefteq_\delta b$ if there exists $\alpha < \delta$ so that $a, b \in L_\alpha$ and $a \trianglelefteq_\alpha b$. □

17.1. **The axiom of constructibility and its consequences.** Is every set constructible? This is the content of:

Definition 17.7 (Axiom of constructibility)**.** The *axiom of constructibility*, abbreviated $V = L$, is the statement that every set is constructible.

We have to be a little bit careful, since the real question in MK would be "is every class constructible?" (whatever that exactly means). Thus the axiom of constructibility really makes sense only as an additional axiom for ZFC (not for MK).

Since the Gödel functions are absolute, the construction of L is completely absolute, hence (writing L_α^M for the unique $x \in M$ satisfying the property $x = L_\alpha$ relativized to M):

Lemma 17.8. Assume that M is an inner model of ZF.

1. If $\alpha \in \mathrm{OR} \cap M$, then $(L_\alpha)^M = L_\alpha$.
2. $L^M = L_{\mathrm{OR} \cap M}$ (using the convention that L_{OR} means L). In particular, L is the $\subseteq$-minimal proper class inner model of ZF.
3. M is a model of the axiom of constructibility if and only if $M = L_{\mathrm{OR} \cap M}$.

Proof. The first two are immediate from the absoluteness of L. The third also immediately follows: suppose for example that M is a proper class. $V^L = L^L = L$, so L is a model of the axiom of constructibility. Conversely, if M is a model of V = L, then absoluteness of L implies that $M = L$. If M is a set the argument is completely similar. □

We have shown the axiom of constructibility cannot be disproven from the axioms of ZF (because L satisfies it). On the other hand, one can also show that it cannot be proven from the axioms, but this is beyond the scope of this class.

The axiom of constructibility specifies the shape of the universe of sets completely, so it has a lot of consequences. One consequence we have already seen is the axiom of choice. It also implies the continuum hypothesis. This is perhaps not too surprising: L is as small as possible, so in it there must be as few subsets of reals as possible. One must be careful with such arguments, however, because we are also adding as few bijections as possible, so wouldn't it be possible that we are missing also a lot of witnessing to the countability of certain sets, and hence get a lot of uncountable sets?

A precise answer is obtained by understanding the construction of subsets of ω. In the V_α hierarchy, at stage $\omega + 1$ all subsets of ω have been constructed, so the V_α hierarchy is completely useless to understand the power set operation. The L_α hierarchy, however, grows much more slowly. At stage $\omega + 1$, not all subsets of ω have been constructed. This can be seen with a simple counting argument: $L_{\omega+1}$ is countable, but there are uncountably-many subsets of ω. Similarly, at any countable ordinal α, L_α does not contain all subsets of ω. Thus the earliest the construction of all subsets of ω can finish is ω_1. That it does finish at this point is established by a clever Löwenheim-Skolem-Tarski argument.

Note that for the theorem below, our base theory is ZFC (not MK).

Theorem 17.9. Assume ZFC together with the axiom of constructibility. If λ is an infinite cardinal, then $\mathcal{P}(\lambda) \subseteq L_{\lambda^+}$. In particular, the generalized continuum hypothesis (GCH) holds.

Proof. The "in particular" part is immediate: $|\mathcal{P}(\lambda)| \leq |L_{\lambda^+}| = \lambda^+$. To see the main part, let $X \subseteq \lambda$. We have to see that $X \in L_{\lambda^+}$. We know that $X \in V = L$. Pick $M \preceq L$ an elementary submodel with $\lambda \cup \{X\} \subseteq M$ and $|M| = \lambda$. This is possible by the Löwenheim-Skolem-Tarski theorem[13] M is wellfounded, set-like, and extensional, so by the Mostowski collapsing lemma there is a set-initial N and an isomorphism $\pi : (M, \in) \cong (N, \in)$. Note that M is a model of ZFC + V = L, hence so is N. Moreover, N is an inner model. By Lemma 17.8, $N = L_\alpha$, where $\alpha = \mathrm{OR} \cap N$. Since N is set-initial, $\alpha \subseteq N$, so because $|N| = |M| = \lambda$, we must have $\alpha < \lambda^+$. Thus $\pi(X) \in N \subseteq L_{\lambda^+}$. To finish, it therefore suffices to see that $\pi(X) = X$.

Observe first that $\pi[\lambda] = \lambda$. To see this, we prove by induction that $\pi(\alpha) = \alpha$ for all $\alpha < \lambda$. This is straightforward from the definition of the collapse and the fact that $\lambda \subseteq M$. It follows immediately that $\pi[X] = X$, but be careful: this is not quite what we have to prove! Instead, assume $\alpha \in X$. As π is an isomorphism, $\alpha = \pi(\alpha) \in \pi(X)$. Thus $X \subseteq \pi(X)$. Conversely, if $\alpha \in \pi(X)$, then as π is an isomorphism again, $\alpha = \pi^{-1}(\alpha) \in X$. Thus $\pi(X) \subseteq X$, so $X = \pi(X)$. □

The axiom of choice in L implies the existence of a very definable well-ordering of the real. This in turn gives very definable non-measurable sets (it is not important here to know what measurability exactly means: basically it means one can integrate the characteristic function of such sets), and serious failures of determinacy.

Fact 17.10. Assume ZFC together with the axiom of constructibility. Then there is a $\boldsymbol{\Delta}^1_2$ well-ordering of $\mathbb{R}$ of type ω_1.

Proof idea. Go back to the proof of Theorem 17.6 and consider the well-ordering $\trianglelefteq$ of L_{ω_1} that it gives. Note that $\trianglelefteq$ has type ω_1, since L_α is a countable initial

[13] We are cheating a little bit here: the Löwenheim-Skolem-Tarski theorem was proven in MK, and here we are working within ZFC where it is provably impossible to prove existence of submodels that are elementary with respect to all properties. We can go around this difficulty by imitating the proof and only getting elementarity for all properties with at most $10^{10^{10}}$ symbols, say. In any case, we ignore this detail.

segment of the well-ordering for each $\alpha < \omega_1$, and $L_{\omega_1} = \bigcup_{\alpha<\omega_1} L_\alpha$. We identify $\mathbb{R}$ with $\mathcal{P}\omega$. By Theorem 17.9, $\mathcal{P}(\omega) \subseteq L_{\omega_1}$. Thus $\trianglelefteq\restriction \mathcal{P}(\omega)$ has type ω_1. Now by looking specifically at how $\trianglelefteq$ was defined, we can check that it is $\mathbf{\Sigma}^1_2$ (this is tedious, so we skip this part).

Since $x \trianglelefteq y$ if and only if $y \not\triangleleft x$ or $x = y$, we also get that it is $\mathbf{\Pi}^1_2$, hence it is $\mathbf{\Delta}^1_2$. □

Corollary 17.11. Assume ZFC together with the axiom of constructibility. There is a $\mathbf{\Delta}^1_2$ nonmeasurable subset of $\mathbb{R}^2$.

Proof. Let A be the graph of a $\mathbf{\Delta}^1_2$ well-ordering $\trianglelefteq$ of $[0,1] \times [0,1]$. That is, $A = \{(x,y) \in [0,1]^2 \mid x \trianglelefteq y\}$. Suppose for a contradiction that A is measurable. We integrate the indicator function f_A of A in two ways. Observe that for a fixed $y \in [0,1]$, $A_y := \{x \in [0,1]^2 \mid x \trianglelefteq y\}$ is countable (because $\trianglelefteq$ has type ω_1). Thus $\int_0^1 f_{A_y}(x)dx = 0$. This is valid for any y, so $\int_0^1 \int_0^1 f_A(x,y)dxdy = 0$. On the other hand, for a fixed $x \in [0,1]$, $A_x := \{y \in [0,1]^2 \mid x \trianglelefteq y\}$ is cocountable, so $\int_0^1 f_{A_x}(y)dy = 1$. This is valid for any x, so $\int_0^1 \int_0^1 f_A(x,y)dydx = 1$. This contradicts Fubini's theorem. □

Corollary 17.12. Assuming ZFC and the axiom of constructibility, not all projective sets are determined. In fact, there is a $\mathbf{\Sigma}^1_1$ set which is not determined.

Proof idea. There is a certain game that establishes measurability not only for the payoff set, but for the projection of the payoff set. Since $\mathbf{\Delta}^1_2$ are projections of $\mathbf{\Sigma}^1_1$ sets, the result follows. □

You should contrast this result with Borel determinacy (Theorem 12.15) and the Martin-Steel theorem (Corollary 12.35): if there are sufficiently large cardinals in the universe, then all projective sets are determined. Thus the constructible universe cannot accomodate certain large cardinals[14], and this is one of the many downsides of assuming the axiom of constructibility in set-theoretic practice. In the next section, we say more about those large cardinal axioms.

18. The measure problem and large cardinals

18.1. **The measure problem.** Recall from Definition 13.7 that a *measure* on a set S is a function $m : \mathcal{P}(S) \to \mathbb{R}$ such that $m(S) > 0$, $A \subseteq B \subseteq S$ implies $m(A) \leq m(B)$, and if A and B are disjoint then $m(A \cup B) = m(A) + m(B)$. We are going to require several further conditions on m:

Definition 18.1. Let m be a measure on a set S.

(1) m is a *probability measure* if $m(S) = 1$.
(2) m is *nontrivial* if $m(\{a\}) = 0$ for any $a \in S$.
(3) m is *countably additive* if whenever $(A_n)_{n<\omega}$ is a countable sequence of pairwise disjoint subsets of S, then $m(\bigcup_{n<\omega} A_n) = \sum_{n=0}^{\infty} m(A_n)$.

In this section, we simply call a nontrivial countably additive probability measure a *measure*.

[14] It is not difficulty to see that if λ is an inaccessible cardinals, then L satisfies the property "λ is an inaccessible cardinal". We have in mind here cardinals that are inaccessible but also satisfy much stronger properties, which are in particular not absolute from V to L.

The condition of being a probability measure can be achieved without loss of generality: given an arbitrary measure m, $m' := m/m(S)$ is a probability measure. The nontriviality condition ensures that we do not have the whole weight of the measure on a single point. For example, one way to build a measure on a set S is to fix $a \in S$ and declare a subset of S to have measure 1 if and only if it contains a. This leads to a principal filter, which we want to avoid. Finally, the countable additivity condition is natural and common in probability and measure theory.

If $S = [0, 1]$, it would be natural to ask also for translation invariance: if $A \subseteq [0, 1]$ and $r \in [0, 1]$ is such that $A + r := \{x + r \mid x \in A\} \subseteq [0, 1]$, then $m(A) = m(A + r)$. This would correspond to a notion of "length". Unfortunately such a measure does not exist.

Theorem 18.2. There is no translation-invariant measure on $[0, 1]$.

Proof. Suppose for a contradiction m is such a measure. We first show there must exist a nonzero natural number $n \geq 1$ such that $m([0, 1 - 1/n)) > 0$. Indeed, if not, then $m([0, 1)) = m(\bigcup_{n \in \mathbb{N} - \{0\}} [0, 1 - 1/n)) \leq \sum_{n=1}^{\infty} m([0, 1 - 1/n)) = 0$, hence $m(\{1\}) = 1$, contradicting non-triviality.

We must have that $m([0, 1 - 1/n)) < 1$. Otherwise it is 1 and by translation invariance, $m([1/n, 1)) = 1$, hence $m([0, 1 - 1/n) \cap [1/n, 1)) = m([1/n, 1 - 1/n)) = 1$ (the value of the measure at a singleton is always zero). Translating this by 1/n further, $m([2/n, 1)) = 1$, and eventually we get that $m([n/n, 1)) = m(\emptyset) = 1$, a contradiction.

We conclude that $m([0, 1/n)) > 0$ for any nonzero natural number n. In particular, $m([0, 1/2)) > 0$.

Now consider the relation $\sim$ on $[0, 1]$ given by $x \sim y$ if $x - y$ is rational. This is an equivalence relation. Let $F : [0, 1]/\sim \to [0, 1]$ be a choice function and let $A := \operatorname{ran}(F) \cap [0, 1/2)$. Let $\alpha := m(A)$. For each $r \in \mathbb{Q} \cap [0, 1/2]$, $A + r$ is disjoint from A (except at possibly one point), and is a subset of $[0, 1]$. Moreover, any $x \in [0, 1/2]$ is in some equivalence class, so is of the form $x = y + r$ for $y \in A$ and $r \in [0, 1/2]$. In other words, $[0, 1/2] \subseteq \bigcup_{r \in \mathbb{Q} \cap [0, 1/2]} A + r \subseteq [0, 1]$. By translation invariance, $m(A + r) = \alpha$, and so $0 < m([0, 1/2]) \leq \sum_{r \in \mathbb{Q} \cap [0, 1/2]} \alpha \leq 1$. This is a contradiction. □

We now drop translation invariance and investigate whether there is even a measure on $[0, 1]$. More generally, we will investigate for which set S there is a measure on S. Observe that, for this problem, the exact set S does not matter: only its cardinality. At this point an straightforward observation is that it makes sense to consider the least cardinal carrying a measure: all the cardinals above it will also have an induced measure.

Lemma 18.3. If there is a measure on a cardinal λ, then there is a measure on θ for any $\theta \geq \lambda$.

Proof. Let m be a measure on λ. Define $m' : \mathcal{P}(\theta) \to [0, 1]$ by $m'(X) := m(X \cap \lambda)$. Check that this works. □

So how big can the least cardinal carrying a measure be? Another easy observation is:

Remark 18.4. If there is a measure on S, then S is uncountable. Indeed, S is not empty (otherwise it would be disjoint from itself, and so would have measure 2, not

1). Moreover, since any singleton has measure zero and the measure is countably additive, any countable set has measure zero.

The smallest uncountable size is $\aleph_1$. However, we have:

Theorem 18.5. There is no measure on $\aleph_1$. In particular, if there is a measure on $[0, 1]$, then the continuum hypothesis fails.

Proof. Let $S := \omega_1$, and suppose for a contradiction $m : \mathcal{P}(S) \to [0, 1]$ is a measure. Let I be the ideal of sets of measure zero: $I := \{X \subseteq S \mid m(S) = 0\}$. Observe that I is indeed an ideal (exercise). In fact, I contains all singletons (by nontriviality of m), and I is countably complete, in the sense that if $(A_n)_{n<\omega}$ is a sequence of sets in I, then $\bigcup_{n<\omega} A_n \in I$ (the definition of an ideal only requires I to be closed under finite unions).

This shows that I has some of the properties of the nonstationary ideal. An important theorem about stationary sets is Solovay's splitting theorem: any stationary set can be split into $\aleph_1$-many pairwise disjoint stationary subsets. Is this true here? The answer is no:

Claim: There is no collection $(S_i)_{i<\omega_1}$ of pairwise disjoint subsets of S such that for all $i < \omega_1$, $m(S_i) > 0$ (i.e. $S_i \notin I$).

Proof of Claim: For each $i < \omega_1$, pick n_i a nonzero natural number such that $m(S_i) > \frac{1}{n_i}$. By the pigeonhole principle, there is $n < \omega$ and an unbounded $X \subseteq \omega_1$ such that $n_i = n$ for all $i \in X$. This is a contradiction, because then by pairwise disjointness $1 = m(S) \geq m(\bigcup_{i\in X} S_i) \geq \sum_{i\in X} m(S_i)$, which is unbounded. $\dagger_{\text{Claim}}$

We now use similar ideas as in the proof of Solovay's splitting theorem to see that the answer is yes after all. We first build an *Ulam matrix*: a matrix $(A_{\alpha n})_{\alpha<\omega_1,n<\omega}$ such that for all $n < \omega$ and all $\alpha < \beta < \omega_1$:

(1) $A_{\alpha n} \cap A_{\beta n} = \emptyset$.
(2) $S - \bigcup_{n<\omega} A_{\alpha n}$ is countable.

Why is this possible? For each $i < \omega_1$, pick a surjection $f_i : \omega \to i$, and let $A_{\alpha n} := \{i < \omega_1 \mid f_i(n) = \alpha\}$. The first property is immediate. As for the second, because each f_i is surjective, $\bigcup_{n<\omega} A_{\alpha n} = [\alpha, \omega_1)$, which is cocountable.

Now that we have built the Ulam matrix, we observe that for every fixed $\alpha < \omega_1$, $S - \bigcup_{n<\omega} A_{\alpha n}$ is in I, so $\bigcup_{n<\omega} A_{\alpha n}$ has measure 1. Thus by countable additivity, there exists $n = n_\alpha < \omega$ such that $A_{\alpha n}$ has positive measure. Again, by the pigeonhole principle there must exist $X \subseteq \omega_1$ of size $\aleph_1$ and $n < \omega$ such that α in X implies $n = n_\alpha$. Thus $(A_{\alpha n})_{\alpha\in X}$ is an uncountable pairwise disjoint collection of positive measure subsets of S, contradicting the claim. □

Remark 18.6. Another proof of the "in particular" part of Theorem 18.5 uses the same idea as Corollary 17.11: suppose for a contradiction there is a measure on $[0, 1]$ and consider the graph of a well-ordering of $[0, 1]$ of type ω_1. To carry out this proof formally one would need to prove Fubini's theorem in enough generality first.

The cardinal $\aleph_2$ does not have a measure either. This will follow from a straightforward generalization of the above proof. It is convenient to make the following definitions first:

Definition 18.7. For κ an infinite cardinal, an ideal I is *κ-complete* if for any $0 < \alpha < \kappa$ and any sequence $(A_i)_{i<\alpha}$ of members of I, then $\bigcup_{i<\alpha} A_i \in I$. Dually define a filter to be κ-complete if it is closed under intersection of size strictly less than κ. Similarly, a measure on a set S is *κ-additive* if for any $0 < \alpha < \kappa$ and any sequence $(A_i)_{i<\alpha}$ of pairwise disjoint subsets of S, $m(\bigcup_{i<\alpha} A_i) = \sum_{i<\alpha} m(A_i)$. Here, infinite sums are defined in terms of the supremums of finite partial sums.

Note that any finitely additive measure is $\aleph_0$-additive, and a measure is countably additive (as in this section) if and only if it is $\aleph_1$-additive. Similarly, any filter or ideal is $\aleph_0$-complete. To check whether a measure is κ-additive, it suffices to check its ideal of measure zero sets:

Lemma 18.8. If m is a measure on a set S and I is its ideal of measure zero sets, then I is κ-complete if and only if m is κ-additive.

Proof. If m is κ-additive and $(A_i)_{i<\alpha}$ is a sequence of fewer than κ-many sets of measure zero, then $m(\bigcup_{i<\alpha} A_i) \leq \sum_{i<\alpha} m(A_i) = 0$, so I is κ-complete. Conversely, assume that I is κ-complete and let $(A_i)_{i<\alpha}$ be a sequence of fewer than κ-many pairwise disjoint subsets of S. All except countably-many must have measure zero (otherwise uncountably-many A_i's have positive measure, and as in the proof of the claim in Theorem 18.5, we get that for some natural number n, uncountably-many of them must have measure at least $1/n$, hence the measure of S is unbounded, contradiction). Without loss of generality, $\alpha \geq \omega$, and we can re-arrange so that $(A_i)_{\omega \leq i<\alpha}$ have measure zero. Then by countable additivity and κ-completeness of the measure zero ideal, $m(\bigcup_{i<\alpha} A_i) = \sum_{i<\omega} m(A_i) = \sum_{i<\alpha} m(A_i)$, as desired. □

Lemma 18.9. If κ is the minimal cardinal such that there is a measure m on κ, then m is κ-additive.

Proof. By Lemma 18.8, it suffices to check that the ideal I of measure zero sets is κ-complete. If not, fix $\lambda < \kappa$ such that there exists $(A_i)_{i<\lambda}$ of measure zero such that $A := \bigcup_{i<\lambda} A_i$ has strictly positive measure. Let $\alpha := m(A)$. Without loss of generality, the A_i's are pairwise disjoint (otherwise define inductively $A_i^* := A_i - \bigcup_{j<i} A_j^*$). Define a surjection $f : A \to \lambda$ by letting $f(a)$ be the unique $i < \lambda$ such that $a \in A_i$. Define $m' : \mathcal{P}(\lambda) \to [0, 1]$ by $m'(X) := \frac{1}{\alpha} m(f^{-1}[X])$. We have that m' is a measure on λ. Most of the properties are easy to check because taking inverse images of sets preserves union and intersections. m is not trivial because for any $b \in \lambda$, $f^{-1}[\{b\}] = A_i$ for some i, which has measure zero. This contradicts minimality of κ. □

Theorem 18.10. If κ is the minimal cardinal such that there is a measure on κ, then κ is weakly inaccessible.

Proof. Let m be a measure on κ. By Lemma 18.9, m is κ-additive. By imitating the proof of Theorem 18.5, we see that κ cannot be a successor cardinal, so κ is a limit cardinal. We have seen already that κ is uncountable and κ is also regular: if $\theta := \mathrm{cf}(\kappa) < \kappa$, then let $(\kappa_i)_{i<\theta}$ be a cofinal sequence. We know every subset of κ of cardinality strictly less than κ has measure zero, by κ-additivity and the fact every singleton has measure zero. In particular, $m(\kappa_i) = 0$ for all $i < \theta$. Since $\theta < \kappa$, κ-addivity implies that $\kappa = \bigcup_{i<\theta} \kappa_i$ also has measure zero, contradiction. □

We have shown in particular that, if there is a measure on $[0, 1]$, then $2^{\aleph_0} \geq \kappa$ for some weakly inaccessible cardinal κ (in other words, the continuum will be *very* big). Recall that the existence of weakly inaccessible cardinals is impossible to prove. Let us now investigate further the properties of the measure. It is natural to ask whether it can take other values than zero or one:

Definition 18.11. Let m be a measure on a set S.

1. m is *two-valued* if $m(A) \in \{0,1\}$ for any $A \subseteq S$.
2. An *atom* (for m) is a subset $A \subseteq S$ such that for any $X \subseteq A$, either $m(X) = 0$ or $m(A - X) = 0$. We say that m is *atomless* if there are no atoms for m.

Remark 18.12. If m is a measure on S and A is an atom, then there is a two-valued measure m' on A given by $m'(X) := \frac{m(X)}{m(A)}$.

We can now answer the original question on whether there is a measure on $[0,1]$:

Theorem 18.13 (Ulam's theorem)**.** If there is a measure, then either there is a two-valued measure, or there is a measure on $2^{\aleph_0}$.

We need one more lemma to prove this.

Lemma 18.14. Let m be an atomless measure on S.

1. For every $\epsilon > 0$ and every $X \subseteq S$ of positive measure, there is $Y \subseteq X$ such that $0 < m(Y) < \epsilon$.
2. For every $X \subseteq S$, there is $Y \subseteq X$ such that $m(Y) = \frac{1}{2}m(X)$.

Proof.

1. Let $X_0 := X$. For each $n < \omega$, we can find $X_{n+1} \subseteq X_n$ such that $m(X_{n+1}) \leq \frac{1}{2}m(X_n)$. Indeed, as X_n is not an atom there is $X' \subseteq X_n$ so that both X' and $X_n - X'$ have positive measure, and so one of these two must have at most half the measure of X_n. In the end, $m(X_n) \leq \frac{1}{2^n}m(X)$, and so by taking n sufficiently big, X_n is as desired.
2. Suppose not. We build $(X_i)_{i<\omega_1}$ as follows. Pick $X_0 \subseteq X$ so that $m(X_0) \leq \frac{1}{2}m(X)$ (this is possible by the previous part). Now given X_i, if $m(X_i) = \frac{1}{2}m(X)$ we are done. Otherwise, pick $X_i \subseteq X_{i+1} \subseteq X$ such that $m(X_{i+1} - X_i) > 0$ and $m(X_{i+1}) \leq \frac{1}{2}m(X)$. This is possible by the previous part also. Finally, given $(X_j)_{j<i}$, with i limit, let $X_i := \bigcup_{j<i} X_j$. As in the claim of the proof of Theorem 18.5, there must exist $0 < n < \omega$ such that for unboundedly-many $i < \omega_1$, $m(X_{i+1} - X_i) \geq \frac{1}{n}$, contradicting the fact that $m(X) \leq 1$.

□

Proof of Ulam's theorem. Let m be a measure on a set S. If m has an atom, then by Remark 18.12 there is a two-valued measure. Assume now that m is atomless. We define $(A_s)_{s \in {}^{<\omega}2}$ such that, for all $s \in {}^{<\omega}2$:

1. A_s is a subset of S of positive measure.
2. $A_{s\frown 0}, A_{s\frown 1}$ is a partition of A_s, and $m(A_{s\frown 0}) = A_{s\frown 1}$.

This is possible: we proceed by induction on the length of s. Let $A_{\langle\rangle} := S$, and given A_s, use Lemma 18.14 to get $A_{s\frown 0} \subseteq A_s$ such that $m(A_{s\frown 0}) = \frac{1}{2}m(A_s)$, and let $A_{s\frown 1} := A_s - A_{s\frown 0}$.

This is enough: for each $f \in {}^{\omega}2$, let $A_f := \bigcap_{n<\omega} A_{f\restriction n}$. By construction, if $f \neq g$ then $A_f \cap A_g = \emptyset$. Moreover, $m(A_f) = 0$ for any $f \in {}^{\omega}2$. We can now define a measure $m' : \mathcal{P}({}^{\omega}2) \to [0,1]$ by $m'(X) := m(\bigcup_{f \in X} A_f)$. This works (nontriviality is precisely because $m(A_f) = 0$ for any f, and the other properties are easy to check). □

In conclusion, if there is a measure at all, then either is a measure on $[0, 1]$ (and hence a weakly inaccessible below $2^{\aleph_0}$), or there is a two-valued measure (or both). Note that in case there is an atomless measure, the measure on ${}^{\omega}2$ obtained from the proof of Ulam's theorem induces a measure μ on $[0, 1]$ (via the map $x \mapsto \sum_{k=1}^{\infty} 2^{-k} x(k)$). This measure satisfies $\mu((a, b)) = b - a$ whenever $0 \leq a < b \leq 1$, hence extends the usual Lebesgue measure.

We investigate this second possibility in the next subsection.

18.2. **Measurable cardinals.** We have seen (Exercise 13.13) that there is a close connection between two-valued measures and ultrafilters: if m is a two-valued measure on a set X then $U := \{A \subseteq S \mid m(A) = 1\}$ is an ultrafilter. Since the measure is nontrivial, the ultrafilter will be nonprincipal. If (as we are assuming now) m is countably complete, we get that U is itself $\aleph_1$-complete in the sense of Definition 18.7: it is closed under countable intersections. Can we build such ultrafilters? We saw that the club filter is $\aleph_1$-complete, and we know that it extends to an ultrafilter. However we may loose the $\aleph_1$-completeness in the extension process. In fact, it turns out that the existence of an $\aleph_1$-complete ultrafilter has very strong consequences.

Lemma 18.15. If κ is the least cardinal such that there is a nonprincipal $\aleph_1$-complete ultrafilter on κ, then any such ultrafilter is κ-complete.

Proof. This is similar to the proof of Lemma 18.9. Let U be an $\aleph_1$-complete ultrafilter on κ. Of course, κ is uncountable (why?). It is easier to work with the dual ideal $I := \{\aleph_1 - A \mid A \in U\}$. Suppose for a contradiction there exists $\lambda < \kappa$ and $(X_\alpha)_{\alpha<\lambda}$ all in I but with $X := \bigcup_{\alpha<\lambda} X_\alpha \notin I$. Without loss of generality, the X_α's are pairwise disjoint. Note that as U is an ultrafilter, $X \in U$. Let $J := \{Y \subseteq \lambda \mid \bigcup_{\alpha\in Y} X_\alpha \in I\}$. Check that J is an $\aleph_1$-complete ideal on λ. It is nonprincipal since $X_\alpha \in I$ for all $\alpha < \lambda$. It is prime: if $Y \subseteq \lambda$ and $Y \notin J$, then $\bigcup_{\alpha\in Y} X_\alpha \notin I$, so $\bigcup_{\alpha\in Y} X_\alpha \in U$, hence $X - \bigcup_{\alpha\in Y} X_\alpha = \bigcup_{\alpha\notin Y} X_\alpha \in J$. □

We give cardinals satisfying the conclusion of Lemma 18.15 a name:

Definition 18.16. An uncountable cardinal κ is *measurable* if there is a κ-complete nonprincipal ultrafilter on κ.

Corollary 18.17. If there is a measure, then either there is a measure on $[0, 1]$ or there is a measurable cardinal.

Proof. By Ulam's theorem and Lemma 18.15. □

We state without proof that existence of a measure on $[0, 1]$ is *equiconsistent* with existence of a measurable cardinal. This means that if one could disprove the existence of measurable cardinals, then one would be able to also disprove existence of measures on $[0, 1]$, and vice-versa:

Fact 18.18 (Solovay)**.** If there is a measure on $[0, 1]$, then there is a model of ZFC with a measurable cardinal. Conversely, if there is a measurable cardinal then there is a model of ZFC with a measure on $[0, 1]$.

Note that, if it weren't for the "uncountable", $\aleph_0$ would be measurable. Thus in a sense we are trying to look for an uncountable cardinal that interacts with the cardinals below it the same way that $\aleph_0$ interacts with the finite cardinals. We will see many senses in which this analogy is realized. In particular, a measurable cardinal will be very big.

Theorem 18.19. Any measurable cardinal is strongly inaccessible.

Proof. Let κ be a measurable cardinal. As in the proof of Theorem 18.10, κ has to be regular. It remains to see that it is strong limit. Suppose not. Then there exists $\mu < \kappa$ such that $2^\mu \geq \kappa$. In particular, there is $S \subseteq {}^\mu 2$ of cardinality κ, hence by definition of a measurable cardinal there must exist a nonprincipal κ-complete ultrafilter V on S. For each fixed $\alpha < \mu$ and each $f \in S$, either $f(\alpha) = 0$ or $f(\alpha) = 1$. Since V is an ultrafilter, there is therefore $b_\alpha \in \{0, 1\}$ and $X_\alpha \in V$ such that $f \in X_\alpha$ implies $f(\alpha) = b_\alpha$. In words, almost all functions in S take on the same value at α. This plays badly with κ-completeness of V: the set $X := \bigcap_{\alpha<\mu} X_\alpha$ must be in V. In particular (since V is not principal), there exists $f \neq g$ both in X. However we know by definition of X_α that $f(\alpha) = b_\alpha = g(\alpha)$ for all $\alpha < \lambda$, contradiction. □

What are other properties of $\aleph_0$ that transfer to κ? One theorem about $\aleph_0$ is Kőnig's tree lemma: any $\aleph_0$-tree has a branch. We saw that it false that any $\aleph_1$-tree has a branch. However, at measurable cardinals the situation is just like at $\aleph_0$:

Theorem 18.20. If κ is a measurable cardinal, then any κ-tree has a branch.

Proof. Let T be a κ-tree. Note that $|T| = \sum_{\alpha<\kappa} |\operatorname{Lev}_\alpha(T)| \leq \kappa$, and the reverse inequality is because every level is non-empty. Thus we can fix a nonprincipal κ-complete ultrafilter U on T. We build a branch just like in the proof of Kőnig's lemma, but we use the ultrafilter. In details, we build $(s_\alpha)_{\alpha<\kappa}$ such that for all $\alpha < \beta < \kappa$:

(1) $s_\alpha \in \operatorname{Lev}_\alpha(T)$.
(2) $s_\alpha \subseteq s_\beta$.
(3) $\{s \in T \mid s_\alpha \subseteq s\} \in U$.

This is possible. For a fixed $t \in T$, let $X_t := \{s \in T \mid t \subseteq s\}$. For the base case, we let $s_0 := \langle\rangle$. Note that $X_{s_0} = T \in U$ as U is a filter. For $\delta < \kappa$ limit, assuming $(s_\alpha)_{\alpha<\delta}$ are given, let $X := \bigcap_{\alpha<\delta} X_{s_\alpha}$. By κ-completeness, $X \in U$. Since U is a filter, X is not empty, so pick $s \in U$. We have that $s_\alpha \subseteq s$ for all $\alpha < \delta$, so s has height at least δ. Let $s_\delta := s \upharpoonright \delta$. Note that $X_{s_\delta} = X \in U$.

For the successor case, assume s_α is given. We know $X_{s_\alpha} \in U$. Let $Y := \{i < \kappa \mid s_\alpha \frown i \in T\}$. We know that $|Y| \leq |\operatorname{Lev}_{\alpha+1}(T)| < \kappa$. Moreover, $X_{s_\alpha} = \{s_\alpha\} \cup \bigcup_{i \in Y} X_{s_\alpha \frown i}$. Since U is not principal, $\{s_\alpha\} \notin U$. By κ-completeness (and since $X_{s_\alpha} \in U$), there must exist $i \in Y$ such that $X_{s_\alpha \frown i} \in U$. Let $s_{\alpha+1} := s_\alpha \frown i$. □

As another example, you saw in the assignment that the infinite Ramsey theorem does not generalize to $\aleph_1$. In fact, there is a coloring of $[\mathbb{R}]^2$ with two colors that has no uncountable homogeneous sets (color according to whether a well-ordering of the reals agrees with the usual ordering). However again for measurable cardinals one can imitate the proof of the infinite Ramsey theorem using the completeness of the ultrafilter to go through limit stages. We can even prove a stronger statement: the number of color can be any cardinal less than κ, and we can color all tuples at once.

Exercise 18.21.

(1) Assume κ is a measurable cardinal. If $\lambda < \kappa$ and $f : [\kappa]^{<\omega} \to \lambda$, there exists $X \subseteq \kappa$ of size κ such that $f \upharpoonright [X]^n$ is constant for all $n < \omega$.

(2) Is the above result true if $\kappa = \aleph_0$?

One can also show that any measurable cardinal is a Mahlo cardinal (in fact much more). A further result relates measurables to the determinacy of projective sets.

Fact 18.22 (Martin's theorem)**.** If there is a measurable cardinal, then any $\mathbf{\Sigma}_1^1$ set is determined.

Corollary 18.23. Assuming ZFC and the axiom of constructibility, there are no measurable cardinals.

Proof. By Fact 18.22 and Corollary 17.12. □

This gives one more shortcoming of the constructible universe: it is "too small" to accomodate large cardinals. Of course it is impossible to prove that measurable cardinals exist, so it just could be that they imply so many nice theorems that eventually one can derive a contradiction from their existence. However so far no such contradiction has been found.

Measurable cardinals and other large cardinals take to an extreme the idea that in order to prove certain results about small objects, one still needs the universe to be "large". For example the determinacy of $\mathbf{\Sigma}_1^1$ sets fails in the constructible universe so in some sense one needs large cardinals to prove it. Even the determinacy of Borel sets provably needs the existence of $\beth_\alpha$ for every $\alpha < \omega_1$. This in turn is a consequence of the axiom of replacement. In a sense, large cardinals are a continuation of this process: to be able to prove stronger results (about determinacy, say, or about existence of measures on the reals), one sometimes needs to add new axioms of infinity, or *large cardinal axioms*, including existence of inaccessible, Mahlo, measurables, and much more. The study of those axioms is an active area of research in set theory.

Appendix A. The axioms of (Morse-Kelley) set theory

We use the letters $A, B, C, ...$ for classes (a primitive concept). For any two classes A and B, we can ask whether $A \in B$ (another primitive concept). We say A is a *member* of B. We say A is a *subclass* of B, written $A \subseteq B$, if any member of A is also a member of B. A *set* is a class A so that there is a class B with $A \in B$. We use lowercase letters for sets. A class that is not a set is called a *proper class*. The axioms below will guarantee the existence of a class with no members, called the *empty class*, and denoted by $\emptyset$. There will also be a class of all sets, denoted SET. The basic operations of pairing, union, complement, and intersection, can also be defined from the axioms. There is also a notion of ordered pairs, and for any two classes A, B of $A \times B$, the class of ordered pairs with first component in A and second component in B. A *class function* from A to B is a triple $F = (A, B, \Gamma)$, where A, B are classes (the domain and codomain), and Γ is a subclass of $A \times B$ (the graph of the function) so that for any $a \in A$ there is a unique $b \in B$ with $(a, b) \in \Gamma$. We write $F : A \to B$ to indicate that F is a class function from A to B. The *range* of F is the set of $b \in B$ so that there is an $a \in A$ with $F(a) = b$.

- Class axioms:
  - (Extensionality) If A and B are classes so that $A \subseteq B$ and $B \subseteq A$, then $A = B$.
  - (Specification) For any property $P(x)$ of sets, there is a class whose members are exactly the sets a satisfying $P(a)$.
- Basic set-building axioms:
  - (Empty set) $\emptyset$ is a set.
  - (Subset) If b is a set and $a \subseteq b$, then a is a set.
  - (Pairing) If a and b are sets, then $\{a, b\}$ is a set.
  - (Union) If a is a set, then $\bigcup a$ is a set.
  - (Power set) If b is a set, then the class $\{a \mid a \subseteq b\}$ is a set.
- "Hard" axioms:
  - (Infinity) There is a set a that contains $\emptyset$ and is closed under successors ($x \in a$ implies $x \cup \{x\} \in a$).
  - (Replacement) If F is a class function whose domain is a set, then the range of F is a set.
  - (Foundation) For any non-empty class A, there is $a \in A$ such that $a \cap A = \emptyset$.
  - (Choice) There exists a class function $F : \text{SET} \to \text{SET}$ such that $F(a) \in a$ for all non-empty sets a.

Appendix B. The real numbers

We will take the following fact, along with some basic facts from calculus, as given (see Abbott's book if you need a refresher about the real numbers):

Fact B.1. There is a set $\mathbb{R}$, called the *set of real numbers*, a relation $\leq_{\mathbb{R}}$ on $\mathbb{R}$ (usually just written $\leq$), and functions $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ (*addition*), $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ (*multiplication*) (written as binary operations, e.g. $a+b$ instead of $+(a, b)$ — we may also write ab instead of $a \cdot b$ and adopt the usual conventions regarding brackets) satisfying the following properties, for any real numbers a, b, c:

- $\mathbb{N} \subseteq \mathbb{R}$ (where $\mathbb{N}$ is as defined in Definition 2.11).
- $(\mathbb{R}, \leq)$ is a linear order.

- Associativity of addition and multiplication: $(a+b)+c = a+(b+c)$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- Commutativity of addition and multiplication: $a+b = b+a$, $a \cdot b = b \cdot a$.
- Zero is the identity element for addition: $a + 0 = a$.
- One is the identity element for multiplication: $a \cdot 1 = a$.
- Distributivity: $a(b+c) = ab + ac$.
- Existence of additive inverse: there is a (unique) element $-a$ so that $a + (-a) = 0$. We write $a - b$ instead of $a + (-b)$.
- Existence of multiplicative inverse: if $a \neq 0$, there is a (unique) element a^{-1} such that $aa^{-1} = 1$. We write $\frac{a}{b}$ (or a/b) instead of ab^{-1}.
- Addition and multiplication respect the order: if $a \leq b$, then $a + c \leq b + c$, and if $c \geq 0$, $ac \leq bc$.
- $a \cdot 0 = 0$, $0 < 1$, and for $n \in \mathbb{N}$ $\mathrm{S}\, n = n + 1$ [in fact, these all follow from the other axioms].
- Completeness axiom: any non-empty set of real that is bounded above has a supremum, and any non-empty set of real that is bounded below has an infimum.

We also define the *absolute value* $|x|$ of a real number x to be x if $x \geq 0$ or $-x$ otherwise. The *square root* $\sqrt{x}$ of a nonnegative real number x is the unique nonnegative real number y such that $y^2 (= yy) = x$.

Definition B.2. The set $\mathbb{Z}$ of *integers* is the set $\{x \in \mathbb{R} \mid |x| \in \mathbb{N}\}$. The set $\mathbb{Q}$ of *rationals* is the set $\{\frac{n}{m} \mid n \in \mathbb{Z}, m \in \mathbb{Z} - \{0\}\}$.

It can be checked that the set of natural numbers, integers, and rationals are all closed under addition and multiplication. The integers and rationals are closed under subtraction as well. The rationals are also closed under division. Finally, the natural numbers are wellfounded under (the restriction of) the ordering $<$: any non-empty subset of natural numbers has a minimal element.

E-mail address: sebv@math.harvard.edu
URL: http://math.harvard.edu/~sebv/

Department of Mathematics, Harvard University, Cambridge, Massachusetts, USA