In this class, we will explore the properties of infinite sets. We take the natural numbers \( \mathbb{N} = \{0, 1, 2, 3, \ldots\} \) as given. We write \( [n] = \{0, 1, \ldots, n-1\} \).

1 Cardinality

**Definition 1.1.** Let \( f : X \to Y \) be a function between two sets. We say \( f \) is *injective* if \( f(x) = f(y) \) implies \( x = y \). We say \( f \) is *surjective* if for all \( y \in Y \), there is \( x \in X \) such that \( f(x) = y \). If \( f \) is both surjective and injective, we say \( f \) is *bijective*.

If \( f : X \to Y \) is a bijection, then for every \( y \in Y \) there is a unique \( x \in X \) such that \( f(x) = y \). That is, \( f \) matches up all the elements of \( X \) and \( Y \) in a 1–1 fashion. We can thus consider \( X \) and \( Y \) to have the same number of elements. If \( f : X \to Y \) is an injection, then \( f \) matches up all of the elements of \( X \) with only some of the elements of \( Y \), so we can consider \( X \) to be smaller than \( Y \). This motivates the following definition.

**Definition 1.2.** Let \( X \) and \( Y \) be sets. If there exists an injection \( X \to Y \), we write \( |X| \leq |Y| \). If there exists a bijection, we write \( |X| = |Y| \).

If \( |X| = |[n]| \) for some \( n \in \mathbb{N} \), we say \( X \) is *finite*. Otherwise, we say \( X \) is *infinite*. If \( |X| \leq \mathbb{N} \), we say \( X \) is *countable*. Otherwise, we say \( X \) is *uncountable*.

You should think of \( |X| \) as being a “number” which is the number of elements of \( X \), or the *cardinality* of \( X \). However, for now, this is just a formal notation—if \( X \) is infinite, we don’t have any actual object which is the number \( |X| \). It is only meaningful to compare \( |X| \) to \( |Y| \) for other \( Y \) and ask whether \( |X| \leq |Y| \) or \( |X| = |Y| \). Note that since every bijection is an injection, \( |X| = |Y| \) implies \( |X| \leq |Y| \). If \( |X| \leq |Y| \) but \( |X| \neq |Y| \), then we say \( |X| < |Y| \).

Note, however, that with infinite sets, even if there is an injection \( f : X \to Y \) that is not a bijection, we can only say \( |X| \leq |Y| \), not \( |X| < |Y| \). Even though our particular map \( f \) is not a bijection, there might be some other bijection \( g : X \to Y \). For example, let \( X \) be the set of even natural numbers and \( Y \) be the set of natural numbers. Then \( f(n) = n \) is an injection from \( X \) to \( Y \) which is not a bijection, but \( g(n) = n/2 \) is a bijection from \( X \) to \( Y \). That is, \( Y \) is the same size as one of its proper subsets!

**Example 1.3.** We can find a bijection \( f : \mathbb{N} \to \mathbb{Z} \) by defining:

\[
f(0) = 0, f(1) = 1, f(2) = -1, f(3) = 2, f(4) = -2, \ldots
\]

Thus \( |\mathbb{N}| = |\mathbb{Z}| \).

**Example 1.4.** Let \( \mathcal{P}(\mathbb{N}) \) be the *power set* of \( \mathbb{N} \), the set of all subsets of \( \mathbb{N} \). Then \( \mathcal{P}(\mathbb{N}) \) is uncountable. To prove this, suppose that \( f : \mathbb{N} \to \mathcal{P}(\mathbb{N}) \) is a bijection; we will get a contradiction from this assumption.

Define a subset \( A \) of \( \mathbb{N} \) as follows:

\[
A = \{n \in \mathbb{N} : n \notin f(n)\}.
\]

\[1\]The property of being in bijection with a proper subset is even sometimes taken as the definition of when a set is infinite,
Since we assumed \( f \) was a bijection, we must have \( A = f(n) \) for some \( n \). But then by definition of \( A \),
\[
n \in A \iff n \not\in f(n) = A,
\]
which is a contradiction.

Thus there cannot exist a bijection between \( \mathbb{N} \) and \( \mathcal{P}(\mathbb{N}) \). Since there certainly is an injection \( \mathbb{N} \to \mathcal{P}(\mathbb{N}) \) (for instance, send \( n \) to the set \( \{n\} \)), we have \( |\mathbb{N}| < |\mathcal{P}(\mathbb{N})| \).

**Remark 1.5.** A variation on this argument shows that the real numbers \( \mathbb{R} \) are also uncountable.

**Example 1.6.** Nothing about the argument in Example 1.4 was special to the set \( \mathbb{N} \); we could replace \( \mathbb{N} \) by any other set. Thus for any set \( X \), \( |X| < |\mathcal{P}(X)| \). Thus for any infinite set, there is another that is strictly larger.

Although we have defined this notation \( |X| \leq |Y| \) and \( |X| = |Y| \), we must be careful to check that it satisfies the properties we would expect of \( \leq \) and \( = \).

**Proposition 1.7.** The relation \( |X| = |Y| \) is an equivalence relation. That is:

1. \(|X| = |X|\).
2. If \(|X| = |Y|\), then \(|Y| = |X|\).
3. If \(|X| = |Y|\) and \(|Y| = |Z|\), then \(|X| = |Z|\).

**Proof.** For (1), the identity map \( f(x) = x \) is a bijection from \( X \) to itself.

For (2), suppose \( f : X \to Y \) is a bijection. Define its *inverse* \( f^{-1} : Y \to X \) by saying that \( f^{-1}(y) = x \) iff \( f(x) = y \). Then \( f^{-1} \) is a bijection, so \(|Y| = |X|\).

For (3), suppose \( f : X \to Y \) and \( g : Y \to Z \) are bijections. Define \( h : X \to Z \) by \( h(x) = g(f(x)) \). Then \( h \) is a bijection, so \(|X| = |Z|\). \( \square \)

**Proposition 1.8.** The relation \( |X| \leq |Y| \) is a preorder. That is:

1. \(|X| \leq |X|\).
2. If \(|X| \leq |Y|\) and \(|Y| \leq |Z|\), then \(|X| \leq |Z|\).

**Proof.** The proof is identical to that of Proposition 1.7(1,3), substituting “bijection” for “bijection” everywhere. \( \square \)

That much was straightforward. However, there’s a couple other key properties that \( \leq \) should have. First, if \(|X| \leq |Y|\) and \(|Y| \leq |X|\), then we should have \(|X| = |Y|\). Second, for any \( X \) and \( Y \), we should have either \(|X| \leq |Y|\) or \(|Y| \leq |X|\). However, it’s not at all obvious how to prove these from the definitions. The first of these is a famous theorem:

**Theorem 1.9** (Schröder-Bernstein). If \(|X| \leq |Y|\) and \(|Y| \leq |X|\), then \(|X| = |Y|\).

**Proof.** The proof of this is best seen with a picture (see Figure 1).

Suppose \( f : X \to Y \) and \( g : Y \to X \) are injections. We use \( g \) to identify \( Y \) with a subset of \( X \), and we use \( f \) to identify \( X \) with a subset of \( Y \). This allows us to get an infinite chain of subsets of \( X \), as shown in Figure 1. The idea is then to get a bijection from \( X \) to \( Y \) by pushing the shaded parts of the picture in one level in the nesting. Let’s make this more precise. Let \( X_0 = X \), \( Y_0 = g(Y) \subseteq X_0 \), \( X_1 = g(f(X)) \subseteq Y_0 \), \( Y_1 = g(f(g(Y))) \subseteq X_1 \), and so on, as in the picture. We then define \( h : X \to Y_0 \) as follows. If \( x \in X_n \setminus Y_n \) for some \( n \), we let \( h(x) = g(f(x)) \in X_{n+1} \setminus Y_{n+1} \). Otherwise, we let \( h(x) = x \).

It is not hard to see that \( h \) is a bijection, so \(|X| = |Y_0|\). Since \( g \) is a bijection from \( Y \) to \( Y_0 \), we also have \(|Y_0| = |Y|\), so \(|X| = |Y|\). \( \square \)

though this property is usually called “Dedekind infinite”. See also Exercise 1.2
The second property, that either \(|X| \leq |Y|\) or \(|Y| \leq |X|\), is much harder. Let’s prove it in the case that \(X = \mathbb{N}\).

**Proposition 1.10.** For any set \(Y\), either:

1. \(|\mathbb{N}| \leq |Y|\), or
2. \(|Y| < |\mathbb{N}|\), and \(Y\) is finite.

**Proof.** We define an injection \(f : \mathbb{N} \rightarrow Y\) by induction. Given the values of \(f(0), f(1), \ldots, f(n-1)\), pick some element \(y \in Y\) not equal to \(f(0), \ldots, f(n-1)\) and define \(f(n) = y\).

If no such \(y\) exists, of course, then this argument fails. However, in that case, \(f\) must have defined a bijection from \([n]\) to \(Y\), so \(Y\) is finite and \(|Y| < |\mathbb{N}|\). Thus in both cases, we get either a injection from \(\mathbb{N}\) to \(Y\) or an injection from \(Y\) to \(\mathbb{N}\). \(\square\)

The key to this proof is induction (or “counting”). Basically, to build an injection from \(\mathbb{N}\) to \(Y\), we just start matching up elements one at a time by induction. If we run out of elements of \(\mathbb{N}\), this means \(|Y| \leq |\mathbb{N}|\); otherwise, we can define an injection from all of \(\mathbb{N}\) to \(Y\). To put it another way, we’re using \(\mathbb{N}\) to “count” the set \(Y\). If we replace \(\mathbb{N}\) by an arbitrary set \(X\), we can almost make this argument work. We can say:

Define an injection from \(f : X \rightarrow Y\) as follows. First, pick an element \(a_0 \in X\), and pick some \(z_0 \in Y\) and let \(f(a_0) = z_0\). Then pick a different \(a_1 \in X\), and a different \(z_1 \in Y\), and let \(f(a_1) = z_1\). Keep picking new elements of \(X\) and \(Y\) to pair up one-by-one. If we run out of elements of \(X\) to pick, we’ve defined an injection from \(X\) to \(Y\). If we ever run out of elements of \(Y\) to pick, we’ve defined a bijection between a subset of \(X\) and \(Y\), whose inverse gives an injection from \(Y\) to \(X\). That is, we use the elements of \(X\) to “count” the elements of \(Y\), until one of the sets runs out.

What’s wrong with this argument? Well, if \(X\) and \(Y\) are infinite, this process of picking elements one by one takes infinitely long! But still, maybe we can make it work. First, we can by induction match up elements \(a_n \in X\) and \(z_n \in Y\) for all \(n \in \mathbb{N}\). Once we’re done with that, we’re out of numbers to use as subscripts, but there’s no reason we can’t keep picking more elements to match. We could, say, next pick elements we call \(b_0 \in X\) and \(y_0 \in Y\), and then \(b_1\) and \(y_1\), and so on. If we still never run out of elements to pick, we can exhaust the \(b_n\)’s and \(y_n\)’s for \(n \in \mathbb{N}\), and then pick \(c_0 \in X\) and \(x_0 \in Y\). We can just keep going like this! We’ll eventually run out of letters, but we could have been more clever in our choices of variable names. For example, instead of \(a_n\), \(b_n\), and \(c_n\), we could write \(a_{0,n}\), \(a_{1,n}\), and \(a_{2,n}\), and thus effectively give ourselves a letter “\(a_m\)” to use for every \(m \in \mathbb{N}\). After we run out of \(m\)’s, then we can go to \(b_{0,0}\) and do the whole thing over again. Or, if we want to be really clever, instead of \(b_{0,0}\) we can say \(a'_{0,0}\), and keep adding primes so we won’t run out of letters. And then...

OK, so maybe it’s not clear that we can keep making this work forever—we certainly would need a better-thought-out notation scheme than what we’ve been using to make sure we never run out of symbols. But the argument is at least plausible, and if you think about it, there really is no reason that you should ever not be able to continue picking more elements, even if you’ve run out of symbols for them.

In fact, this argument does work, if you’re careful enough! This generalized kind of induction is called *transfinite induction*, and we’ll see how it works in the next section.
1.1 Exercises

**Exercise 1.1.** Let $X$ and $Y$ be sets, with $X$ nonempty. Show that $|X| \leq |Y|$ iff there is a surjection from $Y$ to $X$.

**Exercise 1.2.** Say a set $X$ is Dedekind-infinite if there is a proper subset $Y \subset X$ such that $|X| = |Y|$. Show that a set is Dedekind-infinite iff it is infinite. (Hint: Use Proposition 1.10.)

**Exercise 1.3.** Let $X = \mathbb{N} \times \mathbb{N}$, the set of ordered pairs of natural numbers. Show that $X$ is countable. (Hint: Think of $X$ geometrically as the integer lattice points in the first quadrant of the plane, and find a way to “count” these lattice points.)

**Exercise 1.4.** Let $X_0, X_1, X_2, \ldots$ be sets such that $X_n$ is countable for all $n$. Show that $X = \bigcup_{n \in \mathbb{N}} X_n$ is also countable. (Hint: Find an injection from $X$ to $\mathbb{N} \times \mathbb{N}$.)

**Exercise 1.5.** Show that the rational numbers $\mathbb{Q}$ are countable. (Hint: Use Exercise 1.4.)

**Exercise 1.6.** Prove that $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$. (Hint: Use Schröder-Bernstein. For one direction, use that every real number is determined by the set of rational numbers less than it. For the other direction, consider the set of numbers between 0 and 1 whose decimal expansion uses only the digits 0 and 1.)
2 Well-ordering

Let’s now look at how we can do “induction” on more general sets than just the natural numbers. First, we need to see what we’re really trying to generalize—what is induction? We will generalize a slight variant of what may be the more familiar form of induction, sometimes called “strong” induction. This says:

Fact 2.1 (Strong Induction). Let \( P(n) \) be a property of natural numbers \( n \in \mathbb{N} \). Suppose that if \( P(m) \) holds for all \( m < k \), then \( P(k) \) holds. Then \( P(n) \) holds for all \( n \in \mathbb{N} \).

The key feature of \( \mathbb{N} \) used here is the order \(<\). It is thus the order that we will try to generalize.

Definition 2.2. A (strict) total order on a set \( X \) is a relation \(<\) satisfying:

1. Transitivity: if \( x < y \) and \( y < z \), then \( x < z \).
2. Trichotomy: For any \( x, y \in X \), exactly one of \( x < y \), \( x = y \), and \( x > y \) is true.

A well-ordering is a total order such that for any nonempty \( S \subseteq X \), there is \( s_0 \in S \) such that \( s_0 \leq s \) for all \( s \in S \) (we call \( s_0 \) the least element of \( S \)).

Example 2.3. The natural numbers \( \mathbb{N} \) are well-ordered. Indeed, suppose \( S \subseteq \mathbb{N} \) has no least element. Then we can prove by induction that \( n \notin S \) for all \( n \), so \( S \) is empty. Indeed, suppose \( m \notin S \) for all \( m < n \). Then if \( n \) were in \( S \), it would be the least element, so \( n \notin S \).

In fact, this argument is perfectly reversible.

Theorem 2.4 (Induction). Let \( X \) be well-ordered and let \( P(\alpha) \) be a property of elements \( \alpha \in X \). Suppose that if \( P(\beta) \) holds for all \( \beta < \alpha \), then \( P(\alpha) \) holds. Then \( P(\alpha) \) holds for all \( \alpha \in X \).

Proof. Let \( S = \{ \alpha \in S : P(\alpha) \text{ is false} \} \). We want to show \( S \) is empty; suppose it isn’t. Then it has a least element \( \alpha \in S \). But then \( \beta \notin S \) for all \( \beta < \alpha \), which means that \( P(\beta) \) holds for all \( \beta < \alpha \). By hypothesis, this implies \( P(\alpha) \), contradicting \( \alpha \in S \).

Thus we can do induction on any well-ordered set, not just \( \mathbb{N} \). Let’s look at some more examples of well-ordered sets.

Proposition 2.5. Let \( X \) be well-ordered and \( Y \subseteq X \). Then \( Y \) is well-ordered.

Proof. Let \( S \subseteq Y \) be nonempty. Then \( S \subseteq X \), so it has a least element.

Example 2.6. For \( n \in \mathbb{N} \), \([n]\) is well-ordered by Proposition 2.5. This corresponds to doing induction that you know will finish after \( n \) steps.

Example 2.7. Let \( \mathbb{N} + 1 = \mathbb{N} \cup \{\omega\} \), where we declare that \( \omega > n \) for all \( n \in \mathbb{N} \). Then \( \mathbb{N} + 1 \) is well-ordered. Indeed, if \( S \subseteq \mathbb{N} + 1 \) is nonempty, there are two cases. If \( S \cap \mathbb{N} \neq \emptyset \), then the least element of \( S \cap \mathbb{N} \) will be the least element of \( S \). Otherwise, \( S \) must be just \( \{\omega\} \), so \( \omega \) is its least element.

What does it mean to do induction on \( \mathbb{N} + 1 \)? In the notation of the end of Section 1 this is like taking elements \( a_n \) for all \( n \in \mathbb{N} \), and then after that taking one more element \( b_0 \).

Example 2.8. The integers \( \mathbb{Z} \) are not well-ordered with the usual order. Indeed, \( \mathbb{Z} \) itself has no least element! However, let’s put a different order on \( \mathbb{Z} \). Let’s say that

\[
0 < 1 < 2 < 3 < \cdots < -1 < -2 < -3 < \ldots
\]

With this ordering, \( \mathbb{Z} \) is well-ordered, as you will show in Exercise 2.1(a). What is induction on this well-ordered set? It’s picking elements \( a_n \) for all \( n \in \mathbb{N} \), and then picking elements \( b_n \) for all \( n \in \mathbb{N} \) (where \( b_n \) corresponds to \(-(n + 1) \in \mathbb{Z}\)).
Example 2.9. Consider the set $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$ with the lexicographic order:

$$(0,0) < (0,1) < (0,2) < (0,3) < \ldots$$
$$< (1,0) < (1,1) < (1,2) < (1,3) < \ldots$$
$$< (2,0) < (2,1) < (2,2) < (2,3) < \ldots$$
$$< (3,0) < (3,1) < (3,2) < (3,3) < \ldots$$
$$< \ldots$$

With this ordering, $\mathbb{N}^2$ is well-ordered, as you will show in Exercise 2.1(b). What is induction on this well-ordered set? It’s picking elements $a_{0,n}$ for all $n \in \mathbb{N}$, then $a_{1,n}$, and so on through $a_{m,n}$ for all $m \in \mathbb{N}$.

Example 2.10. Let $X$ be well-ordered. Define $X + 1 = X \cup \{\infty\}$, where we declare that $\infty > x$ for all $x \in X$. Then $X + 1$ is well-ordered for the same reason $\mathbb{N} + 1$ was. Doing induction on $X + 1$ is like doing induction on $X$ and then doing one more step at the end. This shows that whenever you’re doing an induction, you can always just add one more step.

OK, now that we’ve seen that we can build well-ordered sets to approximate the argument given at the end of Section 1 to construct an injection between any two sets, let’s formally prove this, assuming we have a well-ordering on the sets.

Definition 2.11. An isomorphism between two ordered sets $X$ and $Y$ is a bijection $f : X \to Y$ such that $\alpha < \beta$ iff $f(\alpha) < f(\beta)$. In this case, we write $X \cong Y$.

Let $X$ be well ordered and $\alpha \in X$. Then the (proper) initial segment of $\alpha$ is $I(\alpha) = \{\beta \in X : \beta < \alpha\}$.

If $Y$ is well-ordered and $Y \cong I(\alpha)$ for some proper initial segment $I(\alpha) \subset X$, we write $Y < X$.

Suppose $S \subset X$ is a proper subset and whenever $\beta \in S$ and $\gamma < \beta$ for $\gamma \in X, \gamma \in S$. Then we can let $\alpha$ be the least element of $X \setminus S$, and it is not hard to see that $S = I(\alpha)$. Thus the proper initial segments of $X$ are exactly the proper subsets that are “closed on the left”. For convenience, we also will consider $X$ itself to be an (improper) initial segment of $X$.

Theorem 2.12. Let $X$ and $Y$ be well-ordered sets. Then exactly one of $X < Y, X \cong Y, \text{ and } X > Y$ is true. Furthermore, the isomorphisms involved are unique.

Proof. The proof of this is essentially the “proof” by induction we gave earlier that for any two sets $X$ and $Y$, either $|X| \leq |Y|$ or $|Y| \leq |X|$. More precisely, we will attempt to define by induction an isomorphism $f$ from $X$ to an initial segment of $Y$.

To define $f(\alpha)$ by induction for $\alpha \in X$, we first assume that $f$ is defined for all $\beta < \alpha$, i.e. $f$ is defined on $I(\alpha)$. Given this, we define $f(\alpha)$ to be the least element of $Y \setminus f(I(\alpha))$, i.e. the least element of $Y$ we haven’t hit yet. Note that $f$ defined this way will preserve order: for all $\beta < \alpha$, $f(\beta)$ was chosen to be the least element of $Y \setminus f(I(\beta))$. Since $f(\alpha) \in Y \setminus f(I(\alpha)) \subset Y \setminus f(I(\beta))$, this implies $f(\beta) < f(\alpha)$.

If this definition by induction works, we in the end get an isomorphism $f$ from $X$ to some subset $f(X) = S \subset Y$. If $S = Y$, we have $X \cong Y$, so we’re done. If $S \subset Y$, note that $S$ is closed on the left. Indeed, at each step of defining $f$, we picked the least possible element, so if $y = f(\alpha) \in S$ and $z < y$, $z$ must be $f(\beta)$ for some $\beta < \alpha$ or else $f(\alpha)$ would have been defined to be $z$ instead of $y$. By the earlier discussion, this implies $S$ is an initial segment of $Y$. Thus $f$ is an isomorphism from $X$ to an initial segment of $Y$, so $X < Y$.

What if the definition by induction doesn’t work? The only way it could fail is if for some $\alpha$, there is no least element of $Y \setminus f(I(\alpha))$ to choose because $f(I(\alpha)) = Y$. In this case, $f$ gives an isomorphism from $I(\alpha)$ to $Y$, so we have $Y < X$.

Finally, our definition of $f$ is the only possible way we could have defined bijections between initial segments of $X$ and initial segments of $Y$. Indeed, suppose that for some $\alpha$, $f(\alpha)$ is not the least element $y$ of

---

2Technically, we have not shown that we can define functions by induction, only that we can prove theorems by induction. It is technical but not too hard to go from proof by induction to definition by induction; talk to me if you want to see the details of this.
$Y \setminus f(I(\alpha))$. Then since $y$ is the least such element, $f(\alpha) > y$. Furthermore, $f(\beta) > y$ for any $\beta > \alpha$, since $f$ must be order-preserving. Thus $f$ can never hit $y$. But then the image of $f$ cannot be an initial segment of $Y$, since it contains $f(\alpha)$ but does not contain the element $y$ which is smaller than $f(\alpha)$. This shows $f$ is unique, and furthermore that we cannot have more than one of $X < Y$, $X \cong Y$, and $Y < X$ hold.

The argument we used here is exactly the argument we tried to make at the end of the last section to show we could compare any two sets: use one set to “count” the other set until one of the two sets runs out. Thus we have proven:

**Corollary 2.13.** Let $X$ and $Y$ be well-ordered sets. Then $|X| \leq |Y|$ or $|Y| \leq |X|$.

**Proof.** If $X < Y$, then $X$ is in bijection with a subset (in fact, an initial segment) of $Y$, which gives an injection from $X$ to $Y$. The result then follows from Theorem 2.12. □

That is, as long as our sets are well-ordered, we can prove that $\leq$ behaves the way it should.

### 2.1 Exercises

**Exercise 2.1.**

(a): Show that with the ordering of Example 2.8, $\mathbb{Z}$ is well-ordered.

(b): Show that with the ordering of Example 2.9, $\mathbb{N}^2$ is well-ordered.

**Exercise 2.2.** Let $X$ and $Y$ be (disjoint) well-ordered sets.

(a): Define $X + Y$ as $X \cup Y$ with the ordering that every element of $X$ is less than every element of $Y$. Show that $X + Y$ is well-ordered.

(b): Define $X \cdot Y$ as $X \times Y$ with the lexicographic ordering: $(a, b) < (c, d)$ iff $a < c$ or $a = c$ and $b < d$. Show that $X \cdot Y$ is well-ordered.

**Exercise 2.3.** Show that a totally ordered set $X$ is well-ordered iff there is no infinite decreasing sequence $x_0 > x_1 > x_2 > \ldots$ of elements of $X$.

**Exercise 2.4.** Prove or find a counterexample: Let $Y$ be a totally ordered set, and let $X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots \subseteq Y$ be subsets of $Y$ that are well-ordered. Then $\bigcup_{n \in \mathbb{N}} X_n$ is well-ordered.

**Exercise 2.5.** Can you find a well-ordering of $\mathbb{R}$, or prove that no such well-ordering exists? (Warning: This is very hard!)
3 More Well-Ordering

At the end yesterday, we showed that the relation \( \leq \) on cardinalities behaves the way it should as long as our sets are well-ordered. But maybe this isn’t so useful: all of the examples of well-ordered sets we’ve written down so far are actually countable! Maybe there aren’t any uncountable well-ordered sets at all, and this is all pretty useless.

But, the wise reader will know that I would not have made you learn all of this if it were so useless!

Example 3.1. Consider the collection \( C \) of all countable well-ordered sets. For any \( X, Y \in C \), we have \( X \leq Y \) or \( Y \leq X \), via a unique isomorphism to an initial segment. We can use these isomorphisms to “glue together” all the sets in \( C \) to form one big well-ordered set, which we will call \( \omega_1 \).

More explicitly, let \( \Omega \) be the disjoint union of all the sets in \( C \); that is, the set of pairs \( (X, x) \) where \( X \in C \) and \( x \in X \) (we can think of this as the union of all the sets \( X \), but we have forced them to be disjoint from each other). Define an equivalence relation on \( \Omega \) as follows. We say \( (X,x) \sim (Y,y) \) and \( (Y,y) \sim (X,x) \) if \( X \leq Y \) and the unique isomorphism from \( X \) to an initial segment of \( Y \) sends \( x \) to \( y \).

By uniqueness of the isomorphisms between well-ordered sets, it is not hard to see that this is indeed an equivalence relation. We define \( \omega_1 \) as the quotient \( \Omega/\sim \).

Note that \( \omega_1 \) is well-ordered, where we order it by “gluing together” the orders of all the sets of \( C \). Indeed, if \( A \subseteq \omega_1 \) is nonempty, let \( [(X,x)] \in A \) and let \( x_0 \in X \) be minimal such that \( [(X,x_0)] \in A \). Then \( [(X,x_0)] \) will be the least element of \( A \).

Also, each \( X \) injects into \( \omega_1 \) by sending \( x \) to \( [(X,x)] \), and the image is an initial segment. Thus \( X \leq \omega_1 \) for all \( X \in C \). But for any countable well-ordered set \( X \), there is another countable well-ordered set \( X + 1 \) which is strictly longer: simply let \( X + 1 \) be \( X \) together with one more point added at the end. Thus for any countable \( X \),

\[
X < X + 1 \leq \omega_1.
\]

Now suppose \( \omega_1 \) was countable. Then the argument above would give \( \omega_1 < \omega_1 \), which is impossible. Thus \( \omega_1 \) is an uncountable well-ordered set.

This construction is rather sneaky: we take all the countable well-ordered sets, and glue them all together to get a new well-ordered set \( \omega_1 \) that contains all the countable well-ordered sets as initial segments. Since for any countable well-ordered set, you can build a larger one by adding one more element at the end, there can be no single countable well-ordered set that is longer than all others. This implies that \( \omega_1 \) cannot be a countable well-ordered set, and hence is uncountable.

This argument works much more generally; there is nothing special about the collection of all countable well-ordered sets.

Lemma 3.2. Let \( C \) be any set of well-ordered sets. Then there is a well-ordered set \( A \) such that for all \( X \in C, X \leq A \).

Proof. Define \( A \) to be all the sets in \( C \) “glued together”, exactly as in Example 3.1. Just like in Example 3.1, every element of \( C \) is naturally isomorphic to an initial segment of \( A \).

We will sometimes write \( \text{sup} \, C \) for this well-ordered set \( A \).

Just as we used this argument to find an uncountable well-ordered set, we can also use it to find arbitrarily large well-ordered sets.

Lemma 3.3. Let \( S \) be any set. Then there exists a well-ordered set \( A \) such that \( |A| \not\leq |S| \) (i.e., there is no injection from \( A \) to \( S \)).

Proof. Let \( C \) be the set of all well-ordered sets of the form \( X + 1 \) (as in Example 2.10) for which there exists an injection from \( X \) to \( S \). Let \( A = \text{sup} \, C \) be as in Lemma 3.2, i.e. such that \( X + 1 \leq A \) for all \( X \) such that there exists an injection \( X \to S \).

Now suppose there exists an injection \( A \to S \). Then \( A + 1 \) would be an element of the set \( C \), and so we would have \( A + 1 \leq A \). But clearly \( A < A + 1 \), so this is a contradiction.
The well-ordered set $A$ constructed in Lemma 3.3 is sometimes called the Hartogs set of $S$. Roughly, Lemma 3.3 says that there is no set $S$ which is bigger than all well-ordered sets. Since by Proposition ??, any subset of a well-ordered set is well-ordered, this suggests that it should be possible to well-order any set. Indeed, if we use the Hartogs set of $S$ to try to “count” $S$, we can obtain a well-ordering of $S$.

**Theorem 3.4** (Well-Ordering Principle). Any set can be well-ordered.

Proof. Let $S$ be any set, and let $A$ be its Hartogs set. We will attempt to define an injection $f : A \to X$ by induction.

For $\alpha \in A$, suppose we have defined $f(\beta)$ for all $\beta < \alpha$. Let $s$ be some element of $S$ which is distinct from $f(\beta)$ for all $\beta < \alpha$. We then define $f(\alpha) = s$.

This would define an injection from $A$ to $S$, but by definition of $A$, there is no injection from $A$ to $S$. Thus there must be something wrong with the construction above. How could this happen? Well, it might be that for some $\alpha$, there is no $s \in S$ which is distinct from $f(\beta)$ for all $\beta < \alpha$. That is, $f : I(\alpha) \to S$ is surjective. Since $f$ is constructed to be injective, this means $f$ gives a bijection from $I(\alpha)$ to $S$. Since $I(\alpha)$ is well-ordered, we can use this to well-order $S$.

Since any two well-ordered sets are comparable in size, we conclude that any two sets are comparable in size, finally completing our discussion from the first section.

**Corollary 3.5.** Let $X$ and $Y$ be sets. Then either $|X| \leq |Y|$ or $|Y| \leq |X|$.

### 3.1 Exercises

**Exercise 3.1.** Let $C$ be the collection of all finite well-ordered sets. Prove that $\sup C$, the well-ordered set obtained by gluing the sets in $C$ together, is isomorphic to $\mathbb{N}$.

**Exercise 3.2.** Let $C$ be a collection of well-ordered sets and let $A = \sup C$ be obtained by gluing the sets in $C$ together. Let $\alpha \in A$. Show that for some $X \in C$, $I(\alpha) < X$.

**Exercise 3.3.** Let $\omega_1$ be the well-ordered set of Example 3.1. Recall that $\omega_1$ is built by “gluing together” all the countable well-ordered sets, and has every countable well-ordered set as an initial segment.

(a) Show that $\omega_1$ is the shortest possible uncountable well-ordered set. That is, any proper initial segment of $\omega_1$ is countable.

(b) Let $\alpha_0 < \alpha_1 < \alpha_2 < \ldots$ be a countable increasing sequence of elements of $\omega_1$. Show that there exists $\beta \in \omega_1$ such that $\beta > \alpha_n$ for all $n$. (Hint: Use Exercise 1.4, with $X_n = I(\alpha_n)$.)

**Exercise 3.4.** Suppose we have finitely many symbols we know how to write. Let $W$ be the set of all finite strings of these symbols (“words”). Show that $|W| = |\mathbb{N}|$. (Hint: List them by length.)

Exercise 3.4 implies that there is no possible way to give “names” to every element of $\omega_1$! In this sense, it is impossible to “explicitly write down” what $\omega_1$ looks like.

The following is an axiom of set theory (we will talk more about axioms of set theory next week):

**Axiom of Choice.** Let $B$ be a set of nonempty sets. Then there is a “choice function” $\varphi$ defined on $B$ that picks an element of each set in $B$: for all $B \in B$, $\varphi(B) \in B$.

This axiom was implicitly used in the proof of the Well-Ordering Principle, because at each stage of the induction on $A$, we had to choose an element $s \in S$ which was not yet in the image of $f$. Explicitly, we could have let $B$ be the collection of all nonempty subsets of $S$, and for $\varphi$ a choice function on $B$, we inductively define $f(\alpha) = \varphi(S \setminus f(I(\alpha)))$.

In fact, the Axiom of Choice is equivalent to the Well-Ordering Principle, and also to Corollary 3.5.

**Exercise 3.5.**
(a) Assuming the Well-Ordering Principle, prove the Axiom of Choice. (Hint: If a set is well-ordered, it’s easy to choose elements of subsets of it.)

(b) Assuming Corollary 3.5, prove the Well-Ordering Principle. (Hint: Use Lemma 3.3 and Proposition 2.5.)

**Exercise 3.6.**

(a) Let $C$ be any collection of well-ordered sets. Show that one of them must be the shortest: there exists $X \in C$ such that $X \leq Y$ for all $Y \in C$. (Hint: Let $A$ be a well-ordered set such that $X < A$ for all $X \in C$, and write $X \cong I(\alpha)$ for some $\alpha \in A$.)

(b) Let $C$ be any collection of sets. Show that there is a set $S \in C$ such that $|S| \leq |T|$ for all $T \in B$. 
4 Ordinals and Cardinals

We have now mostly finished our general discussion of well-ordering and how it informs our understanding of sets. There's one final observation that makes thinking about well-ordered sets easier.

**Definition 4.1.** Let \( C \) be the collection of *all* well-ordered sets. We can glue these sets together to obtain a well-ordered class \( \text{Ord} \) which contains all well-ordered sets at once, which we call the *ordinal numbers*.

Since the ordinals contain all well-ordered sets as initial segments, instead of having to pick a well-ordered set that is long enough for our purposes, we can always just use the ordinals. That is, there are enough ordinals to “count” as high as we might possibly want.\(^3\) Any well-ordered set is isomorphic to \( I(\alpha) \) for some unique \( \alpha \in \text{Ord} \), which we call the *length* of the well-ordered set.

Let’s now look at what the ordinals actually look like. They themselves are well-ordered by Theorem 2.11, so we can write them down in order. With a bit of thought, we can see that the following are the first few ordinals: Let’s now look at what the ordinals actually look like. From the examples of well-ordered sets we have so far, we can see that they start like this:

\[
\begin{align*}
0 \\
1 \\
2 \\
3 \\
\ldots \\
\omega \\
\omega + 1 \\
\omega + 2 \\
\ldots \\
2\omega \\
2\omega + 1 \\
\ldots \\
3\omega \\
\ldots \\
\omega^2 \\
\ldots 
\end{align*}
\]

In general, given an ordinal \( \alpha \), there is a least ordinal which is greater than \( \alpha \) (since the ordinals are well-ordered), which we write as \( \alpha + 1 \). This is not unrelated to our previous notation of \( X + 1 \) for adding one element to the end of a well-ordered set: if \( X \) is isomorphic to \( I(\alpha) \), the set of ordinals less than \( \alpha \), then \( X + 1 \) is isomorphic to \( I(\alpha + 1) \).

Our other construction of gluing together well-ordered sets to get a bigger well-ordered sets also has an analogue for ordinals.

**Definition 4.2.** If \( C \) is a set of ordinals, we define \( \sup C \) to be the least ordinal \( \beta \) such that \( \alpha \geq \beta \) for all \( \alpha \in C \) (the least upper bound of \( C \)). In this case, \( I(\beta) \) is (isomorphic to) the well-ordered set obtained by gluing together the sets \( I(\alpha) \) for \( \alpha \in C \) (you will verify this in Exercise ??(b)).

**Definition 4.3.** Let \( \alpha \) be an ordinal. If \( \alpha = \beta + 1 \) for some ordinal \( \beta \), we say \( \alpha \) is a *successor ordinal*. If \( \alpha \) is not a successor, we say that \( \alpha \) is a *limit ordinal*.

\(^3\)If you are uneasy about gluing together all well-ordered sets at once, you can let “the ordinals” refer to “a well-ordered set
In some sense “most” ordinals are successors. Indeed, in the list above, the only limit ordinals are 0, ω, and 2ω, 3ω, . . . , ω². These examples (except perhaps for 0) illustrate the name “limit ordinal”: we can think of them as the limit of infinitely many smaller ordinals.

As a final remark, so far we’ve always been doing induction on well-ordered sets as “strong” induction, where our induction hypothesis is that we have already shown what we want for all of them as the limit of infinitely many smaller ordinals.

Definition 4.4. If α is an ordinal, the cardinality |α| of α refers to |I(α)|, the cardinality of the set of ordinals less than α (or, the cardinality of a well-ordered set of length α).

A cardinal number is an ordinal α such that |κ| > |α| for all α < κ. That is, a cardinal is an ordinal that is minimal among ordinals of its cardinality.

From now on, the cardinality |S| of a set S will refer to the unique cardinal number κ such that |S| = |κ| (in our old notation). Equivalently, |S| is the minimal possible length of a well-ordering of S.

Example 4.5. The first few cardinals are 0, 1, 2, 3, . . . , ω, which are the same as the first few ordinals. However, after this the situation rather changes. For example, all the ordinals between ω and ω² are countable and thus not cardinals (since they’re the same size as ω³). Indeed, ω² is naturally in bijection with N × N, which is countable by Exercise 1.3 (and we will also give a proof tomorrow).

The next cardinal after ω will clearly be the first uncountable cardinal. This is the length of the well-ordered set ω₁ we constructed earlier; we will abuse notation and also call this ordinal ω₁.

Definition 4.6. Let X and Y be sets. Then the disjoint union is the set

\[ X \sqcup Y = \{(0, x) : x \in X\} \cup \{(1, y) : y \in Y\} \]

and their (Cartesian) product is the set

\[ X \times Y = \{(x, y) : x \in X, y \in Y\}. \]

Let κ and λ be cardinals. Then we define

\[ \kappa + \lambda = |I(\kappa) \sqcup I(\lambda)| \]

that is large enough for whatever we’re trying to do”.

4Actually, the base case is really a special case of the limit case, since 0 is a limit ordinal.
Some rules of arithmetic are easy to prove from these definitions, and are familiar from ordinary finite arithmetic.

**Proposition 4.7.** Let \( \kappa, \lambda, \) and \( \theta \) be cardinals. Then:

1. \( \lambda + \kappa = \kappa + \lambda \).
2. \( \lambda + (\kappa + \theta) = (\lambda + \kappa) + \theta \).
3. \( \lambda \kappa = \kappa \lambda \).
4. \( \lambda(\kappa \theta) = (\lambda \kappa)\theta \).
5. \( \lambda(\kappa + \theta) = \lambda \kappa + \lambda \theta \).

*Proof.* These identities all come from simple bijections and are mostly left as an exercise. As an example, we prove (3). By definition,
\[
\lambda \kappa = |\mathcal{I}(\lambda) \times \mathcal{I}(\kappa)|
\]
and
\[
\kappa \lambda = |\mathcal{I}(\kappa) \times \mathcal{I}(\lambda)|.
\]
But \( |\mathcal{I}(\lambda) \times \mathcal{I}(\kappa)| = |\mathcal{I}(\kappa) \times \mathcal{I}(\lambda)| \), since there is a bijection between these two sets which sends a pair \((\alpha, \beta)\) to the pair \((\beta, \alpha)\). Hence \( \lambda \kappa = \kappa \lambda \). \( \square \)

### 4.1 Exercises

**Exercise 4.1.** Let \( \alpha \) be an ordinal. Show that \( \alpha = \sup I(\alpha) \) iff \( \alpha \) is a limit ordinal.

**Definition 4.8.** Let \( \alpha \) be an ordinal and \( n \in \mathbb{N} \). We define \( \alpha + 0 = \alpha \), and we have already defined \( \alpha + 1 \) as the least ordinal greater than \( \alpha \). If \( n > 1 \), we inductively define
\[
\alpha + n = (\alpha + (n - 1)) + 1,
\]
so \( \alpha + n \) is the \( n \)th smallest ordinal that is greater than \( \alpha \). We define
\[
\alpha + \omega = \sup \{\alpha + n : n \in \mathbb{N}\}.
\]

**Exercise 4.2.** Show that any ordinal \( \alpha \) can be written uniquely in the form \( \beta + n \), where \( \beta \) is a limit ordinal and \( n \in \mathbb{N} \). (Hint: Use induction on \( \alpha \), splitting into successor and limit cases.)

**Exercise 4.3.**

(a) Show that an ordinal \( \beta \) is a limit ordinal iff for all \( \alpha < \beta \), \( \alpha + n < \beta \) for all \( n \in \mathbb{N} \).

(b) Show that for any \( \alpha \), \( \alpha + \omega \) is a limit ordinal.

(c) Suppose \( \alpha < \beta \) are ordinals and suppose there are infinitely many ordinals between \( \alpha \) and \( \beta \). Show that there is a limit ordinal \( \gamma \) such that \( \alpha < \gamma \leq \beta \).

**Exercise 4.4.** Prove Proposition 4.7

**Exercise 4.5.**

(a) Show that \( \aleph_0 + \aleph_0 = \aleph_0 \).

(b) Show that \( \aleph_0 + \aleph_1 = \aleph_1 \).

**Exercise 4.6.** Let \((\alpha_n)_{n \in \mathbb{N}}\) be a countable sequence of ordinals such that \( \alpha_n < \omega_1 \) for all \( n \). Show that \( \sup \{\alpha_n\}_{n \in \mathbb{N}} < \omega_1 \). That is, \( \omega_1 \) cannot be reached by taking a countable limit of smaller ordinals.
We'll now learn how to actually compute sums and products of infinite cardinals. Let's start with a simple example.

**Proposition 5.1.** \( \aleph_0 \cdot \aleph_0 = \aleph_0 \).

**Proof.** You may have seen a proof of this before (or a proof of the closely related fact that \( |\mathbb{Q}| = \aleph_0 \)), but our proof will be slightly different. We will construct a bijection from \( \mathbb{N} \) to \( \mathbb{N} \times \mathbb{N} \). It suffices to show we can write down the elements of \( \mathbb{N} \times \mathbb{N} \) in an (infinite) list (i.e., “count” the elements). We do this as follows:

\[
(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (2, 1), (0, 2), (1, 2), (2, 2), (3, 0), \ldots
\]

That is, one by one for each \( n \in \mathbb{N} \), we list

\[
(n, 0), (n, 1), \ldots, (n, n-1), (0, n), (1, n), \ldots, (n-1, n), (n, n).
\]

Note that this proof is essentially by induction. For each \( n \) by induction, we are adding

\[
(n, 0), (n, 1), \ldots, (n, n), (0, n), (1, n), \ldots, (n-1, n), (n, n)
\]

to the list. Note that after doing the \((n - 1)\)st stage of this, we have listed exactly the pairs \((m, k)\) with \( m, k < n \). There are exactly \( n^2 \) of these. Thus you could really say that what we’re doing is writing \( \mathbb{N} \times \mathbb{N} \) as the limit (i.e., union) of the subsets \( I(n) \times I(n) \). Since \( n^2 < \aleph_0 \) for all \( n < \aleph_0 \), they all “fit” inside \( \mathbb{N} \), so we get that \( |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}| \).

This discussion suggests generalizing this argument to all infinite cardinals by induction. First, we need a little observation.

**Lemma 5.2.** Every infinite cardinal is a limit ordinal.

**Proof.** Let \( \alpha = \beta + 1 \) be an infinite successor ordinal. Then \( \omega \leq \beta \). We can thus define a bijection \( f : \alpha \to \beta \) as follows. For \( n < \omega \), \( f(n) = n + 1 \). For \( \omega \leq \gamma < \beta \), \( f(\gamma) = \gamma \). Finally, \( f(\beta) = 0 \).

Thus \( |\alpha| = |\beta| \) but \( \beta < \alpha \), so \( \alpha \) is not a cardinal. \( \square \)

**Theorem 5.3.** Let \( \kappa \) be an infinite cardinal. Then \( \kappa^2 = \kappa \).

**Proof.** We prove the theorem by induction on \( \alpha \) for \( \kappa = \aleph_\alpha \). We have already proven the result for \( \alpha = 0 \).

Now suppose \( \kappa > \aleph_0 \) and we know that for every infinite \( \lambda < \kappa \), \( \lambda^2 = \lambda \). This implies that for every \( \lambda < \kappa \), \( \lambda^2 < \kappa \). We will use this fact to construct a well-ordering of \( I(\kappa) \times I(\kappa) \) which has length at most \( \kappa \).

This will imply \( |I(\kappa) \times I(\kappa)| \leq \kappa \), and the reverse inequality is easy.

So here’s our well-ordering. Let \( (\alpha, \beta) \) and \( (\gamma, \delta) \) be in \( I(\kappa) \times I(\kappa) \). We say \( (\alpha, \beta) < (\gamma, \delta) \) if:

1. \( \max(\alpha, \beta) > \max(\gamma, \delta) \), or
2. \( \max(\alpha, \beta) = \max(\gamma, \delta) \) and \( \alpha < \gamma \), or
3. \( \max(\alpha, \beta) = \max(\gamma, \delta) \), \( \alpha = \gamma \), and \( \beta < \delta \).

A bit less precisely, this definition means that we order \( I(\kappa) \times I(\kappa) \) by one-by-one for each \( \eta < \kappa \) tacking on

\[
(\eta, 0), (\eta, 1), \ldots, (\eta, \alpha), \ldots, (0, \eta), (1, \eta), \ldots, (\alpha, \eta), \ldots, (\eta, \eta)
\]

to the end. Here \( \alpha \) ranges over all \( \alpha < \eta \).

It is not hard to see that this is indeed a total order. To see that it is a well-ordering, let \( S \subseteq I(\kappa) \times I(\kappa) \) be nonempty. Then there is some least value \( \eta \) of \( \max(\alpha, \beta) \) for \( (\alpha, \beta) \in S \); let \( S' = \{ (\alpha, \beta) \in S : \max(\alpha, \beta) = \eta \} \); a least element of \( S' \) will be a least element of \( S \). Now \( S' \) is a subset of the ordered set

\[
T = \{ (\eta, 0), (\eta, 1), \ldots, (\eta, \alpha), \ldots, (0, \eta), (1, \eta), \ldots, (\alpha, \eta), \ldots, (\eta, \eta) \},
\]
ordered as written. If \( S' \cap \{(\eta, 0), (\eta, 1), \ldots, (\eta, \alpha), \ldots \} \) is nonempty, \((\eta, \alpha)\) will be the least element of \( S' \) where \( \alpha \) is the least \( \alpha \) such that \((\eta, \alpha) \in S'\). Otherwise, \( S' \subseteq \{(0, \eta), (1, \eta), \ldots, (\alpha, \eta), \ldots, (\eta, \eta)\} \), which is clearly isomorphic to \( \eta + 1 = \{0, 1, \ldots, \eta\} \) and hence well-ordered, so \( S' \) has a least element.

Thus \( I(\kappa) \times I(\kappa) \) is well-ordered. Furthermore, for any \( \eta < \kappa \), the initial segment of \((\eta, 0)\) under this ordering is exactly the subset \( I(\eta) \times I(\eta) \). For any \((\alpha, \beta) \in I(\kappa) \times I(\kappa)\), if \( \eta = \max(\alpha, \beta) + 1 \) then \((\alpha, \beta) < (\eta, 0)\) (here we use that \( \kappa \) is a limit ordinal to be sure that \( \eta < \kappa \)). Thus \( I(\kappa) \times I(\kappa) \) is the union of the initial segments \( I((\eta, 0)) = I(\eta) \times I(\eta) \) for \( \eta < \kappa \).

By our remarks at the beginning of the proof, \( |I(\eta) \times I(\eta)| < \kappa \) for all \( \eta < \kappa \). Hence \( \text{length}(I(\eta) \times I(\eta)) < \kappa \).

This means that every initial segment of \( I(\kappa) \times I(\kappa) \) has length less than \( \kappa \). If \( \text{length}(I(\kappa) \times I(\kappa)) \) were greater than \( \kappa \), then it would contain a proper initial segment of length \( \kappa \) (defined by the element corresponding to \( \kappa \) in the ordering). Thus \( \text{length}(I(\kappa) \times I(\kappa)) \leq \kappa \), so \( |I(\kappa) \times I(\kappa)| \leq \kappa \). The desired result follows. \( \square \)

Most of cardinal arithmetic follows from this result.

**Corollary 5.4.** Let \( \kappa \) and \( \lambda \) be nonzero cardinals, at least one of which is infinite. Then \( \kappa \cdot \lambda = \kappa \lambda = \max(\kappa, \lambda) \).

**Proof.** Assume without loss of generality that \( \kappa \geq \lambda \). We also assume \( \lambda \geq 2 \); the case \( \lambda = 1 \) is easily handled by a similar argument. Then it is easy to see that

\[
\kappa \leq \kappa + \lambda \leq \kappa + \kappa \leq \kappa \cdot 2 \leq \kappa \lambda \leq \kappa^2 = \kappa.
\]

See also Figure 2. \( \square \)

So that’s all there is to addition and multiplication of infinite cardinals, at least if we’re only adding or multiplying together finitely many things (the definitions of these operations are analogous to the binary operations). Exponentiation is more complicated.

**Definition 5.5.** Let \( X \) and \( Y \) be sets. Then we write \( Y^X \) for the set of all functions from \( Y \) to \( X \).

Let \( \kappa \) and \( \lambda \) be cardinals. Then we define

\[
\kappa^\lambda = |I(\lambda) I(\kappa)|.
\]

As with addition and multiplication, we have some familiar properties that are easy to prove.

**Proposition 5.6.** Let \( \kappa, \lambda, \) and \( \theta \) be cardinals. Then:

1. \( \lambda^{\kappa+\theta} = \lambda^\kappa \lambda^\theta \).
2. \( (\lambda \kappa)^\theta = \lambda^\theta \kappa^\theta \).
3. \( (\lambda^\kappa)^\theta = \lambda^{\kappa \theta} \).

**Proof.** We prove (1) and leave the rest as an exercise. The left-hand side is the number of functions from \( I(\kappa) \sqcup I(\theta) \) to \( I(\lambda) \). But giving such a function is the same thing as first defining a function \( I(\kappa) \rightarrow I(\lambda) \) and then defining another function \( I(\theta) \rightarrow I(\lambda) \). This is the same as an element of the cartesian product \( I(\kappa) I(\lambda) \times I(\theta) I(\lambda) \), and the cardinality of this is \( \lambda^\kappa \lambda^\theta \). \( \square \)

**Remark 5.7.** Let \( \kappa \) be a cardinal and \( \mathcal{P}(\kappa) = \{S : S \subseteq \kappa\} \) be the powerset of \( \kappa \). Then \( |\mathcal{P}(\kappa)| = 2^\kappa \). Indeed, given \( S \subseteq \kappa \), we have a function \( 1_S : \kappa \rightarrow 2 \) defined by \( 1_S(\alpha) = 1 \) if \( \alpha \in S \) and \( 1_S(\alpha) = 0 \) otherwise. We call this the characteristic function of \( S \). It is easy to see that this gives a bijection between \( \mathcal{P}(\kappa) \) and \( ^\kappa 2 \).

In Section 1, we showed that for any set \( X \), \( |X| < |\mathcal{P}(X)| \). This can now be restated in the following way.

**Theorem 5.8** (Cantor). For any \( \kappa, \kappa < 2^\kappa \).

**Corollary 5.9.** Let \( \kappa \) be an infinite cardinal and \( 2 \leq \lambda \leq 2^\kappa \). Then \( \lambda^\kappa = 2^\kappa \).
Figure 2: Arithmetic of infinite cardinals (by Bill Watterson)
Proof. We have
\[ 2^\kappa \leq \lambda^\kappa \leq (2^\kappa)^\kappa = 2^{\kappa^2} = 2^\kappa. \]

We now have two different ways to get from a cardinal \( \kappa \) to a bigger cardinal: \( \kappa^+ \) and \( 2^\kappa \). Are these the same?

**Generalized Continuum Hypothesis (GCH).** *For any infinite cardinal \( \kappa \), \( 2^\kappa = \kappa^+ \).*

As the name suggests, this statement is *not* a theorem. In fact, GCH cannot be proved or disproved from the usual axioms of set theory (ZFC)! We say GCH is *independent* from ZFC.

In the case that \( \kappa = \aleph_0 \), GCH is better-known and has a shorter name.

**Continuum Hypothesis (CH).** \( 2^{\aleph_0} = \aleph_1 \).

If you take Susan’s class on the Continuum Hypothesis in weeks 3 and 4, you will learn how to prove that CH is independent from ZFC.

The fact that a statement as simple as \( 2^{\aleph_0} = \aleph_1 \) cannot be answered is very representative of the flavor of the combinatorics of uncountable cardinals. A surprisingly large number of basic questions about infinite combinatorics cannot be answered in ZFC.

### 5.1 Exercises

**Exercise 5.1.** Prove Proposition 5.6

**Exercise 5.2.** Show that \( \aleph_1^{\aleph_0} = 2^{\aleph_0} \).

**Exercise 5.3.** Let \( X \) be an infinite set and let \( S \) be the set of all finite subsets of \( X \). Show that \( |X| = |S| \).

(Hint: Show that \( |S| \leq \aleph_0 \cdot |X| \).)

**Exercise 5.4.**

(a) Let \( \kappa = \lambda^+ \) be a successor cardinal, and \( S \subseteq I(\kappa) \) have cardinality less than \( \kappa \). Show that \( \sup S < \kappa \).

(Hint: Use \( \lambda^2 = \lambda \).

(b) Find a counterexample to (a) for \( \kappa = \aleph_\omega \).

**Exercise 5.5.** Show that Theorem 5.3 implies the Axiom of Choice. Without Choice, the statement of Theorem 5.3 should be interpreted as \( |Y \times Y| = |Y| \) for all infinite sets \( Y \). (Hint: Let \( X \) be a set and \( A \) be its Hartogs set, and let \( Y = X \sqcup A \). Now use Theorem 5.3 to show \( |X| \leq |A| \).)
6 Regular and singular cardinals

In this section we will do a little bit more cardinal arithmetic and get a bit more information about cardinals that will be useful to us later. We begin with a new inequality.

**Theorem 6.1.** \( \aleph_0^{\aleph_0} > \aleph_\omega. \)

*Proof.* Let \( A = I(\aleph_\omega) \) be a set of size \( \aleph_\omega \) and \( B = {}^N A \) be a set of size \( \aleph_0^{\aleph_0} \). It suffices to show there is no bijection \( f : A \to B \).

Suppose \( f : A \to B \) is any function; we will show \( f \) is not surjective. Note that for any \( \alpha \in A \), \( f(\alpha) \) is a function \( f : \mathbb{N} \to A \), so we can write \( f(\alpha)(n) \in A \) for any \( n \in \mathbb{N} \). Note that

\[
A = \bigcup_{n=0}^{\infty} A_n,
\]

where \( A_n = I(\aleph_n) \) is a set of size \( \aleph_n \). For any \( n \in \mathbb{N} \), let

\[
S_n = \{ f(\alpha)(n) : \alpha \in A_n \} \subseteq A.
\]

For each \( n \), we have

\[
|S_n| \leq |A_n| = \aleph_n < |A|.
\]

Thus no \( S_n \) can be all of \( A \). We define a new function \( \varphi : \mathbb{N} \to A \) by picking some \( \varphi(n) \in A \setminus S_n \) for each \( n \).

We claim that \( \varphi \in B \) but \( \varphi \neq f(\alpha) \) for all \( \alpha \in A \), and hence \( f \) is not surjective. Indeed, for any \( \alpha \in A \), \( \alpha \in A_n \) for some \( n \). But then \( f(\alpha)(n) \in S_n \) but \( \varphi(n) \notin S_n \), so \( f(\alpha) \neq \varphi \).

The key thing that made Theorem 6.1 was the fact that \( \aleph_\omega \) could be written as the "limit" of only \( \aleph_0 \) many smaller cardinals. This motivates the following definition.

**Definition 6.2.** Let \( \kappa \) be an infinite cardinal. The cofinality of \( \kappa \) is the smallest cardinal \( \text{cf}(\kappa) = \lambda \) such that \( I(\kappa) \) can be written as a union of \( \lambda \) subsets, all of which have cardinality less than \( \kappa \).

If \( \text{cf}(\kappa) = \kappa \), we say \( \kappa \) is regular. Equivalently, \( \kappa \) is regular if a set of size \( \kappa \) cannot be written as a union of fewer than \( \kappa \) sets of size less than \( \kappa \). If \( \text{cf}(\kappa) < \kappa \), we say \( \kappa \) is singular.

**Example 6.3.** \( \aleph_0 \) is regular, since you can’t get an infinite set as the union of finitely many finite sets.

**Example 6.4.** \( \aleph_\omega \) is singular, since \( I(\aleph_\omega) = \bigcup_n I(\aleph_n) \) is the union of only \( \aleph_0 \) sets of smaller cardinality. Indeed, \( \text{cf}(\aleph_\omega) = \aleph_0 \).

**Example 6.5.** Similarly, \( \aleph_{2\omega} \) is also singular with cofinality \( \aleph_0 \), since \( I(\aleph_{2\omega}) = \bigcup_n I(\aleph_{\omega+n}) \).

**Example 6.6.** Any successor ordinal \( \kappa = \lambda^+ \) is regular. Indeed, any set of size less than \( \kappa \) has size at most \( \lambda \). Thus a union of fewer than \( \kappa \) sets of size less than \( \kappa \) is a union of at most \( \lambda \) sets of size at most \( \lambda \), and thus has cardinality at most \( \lambda^2 = \lambda < \kappa \).

**Example 6.7.** A cardinal that is both regular and limit is called weakly inaccessible. We will see later in the week that (assuming GCH) it is impossible to prove that weakly inaccessible cardinals exist!

The proof of Theorem 6.1 generalizes verbatim to prove the following more general theorem.

**Theorem 6.8.** Let \( \kappa \) be an infinite cardinal. Then \( \kappa^{\text{cf}(\kappa)} > \kappa \).

*Proof.* Let \( A = I(\kappa) \), \( C = I(\text{cf}(\kappa)) \), and \( B = {}^C A \). Write \( A = \bigcup_{c \in C} A_c \), where each \( A_c \) has size less than \( \kappa \). We can then copy the proof of Theorem 6.1 exactly, with \( C \) in place of \( \mathbb{N} \) and \( A_c \) in place of \( A_n \).

This also has the following interesting consequence.

**Corollary 6.9.** Let \( \kappa \) be an infinite cardinal. Then \( \text{cf}(2^\kappa) > \kappa \).
Proof. Suppose that $\text{cf}(2^\kappa) \leq \kappa$. Then we would have:

$$2^\kappa < (2^\kappa)^{\text{cf}(2^\kappa)} \leq (2^\kappa)^\kappa = 2^{\kappa^2} = 2^\kappa,$$

a contradiction. Hence $\text{cf}(2^\kappa) > \kappa$.

Corollary 6.10. $2^{\aleph_0} \neq \aleph_\omega$.

In particular, $2^{\aleph_0}$ has uncountable cofinality. Thus while we can’t determine the value of $2^{\aleph_0}$, we can say that it can’t be exactly $\aleph_\omega$, or any other limit cardinal of cofinality $\aleph_0$. In fact, this is all we can say about the value of $2^{\aleph_0}$.

Fact 6.11. It is consistent with the axioms of set theory for $2^{\aleph_0}$ to be any cardinal having uncountable cofinality.

Susan’s Continuum Hypothesis class will give a more precise statement as well of a proof of this fact, or perhaps a somewhat weaker version of it.

In fact, it is more generally true that virtually every inequality that can be proven (in ZFC) about cardinal arithmetic follows from Theorem 6.8 (or its generalization Exercise 6.3). The only exception is that there are some things that can be proven about exponentiation of singular cardinals using a sophisticated technique (which I don’t know!) called $\text{pcf theory}$. For example, it is possible to prove that $\aleph_\omega^{\aleph_0} \leq 2^{\aleph_0} + \aleph_\omega$. I have no idea how to prove this.

6.1 Exercises

Exercise 6.1. Show that for any infinite cardinals $\kappa$ and $\lambda$, $\text{cf}(\kappa^\lambda) > \lambda$.

Exercise 6.2. Show that for any infinite cardinal $\kappa$, $\text{cf}(\text{cf}(\kappa)) = \text{cf}(\kappa)$. That is, $\text{cf}(\kappa)$ is regular.

Exercise 6.3 (König’s Theorem). Let $\lambda_\alpha$ and $\kappa_\alpha$ be cardinals such that $\lambda_\alpha < \kappa_\alpha$ for all $\alpha \in I$ (for some index set $I$). Prove that

$$\sum_{\alpha \in I} \lambda_\alpha < \prod_{\alpha \in I} \kappa_\alpha.$$

(Hint: Imitate the proof of Theorem 6.1, with $I$ in place of $\mathbb{N}$, $\lambda_\alpha$ in place of $\aleph_n$, and $\kappa_\alpha$ in place of $\aleph_\omega$.)

Exercise 6.4.

(a) Use Exercise 6.3 to prove Theorem 6.8.

(b) Use Exercise 6.3 to prove $\kappa < 2^\kappa$ for all $\kappa$. (Hint: Let $\lambda_\alpha = 1$ and $\kappa_\alpha = 2$.)

Exercise 6.5. Let $\lambda$ and $\kappa$ be cardinals with $\lambda < \text{cf}(\kappa)$. Show that any function $f : I(\lambda) \to I(\kappa)$ is bounded: that is, there exists $\beta \in I(\kappa)$ such that $f(\alpha) < \beta$ for all $\beta \in I(\lambda)$.

Exercise 6.6. Let $\kappa$ be infinite and $\lambda > 0$. Show that

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

(Hint: If $\lambda \geq \kappa^+$, this is easy. If $\lambda \leq \kappa$, note that $\text{cf}(\kappa^+) = \kappa^+ > \lambda$. Now use Exercise 6.5 to show $(\kappa^+)\lambda \leq \kappa^+ \cdot \kappa^\lambda$.)

Exercise 6.7. Let $\kappa^{< \lambda} = \sup_{\theta < \lambda} \kappa^\theta$. Show that for $\kappa$ infinite, $2^\kappa = (2^{< \kappa})^{\text{cf}(\kappa)}$.

Exercise 6.8. Say a cardinal $\kappa$ is strong limit if $\lambda < \kappa$ implies $2^\lambda < \kappa$.

(a): Show that any strong limit cardinal is a limit cardinal.

(b): Show that for any $\lambda$, there is a strong limit cardinal $\kappa$ such that $\kappa > \lambda$.

(c): Show that if $\kappa$ is strong limit, then $2^\kappa = \kappa^{\text{cf}(\kappa)}$. 

19
7 The language and axioms of set theory

As has been occasionally alluded to earlier, there are axioms, the Zermelo-Fraenkel Axioms with Choice (ZFC) which are used as the basis for set theory. Since set theory can be used to do all of mathematics, these axioms are the fundamental assumptions that all of mathematics starts from.5 It should be noted that in this setup, all mathematical objects are taken to be sets, so our sets are sets of sets.

The basic reason for carefully choosing axioms is that if we don’t, we can find paradoxes and prove contradictions. For example, suppose that we just say that any property \( P(x) \) can be used to define a set \( \{ x : P(x) \} \). Then we can define a set \( R = \{ x : x \notin x \} \). Is \( R \in R \)? Clearly \( R \in R \) iff \( R \notin R \), which is a contradiction! This is known as Russell’s Paradox. Similarly, if we allow ourselves to talk about a set of all ordinals, we can prove a contradiction, by considering \( Ord + 1 \), which would be a well-ordered set longer than \( Ord \).

The solution is to not allow us to talk about just any set, but to pick certain axioms that tell us some ways we are allowed to make sets. We want to allow ourselves to make all the sets we really need to do math, but not allow ourselves to prove contradictions. Within this framework, Russell’s Paradox is not a paradox but merely a proof by contradiction that there does not exist a set \( R = \{ x : x \notin x \} \). That is, we can talk about the collection of all sets \( x \) such that \( x \notin x \), but this collection is not a set. Thus \( R \notin R \), because \( R \) is not a set, and \( R \) is the collection of all sets that are not elements of themselves. We similarly say that there is no set of all ordinals. An arbitrary collection of sets is called a class; a class which is not a set is called a proper class.

In this section we learn both what the ZFC axioms are and the framework, called first-order logic, in which they operate. First-order logic will be talked about in more detail in Susan and Steve’s Model Theory class this week.

In axiomatic set theory, we must carefully state exactly what axioms we are using and what rules we are using to prove theorems from them. First, we must say what it even means to state a theorem and what language we are allowed to use to express ideas.

Definition 7.1. Fix a countably infinite collection of symbols to be used as variables (e.g., \( a, a', a'', \ldots \)). A formula (in the language of set theory) is defined inductively as follows:

1. If \( x \) and \( y \) are variables, then \( x = y \) and \( x \in y \) are formulas (these are called atomic formulas).
2. If \( \varphi \) and \( \psi \) are formulas, then so are \( \varphi \land \psi \), \( \varphi \lor \psi \), and \( \neg \varphi \).
3. If \( \varphi \) is a formula and \( x \) is a variable, then \( \forall x \varphi \) and \( \exists x \varphi \) are formulas.

Let’s explain this notation a bit. A formula is just any statement that we are allowed to make; anything that cannot be expressed by a formula is declared to be outside the realm of set theory and mathematics. The basic things we can talk about are whether two things are equal or whether one set is an element of another set. Given formulas, we can build other formulas out of them using the basic notions of logic: \( \varphi \land \psi \) means “\( \varphi \) and \( \psi \)” \( \varphi \lor \psi \) means “\( \varphi \) or \( \psi \)” and \( \neg \varphi \) means “not \( \varphi \)” \( \forall x \varphi \) means that \( \varphi \) is true for all possible values of the variable \( x \) and \( \exists x \varphi \) means \( \varphi \) is true for some value of the variable \( x \). We will also write \( \varphi \Rightarrow \psi \) and \( \varphi \Leftrightarrow \psi \) as abbreviations for \( \neg \varphi \lor \psi \) and \( (\varphi \land \psi) \lor (\neg \varphi \land \neg \psi) \), respectively.

A variable appearing in a formula which is being quantified over is bound, and one which is not is free. For example, in \( x \in y \land \exists z (z = z) \), \( z \) is bound but \( x \) and \( y \) are free. Note that in \( x = x \lor \exists x (x \in x) \), the first two \( x \)'s are free but the second two are bound. We write \( \varphi(x_1, \ldots, x_n) \) or \( \varphi(\bar{x}) \) to denote that all free variables occurring in \( \varphi \) are among the \( x_1, \ldots, x_n \). A class on formal logic would give a more precise definition of free and bound variables, but we will say no more and hope the meaning is clear.

If a formula has no free variables, it is called a sentence. A sentence is something that can be true or false. For example, \( \forall z (z = z) \) is true, while \( \forall z (x \in z) \) is false. On the other hand, a formula with free variables has no intrinsic truth values, and can only be interpreted if specific sets are substituted for the variables. For examples, \( x \in y \) is neither true nor false, since we don’t know what \( x \) and \( y \) are.

---

5There do exist alternate approaches to the foundations of mathematics, but axiomatic set theory is the standard approach.
There are some other things we ought to be able to talk which cannot obviously be written as formulas. For example, if \( A = \{ x : \varphi(x) \} \) is a specific set, we ought to be able to say \( x \in A \), \( x = A \), or \( A \in x \). We express these formally as \( \varphi(x) \), \( \forall y(y \in x \iff \varphi(y)) \), and \( \forall y(y = A \Rightarrow y \in x) \), respectively (where in the last of these, we use the second expression for “\( y = A \)”).

Now that we have the language of set theory, we need to say how we are allowed to prove things. In a class on formal logic, you would learn very precise, algorithmic rules for what kinds of proofs are allowed, so that a computer could check the validity of a proof. Since this is not a class on formal logic, we will not learn such rules for the whole language of set theory.

### Axiom (Extensionality)

If \( x \in X \) iff \( x \in Y \) for all \( x \), then \( X = Y \).

This axiom does not tell us how we can construct sets, but it tells us what sets are: a set is determined by its elements. Alternatively, this says that everything is a set: there are no “atoms” which have no elements but are distinct from the empty set.

The next four axioms are self-explanatory and allow us to construct certain sets.

### Axiom (Empty Set)

There exists a set \( \emptyset \) which has no elements.

### Axiom (Pairing)

For any \( x \) and \( y \), there is a set \( \{ x, y \} = \{ z : z = x \text{ or } z = y \} \).

### Axiom (Union)

For any \( X \), there is a set \( \bigcup X = \{ a : a \in x \text{ for some } x \in X \} \).

Note that Pairing and Union together imply that binary unions \( x \cup y = \bigcup \{ x, y \} \) exist.

### Axiom (Power Set)

For any \( X \), there is a set \( \mathcal{P}(X) = \{ Y : Y \subseteq X \} \) (where \( Y \subseteq X \) means every element of \( Y \) is also an element of \( X \)).

### Axiom (Separation Schema)

Let \( \varphi(x, p_1, \ldots, p_n) \) be any formula and \( X \) be any set. Then for any sets \( p_1, \ldots, p_n \), there exists a set \( \{ x \in X : \varphi(x, p_1, \ldots, p_n) \} \).

In naive (non-axiomatic) set theory, one typically works with the assumption that for any \( \varphi \), we can construct a set \( A = \{ x : \varphi(x) \} \). This is false by Russell’s Paradox. However, with Separation we can get something almost as good: we just have to restrict \( A \) to only contain elements of some set \( X \). Briefly, Separation says that any subclass of a set is a set.

As a final note, technically Separation is not one axiom but infinitely many, one for each formula \( \varphi \).

### Axiom (Replacement Schema)

Let \( \varphi(x, y, p_1, \ldots, p_n) \) be any formula and \( X \) be any set. Let \( p_1, \ldots, p_n \) be sets and suppose that for any \( x \), there is at most one \( y \) such that \( \varphi(x, y, p_1, \ldots, p_n) \) is true. Then there exists a set \( \{ y : \varphi(x, y, p_1, \ldots, p_n) \text{ for some } x \in X \} \).

The condition on the formula \( \varphi \) is just that it defines a function (on some domain). Replacement thus says that the image of any set under a function is again a set. This implies that any class which is the same size or smaller than a set is also a set. Thus the only way a class can fail to be a set is if it is too big to be a set.

Like Separation, Replacement is actually an infinite collection of axioms.

As we will later see, we cannot prove the existence of an infinite set from the other axioms without just explicitly assuming it. There are various equivalent ways of formalizing the statement “there exists an infinite set”: the following is the standard one.

### Axiom (Infinity)

There exists a set \( X \) such that \( \emptyset \in X \) and such that for all \( x \in X \), \( x \cup \{ x \} \in X \).
Such an $X$ is called an \textit{inductive set}. Note that the ordinal $\omega$ is inductive, since for ordinals $x \cup \{x\}$ is just adding 1. In fact, it’s not hard to see that $\omega$ is the intersections of all inductive sets. Thus Infinity and Separation together imply the existence of $\omega$ (since $\omega$ is a subset of any inductive set).

\textbf{Axiom (Foundation).} The relation $\in$ is \textit{well-founded}: for any nonempty $x$, there exists $y \in x$ such that there is no $z \in x$ such that $z \in y$.

We will explore the meaning of this axiom later. Note that if $\in$ were a total order, Foundation would say exactly that it is well-ordering (for any set $x$, there is a least element $y \in x$). For the exact same reason as for well-orderings, Foundation implies that we can use induction on the relation $\in$. Intuitively, Foundation says that all sets can be built up from the empty set by (transfinite) induction. We call an element $y \in x$ which contains no $z \in x$ an $\in$-\textit{minimal} element of $x$.

\textbf{Axiom (Choice).} Let $X$ be a set of nonempty sets. Then there exists a set $Y$ such that $Y \cap x$ has exactly one element for all $x \in X$.

That is, given a set of sets, we can pick one element from each simultaneously. As we have seen in the exercises in previous sections, Choice is equivalent to the Well-Ordering Principle as well as several other important facts about the behavior of infinite sets.

Unlike all the other axioms (except perhaps Foundation), Choice asserts the existence of sets without explicitly saying what they are. That is, Choice allows \textit{nonconstructive} proofs in which we show something exists without being able to write it down.

Using all the axioms together, it is possible to construct any standard set in mathematics. We will not always justify the existence of sets from the axioms; in such cases we encourage the reader to find such a justification. Some examples of constructing sets from the axioms are given in the exercises.

We now explain the meaning of the Axiom of Foundation. Roughly, it allows us to use the axioms to build up the universe of sets inductively. Let’s make this precise.

\textbf{Definition 7.2.} By induction on $\alpha \in \text{Ord}$, define sets $V_\alpha$ as follows. First, $V_0 = \emptyset$. If $\alpha + 1$ is a successor ordinal, define $V_{\alpha+1} = P(V_\alpha)$. Finally, if $\alpha$ is a limit ordinal and we have defined $V_\beta$ for all $\beta < \alpha$, then define $V_\alpha = \bigcup_{\beta < \alpha} V_\beta$. We call the collection of the $V_\alpha$ the \textit{von Neumann hierarchy} (or the \textit{cumulative hierarchy}).

Note that we use Empty Set to define $V_0$, Power Set to define $V_{\alpha+1}$, and Replacement and Union for limit ordinals (by Replacement, if $\alpha = \{\beta\}_{\beta < \alpha}$ is a set then $\{V_\beta\}_{\beta < \alpha}$ is a set).

Intuitively, $V_\alpha$ is the set of sets that can be built in less than $\alpha$ steps. No sets can be built in less than 0 steps, so $V_0 = \emptyset$. The empty set can be built in 0 steps (no other sets have to exist before building the empty set), so $V_1 = P(V_0) = \{\emptyset\}$. In general, for $n < \omega$, it is easy to see that a set is in $V_n$ iff you can list its elements using only \{ and \} and there are never more than $n$ nested braces at a time. For infinite ordinals $\alpha$, the description is a bit more complicated. Note that if $\alpha < \beta$ then $V_\alpha \subseteq V_\beta$. This can be proven by induction on $\beta$: for $\beta$ limit this is immediate from the definition, and for $\beta = \gamma + 1$, by the induction hypothesis it suffices to show that $V_\gamma \subseteq V_\beta = P(V_\gamma)$. But every element of $V_\gamma$ is a subset of $V_\alpha$ for some $\alpha < \gamma$ and by induction $V_\alpha \subseteq V_\gamma$, so it is true that $V_\gamma \subseteq P(V_\gamma)$.

The Axiom of Foundation says that every set is in $V_\alpha$ for some ordinal $\alpha$. That is, $V = \bigcup_\alpha V_\alpha$, where $V$ is the class of all sets. In the homework you will prove that this actually is equivalent to the statement of the Axiom of Foundation we gave earlier, namely that for any nonempty set $x$, there exists $y \in x$ such that $x \cap y = \emptyset$.

\textbf{Definition 7.3.} The \textit{rank} of a set $x$ is the least $\alpha$ such that $x \in V_{\alpha+1}$.

Note that rank$(x)$ is the least ordinal which is greater than rank$(y)$ for all $y \in x$. That is, rank$(x) = \sup_{y \in x} \text{rank}(y) + 1$. Also, rank$(x)$ is the least $\alpha$ such that $x \subseteq V_\alpha$. This notion of rank is useful because it allows us to prove things about all sets by induction on rank.

The von Neumann hierarchy is a useful way to understand the function of each axiom. Empty Set lets us get started with $V_0$. Power Set allows us to continue the hierarchy at successor steps. Replacement and
Union let us continue at limit steps. Infinity tells us that $V_\omega$ exists, allowing us to get to the infinite ordinals at all.

The other axioms do not directly help us build the hierarchy but tell us more about what sets as a whole look like. Foundation tells us that the hierarchy gives us all the sets. Extensionality tells us what sets are. Separation tells us that we can build any subset, which clarifies the meaning of Power Set. Choice tells us that every set can be put in bijection with an ordinal and basically tells us that infinite sets are well-behaved.

7.1 Exercises

Exercise 7.1. For two sets $x$ and $y$, define the ordered pair $(x,y)$ to be the set $\{\{x\}, \{x,y\}\}$. This exists by Pairing.

(a): Show that $(x,y) = (a,b)$ iff $x = a$ and $y = b$.

(b): Show that for any two sets $X$ and $Y$, there exists a set $X \times Y = \{(x,y) : x \in X \text{ and } y \in Y\}$. (Hint: Use Pairing, Union, Power Set and Separation.)

Exercise 7.2. A function $f : X \to Y$ is a subset $f \subseteq X \times Y$ such that for each $x \in X$, there is a unique $y \in Y$ such that $(x,y) \in f$. Show that for any $X$ and $Y$, there exists a set $Y^X$ of all functions $X \to Y$.

Exercise 7.3. Show that the class $V = \{x : x = x\}$ of all sets is a proper class. (Hint: Use Separation and Russell’s Paradox.)

Exercise 7.4. Compute rank$(X \times X)$ in terms of rank$(X)$. (Hint: You should separate into the cases where rank$(X)$ is successor and the cases where rank$(X)$ is limit.)

Exercise 7.5. Let $N$ be a set of rank $\omega$, $Z = N \times N/\sim$ for some equivalence relation $\sim$, and $Q = Z \times (Z \setminus \{0\})/\simeq$ for some other equivalence relation $\simeq$. Define $\mathbb{R}$ to be a certain subset of $\mathcal{P}(\mathbb{Q})$ and $\mathbb{C} = \mathbb{R} \times \mathbb{R}$.

(a): Show that all of these sets have rank $< 2^\omega$.

(b): Compute the exact rank of each of these sets.

In fact, any set that you have ever encountered in a non-set theory class has rank $< 2^\omega$. While the set-theoretic universe is enormous, virtually all of mathematics can be done just in $V_{2\omega}$.

Exercise 7.6. Show that every set in $V_\omega$ can be proven to exist using only Empty Set, Pairing, and Union. (Hint: Every $x \in V_\omega$ is finite and has finite rank. Use induction on both rank$(x)$ and $|x|$.)

Recall that the Axiom of Foundation’s original statement was:

Axiom (Foundation). The relation $\in$ is well-founded: for any nonempty $x$, there exists $y \in x$ such that there is no $z \in x$ such that $z \in y$.

Exercise 7.7 ($\in$-Induction). Use the Axiom of Foundation to prove the following: Suppose that whenever $\varphi(y)$ is true for all $y \in x$, then $\varphi(x)$ is true. Then $\varphi(x)$ is true for all sets $x$.

Exercise 7.8.

(a) Use Exercise 7.7 to prove that $V = \bigcup V_\alpha$ (i.e., any set is in $V_\alpha$ for some ordinal $\alpha$).

(b) Use $V = \bigcup V_\alpha$ to prove the Axiom of Foundation. (Hint: Let $y \in x$ have minimal rank.)
7.2 Appendix: Reference sheet of the ZFC axioms

Here is a list of all the axioms of ZFC, both in readable English and in the language of set theory.

**Axiom (Extensionality).** \(\forall X\forall Y(\forall x(x \in X \leftrightarrow x \in Y) \Rightarrow X = Y)\)

If \(x \in X\) iff \(x \in Y\) for all \(x\), then \(X = Y\).

**Axiom (Empty Set).** \(\exists x\forall y(\neg y \in x)\)

There exists a set \(\emptyset\) which has no elements.

**Axiom (Pairing).** \(\forall x\forall y\exists z(z \in S \Leftrightarrow (z = x \lor z = y))\)

For any \(x\) and \(y\), there is a set \(\{x, y\}\).

**Axiom (Union).** \(\forall X\exists U\forall a(a \in U \Leftrightarrow \exists x(x \in X \land a \in x))\)

For any \(X\), there is a set \(\bigcup X = \{a : a \in x\text{ for some } x \in X\}\).

**Axiom (Power Set).** \(\forall X\exists P\forall Y(Y \in P \Leftrightarrow \forall x(x \in Y \Rightarrow x \in X))\)

For any \(X\), there is a set \(\mathcal{P}(X) = \{Y : Y \subseteq X\}\) (where \(Y \subseteq X\) means every element of \(Y\) is also an element of \(X\)).

**Axiom (Separation Schema).** For any formula \(\varphi(x, p_1, \ldots, p_n)\), the following is an axiom:

\[\forall X\forall p_1\ldots\forall p_n\exists S\forall x(x \in S \Leftrightarrow (x \in X \land \varphi(x, p_1, \ldots, p_n)))\]

Let \(\varphi(x, p_1, \ldots, p_n)\) be any formula and \(X\) be any set. Then for any sets \(p_1, \ldots, p_n\), there exists a set \(\{x \in X : \varphi(x, p_1, \ldots, p_n)\}\).

**Axiom (Replacement Schema).** For any formula \(\varphi(x, y, p_1, \ldots, p_n)\), the following is an axiom:

\[\forall X\forall p_1\ldots\forall p_n\forall x\forall y\forall y'(\varphi(x, y, p_1, \ldots, p_n) \land \varphi(x, y', p_1, \ldots, p_n) \Rightarrow y = y') \Rightarrow \exists S\forall y(y \in S \Leftrightarrow \exists x(x \in X \land \varphi(x, y, p_1, \ldots, p_n)))\]

Let \(\varphi(x, y, p_1, \ldots, p_n)\) be any formula and \(X\) be any set. Let \(p_1, \ldots, p_n\) be sets and suppose that for any \(x\), there is at most one \(y\) such that \(\varphi(x, y, p_1, \ldots, p_n)\) is true. Then there exists a set \(\{y : \varphi(x, y, p_1, \ldots, p_n)\text{ for some } x \in X\}\).

**Axiom (Infinity).** \(\exists X(\forall S(\forall x(\neg x \in S) \Rightarrow S \in X) \land \forall x(x \in X \Rightarrow \forall y(\forall a(a \in y \Leftrightarrow (a \in x \land a = x)) \Rightarrow y \in X)))\)

There exists a set \(X\) such that \(\emptyset \in X\) and such that for all \(x \in X\), \(x \cup \{x\} \in X\).

**Axiom (Foundation).** \(\forall x(\exists y(y \in x) \Rightarrow \exists y(y \in x \land \forall z(z \in x \Rightarrow \neg z \in y)))\)

The relation \(\in\) is well-founded: for any nonempty \(x\), there exists \(y \in x\) such that there is no \(z \in x\) such that \(z \in y\).

**Axiom (Choice).** \(\forall X(\forall x(x \in X \Rightarrow \exists a(a \in x)) \Rightarrow \exists Y \forall x(x \in X \Rightarrow (\exists a(a \in Y \land a \in x) \land \forall a\forall a'(a \in Y \land a \in x \land a' \in Y \land a \neq a' \Rightarrow a = a'))))\)

Let \(X\) be a set of nonempty sets. Then there exists a set \(Y\) such that \(Y \cap x\) has exactly one element for all \(x \in X\).
8 Models of Set Theory

In Section 6, we showed that Empty Set, Pairing, and Separation can be proved from the other axioms. It turns out that the remaining 7 axioms are all irredundant—none of them can be proven from the others. Our goal now is to prove this for as many of them as possible.

How can you prove that something cannot be proven? The answer is to find a model where it fails. For example, consider the axioms of group theory:

1. \( \forall a \forall b \forall c ((ab)c = a(bc)) \)
2. \( \forall a (a1 = a \land 1a = a) \)
3. \( \forall a (aa^{-1} = 1 \land a^{-1}a = 1) \)

One could ask whether \( \forall a \forall b (ab = ba) \) can be proven from these axioms—that is, whether the axioms for groups imply that groups must be abelian. The answer, of course, is no, and the way you prove it is to provide a model in which the axioms for groups hold but the abelian axiom does not hold. That is, we just have to give an example of a nonabelian group. The smallest such example is the symmetric group \( S_3 \).

Let’s analyze what’s going on from the point of view of a logician. If we could prove the abelian axiom from the axioms for groups, then we could prove that the abelian axiom was true for any structure in which the axioms for groups were true. Thus if there exists a group which is not abelian, the abelian axiom cannot be proven.

Let’s now apply this to set theory.

Definition 8.1. A structure (over the language of set theory) is a set \( M \) together with a binary relation \( \epsilon \) on \( M \). For any sentence \( \varphi \) in the language of set theory, we say that \( M \) is a model of \( \varphi \) (or \( M \) “believes” \( \varphi \)) if \( \varphi \) is true when we interpret \( \epsilon \) as the relation \( \epsilon \) and consider all quantified variables as taking values in the set \( M \). We write this as \( M \vDash \varphi \) or \( \varphi \vDash M \).

For example, \( M \vDash \forall x (\neg x \in x) \) iff for all \( x \in M \), \( x \epsilon x \) is false. To be completely precise, we should really define the meaning of \( M \vDash \varphi \) by induction on the length of the formula \( \varphi \), but this is not a logic or model theory class, so we will just trust that the meaning of this informal definition is clear. In fact, for all structures we will actually consider, \( \epsilon \) will be the same as the \( \in \) relation we already have, so we will not specify it.

Note that the meaning of \( M \vDash \varphi \) can sometimes be subtle. For example, \( M \vDash \forall x (\neg x \in x) \) does not mean that some element of \( M \) is uncountable. Rather, it means that there is a set \( x \in M \) such that there is no injection in \( M \) from \( x \) to \( \omega \). In particular, it is possible that there is an element \( x \in M \) such that \( M \vDash x = \aleph_1 \), but \( x \) is actually countable. In cases like this we write \( x = \aleph_1^M \).

We wish to relate these models to provability.

Definition 8.2. Let \( T \) be a set of sentences in the language of set theory and \( \varphi \) is a sentence, we write \( T \vdash \varphi \) if \( \varphi \) can be proven if we take \( T \) as our axioms.

Fact 8.3. Let \( M \) be a structure, \( \varphi \) be a sentence, and \( T \) be a set of sentences. If \( M \vDash T \) and \( T \vdash \varphi \), then \( M \vDash \varphi \).

We cannot prove this fact because we have not given a full definition of the notion of a “proof” and the relation \( \vdash \), but it is not hard to prove, and is true basically because the rules governing proofs are logically valid.

In fact, the converse is also true.

Theorem 8.4. [Completeness of First-Order Logic] Let \( \varphi \) be a sentence and let \( T \) be a set of sentences. Suppose that for all structures \( M \) such that \( M \vDash T \), \( M \vDash \varphi \) as well. Then \( T \vdash \varphi \).

We certainly are not equipped to prove this theorem, and will not actually need it. However, it nicely
completes the picture of what is going on: if it’s not possible to prove \( \varphi \) from \( T \), then we are guaranteed to be able to find a model which witnesses this. That is, constructing models is in some sense the only way we can show that something cannot be proven.

An important special case of the Theorem 8.4 is the following. Let \( \bot \) denote any sentence which is a contradiction, such as \( \exists x(x \in x \land x \notin x) \). It does not really matter what contradiction we choose, since anything can be proven from a contradiction; any necessarily false sentence would serve equally well.

**Definition 8.5.** A set of sentences \( T \) is **consistent** if \( T \not\models \bot \). We abbreviate this as \( \text{Con}(T) \).

**Corollary 8.6.** A set of sentences \( T \) is consistent iff it has a model (i.e., there exists a structure \( M \) such that \( M \models T \)).

**Proof.** If \( T \) is inconsistent, it certainly cannot have a model since \( \bot \) cannot be true in any structure. Conversely, if \( T \) has no model, then it is vacuously true that \( M \models T \) implies \( M \models \bot \). By Theorem 8.4, this means that \( T \models \bot \), so \( T \) is inconsistent.

How do we find a model where all axioms but one hold? One thing we can do is look at how we use the axioms to build the von Neumann hierarchy. If we then just don’t allow ourselves to use that axiom and cut off the hierarchy, we might find a model where all axioms but that one are true.

Let’s start simple. The first axiom we used when constructing the von Neumann hierarchy was Empty Set, which gave us \( V_0 = \emptyset \). Empty Set follows from Infinity (together with Separation), so Infinity will also have to fail in our model. If we don’t allow ourselves the empty set, then we can’t prove we have any sets!

**Theorem 8.7.** \( V_0 \models ZFC \setminus \{\text{Empty Set, Infinity}\} \).

**Proof.** Every axiom except Empty Set and Infinity starts “For all sets, . . .”. If there are no sets, they are all vacuously true!

**Corollary 8.8.** \( ZFC \setminus \{\text{Empty Set, Infinity}\} \) cannot prove the existence of a set (and in particular, cannot prove Empty Set).

What’s the next axiom we use? We use Power Set at successor steps. However, for finite \( n \), we don’t actually need Power Set to construct \( V_{n+1} \) from \( V_n \); Pairing and Union are enough (see Exercise 7.6). Thus, for example, \( V_1 \) will fail to be a model of Power Set, but it will also fail to be a model of Pairing, so it is not a very interesting example of getting almost all the axioms to hold. In Exercise 8.4, you will see exactly which axioms hold in \( V_n \) for \( 0 < n < \omega \).

The next axiom we use in constructing the von Neumann hierarchy is Infinity, which we need to prove that \( \omega \) and hence \( V_\omega \) exists. In fact, Infinity really is needed for this:

**Theorem 8.9.** \( V_\omega \models ZFC \setminus \{\text{Infinity}\} \).

**Proof.** By Exercises 8.2 and 8.1, Foundation holds in any structure where the relation is \( \in \) and Extensionality holds in any transitive set (and hence in \( V_\alpha \) for any \( \alpha \)). It is straightforward to check that all the other axioms are also true, using the fact that \( V_\omega \) is exactly the set of hereditarily finite sets. For example, Replacement is true because any finite subset of \( V_\omega \) is an element of \( V_\omega \) (since there is some finite bound on the rank of all of its elements), and any set we can construct from an element of \( V_\omega \) using Replacement will be finite. Similarly, Union is true because a union of finitely many finite sets is finite. Empty Set is trivial and Power Set follows from \( P(V_n) = V_{n+1} \subset V_\omega \). Pairing and Separation follow from the other axioms. Finally, Choice is true since it is true in \( V \), and if the sets in question are all in \( V_\omega \), then the set whose existence is asserted by Choice is also in \( V_\omega \) (being a subset of their union).

**Corollary 8.10.** \( ZFC \setminus \{\text{Infinity}\} \) cannot prove the existence of an infinite set (and in particular, cannot prove Infinity).
8.1 Exercises

Exercise 8.1. A set \( M \) is transitive if \( x \in M \) and \( y \in x \) implies \( y \in M \). Let \( M \) be any transitive set (with \( \epsilon = \in \)). Show that \( M \vDash \text{Extensionality} \).

Exercise 8.2. Let \( M \) be any structure where the relation \( \epsilon \) is \( \in \). Show that \( M \vDash \text{Foundation} \).

Definition 8.11. Let \( x \) be a set. By induction on \( \text{rank}(x) \), we say that \( x \) is hereditarily finite if:

1. \( x \) is finite.
2. For all \( y \in x \), \( y \) is hereditarily finite.

Exercise 8.3. Show that \( x \) is hereditarily finite iff \( x \in V_\omega \). (Hint: For both directions, use induction on \( \text{rank}(x) \).)

Exercise 8.4. Which axioms are true in \( V_n \) for \( 0 < n < \omega \)?

Exercise 8.5. Which axioms are true in \( V_{\alpha+1} \) for \( \alpha \) infinite?

Exercise 8.6. Which axioms are true in \( V_{2\omega} \)?
9 Hereditarily small sets

Let’s now try to find a model where Power Set fails. The first time we really need Power Set is to go from $V_{\omega+1}$ to $V_{\omega+2}$ ($V_{\omega+1}$ itself just comes from Infinity). A bit of thought shows that $V_{\omega+2}$ is also the first stage that contains uncountable sets. Indeed, countable sets are closed under the operations that the other axioms gives you—a union of countably many countable sets is countable, a subset of a countable set is countable, etc. Thus the right way to get a model without Power Set is to look at a world where everything is countable.

**Definition 9.1.** Let $\kappa$ be a cardinal. By induction on rank($X$), we say $X$ is hereditarily of size $< \kappa$ if:

1. $|X| < \kappa$.
2. Every element of $X$ is hereditarily of size $< \kappa$.

We write $H_\kappa$ for the class of all sets that are hereditarily of size $< \kappa$.

For $\kappa = \aleph_0$, $H_\aleph_0 = V_\omega$ is the hereditarily finite sets of Exercise 8.3. For $\kappa = \aleph_1$, $H_\aleph_1$ is the collection of “hereditarily countable sets”.

It’s not obvious that $H_\kappa$ is actually a set.

**Proposition 9.2.** Let $\kappa$ be a regular cardinal. Then $H_\kappa \subseteq V_\kappa$.

**Proof.** Let $X \in H_\kappa$. We prove by induction on rank($X$) that $X \in V_\kappa$, i.e. rank($X$) $< \kappa$. Suppose we already know that for all $Y \in H_\kappa$ with rank($Y$) $< \kappa$, $Y \in V_\kappa$. Then every element of $X$ is in $V_\kappa$. Since $|X| < \kappa$, the ranks of elements of $X$ form a set of less than $\kappa$ ordinals which are all less than $\kappa$. Since $\kappa$ is regular, the supremum of all these ordinals is still less than $\kappa$. It follows that rank($X$) $< \kappa$, so $X \in V_\kappa$.

**Corollary 9.3.** For any $\kappa$, $H_\kappa$ is a set.

**Proof.** If $\kappa$ is regular, this follows from Proposition 9.2 and Separation. For $\kappa$ singular, it again follows from Separation using the fact that $\kappa^+$ is singular and hence $H_\kappa \subseteq H_{\kappa^+} \subseteq V_{\kappa^+}$. □

**Theorem 9.4.** $H_{\omega_1} \models ZFC \setminus \{\text{Power Set}\}$, but $H_{\omega_1} \not\models \text{Power Set}$.

**Proof.** Foundation, Extensionality, Empty Set, and Infinity are easy. For Replacement, any countable set of hereditarily countable sets is again a hereditarily countable set. For Union, a union of countably many countable sets is countable, and it’s easy to see that the same is true for hereditarily countable sets. Pairing and Separation are similarly easy, and Choice is true for the same reason as in all the previous models.

Power set fails because every subset of $\omega$ is in $H_{\omega_1}$ but there is no set that contains them all. Indeed, $H_{\omega_1} \models \text{“every set is countable”}$.

**Corollary 9.5.** $ZFC \setminus \{\text{Power Set}\}$ cannot prove the existence of an uncountable set (and in particular, cannot prove Power Set).

The only nontrivial part of the proof that everything except Power Set holds was that a countable union of countable sets is countable, which is just regularity of $\omega_1$. Thus we more generally have:

**Theorem 9.6.** Let $\kappa$ be a regular uncountable cardinal. Then $H_\kappa \models ZFC \setminus \{\text{Power Set}\}$.

One way to understand this is to say that the real function of Power Set is to prove that for any set, there is always a set of strictly larger cardinality. Thus without Power Set, we can set an arbitrary cardinality bound of $\kappa$ on sets and obtain a model $H_\kappa$.

Let’s now look at Union. When is Union needed to build sets? Well, when building the $V_\alpha$, it is used at limit steps, but we saw that it still does hold in $V_\alpha$ for $\alpha$ limit. Indeed, it is not hard to show that it holds in any $V_\alpha$. What about $H_\kappa$? Power Set turned out to be all about being able to get sets of larger cardinalities. Power Set tells us that sets of size $\beth_1 = 2^{\beth_0}$, $\beth_2 = 2^{2^{\beth_0}}$, $\beth_3 = 2^{2^{2^{\beth_0}}}$, ... must exist. Write $\beth_\omega = \sup_\alpha \beth_\alpha$.\(^6\) Then Union can tell us that sets of size $\beth_\omega$ exist, but Power Set cannot!

\(^6\)More generally, we can define cardinals $\beth_\alpha$ for all ordinals $\alpha$ similarly. GCH then says that $\aleph_\alpha = \beth_\alpha$ for all $\alpha$. 

28
Theorem 9.7. $H_{\beth_\omega} \models ZFC \setminus \{\text{Union}\}$, but $H_{\beth_\omega} \not\models \text{Union}$.

Proof. The proof of Theorem 9.4 works for every axiom except Union. Furthermore, Power Set now holds since if $|X| < \beth_\omega$, $2^{|X|} < \beta_\omega$. Union fails since the set $\{\beta_n\}_{n \in \omega}$ exists but its union $\beth_\omega$ does not. 

Corollary 9.8. $ZFC \vdash \text{Con}(ZFC \setminus \{\text{Union}\})$. $ZFC \setminus \{\text{Union}\}$ cannot prove the existence of a set of size $\beth_\omega$ (and in particular, cannot prove Union).

As in our previous models, there is only one special property of $\beth_\omega$ we really needed.

Definition 9.9. A cardinal $\kappa$ is strong limit if for any $\lambda < \kappa$, $2^\lambda < \kappa$.

The proof of Theorem 9.7 more generally shows the following.

Theorem 9.10. Let $\kappa$ be an uncountable strong limit cardinal. Then $H_\kappa \models ZFC \setminus \{\text{Union}\}$.

Something remarkable now happens if we combine Theorems 9.6 and 9.10.

Definition 9.11. We say a cardinal $\kappa$ is inaccessible if it is uncountable, strong limit, and regular.

Corollary 9.12. Let $\kappa$ be inaccessible. Then $H_\kappa = V_\kappa \models ZFC$.

Proof. Since $\kappa$ is both regular and strong limit, $H_\kappa$ is a model of both $ZFC \setminus \{\text{Power Set}\}$ and $ZFC \setminus \{\text{Union}\}$, and hence of all of ZFC.

9.1 Exercises

Exercise 9.1. Let $\kappa$ be inaccessible. Show that $H_\kappa = V_\kappa$. (Hint: We already know $H_\kappa \subseteq V_\kappa$. Show the other direction by induction on rank.)

Exercise 9.2.

(a) Show that $|V_\omega| = \aleph_0$.

(b) Show that $|V_{\omega+n}| = \beth_n$ for all $n \in \mathbb{N}$.

Exercise 9.3. Let $\kappa$ be a strong limit cardinal and let $\lambda, \theta < \kappa$. Show that $\lambda^\theta < \kappa$.

Exercise 9.4. Prove that if $H_\kappa = V_\kappa$ and $\kappa$ is uncountable, then $\kappa$ is inaccessible.

Exercise 9.5. Show that $|H_{\kappa_1}| = 2^{2^{\aleph_0}}$. (Hint: For $\alpha < \omega_1$, let $A_\alpha = H_{\kappa_1} \cap V_\alpha$. Prove by induction on $\alpha$ that $|A_\alpha| \leq 2^{\aleph_0}$.)
10 Gödel’s Incompleteness Theorem and relative consistency

At the end of class last time, we saw that an inaccessible cardinal allows us to construct a model of ZFC. Since a set of axioms is consistent iff it has a model, we see that if an inaccessible cardinal exists, we can prove (in ZFC) that ZFC is consistent. This would be nice—it would be nice to be able to prove that we will never prove a contradiction. However, there’s a very general problem with being able to do this.

Theorem 10.1. [Gödel’s (Second) Incompleteness Theorem] Let $T$ be a set of axioms which is computably enumerable and which is strong enough to provide a foundation for mathematics. Then if $T$ is consistent, $T$ cannot prove its own consistency.

Here “computably enumerable” is used in the same sense as in Steve’s computability theory class: you can write down an algorithm that will one by one list all the axioms. We do not have time here to discuss what it means for a set of axioms to be “strong enough to provide a foundation for mathematics”, but any reasonably non-stupid subset of the ZFC axioms will satisfy this. In particular, if we are working in ZFC, we can never prove that ZFC is consistent, unless ZFC is actually inconsistent (in which case, of course, we can prove anything!).

This gives a proof that we cannot prove inaccessible cardinals exist.

Theorem 10.2. If ZFC is consistent, it cannot prove that inaccessible cardinals exist.

Proof. If $\kappa$ is inaccessible, $H_\kappa \models ZFC$, so ZFC is consistent. Thus if we can prove an inaccessible cardinal exists, then we can prove ZFC is consistent. By Gödel’s Incompleteness Theorem, ZFC cannot prove it is consistent unless it is inconsistent. Hence if ZFC is consistent, it cannot prove that inaccessibles exist.

So we can’t prove that inaccessibles exist. But maybe we can at least prove that assuming they exist doesn’t give us a contradiction. That is, we want to prove:

If ZFC is consistent, then so is ZFCI,

where ZFCI is the set of axioms consisting of ZFC plus one more axiom saying that there exists an inaccessible cardinal. Unfortunately, even this much is impossible!

Theorem 10.3. If ZFC is consistent, then ZFC cannot prove that if ZFC is consistent, so is ZFCI.

Proof. Let’s assume that we have a proof in ZFCI that if ZFC is consistent, then ZFCI is consistent. We will prove that ZFC is inconsistent.

Let’s start by working in the axiom system ZFCI. Since we are working in ZFCI, an inaccessible cardinal exists, so we can prove that ZFC is consistent. But we also have proven that if ZFC is consistent, then ZFCI is consistent. Thus we can prove that ZFCI is consistent.

Thus ZFCI can prove its own consistency. By Gödel’s Incompleteness Theorem, this implies that ZFCI is inconsistent! By our hypothesis, this implies that ZFC is also inconsistent.

That is, it is strictly harder to prove that ZFCI is consistent than it is to prove that ZFC is consistent. This same analysis applies to the earlier models we constructed where all but one axiom holds. That is, we have:

Theorem 10.4. If $ZFC \setminus \{\text{Infinity}\}$ is consistent, it cannot prove Infinity. Furthermore, it cannot prove that if $ZFC \setminus \{\text{Infinity}\}$ is consistent, then so is ZFC.

Theorem 10.5. If $ZFC \setminus \{\text{Power Set}\}$ is consistent, it cannot prove Power Set. Furthermore, it cannot prove that if $ZFC \setminus \{\text{Power Set}\}$ is consistent, then so is ZFC.

Theorem 10.6. If $ZFC \setminus \{\text{Union}\}$ is consistent, it cannot prove Union. Furthermore, it cannot prove that if $ZFC \setminus \{\text{Union}\}$ is consistent, then so is ZFC.
That is, each of Infinity, Power Set, and Union not only allow us to prove more things, but also make it strictly harder to prove that our axioms are consistent.

The same is also true of Replacement:

**Theorem 10.7.** $V_{2\omega} \models ZFC \setminus \{\text{Replacement}\}$, but $V_{2\omega} \not\models \text{Replacement}$. 

*Proof.* Foundation, Extensionality, Empty Set, and Infinity are easy. Union and Separation are clear because if rank$(x) < 2\omega$, then rank$(\bigcup x) \leq \text{rank}(x) < 2\omega$ and rank$(y) \leq \text{rank}(x) < 2\omega$ for any $y \subseteq x$. Power Set is true because if rank$(x) < 2\omega$, rank$(P(x)) = \text{rank}(x) + 1 < 2\omega$, and Pairing then follows. Choice is true for the same reason as it was for $V_{\omega}$. 

Finally, Separation fails because we can define a bijection from $\mathbb{N}$ to some countable set $X$ of rank $2\omega$ (since every element of $X$ will be in $V_{2\omega}$), but its image $X$ is not an element of $V_{2\omega}$. \qed

**Corollary 10.8.** $ZFC \setminus \{\text{Replacement}\}$ cannot prove sets of rank $2\omega$ exist.

**Theorem 10.9.** If $ZFC \setminus \{\text{Replacement}\}$ is consistent, it cannot prove Union. Furthermore, it cannot prove that if $ZFC \setminus \{\text{Replacement}\}$ is consistent, then so is $ZFC$. 

In fact, all that we used in the proof of Theorem 10.7 was that $2\omega$ is a limit ordinal.

**Theorem 10.10.** Let $\alpha > \omega$ be a limit ordinal. Then $V_\alpha \models ZFC \setminus \{\text{Replacement}\}$. 

The remaining axioms are Extensionality, Foundation, and Choice. For each of these, it turns out that adding either them or their negations to your set of axioms doesn’t make it any harder to be consistent. However, this means that $ZFC$ cannot prove there is a model where they fail; rather, $ZFC$ can prove that if $ZFC$ has a model, then there is a model where they fail.

Let’s start with Foundation. Since Foundation just says $V = \bigcup_\alpha V_\alpha$, there is a simple way to get from a model of $ZFC \setminus \{\text{Foundation}\}$ to a model of $ZFC$. 

**Theorem 10.11.** If $ZFC \setminus \{\text{Foundation}\}$ is consistent, so is $ZFC$. 

*Proof.* Let $M$ be a model of $ZFC \setminus \{\text{Foundation}\}$. Then we can define $V_\alpha^M \in M$ for each $\alpha \in \text{Ord}^M$, the element of $M$ which $M$ thinks is $V_\alpha$. There is then a subset $U = \{x \in M : M \models x \in \alpha \text{ for some } \alpha \in \text{Ord}\}$ of $M$. It is then straightforward to check that $U \models ZFC$. This is because $U = \bigcup V_\alpha$ is closed under taking sets you can define from the axioms: any subset of a set of $U$ is in $U$, as is any set in $U$ that can be defined using Replacement, $U$ is closed under taking unions and power sets, etc. Finally, Foundation holds in $U$ since $U \models V = \bigcup V_\alpha$, and this is equivalent to Foundation by Exercise 7.8. \qed

Basically, Foundation just says that only certain kinds of sets exist; if more sets exist, we can throw them away and get a model where Foundation does hold.

To get a model where Foundation fails requires more cleverness. The idea is to just add a new set $x$ such that $x = \{x\}$, violating Foundation.

**Theorem 10.12.** If $ZFC$ is consistent, so is $ZFC \setminus \{\text{Foundation}\} \cup \{\neg \text{Foundation}\}$. 

*Proof.* Let $(M, \epsilon)$ be a model of $ZFC$. Consider the elements $0 = \emptyset^M$ and $1 = \{\emptyset\}^M$, and define $F : M \to M$ by $F(0) = 1$, $F(1) = 0$, and $F(x) = x$ for $x \neq 0, 1$. Now define a relation $E$ on $M$ by $x \approx y$ if $x \epsilon F(y)$. We claim that $(M, E)$ is a model of $ZFC \setminus \{\text{Foundation}\} \cup \{\neg \text{Foundation}\}$. 

Note that 1 now acts like the empty set (i.e., has no elements), while 0 now satisfies $0 \approx 0$ (i.e., $0 = \{0\}$). It is straightforward to check that every axiom except Foundation still holds, using the fact that they held in $(M, \epsilon)$. For example, Extensionality still holds since $F$ is a bijection, so two elements of $M$ have the same $E$-elements if they had the same $\epsilon$-elements. More generally, if we want to construct some set in $(M, E)$ from the axioms, we basically conjugate the construction in $(M, \epsilon)$ by the bijection $F$.

However, Foundation fails in $(M, E)$, since $0 = \{0\}$ has no $E$-minimal element. \qed

Now let’s look at Extensionality. Again, we’ll first show that if we have a model without Extensionality,
we can get one with it. The idea behind this is simple: just mod out by the equivalence relation that equates sets which have the same elements. Note, though, that there is some subtlety—if $A$ and $B$ have the same elements, we must equate not just $A$ and $B$, but $\{A\}$ and $\{B\}$ as well. Thus the right way to define the desired equivalence relation is by induction on rank.

**Theorem 10.13.** If $\text{ZFC} \setminus \{\text{Extensionality}\}$ is consistent, so is $\text{ZFC}$.

**Proof.** Let $(M, \epsilon)$ be a model of $\text{ZFC} \setminus \{\text{Extensionality}\}$. Internal to $M$, define equivalence relations $\sim_\alpha$ on each $V_\alpha$ by induction. We must do this inside $M$ since the ordinals of $M$ might not really be well-ordered; $M$ might just think they are. This could happen if there is a set which has no least element, but that set is not in $M$.

By induction we say $X \sim_\alpha Y$ if for every $x \in X$ there is a $y \in Y$ such that $x \sim_\beta y$ for some $\beta < \alpha$, and similarly if we reverse the roles of $X$ and $Y$. Note that each $\sim_\alpha$ is actually an element of $M$, not necessarily a real set. Now external to $M$, define an equivalence relation on the set $M$ by $X \sim Y$ if $X \sim_\alpha Y$ for some $\alpha$. Let $U = M/\sim$.

Note that the relation $\epsilon$ is well-defined on $U$: if $x \sim y$ and $x \epsilon X$, then $X \sim X \cup \{y\}$ so $y \epsilon X$ in $U$. Similarly, if $x \epsilon X$ and $X \sim Y$, then there is some $y \epsilon Y$ such that $x \sim y$.

We now claim that $U$ is a model of $\text{ZFC}$. Extensionality is true basically by definition of $U$. For all the other axioms, we can construct the desired sets in $M$, and those same sets will work in $U$, with perhaps a little tweaking. For Separation and Replacement, we must be a bit careful about the interpretations of the defining formulas; wherever $x \in y$ appears we should replace it with $\exists z (z \sim x \land z \in y)$. With Replacement, this might make the values of functions no longer unique, but we can use Choice to pick one possible value for each input of the function (if there is a proper class of possible values, we pick one of minimal rank, which might make the values of functions no longer unique, but we can use Choice to pick one possible value). For Foundation, note that $\sim$ respects rank, so elements of $U$ have the same rank as they did in $M$ (so $V = \bigcup V_\alpha$ still holds). For axioms like Power Set, note that any $A$ which is a subset of $X$ mod $\sim$ is equivalent to some $B$ which is actually a subset of $X$ in $M$: for each $a \epsilon A$ there is some $x \epsilon x$ such that $a \sim x$; use Choice to pick a specific such $x_a$ for each $a$ and use Replacement to define $B = \{x_a\}$.

Since Extensionality says that everything is a set, to get a model without Extensionality we just have to add some things that are not sets. That is, we can add some atoms which have no elements but are not equal to the empty set.

**Theorem 10.14.** If $\text{ZFC}$ is consistent, so is $\text{ZFC} \setminus \{\text{Extensionality}\} \cup \{\neg \text{Extensionality}\}$.

**Proof.** Let $(M, \epsilon)$ be a model of $\text{ZFC}$, let $A$ and $O$ be distinct elements of $M$ of the same rank. We want to build a new model in which $O$ acts like $\emptyset$ and $A$ is an atom. Define (internal to $M$) sets $U_\alpha$ (for $\alpha > 0$) by induction. Start with $U_1 = \{A, O\}$. Given $U_\alpha$, let $U_{\alpha+1} = \mathcal{P}(U_\alpha) \setminus \{\emptyset\} \cup U_\alpha$. For a limit, let $U_\alpha = \bigcup_{\beta < \alpha} U_\beta$.

Now external to $M$, define $U = \{x \in M : x \epsilon U_\alpha$ for some $\alpha \in \text{Ord}^M\}$. It is then easy to check that $U$ is a model of all of $\text{ZFC}$ except Extensionality, using that $M$ is a model of $\text{ZFC}$. For Separation and Replacement, you must require that all quantified variables should take values in $U$, but we can do this because $U$ is definable as a class in $M$. Note that either $O$ or $A$ can serve as the empty set in $U$, because by induction every element of $U$ has rank greater than that of $O$ or $A$ so they have no elements in $U$.

Extensionality fails because $A$ and $O$ are not equal but they both have no elements.

The last axiom to consider is Choice. Here, the proofs are much more difficult, and we mostly omit them.

**Theorem 10.15.** If $\text{ZF} = \text{ZFC} \setminus \{\text{Choice}\}$ is consistent, so is $\text{ZFC}$.

**Theorem 10.16.** If $\text{ZFC}$ is consistent, so is $\text{ZF} \cup \{\neg \text{Choice}\}$.

Here is the basic idea of the proof of Theorem 10.15. We define a hierarchy of sets $L_\alpha$ similar to the $V_\alpha$. We have $L_0 = \emptyset$ and $L_\alpha = \bigcup_{\beta < \alpha} L_\beta$ for $\beta < \alpha$. However, instead of having $L_{\alpha+1} = \mathcal{P}(L_\alpha)$, $L_{\alpha+1} = C(L_\alpha)$, where $C(X)$ is defined as the set of all subsets of a set $X$ which can be constructed from $X$ by iterating a certain finite set of operations a finite number of times. These operations are carefully chosen so that any
subset we could define (using Separation) could be constructed by using these operations cleverly. We write \( L = \bigcup_{\alpha \in \text{Ord}} L_\alpha \) and call \( L \) the universe of constructible sets.

It can be shown that \( L \) is transitive and all the axioms of ZF hold in \( L \), basically because the operations used in \( C(X) \) are powerful enough to do all the constructions that the axioms require. What’s more, there is a canonical well-ordering \( \prec \) of \( L \). For \( x \in L \), let \( \rho(x) \) be the least \( \alpha \) such that \( x \in L_\alpha + 1 \). If \( \rho(x) < \rho(y) \), we say \( x \prec y \). Now assuming \( \prec \) is defined on all of \( L_\alpha \), we extend it to \( L_\alpha + 1 \). Every set of \( L_\alpha + 1 \setminus L_\alpha \) can be written as some finitely iterated operations applied to elements of \( L_\alpha \). It follows that there is a natural injection \( L_\alpha + 1 \setminus L_\alpha \to \omega \times (\bigcup_{n \in \omega} L_\alpha^n) \); the element of \( \omega \) encodes exactly which iterated sequence of operations are used, and the element of some \( L_\alpha^n \) tells which elements of \( L_\alpha \) are taken as input into the operations. Since \( L_\alpha \) has a well-ordering \( \prec \), we can construct a well-ordering on \( L_\alpha^n \) for all \( n \) as in the proof of Theorem 5.3. From this, it is not hard to construct a well-ordering on \( \omega \times (\bigcup_{n \in \omega} L_\alpha^n) \). Our injection then induces a well-ordering on \( L_{\alpha+1} \setminus L_\alpha \), which we call \( \prec \).

Thus there is a well-ordering on all of \( L \), and in particular it restricts to a well-ordering on any element of \( L \). Thus \( L \models \text{Choice} \), so \( L \) is a model of ZFC.

But wait! We said before that a proper class does not count as a model, and \( L \) is certainly a proper class (it contains all ordinals, for example). However, remember that we’re not trying to prove that ZFC is consistent, we’re just trying to prove that if ZF is consistent, so is ZFC. So we can assume we have some model \( M \) of ZF, and then take \( L^M \), the subset of \( M \) which \( M \) thinks is the class \( L \). Then \( L^M \) will be a model of ZFC.

In fact, \( L \) is a very special model of ZFC. Since \( L_\alpha + 1 \) is obtained from \( L_\alpha \) by just applying finitely many operations a finite number of times, \( |L_{\alpha+1}| \leq |L_\alpha| \cdot \aleph_\alpha \). In particular, for example, it follows that \( |L_\omega| = \aleph_0 \) for any countably infinite \( \alpha \).\(^7\) Now it can be shown that every subset of \( \omega \) which is in \( L \) is actually in \( L_\omega \). This is basically because any transfinite procedure to define a subset of a countable set must at some step have decided whether each element is in the subset or not. In particular, the procedure has countable cofinality, and thus will run out before you reach \( \omega \).

Now \( L_\omega = \bigcup_{\alpha < \omega} L_\alpha \), and each \( L_\alpha \) is countable. Thus \( |L_\omega| \leq \aleph_1 \cdot \aleph_0 = \aleph_1 \). But we said above that in \( L \), \( \mathcal{P}(\omega) \subseteq L_\omega \). Thus \( 2^{\aleph_0} \leq \aleph_1 \) so \( 2^{\aleph_0} = \aleph_1 \), i.e. the continuum hypothesis is true in \( L \)! In fact, similar arguments can be used to show that the generalized continuum hypothesis also holds in \( L \). Thus if ZF is consistent, ZFC \( \cup \{ \text{GCH} \} \) is also consistent.

The constructible universe was invented by Gödel and was the first method used to prove the consistency of Choice and CH (or GCH). Later, another method called forcing was discovered to give less ad hoc and more easily generalized proofs of the consistency of CH. This method will be the subject of Susan’s class on the Continuum Hypothesis in weeks 3 and 4. Forcing is extremely versatile and can also be used to prove the consistency of a very wide variety of statements with ZFC (or ZF). In particular, forcing is the standard way of proving Theorem 10.16, the consistency of \( \neg \text{Choice} \).

### 10.1 Exercises

**Exercise 10.1.** Let \( M = \mathbb{Z} \), and define a total order \( \epsilon \) on \( M \) by \( 0 \epsilon 1 \epsilon 2 \epsilon \ldots \epsilon -3 \epsilon -2 \epsilon -1 \).

(a): Show that the relation \( \epsilon \) is not well-founded (i.e., there exists some subset \( S \subseteq M \) that does not have a \( \epsilon \)-minimal element).

(b): Show that \( M \models \text{Foundation} \).

Exercise 10.1 shows how a model can think it is well-founded without actually being well-founded. In fact, a basic theorem of model theory implies that if ZFC is consistent, then there do exist models of ZFC which are not actually well-founded.

The following exercise shows that if the \( \epsilon \) relation of a model really is well-founded (rather than the model just thinking it is well-founded), then we may as well assume \( \epsilon = \in \). Recall that a set \( S \) is transitive if \( x \in S \) and \( y \in x \) implies \( y \in S \).

---

\(^7\)This is in stark contrast with the \( V_\alpha \); \( V_{\omega+1} = \mathcal{P}(V_\omega) \) is already uncountable!
Exercise 10.2. Let \((M, \epsilon)\) be a set together with a well-founded relation \(\epsilon\). Define a function \(f\) on \(M\) called the Mostowski collapse by \(\epsilon\)-induction: if we have defined \(f(y)\) for all \(y \in x\), define \(f(x) = \{f(y) : y \in x\}\). Let \(U = f(M)\).

(a): Show that \(U\) is transitive.
(b): If \(M\) is transitive and \(\epsilon = \epsilon\), show that \(U = M\) and \(f(x) = x\) for all \(x\).
(c): If \(M \models\) Extensionality, show that \(f : M \to U\) is an isomorphism, i.e. a bijection such that \(x \epsilon y\) iff \(f(x) \in f(y)\).
(d): Show that if \(M \models\) Extensionality, then \(f : M \to U\) is the only isomorphism from \(M\) to a transitive set.

That is, a well-founded model of Extensionality is uniquely isomorphic to a transitive set with \(\epsilon = \epsilon\). In particular, if there exists a well-founded model of ZFC, then there exists a transitive model (with \(\epsilon = \epsilon\)).

If you solve Exercise 10.3, you will earn \(N\) Waffle Points. It is hard; I at one point spent several days thinking about it and couldn’t figure it out!

Exercise 10.3. Can you construct a model in which Separation and Replacement both fail but the other axioms all hold? The idea would be that all the other axioms only assert that your model is closed under finitely many operations (taking power sets, taking unions, etc.). All you have to do is iterate these operations countably many times and take the union. More precisely, we start with some set \(U_0\) (say, \(V_\omega \cup \{\mathbb{N}\}\)). Given \(U_n\), we define a larger set \(U_{n+1}\) by adding in the pairs, unions, power sets, etc. of all elements of \(U_n\). We then let \(U = \bigcup U_n\), and it is not too hard to show that \(U\) satisfies every axiom except Separation and Replacement (though you have to be careful about making sure Extensionality holds, since when you add a set you should not add all of its elements).\(^8\)

However, I don’t know how to prove that Separation fails in such a \(U\). Note that this \(U\) will be countable: we took the countable set \(U_0\), and then applied finitely many operations to its elements to get new elements of \(U_1\), so \(U_1\) is also countable. Similarly every \(U_n\) is countable, and then \(U\) is countable. Note that \(U\) contains \(\mathcal{P}(\mathbb{N})\), but does not contain all of its elements (since there are uncountably many of them). Thus \(U\) is not transitive.

Since \(U\) contains \(\mathbb{N}\) but not all subsets of \(\mathbb{N}\), you might think that clearly Separation fails in \(U\). But Separation only says that any subset you can define must exist. The way you prove \(\mathcal{P}(\mathbb{N})\) is uncountable is by taking any countable subset \(S \subseteq \mathcal{P}(\mathbb{N})\) and constructing some new set \(A \subseteq \mathbb{N}\) which is not in \(S\). However, if we don’t know that the set \(S\) (and a bijection between it and \(\mathbb{N}\)) exists, we can’t define \(A\)! So it could be that Separation really does hold, and it’s OK because in \(U\), there is no bijection between \(\mathbb{N}\) and \(\mathcal{P}(\mathbb{N}) \cap U\). I find this highly unlikely, but I can’t prove it’s impossible. Can you?

\(^8\)You might ask, why can’t we do this with Separation and Replacement too? After all, they are only countably many more axioms, so we should be able to construct a model of all of ZFC by taking the closure under taking sets you can define with Replacement and Separation in addition to the operations for the other axioms.

The reason that this fails is that which set a definition by Separation or Replacement is referring to depends on the entire model that you’re in. For example, suppose the continuum hypothesis is true, and define a set \(A = \{S : S \subseteq \mathcal{P}(\mathbb{N})\text{ and }|S| = \aleph_1\}\). Now just because the continuum hypothesis is true in the real universe, it doesn’t necessarily have to be true in your model. In particular, it could turn out in the sets you’ve put into \(U_n\) so far, there is no bijection from \(2^{\aleph_0}\) to \(\aleph_1\). Thus when we add \(A\) to \(U_{n+1}\), do we say that \(\mathcal{P}(\mathbb{N}) \subseteq A\) or not? If we don’t say \(\mathcal{P}(\mathbb{N}) \in A\), we might get in trouble later if there is a bijection from \(2^{\aleph_0}\) to \(\aleph_1\) in \(U_N\) for some \(N > n\). Despite this, it is in fact very very nearly possible to make this idea work with a bit more cleverness and something called Skolem functions. However, it still doesn’t quite work, and the reason it doesn’t work is closely related to the reason that \(V\) itself does not count as a model of ZFC.

The sets we can define with axioms other than Separation and Replacement, on the other hand, are simple and limited enough that problems like this cannot come up.