# 11. The Axiom of Choice

## Contents


1 Motivation 2

2 The Axiom of Choice 2

3 Two powerful equivalents of AC 5

4 Zorn's Lemma 5

5 Using Zorn's Lemma 7

6 More equivalences of AC 11

7 Consequences of the Axiom of Choice 12

8 A look at the world without choice 14

# 1 Motivation

Most of the motivation for this topic, and some explanations of why you should find it interesting, are found in the sections below. At this point, we will just make a few short remarks.

In this course specifically, we are going to use Zorn's Lemma in one important proof later, and in a few Big List problems. Since we just learned about partial orders, we will use this time to state and discuss Zorn's Lemma, while the intuition about partial orders is still fresh in the readers' minds. We also include two proofs using Zorn's Lemma, so you can get an idea of what these sorts of proofs look like before we do a more serious one later.

We also implicitly use the Axiom of Choice throughout in this course, but this is ubiquitous in mathematics; most of the time we do not even realize we are using it, nor do we need to be concerned about when we are using it. That said, since we need to talk about Zorn's Lemma, it seems appropriate to expand a bit on the Axiom of Choice to demystify it a little bit, and just because it is a fascinating topic.

Many of the examples contained below, specifically in the last three sections, involve material that is well outside the scope of this course, and the reader should not feel that they have to understand everything immediately. The intention is to plant some ideas in the readers' minds, so that when they encounter these topics in other mathematical context the Axiom of Choice will come back to mind. The reader is not expected to know what the Hahn-Banach Theorem is, for example, but when it comes up while studying functional analysis we hope the reader remembers hearing about it in this context.

**If the reader is not at all interested in learning about the Axiom of Choice, and only wants to learn those things that are directly relevant to this course, they only need to read sections 4 and 5.**

# 2 The Axiom of Choice

The Axiom of Choice can be stated in many ways, and there are a very large number of unrelated-looking statements that turn out to be equivalent to it. We give the most straightforward statement here, which requires a definition first.

**Definition 2.1.** *Let $\mathcal{A}$ be a non-empty set of non-empty sets. A function $f : \mathcal{A} \to \bigcup \mathcal{A}$ is called a* <u>*choice function for $\mathcal{A}$*</u> *if $f(A) \in A$ for all $A \in \mathcal{A}$.*

**Example 2.2.**

1. Let $\mathcal{A}$ be the set of countries on Earth, thinking of each country as a collection of cities. Then $\bigcup \mathcal{A}$ is the set of all cities on Earth, and the function $f$ that assigns to each country its capital city is an example of a choice function for $\mathcal{A}$.

2. (The classic example.) Let $\mathcal{A}$ be the collection of all pairs of shoes in the world. Then the function that picks the left shoe out of each pair is a choice function for $\mathcal{A}$.

3. Let $\mathcal{A} = \mathcal{P}(\mathbb{N}) \setminus \{\emptyset\}$. The function $f(A) = \min(A)$ is a choice function for $\mathcal{A}$.

4. In fact, we can generalize the above to any well-order! Let $(W, \leq)$ be a well-order, and let $\mathcal{A} = \mathcal{P}(W) \setminus \{\emptyset\}$. Again, the function $f(A) = \min(A)$ is a choice function on $\mathcal{A}$.

   (This is the essence of the proof that the Well-Ordering Theorem implies the Axiom of Choice. More on that later.)

Notice that in these examples, and in most other examples the reader might think of, the choice function $f$ is defined by some rule: "choose the left shoe", "choose the least element", "choose the capital city", and so on. In all of these examples we simply *define* the choice function. No matter how large or complicated the well-order in the last example is, the function $f$ that we defined makes perfect sense. No fancy axioms are necessary for situations like this.

What the Axiom of Choice does is guarantee that there always exist choice functions, even in the absence of such a "rule" for defining one easily. As a warm up, think of this small variation on the example about shoes above: let $\mathcal{A}$ be the collection of all pairs of *socks* in the world. It seems harder to define a choice function now, right? Most pairs of socks do not have a distinguished left or right sock. Now what do we do? Well, the Axiom of Choice is here to help us.

(AC) Let $\mathcal{A}$ be a non-empty set of non-empty sets. There exists a choice function for $\mathcal{A}$.

The statement seems simple, but this is a subtle idea. To a non-mathematician, we might summarize this statement as, "If you have a bunch of choices, each of which is possible to make, then you can make all of them." Two main questions should arise here.

**Why this not just obvious, and in particular why does it need to be an axiom?**
The Axiom of Choice seems obviously true to most people, because we imagine ourselves making choices and it seems easy. In reality, we tend to only imagine making finitely many choices, which actually *is* easy, or we imagine making infinitely many choices for which there is a simple rule like the examples above.

Getting into the formality of mathematical logic for just a minute, the statement that a set $B$ is non-empty amounts to the first-order formula $\phi(B) := (\exists x)\,(x \in B)$ being true. If this formula is true, one can actually find a witness to it. In other words, if a set is known to be non-empty, we can infer that there is something in it, give it a name, and work with it. This rule of inference an example of existential instantiation.

Now in mathematical logic, we can "conjoin" finitely many formulas with "and" symbols. That is, if $B_1, \ldots, B_n$ are sets and we know each of them is non-empty, the following is also a formula of first-order logic.

$$\phi(B_1) \wedge \cdots \wedge \phi(B_n).$$

(The symbol $\wedge$ means "and".) This allows us to find an element of each one, by instantiating a $n$-tuple of elements of the respective sets. We cannot, however, conjoin infinitely many statements. The system of logic in which mathematics is defined does not allow for infinite conjunctions. Roughly speaking, this is why we cannot simply produce an element from each set in an infinite collection of non-empty sets in general, but we can do it for any finite collection of sets.

AC needs to be an axiom simply because it cannot be proved from the other axioms. In fact, we have *proved* that we cannot prove it from the other axioms! It is such a natural idea, however, and the mathematical world without it is so weird and unpleasant, that mathematicians decided long ago to assume it is true all the time. Attempts have been made to design a better axiom—one which gives us all or most of the nice things that AC does but without all the weird things—but none have proved worthy of replacing AC.

**Why do people care about not using it sometimes?**

This is a very interesting question. The answer is essentially that the AC introduces "non-constructiveness" into mathematical proofs. Proofs that make no use of AC are usually very constructive, in the sense that if they prove some object exists they tend to *explicitly* construct it. Proofs that do use AC often involve steps where we just ask AC to give us elements of sets, and then we work with them even though we know nothing else about them. Some mathematicians prefer to do things as constructively as possible, and an easy way to do that is to avoid using AC as much as possible.

Digressing into algebra for a bit, we can get some sense for this. In linear algebra, we say that two vector spaces are canonically isomorphic if one can define an isomorphism between them without choosing bases for either space.

A good example of this involves dual spaces. Recall that if $V$ is a real vector space, its dual space is the collection $V^*$ of all linear functions $f : V \to \mathbb{R}$, which is itself a real vector space using the natural definitions of function addition and scalar multiplication. If $V$ is finite dimensional then $V$ is always isomorphic to $V^*$, but defining an actual isomorphism requires selecting a basis of $V$ or defining an inner product on $V$, which involves using choice.

On the other hand, every vector space $V$ (regardless of its dimension) is *canonically isomorphic* to its double dual $V^{**}$—the dual of its dual. The map that sends $v \in V$ to the function $g_v \in V^{**}$ defined by $g_v(f) = f(v)$ is a canonical isomorphism—an isomorphism that does not require having any specific elements of $V$ to work with (such as a basis), and whose definition is particularly elegant. We prove these spaces are isomorphic by simply *defining* an isomorphism, rather than having to first fix bases or anything else.

# 3 Two powerful equivalents of AC

**Theorem 3.1.** *The following are equivalent:*

1. *The Axiom of Choice.*
2. *The Well-Ordering Theorem.*
3. *Zorn's Lemma.*

We will not prove this theorem in this note. At least not all of it; we more or less proved above that (2) $\Rightarrow$ (1). To be clear:

*Proof that* (2) $\Rightarrow$ (1). Let $\mathcal{A}$ be a non-empty collection of non-empty sets. Let $W = \bigcup \mathcal{A}$, and let $\leq$ be a well-ordering of $W$. Then the function $f : \mathcal{A} \to W$ defined by $f(A) = \min(A)$ is a choice function on $\mathcal{A}$. □

The most natural proof that (1) $\Rightarrow$ (2) involves defining a bijection between $\bigcup \mathcal{A}$ and a well-ordered set, which requires knowing about cardinal and ordinal numbers, falling outside the scope of these lectures.

The reader should take a moment to think about the gap in intuitiveness between the Axiom of Choice and the Well-Ordering Theorem. As we said earlier, most people feel that the Axiom of Choice is obviously true. On the other hand, the Well-Ordering Theorem seems to be obviously false to most people, because for example we cannot even imagine well-orderings of very familiar sets like $\mathbb{R}$.

As for Zorn's Lemma...

# 4 Zorn's Lemma

Zorn's Lemma is an unusual-looking statement about partial orders. Luckily, we recently learned about partial orders, so before going on, the reader should remind themselves of the definition of a partial order, and the definition of a chain.

We will need two more definitions before continuing, but these are easy ones; they mean what they sound like they mean.

**Definition 4.1.** *Let $(\mathbb{P}, \leq)$ be a partial order, and let $A \subseteq \mathbb{P}$. An element $p \in \mathbb{P}$ is an upper bound for $A$ if $a \leq p$ for all $a \in A$.*

**Definition 4.2.** *Let $(\mathbb{P}, \leq)$ be a partial order. An element $m \in \mathbb{P}$ is said to be maximal if there is no $p \in \mathbb{P}$ such that $m < p$.*

One subtlety in the second definition is that maximal elements are not necessarily above everything in the partial order, but rather nothing in the partial order is above them. A partial order can have many maximal elements. Speaking of which, here is a simple exercise:

**Exercise 4.3.** Suppose $(\mathbb{P}, \leq)$ is a partial order and $p, q \in \mathbb{P}$ are distinct maximal elements. Show that $p$ and $q$ are incomparable.

These concepts are intuitive, but here are some examples just to be sure:

**Example 4.4.**

1. Let $X$ be a set, and consider the partial order $(\mathcal{P}(X), \subseteq)$. $X$ is a maximal element in this order, and there are no others. In fact, $X$ is a "global" maximal element, in the sense that it is above every element of the partial order.

2. $(\mathbb{N}, \leq)$ has no maximal elements.

3. $\omega$ is a maximal element of $\omega + 1$.

4. Let $\preceq$ be the relation "is divisible by" on $\mathbb{N}$. Note, this is the reverse of the divisibility relation discussed in the notes on orders. To be clear, we mean

$$n \preceq m \text{ if and only if } m|n,$$

so for example $15 \preceq 3$, $10 \preceq 5$, and so on. Larger numbers are lower in this order.

In the partial order $(\mathbb{N} \setminus \{1\}, \preceq)$, every prime number is a maximal element. In $(\mathbb{N}, \preceq)$, 1 is the unique (and global) maximal element.

With these concepts solidified, here is the statement of Zorn's Lemma.

**Theorem 4.5** (Zorn's Lemma)**.** *Let $(\mathbb{P}, \leq)$ be a non-empty partial order such that every chain in $\mathbb{P}$ has an upper bound. Then $\mathbb{P}$ has a maximal element.*

We will not prove this here. It is relatively easy to show that Zorn's Lemma implies the AC (we will mention that later), but the proof that AC implies Zorn's Lemma is quite involved.

The statement of this theorem might look cryptic at the moment, but this is natural. In fact there is a famous saying about the result of Theorem 3.1:

> The Axiom of Choice is obviously true, the Well-Ordering Principle is obviously false, and nobody knows about Zorn's Lemma.

The strength of Zorn's Lemma is hard to see at first. The Axiom of Choice and the Well-Ordering Principle, even seeing them for the first time, seem like powerful statements because they both explicitly say something about *all sets*. For example, the Well-Ordering Principle says that *every set* can be well-ordered. Zorn's Lemma seems to only be saying something about partial orders with a specific, unfamiliar property. The strength of Zorn's Lemma is hidden in the fact that a great many things can be *coded* by or with partial orders.

 

# 5 Using Zorn's Lemma

In this section we state and prove two results, to give you a feeling for how Zorn's Lemma is used to prove things. Both of these proofs are simple ones, as Zorn's Lemma proofs go. These proofs will be unusually wordy, so that the reader can clearly see the justifications for every step.

When we set out to use Zorn's Lemma to prove some object exists, our goal is to design a partial order $(\mathbb{P}, \leq)$ such that a maximal element of $\mathbb{P}$ gives us (and in most cases actually *is*) the object we are looking for. Your intuition should be that we are trying to construct or find some sort of complicated object; one that contains a large or somehow maximal amount of information. Our partial order will usually consist of partial (often finite) approximations to this complicated object, ordered by inclusion, so that going up in the partial order corresponds to containing more information. You will get a feeling for this during the next two proofs.

**Corollary 5.1.** *Every vector space has a basis.*

This is a result you certainly took for granted when you studied linear algebra, but if you think about it, it is not at all obvious. It *is* obvious for finite-dimensional vector spaces, in the same way that finite choice is true. In a finite-dimensional vector space we can pick a vector, then pick another vector outside of its span, then pick a third vector outside of the span of the first two, and continue in this way until we cannot pick any more vectors (the number of vectors you have picked is the dimension of the space). No AC or Zorn's Lemma needs to be involved there. As with AC, this result is only non-obvious when we have a very large vector space. For example, consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Not much hope of doing that same proof here.

This result has nothing to do with topology, of course, but we choose to present it here because it is a very simple use of Zorn's Lemma in the familiar setting of vector spaces.

*Proof of Corollary 5.1.* Let $V$ be a vector space. Recall that a basis of $V$ is a linearly indepedent collection of vectors that spans $V$. Also recall that a set of vectors $A \subseteq V$ is linearly independent if no element of $A$ can be expressed as a finite linear combination of other elements of $A$. Importantly for us, the failure of a set of vectors to be linearly independent is witnessed by *finitely many* vectors, even if the set itself is infinite. The reader should make sure they are clear on this point before proceeding; no matter how large a set of vectors $A$ is, if $A$ is not linearly independent then this is witnessed by *finitely many* elements of $A$.

Now, let $\mathbb{P} = \{ A \subseteq V : A \text{ is linearly independent} \}$, and order $\mathbb{P}$ with the usual inclusion relation $\subseteq$. Then $(\mathbb{P}, \subseteq)$ is a partial order. Note that $\mathbb{P}$ is non-empty, since any singleton is trivially linearly independent.

**Claim.** *If $B \in \mathbb{P}$ is maximal, then $B$ is a basis for $V$.*

*Proof.* Suppose $B \in \mathbb{P}$ is maximal. By definition of $\mathbb{P}$, $B$ is linearly independent, so it only remains to show that $B$ spans $V$.

Suppose for the sake of contradiction that $v \in V \setminus \operatorname{span}(B)$. Then $C := B \cup \{v\}$ is linearly independent, and therefore $C \in \mathbb{P}$. Clearly $B \subseteq C$, and so this contradicts the maximality of $B$. That is, $C$ is an element of $\mathbb{P}$ that is strictly larger than $B$, which is impossible since $B$ is maximal. □

With the claim established, it remains to show that $(\mathbb{P}, \subseteq)$ satisfies the hypotheses of Zorn's Lemma. Once we have shown this, Zorn's Lemma will tell us there is a maximal element, which will be our basis.

We already mentioned that $\mathbb{P}$ is non-empty, since for example singletons are linearly independent.

**Claim.** *Every chain in $\mathbb{P}$ has an upper bound.*

*Proof.* Take a minute to think about what a chain in $\mathbb{P}$ is. A chain $\mathcal{C} \subseteq \mathbb{P}$ is a collection $\mathcal{C} = \{ C_\alpha : \alpha \in I \}$ of linearly independent sets such that for all $\alpha, \beta \in I$, $C_\alpha \subseteq C_\beta$ or $C_\beta \subseteq C_\alpha$. In particular, notice that $\mathcal{C}$ is a collection of sets of vectors.

The result is vacuously true for empty chains, so let $\mathcal{C} \subseteq \mathbb{P}$ be a non-empty chain and define $X = \bigcup \mathcal{C}$. This $X$ is a set of vectors—the union of all the sets of vectors in $\mathcal{C}$. We show that $X$ is an upper bound for $\mathcal{C}$.

To do this, we must show two things: $X \in \mathbb{P}$, and $C \subseteq X$ for all $C \in \mathcal{C}$. The latter is immediate from the definition of $X$—we defined $X$ to be the union of all the elements of $\mathcal{C}$, so certainly every element of $\mathcal{C}$ is a subset of $X$. It remains only to show that $X$ is linearly independent, and therefore an element of $\mathbb{P}$.

Suppose for the sake of contradiction that $\{v_1, \ldots, v_n\} \subseteq X$ is a linearly dependent set of vectors (remember, any such set of witnesses must be finite). By definition of $X$, for each $i = 1, \ldots, n$, we have $v_i \in C_i$ for some $C_i \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, there must be some $k$ between 1 and $n$ such that $C_i \subseteq C_k$ for all $i = 1, \ldots, n$, since any *finite* linear order—in this case collection $C_1, \ldots, C_n$—has a largest element.

But then $\{x_1, \ldots, x_n\} \subseteq C_k$, contradicting the assumption that $C_k \in \mathbb{P}$, or in other words that $C_k$ is linearly independent. □

Therefore, by Zorn's Lemma $\mathbb{P}$, has a maximal element and that maximal element is a basis of $V$. □

Note that in this proof, we were looking for an object containing a "maximal" amount of information in some sense—it is linearly independent, and no larger set is linearly independent. To find it, we designed a partial order in which a maximal element has precisely that property.

The part of the second claim where we took the union of the chain $\mathcal{C}$ is a very common move in Zorn's Lemma proofs. We often do this, then have to prove that the resulting union is actually an element of the partial order. This usually involves a similar argument to the one we gave above, in which any witness to the union not being in $\mathbb{P}$ must actually occur in an element of the chain, which is an element of $\mathbb{P}$ by assumption.

Next, we prove a topological fact about $\omega_1$ using Zorn's Lemma. This proof can be done with elementary techniques, but we present it here as a simple use of Zorn's Lemma.

**Corollary 5.2.** *$\omega_1$ does not have the countable chain condition.*

*Proof.* Our goal here is to design a partial order in which any maximal element is an uncountable collection of mutually disjoint non-empty open subsets of $\omega_1$. Therefore, it is natural to use a partial order consisting of collections of mutually disjoint non-empty open sets, in hopes that a maximal one will have to be uncountable. Fortunately we actually do not have to worry about the cardinalities of the collections in our poset, but we do care about the size of each set.

Let $\mathcal{T}$ be the order topology on $\omega_1$, and let $\mathcal{U} \subseteq \mathcal{T}$ be the collection of all *countable* open subsets of $\omega_1$. Define

$$\mathbb{P} = \{\, \mathcal{A} \subseteq \mathcal{U} \setminus \{\emptyset\} \,:\, A \cap B = \emptyset \text{ for all } A, B \in \mathcal{A} \,\},$$

and order it with the usual inclusion relation $\subseteq$. This $\mathbb{P}$ is exactly what we described above—the partial order of all collections of mutually disjoint non-empty open subsets of $\omega_1$—with the additional requirement that each open set is countable. The elements of $\mathbb{P}$ need not be countable, but each element of $\mathbb{P}$ is a set of countable open subsets of $\omega_1$.

**Claim.** *If $\mathcal{M} \in \mathbb{P}$ is maximal, then $\mathcal{M}$ is uncountable.*

This claim is why we require the open sets in elements of $\mathbb{P}$ to be countable. We want maximal elements of this partial order to witness that $\omega_1$ is not ccc, but if we allow *any* open sets in the elements of the partial order, we get stuff like $\{\omega_1\} \in \mathbb{P}$ which is obviously maximal but not uncountable.

*Proof of Claim.* Suppose $\mathcal{M}$ is maximal, and suppose for the sake of contradiction that $\mathcal{M}$ is countable. Then $M := \bigcup \mathcal{M} \subseteq \omega_1$ is a countable union of countable sets, and is therefore countable. Using one of the basic facts about $\omega_1$, we can find an element $\alpha \in \omega_1$ such that $m \leq \alpha$ for all $m \in M$—an upper bound of $M$, in other words. But then we can easily find a countable open subset $U$ of $\omega_1$ that lives entirely above $\alpha$. For example if we define:

$$\alpha + 1 := \min\Big(\omega_1 \setminus (\operatorname{pred}(\alpha) \cup \{\alpha\})\Big),$$

then $U := \{\alpha + 1\} = (\alpha, \alpha + 1]$ is such an open set. Having found such a set, the collection $\mathcal{M} \cup \{U\}$ consists of mutually disjoint countable non-empty open sets, and strictly contains $\mathcal{M}$ (and is therefore strictly larger in the order on $\mathbb{P}$), contradicting the maximality of $\mathcal{M}$. □

So again, a maximal element of $\mathbb{P}$ does what we need; a maximal element of $\mathbb{P}$ is an uncountable collection of mutually disjoint non-empty open sets, witnessing that $\omega_1$ is not ccc. It again remains to show that $(\mathbb{P}, \subseteq)$ satisfies the hypotheses of Zorn's Lemma, which will then furnish us with a maximal element.

Before reading on, take a moment to verify that $\mathbb{P}$ is non-empty.

**Claim.** *Every chain in $\mathbb{P}$ has an upper bound.*

*Proof.* Let $\mathcal{C} \subseteq \mathbb{P}$ be a non-empty chain. We again claim that $X := \bigcup \mathcal{C}$ is an upper bound for $\mathcal{C}$. Take a moment here to consider what each of these sets consists of. The elements of $\mathbb{P}$ (and in turn the elements of $\mathcal{C}$) are sets of open subsets of $\omega_1$. Therefore $\mathcal{C}$ is a set of sets of open subsets of $\omega_1$, and in turn $X$ is a set of open subsets of $\omega_1$. This sort of thing can get confusing, and drawing a picture may be helpful.

Again, it is obvious from the definition of $X$ that $C \subseteq X$ for all $C \in \mathcal{C}$. It remains only to show that $X \in \mathbb{P}$, or in other words that $X$ is a collection of mutually disjoint countable open subsets of $\omega_1$.

Since every open set in $C$ is non-empty and countable for every $C \in \mathcal{C}$, it is clear that all the elements of $X$ are non-empty and countable. The only thing that can go wrong is the disjointness. So suppose for a contradiction that $U \cap V \neq \emptyset$ for some $U, V \in X$. By definition of $X$, $U \in C_1$ and $V \in C_2$ for some $C_1, C_2 \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, we must have that $C_1 \subseteq C_2$ or $C_2 \subseteq C_1$. Suppose without loss of generality that $C_1 \subseteq C_2$. This means $U, V \in C_2$, and therefore they must be disjoint since $C_2 \in \mathbb{P}$.

This shows that $X \in \mathbb{P}$, and therefore that it is an upper bound for $\mathcal{C}$. □

□

There we have two relatively straightforward applications of Zorn's Lemma. They may have been a little bit tricky to follow due to the many "layers" of sets involved. The chain $\mathcal{C}$ in the last proof, for example, was a set of sets of open subsets of $\omega_1$. That's confusing! But otherwise, the arguments are very simple.

In general, when a partial order $\mathbb{P}$ consists of collections of objects ordered by inclusion, as the partial orders in both of these examples did, this idea of taking the union of a chain to form an upper bound usually works.

Many of the most familiar uses of Zorn's Lemma are similar to these. For example, Zorn's Lemma can be used to prove the following facts. The reader should not be concerned if they do not know what some of these words mean, as many of them are outside the scope of this course.

- In a ring $R$, every ideal $I$ is contained in a maximal ideal. To prove this, you fix a ring $I$ and form the partial order $\mathbb{P} = \{ J \subseteq R : J \text{ is an ideal on } R \text{ and } I \subseteq J \}$, ordered by inclusion. The hypothesis of Zorn's Lemma is easily satisfied since the union of a nested collection of ideals is again an ideal.

- Every field has an algebraic closure. This is easy to prove by [regular] induction for countable fields, since there are only countably many polynomials over a countable field.

For a general field, we need something like Zorn's Lemma. In this case, Zorn's Lemma is used to prove that every field $F$ is contained inside an algebraically closed field $K$, from which point it is simple to define the algebraic closure of $F$ as a subfield of $K$.

- The Well-Ordering Theorem: every set can be well-ordered. To prove this, fix a set $X$, and define the partial order $\mathbb{P}$ of well-orderable subsets of $X$. That is, $(Y, \leq) \in \mathbb{P}$ if and only if $Y \subseteq X$ and $\leq$ is a well-ordering of $Y$. These are like pieces of, or attempts at defining, the final well-order of $X$ that we want.

  We order these by saying that $(Y_1, \leq_1) \preceq (Y_2, \leq_2)$ if and only if the first ordering is an initial segment of the other. That is, $Y_1 \subseteq Y_2$, $\leq_2\restriction_{Y_1} = \leq_1$, and every element of $Y_2 \setminus Y_1$ is above every element of $Y_1$ according to $\leq_2$.

  Then $(\mathbb{P}, \preceq)$ is a partial order, and by taking unions of chains it is relatively simple to show that every chain has an upper bound. Then Zorn's Lemma furnishes us with a maximal element of $\mathbb{P}$, which is a well-order of all of $X$ (since if a maximal element $(M, \leq)$ of $\mathbb{P}$ omits some element $x \in X$, we can create a larger well-ordered subset $M \cup \{x\}$ of $X$ by declaring that $x$ is above every element of $M$.)

- The Axiom of Choice: every non-empty collection of non-empty sets admits a choice function. To prove this, fix a non-empty collection of non-empty sets $\mathcal{A}$, and define the collection of *partial* choice functions for $\mathcal{A}$. That is, choice functions that only make choices for some subcollection of the sets in $\mathcal{A}$. Order this collection by function extension ($f$ extends $g$ if $\mathrm{dom}(g) \subseteq \mathrm{dom}(f)$ and $f(x) = g(x)$ for all $x \in \mathrm{dom}(g)$). Again by taking unions of chains we can prove that every chain has an upper bound, and a maximal element of this partial order is a function that cannot be extended, or in other words one which makes a choice from every element of $\mathcal{A}$.

We finish this note with some novel equivalences and corollaries of the Axiom of Choice, along with some examples of how hideous the mathematical world is without the Axiom of Choice.

## 6 More equivalences of AC

**Theorem 6.1.** *The following are equivalent.*

1. *The Axiom of Choice.*
2. *The Well-Ordering Principle.*
3. *Zorn's Lemma.*
4. *Tychonoff's Theorem.*
5. *Every vector space has a basis.*

6. *Every nontrivial, unital ring has a maximal ideal.*
7. *Every non-empty set can be given a group structure.*
8. *If* $\{ A_\alpha : \alpha \in I \}$ *is a collection of non-empty sets, then their Cartesian product* $\prod_{\alpha \in I} A_\alpha$ *is non-empty.*
9. *Every surjection has a right inverse. That is, if* $f : X \to Y$ *is a surjection, then there is a function* $g : Y \to X$ *such that* $f(g(y)) = y$ *for all* $y \in Y$.

Some remarks about these:

4. This is a major theorem we will prove later in the course. It says that any product of compact topological spaces is compact.
5. We saw one direction of this above, but the other is very difficult and was only proved in 1984. You can read the proof in this paper.
7. Wikipedia has a nice treatment of this proof.
8. This is my favourite equivalence to AC. On the one hand it seems *so obvious* that a product of non-empty sets should be non-empty, but on the other hand it is *very easy* to prove this is equivalent to AC. An element of the Cartesian product is exactly a choice of an element from each $A_\alpha$.
9. A right inverse for $f$ is exactly a choice of an element from $f^{-1}(y)$ for each $y \in Y$.

# 7 Consequences of the Axiom of Choice

In this section we list some results that require the Axiom of Choice to prove. For each statement you understand, try to think about what the statement says and find the "non-constructiveness" contained within it.

**Theorem 7.1.** *The following results all require AC.*

1. *In a first countable topological space, if* $x \in \overline{A}$, *then there is a sequence of elements of* $A$ *converging to* $x$.
2. *A countable union of countable sets is countable.*
3. *Gödel's Completeness Theorem (at least the strongest version of it).*
4. *The Hahn-Banach Theorem (from real analysis).*
5. *Every* $T_{3.5}$ *topological space has a Stone-Čech compactification.*
6. *The Baire Category Theorem.*
7. *Every field has an algebraic closure.*

 

8. *The existence of non-Lebesgue measurable subsets of* $\mathbb{R}^n$.
9. *Every set has a well-defined cardinality.*
10. *The Banach-Tarski "Paradox".*
11. *If* $X$ *is an infinite set, then there exists an injection* $f : \mathbb{N} \to X$.
12. *Every filter is contained in an ultrafilter.*

Again, some remarks about these results:

2. Using the fact that a set is countable involves choosing a bijection from that set to $\mathbb{N}$. Choosing infinitely many such bijections requires AC.

3. This result is central to the study of logic (and in particular mathematical logic). One way of stating this result is that if you have some axioms and a statement that is true in any model of those axioms, then the statement is provable from those axioms.

   To give an analogy in more familiar terms, this is like saying that if some statement is true of every vector space, then you can *prove* it follows from the axioms that define vector spaces.

   This probably seems obvious, but that is because it is fundamental to the structure of first-order logic. Without this result, returning to our analogy, it could be that every vector space has some property $\phi$, while being impossible to *prove* that every vector space has $\phi$.

   You only need choice to prove the strongest version of this theorem. You often hear this theorem stated about well-orderable first-order languages, and choice enters the picture of you want to push this to *all* first-order langauges.

5. We will (hopefully) learn about this at the very end of the course. Roughly speaking, this says that any reasonably nice topological space can be embedded as a dense subset of a compact space called a compactification, and moreover that there is a particular compactification with many nice properties.

6. This appeared as a three-star problem in an early section of the Big List, but only for $\mathbb{R}$. It turns out to be true of any complete metric space, as we will soon see. Try to see the use(s) of choice in your proof for the reals.

8. This is one of the reasons people tried for so long to avoid using AC. Mathematicians at the time felt as though every subset of $\mathbb{R}^2$ should have a well-defined area, but this cannot be true if AC is true, however. Non-Lebesgue measurable subsets of $\mathbb{R}$ are called Vitali sets, and the proof that they exist is quite straightforward. You are encouraged to look into it yourself.

10. This is the most popular "weird" consequence of AC. Roughly speaking, it says that you can take a solid sphere in $\mathbb{R}^3$, split it up into five disjoint pieces, and reassemble those pieces into two exact copies of your original sphere without stretching or warping any of them. It certainly seems weird.

    The name is not accurate though as it is in no way a paradox. The weirdness comes from the existence of non-Lebesgue measurable subsets of $\mathbb{R}^3$. When we think of this as paradoxical, we are imagining that something was created from nothing—a whole sphere's worth of "stuff" seemed to come out of nowhere. The reason it is not a paradox is that when we split the sphere up into pieces to do this, some of them were necessarily non-measurable; they had no well-defined volume. So there is no reason to believe there should be any connection between the volume of the shape we start with and the volume of the shapes we end up with. In the intervening stages, all common-sense notions of volume were destroyed.

12. Filters and ultrafilters were defined in the supplementary notes on nets and filters. This proof is one of the classic uses of Zorn's Lemma. One defines the partial order of all filters that contain your given filter, and a maximal element is a filter that cannot be extended, which is precisely an ultrafilter.

## 8 A look at the world without choice

Weird consequences of AC like the Banach-Tarski Paradox are often cited as reasons we should not be happy using AC. The world without AC is not described nearly as often, and it is *awful*. Here are just a few examples the reader can use to scare people who think we should not assume AC.

**Proposition 8.1.** *The following things are possible if you do not assume AC (some of these require assuming you do not have any weak form of AC, like the Axiom of Countable Choice which says you can make countably many choices).*

1. *There exists an infinite set $X$ such that there does not exist an injection $f : \mathbb{N} \to X$.*
2. *Some sets $X$ can be partitioned into a collection of disjoint, non-empty pieces such that there are more pieces in the partition than elements of the set ("more" in the sense of cardinality).*
3. *A Cartesian product of non-empty sets might be empty.*
4. *There are vector spaces without bases.*
5. *The real numbers might be a countable union of countable sets.*

    *In this world, most of real analysis as we know it breaks down. Some things can be salvaged, but it's ugly. Note that without AC, this is not the same as saying that the reals are countable, since without AC a countable union of countable sets need not be countable.*

 

6. *In particular,* $\mathbb{R}$ *can be written as a union of two subsets with strictly smaller cardinality.*
7. *Every ultrafilter on* $\mathbb{N}$ *is principal.*