The axiom of choice and Zorn’s lemma

1 The axiom of choice

The axiom of choice is one of the basic axioms of set theory. Recall that, if $A$ is a nonempty set and $X_\alpha$ is another set for every $\alpha \in A$ (that is, $\langle X_\alpha \rangle_{\alpha \in A}$ is an indexed family of sets), then the Cartesian product $\prod_{\alpha \in A} X_\alpha$ is the set of all indexed families $\langle x_\alpha \rangle_{\alpha \in A}$ such that $x_\alpha \in X_\alpha$ for every $\alpha$. Formally, $\langle x_\alpha \rangle_{\alpha \in A}$ is really a function

$$A \longrightarrow \bigcup_{\alpha \in A} X_\alpha : \alpha \mapsto x_\alpha$$

which satisfies $x_\alpha \in X_\alpha$ for every $\alpha$.

The Axiom of Choice, ‘AC’. If $A$ is nonempty and $X_\alpha$ is nonempty for each $\alpha \in A$, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.

Most mathematicians feel strongly that this should be true: if all the sets in some collection are nonempty, then there must be a way of picking an element from each one! But famous theorems of Gödel and Cohen show that AC is logically independent of the other standard (meaning ‘Zermelo–Fraenkel’) axioms of set theory, and so we accept AC as a new axiom. There is some interest in what can be proved without using AC, but we do not address this matter in this course.

2 Partially ordered sets and Zorn’s lemma

The axiom of choice is rarely used in the form stated above. It is known to be equivalent to various other statements about ‘partially ordered sets’, and those formulations are much easier to apply in the uses that arise in real analysis.

Definition 2.1. Let $X$ be a set. A partial order on $X$ is a relation $\leq$ which is reflexive, antisymmetric and transitive. It is a total order if any two elements
$x, y \in X$ are comparable: either $x \leq y$ or $y \leq x$. Within a partially ordered set $(X, \leq)$, an element $x$ is **maximal** (resp. **minimal**) if the only $y \in X$ such that $y \geq x$ (resp. $y \leq x$) is $y = x$. If $A \subset X$ and $x \in X$, then $x$ is an **upper** (resp. **lower**) bound for $A$ if every $y \in A$ satisfies $x \geq y$ (resp. $x \leq y$).

**Examples 2.2.**

1. For any set $S$, the collection of its subsets $\mathcal{P}(S)$ is partially ordered by set inclusion $\subset$. This is not a total order unless $S$ is empty or a singleton.

2. The usual ordering of $\mathbb{R}$, or the restriction of that ordering to any subset of $\mathbb{R}$, is a total order.

3. Let $U \subset \mathbb{R}^2$, and for points in $U$ define
   \[(x_1, y_1) \preceq (x_2, y_2) \text{ to mean } [ y_1 \leq y_2 \text{ and } |x_2 - x_1| \leq |y_2 - y_1| ] .\]
   This is a partial order on $U$ (exercise!). Depending on $U$ it may have no maximal element or upper bound, several maximal elements but no upper bound, or an upper bound.

Another useful term: if $(X, \leq)$ is a partially ordered set, then a **chain** in $X$ is a subset $Y \subset X$ such that the restriction of $\leq$ to $Y$ is a total order.

Here are two of those statements that are equivalent to AC:

**Theorem 2.3** (The Hausdorff Maximal Principle). Any nonempty partially ordered subset $(X, \leq)$ has a maximal chain $Y$, meaning that if $Z$ is another chain and $Z \supset Y$ then $Z = Y$.

**Theorem 2.4** (Zorn’s lemma). If $(X, \leq)$ is a nonempty partially ordered set in which every chain has an upper bound, then $X$ has a maximal element.

I have written these as theorems to be deduced from AC. Each of them also implies AC, but I omit the proofs in that direction. In view of all these implications, some authors simply introduce Theorem 2.3 or 2.4 as axiom in the place of AC and proceed without proving any of these implications. I find this unsatisfactory: while AC seems very intuitive to me and I am happy to take it as an axiom, neither of Theorems 2.3 or 2.4 strikes me as obvious.

Another famous principle which is equivalent to AC is the ‘well ordering principle’. A proper account of this branch of set theory should really cover well orderings too: they greatly clarify the implications between all these various principles, and in some applications they are the most convenient re-formulation of AC. But we do not use them in this course, so I set them aside here.
3 Sketch proof of Hausdorff and Zorn from AC

We start with another result that we can deduce from AC (and which could in fact be taken as another equivalent axiom), and then show how it implies Theorems 2.3 and 2.4.

**Theorem 3.1.** Let $S$ be a nonempty set, and let $\mathcal{F}$ be a nonempty family of subsets of $S$, ordered by set inclusion. Assume that $\mathcal{F}$ is

i. downwards closed: if $A \in \mathcal{F}$ and $B \subset A$ then $B \in \mathcal{F}$;

ii. chain-closed: if $C$ is a chain contained in $\mathcal{F}$, then $\bigcup C \in \mathcal{F}$.

Then $\mathcal{F}$ has a maximal element.

**Proof sketch of Theorem 3.1.** Since $\mathcal{F}$ is nonempty and satisfies property (i), it contains $\emptyset$.

Suppose, towards a contradiction, that $\mathcal{F}$ has no maximal element. Then for every $A \in \mathcal{F}$ there exists a nonempty $B \subset S \setminus A$ such that $A \cup B \in \mathcal{F}$. By property (i) again, it follows that $A \cup \{s\} \in \mathcal{F}$ for every $s \in B$, so we may assume that $B$ is a singleton.

Thus, by AC, there is a *choice function* $f : \mathcal{F} \to S$ such that for every $A \in \mathcal{F}$ we have $f(A) \in S \setminus A$ and $A \cup \{f(A)\} \in \mathcal{F}$.

Let us call a sub-family $\mathcal{T} \subset \mathcal{F}$ a *tower* if

T1. $\emptyset \in \mathcal{T}$,

T2. if $A \in \mathcal{T}$ then $A \cup \{f(A)\} \in \mathcal{T}$,

T3. $\mathcal{T}$ is chain-closed.

Towers exist, because the whole of $\mathcal{F}$ is a tower (this is where we use assumption (ii) of the theorem). One can check immediately that any intersection of towers is a tower. As a result, the intersection of all towers is itself a tower, minimal in the sense that it is contained in any other tower. Call this intersection $\mathcal{T}_{\text{min}}$. It is nonempty by property T1.

The heart of the proof is to show that $\mathcal{T}_{\text{min}}$ is a chain, as well as a tower. To this end, call an element $A \in \mathcal{T}_{\text{min}}$ a *bottleneck*\(^1\) if it is comparable to every other element of $\mathcal{T}_{\text{min}}$: that is,

\[
\text{every } B \in \mathcal{T}_{\text{min}} \text{ satisfies either } B \subset A \text{ or } A \subset B. 
\]

\(^1\)not standard terminology
Claim. Every element of $T_{\text{min}}$ is a bottleneck.

Proof of the claim. We show that the set of bottlenecks in $T_{\text{min}}$ is itself a tower. It must then equal all of $T_{\text{min}}$ because $T_{\text{min}}$ is the minimal tower.

This assertion about bottlenecks follows from several smaller deductions that we leave as exercises to the reader:

1. Given a bottleneck $A \in T_{\text{min}}$, let $G(A)$ be the subset of all $B \in T_{\text{min}}$ which satisfy
   
   either $B \subset A$ or $A \cup \{f(A)\} \subset B$.

   Claim: $G(A)$ is a tower, and therefore $G(A) = T_{\text{min}}$, since $T_{\text{min}}$ is the minimal tower. [Hint: this part of the proof depends on the fact that $A \cup \{f(A)\}$ is larger than $A$ by exactly one element.]

2. Claim: if $A \in T_{\text{min}}$ is a bottleneck, then so is $A \cup \{f(A)\}$. [Hint: this part of the proof is based on the claim in Step 1.]

3. Claim: any union of a chain of bottlenecks is a bottleneck.

This completes the proof of the claim.

So every element of $T_{\text{min}}$ is a bottleneck, which shows that $T_{\text{min}}$ is a chain. Therefore $\bigcup T_{\text{min}}$ is a member of $T_{\text{min}}$, by property T3. But now, by property T2 $T_{\text{min}}$, must also contain the set

$$\bigcup T_{\text{min}} \cup \{f\left(\bigcup T_{\text{min}}\right)\},$$

except this set is properly larger than any element of $T_{\text{min}}$ — contradiction. □

Proof of Theorem 2.3. Let $F$ be the family of all chains in $X$. This family satisfies conditions (i) and (ii) of Theorem 3.1 (exercise: check!), and so $F$ has a maximal element — that is, $X$ has a maximal chain. □

Proof of Theorem 2.4. Consider an upper bound $x$ for a maximal chain $Y$ in $X$. This $x$ must be maximal in $X$: otherwise, we could pick another element $y \in X$ strictly above $x$, and then $Y \cup \{y\}$ would be a chain which properly includes $Y$ — contradiction. □
Remark. The arguments above are close to an early proof by Zermelo in which we first explicitly formulated AC. See Chapter 16 of Halmos' classic textbook *Naive Set Theory* or the Appendix to Rudin's *Real and Complex Analysis* for more complete accounts which also follow Zermelo's proof. [Note: Halmos’ book is freely available online within the UCLA network through Springerlink.]

4 Example: bases for vector spaces

**Proposition 4.1.** Every vector space has a basis.

**Proof.** Let $V$ be a vector space over a field $k$. A subset $C$ of $V$ is **linearly independent** if there are no nontrivial linear relations among the elements of $C$: more explicitly, whenever $v_1, \ldots, v_m \in C$ are distinct, the equation

$$\lambda_1 v_1 + \cdots + \lambda_m v_m = 0 \quad (\lambda_1, \ldots, \lambda_m \in k)$$

has only the trivial solution $\lambda_1 = \cdots = \lambda_m = 0$.

Let $F$ be the collection of all linearly independent subsets of $V$, and partially order $F$ by inclusion of sets.

- The collection $F$ is nonempty, because it contains $\emptyset$.
- If $C$ is a chain in $F$, then $\bigcup C \in F$. Indeed, given distinct vectors $v_1, \ldots, v_m \in \bigcup C$, there are elements $C_1, \ldots, C_m \in C$ such that $v_i \in C_i$. Since $C$ is a chain, one of these elements, say $C_i$, must be the largest: $C_i \supset C_j$ for every other $j = 1, 2, \ldots, m$. So $v_1, \ldots, v_m \in C_i$. Since $C_i \in F$, it follows that $v_1, \ldots, v_m$ are linearly independent.

Given these properties, Zorn’s lemma promises a maximal element $C$ of $F$. By the definition of $F$, $C$ is linearly independent. This particular $C$ must also span $V$ because of its maximality. Indeed, if any $v \in V$ could not be written as a linear combination in $C$, then $C \cup \{v\}$ would be a member of $F$ which is properly larger than $C$ — contradiction. So $C$ is a basis. \qed

**Remark.** In the application above, the properties of the family $F$ that we use are exactly as in Theorem 3.1, so we could have applied that result instead of Zorn’s lemma itself.

**Remark (Impossible exercise).** The set of real numbers $\mathbb{R}$ is a vector space over $\mathbb{Q}$, so it has a basis over $\mathbb{Q}$. Try to imagine what such a basis could possibly look like as a subset of $\mathbb{R}$.