SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The axiom of choice

av

Erik Adolfsson

2021 - No K17

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

The axiom of choice

Erik Adolfsson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Gregory Arone

2021

Acknowledgements

This work is largely based on the book Set theory and metric spaces by Irving Kaplansky [3]. Two great sources for inspiration before and during the project are the YouTube channels Vsauce by Michael Stevens and The Bright Side of Mathematics by Julian p. Grossmann.

Lastly I would like to say great thanks to my thesis advisor Gregory Arone for guiding me along this journey down the mathematical rabbit-hole.

Contents

1 A brief history **3**

2 Set theoretic background **3**
2.1 Basic notation 4
2.2 Partially ordered sets 5
2.3 Functions 6

3 The axiom of choice and equivalent statements **6**
3.1 The axiom of choice, AC 6
3.2 Zorn's lemma, ZL 8
3.3 The well ordering theorem, WO 8

4 Proof of the equivalence of AC, ZL and WO **9**
4.1 AC $\Longrightarrow$ ZL 9
4.2 ZL $\Longrightarrow$ WO 11
4.3 WO $\Longrightarrow$ AC 13

5 Applications **13**
5.1 Existence of bases in vector spaces 14
5.2 Maximal ideals in rings 15
5.3 Non-measurable sets and the Banach-Tarski paradox 16

1 A brief history

In the late 1800's mathematicians began the development of modern set theory with the intentions that it would be a solid foundation for all other parts of mathematics. While doing so they encountered inconsistencies and paradoxes within their formulations. One example is Russell's paradox which can be stated as follows: Let A be the set of all sets that are not members of themselves. Is A a member of itself? If A is a member of itself it would not fit the definition of A and thus not be a member of itself, on the other hand if A is not a member of itself then by definition it must be a member of itself.

These issues were solved by constructing an axiomatic theory, called Zermelo-Fraenkel set theory (ZFC). An axiomatic theory consist of a number of statements called axioms, which are regarded as true. A good axiomatic theory should be self-consistent, the axioms must not contradict each other. Secondly the theory should not contain any redundant information as an axiom, i.e: if a statement can be proven from other axioms it does not need to be an axiom itself. Thirdly the axioms should be strong enough to have interesting consequences.

ZFC consist of nine axioms. One of them, called the axiom of choice, has a special status among the other axioms. The axiom of choice says that given a collection of nonempty sets, it is possible to choose one element from each set. Even more concretely you could imagine jars with marbles in them, the axiom of choice lets you pick one marble from each jar. This might feel like a very obvious statement if we are considering finitely many jars. The power of the axiom of choice is that it can be applied to collections of arbitrarily many sets or jars.

The axiom of choice is regarded as the most controversial axiom of ZFC. This is because it has some consequences that seem counter intuitive or hard to accept. One example is that the axiom of choice implies that any set can be well ordered, which we will also show later. In this work we will investigate the axiom of choice and some of its consequences. We will not discuss any other axioms from ZFC. Instead we will take for granted that ZFC enables us to use the basic concepts needed from set theory.

2 Set theoretic background

Before we start talking about the axiom of choice and its equivalent statements I would like to present some of the concepts and tools from set theory which will be important for the later discussions.

2.1 Basic notation

A set is uniquely determined by its elements or members. We use the notation $x \in A$ to denote that x is a member of the set A. Two sets, A and B, are equal if they have the same elements. In symbols we say $A = B$ if and only if $x \in A \iff x \in B$ for all x. Curly brackets containing a list of elements can be used to represent a set, for example $\{1, 2, 3\}$ and $\{2, 4, 6, ...\}$ where the dots mean that the pattern extends infinitely. We will construct sets by inferring some rule that determines its members. For example can we construct the natural numbers from the integers in the following way: $\mathbb{N} = \{x \in \mathbb{Z} \mid x \geq 0\}$. To represent a set of sets we can use the notation $\{A_i\}$, where there is one set A_i for every i in some index set I. The empty set is a set that has no members and we denote it by $\emptyset$ or $\{\}$.

We adopt the following convention regarding notation for subsets: $A \subset B$ will be used when A is a proper subset of B which means that every element in A is also an element of B and that $A \neq B$. In the case that equality may also hold we use $C \subseteq D$ which means that either $C \subset D$ or $C = D$.

Definition: Let A be a set. The *power set* of A is the set of all subsets of A. We write the power set of A as $\mathcal{P}(A)$.

Example: If $A = \{x, y, z\}$ then
$\mathcal{P}(A) = \{\emptyset, x, y, z, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}$. Note that $\emptyset \in \mathcal{P}(A)$ and $A \in \mathcal{P}(A)$, which is true in general for power sets. Another thing to note is that $\mathcal{P}(A)$ has more elements than A, this is also true in general for sets.

The two most basic operations on sets are intersection and union. We write the intersection of A and B as $A \cap B$ and it is equal to the set of all x such that $x \in A$ and $x \in B$. Or in symbols: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

The union of two sets C and D is defined as the set of all x such that $x \in C$ or $x \in D$ and is written $C \cup D$. In symbols: $C \cup D = \{x \mid x \in C \text{ or } x \in D\}$. The union uses inclusive or which means that if x is in both C and D it is also an element of their union.

Example: For the real and natural numbers we have $\mathbb{N} \subset \mathbb{R}$ and thus their intersection and union is $\mathbb{R} \cap \mathbb{N} = \mathbb{N}$ and $\mathbb{R} \cup \mathbb{N} = \mathbb{R}$ respectively.

As a numerical example we have $A = \{x, y, z\}$, $B = \{a, b, x, y\}$. This gives $A \cap B = \{x, y\}$ and $A \cup B = \{a, b, x, y, z\}$.

Definition: Let U be a set and $A \subseteq U$. The *complement* of A (in U) is written A^c and is defined by $A^c = \{x \in U \mid x \notin A\}$.

Often the set U is implied by the context and does not need to stated explicitly.

Definition: Let A and B be sets. The *set difference* is denoted by $A \setminus B$ and is defined by $A \setminus B = \{x \in A \mid x \notin B\}$. The difference $B \setminus A$ is defined analogously.

2.2 Partially ordered sets

We are now going to start working with the symbols $\leq$ and $\geq$. Their meaning is very intuitive in the real numbers but we will generalise this notion so that we can compare elements of any set.

Definition: Let T be a set equipped with a relation $\leq$. We say that T is a partially ordered set if the following properties hold for all $a, b, c \in T$:

1. $a \leq a$.
2. If $a \leq b$ and $b \leq a$, then $a = b$.
3. If $a \leq b$ and $b \leq c$, then $a \leq c$.

When talking about partially ordered sets we may write $a < b$ when $a \leq b$ and $a \neq b$.

Example: The set of real numbers with their natural ordering is a partially ordered set.

The power set $\mathcal{P}(A)$ of any set A together with the inclusion relation $\subseteq$ is also a partially ordered set.

Definition: A partially ordered set T is called a *chain* if it has the property that for every a and b in T either $a \leq b$ or $b \leq a$ holds.

A chain T may also be referred to as a *totally ordered* set.

Example: The set of real numbers satisfy the stronger definition of a chain. The power set $\mathcal{P}(A)$ of a given set A with the inclusion relation $\subseteq$ is not a chain however. For subsets B, C of A neither $B \subseteq C$ nor $C \subseteq B$ need to be true. One example is if B and C are disjoint, i.e. $B \cap C = \emptyset$. Then the two sets have no elements in common and clearly none of them could be a subset of the other.

Definitions: Let T be a partially ordered set and let U be a subset of T. We say that an element $x \in T$ is an *upper bound* of U if $u \leq x$ for all $u \in U$. If in addition to this x also has the property that $x \leq v$ where v is any upper bound of U, we call x the *least upper bound* of U.

We also define *lower bound* and *greatest lower bound* in a similar fashion. Let $y \in T$, then y is a *lower bound* of U if $u \geq y$ for all $u \in U$. If y also fulfills that $y \geq w$ for any lower bound w of U, then y is the *greatest lower bound* of U.

Least upper bound and greatest lower bound are also called supremum and infimum respectively.

2.3 Functions

As I will formulate it in this paper the axiom of choice is a statement about the existence of a certain function. With this in mind we will use set theory to define the necessary concepts to work with functions.

Definition: An *ordered pair* (x, y) is defined by $(x, y) = \{\{x\}, \{x, y\}\}$.

In a set the order of the elements does not matter but with this definition we have constructed an object that has an intrinsic order. Namely this definition ensures that if $(x, y) = (x', y')$ then we must have $x = x'$ and $y = y'$. This notion of an ordered pair will enable us to define functions. First however we will define the Cartesian product of two sets.

Definition: For two sets A and B their *Cartesian product* is denoted as $A \times B$ and defined by $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$.

Definition: Let A and B be sets. The subset $G_f \subseteq A \times B$ is called a *function* if the following is true: For all $x \in A$ and all y, $y' \in B$ we have that $(x, y) \in G_f$ and $(x, y') \in G_f$ implies $y = y'$.

This definition ensures that every input x can get mapped to at most one output y.

If $G_f \subseteq A \times B$ is a function and additionally for all $x \in A$ there is a $y \in B$ such that $(x, y) \in G_f$ we may use the (hopefully familiar) notation $f : A \to B$ and $f(x) = y$.

Suppose that X is a subset of A, then $f(X)$ is the set of all $f(x)$ in B such that $x \in X$. We call $f(X)$ the *image* of X.

3 The axiom of choice and equivalent statements

3.1 The axiom of choice, AC

We are now ready to state the axiom of choice:

The axiom of choice: Let X be any set. Then there exist a function $f : \mathcal{P}(X) \to X$, defined on X, such that $f(A) \in A$ for every non-empty subset A of X.

This type of function is called a *choice function.* For a finite set X a choice function obviously exists, just complete the finite process of choosing one element from every (nonempty) subset of X.

For an infinite set it is not obvious whether we should allow this, thus one way to interpret the axiom of choice is that it enables you to complete an infinite process of "choosing". We state an alternative version of the axiom that we call AC*.

AC*: Let $\{X_i\}$ be a collection of disjoint nonempty sets. There exist a set Y such that every $y \in Y$ is an element of exactly one of the sets X_i and furthermore for every i there is a unique element $y' \in Y$ such that $y' \in X_i$.

In other words this means that the set Y has exactly one element in common with each X_i and contain no other elements. We now proceed to prove that AC is equivalent to AC*.

Proof AC $\iff$ AC*: First we prove that AC $\implies$ AC*. Let $\{X_i\}$ be a collection of disjoint nonempty sets where i are members of some index set I. Now form $Z = \bigcup_{i \in I} X_i$. From AC we have a choice function f on Z and we can let $Y = \bigcup_{i \in I} f(X_i)$. Since the sets X_i are disjoint every $f(X_i)$ is in exactly one X_i and for every i there is a unique element $f(X_i) \in Y$ such that $f(X_i) \in X_i$.

Next we prove AC* $\implies$ AC. Let W be a set. To be able to apply AC* we need to construct a collection of disjoint nonempty subsets. Let us consider $\mathcal{P}(W)$, this is however not a collection of disjoint sets.

Suppose that $j \in \mathcal{P}(W) \setminus \emptyset$. Let V_j be the set of ordered pairs (v, j) where $v \in j$. There is a one-to-one correspondence between V_j and j, namely $v \mapsto (v, j)$ and thus we can treat V_j as a copy of j.

Now we can form the set $\{V_j \mid j \in \mathcal{P}(W),\ j \neq \emptyset\}$ which is a collection of disjoint copies of all subsets of W. AC* now gives that there is a set T such that every $t \in T$ is in some V_j and exactly one element from V_j is in T for all $j \in \mathcal{P}(W) \setminus \emptyset$. Hence for every nonempty subset $j \subseteq W$ there exist exactly one element $(v, j) \in V_j$ that is also an element of T. Denote this specific v by v_j. By our definition of the ordered pairs (v, j) we have that $v_j \in j$. We define a function $f : \mathcal{P}(W) \setminus \emptyset \to W$ such that $f(j) = v_j$. Since $v_j \in j$ we get that $f(j) \in j$ and j is an arbitrary nonempty subset of W. Thus f is a choice function. $\square$

The axiom of choice has been proven to be independent of the other ZF axioms [1], meaning that ZF is consistent with the axiom of choice and its negation. We should also note that this is only a statement about the existence of a choice function and it tells us nothing about what the function may look like or how to find such a function.

3.2 Zorn's lemma, ZL

Next we wish to state Zorn's lemma. For this we have to introduce a new definition:

Definition: x is a *maximal element* of a partially ordered set L if $x < y$ is not true for any $y \in L$.

Since L is only required to be partially ordered the maximal element may not be unique. Intuitively a maximal element x is what it sounds like, namely there is no element that is "greater" than x. For example the real numbers have no maximal element and any closed interval has a maximal element.

Now we can state Zorn's lemma which, as I have previously mentioned, is equivalent to the axiom of choice.

Zorn's lemma: If L is a partially ordered set and furthermore every chain in L has an upper bound in L, then L contains a maximal element.

When we say every chain in L here we mean every subset of L that is a chain.

To get somewhat of an intuitive understanding of this lemma take any element $l_1 \in L$. If this is the maximal element ZL is satisfied. If not, there exist an element $l_2 \in L$ such that $l_1 < l_2$. If again l_2 is not maximal we can continue the process until we have the chain $l_1 < l_2 < l_3 < \dots$. By the assumption of ZL this chain has an upper bound l_{up} in L and we can now repeat the process starting with l_{up}. And we can again repeat this process until we find a maximal element of L.

Now we begin to see the connection to AC as ZL tells us that this process of finding "larger and larger" elements of L must come to an end, and thus there is a maximal element.

Another similarity between AC and ZL is that ZL gives only a statement about the existence of a maximal element with no clue about how to find it or what it looks like.

3.3 The well ordering theorem, WO

Our third statement equivalent to AC is WO and as the name suggest we need to define well-ordered sets before we can state the axiom.

Definition: A chain X is called *well-ordered* if every nonempty subset of X has a smallest element.

By smallest element of a set X we mean an element a such that $a \leq x$ for all $x \in X$.

Note that for example the natural numbers are well ordered with their natural ordering while the real numbers are not. For example, an open interval does not have a smallest element.

With this definition in mind we can state WO.

The well ordering theorem: Any set can be well ordered.

You might begin to notice a pattern, WO does not tell us how to find the ordering in question. Only that for any set there exist an ordering such that the set becomes well ordered.

Our example with the real numbers of course still holds. If we accept WO however we know that the real numbers can be well ordered, but that ordering is certainly very hard to find if not impossible.

We have now made three different statements about sets and our main goal now is to prove that these are all, in fact, equivalent.

4 Proof of the equivalence of AC, ZL and WO

In order to prove the equivalence of these three statements we will first prove AC $\Rightarrow$ ZL, then ZL $\Rightarrow$ WO and lastly WO $\Rightarrow$ AC. Thus the argument comes full circle and we have proved the desired equivalence.

4.1 AC $\implies$ ZL

Before this first proof we start with two definitions that will come in handy.

Definition: Given a partially ordered set T, a subset I is called an *ideal* in T if $x \in I$ and $y \leq x$ implies that $y \in I$.

Definition: Again let T be a partially ordered set and $t \in T$. Define $S(t)$ to be the set of all $x \in T$ such that $x < t$. $S(t)$ is then called the *segment* defined by t.

We will write $S_T(t)$ to specify that the segment is to be calculated in the set T. We state a lemma before proving the main result of the section.

Lemma 1: An ideal I in a well ordered set A is either all of A or a segment in A.

Proof lemma 1: Let I be an ideal in a well-ordered set A. Suppose that $I \neq A$. Since A is well ordered I^c has a smallest element a. We can now prove that $I = S(a)$. If $x < a$ then $x \in I$ because a is the smallest element that is not in I and hence every $x < a$ is in I. Conversely, $x \in I$ implies that $x < a$. This is because otherwise $a \leq x$ which implies $a \in I$ since I is an ideal. $\square$

Proof AC $\implies$ ZL: we take the conditions of ZL to be true: let L be a partially ordered set in which every chain has an upper bound. From this we need to prove that L has a maximal element. We will make a proof by contradiction so we begin by assuming the negated statemen, i.e. we assume that L has no maximal element.

For any chain C we let C' be the set of upper bounds u of C such that $u \notin C$. We can show that $C' \neq \emptyset$. For any chain C in L take u to be an upper bound of C. Our assumption was that L has no maximal element and thus there exists $v \in L$ such that $u < v$. Thus v is obviously also an upper bound of C and $v \notin C$.

Let f be a choice function on L (provided to us by AC) and define $g(C) = f(C')$. Hence g is defined on every chain in L and $g(C)$ is an upper bound of C which is not contained in C.

For the rest of the proof we fix an element $x_L \in L$. Let B be a subset of L with a well-ordering. We say that B is "special" if x_L is the smallest element of B and for any other element $y \in B$ we have $y = g(S_B(y))$. We now make a key claim: if B and D are special subsets of L, then either B is an ideal in D or D is an ideal in B. We now proceed to prove this key claim.

Let E be the set of elements x defined by the following two properties: $x \in B \cap D$ and $S_B(x) = S_D(x)$. For any $x \in E$ by definition every $y \in S_E(x)$ is also in E. I.e. $x_L \leq y \leq x$ implies $y \in E$ and thus E is an ideal in B and D.

First suppose that $E \subset B, D$, which means that E is neither equal to B nor D. Then, by lemma 1, $E = S_B(v)$ and $E = S_D(w)$ for some $v \in B$ and $w \in D$ respectively. We remember that B and D are special and thus $v = g(E)$ and $w = g(E)$ which gives $v = w$. By the definition of E we have $v \in E$ but also $E = S_B(v)$. By the definition of a segment $v \notin E$, a contradiction. Thus we can conclude that either $E = B$ or $E = D$. Since E is an ideal in B and D we have two cases: if $E = B$ then B is an ideal in D and vice versa if $E = D$. That concludes the proof of the key claim and we can carry on with the rest of the proof.

Let G be the union of all special subsets of L. G is a subset of L and thus it has a partial ordering that coincides with L. Our next goal is to prove that G is special.

We start by proving that G is a chain. Let $x, y \in G$, the definition of G gives that $x \in B$ for some special set B. Analogously $y \in D$ for some special set D. The key claim we proved says that one of B and D is an ideal in the other. Thus we have either $B \subseteq D$ or $D \subseteq B$. Hence we know that either x and y are both elements in B or both are elements in D. The sets B and D are special and hence they are well-ordered. The well ordering $\leq$ gives us that either $x \leq y$ or $y \leq x$ must hold. Thus G is a chain.

Before proving that G is well ordered we prove an intermediate claim. Again let $x \in G$ and hence $x \in B$ for some special set B. We will prove that for $y \in G$ such that $y < x$, y must be an element of B. In other words $S_G(x) = S_B(x)$. Since y is a member of G we have $y \in D$ for some special set D. The key claim now give that one of B and D is an ideal in the other. We treat the two cases separately. If D is an ideal in B we have $D \subseteq B$ and since $y \in D$ we get that $y \in B$. In the other case, where B is an ideal in D, we have $x \in B$ and $y < x$. The definition of an ideal now gives that $y \in B$.

We can now prove that G is well ordered. We need to prove that A has a minimal element for any nonempty $A \subseteq G$. Let x be an element of A. If x is a minimal element of A, we are done. Lets assume that x is not minimal in A. The segment $S_A(x)$ contains all elements of A smaller than x, hence if $S_A(x)$ has a minimal element y then it is also a minimal element of A. Thus it is enough to prove that $S_A(x)$ has a minimal element. Since $x \in A$ and hence $x \in G$ there is a special set B with $x \in B$. Now our intermediate claim tells us that every element of $S_A(x)$ is also an element of B, i.e. $S_A(x) \subseteq B$. Since B is well ordered all of its subsets has a minimal element. This proves that G is well ordered.

To finally prove that G is special we must show that x_L is the smallest element of G and that for any other $y \in G$, $y = g(S_G(y))$. Since G is a union of sets that all have the x_L as a minimal element, x_L must also be a minimal element of G.

Now let x be an element of G, we have $x \in B$ for some special set B. Since B is special, $x = g(S_B(x))$. The intermediate claim gives that $S_B(x) = S_G(x)$ and hence $x = g(S_G(x))$.

Let us now consider the set $G' = G \cup \{g(G)\}$. Since $g(G)$ is a strict upper bound of G the set G' is well ordered and additionally x_L is the smallest element of G'. All x in G that satisfied $x = g(S_G(x))$ will satisfy $x = g(S_{G'}(x))$ and we just have to check that this holds for the element $g(G)$. By the definition of G' we have $S_{G'}(g(G)) = G$ and thus $g(S_{G'}(g(G))) = g(G)$. Thus G' is a special set which contradicts the assumption that G is the union of all special subsets. By this contradiction we can conclude that L must have a maximal element. □

4.2 ZL $\Longrightarrow$ WO

We will need a special property of chains for the final proof and thus we start by proving two lemmas before dealing with the main course.

Lemma 2: A chain is well ordered if and only if it does not contain an infinite descending sequence.

Proof lemma 2: Any given chain C that is well ordered can not contain an infinite descending sequence as this would be a subset of C with no smallest element.

On the other hand, if we assume that C is not well ordered, there exist a set $B \subseteq C$ with no smallest element. We can now construct an infinite descending sequence. First pick any element b_1 in B. Since b_1 is not the smallest element we can pick another $b_2 \in B$ such that $b_2 < b_1$. There is also $b_3 \in B$ such that $b_3 < b_2$. Thus we have the infinite descending sequence $b_1 > b_2 > b_3 > \dots$. $\Box$

Lemma 3: If C is a chain and every segment in C is well-ordered, then C is well-ordered.

Proof Lemma 3: We will make a proof by contraposition. I.e. if C is not well ordered, Lemma 2 states that C contains an infinite descending sequence $b_1 > b_2 > b_3 > \dots$. The segment $S(b_1)$ is thus not well-ordered. $\Box$

Proof ZL $\implies$ WO: We start the proof by defining a certain set L that we will make use of in the rest of the proof.

Definition: Let A be a set. Now let L be the set of all ordered pairs (S, ρ) where S is a nonempty subset of A and ρ is a well-ordering on S.

For the rest of the proof when referring to A and L we mean the way they are defined above. Note that L contains multiple copies of any subset of A, each with a different well-ordering.

We will now proceed to prove WO by showing that A itself is a set such that $(A, \delta) \subset L$ for some well ordering δ. We define a partial ordering $\leq_L$ on L such that for (B, β), $(C, \gamma) \in L$ we take $(B, \beta) \leq_L (C, \gamma)$ to be true if and only if B is an ideal in C and γ coincides with β when restricted to B. In the rest of the proof we will only write out the sets in the relation $\leq_L$ and just remember that each of them have a well ordering.

We wish to apply ZL to the set L and hence need to prove that every chain in L has an upper bound. Let $\{B_i\}$ be a chain in L where i are elements of some arbitrary index set I. We claim that an upper bound for $\{B_i\}$ is $B = \bigcup_{i \in I} B_i$. To prove that B is an upper bound of the chain $\{B_i\}$ we must show that $B_i \leq_L B$ holds for all $i \in I$. To do this we must show that B has a well ordering and then prove that every $\{B_i\}$ is an ideal in B.

We begin by defining a partial order on B. For any $x, y \in B$ we know that $x \in B_j$ and $y \in B_k$ for some $B_j, B_k \in \{B_i\}$. Since $\{B_i\}$ is a chain in L either $B_j \leq_L B_k$ or $B_k \leq_L B_j$ must be true, let us say that $B_j \leq_L B_k$. In this case both x and y are elements of B_k and thus for $x \neq y$ either $x < y$ or $y < x$ must hold. Hence if we can prove that B is partially ordered we also automatically know that it is a chain. The argument works analogously if $B_k \leq_L B_j$. If you compare x and y in some other set B_l the same relation still holds. This is because $\{B_i\}$ is a chain with respect to the partial ordering defined on L and thus all sets B_i must agree on either $x < y$ or $y < x$. Note here that we do not use $\leq_L$ when comparing x and y but rather the well orderings of the different sets in $\{B_i\}$.

Now we can prove that B is partially ordered. $x \leq x$ holds because all B_i are partially ordered. If $x \leq y$ and $y \leq x$ this would hold in one $B_i \in \{B_i\}$ and since B_i is partially ordered we get $x = y$. To prove that $x \leq y$ and $y \leq z$ implies $x \leq z$ just take a B_i containing all three of them. Since B_i is partially ordered we get the desired property. We have thus proved that B is partially ordered and hence from our previous claim it is also a chain.

Let us now prove that B is well-ordered. By Lemma 3 it is enough to prove that every segment in B is well ordered. Let $x \in B_i$. It is enough to prove that $S_B(x) \subseteq B_i$ since B_i is well ordered. Thus we need to show that $y \in B_i$ for any $y \in B$ such that $y < x$. Since $y \in B$ we have $y \in B_j$ for some B_j. If $B_j \leq_L B_i$ we get $y \in B_i$. If $B_i \leq_L B_j$ we have from the definition of $\leq_L$ that B_i is an ideal in B_j. Since we have $x \in B$ and $y < x$ we get $y \in B_i$. This proves that B is well ordered.

This also proves that B_i is a segment in B for every $i \in I$.

We have now showed that every chain in L has an upper bound. Zorns' lemma now gives us that L has a maximal element M. Since M is an element of L it is also a well-ordered subset of A. We can now show that $M = A$ via a contradiction. So assume that $M \subset A$. Then there exist an element $w \in A$ such that $w \notin M$. we can construct another well-ordered set $M \cup \{w\}$ by taking the well-order relation on M and letting $x \leq w$ for all $x \in M$. We also get that M is an ideal in $M \cup \{w\}$. This contradicts the fact that M is the maximal element of A and thus we must have $M = A$. □

4.3 WO $\Longrightarrow$ AC

Proof: here we take WO to be true and we wish to prove AC. Let X be a set, from WO we know that X can be well ordered. For an arbitrary subset $A \subseteq X$ we define a function f such that $f(A)$ is the smallest element of A. The function f is well defined because X is well ordered and hence every subset A has a smallest element. Thus f is a choice function on X. □

5 Applications

Now we have went through a lot of work to prove that the axiom of choice is equivalent to Zorn's lemma and the well ordering theorem and we might ask ourselves why we even need the axiom of choice to begin with. To give some motivation for accepting the axiom of choice we will prove two theorems that require the axiom of choice. Or more accurately, we will use Zorn's lemma to prove two important theorems from linear algebra and ring theory. We will not give a complete introduction to the two subjects but rather just give enough background to prove the desired theorems. Lastly we will briefly discuss another consequence of the axiom of choice called the Banach-Tarski paradox.

5.1 Existence of bases in vector spaces

We wish to prove that every vector space has a basis. We explicitly define the concepts of linearly independent sets and bases.

Definition: Let V be a vector space over a field $\mathbb{F}$. A *finite* set of vectors $\mathbf{v}_i$ in V is *linearly independent* if $\sum_{i \in I} \lambda_i \mathbf{v}_i = 0$ implies that $\lambda_i = 0$ for all i. Where λ_i are elements of $\mathbb{F}$ and i are members of some finite index set I.

Definition: Let V be a vector space over a field $\mathbb{F}$. An *infinite* subset $W \subseteq V$ is said to be *linearly independent* if every finite subset of W is linearly independent.

Definition: Let V be a vector space over the field $\mathbb{F}$. A linearly independent subset B of V is called a *basis* if every vector $\mathbf{v}$ in V can be expressed as a linear combination of vectors in B.

Theorem 1: Every vector space has a basis.

Proof Theorem 1: Let V be a vector space over $\mathbb{F}$ and let A be the collection of all linearly independent subsets of V. The set A is partially ordered by inclusion. To use Zorn's lemma we need to prove that every chain in A has an upper bound in A. Let $\{X_i\}$ be a chain in A where i are members of some finite index set I. I.e. the sets X_i are linearly independent and for all $X_j, X_k \in \{X_i\}$ either $X_j \subseteq X_k$ or $X_k \subseteq X_j$ holds. We form the union $X = \bigcup_{i \in I} X_i$. The set X is an upper bound to $\{X_i\}$. Next we must show that X itself is linearly independent. We form the equation $\sum_{j \in J} \lambda_j \mathbf{x}_j = 0$ which we call $\star$, where $j \in J$ for some finite index set J, $\lambda_j \in \mathbb{F}$ and $\mathbf{x}_j \in X$. We will now make a proof by contradiction to show that $\star$ only has the trivial solution. So suppose there is a set $\{\lambda_j\}_{j \in J} \subseteq \mathbb{F}$ of scalars that are not all zero and satisfies equation $\star$. Since J is finite and $\{X_i\}$ is a chain there is an $m \in I$ such that $\mathbf{x}_j \in X_m$ for all $j \in J$. Since X_m is a member of $\{X_i\}$ it is linearly independent and hence $\lambda_j = 0$ in equation $\star$ for all $j \in J$.

Now that we have shown that X is linearly independent we can conclude that every chain in A has an upper bound in A. Hence Zorn's lemma gives that there is a maximal element in A, call it B. By the definition of A we get that B is the maximal linearly independent subset of V. We now wish to prove that B is a basis for V, i.e. we must show that any vector $\mathbf{v} \in V$ can be written as a linear combination of vectors in B. We make a proof by contradiction so suppose that $\mathbf{v}$ is not a linear combination of elements in B. Since B is linearly independent we have that the equation $\sum_{k\in K}\lambda_k\mathbf{b}_k = 0$ holds true only when all $\lambda_k = 0$. Here $\lambda_k \in \mathbb{F}$, $\mathbf{b}_k \in B$ and k are members of some finite index set K. Next we look at the equation $\lambda_v\mathbf{v} + \sum_{k\in K}\lambda_k\mathbf{b}_k = 0$, call it $\triangle$, where λ_v is some element of $\mathbb{F}$. Since $\mathbf{v}$ is not a linear combination of vectors in B we have $\sum_{k\in K}\lambda_k\mathbf{b}_k \neq \mathbf{v}$ for all $\lambda_k \in \mathbb{F}$ and $\mathbf{v} \neq 0$. Thus equation $\triangle$ has only the trivial solution where all λ-factors are zero. Hence the set $B \cup \{\mathbf{v}\}$ is linearly independent. This violates the maximality of B, we have a contradiction. □

5.2 Maximal ideals in rings

In this section we will prove that under certain conditions a ring always has a maximal ideal. We start by defining what we mean by an ideal in a ring.

Definition: Let $(R, +, \cdot)$ be a ring. We say that I is an *ideal* in R if it fulfills the following properties:
1. $(I, +)$ is a subgroup of $(R, +)$ with the addition operation inherited from $(R, +, \cdot)$.
2. For every $x \in R$ and every $a \in I$, $x \cdot a$ or $a \cdot x$ is in I.

Note that this definition guarantees the existence of ideals, namely the set consisting of only the zero-element is an ideal.

Definition: Let $(R, +, \cdot)$ be a ring and let I be an ideal in R. If $I \subset R$ we say that I is a *proper ideal* in R.

Lemma 4: Let $(R, +, \cdot)$ be a ring with a multiplicative unit element 1. An ideal of R is proper if and only if it does not contain 1.

Proof Lemma 4: First we prove that a proper ideal must not contain 1. Let $(R, +, \cdot)$ be a ring and let I be a proper ideal in R. We make a proof by contradiction, so assume that $1 \in I$. By property 2 in the definition of an ideal in a ring we have that for every $x \in R$, $x \cdot 1 = 1 \cdot x = x$ is an element of I. Thus we have that $x \in R$ implies $x \in I$ which means that I is not a proper subset of R. This violates the definition of a proper ideal. We have a contradiction.

Next we prove that an ideal that does not contain 1 is necessarily proper. Let $(R, +, \cdot)$ be a ring with a multiplicative unit 1. Let J be an ideal such that $1 \notin J$. Since R contain 1 but J does not, clearly they can not be equal and hence $J \subset R$. □

For the main proof of the section we will use a theorem from group theory. But since our main focus is to connect Zorn's Lemma with ring theory we will skip the detour into group theory to prove the theorem. Instead we will simply state it without proof.

Lemma 5: Let $(R, +)$ be a group and let I be a non-empty subset of R. Then $(I, +)$ is a subgroup of $(R, +)$ if and only if: for all $x, y \in I$ we have $(x - y) \in I$.

Theorem 2: Let $(R, +, \cdot)$ be a ring with a multiplicative unit 1. There is an ideal I such that $I \subset J \subset R$ is not true for any proper ideal J. We call I a maximal ideal in R.

Proof Theorem 2: Let X be the set of all proper ideals in R. As we noted earlier X is non-empty and it is furthermore partially ordered by inclusion. We wish to use Zorn's lemma and hence need to show that every chain in X has an upper bound. A chain in X is a totally ordered set of ideals $\{I_x\}$. We let $I = \bigcup_x I_x$ where the union runs over all ideals in $\{I_x\}$. Next we must show that I is a proper ideal and thus an upper bound of $\{I_x\}$. We go through the three properties defining a proper ideal one by one.

1. Since every I_x is an ideal they are also by definition subgroups of $(R, +)$. Let $s, t \in I$. Then there must be two ideals I_a and I_b in $\{I_x\}$ such that $s \in I_a$ and $t \in I_b$. Since $\{I_x\}$ is a chain either $I_a \subseteq I_b$ or $I_b \subseteq I_a$ holds. We can suppose that $I_a \subseteq I_b$ without loss of generality. Thus s and t are elements of I_b. Since I_b is a group we get that $s - t \in I_b$ and hence $s - t \in I$. Lemma 5 now gives that $(I, +)$ is a subgroup of $(R, +)$.

2. All sets $I_x \in \{I_x\}$ are by definition ideals. Hence for every $a \in I$ there is an ideal I_a containing a. Since I_a is an ideal we get that for every $r \in R$ either $a \cdot r$ or $r \cdot a$ is in I_a. Since I_a is a subset of I either $a \cdot r$ or $r \cdot a$ is an element of I.

3. Lemma 4 immediately gives that since I does not contain 1 it is a proper ideal in R.

We have now shown that I itself is an ideal and hence is an upper bound for the arbitrary chain $\{I_x\}$ in X. Thus we can conclude that every chain in X has an upper bound. Now Zorn's lemma gives that X has a maximal element. But X is the set of all proper ideals in R and thus R must have a maximal ideal. □

5.3 Non-measurable sets and the Banach-Tarski paradox

The field of measure theory deals with the problem of defining a size or generalized volume of sets. It is possible to define a measure, called the Lebesgue measure, such that almost all subsets of $\mathbb{R}^n$ have a volume. However the axiom of choice implies the existence of sets whose measure is undefined, these are called non-measurable sets. We demonstrate this with an example, but since we are not attempting a deep dive into measure theory we will keep the discussion brief.

A measure is a function μ that assigns a non-negative volume to the subsets of a given set. One important property of measures μ for this discussion is *countable additivity.* This means that for a collection of countably many disjoint sets $\{E_i\}$ we have:

$$\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i).$$

Intuitively this is obvious, the volume of an object should be the same if we measure it directly or split it into pieces that we measure separately. Let D be the unit circle in the plane. Suppose that we have a measure on D that is invariant under rotation. Let G be the group of rotations of the circle by angles that are rational multiples of π. Note that G is countable. The set D is uncountable so by letting G act on D we get uncountably many orbits under G. By the axiom of choice there exist a set $H \subset D$ containing one element from every orbit. Furthermore all translated copies, i.e. all the sets we can aquire from letting G act on H, are disjoint. Since G is countable this partitions the circle into a countable collection of shifted copies of H. From countable additivity we get that if H has zero measure then all of its copies and hence the circle D has zero measure. On the other hand if H has non-zero measure then D has infinite measure.

The fact that some sets have no measure might seem counterintuitive in its own right but it also has some interesting consequences. The most famous is probably the Banach-Tarski paradox which states that it is possible to take a solid sphere, cut it up into finitely many pieces, rotate and translate these pieces and put them back together such that you end up with two identical copies of the original sphere. Of course this is not a ”real” paradox, it just goes against our intuition from the physical world. What actually happens is that the sphere is split up into non-measurable pieces and hence countable additivity does not hold true. In other words, the volume is not conserved during this process and that is why another sphere could be created seemingly out of nowhere. [2]

References

[1] Paul J. Cohen. The independence of the axiom of choice. `https://stacks.stanford.edu/file/druid:pd104gy5838/SCM0405.pdf`. Retrieved 2021.

[2] Lukas Enarsson, Oskar Johansson, Vincent Molin, and Emil Timlin. The banach–tarski paradox and its implications on the problem of measure. *Bachelors thesis, Chalmers University of Technology, Gothenburg*, 2020.

[3] Irving Kaplansky. *Set theory and metric spaces.* American Mathematical Soc., second edition, 2008.