Axiom of Choice and Zorn's Lemma

Many algebraic proofs rely indirectly on the Axiom of Choice. This axiom is not part of the standard (Zermelo-Fraenkel) axioms of set theory and is independent of these axioms. The axiom of choice can be stated in several different forms:

Axiom. *Let S_i be a collection of non-empty sets indexed by $i \in I$. Then there exists a "choice" function $f : I \to \cup S_i$ such that $f(i) \in S_i$ for $i \in I$.*

Axiom. *Let S_i be a collection of non-empty sets indexed by $i \in I$. Then the Cartesian product $\prod_{i\in I} S_i$ is not empty.*

The second form of the Axiom of Choice is most often used in certain proofs such as the proof of Tychonoff's compactness theorem.

Remark 1. In many algebraic constructions the sets S_i have some extra structure, e.g., they are groups/fields/vector spaces, and have distinguished elements like 0. Relying on these distinguished elements one can construct a "choice" function which sends i to the distinguished element in the set S_i without using the Axiom of Choice. For example, one does not need the Axiom of Choice to show that the direct product (or direct sum) of any family of vector spaces is not empty (it is not empty since it is a vector space and contains 0).

There are only a few algebraic arguments that directly use the Axiom of Choice. However, there are several other (more than 10) statements which are equivalent to the Axiom of Choice which are often used in algebra.

Definition 2. A relation $\prec$ on pairs of elements in A is called an *order* if

- for any two elements a and b in A, exactly one of the following is true $a = b$ or $a \prec b$ or $b \prec a$;
- the relation $\prec$ is transitive, i.e., $a \prec b$ and $b \prec c$ imply $a \prec c$.

An order is called a *well order*, if there are no infinite decreasing sequences of elements, i.e., it is not possible to find elements a_i such that

$$\cdots \prec a_n \prec a_{n-1} \prec \cdots \prec a_3 \prec a_2 \prec a_1.$$

Example 3. The sets $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ have natural orders which satisfy the properties above. However, only $\mathbb{N}$ is well ordered since the sequence $-1, -2, \ldots, -n, \ldots$ is an infinite decreasing sequences of elements in $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$.

Theorem 4 (Well-ordering theorem)**.** *There exists a well ordering on any set A.*

Remark 5. One can directly construct a well ordering on any finite or countable set A using the natural order on $\mathbb{N}$ and a bijection between A and $\mathbb{N}$. On the other hand there are no known explicit constructions of a well ordering of $\mathbb{R}$.

The well-ordering theorem is an essential building block in the theory of cardinals and ordinals. One can use the well-ordering theorem to do an analog of mathematical induction on sets larger than $\mathbb{N}$.

Another statement equivalent to the axiom of choice which is often used in set theory is:

Theorem 6. *Let A and B be non-empty sets. Then there exists either an injective function f from A to B or an injective function g from B to A.*

A third equivalent form of the axiom of choice is Zorn's Lemma. We will start with some definitions:

Definition 7. The relation $\preceq$ is a *partial order* on some set A if it satisfies the following:

- the relation $\preceq$ is reflexive, i.e, $a \preceq a$;
- the relation $\preceq$ is transitive, i.e., $a \preceq b$ and $b \preceq c$ imply $a \preceq c$;
- the relation $\preceq$ is antisymmetric, i.e., $a \preceq b$ and $b \preceq a$ imply $a = b$.

Definition 8. A subset T of a partially ordered set A is *totally ordered* if for any s, t in T we have either $s \preceq t$ or $t \preceq s$. Such a set T has an *upper bound* u in A if $t \preceq u$ for all t in T. Note that u is an element of A but need not be an element of T.

Definition 9. A *maximal element* of A is an element m such that the only element x with $m \preceq x$ is m itself. **Notice** that we do not require that $x \preceq m$ for all x!

Example 10. Many partial orders in algebra arise from inclusion of sets: the set A consist of all subsets of some object O which satisfy a certain property P and we define $S \preceq S'$ iff $S \subseteq S'$. It is very easy to verify that this gives a partial order on A. The maximal elements in this partial ordered set are precisely the subsets M which have the property P but any M' strictly containing M does not have property P.

A typical example of a totally ordered subset of A is an infinite chain of sets S_i such that

$$S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n \subseteq S_{n+1} \subseteq \cdots,$$

however sometimes it is possible to construct "uncountable" chains (for example, for $r \in \mathbb{R}$ the open sets $S_r = (-\infty, r) \subset \mathbb{R}$).

Lemma 11 (Zorn's Lemma)**.** *Every non-empty partially ordered set in which every totally ordered subset has an upper bound contains at least one maximal element.*

A typical application of Zorn's lemma is the following result from functional analysis (the theorem below is weaker than the classical Hahn-Banach Theorem because it ignores the topology on the vector space V).

Theorem 12 (Hahn-Banach Theorem)**.** *Let V be a non-zero vector space over a field F and let u be a non-zero element in V. Then there exists a linear transformation $g : V \to F$ such that $g(u) = 1$.*

Proof. Consider the set A of all subspaces W of V which do not contain the vector u. One can define a partial order on A using inclusion, i.e., $W \preceq W'$ if and only if $W \subseteq W'$.

Before applying Zorn's lemma one needs to show that any (non-empty) totally ordered subset $\{W_i\}_{i\in I}$ of A has an upper bound. Define the subset $W = \bigcup_{i\in I} W_i$. First we will show that W is an element in A: it is clear that W is closed under multiplication by field elements since all W_i are. We will show that W is closed under addition: let w and w' be elements in W, therefore there exists indices i and i' such that $w \in W_i$ and $w' \in W_{i'}$. Using the total ordering gives us that either $W_{i'} \preceq W_i$ or $W_i \preceq W_{i'}$. Without loss of generality we can assume that $W_{i'} \preceq W_i$, thus both w and w' are elements of the subspace W_i. This shows that the sum $w + w'$ is also an element of W_i and therefore an element of W, which completes the proof that W is a subspace of V.

In order to show that W is an element of A we also need to check that u is not in W. However, u is not an element of any W_i therefore it is not an element of the their union.

By construction the $W_i \subseteq W$ therefore $W_i \preceq W$ for $i \in I$. Thus we have constructed an upper bound W for the totally ordered set $\{W_i\}_{i\in I}$, therefore any totally ordered subset of A has an upper bound.

Applying Zorn's Lemma we obtain some maximal element M in A. Let u' be a vector in V which is not in M. Consider the subspace $M' = \mathrm{Span}(M \cup \{u'\})$. If this subspace does not contain u we will get a contradiction since M' is in A and $M \preceq W'$ but $M' \neq M$. Therefore the vector u is an element of M', i.e.,

$$u = m + \alpha u'$$

for some $w \in W$ and $\alpha \in F$. Notice that the scalar α is not zero, since $u \notin M$, which allows us to write

$$u' = -\alpha^{-1}m + \alpha^{-1}u$$

This shows that every element $v \in V$ can be written uniquely (the uniqueness follows from $u \notin M$) as $v = \beta_v u + m_v$, where $\beta_v \in F$ and $m_v \in M$. Equivalently we have the decomposition $V = M \oplus Fu$ as an inner direct sum. Finally one can define the function g from V to F by $g(v) = \beta_v$. One can easily verify that this map is a linear transformation such that $g(u) = 1$. □

A slight modification of the above proof gives the following:

Theorem 13. *Any subspace W of V has a complementary subspace, i.e., there exists another subspace U of V such that $U \cap W = \{0\}$ and $\mathrm{Span}(U \cup W) = V$, which gives a decomposition of $V = U \oplus W$ as an inner direct sum.*

An other classical result which requires the use of Zorn' lemma is:

Theorem 14. *Every vector space V has a basis $\mathcal{B}$.*

Proof. Consider the collection A of all linearly independent subsets S of V. There is a natural partial order on A using inclusion, i.e., $S \preceq S'$ if and only if $S \subseteq S'$. We show that any maximal element in the partially ordered set A is a basis for V. The existence of such a maximal element is not "obvious" and uses Zorn's lemma.

Before applying Zorn's lemma one needs to show that any (non-empty) totally ordered subset $\{S_i\}_{i\in I}$ of A has an upper bound. Define the subset $S = \bigcup_{i\in I} S_i$. First we will show that S is an element in A: it is clear that S is a subset of V and we only need to show that S is linearly independent. Assume that there is a non trivial linear dependence $\sum_{k=1}^{N} \alpha_k v_k = 0$ between the elements of S, where N is some finite number. The vectors v_k are in S, therefore they are in some S_{i_k} for some indices i_k. It is clear that any total order on a finite set has a "largest" element, i.e., there exists some j between 1 and N such that $S_{i_k} \preceq S_{i_j}$ for all k. This shows that all vectors v_k are in the linearly independent set S_{i_j}, which contradicts the existence of a non trivial linear dependence among them. Thus, the set S is linearly independent and it is an upper bound for the totally ordered subset $\{S_i\}_{i\in I}$.

By Zorn's lemma there exists some maximal element $\mathcal{B}$ in A. By construction $\mathcal{B}$ is a linearly independent subset of V and in order to show $\mathcal{B}$ is a basis for V it is enough to check that $\mathrm{Span}\{\mathcal{B}\} = V$. Assume that $\mathrm{Span}\{\mathcal{B}\}$ is a proper subspace of V, then there exists a vector $v \in V \setminus \mathrm{Span}\{\mathcal{B}\}$. One can easily verify that the set $\mathcal{B} \cup \{v\}$ is also linearly independent, however this contradicts the maximality of $\mathcal{B}$. Therefore $\mathcal{B}$ spans V and is a basis of V. □

There are several extensions of Theorem 14, which have similar proofs:

Theorem 15. *Any linearly independent subset S of a vector space V can be extended to a basis $\mathcal{B}$.*

Theorem 16. *Let T be a subset of a vector space V. Then there exists a linearly independent subset S of T such that $\mathrm{Span}(S) = \mathrm{Span}(T)$. Equivalently any spanning set of V contains a basis.*

Once can summarize the above proofs as follows: First, consider all "sub-objects" of a given algebraic object which satisfy a certain property (not containing the vector u in the above example) and define a partial order on this using inclusion. Second, show

that any totally ordered set has a maximal element – typically the set theoretic union is an upper bound, but one needs to show that this is a "sub-object" which satisfies the required properties. Finally, use a maximal "sub-object" given by Zorn's lemma to complete the proof.

Cardinal Numbers

We now give some methods of describing and dealing with very large sets. We begin with a way of "counting" their elements to describe just how large they are.

Definition 17. Two set A and B are said to have the same *cardinality* if there exists a one-to-one, onto function $f: A \longrightarrow B$. This will be denoted by writing $|A| = |B|$. The symbol $|A|$ is called the *the cardinality of* A.

Remark 18. Cardinality defines the analogue of an equivalence relation on the class of all sets: It is reflexive (via the identity function), symmetric (via the inverse of a one-to-one, onto function), and transitive (via composition of fuctions). Thus writing "$=$" to mean "has the same cardinality" behaves the usual way an equals sign behaves.

Definition 19. If there exists a one-to-one function $f: A \longrightarrow B$, the cardinality of A is said to be *less than or equal to* the cardinality of B, and symbolized by $|A| \leq |B|$. If in addition there exists no one-to-one onto function $g: A \longrightarrow B$, the cardinality of A is said to be *less than* the cardinality of B, and symbolized by $|A| < |B|$.

Proposition 20. *Let A and B be sets.*

a. *If there exists a one-to-one function $f: A \longrightarrow B$, then there exists an onto function $g: B \longrightarrow A$ such that $g \circ f = 1_A$.*

b. *If there exists an onto function $g: B \longrightarrow A$, then there exists a one-to-one function $f: A \longrightarrow B$ such that $g \circ f = 1_A$.*

That is, the following statements are equivalent:

1. $|A| \leq |B|$,
2. *There exists a one-to-one function $f: A \longrightarrow B$,*
3. *There exists an onto function $g: B \longrightarrow A$.*

Proof. If either A is empty, the result is trivial from the defintion of function. If that is not clear, see the formal description and discussion of functions given below. We now assume that A is non-empty and choose some element $a_0 \in A$. For $b \in f(A)$ there exists a unique $a \in A$ with $f(a) = b$ as f is one-to-one. Define

$$g(b) = \begin{cases} a_0 & \text{if } b \notin f(A) \\ a & \text{if } b \in f(A) \text{ and } f(a) = b \end{cases}.$$

Then $g(f(a)) = a$ for all $a \in A$.

Let $g^{-1}(a) = \{\, b \in B \mid g(b) = a \,\}$. As g is onto, this is a non-empty set for every $a \in A$. Further, B is clearly the disjoint union of these subsets $g^{-1}(a)$, $a \in A$. By the Axiom of Choice, there exists a choice function c which picks exactly one element from each of these sets $g^{-1}(a)$. Now define

$$f(a) = c(g^{-1}(a)) \ .$$

Clearly $g(f(a)) = a$ for all $a \in A$ (as $g(x) = a$ for every $x \in g^{-1}(a)$).

The last statement is just a summary of the result. □

Theorem 21 (Schroeder-Bernstein)**.** *If there exist one-to-one functions* $f : A \longrightarrow B$ *and* $g : B \longrightarrow A$*, then there exists a one-to-one, onto function* $h : A \longrightarrow B$*.*
That is, $|A| \leq |B|$ *and* $|B| \leq |A|$ *imply that* $|A| = |B|$*.*

Proof. Given $a \in A$, if there exists an element $b \in B$ such that $g(b) = a$ (necessarily unique), we will call b and *ancestor* of a. Similarly, if $b \in B$ and there is an element $a \in A$ with $f(a) = b$, a is called an ancestor of b. An element $a = a_0$ of A is said to have an *odd* number $(2k + 1)$ of ancestors, if there exist elements

$$a_0, a_1, \ldots, a_{2k} \in A$$

and elements

$$b_1, b_3, \ldots, b_{2k+1} \in B$$

with

$$\begin{aligned} g(b_{2i+1}) &= a_{2i} \text{ for } 0 \leq i \leq k \\ f(a_{2i}) &= b_{2i-1} \text{ for } 0 < i \leq k \\ f^{-1}(b_{2k+1}) &= \phi \end{aligned}$$

Similarly, the number of ancestors of a is *even* $(2k)$, if the process stops with $g^{-1}(a_{2k})$ being empty. The number of ancestors is *infinite*, if the process never stops (a_{2i} and b_{2i+1} exist for all positive integers i).

We denote by $\mathcal{E}_A$ the elements of A with an even number of ancestors, $\mathcal{O}_A$ those with an odd number of ancestors, and $\mathcal{I}_A$ those with an infinite number of ancestors. Clearly A is the disjoint union of these three sets. Similarly there is a decomposition of B into three disjoint subsets $\mathcal{E}_B$, $\mathcal{O}_B$ and $\mathcal{I}_B$.

It is now possible to restrict f of g to these three subsets and obtain one-to-one functions (denoted by a "bar"):

$$\begin{aligned} \overline{f} &: \mathcal{O}_A \longrightarrow \mathcal{E}_B \\ \overline{f} &: \mathcal{E}_A \longrightarrow \mathcal{O}_B \\ \overline{f} &: \mathcal{I}_A \longrightarrow \mathcal{I}_B \end{aligned}$$

and similarly

$$\begin{aligned} \overline{g}: \mathcal{O}_B &\longrightarrow \mathcal{E}_A \\ \overline{g}: \mathcal{E}_B &\longrightarrow \mathcal{O}_A \\ \overline{g}: \mathcal{I}_B &\longrightarrow \mathcal{I}_A \ . \end{aligned}$$

By definition the sets $\mathcal{O}_A$ and $\mathcal{O}_B$ are in the image of f or g while $\mathcal{E}_A$ and $\mathcal{E}_B$ may not be as they contain the elements with 0 ancestors.

Thus we have found four one-to-one, onto functions:

$$\begin{aligned} \overline{f}: \mathcal{E}_A &\longrightarrow \mathcal{O}_B \\ \overline{f}: \mathcal{I}_A \longrightarrow \mathcal{I}_B \overline{g}: \mathcal{E}_B &\longrightarrow \mathcal{O}_A \\ \overline{g}: \mathcal{I}_B &\longrightarrow \mathcal{I}_A \ . \end{aligned}$$

We now "piece" together the sought-after one-to-one, onto function:

$$h(a) = \begin{cases} f(a) & \text{if } a \in \mathcal{E}_A \\ g^{-1}(a) & \text{if } a \in \mathcal{O}_A \\ f(a) & \text{if } a \in \mathcal{I}_A \ . \end{cases}$$

□

The Notes for the course *Math 4330, Honors Linear Algebra* at Cornell University have been developed over the last ten years or so mainly by the following (in chronological order):

Gerhard O. Michler

R. Keith Dennis

Martin Kassabov

W. Frank Moore

and

Yuri Berest.

Most sections have been revised so many times the original author may no longer recognize it. The intent is to provide a modern treatement of linear algebra using consistent terminology and notation. Some sections are written simply to provide a central source of information such as those on "Useful Definitions", "Subobjects", and "Universal Mapping Properties" rather than as a chapter as one might find in a traditional textbook. Additionally there are sections whose intent is to provide proofs of some results which are not given in the lectures, but rather provide them as part of a more thorough development of a tangential topic (e.g., Zorn's Lemma to develop cardinal numbers and the existence of bases and dimension in the general case).

A large number of challenging exercises from many different sources have been included. Although most should be readily solvable by students who have mastered the material, a few even more challenging ones still remain.

Much still remains to be done. Corrections and suggestions for additional exercises, topics and supplements are always welcome.