The Axiom of Choice, Zorn's Lemma, and all that

When set theory was formalized in the early 1900's, and a system of axioms set down, it was found (as for Euclidean geometry centuries earlier!) that one of the axioms proposed was not quite as ''obvious'' as the others. This was the axiom that allowed one, in a construction, to simultaneously make an arbitrarily large set of choices. The formal statement was that given a family $(X_i)_{i\in I}$ of nonempty sets, one could choose a family $(x_i)_{i\in I}$ with each $x_i \in X_i$ – in other words, that the product $\Pi_{i\in I} X_i$ of a family of nonempty sets X_i is nonempty!

If the index set I is finite, there is no problem with this: The statement that X_i is nonempty says that there exists (which is all that is meant by ''one can choose'') an element $x_i \in X_i$; a finite number of applications of this observation gives an element $(x_i) \in \Pi_{i\in I} X_i$. (One can make this argument precise using induction.) But some mathematicians argued that if I was infinite, we had no way of knowing that the same was true, so that the general assertion, called the Axiom of Choice, should be rejected. And indeed, some applications of that Axiom led to conclusions that mathematicians found surprising, and possibly unreasonable. (For instance, that there exists a set S of real numbers containing exactly one representative of each coset of the additive subgroup of rational numbers.) We will not go further into the history here. The Axiom is part of the set-theoretic foundations accepted by most mathematicians today, ''Zermelo-Fraenkel set theory with Choice'', abbreviated ZFC.

However, certain reformulations are generally more convenient to use than the Axiom in its raw form. A particularly useful formulation is Zorn's Lemma, stated as part (ii) of the Theorem on the next page, which allows one to carry out constructions that may require infinite sequences of choices, each of which depends on the preceding ones, so that one does not know initially just what choices are to be made and in what order.

In what follows, I assume you are familiar with the concepts of *partially ordered set* and *totally ordered set*. (Many authors shorten one or the other of these terms to *ordered set*; e.g., in Appendix 2.1 Lang [**2**], a partially ordered set is so called.) An element m of a partially ordered set P is called *maximal* if no element of P is strictly greater than m. A subset S of a partially ordered set P is called *bounded* if there exists $t\in P$ such that t is $\geq$ every element of S. A subset S of a partially ordered set P is called a *chain* if it is totally ordered under the induced ordering. An *initial segment* of a chain S means a subset $T \subseteq S$ such that if $u \leq v$ are members of S, and $v \in T$, then $u \in T$.

A *well-ordered* set means a totally ordered set in which every nonempty subset has a *least* element. You should verify the following observations that we will need:

(1) If S is a well-ordered subset of a partially ordered set P, and $t \notin S$ is an upper bound for S in P, then $S \cup \{t\}$ is well-ordered.

(2) If C is a set of well-ordered subsets of a partially ordered set P, such that for all $X, Y \in C$, either X is an initial segment of Y, or Y is an initial segment of X, then $\cup\, C$ (i.e., $\cup_{X\in C} X$) is well-ordered.

In the Theorem below, we assume the axioms of ZFC *other than* the Axiom of Choice, and sketch a proof that under these assumptions, four statements, one of which is that Axiom, and another of which is Zorn's Lemma, are equivalent. We will not list the other Axioms of ZFC, but simply allow ourselves to use familiar set-theoretic techniques other than infinite systems of choices. For a brief discussion of these axioms see [**1**, §4.4]; for a detailed development, see a text on set theory. In Lang [**2**] the Axiom of Choice is assumed without being stated, so that in his Appendix on Set Theory, statements (ii) and (iii) below are simply ‘‘proved’’. The development here is shorter than Lang’s, and explicitly shows the relation with the Axiom of Choice.

To indicate how assertion (ii) below is useful, let me note that in the typical situation where it is to be applied, P represents a set of ‘‘partially completed’’ constructions, and the relation $q \geq p$ means ‘‘q is a partial construction which extends (carries further) the partial construction p’’. For a given application, one has to set up the P that represents one’s method of construction (sometimes tricky), verify that it satisfies the hypotheses of Zorn’s Lemma (usually easy), and then show that a *maximal* element of P corresponds to the sort of completed construction one wants (frequently a nontrivial argument in the area of mathematics one is applying the Lemma to). We will see examples of this procedure below in the proofs of the three other conditions from (ii).

Theorem 1. *The following statements are equivalent* (*assuming Zermelo-Fraenkel Set Theory but not the Axiom of Choice*).

(i) (Axiom of Choice) *If* $(X_i)_{i\in I}$ *is a family of nonempty sets, then* $\Pi_I X_i$ *is also nonempty.*

(ii) (Zorn’s Lemma) *If* P *is a nonempty partially ordered set with the property that every chain in* P *is bounded, then* P *has a maximal element.*

(iii) (Well-Ordering Principle) *Every set* X *can be well-ordered.* (*I.e., for every* X *there exists a binary relation* $\leq$ *on* X *which makes it a well-ordered set.*)

(iv) (Comparability of cardinalities, or Bernstein’s Theorem) *Given any two sets* X *and* Y, *there exists either a bijection between* X *and a subset of* Y, *or a bijection between* Y *and a subset of* X.

Sketch of Proof. We shall first prove (i)⇒(ii) (which is the hard step, and is the important implication for our purposes, since it means that assuming ZFC, we can use Zorn’s Lemma), then show that (ii) implies each of (i), (iii) and (iv), next prove (iii)⇒(i), and finally, very roughly, indicate how (iv)⇒(iii) is proved. If each part of this argument were complete, the arguments showing (ii)⇒(i) and (ii)⇒(iii) could be omitted, and we would have a cyclic proof of equivalence of our four conditions. The present arrangement allows us to get the equivalence of (i), (ii) and (iii) without needing the details of (iv)⇒(iii), which we haven’t time to go into, and also gives three instructive illustrations of how Zorn’s Lemma is used.

(i)⇒(ii). Let P satisfy the hypothesis of (ii). If S is any chain in P, then by assumption, the set of upper bounds to S in P is nonempty; from this it is not hard to see that if S does *not* contain a maximal element of P, the set of upper bounds to S which do not lie in S is nonempty. Let us call the latter set $B(S)$. Assuming (i), let us choose a function φ on the set of all chains in P which do not contain maximal elements of P, associating to each such chain S a member of $B(S)$.

The *idea* of what we want to do next is to start with any element $p \in P$; if it is not maximal, let $p' = \varphi(\{p\}) > p$; if p' is not maximal in P let $p'' = \varphi(\{p, p'\}) > p'$; if p'' is not maximal, let $p''' = \varphi(\{p, p', p''\}) > p''$, and so on. If we do not get a maximal element after finitely many of these steps, we let $p^* = \varphi(\{p, p', \dots, p^{(i)}, \dots\})$; if p^* is not maximal in P, we let $p^{*\prime}$ denote the result of applying φ to the chain gotten by appending p^* to the above chain – ‘‘and so on’’! Intuitively, this process can ‘‘go on running’’ indefinitely, producing a chain of ever growing cardinality, and would run beyond the cardinality of P if it never terminated; hence it must eventually terminate, which can only happen if we hit a maximal element of P. But we do not have at hand the machinery needed to formalize how to ‘‘set such a construction running’’ (though this is developed in [**1**]). What we shall use instead is a definition which *characterizes* the set of chains that *would* have been given by the above construction, if we had been in a position to formalize it.

Namely, having fixed an element $p \in P$, let C denote the set of subsets $S \subseteq P$ which have the properties:

(a) S is a well-ordered chain in P.

(b) p is the least element of S.

(c) For every proper nonempty initial segment $T \subseteq S$, the least element of $S - T$ is $\varphi(T)$.

We see that C is nonempty, because it contains $\{p\}$. We claim that if S and S' are two members of C, then one of them is an initial segment of the other. Indeed, let R denote the union of all sets that are initial segments both of S and of S'; thus, R is the greatest common initial segment of S and of S'. If R were a *proper* initial segment both of S and of S', then by (c), $\varphi(R)$ would be the least element both of $S - R$ and of $S' - R$; but this would make $R \cup \{\varphi(R)\}$ an initial segment both of S and of S', contradicting the choice of R. So R cannot be proper in both S and in S', i.e., it must equal one of them. That one will thus be an initial segment of the other, as claimed.

Hence if we now denote by U the union of all members of C, then by (2) U is well-ordered. Clearly, also, all members of C are initial segments of U, and the least element of U is p. We claim that U also satisfies (c) above. Indeed, if T is a proper nonempty initial segment of U, there exists some element $u \in U - T$. By construction of U, u belongs to some $S \in C$, and by what we have proved about C, T must be a proper initial segment of S. Hence (c) says that $\varphi(T)$ is the least element of $S - T$, so as S is an initial segment of U, $\varphi(T)$ is also the least element of $U - T$, as required.

This proves that U is itself a member of C. Now suppose U does not contain a maximal element of P. Then $U \cup \{\varphi(U)\}$ will be an element of C (cf. (1) above), which is not a subset of U, contradicting the definition of U. So U must contain a maximal element of P; so such a maximal element exists, establishing (ii).

The above argument was intricate; but the elegance of Zorn's Lemma will be seen in the fact that the proofs of all three other conditions *from* that Lemma are quite direct. I leave it to you to verify that in each case, the partially ordered set described satisfies the hypothesis of Zorn's Lemma, and that a maximal element, whose existence is the conclusion of that Lemma, will be an object of the sort we are trying to get in the case in question. If you have not seen Zorn's Lemma proofs before, I urge you to think through all these verifications carefully!

(ii)$\Rightarrow$(i). Let $(X_i)_{i \in I}$ be given as in (i). We let P be the set of ''partial functions'' on I carrying each $i \in I$ to a member of X_i; formally, we define this to be the set of all subsets $\varphi \subseteq I \times (\bigcup_I X_i)$ such that for each $i \in I$, any element of φ with first component i has second component in X_i, and for each i there is *at most* one such element. We let P be partially ordered by inclusion (i.e., we consider the partial order relation $\subseteq$ on P). P is nonempty because the *empty* partial function is a member of it; and given any chain S in P, one can verify that the *union* of S (as a family of sets) is an upper bound to S in P. By Zorn's Lemma we conclude that P has a maximal element φ under this ordering; we then verify that such a φ will be a *function*, i.e., a member of $\prod_I X_i$, as required.

(ii)$\Rightarrow$(iv). (We do this next because the method is very similar to the preceding.) Given sets X and Y, let P be the set of all subsets $\varphi \subseteq X \times Y$ such that for each $x \in X$, φ contains at most one element with first component x, and for each $y \in Y$, φ contains at most one element with second component y, and let us order P by inclusion. As above, Zorn's Lemma gives us a maximal $\varphi \in P$. In this case, we find that maximality implies that *either* φ is a one-to-one function from X to Y *or* the set $\varphi^* = \{(y, x) \mid (x, y) \in \varphi\}$ is a one-to-one function from Y to X; so *one* of these exists, as desired.

(ii)$\Rightarrow$(iii). Given X, we define the set P to consist of all pairs $(A, \preccurlyeq)$ such that A is a subset of X, and $\preccurlyeq$ is a well-ordering of A. We then partially order P by letting $(A, \preccurlyeq) \leq (A', \preccurlyeq')$ if $A \subseteq A'$, $\preccurlyeq$ is the restriction of $\preccurlyeq'$ to A, and A is an *initial segment* of A' under $\preccurlyeq'$. Given a chain $\{(A_i, \preccurlyeq_i) \mid i \in I\}$ in P, one gets an upper bound by taking the union of the A_i, ordered by the union of the $\preccurlyeq_i$. The ordering on this union is a well-ordering by (2). (If in defining the order relation $(A, \preccurlyeq) \leq (A', \preccurlyeq')$ on P, we had left out condition that A be an initial segment of A', this would not be true!) By (ii), P has a maximal element. Now let us note that for any element $(A, \preccurlyeq) \in P$ with A a *proper* subset of X, we can add one more element of X to A to get a larger set A', extend $\preccurlyeq$ to an ordering $\preccurlyeq'$ in which this new element is greater than all others, and in view of (1), the result is an element of P above our given element. Hence for $(A, \preccurlyeq) \in P$ that is *maximal*, we must have $A = X$, so $\preccurlyeq$ is a well-ordering of the whole set X, as required.

(iii)$\Rightarrow$(i). Given $(X_i)_{i \in I}$ as in (i), let us choose, using (iii), some well-ordering $\leq$ of the set $X = \bigcup_I X_i$. For each $i \in I$ let x_i be the *least* element of the nonempty subset $X_i \subseteq X$ under this well-ordering. Then $(x_i)_{i \in I}$ is a member of $\prod_I X_i$.

(iv)$\Rightarrow$(iii) requires a bit of machinery that we will not develop here: Given any set X, one can show that there exists a well-ordered set Y which cannot be put in bijective correspondence with any subset of X. (One constructs this, loosely, by ''welding together'' copies of *all* well-ordered sets whose underlying sets are subsets of X, so that each is embedded in Y as an initial segment, and then adding one more element at the top.) But (iv) tells us that *either* Y can be put in bijective correspondence with a subset of X, or X with a subset of Y. Hence the latter must hold; and the resulting bijection between X and a well-ordered set induces a well-ordering on X, as required. □

I will assume the equivalent statements (i)-(iv) for the remainder of this course. The one we will use by far the most often is Zorn's Lemma.

Let us note that the formulation of Zorn's Lemma is slightly redundant: the empty set $\varnothing$ is a chain in any partially ordered set P, and an upper bound in P to that chain means an arbitrary element of P. Hence if *every* chain, including $\varnothing$, has an upper bound, it is unnecessary to add a separate assumption that P is nonempty. I have stated Zorn's Lemma as I did to conform with common usage, but in fact, you can see that in the proof of (ii)$\Rightarrow$(i) above, the verification of the nonemptiness of P used precisely the element that would have been gotten by applying our construction of upper bounds to the empty chain; and this is also true of the corresponding verifications (not given explicitly) in the proofs of (ii)$\Rightarrow$(iii) and (ii)$\Rightarrow$(iv). In almost every proof using Zorn's Lemma that I have seen, either the same is true, or the proof could be made a bit neater by using that approach! Only very rarely does one need different methods to find upper bounds for nonempty chains and for the empty chain. (However, some authors muddy the waters by defining a ''chain'' to be a *nonempty* totally ordered subset.) In our proof of Zorn's Lemma from the Axiom of Choice, it would likewise have been more elegant not to bring in p, but to replace (b) and (c) by the single condition gotten by removing ''nonempty'' from (c).

References

[**1**] George M. Bergman, *An Invitation to General Algebra and Universal Constructions*, pub. Henry Helson, Berkeley, CA, 1998. ii+398 MR **99h**:18001.

[**2**] Serge Lang, *Algebra*, revised third edition, Springer GTM v.211, 2002.