ZF, Choice, Zorn, Ordinals, Ultrafilters

Rich Schwartz

September 11, 2018

1 Introduction

There are various foundations for what you could call everyday mathematics. These foundations usually come as a list of axioms describing the allowable kinds of mathematical reasoning and constructions. The most popular foundational axioms are the *Zermelo-Frankel Axioms*. Most of them sound completely obvious. However, the ZF axioms are not really stand-alone axioms. In order to understand and use them, you need language and logic. Which comes first, the language, the logic, or the axioms? I don't know. In practice, when you do mathematics, you freely use all three things in ways that are not so easy to disentangle.

There are some additional axioms which do not follow from the ZF axioms (and language and logic). The most famous is the *Axiom of Choice*, an axiom which has many reformulations – e.g. Zorn's Lemma. One purpose of these notes is to discuss the ZF axioms, with a view towards putting the Axiom of Choice in context. The main purpose is to show the equivalence of the Axiom of Choice, Zorn's Lemma, and the Well-Ordering Principle, given ZF. Finally, the last purpose is to derive some nice applications of these axioms – e.g., the existence of nonprinciple ultrafilters.

I got many of the details about the ZF axioms from Wikipedia and from *An introduction to Set Theory* by William A.R. Weiss, though these notes are not a perfectly faithful account of what is written there. In particular, the arguments in the notes do not "go all the way back to the axioms", though of course they could be re-written (at length) to do so. Overall, I wanted to write down how I think about the foundations, and to get to some interesting highlights fairly efficiently.

2 Sets

The foundational axioms of mathematics concern properties of *sets*. Here is an informal definition:

A set is a collection of things, the things themselves being sets.

This is a very beautiful idea: There is only one kind of "thing" (or "object" or "entity" or "structure" or "gadget" or "gizmo") in the mathematical universe, and it is made of parts which are "things" of the same kind.

On the down side, a set has no formal definition. It is considered to be a basic and undefinable concept. How would you define it? As a collection of things? What are collections? What are things? Any attempt will just lead in circles. Another problem is that the mathematical universe turns out to contain "collections of things" which are not actually sets – e.g., the *class* of all sets. A class is like a set, but bigger. The ZF axioms do not apply to classes, though you sometimes want to refer to classes when making a logical argument which otherwise uses the ZF axioms. Yet another problem is that the axioms assert the *existence* of sets, without saying in what sense sets exists. Probably they don't exist in the same sense that, say, New Jersey exists. I don't know any satisfying resolution to these problems, but if you don't move past them you won't actually get to mathematics.

Once you grant the existence of sets (whatever they are), you have to swallow the concept of *equality* and *membership*. One writes $A = B$ to indicate that A and B are the same set. One writes $s \in S$ to indicate s is a member of S. That is, s is one of the sets of which S is constituted. Even though the members of a set are themselves sets, they are usually called *members*, or *elements*. Again, ultimately everything (except classes) is a set.

The symbols { and } are often used to denote the members of a set. For instance $\{1, 2, 3\}$ denotes the set whose members are 1 and 2 and 3 (whatever they are.) Other useful notions are reducible to these, such as the notion of *subset*. One writes $A \subset B$ if every member of A is also a member of B. That is $a \in A$ implies $a \in B$. Notice that even to talk about the basic definition of a subset, we need a certain amount of language. For instance, we have to know what the word *implies* means. This is what I mean when I say that language and logic are part of the foundations of mathematics just as much as the set-theoretic axioms.

3 The Zermelo-Frankel Axioms

Now I'm going to list some version of the ZF axioms. Some of them are redundant, though any list of the ZF axioms can be deduced from these.

Existence of the Empty Set: There exists a set with no members. It is called $\emptyset$, the empty set. This axiom is usually considered a consequence of the others, but I think it is a good place to start.

Axiom of Extension: Two sets are equal if and only if they have the same members. This one seems completely obvious.

The Axiom of Foundation: Given any nonempty set S, there exists an element $T \subset S$ such that $T \cap S = \emptyset$. Note that the members of sets are also sets, so this makes sense. As I'll explain in the next section, this axiom prevents a set from having itself as a member.

Axiom of Union: Given a set S, there is a set U_S so that $t \in U_S$ if and only if $t \in A$ for some $A \subset S$. So, U_S is the union of the elements in subsets of S. For instance, if

$$S = \{\{\{A, B\}, D\}, \{A, B\}, \{B, C\}\}$$

then $U_S = \{\{A, B\}, A, B, C, D\}$.

Existence of Power Sets: Given a set S there exists a set T whose members are the subsets of S. For instance, if $S = \{1, 2, 3\}$, then

$$T = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Axiom of Pairing: Given sets A and B there exists a set C whose members are A and B. That is $C = \{A, B\}$ exists. This axiom turns out to be redundant, but it is convenient to have it explicitly listed.

Axiom of Infinity There exists a set S which contains the emptyset and has the following property. If T is a member of S, then the successor of T is also a member of S. Here, the *successor* of T is defined to be the set $T \cup \{T\}$, the union of T with the set whose only member is T.

The Language of Set Theory: There is one more axiom - really an infinite list of them - but it requires a buildup of terminology. The *language of set theory* is a certain collection of *formulas* involving variables, logical connectives, and quantifiers.

The simplest formulas have the form $A = B$ or $A \in B$. Here A and B are the variables. These are called the *atomic formulas*. Then we have the following rules.

- If S is a formula, so is NOT S.
- If S and T are formulas, so is S AND T.
- One can universally quantify over a formula. Given a formula S, one also has the formula $\forall AS$, which means "S is true for all A."

One can make other constructions based on these rules. For instance S OR T is logically equivalent to NOT((NOT S) AND (NOT T)). Likewise $\exists AS$ is equivalent to NOT ($\forall A$ (NOT S)).

The formulas acquire truth values when one substitutes sets for all the variables. A formula like $\forall A(A \subset B)$ would say that, for all sets A, the set A is a subset of B. This is clearly false.

Free Variables: Given a formula F and a varaible A which occurs in F, you want to minimize the number of times A occurs without changing the (potential) truth value of F. You can do this by replacing all occurances of $\forall A(..A...)$ and $\exists A(...A...)$ with $\forall \widehat{A}(...\widehat{A}...)$ and $\exists \widehat{A}(...\widehat{A}...)$, where $\widehat{A}$ is some new variable what does not occur in F. The remaining instances of A are called *free variables*. Call F *clean* if all such replacements have been made.

A *set function* is a clean formula F, with free variables $S, x, y_1, ..., y_n, z$, such that for each set x, and each choice of sets $y_1, ..., y_n, S$, there is a unique set z such that F is true. One then defines $f(x) = z$.

Axiom Schema of Replacement: Given any set S and a set formula f on S as above, there exists a set T such that $z \in T$ if and only if $z = f(x)$ for some $x \in S$.

This is really an infinite list of axioms, because it makes one statement per formula, and there are infinitely many formulas. The Axiom Schema of Replacement basically says that you get a new set if you replace each element of S by its image under f.

4 Some Consequences

No Self-Membership: The Axiom of Foundation prevents a situation where a set has itself as a member. Suppose that there is a set S such that $S \in S$. By the Axiom of Pairing applied to $A = B = S$, there is a set $\{S\}$ whose only member is S. But then the Axiom of Foundation says that there is a member of $\{S\}$ which has no members in common with $\{S\}$. The only member to choose from is S, so S has no members in common with $\{S\}$. But if $S \in S$, then S is a member common to both S and $\{S\}$. This is a contradiction.

Set Building: Given a formula F with free variables $S, x, y_1, ..., y_n$, we get a function f from S into $\{\text{True}, \text{False}\}$ once $y_1, ..., y_n$ have been specified as sets. The *Axiom Schema of Comprehension* says that there exists a subset $T \subset S$ consisting exactly of those $s \in S$ for which $f(s)$ is true. That is,

$$T = \{s \in S |\ f(s) = \text{True}\}. \tag{1}$$

This is the usual way that one builds new sets out of old ones.

This axiom schema is a consequence of the Axiom Schema of Replacement and the existence of the emptyset. If there is no $s \in S$ for which $f(s)$ is true, we set $T = \emptyset$. Otherwise there is some $s \in S$ for which $f(s)$ is true. In this case we define $g : S \to S$ by the rule that $g(t) = t$ if $f(t)$ is true and $g(t) = s$ if $f(t)$ is false. We can write g as a set-valued function on S and the set given by the Replacement Axiom is as in Equation 1.

Intersections: Suppose that S is a nonempty set and $T \in S$. Then we use the language of set theory and the Axiom Schema of Comprehension to define the intersection of all the members of S which are also members of T. I'll denote this as

$$\bigcap^T S = \{s \in T |\ \forall t \in S, s \in t\}.$$

We think of T as the "container" for the intersection. For example, if

$$S = \{\{1,2,3,4\},\{1,2\},\{1,3\},\{1,4\}\}, \qquad T = \{1,2,3,4\},$$

then

$$\bigcap^T S = \{1\}.$$

5 The Natural Numbers

In the preceding sections, we have been using the symbols $1, 2, 3, ...$ informally to denote sets, and also as part of the language of set theory. We might have used other symbols like $A, B, C, ...$ for these purposes, but still you might wonder how the natural numbers are related to the ZF axioms. Here I'll sketch how one defines the natural numbers in terms of the axioms.

Let S be the set from the Axiom of Infinity. Let 2^S denote the set of subsets of S. This set is guaranteed by the Existence of Power Sets. Let $\Sigma \subset 2^S$ denote those elements t such that $\emptyset \in t$ and $u \in T$ implies $u' \in T$. As above u' is the successor of u. In other words Σ consists of all the subsets of S which also satisfy the Axiom of Infinity. The set Σ is guaranteed by the Axiom Schema of Comprehension. Note that $S \in \Sigma$. We define

$$\boldsymbol{N} = \bigcap^{S} \Sigma. \tag{2}$$

That is, $\boldsymbol{N}$ is the smallest set satisfying the Axiom of Infinity.

One can prove that $\boldsymbol{N}$ consists precisely of $\emptyset$ and the successors of $\emptyset$. If $\boldsymbol{N}$ contains any other element s, one can produce a smaller set satisfying the Axiom of Infinity by removing s and all its "predecessors".

Once $\boldsymbol{N}$ is defined, we can give the elements of $\boldsymbol{N}$ their more traditional names.

- 0 is a name for $\emptyset$.
- 1 is a name for $(0)'$ or $\{\emptyset\}$.
- 2 is a name for $(1)'$, or $\{\emptyset, \{\emptyset\}\}$.
- 3 is a name for $(2)'$, or $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$.

and so on. This definition has a special beauty: A natural number is a set, with fintely members, each of which is a natural number. Given any two natural numbers, one is a subset of the other. The natural numbers are literally built out of nothing!

With the above definition of natural numbers, you can define $m < n$ if $m \subset n$. Also, you can define $n + 1 = n'$. From there, you can derive all the usual rules of arithmetic. Then you can define the integers, the rationals, the reals, complex numbers, matrices, vector spaces, manifolds, etc.

6 The Axiom of Choice

The Axiom of Choice is another axiom one can introduce for sets. Say that a *partition* is a set S whose members are pairwise disjoint. That is, if $A, B \in S$, then $A \cap B = \emptyset$. We insist that the emptyset is not a member of S.

Here is an example. The set

$$\{\{1,2\},\{3,4\}\}$$

is a partition but $\{\{1,2\},\{2,3\}\}$ is not. It is worth pointing that the set S is usually considered to be a partition *of* the union set U_S. For instance, the partition above is a partition of $\{1,2,3,4\}$.

Axiom of Choice: If S is a partition, then there exists a set T such that T has one member in common with each member of S.

For instance, if S is the partition above, we could take $T = \{1,3\}$.

Sometimes, you can get the result of the Axiom of Choice just using the ZF axioms. For instance, suppose that S is a partition of $\boldsymbol{N}$ the natural numbers, and each member of S is a finite set of natural numbers. We could then define T to be the set consisting of least elements of members of S. That is, $a \in T$ if and only if a is the least element of the member of S which contains a.

You need to invoke the Axiom of Choice when you do not have a clear selection process like this. For instance, (jumping back to normal mathematical usage for a minute) say that two real numbers are equivalent if and only if their difference is a rational number. The set S of equivalence classes gives a partition of the set $\boldsymbol{R}$ of real numbers, and here you really need the Axiom of Choice to get the set T.

You might ask: What is the big deal about the Axiom of Choice? Well, let B_r denote the ball of radius r about the origin in $\boldsymbol{R}^3$. Here is a famous theorem.

Theorem 6.1 (Banach-Tarski) *Assume ZF and also that the Axiom of Choice holds for all partitions of* $\boldsymbol{R}$. *Then there are finitely many subsets* $S_1, ..., S_n \subset B_1$ *and rigid motions* $T_1, ..., T_n$ *of space, so that*

$$B_2 = T_1(S_1) \cup ... \cup T_n(S_n).$$

That is, you can cut apart the unit ball into finitely many pieces, and rearrange the pieces so that they equal the ball of radius 2.

7 Zorn's Lemma

The ZF axioms allow one to define the Cartesian product $S \times S$ consisting of ordered pairs (a, b) with $a, b \in S$. A *relation* on S is a subset $R \subset S \times S$. One often writes aRb if and only if $(a, b) \in R$. The relation R is called a *partial order* if

- aRa for all a.
- aRb and bRa imply $a = b$.
- aRb and bRc implies aRc.

In this situation one typically writes $a \leq b$ in place of aRb. This is what we do. We write $a < b$ if and only if $a \leq b$ and $a \neq b$. A set with a partial order on it is often called a *POSET*.

Two elements $a, b \in S$ are called *comparable* if either $a \leq b$ or $b \leq a$ (or both.) The partial order on a POSET is called *total* if every two elements are comparable. Here are some examples of POSETS.

1. The usual ordering on the integers is a total ordering.
2. On the natural numbers $\boldsymbol{N} = \{1, 2, 3, ...\}$ define $a \leq b$ if and only if a divides b. For instance $6 < 12$ in this ordering but 6 and 11 are not comparable.
3. Given any set T, let S be the set of subsets of T. For two subsets $A, B \subset T$, we write $A \leq B$ if and only if $A \subset B$. This makes S a POSET.

Maximal Elements: An element $x \in S$ is *dominated* if there is some $y \in S$ so that $x < y$. A *maximal element* of S is an element which is not dominated. Of the three examples above, only the third one has a maximal element. In the third example, the maximal element is the set T itself.

Chained Posets: Suppose that S is a POSET. Any subset $T \subset S$ inherits a partial order from S in the obvious way. A *chain* in S is a nonempty subset $T \subset S$ which is totally ordered. In Example 2 above, the powers of 2 form a chain. Given any set $T \subset S$, an *upper bound* for T is some element $s \in S$ so that $t \leq s$ for all $t \in T$. Note that s need not belong to T. Say that S is a

chained POSET if every chain of S has an upper bound. Note that a set with a maximal element is automatically chained, because this maximal element is an upper bound for every chain. Zorn's Lemma says that the converse is true.

Zorn's Lemma: A chained POSET has a maximal element.

We will see at the end of these notes that Zorn's Lemma is a consequence of ZF and the Axiom of Choice. However, we'll need some more machinery before proving this. The converse direction is much easier, and perhaps that is what makes Zorn's Lemma such a nice formulation of the Axiom of Choice.

Zorn Implies Choice: Let's first use Zorn's Lemma (and the ZF axioms) to prove the Axiom of Choice. Let S be a partition. Σ denote the set of all sets T such that T has at most one member in common with each element of S. Σ is nonempty because it contains sets of the form $\{a\}$ where a is some member of some member of S. By definition. $T \in \Sigma$.

We put a partial ordering on Σ by inclusion: $T_1 \leq T_2$ if and only if $T_1 \subset T_2$. This makes Σ into a POSET.

Lemma 7.1 Σ *is chained.*

Proof: Suppose that C is some chain. So for every two elements $A, B \in C$ we have either $A \subset B$ or $B \subset A$. Let T_C denote the union of all the members of T. We claim that $T_C \in \Sigma$. To see this, suppose that $T_C \cap A$ contains at least 2 elements a, b, for some $s \in S$. Then there are two elements $A, B \in C$ so that $a \in A$ and $b \in B$. Say that $A \subset B$. Then $a, b \subset B$ and $B \cap s$ contains two elements. This contradicts $B \in \Sigma$. This contradiction shows that $T_C \in \Sigma$. But then T_C is an upper bound for C. Hence Σ is a chained POSET. ♠

By Zorn's Lemma, Σ has some maximal element T. We claim that T has one member in common with each member of S. If this is false, then there is some member $A \in S$ which is disjoint from T. But then $T \cup \{a\}$ is also a member of Σ and is larger than T. This is a contradiction. Hence T has one element in common with each member of S.

8 Common Applications

Here are some common applications of Zorn's Lemma.

Theorem 8.1 *Suppose that S and T are two sets. Then either there is an injective map from S into T or an injective map from T into S.*

Proof: Let Σ denote the set of injective maps $f : A \to T$ where $A \subset S$. We make Σ into a POSET by saying that $(f, A) \leq (g, B)$ if $A \subset B$ and the restriction of g to A equals f. An argument very much like the one given in Lemma 7.1 shows that Σ is chained. By Zorn's Lemma, Σ has a maximal element (f, A). If $A \neq S$ then there are two cases to consider.

- If $f(A) = T$ then the inverse map $f^{-1} : T \to A$ gives an injective map from T into S.
- If $f(A) \neq T$ there is some $t \in T - F(A)$. But then we can choose some $s \in S - A$ and extend f by defining $f(s) = t$. This gives a new pair $(g, A \cup \{s\}) \in \Sigma$, contradicting the maximality of (f, A).

So, either $A = S$ and $f : S \to T$ is an injective map or else $f(A) = T$ and $f^{-1} : T \to S$ is an injective map. ♠

The next result assumes some basic knowledge of linear algebra.

Theorem 8.2 *Every vector space has a basis.*

Proof: Let Σ denote the set of linearly independent subsets of V. The elements of Σ are ordered by inclusion. $\{v_1, ..., v_n\} \subset \{v_1, ..., v_n, w_1, ..., w_m\}$. An argument similar to Lemma 7.1 shows that Σ is a chained POSET. Hence Σ has a maximal element β. If β is not a basis for Σ, then there is some $v \in V - \text{span}(\beta)$. But then $\beta \cup \{v\}$ is an element of Σ which is larger than β. This is a contradiction. Hence β is a basis for V. ♠

Theorem 8.2 is a prototype of many similar kinds of results. For instance, Zorn's Lemma also implies that every field has an algebraic closure and every ring-with-1 has a maximal ideal. In practice, if you don't want to use Zorn's Lemma, you can usually work with subspaces (or subrings or subfields) on which the existence of the desired object does not require it.

9 Well Ordering

Let S be a totally ordered set. We say that the ordering on S is a *well ordering* if every nonempty subset $T \subset S$ has a least element. That is, there is some $t \in T$ so that $t \leq u$ for all $u \in T$. The prototypical example is $\boldsymbol{N}$, with the usual ordering, but I'll give many other examples below.

Well Ordering Principle: Every set has a well ordering.

Zorn Implies Well Ordering: Let Σ denote the set of pairs $(T, \leq)$, where $T \subset S$ and $\leq$ is a well ordering on T. Note that Σ is nonempty because we can just take T to be a single element of S with the only possible ordering. We make Σ into a POSET by writing $A \leq B$ if $A \subset B$ and the ordering on B is compatible with the ordering on A, and $a < b$ for all $a \in A$ and $b \in B - A$.

Lemma 9.1 *Σ is chained.*

Proof: Let C be a chain of Σ. Let T_C be the union of all the members of C. Given $a, b \in C$ there are sets $A, B \in C$ so that $a \in A$ and $b \in B$. Say that $A \subset B$. Then $a, b \in B$. We use the ordering on B to compare a and b. This definition is independent of all choices. Thus T_C has a total ordering. By construction, the ordering on T_C extends the ordering on each member of C.

To show that $T_C \subset \Sigma$, we just have to show that our ordering on T_C is a well ordering. Let $\tau \subset T_C$ be any nonempty subset. There is some $A \in C$ so that $\tau \cap A$ is nonempty. Since A is well ordered, there is a least element a of $\tau \cap A$. Suppose a is not a least element of τ. Then $b < a$ for some $b \in \tau$. Note that b cannot belong to A, by definition of a. But then there exists $B \in C$ such that $b \in B - A$. But then $a < b$, by definition of our ordering on Σ. This contradiction shows that a is a least element for τ. Hence T_C is well ordered. Hence $T_C \subset \Sigma$. Hence Σ is chained. ♠

By Zorn's Lemma, Σ has some maximal element T, which is well ordered. If T is a proper subset of S, then there is some $a \in S - T$ and we can well order $T \cup \{a\}$ by using the ordering already on T and declaring that $t < a$for all $t \in T$. But then $T \cup \{a\} \subset \Sigma$ and $T \leq T \cup \{a\}$. This contradicts the maximality of T. Hence $T = S$. Hence S has a well ordering.

Well Ordering Implies Choice: Suppose that S is a partition. We can put a well ordering on each member of S. We then define T so that $t \in T$ if and only if t is the least element of some member of S. The set T then has one member in common with each member of S.

Transfinite Recursion: Suppose that S and T are nonempty sets with S being well ordered. Let s_0 be the least element of S and let t_0 be some element T. Let 2^T be the set of subsets of T. Suppose that $g : 2^T \to T$ is a function. One version of the *Principle of Transfinite Recursion* asserts that there exists a unique function $f : S \to T$ so that $f(s_0) = t_0$ and

$$f(a) = g(f(\{b|b < a\})), \qquad \forall a \in S. \tag{3}$$

First, let's parse this sentence. The set $\{b|b < a\}$ is the set of elements $b \in S$ with $b < a$. The set $f(\{b|b < a\})$ is a subset of T and hence a member of 2^T, so we may apply g to it and this gives us a specific element of T which we declare to be the value of $f(a)$.

To establish the Transfinite Recursion Principle, we call an element $c \in S$ *good* if there exists a function $f_c : S_c \to T$ satisfying the above relations on the set

$$S_c = \{d \in S : \ d \leq c\}.$$

We call c *excellent* if such a function is unique. Suppose c is good, and $f_{c,1}$ and $f_{c,2}$ are two such functions. If these functions do not agree, then there is some smallest $a \in S_c$ on which they disagree. But then Equation 3 says that they also agree at a. This is a contradiction. So, if c is good, then c is excellent.

Suppose all $c \in S$ are good. Then all $c \in S$ are excellent. We define $f(c) = f_d(c)$, where $d \in S$ is any element such that $c \leq d$. This well defined because the restrictions of f_{d_1} and d_{d_2} to S_c agree. But then our function f has the desired properties.

Alternatively, suppose that some $c \in S$ is not good. Then there is some smallest bad element, a. But then we can use Equation 3 to extend our function to a. This is a contradiction. Hence, all $c \in S$ are good, and we have our map.

In short, once we have a well ordering on S, we can also use transfinite recursion.

Well Ordering Implies Zorn: Let S be a chained POSET. We're going to assume that S can be well ordered, and we will deduce that S has a maximal element. Since the Well Ordering Principle implies the Axiom of Choice, we can also use the Axiom of Choice.

Lemma 9.2 *If S does not have a maximal element, then there exists a function $h : S \to S$ such that $h(x) > x$ for all $x \in S$.*

Proof: Let $a \in S$. Since S has no maximal element, the element a is not a maximal element. This means that there exists an element b such that $a < b$. We can simply define $h(a) = b$. That is the end of the proof, but let's reveal the argument more directly to be an application of the Axiom of Choice. Consider the subset of $P \subset S \times S$ consisting of pairs (a, b) where $a < b$. We get a partition of P into members, each of the form $\{a\} \times S$. By the Axiom of choice, we can choose one element from each member of this partition, and this defines the function h. ♠

Now we well order S, and call this well ordering $\prec$. Let s_0 denote the least element of S in the well ordering. We define $g : 2^S \to S$ as follows.

- If $C \in 2^S$ is a chain, we let $g(C) = h(b)$ where b is an upper bound relative to the POSET order. So $c < f(C)$ for all $c \in C$.
- If $C \in 2^S$ is not a chain we define $g(C) = s_0$. In fact, we never will consider non-chains in our application of transfinite recursion below.

By transfinite recursion, there is a unique function $f : S \to S$ so that $f(s_0) = s_0$ and $f(b) = g(f(\{b \prec a\}))$. We claim that

$$b \prec a \qquad \Longrightarrow \qquad f(b) < f(a).$$

Say that $a \in S$ is *bad* if this implication fails for some $b \prec a$. Suppose that some elements are bad. Let a be the least bad element in the well ordering. By definition $C = f(\{b | b \prec a\})$ is a chain in S. Hence $f(a) = g(C)$ is larger than all $c \in C$. This is a contradiction. Hence, there are no bad elements of S. This proves the claim. The same argument shows that $a \leq f(a)$ for all $a \in S$.

Since $f(S)$ is a chain, there is some upper bound b such that $f(a) \leq b$ for all $a \in S$. But then $a \leq f(a) \leq b$ for all $a \in S$. Hence b is a maximal element. This establishes Zorn's Lemma.

10 Comparing Well Ordered Sets

Two well ordered sets are *equivalent* if there is a bijection between them which respects the orders. Given a well ordered set S, a subset $T \subset S$ is *initial* if $a < b$ for all $a \in S$ and $b \in T - S$. Note that S is an initial subset of itself.

Lemma 10.1 *Suppose that S and T are well ordered sets. Then either S is equivalent to an initial subset of T or T is equivalent to an initial subset of S,*

Proof: It suffices to consider the case when both sets are nonempty. Let s_0 and t_0 be the least elements of S and T respectively. Let ξ be some element not in T and consider the new set $U = T \cup \xi$. We well order T^* by making ξ greater than all other element of T. We define a function $g : 2^{T^*} \to T^*$ as follows:

- If C is a proper subset of T then $g(C)$ is the least element of $T - C$.
- If $C = T$ then $g(T) = \xi$.
- If $\xi \in C$ then $g(C) = \xi$.

By transfinite recursion, there is a function $f : S \to T^*$ which satisfies Equation 3. By construction f maps initial subsets of S to initial subsets of T^*. If $\xi \notin f(S)$ then either $f(S) = T$ or $f(S)$ is an initial subset of T. In this case, f gives an equivalence between S and an initial subset of T. If $\xi \in f(S)$, there is some least element $a \in S$ so that $f(a) = \xi$. But then f gives equivalence between the initial subset $\{b|b < a\} \subset S$ and T. ♠

Lemma 10.2 *It is impossible for a well ordered set to be equivalent to a proper initial subset.*

Proof: Suppose that S is equivalent to a proper subset $T \subset S$. Note that $T = \{b|b < a\}$ for some $a \in S$. Call the element $a \in S$ *bad* if $\{b|b < a\}$ is isomorphic to S. Let a be the least bad element and let $T = \{b|b < a\}$. $f : T \to S$ be the order isomorphism. Note that $f^{-1}(T)$ is equivalent to T which is equivalent to S. But then $f^{-1}(a)$ is bad and $f^{-1}(a) < a$. This is a contradiction. ♠

11 Von Neumann Ordinals

A set S is a *Von Neumann Ordinal* if

1. The inclusion ordering on the elements of S is a well ordering.
2. Every member of S is an initial subset of S.
3. Every initial subset of S is a member of S.

The first condition requires some explanation. Any two elements $s, t \in S$ are sets in themselves, and we write $s \leq t$ if and only if $s \subset t$. We are saying that this ordering is a well ordering of S.

Every natural number is a Von Neumann ordinal. Below we will show that there are many more Von Neumann ordinals. First we prove some preliminary results.

Lemma 11.1 *Two equivalent Von Neumann Ordinals are identical.*

Proof: Let S and T be VNOs. Let $f : S \to T$ be the isomorphism constructed in the proof of Lemma 10.1. Note that $\emptyset$, the empty set, must be the least element of both S and T, and $f(\emptyset) = \emptyset$. Let $a \in S$ be the smallest element such that $f(a) \neq a$. The set $\{b | b < a\}$ contains all proper initial subsets of a (considered as a well ordered set) and so $\{b | b < a\}$ is just a. Since $f(b) = b$ for all $b < a$, we have $f(\{b | b < a\}) = \{b | b < a\}$. But then $f(a) = g(\{b | b < a\}) = a$. This is a contradiction. ♠

Lemma 11.2 *If S and T are Von Neumann ordinals, then either $S \subset T$ or $T \subset S$.*

Proof: An immediate consequence of the definition is that every member of a VNO is another VNO. Since S and T are both well ordered, we can swap these sets if necessary to guarantee that S is isomorphic to some initial subset of T. An initial subset of T is also a VNO. So, without loss of generality, it suffices to consider the case when S and T are equivalent. We have already treated this case. ♠

Now we can make a statement about well ordered sets which perhaps are not VNOs.

Lemma 11.3 *Any well ordered set is equivalent to a Von Neumann ordinal.*

Proof: Let S be a well ordered set. Suppose that this lemma is false. Then, for each Von Neumann ordinal T, there is some proper initial subset S_T which is equivalent to T. The set S_T has the form $\{b|b < a\}$ for some $a \in S$. The choice of a is unique, by Lemma 10.2.

There are two cases to consider. Either there is some $a \in S$ such that $a \notin S_T$ for any such T or else there is not. In the former case, we can replace S by the set $\{b|b < a\}$. That is, it suffices to consider the case when every $a \in S$ corresponds to some VNO.

So we have a set valued function $f(a)$ is the VNO such that $\{b|b < a\}$ is equivlent to $f(a)$. The Axiom of Replacement now allows us to construct a set $\widehat{S}$ whose members are precisely the Von Neumann Ordinals. Essentially, we are replacing each $a \in S$ by $f(a)$.

The contradiction is the the collection of all VNOs is a class, not a set. The way to see the problem is that $\widehat{S}$ satisfies all the properties needed to be a VNO. So $\widehat{S}$ is also a VNO. But then $\widehat{S}$ is a member of itself. This contradicts the Axiom of Foundation. ♠

One definition of an *ordinal number* is that it is an equivalence class of well ordered sets. This definition is somewhat problematic because ZF does not all one to take "collections of all sets" in an unrestricted manner. A better definition is that an ordinal number is a Von Neumann Ordinal, and then one could say that a well ordered set represents the VNO to which it is equivalent. It is sometimes nice to work with well ordered sets other than VNOs because sometimes they are easier to specify.

$\boldsymbol{N}$ with its usual ordering represents a VNO. However, there are many other well orderings of a countable set. For instance $\boldsymbol{N}$ is bijective with $\boldsymbol{N} \times \{1, 2\}$, and we define a well ordering on $\boldsymbol{N} \times \{1, 2\}$ with the rules

- $(m, 1) \leq (j, 2)$ for all n, m.
- $(m, j) \leq (n, j)$ if and only if $m \leq n$.

This new well ordering is not equivalent to the first one, and thus represents a different (larger) ordinal.

Indeed, the set of countable ordinals is uncountable; otherwise we could get the same contradiction as in Lemma 11.3. The set of countable ordinals is a beautiful thing. It is, itself, a well ordered set, and it represents the least uncountable ordinal.

12 Choice implies Zorn

Finally, we deduce Zorn's Lemma from the Axiom of Choice. Let S be a chained POSET. Suppose for the sake of contradiction that S does not have a maximal element. Lemma 9.2, which only uses the Axiom of Choice, gives us a function $h : S \to S$ such that $a < h(a)$ for all $a \in S$.

Let α be some Von Neumann Ordinal. Note that α is well ordered. The same argument as in "Well Ordering implies Zorn", namely a transfinite recursion argument based on the function h, gives us a function $f_\alpha : \alpha \to S$ with the property that $a < b \in \alpha$ implies that $f_\alpha(a) < f_\alpha(b) \in S$. The function is unique in the following sense. If α, β, γ are VNOs with $\alpha < \beta$ and $\alpha < \gamma$ then $f_\beta(\alpha) = f_\gamma(\alpha)$. For this reason, we can simply write $f(\alpha) = f_\beta(\alpha)$ where β is any VNO with $\alpha < \beta$.

Let $T \subset S$ denote those elements s such that $s = f(\alpha)$ for some ordinal α. Given that $\alpha < \beta$ implies that $f(\alpha) < f(\beta)$, there is at most one VNO α such that $f(\alpha) = a$. This means that we have set valued function f^{-1} defined on T, so that $f^{-1}(T)$ is a VNO, and every VNO arises in this way. But now we have the same contradiction as in the proof of Lemma 11.3. The only way out of the contradiction is that S has a maximal element.

13 Ultrafilters

Basic Definition: An *ultrafilter* on a set S is a map $F : S \to \{0, 1\}$ with the following properties:

1. $F(\emptyset) = 0$ and $F(S) = 1$.
2. If $A \subset B \subset S$ and $F(B) = 0$ then $F(A) = 0$.
3. If $A = B_1 \cup B_2$ and $B_1 \cap B_2 = \emptyset$ and $F(A) = 1$, then $F(B_1) + F(B_2) = 1$.

Call these properties U1,U2,U3. Intuitively, a "ghost" is hiding somewhere in the set S and the ultrafilter tells you whether or not that ghost is hiding inside one of the sets. If $F(A) = 1$, then the ghost is hiding inside A.

An ultrafilter F is called *principle* if there is some $s \in S$ so that $F(A) = 1$ if and only if $s \in A$. If F is not a principle ultrafilter, then F is a called a *nonpriciple ultrafilter*. When F is a nonprinciple ultrafilter, there is no set $A \subset S$ such that $F(B) = 1$ if and only if $A \subset B$. This situation is ruled out by the third condition above.

Reformulation: Let 2^S be the set of subsets of S. Given an ultrafilter F one can define a subset $\widehat{F} \subset 2^S$ by the rule that $A \in \widehat{F}$ if and only if $F(A) = 1$. The set $\widehat{F}$ has the following properties.

1. $\emptyset \notin \widehat{F}$ and $S \in \widehat{F}$.
2. If $A \in \widehat{F}$ and $A \subset B$ then $B \in \widehat{F}$.
3. If $A, B \in \widehat{F}$ then $A \cap B \in \widehat{F}$.
4. For any $A \subset S$, exactly one of the two sets A and $S - A$ belongs to $\widehat{F}$.

Call the properties above P1,P2,P3,P4. Properties P1 and P2 follow immediately from Conditions U1 and U2. Property P4 follows Condition U3, applied to the triple $(S, A, S - A)$. Here is a proof of P3: If $A, B \in \widehat{F}$ then $F(A) = F(B) = 1$. Suppose $F(A \cap B) = 0$. Since $A - B$ and $A \cap B$ partition A, we must have $f(A-B) = 1$. Likewise $f(B-A) = 1$. But then consider the set $C = (A-B) \cup (B-A)$. We have $f(C) = 1$ but $F(A-B)+F(B-A) = 2$, which violates U3.

Conversely, given $\widehat{F}$ which satisfies P1-P4 above. We define a function $F : S \to \{0, 1\}$ by the rule that $F(A) = 1$ if and only if $A \in \widehat{F}$.

Lemma 13.1 *The function F is an ultrafilter.*

Proof: Conditions U1 and U2 are immediate from Properties P1 and P2. Finally, let's verify U3. Suppose $A = B_1 \cup B_2$ and $B_1 \cup B_2 = \emptyset$ and $F(A) = 1$. Suppose $F(B_1) + F(B_2) = 2$. Then $B_1, B_2 \in \widehat{F}$, which means that $\emptyset = B_1 \cap B_2 \in \widehat{F}$. This is a contradiction.

Suppose that $F(B_1) + F(B_2) = 0$. For any set X let $X^c = S - X$. Since $B_j \notin \widehat{F}$ we have $B_j^c \in \widehat{F}$ for $j = 1, 2$, by P3. But then

$$A^c = (B_1 \cup B_2)^c = B_1^c \cap B_2^c \in \widehat{F}$$

by P4. But then both A and A^c belong to $\widehat{F}$, which contradicts P3. The only option left is that $F(B_1) + F(B_2) = 1$, as desired. ♠

So, function on S satisfying U1,U2,U3 is equivalent to a subset of 2^S satisfying P1,P2,P3,P4. This turns out to be a useful reformulation.

Filters: A *filter* is a subset $F \subset 2^S$ satisfying P1, P2, P3, but not necessarily P4. (For ease of notation we use F in place of $\widehat{F}$.) F is called *maximal* if there is no other filter G such that $F \subset G$ and $F \neq G$.

Lemma 13.2 *A maximal filter is an ultrafilter.*

Proof: Let F be a maximal filter. If F is not ultrafilter than there is some $D \subset S$ so that neither D nor D^c belongs to F. In this case we augment F by adding in all sets T such that $D \cap A \subset T$ and $A \in F$. Call the resulting subset G. Since we never add in the emptyset G satisfies P1.

To show that G satisfies P2, suppose that $T \in G$ and $T \subset U$. If $T \in F$ then we have $U \in F \subset G$. If $T \supset A \cap D$ for some $A \in F$, then $U \supset A \cap D$ as well, and $U \in G$. Hence G satisfies P2.

Now we show that G satisfies P3. Let $T, U \in G$. If $T, U \in F$, than the $T \cap U \in F$ because F is a filter. There are two remaining cases.

1. If $T \in F$ and $A \cap D \subset U$ for some $A \in F$, then $T \cap U = (A \cap T) \cap D$. Since $A, T \in F$, we have $A \cap T \in F$. Hence $T \cap U \in G$.

2. Suppose $A_1 \cap D \subset T$ and $A_2 \cap D \subset U$ for some $A_1, A_2 \in F$. Then
$$(A_1 \cap A_2) \cap D \subset T \cap U.$$
Since $A_1 \cap A_2 \in F$ we have $T \cap U \in G$ by definition.

Hence G is a filter which properly contains F. This is a contradiction. ♠

Theorem 13.3 *Let S be any set and let $\widehat{F}$ be a filter on S. Then there exists an ultrafilter $\widehat{\mu}$ such that $\widehat{F} \subset \widehat{\mu}$.*

Proof: Let Σ denote the set of filters which contain $\widehat{F}$. We order the elements of Σ by inclusion. Given any chain in Σ, we simply take the union of all the filters in the chain. This is again a filter containing $\widehat{F}$. Hence Σ is a chained POSET. By Zorn's Lemma, Σ has a maximal element. By Lemma 13.2, a maximal element of Σ is an ultrafilter. ♠

Corollary 13.4 *An infinite set S has a nonprinciple ultrafilter.*

Proof: Let $\widehat{F}$ denote the filter which contains precisely the sets whose complements are finite. $\widehat{F}$ is easily checked to be a filter. Now apply the theorem. ♠

14 Ultralimits

Now I'll give an example of what a nonprinciple ultrafilter can do for you. Let X be a compact metric space. We call an ultrafilter μ on X *focused* if there is some $x \in X$ such that $\mu(U) = 1$ for every open set $U \subset X$ which contains x. We call x the *focal point* of μ.

Lemma 14.1 *Let X be a compact metric space. Any ultrafilter on X is focused, and the focal point is unique.*

Proof: Let F be the ultrafilter. For any n, we can partition X into finitely sets which have diameter at most $1/n$. Exactly one of the sets, say S_n, is such that $F(S_n) = 1$. For different indices m and n, the two sets S_n and S_m cannot be disjoint. Therefore, there is a unique point $x \in X$, such that every neighborhood of x contains infinitely many of the sets S_n. In particular, $\mu(U) = 1$ for every open neighborhood U of x. This shows that x is a focal point of μ.

Suppose that y is a different focal point. Then we can find disjoint open sets U_x and U_y such that $x \in U_x$ and $y \in U_y$, and $\mu(U_x) = \mu(U_y) = 1$. This contradicts the definition of an ultrafilter. ♠

Suppose now that μ is an ultrafilter on $\boldsymbol{N}$ and $f : \boldsymbol{N} \to X$ is any map. That is, f defines a sequence in X. We define the *push-forward* ultrafilter $f_*\mu$ on X as follows:

$$f_*\mu(S) = \mu(f^{-1}(S)), \qquad \forall S \subset X. \tag{4}$$

This defines an ultrafilter on X. The ultrafilter $f_*\mu$ is focused at some point $x \in X$, and we write $x = \lim_\mu(f)$. We call this point the *ultralimit* of f. In this way, an ultrafilter picks out a limit to any sequence in a compact metric space.

When μ is a principle ultrafilter, the ultralimit is kind of silly. Say $\mu = 10$. In this case, the ultralimit of f is just $f(10)$. When μ is a non-principle ultrafilter, the limit is much nicer.

Lemma 14.2 *Suppose $f : \boldsymbol{N} \to X$ is a convergent sequence. Then*

$$\lim_{n\to\infty} f(n) = \lim_\mu f$$

for any non-principle ultrafilter μ.

Proof: Let $x = \lim_{n\to\infty} f(n)$. Let U be any open set containing x. The set $f^{-1}(U)$ is cofinite. Hence μ assigns 1 to this set. This shows that $f_*\mu$ focuses at x. ♠

We can apply this to the case of ordinary bounded sequences. In this case, we take $X = [-N, N]$, where N is chosen large enough so that it X contains all the sequences of interest to us. In this case, the ultralimit obeys all the usual limit laws:

$$\lim_F (a_n * b_n) = \lim_F a_n * \lim_F b_n, \qquad * \in \{+, -, \times, \div\}.$$

In the last case we require (as usual) that $\lim_F b_n$ is nonzero.

On the other hand, there is something a bit crazy about the ultralimit, because it picks out a limit for the sequence $0, 1, 0, 1, 0, 1, ...$ Also, for general sequences, $\{a_n\}$ the ultralimit is not "shift invariant". If $\{b_n\}$ is the sequence with $b_n = a_{n+1}$, then $\lim_F b_n$ need not equal $\lim_F a_n$.

15 The Space of Ultrafilters

Now I'm going to explain what is often called the *Stone-Cech compactification* of the natural numbers. Let $\boldsymbol{N}$ denote the set of natural numbers, as usual. The set $\boldsymbol{N}$ is equipped with the discrete topology: every subset of $\boldsymbol{N}$ is open.

Let $\beta\boldsymbol{N}$ denote the set of ultrafilters on $\boldsymbol{N}$. The set $\beta\boldsymbol{N}$ has a topology, which we'll now describe. We first define some distinguished subsets of $\beta\boldsymbol{N}$. For any subset $S \subset \boldsymbol{N}$, we define U_S to be the set of ultrafilters F such that $F(S) = 1$.

Lemma 15.1 *The sets U_S form a basis for a topology.*

Proof: We need to check two properties. First, we need to check that $\beta\boldsymbol{N}$ is covered by these sets. Each $F \in \beta\boldsymbol{N}$ is contained in some U_S, because we just take S to be any set where $F(S) = 1$. This takes care of the covering property.

Now we need to check the following: For any $F \in U_S \cap U_T$, there is some $U_R \subset U_S \cap U_T$ such that $F \in U_R$. For this, we note that $F(S) = 1$ and $F(T) = 1$, so that $F(S \cap T) = 1$. We take $R = S \cap T$, and we are done. ♠

A set $X \subset \beta\boldsymbol{N}$ is declared to be open if and only if, for each $x \in X$, there is some U_F such that $x \in U_F \subset X$. Given that our sets form a basis, this definition gives us a topology on $\beta\boldsymbol{N}$.

Lemma 15.2 *$\beta\boldsymbol{N}$ is compact.*

Proof: We want to prove that any covering of $\beta\boldsymbol{N}$ has a finite subcovering. Suppose that we have a covering of $\beta\boldsymbol{N}$ which has no finite subcovering. Let $\widehat{F}$ denote the set of subsets $S \subset \boldsymbol{N}$ such that U_{S^c} has a finite covering. Here $S^c = \boldsymbol{N} - S$.

We claim that $\widehat{F}$ is a filter. First, $\emptyset \notin \widehat{F}$ by assumption and obviously $\boldsymbol{N} \in \widehat{F}$. Next, if $A \in \widehat{F}$ and $A \subset B$, then U_{A^c} has a finite cover. Since $B^c \subset A^c$, we see that U_{B^c} has a finite cover. Hence $B \in \widehat{F}$. Finally, suppose that $S, T \in \widehat{F}$. Then U_{S^c} and U_{T^c} have finite covers. But $U_{S^c \cup T^c} \subset U_{S^c} \cup U_{T^c}$, so $U_{S^c \cup T^c}$ has a finite cover. Since $S^c \cup T^c = (S \cap T)^c$, we get $S \cap T \in \widehat{F}$. This proves the claim.

By Theorem 13.3, there is some ultrafilter $\widehat{\mu}$ which contains $\widehat{F}$. We claim that if $S \in \widehat{\mu}$ then U_S does not have a finite cover. To see this, note that $S^c \notin \widehat{\mu}$. Hence $S^c \notin \widehat{F}$. Hence U_S has no finite subcover. This proves the claim. Now we think of $\widehat{\mu}$ as a point in $\beta\boldsymbol{N}$, and we call this point μ. Since we have a covering, we have $\mu \in V$ for some element V of our covering. Hence, there is some set S such that $\mu \in U_S \subset V$. But then $S \in \widehat{\mu}$. But then U_S has no finite cover. This contradicts the fact that $U_S \subset V$. ♠

Lemma 15.3 *$\beta\boldsymbol{N}$ is Hausdorff and totally disconnected.*

Proof: Choose two ultrafilters μ_1 and μ_2. Since these ultrafilters do not coincide, there is some set S such that $\mu_1(S) = 0$ and $\mu_2(S) = 1$. But then $\mu_1(S^c) = 1$. Hence $\mu_1 \subset U_{S_c}$ and $\mu_2 \subset U_S$. We have placed our two points in disjoint open sets. Therefore, points in $\beta\boldsymbol{N}$ may be separated by open sets. Since $\beta\boldsymbol{N}$ is compact, this means that disjoint closed sets can also be separated by open sets.

Notice that, during the course of our proof, we showed that any two distinct points of $\beta\boldsymbol{N}$ can be placed in disjoint open sets whose union of $\beta\boldsymbol{N}$. This proves that $\beta\boldsymbol{N}$ is totally disconnected. ♠

We think of $\boldsymbol{N}$ as a subset of $\beta\boldsymbol{N}$: The number n is identified with the principle ultrafilter associated with n.

Lemma 15.4 *$\boldsymbol{N}$ is a dense subset of $\beta\boldsymbol{N}$ and the subset topology on $\boldsymbol{N}$ is just the discrete topology.*

Proof: Let F be any ultrafilter. We just need to show that every basis neighborhood of F contains a principle ultrafilter. Let U_S be such a neighborhood. We simply choose some $n \in S$ and then the principle ultrafilter associated to n is contained in U_S.

By construction, the set $U_{\{n\}}$ just consists of the single ultrafilter associated to n. Thus, these principle ultrafilters are open sets in the topology on $\beta\boldsymbol{N}$. ♠

Lemma 15.5 *Let Y be a compact metric space and let $f : \boldsymbol{N} \to Y$ be any map. Then f extends uniquely to a continuous map from $\beta\boldsymbol{N}$ to Y.*

Proof: Let μ be any ultrafilter. We simply define $\beta f(\mu) = \lim_\mu f$, the ultralimit of the map. To see that this is continuous, let $U \subset Y$ be open and consider the set $\beta f^{-1}(U)$. Choose any ultrafilter μ in this set. So, $\beta f(\mu) \in U$. But then there is some subset $S \subset \boldsymbol{N}$ such that $\mu(S) = 1$ and $f(S) \subset U$. By construction $U_S \subset \beta f^{-1}(U)$. ♠