ALL THE FILTER STUFF IN ONE PLACE
AND SEVERAL THINGS EQUIVALENT TO THE CHOICE AXIOM

1. *Definition and Fundamental Properties of Filters.*

Definition: Let X be a nonempty set. A **filter base** in X is a subfamily $\mathfrak{B} \subseteq \mathbf{2}^X$ satisfying the following requirements:

(1) $\emptyset \notin \mathfrak{B}$;

(2) For every two sets B_1, $B_2 \in \mathfrak{B}$ there exists a set $B_3 \in \mathfrak{B}$ with $B_3 \subseteq B_1 \cap B_2$ (and thus necessarily $B_1 \cap B_2 \neq \emptyset$).

Note that the sets in a filter base are not required to "get small" any way other than set-theoretically: indeed, we're not in a metric space and have no way to say "small" except under the relation of set-inclusion.

The next definitions may seem unnatural, strange and technical. They are also extremely handy.

Definition: Let X be a nonempty set. A **filter** in X is a subfamily $\mathfrak{F} \subseteq \mathbf{2}^X$ satisfying the following requirements:

(1) $\emptyset \notin \mathfrak{F}$;

(2) For every two sets F_1, $F_2 \in \mathfrak{F}$ the set $F_1 \cap F_2$ belongs to $\mathfrak{F}$ (so $\mathfrak{F}$ is closed under the operation of intersecting finitely many of its elements);

(3) If $S \in \mathbf{2}^X$ is a superset of some $F \in \mathfrak{F}$, then $S \in \mathfrak{F}$.

Definition: Let X be a nonempty set and $\mathfrak{B}$ a filter base in X. The **filter based on** or **generated by** $\mathfrak{B}$ is the family

$$\mathfrak{F}(\mathfrak{B}) = \{S \in \mathbf{2}^X : \text{For some } B \in \mathfrak{B},\ S \supseteq B\}\ .$$

In words, $\mathfrak{F}(\mathfrak{B})$ is the family of all supersets of elements of $\mathfrak{B}$.

Definition: Let X be a nonempty set and $Y \subseteq X$ a nonempty subset of X. A filter $\mathfrak{G} \subseteq \mathbf{2}^X$ is said to be **based in** Y if there exists a filter base $\mathfrak{B} \subseteq \mathbf{2}^Y$ for which $\mathfrak{F}(\mathfrak{B}) = \mathfrak{G}$.

Of these definitions, the first and third may make some sense, since the definition of a filter base describes the family $\mathfrak{C}$ of sets in the "generalized Cantor intersection theorem" except for dropping the requirement that they be closed (which would not make sense in a general set X anyway). The definition of a filter is motivated in two ways. The first is that, since one is only interested in the "set-theoretically small" sets in a filter base, one can throw in the intersections and the supersets without making any difference in the "small" sets—which are the ones in which one is really interested—one can look at $\mathfrak{F}(\mathfrak{B})$ as just being a "maximal version of $\mathfrak{B}$." What is really going on when one makes this definition, however, is the fact that a filter as thus defined is a **proper dual ideal** in the Boolean algebra $(\mathbf{2}^X, \cup, \cap)$. That is: consider the family $\mathfrak{J}$ of complements in X of the elements of a given filter $\mathfrak{G}$. The duals (*i.e.*, statements in terms of complements in X) of the defining properties of a filter read:

(1) $X \notin \mathfrak{J}$;

(2) For every two sets S_1, $S_2 \in \mathfrak{J}$ the set $S_1 \cup S_2$ belongs to $\mathfrak{J}$ (so $\mathfrak{J}$ is closed under the operation of uniting finitely many of its elements);

(3) If $S \in \mathbf{2}^X$ is a subset of some $J \in \mathfrak{J}$, then $S \in \mathfrak{J}$.

Now (2) says that $\mathfrak{J}$ is closed under finite union = Boolean addition and (3) is easily seen to be equivalent to saying: if $S \in \mathbf{2}^X$ and $J \in \mathfrak{J}$, then the "Boolean-algebra product" $S \cap J \in \mathfrak{J}$. These are the properties of an **ideal** in the Boolean algebra $\mathbf{2}^X$. Now (1) is seen to say that the identity element X for "multiplication" = $\cap$ does not belong to $\mathfrak{J}$, which is then a **proper ideal** in $\mathbf{2}^X$. Hence the notion of a filter is that of a *proper* *dual ideal.*[1] The complements of the elements of a filter base $\mathfrak{B}$ then will appear as a **set of generators** for the ideal formed by the complements of the elements of the filter $\mathfrak{F}(\mathfrak{B})$. We don't intend to make a big fuss about this in this course, but filters virtually provide one with a way to turn the notion of convergence in a metric (or, more generally, in a topology) into algebra in the Boolean algebra $\mathbf{2}^X$.

The following definition of convergence of a filter base (or filter: a filter satisfies the definition of filter base) in a metric space should seem quite natural.

Definition: Let (X, d) be a metric space. A filter base $\mathfrak{B}$ in X **converges to** a point $x_0 \in X$ if for every $\epsilon > 0$ there exists a set $B_\epsilon \in \mathfrak{B}$ such that $B_\epsilon \subseteq B(x_0; \epsilon)$.

This should *look like convergence.* For every $\epsilon > 0$ there is a set $B_\epsilon \in \mathfrak{B}$ all of whose points are within ϵ distance of x_0 in the metric d. The points of other B's in $\mathfrak{B}$ that are contained in B_ϵ will be at least ϵ-close, and maybe even closer. But there is a neat way to rewrite this definition in terms of pure set-inclusion.

Definition: Let (X, d) be a metric space and $x \in X$. The **neighborhood filter of** x is the filter

$$\mathfrak{N}(x) = \{S \in \mathbf{2}^X : \text{there exists } \eta > 0 \text{ such that } B(x; \eta) \subseteq S\} .$$

An alternative way to say this is: observe that the open d-balls centered at x form a filter base: then $\mathfrak{N}(x)$ is the filter based on the family of open d-balls centered at x. This language agrees with the standard definition of **neighborhood of a point** x to mean a set S that contains an open ball centered on x.

Proposition (or alternate definition): Let (X, d) be a metric space. A filter base $\mathfrak{B}$ in X **converges to** a point $x_0 \in X$ if and only if $\mathfrak{N}(x) \subseteq \mathfrak{F}(\mathfrak{B})$.

Proof. Suppose $\mathfrak{B}$ converges to x. Then for any $\eta > 0$ there exists $B_\eta \in \mathfrak{B}$ with $B_\eta \subseteq B(x; \eta)$. Since $\eta > 0$ was arbitrary, we see that every set S that contains some $B(x; \eta)$—*i.e.*, every $S \in \mathfrak{N}(x)$—contains some $B_\eta \in \mathfrak{B}$, and so $S \in \mathfrak{F}(\mathfrak{B})$ by definition. We have shown that $\mathfrak{N}(x) \subseteq \mathfrak{F}(\mathfrak{B})$. On the other hand, if that inclusion holds then given any $\epsilon > 0$, since the ball $B(x; \epsilon) \in \mathfrak{N}(x)$, the ball $B(x; \epsilon)$ must belong to $\mathfrak{F}(\mathfrak{B})$ and the sets of that filter are precisely those that contain some element of $\mathfrak{B}$: so there must exist some $B_\epsilon \in \mathfrak{B}$ with $B_\epsilon \subseteq B(x; \epsilon)$. Since $\epsilon > 0$ was arbitrary, the condition defining "$\mathfrak{B}$ converges to x" is satisfied.

The next definition and proposition should be no surprise.

Definition: Let (X, d) be a metric space. A filter base $\mathfrak{B}$ in X is a **Cauchy filter base** if for every $\epsilon > 0$ there exists a set $B_\epsilon \in \mathfrak{B}$ such that $\operatorname{diam}(B_\epsilon) < \epsilon$. A filter $\mathfrak{F}$ is a **Cauchy filter** if $\mathfrak{F} = \mathfrak{F}(\mathfrak{B})$ for some Cauchy filter base, or (which comes to the same thing) if $\mathfrak{F}$ satisfies the defining condition for a Cauchy filter base.

Proposition: If (X, d) is a complete metric space, then every Cauchy filter (base) in X has a limit, *i.e.*, converges to some point of X.

Proof. Let $\mathfrak{B}$ be a Cauchy filter base, and let $\mathfrak{C} = \{\overline{B} : B \in \mathfrak{B}\}$, the set of closures of the elements of $\mathfrak{B}$. Since[2] $\operatorname{diam}(B) = \operatorname{diam}(\overline{B})$ for any $B \subseteq X$, the family $\mathfrak{C}$ satisfies the hypotheses of the "generalized Cantor intersection theorem," and so there is a unique point $x_0 \in \bigcap \mathfrak{C} = \bigcap\{\overline{B} : B \in \mathfrak{B}\}$. Given any $\epsilon > 0$ there is a $B_\epsilon \in \mathfrak{B}$ with $\operatorname{diam}(\overline{B_\epsilon}) < \epsilon$, so $B_\epsilon \subseteq \overline{B_\epsilon} \subseteq B(x_0; \epsilon)$, and that is the condition defining convergence of $\mathfrak{B}$ to x_0.

The definition of "convergent sequence" that we all know and love is easily subsumed into the filter(-base) framework. If $\{x_k\}_{k=1}^\infty$ is a sequence in a set X, then let $T_n = \{x_k : k \geq n\}$ for each $n \in \mathbb{N}$; this set is called the **tail of** $\{x_k\}_{k=1}^\infty$ **beginning at** n. Evidently these sets form a filter base, since $T_m \cap T_n = T_{\max\{m,n\}}$;

[1] As everyone who knows a little ring theory knows, *maximal ideals* in commutative rings are very interesting objects. Maximal filters, called *ultrafilters*, are very interesting objects in this context: in a Hausdorff topological space, compactness can be characterized in terms of these filters. That every filter is contained in an ultrafilter follows easily from Zorn's lemma, a statement equivalent to the axiom of choice that is to be discussed later in this course. On the other hand, it's easy to exhibit ultrafilters: for every $x \in X$, the *principal ultrafilter at* x, namely $\{S \subseteq X : x \in S\}$, is clearly an ultrafilter.

[2] The reader should verify this.

the filter generated by this filter base is called the **elementary filter in X generated by the sequence** $\{x_k\}_{k=1}^\infty$. If (X,d) is a metric space, then the ϵ-N condition defining convergence of $\{x_k\}_{k=1}^\infty$ to x_0 is the same as the condition defining convergence of the elementary filter generated by it to x_0, namely that for every $\epsilon > 0$ there should exist an $N \in \mathbb{N}$ such that $k \geq N \Rightarrow d(x_k, x_0) < \epsilon$, which is the same thing as saying that there should exist an element T_N of the filter base of tails of $\{x_k\}_{k=1}^\infty$ that is contained in $B(x_0;\epsilon)$. The condition defining a Cauchy sequence is the same as (or trivially logically equivalent to) the condition making the filter base of tails of the sequence a Cauchy filter base, namely that for every $\epsilon > 0$ there should exist an $N \in \mathbb{N}$ such that $k,\, m \geq N \Rightarrow d(x_k, x_m) \leq \epsilon$, which is the same thing as saying that there should exist an element T_N of the filter base of tails of $\{x_k\}_{k=1}^\infty$ of diameter $\leq \epsilon$.

The filter-base analogues of "subsequence" and "accumulation point" are again phrased in terms of set-inclusion. The place to start is probably the answer to the question: given a family of filters (or filter bases), when are they all contained in one big filter?

Proposition: Let $\{\mathfrak{B}_\iota : \iota \in I\}$ be a family of filters (or filter bases). A necessary and sufficient condition for all of them to be contained in a filter $\mathfrak{H}$ is that no finite intersection of the elements of $\bigcup\limits_{\iota \in I} \mathfrak{B}_\iota$ be empty.

Proof: Evidently the condition is necessary, because if any such intersection were empty, the family $\mathfrak{H} \supseteq \bigcup_{\iota \in I} \mathfrak{B}_\iota$ would violate one of the defining conditions for a filter (base). On the other hand, if the condition is satisfied, then the family of all intersections $B_1 \cap \cdots \cap B_k$ of finite subfamilies of $\bigcup_{\iota \in I} \mathfrak{B}_\iota$ satisfies the defining condition for a filter base, so the filter $\mathfrak{F}\left(\bigcup\limits_{\iota \in I} \mathfrak{B}_\iota\right)$ is a filter finer than all the filters (generated by the filter bases) $\{\mathfrak{B}_\iota : \iota \in I\}$.

{Remark: Note that the filter just defined is the **coarsest filter finer than all the elements of the family** $\{\mathfrak{B}_\iota : \iota \in I\}$.}

Well, the notion corresponding to subsequence is easy: if $\mathfrak{H}$ is a filter finer than another filter $\mathfrak{G}$, *i.e.*, if $\mathfrak{H} \supseteq \mathfrak{G}$, then "$\mathfrak{H}$ is like a subsequence of $\mathfrak{G}$." Certainly if $\{x_k\}_{k=1}^\infty$ is a sequence in a set X and $\{x_{k_j}\}_{j=1}^\infty$ is a subsequence of the given sequence, then each tail $S_n = \{x_{k_j} : j \geq n\}$ of the subsequence is a subset of the tail $T_n = \{x_k : k \geq n\}$ of the original sequence, so the filter generated by the tails of the subsequence contains—*i.e.*, is finer than—the filter generated by the tails of the original sequence. The analogue of **accumulation point** would then be: a point x_0 in a metric space X is an **adherent point** of a filter base $\mathfrak{B}$ if there is a filter $\mathfrak{G}$ that contains both $\mathfrak{B}$ and the neigborhood filter $\mathfrak{N}(x_0)$. It's obvious that this happens if and only if it happens for the filter $\mathfrak{F}(\mathfrak{B})$ generated by $\mathfrak{B}$, and in fact we can state the following proposition, all of whose assertions are more-or-less "true by definition."

Proposition: The following conditions are equivalent for a filter base $\mathfrak{B}$ (or the filter $\mathfrak{F}(\mathfrak{B})$ it generates) in a metric space (X,d):

(1) x_0 is an adherent point of $\mathfrak{B}$;

(2) Some filter $\mathfrak{G}$ finer than $\mathfrak{F}(\mathfrak{B})$ converges to x_0;

(3) Every ball $B(x_0;\delta)$ has nonempty intersection with every $B \in \mathfrak{B}$;

(4) $x_0 \in \bigcap\{\overline{B} : B \in \mathfrak{B}\}$.

Proof. (1) $\Rightarrow$ (2) is immediate: to say that x_0 is an **adherent point** of $\mathfrak{B}$ is to say that there is a filter $\mathfrak{G}$ that contains both $\mathfrak{B}$ and the neigborhood filter $\mathfrak{N}(x_0)$, and $\mathfrak{G} \supseteq \mathfrak{N}(x_0)$ is the definition of "$\mathfrak{G}$ converges to x_0." (2) $\Rightarrow$ (3): the filter $\mathfrak{G}$ contains every $B(x_0;\delta)$ and every $B \in \mathfrak{B}$ and the intersection of two sets of $\mathfrak{G}$ is never empty, so (3) must hold. (3) $\Rightarrow$ (4): if (4) fails then there is some $B \in \mathfrak{B}$ for which $x_0 \in X \setminus \overline{B}$, and since that is an open set there is a ball $B(x_0;\delta) \subseteq X \setminus \overline{B}$, but then obviously $B(x_0;\delta) \cap B = \emptyset$ and so (3) cannot hold. (4) $\Rightarrow$ (1): if (1) fails then the family of all finite intersections of members of the family $\mathfrak{B} \cup \mathfrak{N}(x_0)$ fails to be a filter base, and so some set of the form $B_1 \cap \cdots \cap B_k \cap U(x_0) = \emptyset$, where the $B_j \in \mathfrak{B}$ and $U(x_0) \in \mathfrak{N}(x_0)$. If $B_0 \in \mathfrak{B}$ is such that $B_0 \subseteq (B_1 \cap \cdots \cap B_k)$ and if $B(x_0,\delta) \subseteq U(x_0)$ then also $B_0 \subseteq X \setminus B(x_0,\delta)$ and thus also $\overline{B_0} \subseteq X \setminus B(x_0,\delta)$ (the complement of an open ball is closed). It follows that $x_0 \notin \bigcap\{\overline{B} : B \in \mathfrak{B}\}$, *i.e.*, that (4) fails.

One can thus bring the well-known fact that "a Cauchy sequence with an accumulation point converges to it" into the filter framework: cf. the proposition on p. 2 above.

Proposition: For a Cauchy filter (base) $\mathfrak{B}$ to converge to a point $x_0 \in X$ it is necessary and sufficient that x_0 be an adherent point of $\mathfrak{B}$.

Proof. Evidently (why?) any convergent filter (base) has its limit as an adherent point. On the other hand, if $\mathfrak{B}$ is a Cauchy filter (base), then if (as before) we let $\mathfrak{C} = \{\overline{B} : B \in \mathfrak{B}\}$, the set of closures of the elements of $\mathfrak{B}$, we see that the adherent points of $\mathfrak{B}$ are just the elements of $\bigcap \mathfrak{C} = \bigcap \{\overline{B} : B \in \mathfrak{B}\}$. To assume that $\mathfrak{B}$ has an adherent point is then to assume the existence of $x_0 \in \bigcap \mathfrak{C} = \bigcap \{\overline{B} : B \in \mathfrak{B}\}$. Given any $\epsilon > 0$ there is a $B_\epsilon \in \mathfrak{B}$ with $\operatorname{diam}(\overline{B_\epsilon}) = \operatorname{diam}(B_\epsilon) < \epsilon$, and since $x_0 \in \overline{B_\epsilon}$ one has $B_\epsilon \subseteq \overline{B_\epsilon} \subseteq B(x_0; \epsilon)$; thus, since $\epsilon > 0$ was arbitrary, $\mathfrak{B}$ satisfies the condition defining its convergence to x_0.

Chasing definitions further, we can characterize compactness in terms of filter bases:

Proposition: A metric space (X, d) is compact if and only if every filter (base) in X has an adherent point.

Proof. This is just a restatement of the fact that a metric space (X, d) is compact if and only if: for each family $\mathfrak{F}$ of closed subsets of X that has the finite intersection property, the intersection $\bigcap \{F : F \in \mathfrak{F}\}$ of *all* the elements of the family is nonempty. Any such family $\mathfrak{F}$ is a filter base, and its intersection is the set of all adherent points of that filter base: so if every filter base in X has an adherent point, then the intersection of any family of closed sets with the finite intersection property is nonempty. Conversely, if that characterization of compactness of X is known to hold, then since the family of closed sets $\{\overline{B} : B \in \mathfrak{B}\}$ has *(a fortiori!)* the finite intersection property whenever $\mathfrak{B}$ is a filter (base), one must have $\bigcap \{\overline{B} : B \in \mathfrak{B}\} \neq \emptyset$ and so the set of adherent points of $\mathfrak{B}$ must be nonempty.

One can do lim-inf and lim-sup for filters in $\mathbb{R}$ in pretty much the way one does them for sequences. Given a filter (base) $\mathfrak{B} \subseteq \mathbf{2}^{\mathbb{R}}$, we can define its lim-sup and lim-inf in $\overline{\mathbb{R}}$ by

$$\limsup \mathfrak{B} = \inf\{\sup(B) : B \in \mathfrak{B}\} \ ;$$
$$\liminf \mathfrak{B} = \sup\{\inf(B) : B \in \mathfrak{B}\} \ .$$

It is easy to see that these definitions give the same result for $\mathfrak{B}$ and for $\mathfrak{F}(\mathfrak{B})$. Say that a filter (base) $\mathfrak{B} \subseteq \mathbf{2}^{\mathbb{R}}$ is **bounded above** (or **below**, respectively, or just **bounded**) if there is some set $B \in \mathfrak{B}$ that is bounded above (or below, or bounded, respectively); this comes to the same condition for $\mathfrak{B}$ or for $\mathfrak{F}(\mathfrak{B})$. It is easy to see that if $\mathfrak{B}$ is bounded above, then $\limsup \mathfrak{B} < +\infty$ and that if also $\limsup \mathfrak{B} \in \mathbb{R}$, then $\limsup \mathfrak{B}$ is the greatest adherent point of $\mathfrak{B}$. On one hand $M = \limsup \mathfrak{B}$ is an adherent point: given any $\epsilon > 0$, $M + \epsilon$ is "too large for an inf" so there is some $B_1 \in \mathfrak{B}$ with $\sup B_1 < M + \epsilon$; on the other hand, since $M - \epsilon < \sup B$ for every $B \in \mathfrak{B}$, any set $B_2 \cap B_1$, where $B_2 \in \mathfrak{B}$, contains points $> M - \epsilon$. Thus $B \cap B(M; \epsilon) \neq \emptyset$ for every $B \in \mathfrak{B}$: $M = \limsup \mathfrak{B}$ is an adherent point of $\mathfrak{B}$. On the other hand, if $p > 0$ then $M + p \in \mathbb{R}$ cannot be an adherent point of $\mathfrak{B}$, because there is some $B \in \mathfrak{B}$ with $\sup B < M + p$ (since $M + p$ is not a lower bound for the sup's); if $\epsilon > 0$ is such that $\sup B < M + p - \epsilon$ then obviously $\overline{B} \cap B(M + p, \epsilon) = \emptyset$ and $M = p$ cannot be an adherent point of $\mathfrak{B}$. The lim-inf is dually characterized as the smallest adherent point of $\mathfrak{B}$. It is routine to verify that $\lim \mathfrak{B} = x_0 \in \mathbb{R}$ if and only if $\liminf \mathfrak{B} = x_0 = \limsup \mathfrak{B}$, and the detail-checking can safely be left to the reader.

2. *Directed Sets and Nets.*

It is handy to have things that are just a little more general than sequences with which one can define "filter base of tails" and such like things. **Nets** are an example.

Definition: Let J be a set. A **(weak) partial order** on J is a binary relation $\geq$ on J, usually read "greater than," satisfying the conditions

(1) For j, k, $m \in J$: if $j \geq k$ and $k \geq m$, then $j \geq m$;

(2) For $j \in J$: $j \geq j$.

The possibility that $j \geq k$ and $k \geq j$ without $j = k$ is specifically *permitted.* Devotees of axiomatic set theory will want to regard relations as subsets of $J \times J$. If a relation $\geq$ is present in a context, then $j \leq k$ will mean $k \geq j$.

Definition: The relation $\geq$ **directs the set** J **upward** if for any two members $j, k \in J$ there exists $m \in J$ such that $m \geq j$ and $m \geq k$. {It directs J **downward** if $\leq$ directs the set upward. Since there is an obvious order-reversing way to turn relations going one way to relations going the other, we shall make most of our constructions with upward-directed sets.}

Definition: A **net in a set** X is a triple $(x(\cdot), J, \geq)$ consisting of a function $x : J \to X$, an "index set" J, and a partial order $\geq$ under which J is directed upward.

It is customary to write the argument of $x(\cdot)$ as a subscript and to indicate a net simply as $\{x_j : j \in J\}$. {At this point I used to apologize for this notation: it is customary to write nets as if they were sequences and to use lower-case Greek letters for the elements of the (directed) index set, so that $\{x_j : j \in J\}$ would appear, *e.g.*, as $\{x_\iota\}_{\iota \in J}$. But I have given up apologizing: the notation as you see it here looks just enough like a sequence to make one realize that nets are "almost the same as sequences" and just enough not like a sequence to make one stop and think and perhaps check that the intuitively-reasonable argument one wants to use is actually valid in the more-general-than-sequence context.} The nets that are the model examples for all nets are the sequences $\{x_n : n \in \mathbb{N}\}$ indexed by the natural numbers under their usual order (which is total and thus directed upward).[3] The notion of a net is simply a device to enable one to talk about convergence in a sequence-like way but to keep things natural even when the "converging objects" cannot be put into a sequence in a natural way.

The following definition mimics the model.

Definition: Let (X, d) be a metric space and $\{x_j : j \in J\}$ be a net in X. One says that $\{x_j : j \in J\}$ **converges to** $x_0 \in X$, or writes

$$\lim_{j \in J} x_j = x_0$$

{or things that are obviously similar} if for every $\epsilon > 0$ there exists $K \in J$ such that

$$j \geq K \quad \Rightarrow \quad d(x_0, x_j) < \epsilon\ .$$

One says that a net is a **Cauchy net** {or satisfies the Cauchy condition, etc.} if for every $\epsilon > 0$ there exists $K \in J$ such that

$$j,\, k \geq K \quad \Rightarrow \quad d(x_j, x_k) < \epsilon\ .$$

It is easily seen (but seeing it requires use of the directedness of the index set) that if a net in X has a limit then that limit is unique. The notion of a net that converges to a point can be defined in a general topological space, just as it can for a sequence ("Cauchy" has no meaning in that setting); the uniqueness of limits of nets is equivalent to the Hausdorff property for the topology.

These definitions transcribe those one makes for sequences in the standard treatment of metric spaces, and there are a number of standard exercises that the reader can perform with them. For example, it is true that if (X, d) and (Y, r) are metric spaces, $x_0 \in X$, and $f : X \to Y$ is a function with $y_0 = f(x_0)$, then f is continuous at x_0 if and only if whenever $\{x_j : j \in J\}$ is a net in X that converges to x_0, then $\{f(x_j) : j \in J\}$ is a net that converges to y_0. Similarly, a subset $A \subseteq X$ is closed if and only if it contains the limit of any convergent net whose points lie in A. The advantage of the use of nets over the use of sequences in general topology is that there are "enough" nets to describe the topology, because closed subsets can be characterized as we just did; there usually aren't enough sequences to characterize closed sets, although

[3] In fact, nets are called "generalized sequences" in N. Dunford and J. T. Schwartz, *Linear Operators*, Part I, Ch. I.7, p. 26 ff., which is a place to find some of this material between covers. They are also discussed in J. L. Kelley, *General Topology*, Ch. 2, "Moore-Smith Convergence." His exercises G and H, pp. 77–80, relate to the material at hand as well as to integration along paths. I find it embarrassing to give forty-year-old references for this material, but newer ones are hard to find. It all seems to have become folklore: everybody knows it, but nobody bothers to write it down.

in a metric space there are. Similarly, there are enough nets in general topological spaces to characterize continuity in terms of preservation of convergence of nets.

The reason for this, of course, is that one can make filters from nets. Given a net $\{x_j : j \in J\}$ in X, let $T_n = \{x_j : j \geq n\}$ for each $n \in J$; this set is called the **tail of** $\{x_j : j \in J\}$ **beginning at** n. Evidently these sets form a filter base, since given k, $m \in J$ there will be some $n \in J$ with $n \geq m$ and $n \geq k$, and then $T_k \cap T_m \supseteq T_n$; the filter generated by this filter base is called the **filter of tails of the net** $\{x_j : j \in J\}$. If (X, d) is a metric space, then the "ϵ-n" condition defining convergence of $\{x_j : j \in J\}$ to x_0 is the same as the condition defining convergence of an elementary filter generated by a sequence to x_0, namely that for every $\epsilon > 0$ there should exist an $n \in J$ such that $j \geq n \Rightarrow d(x_j, x_0) < \epsilon$, which is the same thing as saying that there should exist an element T_n of the filter base of tails of $\{x_j : j \in J\}$ that is contained in $B(x_0; \epsilon)$. The condition defining a **Cauchy net** is the same as (or trivially logically equivalent to) the condition making the filter (base) of tails of the net a Cauchy filter base, namely that for every $\epsilon > 0$ there should exist an $n \in J$ such that k, $m \geq N \Rightarrow d(x_j, x_m) \leq \epsilon$, which is the same thing as saying that there should exist an element T_n of the filter base of tails of $\{x_j : j \in J\}$ of diameter $\leq \epsilon$.

There are people who build entire theories of generalized convergence on nets rather than filters. I think this is unwise, because while the notions of "finer filter in X" and "all filters in X finer than $\mathfrak{G}$" only involve the power set of $\mathbf{2}^X$, the notion of "subnet" is less clear. *A priori,* the notion of a **subnet of** $\{x_j : j \in J\}$ should begin with a net $\alpha \mapsto j_\alpha$ defined on a(n upward-)directed set $(A, \geq)$, such that $\beta \geq \alpha \Rightarrow j_\beta \geq j_\alpha$; then $\{x_{j_\alpha} : \alpha \in A\}$ is a subnet. There is no problem with this, but the "set of all subnets" of a given net, unlike the "set of all subsequences" of a given sequence, gets into potential trouble with paradoxes associated with the "set of all sets"—*any* set A might have a partial order under which it was directed upward. In any event, because of the filter bases than occur naturally in Riemann-Stieltjes integration, I prefer to make filters the primary notion of convergence generalizing that of sequences. Of course I will use subnets in situations in which they may be convenient.

There is at least one situation in which nets have an advantage over filters: it is easy to say what a monotone net is. If $\{x_j : j \in J\}$ is a net in $\mathbb{R}$, then we shall say it is **monotone** (**increasing**) if $j \geq i$ in $J \Rightarrow x_j \geq x_i$ in $\mathbb{R}$. (**Monotone decreasing** nets are defined dually, of course.) It is routine to show, by imitating what everybody has seen done with increasing real sequences, that a monotone-increasing net in $\mathbb{R}$ converges if and only if it is bounded above, and necessarily converges to $\sup\{x_j : j \in J\}$.

3.A. *Some Elementary Considerations about Order.*

A **relation** from a set X to a set Y is a *nonempty* subset $R \subseteq X \times Y$. As in the case of functions (which are just a special case of relations), this definition makes the notion that one relation extends another (or that one relation is a restriction of another) a simple matter of set-inclusion: if R is a relation from X to Y and S in another, then S **extends** R, or equivalently R **is a restriction of** S, is defined to mean $R \subseteq S$. The statement $(x, y) \in R$ is usually written xRy. The situation of greatest interest to us at the moment will be the one where "X" and "Y" are the same set X, and R is a **partial order(ing) of** (or **on**) X. These relations are usually written with something resembling the "less-than-or-equal-to" sign, so R might be written as $\preceq$ and therefore $(x, y) \in \preceq$ has the more familiar form $x \preceq y$. Formally, a relation $\preceq$ on a set X is a partial order on X if it has the two properties of

(1) **antisymmetry**: $x \preceq y$ and $y \preceq x \Rightarrow x = y$;

(2) **transitivity**: $x \preceq y$ and $y \preceq z \Rightarrow x \preceq z$.

If $Z \subseteq X$ is a nonempty subset, then by the **relativized** ($\preceq$-)**order on** Z we simply mean $(\preceq) \cap (Z \times Z)$, *i.e.*, what one gets by looking at the order relation (given on X) only as it pertains to comparing elements of the subset Z. If there is a $\preceq$ in a context, then $x \succeq y$ means the same thing as $y \preceq x$; $x \prec y$ means $x \preceq y$ but $x \neq y$, and similarly for $\succ$.

A **linear** (or **total** or **simple**) **ordering** $\preceq$ of a set X is one that has the additional property of **trichotomy**, *i.e.*, given x, $y \in X$ exactly one of the conditions $x \prec y$, $x = y$ or $x \succ y$ holds. Even if $(X, \preceq)$ is not linearly ordered, however, there may be nonempty subsets $C \subseteq X$ for which C is linearly ordered under the relativized order; such sets C are called **chains** (or $\preceq$**-chains** if it is necessary to be explicit about what ordering is being used). A useful property of chains, even if it's a bit awkward to state, is the

Proposition: Let $(X, \preceq)$ be a partially ordered set. Suppose $\mathcal{C}$ is a family of nonempty subsets of X, such that

(1) Every $C \in \mathcal{C}$ is a $\preceq$-chain;

(2) $\mathcal{C}$ itself is a chain in $\mathbf{2}^{\mathbf{2}^X}, \subseteq)$; *i.e.*, given two (distinct) elements C_1 and C_2 of $\mathcal{C}$, either $C_1 \subset C_2$ or $C_2 \supset C_1$.

Then the union $\bigcup\{C : C \in \mathcal{C}\}$ is a chain in $(X, \preceq)$. Loosely but concisely: the union of a chain of chains is a chain.

Proof. Let x and y be two elements of the union. Then there exist members C_x and C_y of $\mathcal{C}$ with $x \in C_x$ and $y \in C_y$. One of C_x and C_y contains the other: suppose, say, that $C_x \subset C_y$. Then since $x \in C_y$ and since C_y is a chain, exactly one of the relations $x \prec y$, $x = y$, $y \prec x$ must hold, as desired.

{Remark: the proposition would be true under the weaker hypothesis that $\mathcal{C}$ is **upward-directed under** $\subseteq$, *i.e.*, that given any two members C_1 and C_2 of $\mathcal{C}$ there exists some $C_3 \in \mathcal{C}$ that contains both C_1 and C_2. The reader can easily supply the simple modifications in the proof.}

A **well-ordered** set $(X, \preceq)$ (see, *e.g.*, Royden pp. 27–28 and ex. 31b) is a *linearly* ordered set with the property that every nonempty subset $S \subseteq X$ contains a $\preceq$-smallest element, *i.e.*, an element s_1 such that $s_1 \preceq s$ holds for every $s \in S$. (This element is clearly unique if it exists, since $\preceq$ is assumed antisymmetric.) $(\mathbb{N}, \leq)$ is the most famous example. Not every linearly ordered set is well-ordered, as $(\mathbb{Q}, \leq)$ shows. Moreover, the union of a chain of subsets of a given partially ordered set $(X, \preceq)$, each of which is well-ordered in the relativized order, can fail to be well-ordered. For example, consider the sets

$$Q_q = \left\{ \frac{p}{q!} \in \mathbb{Q} : p \in \mathbb{N} \right\} \quad (! = \text{the usual factorial}) \;\; \text{for } q \in \mathbb{N}.$$

Evidently $Q_1 \subset Q_2 \subset \cdots$ and each Q_q is order-isomorphic (in the obvious sense) to $\mathbb{N}$, which is well-ordered; however, the union of the Q_q's is the set of all positive rationals, which has many subsets having no smallest element—for instance, the whole set.

To find a way to unite well-ordered sets into bigger well-ordered sets, as well as for other purposes, it will help to think about **segments** (also called **intervals**) in general partially ordered sets $(X, \preceq)$ (cf. Royden, p. 27, ex. 33. The definitions are the same ones we're all used to for the real numbers:

$$\begin{aligned}
(x, y) &= \{z \in X : x \preceq z \preceq y\} \;\; \text{(open)} \\
[x, y) &= \{z \in X : x \preceq z \preceq y)\} \;\; \text{(closed on the left, open on the right)} \\
(x, y] &= \{z \in X : x \preceq z \preceq y\} \;\; \text{(open on the left, closed on the right)} \\
[x, y] &= \{z \in X : x \preceq z \preceq y\} \;\; \text{(closed).}
\end{aligned}$$

If X has a smallest element x_0 then $[x_0, x)$ and $[x_0, x]$ are called **open** and **closed initial segments** respectively. Note that even in $\mathbb{N}$ things are not exactly what one is used to in $\mathbb{R}$: there is no difference between the open initial segment $[1, n+1)$ and the closed initial segment $[1, n]$. Moreover, in a general partially ordered set two intervals $[x, y)$ and $[x, z)$ may be equal as sets even[4] though $y \neq z$; so some care in thinking is necessary. In any partially ordered $(X, \preceq)$ a set $S \subseteq X$ is said to be **order-convex** if S contains with every pair of "comparable" elements—x, $y \in S$ with $x \preceq y$—the closed order interval $[x, y]$. Obviously any order interval is order-convex. In the case of well-ordered sets, it is useful to know the following fact, whose proof the reader can easily supply.

Proposition: Let $(Y, \preceq)$ be a well-ordered set and let y_0 be its smallest element. If Y has a largest element y_1, then $Y = [y_0, y_1]$ is the only closed initial segment which is not open: in all other cases, every

[4] Can this happen in a linearly ordered set?

closed initial segment $[y_0, y]$ equals the open initial segment $[y_0, y^*)$, where y^* is the (uniquely determined) smallest element that is $\preceq$-larger than y. (Yes, the case in which $y_0 = y_1$ is covered.)

Let $(X, \preceq)$ be a partially ordered set. We can put a "stronger" order than simple set-inclusion on the subsets Y of X that are well-ordered in the relativized order, as follows: for two subsets Y_1 and Y_2 of X that are well-ordered in the relativized order, let us write $Y_1 \sqsubseteq Y_2$ if Y_1 is *equal to an initial segment of* Y_2. Evidently $Y_1 \sqsubseteq Y_2$ implies $Y_1 \subseteq Y_2$, so this relation is reflexive and anti-symmetric, and transitivity is easy to check.

Proposition: Let $(X, \preceq)$ be a partially ordered set, let $x_0 \in X$, and let $\mathcal{W}$ be a family of subsets of X such that

(1) Each $Y \in \mathcal{W}$ is well-ordered in the relativized order, with x_0 as its smallest element;

(2) $\mathcal{W}$ is a chain under the ordering $\sqsubseteq$.

Then $W = \bigcup\{Y \in \mathcal{W}\}$ is well-ordered in the relativized order, with smallest element x_0.

Proof. By what we already know, W is a chain under $\preceq$, and it is obvious that x_0 is smaller than any other element of W. To see that W is well-ordered, let S be any proper subset of W. Then there is some $Y \in \mathcal{W}$ for which $Y \cap S \neq \emptyset$; let s be the smallest element of $Y \cap S$ in Y. Then s must be the smallest element of S in W, for if there were some $t \in S$ with $t \prec s$ one would still have $t \in Z$ for some $Z \in \mathcal{W}$. One of Y and Z is an initial segment of the other. If $Y \sqsubseteq Z$ then $t \in Y$ also, and s was incorrectly chosen. If $Z \sqsubseteq Y$ then the choice of s was even more blatantly incorrect! (or else that of t was), so s is indeed the smallest element of S in W.

3.B. *The Hausdorff Maximal Principle and Zorn's Lemma.*

These two statements—each of which is logically equivalent to the choice axiom—look almost alike and are easily shown to be equivalent to each other. It is largely a matter of taste which one to use when we make nontrivial use of the choice axiom, but we should examine them both.

Hausdorff Maximal Principle: Let $\preceq$ be a partial ordering on a set X. Then there is a maximal linearly ordered subset S of X, *i.e.*, a subset $S \subseteq X$ which is linearly ordered by $\preceq$ and has the property that if $S \subseteq T \subseteq X$ and T is linearly ordered by $\preceq$, then $S = T$.[5]

To state Zorn's lemma we need one additional definition that we all really know already:

Definition: Let $(X, \preceq)$ be a partially ordered set and $\emptyset \neq A \subseteq X$. An **upper bound for** (or **of**) A (**relative to** $\preceq$) is an element $a \in X$ such that $x \preceq a$ holds for every $x \in A$. **Lower bound** is defined dually.

Sometimes one makes the following additional definition: a partially ordered set $(X, \preceq)$ is **inductively ordered** (or **inductive**) if every chain in X has an upper bound.

Zorn's Lemma: If $(X, \preceq)$ is a partially ordered set in which every chain has an upper bound, then X has a maximal element, *i.e.*, there exists an element $s \in X$ for which $s \preceq x \Rightarrow x = s$.

Hausdorff $\Rightarrow$ Zorn: Given an inductively ordered set $(X, \preceq)$, let S be a maximal linearly ordered subset of X. Because $(X, \preceq)$ is inductive, there exists some upper bound s of S. We claim s is a maximal element with respect to $\preceq$, for if $x \in X$ satisfies $s \preceq x$, then for every $y \in S$ we have $y \preceq x$; if $x \neq s$, then $x \in S$ is impossible because it would imply $x \preceq s \preceq x$ and so $x = s$. But it would then follow that $T = S \cup \{x\}$ was a set linearly ordered under $\preceq$ and containing S properly, contrary to the maximality of S. The element s must therefore be maximal.

Zorn $\Rightarrow$ Hausdorff: Order the chains in $(X, \preceq)$ by set-inclusion. As we saw above, the union of a totally ordered family of chains is itself a chain, and that union is clearly an upper bound (under set-inclusion) for the given totally ordered family. So the set of chains in X is inductively ordered by set-inclusion. Zorn then gives us a chain which is maximal under set-inclusion, exactly what Hausdorff requires.

[5] Cf. Royden, p. 25.

3.C. *Using Zorn's Lemma.*

Since we want to show that Zorn's Lemma and the choice axiom are equivalent, the first thing to do might well be to prove the latter from the former.

{Remark: The axiom of choice is variously stated. One formulation goes:

Let $\{X_\iota\}_{\iota\in I}$ be an indexed family of nonempty sets (indexed by a [nonempty] index set I). Then there exists a function $c : I \to \bigcup_{\iota\in I} X_\iota$ that assigns to every $\iota \in I$ an element $c(\iota) \in X_\iota$.

This is a nice formulation because it asserts the **nonemptiness of Cartesian products**. In general, the **Cartesian product**—usually written as $\prod_{\iota\in I} X_\iota$—**of an indexed family** $\{X_\iota\}_{\iota\in I}$ of nonempty sets indexed by a [nonempty] index set I is the set of all functions $x : I \to \bigcup_{\iota\in I} X_\iota$ that assign to every $\iota \in I$ an element $x(\iota) \in X_\iota$. The functions are frequently written in the form $(\ldots, x_\iota, \ldots)$ to make them look like "-tuples"; but even an element (x_1, x_2, x_3) of a product of three sets X_1, X_2, X_3—which one might be able to define in various ways—can be construed as a function from the set $\{1, 2, 3\}$ to $X_1 \cup X_2 \cup X_3$. The problem with this definition, when I is infinite and the sets $\{X_\iota\}_{\iota\in I}$ quite arbitrary, is that one doesn't know that the Cartesian product has anything in it—and indeed, any element of the Cartesian product is a choice function.

Another—and an older—formulation of the choice axiom is

Let $\mathcal{F}$ be a pairwise disjoint family of nonempty sets (so $F \neq G \Rightarrow (F \cap G) = \emptyset$). Then there exists a set $C \subseteq \bigcup\{F : F \in \mathcal{F}\}$ with the property that for each $F \in \mathcal{F}$ the set $F \cap C$ contains exactly one point.

The reader should try to prove the equivalence of these formulations.}

Theorem: Zorn's Lemma implies the Axiom of Choice.

Proof. We think of the choice axiom in the first form given above. Let $\{X_\iota : \iota \in I\}$ be a family of nonempty sets indexed by a nonempty index set I. Let $X = \bigcup_{\iota\in I} X_\iota$ and consider the family of all subsets f of $I \times X$ satisfying the following two conditions:

(1) f is a function;

(2) For every $\iota \in \text{domain}(f)$ the value $f(\iota) \in X_\iota$.

Evidently there exist such functions: since each X_ι is nonempty, for each $\iota \in I$ there is a function of this kind with one-point domain. Order the family of all such functions by set-inclusion (as subsets of $I \times X$). It is obvious that the union of a chain of functions is a function, so this family is inductively ordered under set-inclusion. Assuming Zorn's Lemma, there is a maximal such function, call it c. I claim that its domain must be the entire index set I. If that were false, then we could take any $\kappa \in I$ not belonging to domain(c), take any $x \in X_\kappa$, and consider $c \cup \{(\kappa, x)\} \subseteq I \times X$. This relation is still a function, because we did not "redefine the value at κ"—c was not defined at κ. The new function still sends each point ι of its domain to a point of the corresponding X_ι. But since the new function is clearly stricly larger than c with respect to set-inclusion in $I \times X$, its existence would violate the maximality of c. Hence domain(c) $= I$, and c is the choice function we sought.

A similar use of Choice in the form of Zorn shows that given any two nonempty sets, one has at least as many elements as the other.

Proposition: Let A and B be two nonempty sets. Then either there is a 1-1 correspondence between A and a subset of B, or vice versa, or both.

Proof. Consider the family of nonempty subsets $f \subseteq A \times B$ that are "functions in both directions," *i.e.*, both f and $f^{-1} = \{(y, x) \in B \times A : (x, y) \in f\}$ are functions. (These are 1-1 correspondences between their domains and ranges.) It is obvious that the union of a chain (under inclusion) of such relations is again a relation of the same kind. By Zorn's Lemma, there is a maximal such relation F. If its domain is A or its range is B, then F is a 1-1 correspondence of A with a subset of B (namely, its range) or F^{-1} is a

1-1 correspondence of B with a subset of A respectively. If neither of these possibilities were to hold, then there would exist $a \in A$ and $b \in B$ which were outside the domain and range of F respectively. But then $F \cup \{(a, b)\}$ would be a 1-1 correspondence that properly included F, contradicting the maximality of F; therefore this cannot occur and the proposition is established.

{Long Remark: Although we shall not give a proof of it, it is important for everyone to know the truth of the

[Schröder-Bernstein] **Theorem**: If there exist a 1-1 correspondence between A and a subset of B and a 1-1 correspondence between B and a subset of A, then there can be constructed from them a 1-1 correspondence between A and B.

The phrase "there can be constructed" points out the fact that the construction does not involve the axiom of choice; if the two 1-1 correspondences are given, then the 1-1 correspondence between A and B can be explicity written in terms of the given correspondences. For proof see, *e.g.*, G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, p. 362.

It follows from the two theorems above that if A and B are two nonempty sets, then either A is in 1-1 correspondence with a subset of B but not vice versa, B is in 1-1 correspondence with a subset of A but not vice versa, or A and B are in 1-1 correspondence. In these respective cases we say that A has **smaller** or **larger cardinality** than B, or that their **cardinalities are equal**. This looks like a "linear ordering of cardinal numbers," but there are logical difficulties involved in saying what a cardinal number is that we very much want to avoid examining. For more information in an unspecialized context see, *e.g.*, J. L. Kelley, *General Topology*, p. 274 ff. Kelley's appendix on set theory is in general a good place to look for a straightforward approach to the best-known difficulties in classical set theory.

The following lemma is an exercise in mathematical induction for the reader. While it is trivial and obvious, it still requires a formal proof if we are all to believe it.

Proposition: There exists a 1-1 correspondence between the initial segment $[1, k]$ of $\mathbb{N}$ and a proper subset of the initial segment $[1, n]$ if and only if $k \leq n$, vice versa if and only if $n \leq k$, and a 1-1 correspondence between the two segments if and only if $n = k$. Similarly, there is a function mapping $[1, k]$ *onto* $[1, n]$ if and only if $k \geq n$, and such a function is not 1-1 if $k > n$ (that is the **pigeonhole principle**: if I have more objects than boxes and put every object in a box, then some box must contain more than one object).

A straightforward modification of the proof of the last proposition we proved above will establish

Proposition: If A is a nonempty set then exactly one of the following is true: A is in 1-1 correspondence with an initial segment of the natural numbers $\mathbb{N}$, or A is in 1-1 correspondence with $\mathbb{N}$, or A has a proper subset in 1-1 correspondence with $\mathbb{N}$ but A itself cannot be put in 1-1 correspondence with $\mathbb{N}$.

In these three cases we say respectively that A is **finite with n elements** (where A was in 1-1 correspondence with the initial segment $[1, n]$ of $\mathbb{N}$), A is **countably infinite**, or A is **uncountable** (or **uncountably infinite**).}

Getting back to Zorn's Lemma and things equivalent to the Choice Axiom, one of the equivalent statements is the

Well-Ordering Theorem: Given any nonempty set A there is an ordering $\preceq \subseteq A \times A$ for which $(A, \preceq)$ is a well-ordered set.

The fact that such uncountable sets as $\mathbb{R}$ are subject to well-ordering makes this theorem less than intuitively obvious. Nonetheless, it follows from Zorn's Lemma, and we shall now indicate the proof.

Proof that Zorn's Lemma ⇒ Well-Ordering Theorem.

Pick and fix a point $x_0 \in X$, and consider the family of relations $R \subseteq X \times X$ such that

(1) domain(R) = image(R);

(2) R is a total ordering of its domain (= image) which well-orders that set;

(3) the R-first element of the set is x_0.

Such relations exist, since the set $\{(x_0, x_0)\}$ has the desired properties (though its domain isn't very big). We order these relations (as subsets of $X \times X$) as follows: R_1 with domain A_1 is smaller than R_2 with domain A_2, written $R_1 \sqsubseteq R_2$, if and only if R_2 restricted to A_1 equals R_1 and A_1 is an initial segment of A_2 in the order R_2. By imitating the proof of the proposition on p. 8 above, it is easy to see that given a set of relations that is a chain under the ordering $\sqsubseteq$, the union of those relations will have properties (1)–(3) above: the only at-all-difficult thing to show is that the union is a well-ordering of its domain.[6] By Zorn's Lemma there is a maximal such ordering, call it $\preceq$. If its domain = range is X, we're finished. Otherwise, let $x_1 \in X$ not belong to the domain = range of $\preceq$, and consider $(\preceq \cup \{(x_1, x_1)\}) \cup (\text{domain}(\preceq) \times \{x_1\})$; that is, extend $\preceq$ by making x_1 larger than every element of the domain of $\preceq$. It is easy to see that this relation has properties (1)–(3) and is larger than $\preceq$, contrary to maximality, so $\preceq$ well-orders the entire set X as desired.

To show that this material is not of interest only to analysts and set-theorists, the algebraically-minded reader might use Zorn's Lemma to show that if X is a vector space over a field $\mathbb{K}$ and linear dependence is defined finitistically—that is, a set is linearly independent if no finite subset admits a non-zero linear dependence relation—then Zorn's lemma implies that any given linearly independent set is contained in a maximal (under set-inclusion) linearly independent subset. Maximality implies that any given nonzero vector $x \in X$ is a (finite) linear combination of elements of such a maximal linearly independent subset. One has then proved

Proposition: Every linearly independent subset of a vector space over a field is contained in a basis.

Proving this directly from the choice axiom is not all that pleasant (unlike selecting a set of coset representatives). Note that since the reals $\mathbb{R}$ form a vector space over the rationals $\mathbb{Q}$, this proposition is far from intuitively clear. A similar argument will show that any set of elements of a field algebraically independent over a subfield can be enlarged to a transcendence basis over that subfield.

3.D. *Proving Zorn's Lemma from the Axiom of Choice.*

Something relatively unpleasant always has to be done to get from the intuitively-appealing Choice Axiom to the logically equivalent statements that are more convenient technically. Classically one proved the Well-Ordering Theorem from the Choice Axiom. The following proof of Zorn's Lemma from the Choice Axiom, while it contains the same ingredients as the classical proof of the Well-Ordering Theorem, is the shortest and most understandable argument known to me that establishes this counter-intuitive implication. Except for minor technical changes, the proof is one published in the April 1991 Monthly [Jonathan Lewin, *A simple proof of Zorn's lemma*, Amer. Math. Monthly 98 (1991), 353–354]. The simplification seems to come from the fact that there is a preëxisting order on the set X.

Theorem: The Axiom of Choice $\Rightarrow$ Zorn's Lemma: if $(X, \preceq)$ is an inductively ordered set and $x_0 \in X$, then there is a maximal element $a \in X$ with $x_0 \preceq a$.

Proof. Assume that some instance of the theorem is false, so we have an inductively ordered $(X, \preceq)$ and an x_0 for which the statement of the theorem fails. If C is a $\preceq$-chain in such an X, then it has an upper bound c (which may or may not belong to C) and because c is not maximal there exists $b \succ c$, which is an upper bound for C that is strictly larger than every element of C. If we define for each $\preceq$-chain $C \subseteq X$ the set

$$B(C) = \{b \in X : \text{For all } x \in C,\ x \prec b\}$$

of all "strict upper bounds of C," then each $B(C)$ is nonempty. Using the set of chains in X as an index set, we can then invoke the Axiom of Choice to pick for each chain $C \subseteq X$ a strict upper bound of C. Let f denote the choice function, so $f(C) \in B(C)$ is the chosen strict upper bound.

Consider the family of all subsets $C \subseteq X$ that have the following properties:

(1) $x_0 \in C$ and is its $\preceq$-smallest element;

[6] Detailed checking is left to the reader.

(2) $\preceq$ well-orders C (so in particular C is a chain and x_0 is its $\preceq$-initial element);

(3) For each open initial segment $[x_0, x)$ of C, we have $x = f([x_0, x))$. That is, among the (possibly) many ways that the subset $[x_0, x)$ of C might have the form $[x_0, y)$ for some $y \in X$, the application of the choice function f to that set will produce the $x \in C$ that defines it.

It is easy to see that such chains exist: $\{x_0\}$ is one, and $\{x_0, f(\{x_0\})\}$ is another, etc. Following Lewin, we shall call such chains **conforming**[7] chains.

I claim that if A and B are conforming chains then either one is an initial segment of the other or they are equal. To see this, let $\mathbb{S}$ denote the family of all sets $S \subseteq A \cap B$ which are initial segments—closed or open—both of A and of B. This family is nonempty since $\{x_0\} \in \mathbb{S}$. As a union of initial segments of A, $U = \bigcup \mathbb{S}$ is either equal to A or is a (n open) initial segment of A; similarly, U is either equal to B or is an initial segment of B. In the first case, $A \subseteq B$ and if equality does not hold then A is a(n open) initial segment of B; in the second case the rôles of A and B are reversed. The remaining apparent possiblity is that U is a proper, hence open, initial segment of each of the two conforming chains A and B. Looking at A, we have $U = [x_0, a)$ where $a = f(U) \in A$; similarly $U = [x_0, b)$ where $b = f(U) \in B$. But then $a = f(U) = b$ belongs to $A \cap B$ and so $U \supseteq [x_0, a] = [x_0, b]$: this puts $a = b \in U$ contradicting the choice of $f(U)$ as strictly $\succ$-larger than any element of U. So this apparent possibility cannot in fact occur.

Now, just as in the argument we saw on p. 8 above, the union V of all conforming chains in X is well-ordered by $\preceq$ and has x_0 as smallest element. Where is $v = f(V)$? It cannot belong to V because it is strictly larger than every element of V. But then $V \cup \{v\}$ is well-ordered: v is the smallest element of $\{v\}$, and any other proper subset T of $V \cup \{v\}$ has elements in common with V; the smallest element of $T \cap V$ is evidently smallest in T. The only open initial segment of $V \cup \{v\}$ that is not also an initial segment of one of the conforming chains that were united to form V is V itself, and $V = [x_0, f(V))$ in $V \cup \{v\}$. But now $V \cup \{v\}$ is seen to be conforming, and thus $(V \cup \{v\}) \subseteq V$ and $v = f(V) \in V$, which is absurd since it leads to $v \prec v$, this element being a strict upper bound of V. The contradiction shows that the Axiom of Choice implies Zorn's Lemma.

3.E. *Cleaning Up.*

We now know that of the statements

(1) Axiom of Choice (or existence of a Choice Function)

(2) Non-emptiness of Cartesian products

(3) Hausdorff Maximal Principle

(4) Zorn's Lemma

(5) Well-Ordering Theorem

(6) Tychonov Product Theorem of general topology

(2) $\Leftrightarrow$ (1) $\Rightarrow$ (4) $\Leftrightarrow$ (3) and (4) $\Rightarrow$ (5), the equivalences being rather easy to prove. Proving (1) assuming (5) is quite easy: if $\{X_\iota : \iota \in I\}$ is an indexed family of nonempty sets, then let $\preceq$ well-order their union. Each $X\iota$ has a unique $\preceq$-smallest element, and so we can define a choice function f on I by simply setting $f(\iota)$ equal to the $\preceq$-smallest element of X (this can all be strictly formalized but the formalization would only confuse matters at this stage). To discuss the Tychonov theorem at this point would take us too far afield; this is a real-variables course and we haven't even constructed Lebesgue measure yet!

3.F. *The Peano Axioms for* $\mathbb{N}$.

We have been using $\mathbb{N}$ only as an ordered set but haven't really been to specific about what it might be. Here, for reference, is a copy of (a version of) the Peano axioms, as found in an ancient algebra textbook [Richard E. Johnson, *A First Course in Abstract Algebra*, Prentice-Hall, 1953, p. 14 ff.]:

[7] Presumably, conforming to the choice of the next element.

(1) $\mathbb{N}$ is a set containing an element 1, and $\mathbb{N} \setminus \{1\} \neq \emptyset$;

(2) There is a 1-1 correspondence $^* : \mathbb{N} \mapsto \mathbb{N} \setminus \{1\}$; the image of $n \in \mathbb{N}$ under * is written n^* and called "the successor of n;"

(3) If $S \subseteq \mathbb{N}$ has the properties

   (a) $1 \in S$

   (b) For each $n \in \mathbb{N} : n \in S \Rightarrow n^* \in S$,

   then $S = \mathbb{N}$.

The usual operations and order on $\mathbb{N}$ can be built up by construction from these axioms, and in particular $n+1$ turns out to be n^*. Any two pairs $(\mathbb{N},^*)$ both satisfying these axioms can be put in 1-1 correspondence with 1 corresponding to 1 and * preserved (in the obvious sense), and this can be done in only one way. It is convenient for this course to *take as given* a complete ordered field $\mathbb{R}$, and we have given as an exercise a way to find a subset $\mathbb{N} \subset \mathbb{R}$ that satisfies the Peano axioms, so we can have (copies of) $\mathbb{N}$ (and $\mathbb{Q}$) as subsets of $\mathbb{R}$ and not worry about how $\mathbb{R}$ was constructed.

4. *Compactness and Ultrafilters.*

Talking about lim-inf and lim-sup for bounded filters (filter bases) in $\mathbb{R}$, and the convergence of monotone nets, has brought us around to considering questions of compactness, since the last paragraph of §1 above looks like the Bolzano-Weierstraß theorem that every bounded sequence in $\mathbb{R}$ has a convergent subsequence and the last paragraph of §2 reminds us of the same thing for nets. Now when one talks about convergence with filters, one has the moral equivalent of nuclear weapons to bring to bear on the "subsequence" question:

Definition: A filter $\mathfrak{U}$ is an **ultrafilter** if it is not properly contained in any other filter, *i.e.*, if every filter finer than $\mathfrak{U}$ equals $\mathfrak{U}$.

The principal filters at points of X are ultrafilters. However, in some sense the interesting filters are the ones that are not principal:

Proposition: Every filter $\mathfrak{G}$ is contained in some ultrafilter.

Proof. The class of all filters is ordered by set-inclusion in $\mathbf{2}^{\mathbf{2}^X}$. It is trivial to verify that the union of an inclusion-chain of filters is a filter (for any two sets A, B in the union, the question of whether $A \cap B$ is empty is settled in some filter belonging to the chain). This says that the class of all filters containing $\mathfrak{G}$ is inductively ordered. By Zorn's lemma, that class has a maximal element, which is then an ultrafilter $\mathfrak{U}$ containing $\mathfrak{G}$.

As an example of a nonprincipal ultrafilter, consider the filter of tails of $\mathbb{N}$ ($\mathbb{N}$, or more accurately $\mathrm{id}_{\mathbb{N}}$, is itself a sequence). This is contained in an ultrafilter but such an ultrafilter cannot be principal, since for every $n \in \mathbb{N}$ there is a tail of $\mathbb{N}$, namely $\{k \in \mathbb{N} : k \geq n+1\}$, that excludes it. The mind boggles, but ultrafilters of this kind can be very useful. (You might think of them as being points of $\mathbb{N}$ that nature forgot to include.) The set of all ultrafilters in $\mathbb{N}$ can be given a compact Hausdorff topology and is called the **Čech** or **beta** compactification of $\mathbb{N}$; fortunately, $\beta\mathbb{N}$ is not part of this course.

Note that if x adheres to an ultrafilter $\mathfrak{U}$, then $\lim \mathfrak{U} = x$. The reason is that there is some filter finer than $\mathfrak{U}$ that converges to x—but because is maximal, that "finer" filter must be $\mathfrak{U}$ itself.

Here's the ultimate Bolzano-Weierstraß-type construction:

Theorem: The metric space (X, d) is compact if and only if every ultrafilter in X converges.[8]

Proof. We use the "dual" characterization of compactness, according to which X is compact if and only if for every family $\mathfrak{F}$ of closed subsets of X with the finite intersection property (*i.e.*, intersections of finite subfamilies of are nonempty), the intersection of all the elements of $\mathfrak{F}$ is nonempty.

[8] This assertion is valid in any topological space.

Suppose every ultrafilter in X converges. Let $\mathfrak{F}$ with the finite intersection property be given. The finite intersection property is equivalent to: if $\mathfrak{B}$ is the family of all intersections of finite subfamilies of $\mathfrak{F}$, then $\mathfrak{B}$ is a filter base. Let $\mathfrak{U}$ be an ultrafilter containing (the filter based on) $\mathfrak{B}$; $\mathfrak{U}$ exists by Zorn's lemma. If $\lim \mathfrak{U} = x$ then x certainly adheres to $\mathfrak{B}$, so

$$x \in \bigcap \{\overline{U} : U \in \mathfrak{U}\} \subseteq \bigcap \{F : F \in \mathfrak{F}\} .$$

We have exhibited a point in the intersection of all the elements of $\mathfrak{F}$.

Suppose X is compact. For any filter, the family of closures of its sets has the finite intersection property because the filter itself does. It follows that if $\mathfrak{U}$ is an ultrafilter then the intersection of the closures of its sets is nonempty. But to say $x \in \bigcap\{\overline{U} : U \in \mathfrak{U}\}$ is to say that x adheres to $\mathfrak{U}$ and as we saw above an ultrafilter converges to any point to which it adheres; thus $\lim \mathfrak{U} = x$.

By the way, this must be the most absurd way possible to prove that closed intervals in $\mathbb{R}$ are compact, but: if $\mathfrak{U}$ is an ultrafilter in $[a, b] \subseteq \mathbb{R}$, then it has adherent points (namely, its lim sup and lim inf), therefore it converges; so these intervals must be compact, a well-known (Heine-Borel theorem) result.

Notice that no process corresponding to selecting a subsequence had to be done: the maximality of ultrafilters means that all the choices that had to be made have already been made. Hence the power—and the slightly eerie feeling—of ultrafilter approaches to compactness. {For set-theory fans: I should mention that in terms of the Boolean ring $\mathbf{2}^X$, the statement that every filter is contained in an ultrafilter is equivalent to: every ideal in the Boolean ring $\mathbf{2}^X$ is contained in a *prime* ideal (not *a priori* maximal). The reason for this slightly weaker requirement is that an integral domain in which every element satisfies $x^2 = x$ is necessarily $\mathbb{Z}/(2)$, a field, so primes are automatically maximal. Readers who are not interested in this subject can forget that I said anything.}

There is a way to recognize a filter as being "ultra" that is handy to know; it is the fact that maximal ideals in a (commutative) ring are characterized by the fact that their quotient rings are fields, dualized into a set-theoretic analogue.

Proposition: A filter $\mathfrak{F}$ is an ultrafilter if and only if: whenever a finite union of subsets of X belongs to $\mathfrak{F}$, one of the uniands belongs to $\mathfrak{F}$. More generally, the following conditions are equivalent for a filter $\mathfrak{F}$:

(1) $\mathfrak{F}$ is an ultrafilter;

(2) For each $A \subseteq X$, one of A or $X \setminus A$ belongs to $\mathfrak{F}$;

(3) If A, $B \subseteq X$ are nonempty and $A \cup B \in \mathfrak{F}$, then $A \in \mathfrak{F}$ or $B \in \mathfrak{F}$;

(4) If $\{A_i : 1 \leq i \leq k\}$ are nonempty subsets of X whose union belongs to $\mathfrak{F}$, then (at least) one of the sets A_i belongs to $\mathfrak{F}$.

Proof. (1) $\Rightarrow$ (2): Consider $\{A \cap F : F \in \mathfrak{F}\}$. If the empty set is one of these, then $A \subseteq (X \setminus F)$ so $F \subseteq X \setminus A$ and thus $X \setminus A \in \mathfrak{F}$ (a superset of an element of a filter belongs to the filter). Otherwise, those intersections $A \cap F$ form a filter base (verification easy) for a filter which is finer than $\mathfrak{F}$. If $\mathfrak{F}$ is an ultrafilter this new filter cannot be strictly finer, so each $A \cap F$ belongs to $\mathfrak{F}$; since $A = A \cap X$ is such a set, $A \in \mathfrak{F}$. (2) $\Rightarrow$ (3): Suppose neither A nor B belongs to $\mathfrak{F}$, but that their union does. (2) implies that $(X \setminus A)$ and $(X \setminus B)$ both belong to $\mathfrak{F}$, therefore so does their intersection $X \setminus (A \cup B)$. But then $\mathfrak{F}$ contains both $A \cup B$ and its complement, yet their intersection is empty! (3) $\Rightarrow$ (4): Write

$$\bigcup_{i=1}^{k} A_i = \left(\bigcup_{i=1}^{k-1} A_i \right) \cup A_k$$

and use induction on $k - 1$ ((3) is the case $k - 1 = 1$ and also the induction step). (4) $\Rightarrow$ (1): If $\mathfrak{G}$ were a filter strictly finer than $\mathfrak{F}$ then there would be some $G \in \mathfrak{G}$ that hadn't belonged to $\mathfrak{F}$. By (4), $X \setminus G$ would have belonged to $\mathfrak{G}$, and thus $\mathfrak{G}$ would contain both the sets G and $X \setminus G$, again giving an empty intersection and a contradiction to $\mathfrak{G}$'s being a filter.