

# Ultrafilters, ultraproducts, and the compactness theorem

The goal of this note is to present a proof of the following fundamental result. A theory  $T$  is said to be *satisfiable* iff there is a model of  $T$ .

**Theorem 1 (Compactness)** *Let  $T$  be a first order theory. Suppose that any finite sub-theory  $T_0 \subseteq T$  is satisfiable. Then  $T$  itself is satisfiable.*

As indicated on the set of notes on the [completeness](#) theorem, compactness is an immediate consequence of completeness. Here I want to explain a purely semantic proof, that does not rely on the notion of proof.

The argument I present uses the notion of *ultraproducts*. Although their origin is in model theory, ultraproducts have become an essential tool in modern set theory, so it seems a good idea to present them here. We will require the axiom of choice, in the form of Zorn's lemma.

The notion of ultraproduct is a bit difficult to absorb the first time one encounters it. I recommend working out through some examples in order to understand it well. Here I confine myself to the minimum necessary to make sense of the argument.

## 1 Ultrafilters

**Definition 2** *Let  $X$  be a nonempty set. A collection  $\mathcal{F} \subseteq \mathcal{P}(X)$  of subsets of  $X$  is said to have the finite intersection property (fip) iff the intersection of any (nonzero but) finitely many of the sets in  $\mathcal{F}$  is nonempty.*

Note in particular that if  $\mathcal{F} \subseteq \mathcal{P}(X)$  has the fip, then the sets  $Y \in \mathcal{F}$  are all nonempty, Here are some typical examples of such collections. First, we could fix a nonempty subset  $Y \subseteq X$  and let

$$\mathcal{F} = \{Z \subseteq X \mid Y \subseteq Z\}.$$

Or we could consider an infinite set  $X$  and let

$$\mathcal{F} = \{Y \subseteq X \mid X \setminus Y \text{ is finite}\},$$

this is called the *Fréchet* filter on  $X$ .

For a more interesting example, consider the interval  $[0, 1]$  and let

$$\mathcal{F} = \{Y \subseteq [0, 1] \mid \mu(Y) = 1\},$$

where  $\mu$  denotes Lebesgue measure.

All these examples share the following additional properties: First, if  $A \in \mathcal{F}$  and  $A \subseteq B \subseteq X$ , then  $B \in \mathcal{F}$ . Second, not only it is the case that  $A \cap B \neq \emptyset$  whenever  $A, B \in \mathcal{F}$  but, in fact,  $A \cap B \in \mathcal{F}$  as well. Families like this are called *filters* and will be essential in what follows.

**Definition 3** *Given a nonempty set  $X$ , a filter  $\mathcal{F} \subseteq \mathcal{P}(X)$  is a collection of subsets of  $X$  with the following properties:*

1.  $\mathcal{F} \neq \emptyset$ .
2.  $\emptyset \notin \mathcal{F}$ .
3. Whenever  $A, B \in \mathcal{F}$ , then  $A \cap B \in \mathcal{F}$ .
4. Whenever  $A \in \mathcal{F}$  and  $A \subseteq B \subseteq X$ , then  $B \in \mathcal{F}$ .

Generalizing from the last example above, a typical example of a filter is obtained by considering a complete measure space  $(X, \mathcal{M}, \mu)$ , and letting  $\mathcal{F}$  be the collection of subsets of  $X$  of  $\mu$ -measure 1. The completeness of the measure is needed to ensure that property (4) holds. This is an important example, and a working understanding of filters is obtained by thinking of its elements as being “large” in some sense. A similar example is obtained by considering a, say, complete, separable metric space  $X$ , and letting  $\mathcal{F}$  consist of those subsets of  $X$  whose complement is meager (first-category).

**Lemma 4** *Suppose that  $\mathcal{F} \subseteq \mathcal{P}(X)$  is nonempty and has the fip. Write  $\mathcal{F}'$  for the collection of (nonempty) finite intersections of sets in  $\mathcal{F}$ . Let*

$$\mathcal{G} = \{Y \subseteq X \mid \exists Z \in \mathcal{F}' (Z \subseteq Y)\}.$$

*Then  $\mathcal{G}$  is a filter.*

One usually refers to the filter  $\mathcal{G}$  as *the filter generated by  $\mathcal{F}$* .

PROOF: All properties are trivially verified. (1) follows from  $\mathcal{F}$  being nonempty. (2) follows from the fact that  $\mathcal{F}$  has the fip. Given  $Y \in \mathcal{G}$ , say that  $Z \subseteq X$  is a *witness* iff  $Z \in \mathcal{F}'$  and  $Z \subseteq Y$ . Clearly, if  $A \in \mathcal{G}$  with witness  $Z$ , then  $Z$  is also a witness that  $B \in \mathcal{G}$ , for any  $B$  such that  $A \subseteq B \subseteq X$ . This gives us (4). To see (3), simply notice that  $\mathcal{F}'' = \mathcal{F}'$ , so if  $A, B \in \mathcal{G}$  with witnesses  $Z, W$ , then  $Z \cap W$  witnesses that  $A \cap B \in \mathcal{G}$ .  $\square$

Note that, conversely, any filter has the fip, and if  $\mathcal{F}$  is a filter, then  $\mathcal{G}$  as above is just  $\mathcal{F}$ .

**Definition 5** *An ultrafilter  $\mathcal{U}$  on a set  $X$  is a filter on  $X$  that is maximal under containment, i.e., if  $\mathcal{U} \subseteq \mathcal{F}$ , and  $\mathcal{F}$  is also a filter on  $X$ , then  $\mathcal{U} = \mathcal{F}$ .*

There is an equivalent definition of ultrafilters that is widely used in practice.

**Lemma 6** *A filter  $\mathcal{U}$  on a set  $X$  is an ultrafilter iff for any  $A \subseteq X$ , either  $A \in \mathcal{U}$  or  $X \setminus A \in \mathcal{U}$ .*

PROOF: If a filter  $\mathcal{F}$  is as given in the lemma, it must be maximal: For any  $A \subseteq X$ , either  $A$  is already in  $\mathcal{F}$ , or else  $A$  is disjoint from some element of  $\mathcal{F}$ , namely  $X \setminus A$ . Hence no larger subclass of  $\mathcal{P}(X)$  is a filter.

Conversely, given a filter  $\mathcal{U}$  suppose that neither  $A$  nor  $X \setminus A$  are in  $\mathcal{U}$ . We argue that  $\mathcal{U}$  is not maximal. For this, it suffices to check that at least one of  $\mathcal{U} \cup \{A\}$  and  $\mathcal{U} \cup \{X \setminus A\}$  has the fp. Suppose otherwise, so (since  $\mathcal{U}$  is closed under finite intersections) there are sets  $S, T \in \mathcal{U}$  such that

$$A \cap T = \emptyset$$

and

$$(X \setminus A) \cap S = \emptyset.$$

But then  $T \subseteq X \setminus A$ ,  $S \subseteq A$ , and  $T \cap S = \emptyset$ , contradicting that  $\mathcal{U}$  has the fp.  $\square$

A routine application of Zorn's lemma now shows that any filter is contained in an ultrafilter.

**Lemma 7** *Let  $\mathcal{F}$  be a filter on a set  $X$ . Then there is an ultrafilter on  $X$  extending  $\mathcal{F}$ .*

PROOF: Consider the collection of all filters on  $X$  that contain  $\mathcal{F}$ . This is obviously nonempty, since  $\mathcal{F}$  itself is one of them. Given a  $\subseteq$ -chain of such filters, its union clearly has the fp, and therefore generates a filter that contains all the members of the chain. This shows that Zorn's lemma applies, and there is therefore a maximal member of the collection. It is easy to see that this is indeed a maximal filter on  $X$ , and we are done.  $\square$

A typical example of ultrafilters are the *principal* ones, those ultrafilters  $\mathcal{U}$  such that  $\bigcap \mathcal{U} \neq \emptyset$ . It is easy to check that an ultrafilter  $\mathcal{U}$  on  $X$  is principal iff there is some  $a \in X$  such that

$$\mathcal{U} = \{Y \subseteq X \mid a \in Y\}.$$

Ultrafilters not of this form are called nonprincipal or *em free*. It is obvious that if  $X$  is finite, any ultrafilter on  $X$  is principal. On the other hand, for any infinite set  $X$  there are nonprincipal ultrafilters on  $X$ . For example, consider any ultrafilter extending the Fréchet filter.

## 2 Ultraproducts

At the moment, our interest on ultrafilters is purely utilitarian: They will allow us to define an “average” operation on structures in a given language (think of it as an abstract integration of such structures). The properties of this operation will easily imply the compactness theorem.

Suppose that  $\mathcal{U}$  is an ultrafilter on a set  $X$ . Let  $\mathcal{L}$  be a given first order language, and suppose that for all  $a \in X$  we are given an  $\mathcal{L}$ -structure  $\mathcal{M}_a = (M_a, \dots)$ . We want to define an  $\mathcal{L}$ -structure that somehow averages together all the  $\mathcal{M}_a$ . This will be the *ultraproduct*

$$\prod_a \mathcal{M}_a / \mathcal{U}$$

of the structures  $\mathcal{M}_a$  by  $\mathcal{U}$ .

The definition of the ultraproduct will take us some time.

First, we define the product.

**Definition 8** *The set theoretic product of a collection  $(X_a \mid a \in X)$  of nonempty sets is the set of all “choice functions” on this collection,*

$$\prod_a X_a = \{f : X \rightarrow \bigcup_a X_a \mid \forall a (f(a) \in X_a)\}.$$

The axiom of choice is equivalent to the statement that these products are never empty. Note that if  $X = \{a_1, \dots, a_n\}$  is finite, one can naturally identify  $\prod_a X_a$  and the Cartesian product  $X_{a_1} \times \dots \times X_{a_n}$ .

We can think of the “threads”  $f$  in the product as “variable elements” (thinking of  $X$  as time, we have that  $f$  varies with time, being the element  $f(a) \in X_a$  at time  $a$ ).

Suppose now that rather than just having nonempty sets  $X_a$ , we actually have structures  $\mathcal{M}_a$ . The basic idea is that the ultrafilter will allow us to identify “typical” properties the threads maintain as they change in time, and we will ensure that these properties hold in the “average” structure we are defining.

I proceed to make this precise:

First, we define the *universe* of the ultraproduct structure  $\prod_a \mathcal{M}_a / \mathcal{U}$  to be the quotient of  $\prod_a M_a$  by the equivalence relation

$$f =_{\mathcal{U}} g \text{ iff } \{a \in X \mid f(a) = g(a)\} \in \mathcal{U}.$$

This means that the elements of  $\prod_a M_a / \mathcal{U}$  are the equivalence classes of  $=_{\mathcal{U}}$ . What we are saying is that two threads are identified if they coincide almost everywhere.

This is easily seen to be an equivalence relation. To prove transitivity, note that the intersection of sets in  $\mathcal{U}$  is in  $\mathcal{U}$ .

Now we define the interpretation in the ultraproduct of relation, function, and constant symbols in  $\mathcal{L}$ .

Suppose that  $c$  is a constant symbol. Then for all  $a \in X$ , we have an interpretation  $c^{\mathcal{M}_a} \in M_a$ . Let  $\vec{c}$  be the choice function given by  $\vec{c}(a) = c^{\mathcal{M}_a}$  for all  $a \in X$ . Now define

$$c^{\prod_a \mathcal{M}_a / \mathcal{U}} = [\vec{c}]_{\mathcal{U}},$$

where  $[f]_{\mathcal{U}}$  denotes the  $=_{\mathcal{U}}$ -equivalence class of the choice function  $f$ .

Suppose that  $S$  is a  $j$ -ary relation symbol. Given any  $j$ -tuple of elements of the ultraproduct, say  $([f_1]_{\mathcal{U}}, \dots, [f_j]_{\mathcal{U}})$ , set

$$([f_1]_{\mathcal{U}}, \dots, [f_j]_{\mathcal{U}}) \in S^{\prod_a \mathcal{M}_a / \mathcal{U}}$$

iff

$$\{a \in X \mid (f_1(a), \dots, f_j(a)) \in S^{\mathcal{M}_a}\} \in \mathcal{U}.$$

One needs to verify that this is well defined, i.e., that if  $f_i =_{\mathcal{U}} f'_i$  for  $1 \leq i \leq j$ , then  $\{a \in X \mid (f_1(a), \dots, f_j(a)) \in S^{\mathcal{M}_a}\} \in \mathcal{U}$  iff  $\{a \in X \mid (f'_1(a), \dots, f'_j(a)) \in S^{\mathcal{M}_a}\} \in \mathcal{U}$ .

This is immediate from the fact that

$$\{a \in X \mid (f_1(a), \dots, f_j(a)) = (f'_1(a), \dots, f'_j(a))\} \in \mathcal{U},$$

which follows from the fact that  $\mathcal{U}$  is closed under finite intersections.

Finally, suppose that  $H$  is a  $k$ -ary function symbol. Given a  $k$ -tuple  $([f_1]_{\mathcal{U}}, \dots, [f_k]_{\mathcal{U}})$  of elements of the ultraproduct, **define**

$$H^{\prod_a \mathcal{M}_a / \mathcal{U}}([f_1]_{\mathcal{U}}, \dots, [f_k]_{\mathcal{U}})$$

as the equivalence class of the function that to  $a \in X$  assigns  $H^{\mathcal{M}_a}(f_1(a), \dots, f_k(a)) \in \mathcal{M}_a$ .

As before, one needs to verify that this is well-defined, but the argument is almost identical to the one for relations.

This completes the definition of the ultraproduct structure.

If all the structures  $\mathcal{M}_a$  are equal to some fixed structure  $\mathcal{M}$ , we talk of the *ultrapower* of  $\mathcal{M}$  by  $\mathcal{U}$ , and write  $\mathcal{M}^X / \mathcal{U}$ .

Of course, this notion has been designed with the goal in mind that the ultrapower structure must satisfy any property that holds of almost all the structures (in the sense that the set of  $a \in X$  such that  $\mathcal{M}_a$  has the property is in  $\mathcal{U}$ ). That this is indeed the case is the content of the following result:

**Lemma 9 (Łoś)** *Let  $\varphi(x_1, \dots, x_n)$  be any formula, and let  $f_1, \dots, f_n \in \prod_a \mathcal{M}_a$ . Then  $\prod_a \mathcal{M}_a / \mathcal{U} \models \varphi([f_1], \dots, [f_n])$  iff  $\{a \in X \mid \mathcal{M}_a \models \varphi(f_1(a), \dots, f_n(a))\} \in \mathcal{U}$ .*

**PROOF:** We argue by induction in the complexity of  $\varphi$ . To simplify, assume that  $\neg, \wedge, \exists$  are the only primitive connectives and quantifiers, and all others are abbreviations.

- If  $\varphi$  is atomic, the result holds by definition. (This requires in turn an induction on the complexity of terms.)
- Assume the result for  $\phi, \psi$ . Suppose that  $\varphi \equiv \phi \wedge \psi$ . Then the result also holds for  $\varphi$ , since  $\mathcal{U}$  is both upwards closed and closed under intersections.
- Assume the result holds for  $\phi$ . Suppose that  $\varphi \equiv \neg\phi$ . Then the result also holds for  $\varphi$ , since for any  $Y \subseteq X$ , either  $Y \in \mathcal{U}$  or else  $X \setminus Y \in \mathcal{U}$ .

- Finally, assume the result holds for  $\phi(x, x_1, \dots, x_n)$  and  $\varphi \equiv \exists x \phi(x, x_1, \dots, x_n)$ . Let  $f_1, \dots, f_n$  be elements of the product. Then we have, by definition, that

$$\prod_a \mathcal{M}_a/\mathcal{U} \models \varphi([f_1], \dots, [f_n])$$

iff there is some  $g$  in the product such that

$$\prod_a \mathcal{M}_a/\mathcal{U} \models \phi([g], [f_1], \dots, [f_n]).$$

This is equivalent, by assumption, to the statement that there is some  $g$  in the product such that

$$\{a \in X \mid \mathcal{M}_a \models \phi(g(a), f_1(a), \dots, f_n(a))\} \in \mathcal{U}.$$

But this last statement is equivalent to

$$\{a \in X \mid \mathcal{M}_a \models \varphi(f_1(a), \dots, f_n(a))\} \in \mathcal{U} :$$

In one direction, note that

$$\{a \in X \mid \mathcal{M}_a \models \phi(g(a), \vec{f}(a))\} \subseteq \{a \in X \mid \mathcal{M}_a \models \varphi(\vec{f}(a))\}$$

for any  $g$ . In the other, for each  $a \in X$  such that  $\mathcal{M}_a \models \varphi(\vec{f}(a))$ , pick a witness  $\iota_a \in M_a$ . Define a function  $g : X \rightarrow \bigcup_a M_a$  by  $g(a) = \iota_a$  if  $\mathcal{M}_a \models \varphi(\vec{f}(a))$ , and let  $g(a)$  be an arbitrary element of  $M_a$  otherwise. Then

$$\{a \in X \mid \mathcal{M}_a \models \varphi(\vec{f}(a))\} = \{a \in X \mid \mathcal{M}_a \models \phi(g(a), \vec{f}(a))\},$$

and we are done.

This completes the proof.  $\square$

**Corollary 10**  $\mathcal{M}^X/\mathcal{U}$  and  $\mathcal{M}$  satisfy the same first order theory.  $\square$

This proves, for example, the existence of nonstandard models of  $\text{Th}(\mathbb{N}, +, \times, 0, 1, <)$  : Simply consider the ultrapower  $\mathbb{N}^{\mathbb{N}}/\mathcal{U}$  for any free ultrafilter  $\mathcal{U}$  on  $\mathbb{N}$ , and note that an easy argument shows that  $\mathbb{N}^{\mathbb{N}}/\mathcal{U}$  contains infinite elements.

### 3 Compactness

We are finally ready to prove the compactness theorem. Suppose a theory  $T$  is given with the property that for any finite  $T_0 \subseteq T$  there is a models  $\mathcal{M}_{T_0}$  of  $T_0$ . We want to find a model  $\mathcal{M}$  for the whole of  $T$ .

Begin by replacing  $T$  with the theory consisting of the finite (nonempty) conjunctions of sentences in  $T$ . Obviously,  $T$  has a model iff this new theory has a model, and any finite

subset of the new theory has models. To simplify notation, I will then simply call this extension  $T$  as well.

Now let  $X = [T]^{<\mathbb{N}}$  be the collection of finite subsets of  $T$ . For  $\phi \in T$ , let

$$X_\phi = \{T_0 \in X \mid \mathcal{M}_{T_0} \models \phi\}.$$

The following is then immediate:

**Fact 11**  $\{X_\phi \mid \phi \in T\}$  has the fip.  $\square$

There is then an ultrafilter  $\mathcal{U}$  on  $X$  such that all the sets  $X_\phi$  are  $\mathcal{U}$ -large for  $\phi \in T$ .

Let  $\mathcal{M} = \prod_{T_0 \in X} \mathcal{M}_{T_0} / \mathcal{U}$ . This is the model of  $T$  we were trying to find, for if  $\phi \in T$ , then  $X_\phi \in \mathcal{U}$ , and therefore by Łoś's lemma,  $\mathcal{M} \models \phi$ .

*Typeset using LaTeX2WP.*