KIDDIE TALK: ULTRAPRODUCTS AND SZEMÉREDI'S THEOREM

EVAN WARNER

Outline and references

As promised, this talk will *not* assume any knowledge of ultrafilters, ultraproducts, or any model theory. The plan is as follows:

(1) Ultrafilters
(2) Ultraproducts
(3) Joke
(4) What are ultraproducts good for?
(5) Szeméredi's theorem via ultraproducts

References: Mostly Terence Tao's various blog notes on ultraproducts. For model theory, see David Marker's Model Theory: An Introduction for a very readable exposition.

1. Ultrafilters on $\mathbb{N}$

Without giving any particular context, suppose I want a gadget that tells me whether subsets of the natural numbers $\mathbb{N}$ are "big" or "small." Obviously, if we had some measure on $\mathbb{N}$ this would be one way of providing such a gadget, but a measure is more discerning and also less strict than what we want here. Specifically, we want a set ω of subsets of $\mathbb{N}$. We'll call a set $A \subset \mathbb{N}$ "big" if it lies in ω, and "small" otherwise. We want ω to satisfy the following axioms:

(1) $A \in \omega$ if and only if $\mathbb{N} \setminus A \in \omega$. That is, the complement of a "big" set is "small" and vice-versa.
(2) If $A \in \omega$ and $A \subset B$, then $B \in \omega$. That is, any superset of a "big" set is "big." Equivalently, from the first axiom, any subset of a "small" set is "small."
(3) (The weird one) If $A, B \in \omega$, then $A \cap B \in \omega$. That is, the intersection of two (hence, finitely many) "big" sets is "big." Bigness is somewhat robust.

Such an ω is called an *ultrafilter* (over $\mathbb{N}$).

So let's try to write down an example of an ultrafilter. Here's one: take some $n \in \mathbb{N}$, and let ω_n consist of all subsets that contain the element n. We can quickly check the axioms and see that they are all obviously verified. This is called a *principal ultrafilter*, and they're kind of silly. So silly, in fact, that I am immediately going to outlaw them with a new axiom:

(4) For all $n \in \mathbb{N}$, $\{n\} \notin \omega$.

That is, no set with a single element is a big set. It is an easy consequence of the other axioms that, in fact, no *finite* set can be a big set either. Ultrafilters that satisfy this fourth axiom are called, unsurprisingly, *nonprincipal ultrafilters*.

Question: are there any nonprincipal ultrafilters?

Theorem 1. *Yes.* □

They are far from unique. The proof is straightforward: build up bit by bit using Zorn's Lemma. We are essentially making a whole lot of arbitrary choices as to which sets should be big and which sets should be small, subject to the above axiomatic conditions.

Of course, because we have used Zorn's Lemma, we should not expect to be able to write down any examples of a nonprincipal ultrafilter, and in fact this phenomenon does hold: the existence of ultrafilters is provably equivalent to a strictly weaker version of the axiom of choice called the Boolean prime ideal theorem. In particular, we needn't have used the full strength of Zorn's lemma in the above theorem, but we do need to use something stronger than Zermelo-Fraenkel set theory. (Aside: a general theorem of Gödel asserts that any theorem stated in the language of Peano arithmetic that can be proven in ZFC can be proven in ZF, so if we're proving something concrete with ultrafilters there does exist a finitary proof somewhere.)

2. Ultraproducts

So now that we have our gadget ω, we want to use it to construct things. The construction is the ultraproduct, and we'll first define it on sets.

Let $\{A_i\}_{i\in\mathbb{N}}$ be a collection of sets indexed by the natural numbers. Define an equivalence relation $\sim$ on the Cartesian product as follows: two sequences $(a_1, a_2, \ldots)$ and $(b_1, b_2, \ldots)$ are equivalent if $\{i : a_i = b_i\} \in \omega$. That is, two sequences are equivalent if they agree on a big set. This is a legitimate equivalence relation only because the intersection of two big sets is big! The ultraproduct is then defined to be the quotient by this equivalence relation:

$$\mathbf{A} = \prod_{i\in\mathbb{N}} A_i / \sim .$$

Given a sequence (i.e., element of the Cartesian product) $(a_1, a_2, \ldots)$, we call its equivalence class in the ultraproduct the *ultralimit* of the sequence, and denote it by

$$\lim_{i\to\omega} a_i.$$

This notation is meant to be suggestive. There is a profitable analogy here between the construction of an ultraproduct and the completion of a metric space: in both cases, we take as our objects some collection of sequences modulo an equivalence relation. For the completion of a metric space, we take the collection of Cauchy sequences modulo the sequences which tend to zero; for the ultraproduct we take the collection of all sequences modulo a much stronger equivalence relation. It makes sense to think of the ultraproduct as allowing us to take limits of arbitrary sequences.

So far this is still relatively silly, because sets are boring. Now assume that the collection of objects $\{A_i\}$ all have some more structure: e.g., they all have a binary operation, or they all have a distinguished element called the "identity," for example. What we want is the model-theoretic notion of a *structure*: a set equipped with constants, functions, and relations. If the $\{A_i\}$ all have the same constants, functions, and relations, we can define these on the ultraproduct in the obvious way, by defining them term by term on the product. One needs to check that this is well-defined (respected by the equivalence relation), but that is very straightforward.

We can also define the ultralimit of functions: say we have $f_i : A_i \to B_i$ for each i. Then we define the ultralimit $\lim_{i\to\omega} f_i$ by

$$(\lim_{i\to\omega} f_i)(\lim_{i\to\omega} a_i) = \lim_{i\to\omega} f_i(a_i).$$

One again has to check that this is well defined.

From the model-theoretic point of view, we haven't gotten to "the point" yet, which is the following:

Theorem 2 (Łoś' theorem, informally)**.** *Any statement of first order logic holds true in the in the ultraproduct* $\boldsymbol{A}$ *if and only if it holds true in a big subset of the* A_i*.*

In particular, if the A_i are all the same model, then this model is *elementarily equivalent* to the ultraproduct $\mathbf{A}$: any first-order statement is true in one if and only if it is true in the other. Recall that a first order statement is one that can be written down with the usual logical symbols and quantifiers, but we are only allowed to quantify over elements, not sets (or sets of sets, etc.). So for example, we can say something like "for all elements x, y we have..." but we cannot say "for all subsets X we have...".

This is a powerful theorem. It tells us immediately that, for example, if the A_i are groups, so is $\mathbf{A}$, because the group axioms are expressible as statements in first order logic. But in general it tells us if there is something feature we care about in the A_i, and stating that feature doesn't depend on quantifying over sets, then $\mathbf{A}$ also has that feature (and vice-versa, at least for a big set of the A_i). As an aside, the proof of Łoś' theorem is not difficult, once one has defined what things mean precisely; it proceeds by induction on complexity (of the first order sentence).

Now we need an example to see how this works in practice. Let $A_i = \mathbb{R}$ for each i. The resulting ultraproduct is called an ultrapower (because all of the A_i are the same); in this specific case it is called the *hyperreal numbers* and denoted $^*\mathbb{R}$. By Łoś, we know that the hyperreals form an ordered field. What does this field look like? First note that we can embed $\mathbb{R}$ in $\mathbb{R}^*$, by mapping r to the equivalence class of $(r, r, r, \ldots)$. This forms a subfield, obviously, whose elements we call the *standard reals*. But there are more elements than these. Consider $\lim_{n\to\omega} n$, which is an element greater than any standard real (under the embedding we've just defined). We consider it to be an infinite element of $\mathbb{R}^*$. Similarly, $\lim_{n\to\omega}(n+1)$ is a bigger infinite element (not equal to the first infinite element, because they disagree on every coordinate), and $\lim_{n\to\omega} n^2$ is a much bigger infinite element. On the other side of things, $\lim_{n\to\omega}(1/n)$ is less than every standard real, but still greater than zero, so we consider it an infinitesimal.

One should think of $^*\mathbb{R}$ as a "richer" real number line, where every "ordinary" (standard) real number is surrounded by a "halo" of numbers infinitely close to it and we have a whole host of infinite numbers to boot. In fact (this is not at all hard to see), we can decompose each *finite* element $x \in {}^*\mathbb{R}$ uniquely as $x = \mathrm{st}(x) + (x - \mathrm{st}(x))$, where $\mathrm{st}(x)$ is a standard real and $x - \mathrm{st}(x)$ is infinitesimal. This map st is a ring homomorphism.

As another example, consider $^*\mathbb{N}$, which is just a subset of $^*\mathbb{R}$ containing the natural numbers, a lot of infinitesimals, and a lot of infinite numbers. This also gives us a great example of how to abuse Łoś' theorem: consider the (true) statement that $\mathbb{N}$, under its standard ordering, is well-ordered. Is $^*\mathbb{N}$ well-ordered? NO! The definition of well-ordered contains a quantification over sets, so we cannot expect it to behave well under ultraproducts.

As a final remark here, I'll mention that I'm being a little too cautious here about Łoś' theorem. We can add sets to our language, and then some first order statements involving sets will obey Łoś' theorem just fine. The problem is that lots of sets that we are interested in are not *internal sets* - they cannot be written as an ultralimit of subsets, and no incarnation of Łoś' theorem can possibly apply. For example, $\mathbb{N} \subset {}^*\mathbb{N}$ is not an internal subset (I'll leave this as a fairly interesting exercise).

3. Joke (unrelated)

Q: What do you call a geometer who prefers classical varieties to schemes?
A: A radical idealist.

4. What are ultraproducts good for?

(1) First, some general things about why they might be mathematically useful, from a couple of different perspectives.
   (a) Ultraproducts obey something called *saturation* (or, specifically, $\aleph_1$-saturation). I can't really give a definition of what this means without invoking a fair bit of basic model

theory, but in concrete terms, this implies a countable compactness result for internal sets: namely, if an internal set has a countable cover by other internal sets, then there is a finite subcover.

(b) As alluded to above, we can understand ultraproducts as realizing *arbitrary* limits of sequences. These "logical limits" are sometimes more flexible than, for example, topological or categorical limits.

(c) Paraphrasing Terence Tao: When taking the ultrapower of something, for example $\mathbb{R}$, we get everything we like about $\mathbb{R}$, but we have some new adjectives to work with, like "standard" and "internal".

(2) In model theory proper, one can take ultrafilters and ultraproducts over a set of any specified cardinality, not just $\mathbb{N}$. As an example of an application within model theory, ultraproducts give a quick proof of the compactness theorem in first order logic: a set of first order sentences has a model if and only if every finite subset has a model. In other words, if there is a contradiction, there is a finite contradiction.

(3) The field of *nonstandard analysis*, which reformulates standard notions in analysis in terms of the hyperreals, which sometimes simplifies things (we get actual infinitesimals to work with!). This has been proposed pedagogically, as a good way of teaching rigorous calculus.

(4) In analysis, lengthy proofs that involve juggling lots of epsilons that relate to each other in subtle ways can be much simplified by a judicious use of ultraproducts.

(5) Loeb measures: although the concept of a "measure space" cannot be formulated in first order logic, there is a useful way of putting the structure of a measure space on the ultraproduct of measure spaces. This concept has been used often in the field of stochastic PDEs, and there is a recent series of papers formulating "higher order Fourier analysis" in terms of these measure spaces. We will implicitly be using this construction in our main example below.

(6) Quantitative algebra and algebraic geometry: bounding results in terms of complexity of inputs. As an example, take Gromov's theorem that every finitely generated group of polynomial growth is virtually nilpotent (that is, has a nilpotent subgroup of finite index). One can use ultralimits to prove very quickly that we can bound the nilpotent subgroup's index and nilpotence step in terms of the given polynomial bound (i.e. degree and leading coefficient) on the growth of the group.

(7) A "souped-up" version of this kind of thinking can be used to provide a very efficient proof of the Szeméredi regularity lemma, which is a famous and very useful result of 1978 stating in a precise sense that every large enough graph can be divided into subsets of about the same size so that the edges going between these subsets behave almost randomly.

5. Furstenberg recurrence to Szeméredi, by way of ultraproducts

This is a very, very small part of a much larger story that links additive combinatorics, ergodic theory, graph theory, and Fourier analysis (together with generalizations to a still-poorly-understood "higher order" Fourier analysis). For those of you who were at the first number theory seminar a few weeks ago, for instance, Tamar Ziegler's work on solving linear equations in primes is another part of this subject.

One of the earliest, and most famous, theorems in this story is Szeméredi's theorem in additive combinatorics, which I can state very quickly. Let A be a set of integers. If

$$\limsup_{n\to\infty} \frac{|A \cap [-n,n]|}{|[-n,n]|} \geq \delta$$

for some $\delta > 0$, we say that A has *positive upper density*.

Theorem 3 (Szeméredi's theorem)**.** *Every set of integers of positive upper density contains arbitrarily long arithmetic progressions.*

Szeméredi's original 1975 proof was both intricate and clever, and answered a longstanding conjecture in the positive. Two years later, Furstenberg came along with a totally new proof from ergodic theory, using something called the "Furstenberg correspondence principle" to relate results in ergodic theory with arithmetic progressions. There are now other proofs as well: most famously, Green in 2001 came up with a proof that gave explicit, though rather frighteningly large, bounds.

Our goal here is to exhibit a particular case of this correspondence principle using ultraproducts, deriving the above famous theorem from a result in ergodic theory called the Furstenberg recurrence theorem:

Theorem 4 (Furstenberg recurrence theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, let $A \subset X$ have positive measure, and let k be a positive integer. Then there exists a positive integer r such that*

$$A \cap T^r A \cap \ldots \cap T^{(k-1)r} A$$

is nonempty.

By a measure-preserving system, I mean that $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is a measure-preserving bijection such that both T and T^{-1} are measurable. This is by no means a trivial theorem, by the way, but it is a result belonging wholly to ergodic theory (the proof is entirely ergodic theoretic).

The proof of Szeméredi's theorem will proceed by contradiction. Assume we have a set A of positive upper density and no progressions of length k for some $k \geq 1$. This means that there exists a sequence of integers $\{n_i\}$ tending to infinity and a $\delta > 0$ such that

$$|A \cap [-n_i, n_i]| \geq \delta |[-n_i, n_i]|$$

for each i. Now consider the ultrapower ${}^*A$ with respect to some nonprincipal ultrafilter ω, and the ultralimit $n = \lim_{i \to \omega} n_i$. The former is a subset of ${}^*\mathbb{N}$, while the latter is in ${}^*\mathbb{N}$ (which can be considered a subset of the hyperreals). We claim that

$$|{}^*A \cap [-n, n]| \geq \delta |[-n, n]|,$$

where the absolute value signs are as defined on the hyperreals; i.e., the ultralimit of the usual absolute value function (and $\delta = \lim_{i \to \omega} \delta$ is considered as a hyperreal as well). This is a simple application of Łoś' theorem, but let's be pedantic. Take the signature consisting of the binary operations $+$ and $\cdot$, the binary relation $\geq$, the unary relation (or set) A, and the constant δ. By Łoś' theorem, since the above inequality can be written as a first-order statement and is true for each i, and $\mathbb{N}$ is certainly an element of ω, by taking ultralimits of everything we get a true statement, which is precisely the claim. Additionally, and most crucially, it is true that ${}^*A$ has no progressions of length k, since that property can be expressed in a first-order sentence and is true for A.

Call the space of all finite Boolean combinations of shifts of ${}^*A$ by elements of $\mathbb{Z}$, which includes the empty set and all of ${}^*\mathbb{Z}$, the *definable sets* $\mathcal{D}$. The elements of $\mathcal{D}$ form a field of sets (they are closed under finite Boolean operations). The immediate goal is to define a finitely additive premeasure on $\mathcal{D}$.

Define $\mu' : \mathcal{D} \to {}^*\mathbb{R}$ by

$$\mu'(B) = \frac{|B \cap [-n, n]|}{|[-n, n]|},$$

which is a function from definable sets to the set ${}^*[0, 1]$, the ultrapower of the unit interval, such that $\mu'(\emptyset) = 0$ and $\mu'({}^*\mathbb{Z}) = 1$. Additionally, it is finitely additive: given two disjoint definable sets

B_1 and B_2,

$$\begin{aligned}\mu'(B_1 \cup B_2) &= \frac{|(B_1 \cup B_2) \cap [-n, n]|}{|[-n, n]|} \\ &= \frac{|B_1 \cap [-n, n]| + |B_2 \cap [-n, n]|}{|[-n, n]|} \\ &= \mu'(B_1) + \mu'(B_2).\end{aligned}$$

Finally, it is nearly translation invariant under translation by any standard integer j: we have

$$\begin{aligned}\mu'(B + j) &= \frac{|B \cap [-n - j, n - j]|}{|[-n, n]|} \\ &= \frac{|B \cap [-n, n]|}{|[-n, n]|} + \frac{|B \cap [-n - j, -n - 1]|}{|[-n, n]|} - \frac{|B \cap [n - j + 1, n]|}{|[-n, n]|} \\ &= \mu'(B) + \epsilon_1 + \epsilon_2,\end{aligned}$$

where ϵ_1 and ϵ_2 are both infinitesimal, since both numerators are finite but both denominators are infinite. Define $\mu : \mathcal{D} \to \mathbb{R}$ by

$$\mu(B) = \mathrm{st}(\mu'(B))$$

and recall that st is a homomorphism. It is clear that μ is nonnegative, $\mu(\emptyset) = 1$, $\mu({}^*\mathbb{Z}) = 1$, and μ is finitely additive. Taking standard parts kills infinitesimal quantities, so μ is a finitely additive premeasure on the field of sets $\mathcal{D}$ that is invariant under shifts by finite integers. By construction, $\mu({}^*A) \geq \delta$.

We will use μ to construct a measure-preserving system on which to apply the Furstenberg recurrence theorem. Let $X = 2^{\mathbb{Z}}$, the set of all subsets of the integers. We can view this space as the set of all doubly-infinite sequences of ones and zeroes. Define the *cylinder sets* E_m by

$$E_m = \{B \in 2^{\mathbb{Z}} : m \in B\}.$$

From the viewpoint of the doubly-infinite sequences, E_m consists of all sequences with 1 in the mth position. Let the set of all finite Boolean combinations of the E_m, which is clearly a field of sets, be denoted $\mathcal{E}$. Let $T : X \to X$ be a shift-by-one map, shifting each entry of a doubly-infinite sequence to the left. To define a measure ν on this space, we will need to somehow reference our ultraproduct construction.

This reference is provided by a map $\Phi : \mathcal{E} \to \mathcal{D}$, which is defined on the E_m by

$$\Phi(E_m) = {}^*A + m$$

and then extended to all of $\mathcal{E}$ by Boolean combinations. Let $\nu : \mathcal{E} \to \mathbb{R}$ be defined for $E \in \mathcal{E}$ by

$$\nu(E) = \mu(\Phi(E)),$$

the pullback of μ by Φ. It is simple to verify for ν the properties we want: it is nonnegative, we have $\nu(\emptyset) = \mu(\emptyset) = 0$, $\nu(X) = \mu({}^*\mathbb{Z}) = 1$, and it is clearly invariant under the shift T because μ is invariant under shifts by integers. Finite additivity follows because Φ preserves disjointness of sets, which follows because we have for $E_1, E_2 \in \mathcal{E}$ that $E_1 \cap E_2 = \emptyset \implies \Phi(E_1) \cap \Phi(E_2) = \Phi(E_1 \cap E_2) = \emptyset$. Finally, we of course have $\nu(E_0) \geq \delta$.

By the Carathéodory extension theorem, ν extends to a measure on the σ-algebra $\mathcal{B}$ generated by $\mathcal{E}$ on X. We may now finally apply the Furstenberg recurrence theorem on the measure-preserving system $(X, \mathcal{B}, \nu, T)$ and the set E_0 of positive measure to deduce that there is a positive integer r such that

$$E_0 \cap T^r E_0 \cap \ldots \cap T^{(k-1)r} E_0 \neq \emptyset.$$

By applying Φ, it is immediate that

$$^*A \cap (^*A + r) \cap \ldots \cap (^*A + (k-1)r) \neq \emptyset.$$

Therefore $^*A$ contains an arithmetic progression of length k, which is a contradiction.