1 Model theory; introduction and overview

Model theory, sometimes described as "algebra with quantifiers" is, roughly, mathematics done with attention to definability. The word can refer to definability of a whole class of abstract mathematical objects (like the way we can axiomatically define groups, or metric spaces, or ...) but more often it refers to how we define significant subsets of such objects (like solution-sets of polynomial equations, or ϵ-neighbourhoods of points, ...). Sometimes we will be looking at uniform definability within a class of objects (e.g. we can write down a definition of the centre of a group which will apply in all groups), more often we look at the definable subsets of a specific mathematical object (of, e.g., n-dimensional space over a field, or the complex numbers).

"Definability" means definability in some formal language which is set up to be appropriate to the kind of structure we are interested in. Describing such (finitary, first-order) languages is something that most (but probably not all) of you have seen in a course that includes some Predicate Logic. Having chosen a language $\mathcal{L}$, one describes how to build up the terms and formulas of $\mathcal{L}$. Then one says what is meant by an $\mathcal{L}$-structure and one makes the connection between syntax (formulas of $\mathcal{L}$) and semantics (the interpretation of formulas in $\mathcal{L}$-structures), in particular describing how to read $\mathcal{L}$-formulas in $\mathcal{L}$-structures. If you've not seen this before, or want to remind yourself about it, then see the Appendix, which also defines homomorphisms - the structure-preserving maps which connect $\mathcal{L}$-structures.

In this course, we will emphasise 'semantics' - structure and particular examples, taking the view that the more formal aspects are what we use for proving general results that can then be applied in many different contexts.

We will begin by describing the ultraproduct construction and giving some examples of the weird and wonderful things that we can get by using it and the associated theorem of Łos. We will also use it to give a proof the Compactness Theorem which does not depend on setting up a formal proof system.

The interplay of syntax (the formulas of a formal language) and semantics (the meanings of those formulas in structures) leads to the notion of the theory of a structure (the set of all sentences true in the structure) and the class of models of a theory (the class of structures in which all sentences in the theory are true). The latter is an extension of the usual way of specifying abstract classes by writing down a set of axioms. In this "logic narrative", elementary equivalence - having identical syntactically-expressible properties - is the key relation and the class of models of a theory is the natural context, with a central task being to classify the models.

In the more algebraic narrative, we can start with the observation that we often want to understand the solution sets of equations or inequalities and, if there is more than one variable, the projections of these solution sets to fewer variables. These are subsets which are definable in the kind of language $\mathcal{L}$ referred to above, so understanding the definable subsets becomes a central task.

We'll see ideas and results from both these strands which, in any case, cannot be disentangled. Model theory is abstract - it applies in many different contexts (algebraic, analytic, geometric) - and concept-heavy but we will use a variety of examples - some familiar, some less so - to illustrate and apply the results; this should help to make them more concrete (or 'concrete' - this is pure maths).

2 Ultraproducts and Łos' Theorem

2.1 Producing infinitesimals

Question 2.1. Is $0.\overline{9} = 1.0$?

Consider the difference $\epsilon = 1 - 0.\overline{9}$; what can we say about it? It's easy to see that we must have $\epsilon < 1/n$ for every positive integer n but why shouldn't we be able to add the condition $\epsilon > 0$? We can appeal to the Compactness Theorem, which you might have seen already (and which we will prove using ultraproducts).

Theorem 2.2. *(Compactness Theorem v.1) If you want something and there's no reason you can't have it, then you can get it.*

(There is some small print; we'll get to that.)

The Compactness Theorem applies to our question about ϵ: let's consider the set, $\{\epsilon < 1/n : n \in \mathbb{Z}^+\} \cup \{\epsilon > 0\}$, of conditions on ϵ. If we take any finitely many of these, then there is a solution in the reals $\mathbb{R}$, so "there's no reason we can't have" ϵ. Admittedly the Completeness Property for the reals does exclude there being a solution in $\mathbb{R}$ but, and this is part of the small print, *we might have to move* to get the "something" in the theorem.

What that means in this example is that there is a structure - a non-standard version, $\mathbb{R}^*$, of the reals - which has an **infinitesimal** (a solution to all those conditions). So, in $\mathbb{R}^*$, we will have $0.\overline{9} < 1 - \epsilon < 1$, giving an alternative to the standard answer to the original question. This structure $\mathbb{R}^*$ will share a great many properties with the reals and it will have a copy of the reals sitting nicely inside it. But it will contain an infinitesimal, with all that that implies. For example, since $\mathbb{R}$ is an ordered field, so also will be $\mathbb{R}^*$, therefore $\mathbb{R}^*$ will contain elements, such as ϵ^{-1}, greater than every integer and it will contain many infinitesimals, $\epsilon + \epsilon$, ϵ^2, ϵ^3, ...

The Compactness Theorem, properly stated, says that if we have a set of conditions (of a certain form: namely which can be expressed by formulas of a first order predicate language appropriate for the structure, $\mathbb{R}$ in this case) such that every finite subset has a solution in some structure M, then there will be an "elementary extension" M^* of M which contains a simultaneous solution to all the conditions. Of course it might be that the original structure M already contains a solution but it may be that, as in our example, the original structure contains no solution and we do have to move to a proper elementary extension.

The definition is: M^* is an **elementary extension** of M (equally, M is an **elementary substructure** of M^*) if, whenever $\varphi(\overline{x})$ is a formula (in n free variables and parameters from M) then the solution set $\varphi(M)$ in M is the intersection of M^n with the solution set $\varphi(M^*)$ in M^*. Even if you don't know what is meant by a formula (with parameters), perhaps this gives some flavour of the idea. Here's a specific example. Can -1 be a square in $\mathbb{R}^*$? No, because, if so, then it would be in the solution set in $\mathbb{R}^*$ of the formula $\exists y\,(y^2 = x)$. So it would be in the intersection of that solution set with $\mathbb{R}$, hence in the solution set of the same formula in $\mathbb{R}$ - which is not the case. Here, $\exists y\,(y^2 = x)$ is an example of a "formula"; an example of a formula with parameters (namely $\pi \in \mathbb{R}$) in this case) is $\quad x^3 = \pi \wedge \forall y\,(y^3 = \pi \rightarrow y = x) \quad$ (read $\wedge$ as "and").

Formulas with no free variables are just statements (or "sentences") and if M is an elementary substructure of M^* then M and M^* must satisfy the same sentences. For instance $\mathbb{R}^*$ will be densely ordered because $\mathbb{R}$ satisfies the sentence $\forall y, z\, (y < z \to \exists w\, (y < w \wedge w < z)$ which expresses the "densely ordered" property.

Precise definitions of "language", "term", "formula", "sentence", "structure" *etc.* are given in the Appendix sections on Predicate Logic. I will explain these briefly but, if they are new to you, you should refer to the Appendix for definitions, explanations and examples.

Rather than quote the Compactness Theorem, we will directly produce an elementary extension of the reals which contains an infinitesimal. We will use the ultraproduct construction.

First we form the direct product $\mathbb{R}^{\mathbb{P}}$ where $\mathbb{P}$ denotes the set of positive integers. This is the set of sequences $(r_i)_{i\in\mathbb{P}}$ with each $r_i \in \mathbb{R}$. We define addition and multiplication on $\mathbb{R}^{\mathbb{P}}$ pointwise: $(r_i)_i + (s_i)_i = (r_i + s_i)_i$ and $(r_i)_i \times (s_i)_i = (r_i \times s_i)_i$. Then these operations make $\mathbb{R}^{\mathbb{P}}$ into a commutative ring with multiplicative identity element $1 = (1_i)_i$ and additive zero element $0 = (0_i)_i$. Notice, *exercise*, that, unlike $\mathbb{R}$, $\mathbb{R}^{\mathbb{P}}$ is not a field.

We are going to factor out a (maximal) ideal so as to obtain a field. We need the following definitions.

Definition 2.3. *A set $\mathcal{F}$ of subsets of $\mathbb{P}$ is a* **filter on** $\mathbb{P}$ *if:*
- $\mathbb{P} \in \mathcal{F}$;
- $\emptyset \notin \mathcal{F}$;
- *if* $J \subseteq K \subseteq \mathbb{P}$ *and* $J \in \mathcal{F}$ *then* $K \in \mathcal{F}$;
- *if* $J, K \in \mathcal{F}$, *then* $J \cap K \in \mathcal{F}$.

Given any such filter $\mathcal{F}$, we define the set $Z_{\mathcal{F}} = \{(r_i)_{i\in\mathbb{P}} : \{i : r_i = 0\} \in \mathcal{F}\}$ - the set of elements of $\mathbb{R}^{\mathbb{P}}$ which are zero on a set of coordinates in $\mathcal{F}$. Then, *exercise*, $Z_{\mathcal{F}}$ is a (proper) ideal of $\mathbb{R}^{\mathbb{P}}$. In order that $\mathbb{R}^{\mathbb{P}}/I_{\mathcal{F}}$ be a field, it is necessary and sufficient that $Z_{\mathcal{F}}$ be a maximal ideal (*exercise* if you haven't seen/don't recall how to prove this fact). That will be the case iff $\mathcal{F}$ is a maximal filter - or "**ultrafilter**", meaning a filter $\mathcal{F}$ on $\mathbb{P}$ such that, if $\mathcal{F}'$ is any filter on $\mathbb{P}$ with $\mathcal{F} \subseteq \mathcal{F}'$, then $\mathcal{F} = \mathcal{F}'$. It is the case that, if $\mathcal{F}$ is any filter on $\mathbb{P}$, then there is an ultrafilter $\mathcal{U}$ on $\mathbb{P}$ with $\mathcal{F} \subseteq \mathcal{U}$. This follows from Zorn's Lemma, which we will discuss later when we look at the ultraproduct construction in the general context.

Here's how we can use this to produce infinitesimals in some extension/enriched version of the reals.

Define the filter $\mathcal{F}$ to consist of all the **cofinite** subsets I of $\mathbb{P}$ (those with finite complement); *exercise*: show that this is indeed a filter on $\mathbb{P}$. (*Another exercise*: show that $\mathcal{F}$ is the smallest filter containing all the subsets of $\mathbb{P}$ of the form $\{m : m \geq n\}$ for some $n \in \mathbb{P}$.) Let $\mathcal{U}$ be any ultrafilter containing $\mathcal{F}$ and consider the ideal $Z_{\mathcal{U}}$ of $\mathbb{R}^{\mathbb{P}}$: $Z_{\mathcal{U}} = \{(r_i)_i : \{i : r_i = 0\} \in \mathcal{U}\}$. Then the quotient ring $\mathbb{R}^{\mathbb{P}}/Z_{\mathcal{U}}$ is a field (*exercise*) - either show this directly or show that $Z_{\mathcal{U}}$ is a maximal ideal) denoted $\mathbb{R}^*$ say. Wee show that $\mathbb{R}^*$ contains a copy of $\mathbb{R}$.

Consider $\delta : \mathbb{R} \xrightarrow{\Delta} \mathbb{R}^{\mathbb{P}} \xrightarrow{\pi} \mathbb{R}^*$, where Δ is the diagonal embedding, given by taking an element $r \in \mathbb{R}$ to the constant sequence $(r)_i$, and where π is the canonical projection of the ring $\mathbb{R}^{\mathbb{P}}$ to its factor ring $\mathbb{R}^* = \mathbb{R}^{\mathbb{P}}/Z_{\mathcal{U}}$. Then, we claim, δ is an embedding of rings: we must show that $\delta(r) = \delta(s)$ implies that

$r = s$. That follows since $\delta(r) - \delta(s) = 0$ implies $\Delta(r) - \Delta(s) \in \ker(\pi) = Z_{\mathcal{U}}$, hence that $(\Delta(r))_i = (\Delta(s))_i$ for some (and hence for every) coordinate i, hence that $r = s$.

Finally, we show that $\mathbb{R}^*$ contains an infinitesimal: namely the element $\epsilon = \pi((1/i)_i)$. For that, we refer to the orderings on $\mathbb{R}$ and $\mathbb{R}^*$. These can be defined just using the arithmetic operations ($x \le y$ iff $y - x$ is a square) but we can define the ordering on $\mathbb{R}^*$ directly by setting $\pi((r_i)_i) \le \pi((s_i)_i)$ iff $\{i : r_i \le s_i\} \in \mathcal{U}$. It's an *exercise* to show that this is well-defined, meaning independent of choices of representatives of equivalence classes and that it gives a total ordering on $\mathbb{R}^*$; it's a nice additional *exercise* to show then that the ordering defined this way can also be defined algebraically as stated above.

So, that done, let's check that ϵ is, indeed, an infinitesimal. We will identify $\mathbb{R}$ with the copy, $\delta(\mathbb{R})$, of it sitting inside $\mathbb{R}^*$.

Given $n \in \mathbb{P}$, we have that $1/i < 1/n$ for all $i \ge n+1$, hence $\{i \in \mathbb{P} : (1/i)_i < \Delta(1/n)\} \in \mathcal{U}$, so $\epsilon < 1/n$ (where the latter really means $\delta(1/n)$).

Finally $\epsilon \neq 0$ since clearly $\epsilon \notin Z_{\mathcal{U}}$.

That was all rather fast (and particular); in the next few lectures we will go through the ideas and constructions more slowly, and in great generality (but using many examples to illustrate the general ideas).

2.2 Products, filters and ultraproducts

Suppose that I is some index set, and that, for each $i \in I$ we have a structure M_i, where the M_i all are structures of the same kind (all groups, or rings, or partially ordered sets, or...), that is, all $\mathcal{L}$-structures for some language $\mathcal{L}$. Their **product** is, as a set, the product $\prod_{i \in I} M_i$ of their underlying sets. This has, for its elements, the sequences $(a_i)_{i \in I}$ with $a_i \in M_i$. So these are sequences, indexed by I and with the ith coordinate coming from the structure M_i.[1] We will also use the shorter notations $\overline{a} = (a_i) = (a_i)_i$ for such elements.

The $\mathcal{L}$-structure is defined on this set pointwise. For instance, if we have a binary operation, denoted $+$ say, in each structure (more preciseli, let $+_i$ denote the operation on M_i), then we define the operation $+$ on $\prod_i M_i$ by $(a_i)_i + (b_i)_i = (a_i + b_i)_i$ (more precisely, $(a_i +_i b_i)_i$). Here's the general definition.

Given $\mathcal{L}$-structures M_i $(i \in I)$ we make the product $\prod_i M_i$ into an $\mathcal{L}$-structure as follows.

• For each constant symbol c in $\mathcal{L}$, we define the interpretation $c^{\prod M_i}$ of c in $\prod M_i$ to be the element $(c^{M_i})_i$.

• Given an n-ary function symbol f in $\mathcal{L}$, we define its interpretation $f^{\prod M_i}$ in $\prod M_i$ to be the function given by: if $a^1, \dots, a^n \in \prod M_i$, with $a^j = (a_i^j)_i$, then $f(a^1, \dots, a^n) = (f^{M_i}(a_i^1, \dots, a_i^n))_i$.[2]

• Given an n-ary relation symbol R in $\mathcal{L}$, we define its interpretation $R^{\prod M_i}$ in $\prod M_i$ to be set of all n-tuples $\overline{a} = (a^1, \dots, a^n) \in (\prod M_i)^n$ such that $R^{M_i}(a_i^1, \dots, a_i^n)$ holds for each $i \in I$ (where, as above, the ith coordinate of a^j is written a_i^j).

[1] Formally, the elements of such a product can be defined to be the functions a from I to $\bigcup_{i \in I} M_i$ such that, for every i, $a(i) \in M_i$.

[2] The notation is useful in that it lets us make a precise and general definition but it obscures the idea, and maybe you need to know the idea in order to make sense of the notation! To understand what is meant, take specific cases, like $n = 1$, $n = 2$ and maybe even a small index set I.

Exercise: to make sense of this, take $\mathcal{L}$ to be the language with one binary relation symbol, written $<$, take $I = \{1, 2\}$, take M_1 and M_2 to be respectively the sets $\{0, 1\}$ and $\{3, 4, 5\}$ both with their natural ordering. Figure out the ordering on the product (draw its Hasse diagram for example).

In this way we turn the product of any set of $\mathcal{L}$-structures into an $\mathcal{L}$-structure. If all the component structures M_i are copies of the same structure M, you can check (*exercise*) that the diagonal embedding $\delta' : M \to M^I$ is an embedding of $\mathcal{L}$-structures.

Let's use the following **running example** in this section: take the index set I to be the set of positive prime integers, and the structure indexed by $p \in I$ to be field $\mathbb{F}_p$ with exactly p elements, that is, the ring of integers modulo p, also written $\mathbb{Z}_p$ or $\mathbb{Z}/p$.

So what does the product construction give in this example? We get the structure $\prod_p \mathbb{Z}_p$ where p ranges over the primes. The *structure* on this is that of a ring: there's an addition and a multiplication, both defined coordinatewise, an identity $1 = (1_p)_p$ for the multiplication and an identity $0 = (0_p)_p$ for the addition, where 1_p denotes the congruence class of $1 \in \mathbb{Z}$ in $\mathbb{F}_p$ and similarly for 0. For example, $1{+}1 = (0_2, 2_3, 2_5, 2_7, 2_{11}, \dots)$, $1{+}1{+}1 = (1_2, 0_3, 3_5, 3_7, 3_{11}, \dots)$, *et cetera* (using a hopefully self-explanatory notation).

This makes the product into a commutative ring. For example, to check commutativity of the multiplication, we compute:
$(a_i)_i \times (b_i)_i = (a_i \times_i b_i)_i$ (by definition of the structure on the product)
$= (b_i \times_i a_i)_i$ (since each component structure is commutative)
$= (b_i)_i \times (a_i)_i$ (by definition).
It is not, however, a field: e.g. $(1, 0, 0, 0, \dots) \times (0, 1, 0, 0, \dots) = (0, 0, 0, 0, \dots)$.

We can, however, get a field from this product if we factor out by a maximal ideal, in other words, if we collapse elements appropriately. By which I mean that we will define an equivalence relation on the product, form the set of equivalence classes and induce a structure on that set (rather as we do in forming $\mathbb{F}_5 = \mathbb{Z}_5$ from $\mathbb{Z}$). Because the additive group structure is there, so we have cosets, it's actually enough to specify which elements get collapsed together with 0 - that is, to specify the ideal we factor out by - but, to better illustrate the general process, we'll not use that fact.

In forming an ultraproduct from a product, the idea is that we collapse (declare to be equivalent) elements which agree on a "large" set of coordinates.

What should we mean by a "large" set of coordinates? To recap: we have an index set I.[3] and, for each $i \in I$, we have an $\mathcal{L}$-structure M_i. We form the product $\mathcal{L}$-structure $\prod_{i \in I} M_i$ as above and we are going to identify/collapse elements which agree on a large set of coordinates. But how do we decide which subsets of I should count as "large"?

Certainly I itself should be a large subset (equal elements should be identified) and the empty set $\emptyset$ should not be large (collapsing all elements together would not give an interesting result). If $J \subseteq I$ is large and $J \subseteq K \subseteq I$ then surely K should also be large. If we're going to identify a and b and also identify b and c then we're going to have to identify a and c ("identification" will be an equivalence relation). If $J = \{i \in I : a_i = b_i\}$ and $K = \{i \in I : b_i = c_i\}$ are the, "large", sets where these pairs of elements agree, then all we can really say

[3] Think of I as being infinite; finite index sets won't give anything new.

about the set of coordinates where $a_i = c_i$ is that it contains $J \cap K$; so it looks as if we should require this set to be large. Let's extract those conditions.

Definition 2.4. *If I is a set then a* **filter** *on I is a collection $\mathcal{F}$ of subsets of I such that:*
- *$I \in \mathcal{F}$;*
- *$\emptyset \notin \mathcal{F}$;*
- *if $J \subseteq K \subseteq I$ and $J \in \mathcal{F}$ then $K \in \mathcal{F}$;*
- *if $J, K \in \mathcal{F}$, then $J \cap K \in \mathcal{F}$.*

Note (*exercise*) that, as a consequence of these clauses, if a subset $J \subseteq I$ is large then its complement $J^c = I \setminus J$ cannot be large.

Given a filter $\mathcal{F}$ on I, we define the corresponding equivalence relation $\sim_{\mathcal{F}}$, or $\sim$ for short, on the product $\prod_{i\in I} M_i$ by $(a_i)_i \sim (b_i)_i$ iff $\{i \in I : a_i = b_i\} \in \mathcal{F}$. Denote by $\prod_{i\in I} M_i/\mathcal{F}$ the set of equivalence classes, writing $a/\sim$ for the equivalence class of an element $a \in \prod_{i\in I} M_i$. We can then turn $\prod_{i\in I} M_i/\mathcal{F}$ into an $\mathcal{L}$-structure, defining operations and relations pointwise but paying attention only to what happens on "large" sets of indices. This structure is called the **reduced product** of the M_i with respect to the filter $\mathcal{F}$. If all the structures M_i are copies of the same structure M then we use the notation $M^I/\mathcal{F}$ and refer to this as a **reduced power** of M.

Before doing the general case carefully, let's do this with the running example, using the filter $\mathcal{F}$ of cofinite subsets of the set I of primes. We define the algebraic operations by setting $\big((a_p)_p/\sim\big) + \big((b_p)_p/\sim\big) = \big((a_p + b_p)_p/\sim\big)$ and $\big((a_p)_p/\sim\big) \times \big((b_p)_p/\sim\big) = \big((a_p \times b_p)_p/\sim\big)$. It has to be checked that this is well-defined (e.g. that if $(a'_p)_p \sim (a_p)_p$ and $(b'_p)_p \sim (b_p)_p$ then $(a'_p + b'_p)_p \sim (a_p + b_p)_p)$ but the conditions in the definition of a filter include what we need to do this (*exercise*). We can then check that $(0_p)_p/\sim$ is the zero for addition and $(1_p)_p/\sim$ is the identity for multiplication and, indeed, that all the axioms for a commutative ring are satisfied by $\prod_p \mathbb{F}_p/\mathcal{F}$ (more *exercises*).

Exercise 2.5. Prove that the map $\prod_p \mathbb{F}_p \to \prod_p \mathbb{F}_p/\mathcal{F}$ is a surjective homomorphism of rings and identify its kernel.

So let's do that for general structures M_i, $i \in I$. We're supposing that all these are $\mathcal{L}$-structures for some language $\mathcal{L}$ and we must turn the reduced product into an $\mathcal{L}$-structure. That means that we have to interpret every function, constant and relation symbol of the language. We can do it directly but it's quicker to define it now with reference to the $\mathcal{L}$-structure on $\prod_i M_i$. Recall that $\pi : \prod M_i \to M^* = \prod_{i\in I} M_i/\mathcal{F}$ is the projection map, which takes $a \in \prod M_i$ to its equivalence class $a/\sim$ where $\sim$ means $\sim_{\mathcal{F}}$
- For each constant symbol c in $\mathcal{L}$, we define $c^{M^*} = \pi(c^{\prod M_i})$, that is $(c^{M_i})_i/\sim$.
- Given an n-ary function symbol f in $\mathcal{L}$, we define f^{M^*} to be the n-ary function on M^* given by: if $b^1, \dots, b^n \in M^*$ then choose, for each $j = 1, \dots, n$ some $a^j \in \prod M_i$, $a^j = (a^j_i)_i$ say, with $\pi(a^j) = b^j$ and set $f^{M^*}(b^1, \dots, b^n) = \pi(f^{\prod M_i}(a^1, \dots, a^n))$. That is, choose a representative in $\prod M_i$ for each equivalence class $b^1, \dots, b^n$, evaluate the function f on that n-tuple in $\prod M_i$ and then take the $\sim$-equivalence class of the result. Of course it has to be shown that the result is independent of choice of representatives. You should do that as an important *exercise* - important because managing to do it probably means that you have got behind the notation and understood the idea.

• Given an n-ary relation symbol R in $\mathcal{L}$, we define R^{M^*} to be set of all n-tuples $\bar{b} = (b^1, \ldots, b^n) \in (\prod M_i)^n$ such that, if $a^1, \ldots, a^n \in M^*$ are such that $\pi(a^j) = b^j$ for each j, then $(a^1, \ldots, a^n) \in R^{\prod M_i}$. As in the case of functions, it's an *exercise* to show that this is well-defined (independent of the choices of pre-images/representatives a^j).

In this way, we make any reduced product of $\mathcal{L}$-structures into an $\mathcal{L}$-structure. We already saw this construction in our running example and in the ultrapower of the ordered field $\mathbb{R}$, but this, notationally rather unwieldy, definition shows how to do it in general.

Coming back to our running example using the prime fields $\mathbb{F}_p$, we produced a ring $\prod_p \mathbb{F}_p/\mathcal{F}$, where $\mathcal{F}$ is the filter of cofinite sets, but this is still not a field. To see that, split the primes into two infinite disjoint subsets J and K. Define the element a to be 1 on the indices in J and 0 on those in K; define b *vice versa*. Their product is 0 but neither is 0, so this ring is not even an integral domain, let alone a field.

So we need to go further, and impose a further condition on a filter.

Definition 2.6. *An* **ultrafilter** $\mathcal{U}$ *on a set* I *is a filter on* I *which satisfies the further equivalent conditions (we will prove their equivalence):*
• *for each* $J \subseteq I$ *either* $J \in \mathcal{U}$ *or* $J^{\mathrm{c}} = I \setminus J \in \mathcal{U}$*;*
• *if* $J \cup K \in \mathcal{U}$ *then either* $J \in \mathcal{U}$ *or* $K \in \mathcal{U}$*;*
• $\mathcal{U}$ *is a maximal filter (meaning that no collection of subsets of* I *can be a filter and properly include all the sets in* $\mathcal{U}$*).*

So an ultrafilter splits the subsets of I into "large" ones (those in $\mathcal{U}$) and "small" ones (those not in $\mathcal{U}$, equivalently, those whose complement is in $\mathcal{U}$). "Small" does not really mean small (say in the sense of cardinality), just small according to $\mathcal{U}$. But we do have that the union of two "small" sets is still "small" (this is the second of the equivalent conditions above).

In the case where we collapse using an ultrafilter $\mathcal{U}$ rather than any old filter, we refer to the result $\prod_i M_i/\mathcal{U}$ as an **ultraproduct** or, in the case that the M_i all are equal (or isomorphic), an **ultrapower**.

There is one type of ultrafilter that is not interesting. Suppose that $i_0 \in I$. Set $\mathcal{U}(i_0) = \{J \subseteq I : i_0 \in J\}$. Then, *exercise*, $\mathcal{U}(i_0)$ is an ultrafilter, called the **principal ultrafilter generated by** i_0. You can check that, in the ultraproduct $\prod_{i\in I} M_i /\mathcal{U}(i_0)$, all that matters is what happens at the coordinate i_0 indeed, another *exercise*, this ultraproduct is isomorphic to M_{i_0}, so we got nothing new from the construction. Therefore we will consider only *non-principal* ultrafilters. But first - are there any?

If I is any infinite set then the collection of all cofinite sets is a filter, sometimes called the **Fréchet filter** $\mathcal{F}_0$. If $\mathcal{U}$ is any ultrafilter containing $\mathcal{F}_0$ then $\mathcal{U}$ cannot be principal, and conversely (quick *exercise*). We do need to call on Zorn's Lemma to give the existence of a maximal=ultra filter containing any given filter but that can be proved (and is an *exercise* for those who have seen Zorn's Lemma).

Let's continue our example using, in place of the cofinite filter $\mathcal{F} = \mathcal{F}_0$, an ultrafilter $\mathcal{U}$ containing $\mathcal{F}$. Let's check that the ultraproduct $\prod_p \mathbb{F}_p/\mathcal{U}$, which equals the quotient ring $\prod_p \mathbb{F}_p/Z_{\mathcal{U}}$, is a field. So take any non-zero element $a/\sim\, = (a_p)_p/\sim\, \neq 0$ in the ultraproduct. Since it is not in the same $\sim$ class as 0, it must be that $J = \{p : a_p = 0\} \notin \mathcal{U}$. Since we have an *ultra*filter, it follows that $J^c = \{p : a_p \neq 0\} \in \mathcal{U}$. But whenever $a_p \neq 0$, a_p has an inverse, b_p say. Set $b = (b_p)_p \in \prod_p \mathbb{F}_p$ (where, if $p \in J$, set $b_p = 0$, say). Then $(a/\sim)(b/\sim) = (a_p b_p)/\sim = 1$, since the set of coordinates p where $a_p b_p = 1_p$ is in $\mathcal{U}$. So $a/\sim$ has a multiplicative inverse, $b/\sim$, as required.

Exercise 2.7. Consider the product $R = \prod_{i \in I} R_i$ where, for each i in the index set I, R_i is a ring (with 1, commutative if you like). Show that if $\mathcal{F}$ is a filter on I then $Z_{\mathcal{F}} = \{r = (r_i)_i \in R : \{i \in I : r_i = 0\} \in \mathcal{F}\}$ is a (2-sided) ideal of R, Prove that if $\mathcal{F}$, $\mathcal{G}$ are filters on I then $Z_{\mathcal{F}} \subseteq Z_{\mathcal{G}}$ iff $\mathcal{F} \subseteq \mathcal{G}$. Show that if $Z_{\mathcal{F}}$ is a maximal ideal of R then $\mathcal{F}$ must be an ultrafilter on I; is the converse true?

Exercise 2.8. Take the index set I to be the set of positive integers and, for each i, set M_i to be the ring $\mathbb{Z}$ of integers. Let $\mathcal{U}$ be any non-principal ultrafilter on I. Consider the element $p = (p_i)_i/\sim$ where p_i is the ith prime. Show that p is a prime element of $\mathbb{Z}^* = \prod_{i \in I} \mathbb{Z}/\mathcal{U}$ and is not equal to any standard prime (thinking of those as sitting inside the diagonal copy $\delta(\mathbb{Z})$ of $\mathbb{Z}$ in $\mathbb{Z}^*$).

In contrast, let b_i be the product of the first i primes. Show that $b = (b_i)_i/\sim$ has every standard prime as a factor - so this is an element which has infinitely many prime divisors. Show that b also has a prime divisor different from any standard prime.

You might wonder whether every element of $\mathbb{Z}^*$, apart from ± 1, must have at least one prime divisor. The, not so obvious, answer is "yes"; this will follow (*exercise*) directly from Łos' Theorem.

An informal statement of Łos' Theorem is as follows.

Theorem 2.9. *(Łos' Theorem v.1) If all the component structures (or even just a "large" set of them) have a certain property, then their ultraproduct also has that property.*

The small print is that the "property" must be "definable", that is, expressible in terms of the formal language $\mathcal{L}$, where all the component structures M_i are $\mathcal{L}$-structures. Of course, we need to be more precise in the formulation of Łos' Theorem, but let's proceed a little further with this in the specific examples we considered, since the properties clearly will be definable.

For example, this is why the ultraproduct of fields is a field, not just a ring. And also why our running example (the ultraproduct of the prime fields by a non-principal ultrafilter) produces an infinite field: because, given any number N, all but finitely many, hence a "large" set of, components have cardinality greater than N. By Łos' Theorem, the ultraproduct must have cardinality greater than N. That's true for every N, so the ultraproduct is infinite.

Finally, let's come back to infinitesimals. Recall that the requirement on an infinitesimal is that it should be a solution to the set of conditions:
$0 < x$ and, for every positive integer n, $nx < 1$.

In our construction we took our index set I to be the set of positive integers. For each $n \in I$, we took the structure M_n to be the reals $\mathbb{R}$. We called on Zorn's Lemma to get a non-principal ultrafilter $\mathcal{U}$ on I and formed the corresponding ultrapower, $\mathbb{R}^* = \mathbb{R}^I/\mathcal{U}$. We noted that this contains the diagonally-embedded copy of the reals. We then considered the element $\epsilon = \big((1/n)_n/\sim\big) \in \mathbb{R}^*$. Each component is > 0 so, from the way we define the ordering in the ultraproduct, $\epsilon > 0$. Also, given a positive integer n, for all but finitely many i, the ith component of ϵ is $< 1/n$; so, again by the definition of the ordering relation in the ultraproduct, $\epsilon < 1/n$. Before, we saw this by arguing directly but we can see that the fact that ϵ is an infinitesimal is an immediate consequence of Łos' Theorem.

2.3 Definable sets

We have seen that, in an ultraproduct, equations (like those expressing commutativity for the multiplication in a ring) and conditions built from equations (like having an inverse with respect to multiplication) somehow reflect those from the component structures. This is the key to making the precise statement of Łos' Theorem. So we will look at solution sets of equations - these are examples of definable sets. But we will also consider solution sets of more complicated conditions - conjunctions of equations, inequations, even projected (to some components) solution sets of equations. In fact, if we take a structure M and consider the collection of subsets of powers of M which are solution sets of equations (in any finite number of unknowns), and then close under the operations that we just mentioned (forming finite intersections, complements, and projections - but in general repeatedly, to get increasingly complicated sets), then we obtain the definable sets. Let's give a more formal definition.

Definition 2.10. *Let M be a structure. Let $x_1, \dots, x_n$ be variables ("unknowns") and let t_1, t_2 be two terms built up from these variables, using the algebraic operations and also allowing the constants, if there are any, to appear. We write $t(\overline{x})$ to emphasise the variables which may appear. We refer to the expression "$t_1 = t_2$" as an* **equation** *and we define its* **solution set** *to be* $\{\overline{a} \in M^n : t_1(\overline{a}) = t_2(\overline{a})\}$.

Example 2.11. Suppose that we take a field K for our structure (with the ring operations, $+$, $\times$ and constants 0, 1). Then a term built from variables $\overline{x} = (x_1, \dots, x_n)$ is essentially a polynomial, with integer coefficients, in those variables. So, if p, q are such terms/polynomials, then the solution set of the equation $p = q$ is the subset of K^n consisting of all $\overline{a} \in K^n$ where p and q take the same value. Another way of saying this is that the solution set is the zero-set of the polynomial $p - q$.[4]

Definition 2.12. *Suppose first that M is a purely algebraic structure (meaning the "structure" is given by operations and constants - that is, no relation symbols in the language $\mathcal{L}$). The* **definable subsets** *of (the various finite powers of) M are the sets obtained as follows:*
- *the solution set of every equation $t_1 = t_2$ between terms is a definable subset;*
- *the complement, $M^n \setminus D$ of any definable subset D of M^n is definable;*
- *the intersection of any two definable subsets of M^n is definable (therefore, in view of the previous clause, their union also is definable);*
- *if D is a definable subset of M^n and $i \in \{1, \dots, n\}$ then the image of D under projection along the ith axis, that is*

[4] You may know these already as subvarieties of affine space over K - so these are definable subsets of K or, to say it better, subsets of K^n definable in K. As I've said it, these would just be the subvarieties defined over - that is, zero-sets of polynomials with coefficients in - the prime field, $\mathbb{Q}$ or some $\mathbb{F}_p$. To get more general subvarieties, we should allow elements of K to appear as parameters in our formulas, so that those formulas can refer to polynomials with coefficients in K.

$\{(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) : \exists a \in M$ *with* $(a_1, \ldots, a_{i-1}, a, a_{i+1}, \ldots, a_n) \in D\}$, *is a definable subset of* M^{n-1}.

If the structure M *also has relations then we just add those in at the beginning, along with the solution sets of equations. This does make sense, since an* n*-ary relation on a set* M *is, formally, a subset of* M^n. *For instance, a partial order "<" is treated formally as a set of pairs - exactly those pairs* (a, b) *with* $a < b$. *So* $\{(a, b) \in M^2 : a < b\}$ *would be one of the basic definable subsets.*

Every definable set has a definition, namely it is the solution set $\varphi(M)$ of some formula φ of a language $\mathcal{L}$ such that M is an $\mathcal{L}$-structure. The atomic (=most basic) formulas are those of the form $t(\overline{x}) = t'(\overline{x})$ where t and t' are terms, together with those of the form $R(\overline{x})$ where R is some relation symbol of $\mathcal{L}$. The other formulas are built up using the boolean operations of negation and conjunction, together with the prefixing of existential quantifiers - these exactly correspond to the operations on definable sets of taking the complement, forming the intersection and projecting along a coordinate. (Recall that disjunction, implication and universal quantification can be defined using the other operations so, although we use them, we treat them formally as defined in terms of the others - this is useful when proving things by induction on complexity of formulas.)

Exercise: convince yourself, through a variety of examples, that if D is a definable subset of a structure M then there is a formula φ, with free = unquantified variables (among) $x_1, \ldots, x_n$, such that D is the solution set in M of φ. Part of the point of using formulas like φ is that they can be applied to any $\mathcal{L}$-structure, not just the structure we started with, and its meaning will be 'the same', in the sense that it expresses the same property (though, of course, its solutions will be very different).

So, from now on we will use formulas when we talk about definable sets in more than one structure. For instance, the centre of a group is defined by the formula $\forall x_2 \, (x_1 * x_2 = x_2 * x_1)$, where $*$ denotes the operation in the group. So the centre of any group G is a definable subset of G but the formula $\varphi(x_1)$ that we just wrote down can be read in any group H and its solution set, which we denote $\varphi(H)$, is exactly the centre of that group.

2.4 Łos' Theorem

Łos' Theorem is about what is true in an ultraproduct M^*; more precisely it tells us what the $\mathcal{L}$-theory of M^* is and what are its definable subsets. It says that an element $a = (a_i)_i/\sim$ of an ultraproduct belongs to a definable subset $\varphi(M^*)$ iff, for a "large" set of indices i, the component a_i belongs to the corresponding definable subset $\varphi(M_i)$ of M_i. It says, furthermore, that an ultraproduct has a property which can be expressed by a sentence σ of $\mathcal{L}$ iff "most" of its coordinate structures have that property (that is, satisfy σ). Here "large" and "most" mean with respect to the ultrafilter.

Theorem 2.13. *(Łos' Theorem v.2.1) Suppose that* $M^* = \prod_{i \in I} M_i/\mathcal{U}$ *is an ultraproduct. Suppose that* φ *is a formula (with free variables* $x_1, \ldots, x_n$*). Then* $\overline{a} = (a^1, \ldots, a^n)$ *is in the solution set,* $\varphi(M^*)$*, of* φ *in* M^* *iff* $\{i \in I : \overline{a_i} \in \varphi(M_i)\}$ *is in* $\mathcal{U}$*, where* $a^j = (a^j_i)_i/\sim$ *and* $\overline{a_i} = (a^1_i, \ldots, a^n_i)$.

Proof. So suppose that $M^* = \prod_{i \in I} M_i/\mathcal{U}$ is an ultraproduct.

The assertion is that, given a formula ψ,
$(*)$ for all $\overline{a}$, we have $\overline{a} \in \psi(M^*)$ iff $\{i \in I : \overline{a_i} \in \psi(M_i)\} \in \mathcal{U}$
where
$\overline{x} = (x_1, \ldots, x_n)$, $\overline{a} = (a^1, \ldots, a^n) \in (M^*)^n$, $a^j = (a_i^j)_i/\sim$ and $\overline{a_i} = (a_i^1, \ldots, a_i^n)$.
This is proved by induction on the complexity of ψ. This is "complexity" in the sense that atomic formulas are the least complex formulas (so the starting point of the induction) and then "complexity" is increased each time we apply a boolean operation ("and", "or", "not", "implies") or a quantifier ("there exists", or "for all"). Because "not", "and" and "there exists" are enough to define the others, we only need to consider those. So we have the base cases - where ψ is an equation or a relation - and three types of induction step. Here are the statements that, therefore, have to be proved.

If t and t' are terms built from variables $\overline{x} = (x_1, \ldots, x_n)$ (and perhaps constants) and $\overline{a} = (a^1, \ldots, a^n) \in (M^*)^n$ where $a^k = (a_i^k)_i/\sim$, then $t(\overline{a}) = t'(\overline{a})$ iff $\{i \in I : t(\overline{a_i^j}) = t'(\overline{a_i^j})\} \in \mathcal{U}$. This statement, in turn, has to be proved by induction on complexity of terms (how they are built up from the variables and constant symbols by successively applying function symbols). I will do some, maybe all, the details of this in class.

The other base case is that of a basic relation $R(x_1, \ldots x_n)$ and we need the statement that, with notation $\overline{a}$ etc. as above, $R(a^1, \ldots, a^n)$ holds in M^* iff $\{i \in I : R(\overline{a_i})$ holds in $M_i\} \in \mathcal{U}$. But this is how we defined the $\mathcal{L}$-structure on M^* so (assuming we already proved the well-definedness of this), there is nothing to do here.

That's the base case; the induction steps have the following (three) forms.

If ψ and ψ' are formulas and if each of these satisfies $(*)$, then so does the conjunction $\psi \wedge \psi'$ [the proof uses the closure of $\mathcal{U}$ under intersections].

If ψ is a formula which satisfies $(*)$, then so does the negation $\neg\psi$ (this proof of this uses that $\mathcal{U}$ is actually an *ultra*filter).

These two cases are very straightforward. I'll do the third here.

If ψ is a formula which satisfies $(*)$ and y is a variable then $\exists y\, \psi$ satisfies $(*)$ (it doesn't matter whether or not y actually appears in ψ, though it's rather pointless to stick $\exists y$ in front if y doesn't appear in ψ). Let's look at this one more closely (you might guess that this will use the "closed upwards" property of filters; let's see). Suppose then that $\overline{a} \in (\exists y\, \psi(\overline{x}, y))(M^*)$. Then there is $b \in M^*$ such that $(\overline{a}, b) \in \psi(M^*)$. So, by the induction hypothesis, $\{i \in I : (\overline{a}_i, b_i) \in \psi(M_i)\} \in \mathcal{U}$. Now, this set is certainly contained in $\{i \in I : \overline{a} \in (\exists y\, \psi(\overline{x}, y))(M_i)\}$, so this set is in $\mathcal{U}$, proving one direction of $(*)$ (and, indeed, using the upwards-closed property). We must check the other direction.

So suppose $K = \{i \in I : \overline{a} \in (\exists y\, \psi(\overline{x}, y))(M_i)\} \in \mathcal{U}$. For each index i in this set, choose a "witness" to the existential quantifier. That is, choose some $c_i \in M_i$ such that $(\overline{a}_i, c_i) \in \psi(M_i)$. Define the element $c \in M^*$ to be $(c_i)_i/\sim$ where, for $i \in I \setminus K$, you can choose c_i to be any element in M_i (since $I \setminus K$ is a "small" set, it doesn't matter what happens on any of those components). Then, $\{i \in I : (\overline{a}_i, c_i) \in \psi(M_i)\} \in \mathcal{U}$ and so, by the inductive hypothesis, $(\overline{a}, c) \in \psi(M^*)$. Therefore $\overline{a} \in (\exists y\, \psi(\overline{x}, y))(M^*)$, as required. □

Note that definable subsets of $\mathbb{R}$ include solution sets to equations and inequalities, so each of the requirements that characterise an infinitesimal can be expressed by a suitable formula - say in the language of ordered rings. And any finitely many of these requirements can be satisfied in $\mathbb{R}$ (that is, the intersection of the corresponding finitely many definable sets is nonempty). So the construction that we gave earlier is just a special case of Łos' Theorem.

Notice that, if $\mathcal{U}$ is a non-principal ultrafilter, hence every cofinite subset of the index set is in $\mathcal{U}$, an element $a = (a_i)_i/\sim$ of the ultraproduct will have a property which can be expressed by a formula $\varphi(x)$ if some $\sim_{\mathcal{U}}$-representative[5] of a has all but finitely many of its components satisfying that property. Although this condition is sufficient, it is by no means necessary for a to have the given property but I mention it explicitly because it often is used in particular cases.

We're not done with stating Łos' Theorem yet. The version above applies to properties of elements and n-tuples, but what about properties of the whole structure? For instance, the property that a commutative ring is a field. That particular property can be expressed by a formula, $\forall x\, (x \neq 0 \to \exists y\, (xy = 1))$, which has no free variables, that is, by a **sentence**. Sentences express properties of structures, rather than properties of elements. Here (in part (b)) is the version of Łos' Theorem that applies to them. Part (a) is a restatement of the previous version for the case of a single element rather than an n-tuple (less notation being necessary, perhaps the meaning is clearer).

Theorem 2.14. *(Łos' Theorem v.2.2) Suppose that $M^* = \prod_{i\in I} M_i/\mathcal{U}$ is an ultraproduct of $\mathcal{L}$-structures.*
(a) Suppose that $\varphi(x)$ is a formula of $\mathcal{L}$. Then $a = (a_i)_i/\sim$ is in the solution set, $\varphi(M^*)$, of φ in M^* iff $\{i \in I : a_i \in \varphi(M_i)\}$ is in $\mathcal{U}$.
(b) Suppose that σ is a sentence of $\mathcal{L}$. Then σ is true in M^* iff $\{i \in I : \sigma \text{ is true in } M_i\}$ is in $\mathcal{U}$.

Proof. Part (a) is the case $n = 1$ of the version above. But, for part (b) just notice that since a sentence is a special case of a formula, this is already covered by (a).
[If the argument for (b) seems unconvincing, then do the following *exercise*. Consider the case that σ has the form $\forall x\, \varphi(x)$ and suppose that $M^* \models \sigma$ (recall that $\models$ is the notation for the satisfaction/"true in" relation between structures and sentences/formulas with parameters). This means that the definable set $\varphi(M)$ is all of M. Assume the statement of part (a) for the formula φ and deduce that the set of indices i such that $M_i \models \forall x\, \varphi(x)$ is in $\mathcal{U}$. Be careful in your argument. You could also prove the converse, assuming $\{i \in I : M_i \models \sigma\} \in \mathcal{U}$, and deducing $M^* \models \sigma$.] $\square$

Here is the previous version written using the compact notation $\models$.

Theorem 2.15. *(Łos' Theorem v.2.2.5) Suppose that $M^* = \prod_{i\in I} M_i/\mathcal{U}$ is an ultraproduct of $\mathcal{L}$-structures.*
(a) Suppose that $\varphi(x)$ is a formula of $\mathcal{L}$ and let $a = (a_i)_i/\sim \in M^*$. Then $M^* \models \varphi(a)$ iff $\{i \in I : M_i \models \varphi(a_i)\} \in \mathcal{U}$.
(b) Suppose that σ is a sentence of $\mathcal{L}$. Then $M^* \models \sigma$ iff $\{i \in I : M_i \models \sigma\} \in \mathcal{U}$.

To emphasise: "formula" and "sentence" have very precise meanings here - they refer to formulas constructed from the basic algebraic relations, constants, and relations which give meaning to the phrase "type of structure" and where "constructed" means constructed using the boolean operations and quantifiers

[5] meaning an element $(a'_i)_i$ in the product such that $(a'_i)_i/\sim = a$

(the operations that we use when constructing definable sets). Of course, there has to be some kind of restriction: we know, for instance, that the condition "there are only finitely many elements" cannot be expressed by such a sentence, otherwise we could make an ultraproduct which would contradict Łos' Theorem. (On the other hand, saying "there are no more than N elements" is, for any particular natural number N, certainly expressible in any language - all we need is equality to say that.)

2.5 The Compactness Theorem

Now we derive the Compactness Theorem from Łos' Theorem.

Theorem 2.16. *(Compactness theorem, v.2a) Suppose that we have a set T of sentences in a language $\mathcal{L}$ appropriate for some specific type of structure. Suppose that, for every finite subset S of T, there is an $\mathcal{L}$-structure M_S which satisfies all the sentences in S. Then there is an ultraproduct of the M_S which satisfies all the sentences in T.*

Proof. We take the index set I to be the set of finite subsets S of T, with the structure being indexed by S being (some chosen) M_S. Any old ultrafilter won't do, so we first set up a filter (really a 'filter basis') by taking, for each $S \in I$, the subset $\langle S \rangle = \{S' \in I : S \subseteq S'\}$ of I. Note that if $S, S' \in I$ then $\langle S \rangle \cap \langle S' \rangle \supseteq \langle S \cup S' \rangle$.

It follows (see the *exercise* below) that the set $\mathcal{F} = \{J \subseteq I : J \supseteq \langle S \rangle$ for some $S \in I\}$ of subsets of I which contain some set of the form $\langle S \rangle$, is a filter. We then call on Zorn's Lemma to bring an ultrafilter $\mathcal{U}$ on I (necessarily non-principal) containing $\mathcal{F}$. Form the ultraproduct $M^* = \prod_{S \in I} M_S/\mathcal{U}$. I claim that this does the job.

So take any sentence $\sigma \in T$. Then $\{\sigma\} \in I$, so $\langle \{\sigma\} \rangle = \{S \in I : \sigma \in S\}$ is in the filter base, hence in $\mathcal{F}$, hence in $\mathcal{U}$. Note that if $S \in \langle \{\sigma\} \rangle$ then M_S satisfies σ. Therefore the set of indices S where the structure M_S satisfies σ is in $\mathcal{U}$ and hence, by Łos' Theorem, M^* satisfies σ. As required. $\square$

Exercise 2.17. If I is a(n index) set then a set $\mathcal{B}$ of subsets of I is a **filter basis** if the intersection of any finitely many members of $\mathcal{B}$ is nonempty - we say that $\mathcal{B}$ has the **finite intersection property** (**fip** for short). Show that $\mathcal{F}_{\mathcal{B}} = \{J \subseteq I : J \supseteq K_1 \cap \cdots \cap K_n$ for some $K_1, \ldots, K_n \in \mathcal{B}\}$ is a filter on I and is the smallest filter on I which contains every set in $\mathcal{B}$. It is called the filter **generated** by $\mathcal{B}$.

Theorem 2.18. *(Compactness theorem, v.2b) Suppose that M is an $\mathcal{L}$-structure and that Φ is a set of formulas of $\mathcal{L}$ with free variables (among) $\overline{x} = (x_1, \ldots, x_n)$. Suppose that, for every finite subset S of Φ, there is $\overline{a} \in M^n$ which satisfies all the formulas in S. Then there is an ultrapower M^* of M and $\overline{c} \in (M^*)^n$ which satisfies all the formulas in Φ.*

Proof. The proof is quite similar to that above. In fact, it can be made into a special case by introducing n new constant symbols to replace the variables $x_1, \ldots, x_n$, so that a formula $\varphi(\overline{x})$ can be replaced by a sentence in this slightly enriched language, thus replacing Φ by a set of sentences, which can then be fed into the version above. I'll go through the details in the lecture. $\square$

Definition 2.19. *A set T of sentences (of some language $\mathcal{L}$) is* **finitely satisfiable** *if every finite subset of T has a model; the set T is* **satisfiable** *if it has a model, that is, if there is an $\mathcal{L}$-structure $\mathcal{M}$ such that $\mathcal{M} \models \sigma$ for every $\sigma \in T$.*

Theorem 2.20. *(Compactness Theorem, v2.a.5) If a set of sentences of a language $\mathcal{L}$ is finitely satisfiable then it is satisfiable.*

Here is a corollary of the Compactness Theorem, stated more in the style of the next section, where we consider theories and their models.

Corollary 2.21. *Suppose that T is a set of sentences with arbitrarily large finite models. Then T has an infinite model*

Proof. Our assumption is that, for every n, there is a finite model of T with $\geq n$ elements. So consider the set $T \cup \{\sigma_{\geq n} : n \geq 1\}$ (where $\sigma_{\geq n}$ is a sentence saying there are at least n distinct elements). By our assumption this is finitely satisfiable, hence by Compactness it is satisfiable, meaning that there is an infinite model of T. $\square$

Exercise 2.22. Show that if a definable subset $\varphi(\mathcal{M}) \subseteq M$ of an $\mathcal{L}$-structure $\mathcal{M}$ is infinite then there is an elementary extension $\mathcal{M}' \succ \mathcal{M}$ of $\mathcal{M}$ which contains an element $a \in \varphi(\mathcal{M}') \setminus M$. Deduce that every infinite $\mathcal{L}$-structure has a proper elementary extension.

Exercise 2.23. Let p be a prime integer and let G be the multiplicative group of complex p^nth roots of unity (the union over all n). Note that every element of G is torsion. Show that there is an elementary extension of G which contains an element which is not torsion.

3 Theories and Models

Definition 3.1. *Suppose that $\mathcal{L}$ is some language of the kind that we have been considering. A* **theory** *in $\mathcal{L}$, or $\mathcal{L}$-***theory***, is a set, T, of sentences of $\mathcal{L}$. A* **model** *of an $\mathcal{L}$-theory T is an $\mathcal{L}$-structure $\mathcal{M}$ such that $\mathcal{M}$ satisfies every sentence in T: $\mathcal{M} \models \sigma$ for every $\sigma \in T$; we write $\mathcal{M} \models T$. If T is a set of sentences of $\mathcal{L}$, we let $\mathrm{Mod}(T)$ denote the collection of all models of T.*

Usually we include the requirement that, in order to be allowed as a theory, a set T of sentences must be **consistent**, in the sense that it has some model.

Definition 3.2. *The* **deductive closure** *of T is the set $\overrightarrow{T}$ of all sentences σ which are true in every model of T; we could write $\overrightarrow{T} = \mathrm{Th}(\mathrm{Mod}(T))$ (cf. 3.5 below). Note,* exercise*, that $\mathrm{Mod}(\overrightarrow{T}) = \mathrm{Mod}(T)$.*

Corollary 3.3. *(of the Compactness Theorem) If T is an $\mathcal{L}$-theory and σ is a sentence of $\mathcal{L}$ in the deductive closure of T, then there is a finite subset T' of T such that σ is in the deductive closure of T'.*

Proof. *Exercise* □

In practice we often blur the distinction between a theory and its deductive closure, so "the theory of groups" could mean just some choice of axioms for groups or it could mean everything that follows from those, that is, every sentence (in the chosen language) which is true in every group; officially, its the latter.

The terms "consistent" and "deductive closure" suggest the notion of formal deduction and you may know that the Completeness Theorem for Predicate Calculus implies that a theory is consistent iff there is no contradiction deducible from it and that $\overrightarrow{T}$ is the set of sentences of $\mathcal{L}$ formally deducible from T. We are not going to consider formal deductive systems in this course and, instead, we have defined these notions "semantically" - by reference to truth in models (the Completeness Theorem for Predicate Logic says that the two approaches give the same result).

Definition 3.4. *A theory T is* **complete** *if for every sentence $\sigma \in \mathcal{L}$ either $\sigma \in \overrightarrow{T}$ or $\neg\sigma \in \overrightarrow{T}$.*

Definition 3.5. *If $\mathcal{M}$ is an $\mathcal{L}$-structure then the (***complete***)* **theory of** $\mathcal{M}$ *is the set of all sentences of $\mathcal{L}$ which are true in $\mathcal{M}$:*

$$\mathrm{Th}(\mathcal{M}) = \{\sigma \in \mathcal{L} : \sigma \textit{ is a sentence and } \mathcal{M} \models \sigma\}.$$

Corollary 3.6. *If T is an $\mathcal{L}$-theory and σ is a sentence of $\mathcal{L}$ in the deductive closure of T, then there is a finite subset T' of T such that σ is in the deductive closure of T'.*

Proof. *Exercise* □

The theory of $\mathcal{M}$ is indeed complete because, if σ is any sentence then, either $\mathcal{M} \models \sigma$ (so $\sigma \in \mathrm{Th}(\mathcal{M})$) or, if not, that is if $\mathcal{M} \not\models \sigma$, then $\mathcal{M} \models \neg\sigma$ (so $\neg\sigma \in \mathrm{Th}(\mathcal{M})$).

Definition 3.7. *Two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{M}'$ are* **elementarily equivalent** *if they satisfy exactly the same sentences of $\mathcal{L}$: we then write $\mathcal{M} \equiv \mathcal{M}'$.*

It is immediate from the definitions that $\mathcal{M} \equiv \mathcal{M}'$ iff $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{M}')$. Recall (3.14) that if structures are isomorphic then they are elementarily equivalent.

Lemma 3.8. *The following are equivalent for a consistent theory T:*
(i) T is complete;
(ii) $\mathcal{M}, \mathcal{M}' \models T$ implies $\mathcal{M} \equiv \mathcal{M}'$;
(iii) $\overrightarrow{T} = \mathrm{Th}(\mathcal{M})$ for some $\mathcal{L}$-structure $\mathcal{M}$.

Proof. Direct from the definitions. □

A complete theory is the theory of a single structure; this is not to say that a complete theory has only one model!! In fact, most complete theories have lots of models, for example, ultrapowers of any given infinite model. But there is one exception, as follows.

Proposition 3.9. *If T is complete and has a finite model then T has just one model up to isomorphism.*

Proof. (Outline) Say $\mathcal{M} \models T$ has exactly n elements. Then $\sigma_{=n}$, a sentence saying there are exactly n elements, is in T. Therefore every model of T has exactly n elements. But we have to show that $T = \mathrm{Th}(\mathcal{M})$ specifies everything about $\mathcal{M}$, not just how many elements it has. We do this by taking any $\mathcal{N} \models T$ (so N must have n elements) and then we show that one of the $n!$ bijections between M and N must be an isomorphism - otherwise we produce a sentence true in one of $\mathcal{M}$, $\mathcal{N}$ but false in the other - contradicting completeness of T. □

Let T be a theory in some language $\mathcal{L}$ and suppose that we want to understand the models of T. First we may reduce to the case where T is complete.

Definition 3.10. *Say that T' is a* **completion** *of T if $T \subseteq T'$ (so any model of T' is a model of T) and if T' is complete.*

Note that if $M \models T$ then $\mathrm{Th}(M)$ is a completion of T. Therefore we can express $\mathrm{Mod}(T)$ as the disjoint (why?) union of the classes $\mathrm{Mod}(T')$ where T' runs over the distinct completions of T.

Exercise 3.11. Prove that if T is not complete then it has at least two different completions.

Exercise 3.12. (if you know the relevant cardinal arithmetic from Set Theory) Suppose that T is a theory in a countable language. Show that T has at most $2^{\aleph_0}$ completions.

Exercise 3.13. For each $\kappa \in \{1, 2, \ldots, n, \ldots \aleph_0, 2^{\aleph_0}\}$ give an example of a theory T_κ in a countable language which has exactly κ completions (in fact, these are the only possibilities for the number of completions of a theory in a countable language).

Understanding the class of models of a theory is one theme that runs through model theory. We will prove the Downwards Löwenheim-Skolem Theorem, which produces "small" models, for example saying that if the language $\mathcal{L}$ is countable then every infinite $\mathcal{L}$-structure contains a countable (necessarily infinite) elementary substructure. Later we prove the Upwards Löwenheim-Skolem Theorem, which says that any infinite model (of cardinality at least that of the language) has an elementary extension of any prescribed larger cardinality.

It is a consequence of those theorems that if $\mathcal{L}$ is countable then any $\mathcal{L}$-theory with an infinite model has at least one model of every infinite cardinality; in particular, we can't completely specify an infinite structure using first order logic. So the most we can ask of a theory T is that there be, for each infinite cardinality κ, just one model of cardinality κ up to isomorphism. That property of T is referred to as *categoricity*. A weakening of the condition of categoricity is that there be just one countably infinite model (up to isomorphism) - that is $\aleph_0$-categoricity and we will prove some nice characterisations of such theories. Alternatively we could ask for categoricity in higher, uncountable, cardinalities. Morley (in the 60's) proved a remarkable and extremely influential theorem in model theory that a theory in a countable language which is categorical in one uncountable cardinality is categorical in *all* uncountable cardinalities. We won't get to proving that but the upshot is that, for theories in countable languages, categoricity comes in three varieties: countable categoricity; uncountable categoricity; total categoricity (i.e. both the others).

3.1 Building around a set; the Downwards Löwenheim-Skolem theorem

The simplest example of "building around a set" is how we generate a substructure, $\langle A \rangle$, of an $\mathcal{L}$-structure $\mathcal{M}$ from any subset A of M. It is more difficult to build an elementary substructure of $\mathcal{M}$ containing A (and we shouldn't expect the uniqueness/minimality that we got for $\langle A \rangle$) but it can be done.

Theorem 3.14. *(Downwards Löwenheim-Skolem) Suppose that $\mathcal{M}$ is an infinite $\mathcal{L}$-structure, where $\mathcal{L}$ is a countable language, and that A is a countable subset of M. Then there is a countable elementary substructure $\mathcal{N}$ of $\mathcal{M}$ with $A \subseteq N$.*

Proof. We have to produce a subset of M which contains A and which is (the underlying set of) an elementary substructure of $\mathcal{M}$ (we'll worry about its countability later). Certainly any such subset must be a substructure, so we have to include, in addition to A, any interpretations of constant symbols and then "close under the ($\mathcal{L}-$)functions". Doing that would give us a substructure of $\mathcal{M}$ (we proved this in 10.2) but not necessarily an elementary substructure of $\mathcal{M}$. But Tarski's Lemma 10.17 says that a substructure is an elementary one if it contains "witnesses for existential quantifiers". So we list all formulas $\varphi(x)$ with parameters from $A \cup \{\text{constants}\}$ and then, for each φ with $\mathcal{M} \models \exists x \varphi(x)$ we choose some element $b \in M$ with $\mathcal{M} \models \varphi(b)$ (choose just one for each such φ). Add in all these "witnesses" b to get a new set $A \cup \{\text{witnesses}\}$. Now we realise that we have to repeat the process (because we have added new parameters). In fact we have to repeat the process ω-many times but no more (since a formula is a finite object). We end up with a subset of M which satisfies the criterion of Tarski's Lemma for being an elementary substructure of $\mathcal{M}$. Cardinal arithmetic and the countability assumptions we made ensure that this set is indeed countable (and necessarily infinite since it is elementarily equivalent to $\mathcal{M}$).

That was the recipe; let's make the proof now.

Set $A_0 = A$. For each formula $\varphi(x)$, with free variable x and parameters in A_0, such that $\mathcal{M} \models \exists x\, \varphi(x)$ choose some $b_\varphi \in M$ such that $\mathcal{M} \models \varphi(b_\varphi)$. Set $A_1 = A_0 \cup \{b_\varphi\}_\varphi$ (φ running over such formulas). Note that A_1 contains the interpretations in $\mathcal{M}$ of each constant symbol c (take φ to be the formula $x = c$) and, if f is an n-ary function symbol and $a_1, \ldots, a_n \in A_0$ then $f^{\mathcal{M}}(a_1, \ldots, a_n) \in A_1$ (consider the formula with parameters in A_0, $x = f(a_1, \ldots, a_n)$).

We repeat this process: having defined A_i, for every formula $\varphi(x)$, with free variable x and parameters in A_i, such that $\mathcal{M} \models \exists x\, \varphi(x)$ choose some $b_\varphi \in M$ such that $\mathcal{M} \models \varphi(b_\varphi)$. Then set $A_{i+1} = A_i \cup \{b_\varphi\}_\varphi$ (φ running over such formulas); in particular $A_i \subseteq A_{i+1}$.

Set $A_\omega = \bigcup_{i \geq 0} A_i$. We will show that we can take $N = A_\omega$ - that A_ω is the underlying set of an elementary substructure of $\mathcal{M}$. First we show it's the underlying set of a substructure.

We've already seen that A_1, hence A_ω contains all the $c^{\mathcal{M}}$ - the first requirement in 10.1 - so we check the second requirement of that lemma. Let f be an n-ary function symbol of $\mathcal{L}$ and $a_1, \ldots, a_n \in A_\omega$. For each i, $a_i \in A_{m(i)}$ for some $m(i)$. Let $m = \max\{m(1), \ldots, m(n)\}$ - so, since $A_i \subseteq A_j$ if $i \leq j$, each a_i is in A_m. Then, by construction (and as already argued above), $f^{\mathcal{M}}(a_1, \ldots, a_n) \in A_{m+1}$ and hence $f^{\mathcal{M}}(a_1, \ldots, a_n) \in A_\omega$. Therefore A_ω is indeed the underlying set of a substructure, we label it $\mathcal{N}$, of $\mathcal{M}$. To show that it is an elementary substructure, we check the criterion of 10.17.

So let $a_1, \ldots, a_n \in A_\omega$ and suppose that $\psi(y_1, \ldots, y_n, x)$ is such that $\mathcal{M} \models \exists x\, \psi(a_1, \ldots, a_n, x)$. Choose the $m(i)$ and m as we did in the previous paragraph. Then, by construction (with $\varphi(x)$ being $\psi(a_1, \ldots, a_n, x)$), there is $b_\varphi \in A_{m+1}$, hence $b_\varphi \in A_\omega$, such that $\mathcal{M} \models \varphi(b_\varphi)$, that is, $\mathcal{M} \models \psi(a_1, \ldots, a_n, b_\varphi)$. So, by 10.17, $\mathcal{N} \prec \mathcal{M}$.

Clearly $A \subseteq N = A_\omega$ so it remains to count the elements of A_ω. We claim that, for each i, A_i is countable; we prove this by induction. The base case is the assumption that $A_0 = A$ is countable. So suppose inductively that A_i is countable. There are only countably many formulas $\psi(y_1, \ldots, y_n, x)$ (since $\mathcal{L}$ is countable) and, for each of these, only countably many choices for replacing y_i by an element of A_i (since A_i is countable). Hence there are only countably many formulas with parameters from A_i, hence (since we add at most one witness for each), only countably many elements b_φ added to A_i to get A_{i+1}. Finally, a countable union (A_ω) of countable sets is countable. And we're done. $\square$

Corollary 3.15. *Suppose that $\mathcal{L}$ is a countable language and that $\mathcal{M}$ is an infinite $\mathcal{L}$-structure. Then $\mathcal{M}$ has a countable elementary substructure.*

Corollary 3.16. *Suppose that $\mathcal{L}$ is a countable language and T is a consistent $\mathcal{L}$-theory. Then T has a countable model.*

A little cardinal arithmetic

If X and Y are sets and there is a bijection from X to Y then we write $|X| = |Y|$ and say that X and Y have **the same cardinality** (or **the same size**). This is an equivalence relation (and you might note that this is just isomorphism as $\mathcal{L}_0$-structures).

You might already have seen the proofs that $|2\mathbb{N}| = |\mathbb{N}|$, that $|\mathbb{N}| = |\mathbb{Z}|$ and that $|\mathbb{N}| = |\mathbb{Q}|$ - all these sets are **countably infinite** but $|\mathbb{N}| \neq |\mathbb{R}|$ (by Cantor's Diagonal Argument).

Just as in the finite case, it is useful to abstract actual numbers from classes of sets all of the same size. So we set $\aleph_0$ to be the "number of elements in any countably infinite set - this is the smallest infinite number - actually we call these **cardinal numbers** and refer to the **cardinality** of a set.

We can order cardinal numbers: if $\kappa = |X|$ and $\lambda = |Y|$, then we set $\kappa \leq \lambda$ if there is an injection from X to Y. The Cantor-Schröder-Bernstein Theorem says that if there is an injection from X to Y and an injection from Y to X, then there is a bijection between X and Y. That is, if $\kappa \leq \lambda$ and $\lambda \leq \kappa$, then $\lambda = \kappa$. So we get an ordering on the collection of cardinal numbers.

A generalisation of Cantor's Diagonal Argument shows that for every cardinal κ, we have $2^\kappa > \kappa$ (so there is no greatest cardinal number).

Also, assuming the Axiom of Choice, one can prove that the cardinal numbers are totally ordered: if κ and λ are cardinal numbers then either $\kappa \leq \lambda$ or $\lambda \leq \kappa$. Indeed, the collection of cardinal numbers is **well-ordered**, meaning that every non-empty set of cardinal numbers has a least element; it follows that, given κ, there is a least cardinal strictly greater than κ - called the **successor** of κ and written κ^+. We set $\aleph_1 = \aleph_0^+$, $\aleph_2 = \aleph_1^+$, $\ldots$.

Cardinal arithmetic We define addition, multiplication and exponentiation of cardinals in terms of representative sets - of course one then has to show the definitions are independent of choice of representatives, but that's not hard.

Given cardinals κ, λ, choose sets X, Y with $|X| = \kappa$ and $|Y| = \lambda$. Define $\kappa + \lambda = |X \cup Y|$ provided X and Y are disjoint (which can be arranged by re-choosing one of them if necessary). Also define $\kappa \times \lambda = |X \times Y|$ (the cartesian product of X and Y) and $\kappa^\lambda = |X^Y|$ where by X^Y we mean the set of all functions from Y to X. (You should check that these do actually give the right numbers when κ and λ are finite.) Then we can prove the following arithmetic rules:

$\kappa + \lambda = \lambda + \kappa$, $\kappa \times \lambda = \lambda \times \kappa$, $\kappa \times (\lambda \times \mu) = (\kappa \times \lambda) \times \mu$, $\kappa \times (\lambda + \mu) = (\kappa \times \lambda) + (\kappa \times \mu)$, *et cetera*.

Also if κ is infinite then $\kappa + \kappa = \kappa$ and $\kappa \times \kappa = \kappa$.

And, more generally, if $\kappa \leq \lambda$ then $\kappa + \lambda = \lambda$ and $\kappa \times \lambda = \lambda$.
For example, the fact that $\aleph_0 \times \aleph_0 = \aleph_0$ implies that a countable union of countable sets is countable.