1 Ultrafilters, Ultrapowers, ${}^*\mathbb{R}$

Our goal is to formulate a set of numbers, the hyperreals, denoted ${}^*\mathbb{R}$. We want this set to include both the real numbers $\mathbb{R}$ *and* some number of infinitesimal elements, elements that are "infinitely close to" a given real number r but still not *equal* to r. To do this, we will take an *ultrapower* of $\mathbb{R}$. This ultrapower will be a quotient ring of $\mathbb{R}^{\mathbb{N}}$, the ring of sequences of real numbers, divided by an equivalence relation. With our equivalence relation we hope to capture the idea that two sequences have the same values "*almost* everywhere." And to do this, we will need an *ultrafilter* on $\mathbb{N}$.

1.1 The Ultrafilter

The idea is that we will define an *ultrafilter* $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$ that includes all of the "very large" subsets of $\mathbb{N}$, and then we will write $\langle r_n \rangle \equiv \langle s_n \rangle$ when $\{n \in \mathbb{N} \,|\, r_n = s_n\} \in \mathcal{F}$. Here are the criteria we will use:

Definition ([1, p. 18])**.** $\mathcal{F} \subseteq \mathbb{N}$ is a (non-principal) ultrafilter on $\mathbb{N}$ if:

- whenever $A, B \in \mathcal{F}$, we have $A \cap B \in \mathcal{F}$, and
- whenever $A \in \mathcal{F}$ and $A \subseteq B$, we have $B \in \mathcal{F}$, and
- for any $A \in \mathcal{P}(\mathbb{N})$, *either* A or $A^c = \mathcal{P}(\mathbb{N}) - A$ are in $\mathcal{F}$, and
- no finite set is in $\mathcal{F}$.

A subset of $\mathcal{P}(\mathbb{N})$ that satisfies the first two bullets is a *filter*. For instance, $\mathcal{P}(\mathbb{N})$ is a filter on $\mathbb{N}$, as is $\emptyset$. A *proper* filter is one that does not contain $\emptyset$ (as if $\emptyset \in \mathcal{F}$, then $\mathcal{P}(\mathbb{N}) = \mathcal{F}$ by the second bullet point). Stricly speaking, any proper filter that satisfies the third bullet point is an ultrafilter, but we will henceforth use "ultrafilter" to refer only to filters that satisfy all four bullets.

Call a set *cofinite* if its complement is finite. Every cofinite subset of $\mathbb{N}$ is in $\mathcal{F}$, by the second and third bullets.

If we think of $\mathcal{F}$ as being the subsets of $\mathbb{N}$ that include "almost all" of the natural numbers, then these make intuitive sense. If A and B both include "almost all" of the natural numbers, then surely $A \cap B$ does too—"basically no" elements are in $A - B$ or $B - A$ since "basically no" elements are outside A or B. If A includes "almost all" of $\mathbb{N}$, and $A \subseteq B$, then surely B includes "almost all" of $\mathbb{N}$ too. Etc.

Of course, there are some unintuitive things about $\mathcal{F}$: it contains either the set of even numbers $2\mathbb{N}$ or the set of odd numbers $2\mathbb{N} + 1$, but not both, by the third bullet point, for example.

Theorem 1.1 ([1, Corollary 2.6.2])**.** *There is at least one ultrafilter $\mathcal{F}$ on $\mathbb{N}$.*

Proof adapted from [1, pp. 20–21]. Let $\mathcal{F}^{co} \subset \mathcal{P}(\mathbb{N})$ denote the collection of cofinite subsets of $\mathbb{N}$. Let P denote the collection of all proper filters on $\mathbb{N}$ that include $\mathcal{F}^{co}$. Since $\mathcal{F}^{co}$ is itself a filter, $P \neq \emptyset$.

P is partially ordered by $\subseteq$: our approach is to apply Zorn's Lemma. Let $T \subset P$ be totally ordered by $\subseteq$. Then $\cup T$ is clearly an upper bound of T, but we need to show that $\cup T$ is a filter (this is [1, Example 2.4(4)], but is not proven). If $A, B \in \cup T$, then $A \in T_1$ and $B \in T_2$ for some $T_1, T_2 \in T$. Since T is totally ordered by $\subseteq$, either $T_1 \subseteq T_2$ or $T_2 \subseteq T_1$. If $T_1 \subseteq T_2$, then $A \in T_2$, and so since T_2 is a filter $A \cap B \in T_2$ and thus $A \cap B \in \cup T$. The proof is similar if $T_2 \subseteq T_1$. Similarly, if $A \in \cup T$ and $A \subseteq B$, then $A \in T_1 \in T$ and so $B \in T_1$ since T_1 is filter, and so $B \in \cup T$.

So, by Zorn's Lemma, P has a maximal element—call it $\mathcal{F}$. We want to show $\mathcal{F}$ is an ultrafilter (this is [1, Exercise 2.5(6)]). By the definition of P, $\mathcal{F}$ is a proper filter. Since $\mathcal{F}^{co} \subseteq \mathcal{F}$, if we can show that for every $A \subseteq \mathbb{N}$ either A or A^c is in $\mathcal{F}$ we will be done.

Take $A \subseteq \mathbb{N}$. We cannot have $A \in \mathcal{F}$ and $A^c \in \mathcal{F}$, for then we'd have $A \cap A^c = \emptyset \in \mathcal{F}$, which is impossible since $\mathcal{F}$ is proper. Now, assume for a contradiciton that $A \notin \mathcal{F}$ and $A^c \notin \mathcal{F}$. We will show that $\mathcal{F}$ can be extended by adding A or A^c, showing that $\mathcal{F}$ is not a maximal element of P and obtaining our contradiction.

Let $\mathcal{F}'$ be the filter obtained by adding to $\mathcal{F}$ the the intersection $A \cap B$ for any $B \in \mathcal{F}$, and any superset of $A \cap B$ (this technique adapted from [2, Appendix A]). Note that $A = A \cap \mathbb{N}$ and $\mathbb{N} \in \mathcal{F}$, so $A \in \mathcal{F}'$. We will show that $\mathcal{F}' \in P$. First, for any $B \in \mathcal{F}$, we have $B \nsubseteq A^c$ (as $A^c \notin \mathcal{F}$) and so $A \cap B \neq \emptyset$, showing $\mathcal{F}'$ is proper. Next, if $X, Y \in \mathcal{F}' - \mathcal{F}$, where $A \cap B \subseteq X$ and $A \cap C \subseteq Y$, then $(A \cap B) \cap (A \cap C) = A \cap (B \cap C) \subseteq X \cap Y$, with $B \cap C \in \mathcal{F}$, so $X \cap Y \in \mathcal{F}'$. A similar proof shows that for any $X \in \mathcal{F}$ and $Y \in \mathcal{F}' - \mathcal{F}$, $X \cap Y = Y \cap X \in \mathcal{F}'$. Similarly, if $X \in \mathcal{F}' - \mathcal{F}$ and $X \subseteq Y$, where $A \cap B \subseteq X$, then $A \cap B \subseteq Y$ and so $Y \in \mathcal{F}'$. So $\mathcal{F}'$ is a proper filter, hence $\mathcal{F} \subseteq \mathcal{F}' \in P$, violating our assumption that $\mathcal{F}$ is maximal in $(P, \subseteq)$. Hence, our assumption that $A \notin \mathcal{F}$ and $A^c \notin \mathcal{F}$ is incorrect. Since we know from earlier that we can't have $A, A^c \in \mathcal{F}$, we conclude that either A or A^c is in $\mathcal{F}$. Since this holds for any $A \subseteq \mathbb{N}$, we conclude $\mathcal{F}$ is an ultrafilter on $\mathbb{N}$. $\square$

Theorem 1.2 ([1, Exercise 2.5(4)]). *Let $\mathcal{F}$ be an ultrafilter and $\{A_1, \ldots, A_n\}$ a finite collection of pairwise disjoint sets such that*

$$A_1 \cup \cdots \cup A_n \in \mathcal{F}$$

Then $A_i \in \mathcal{F}$ for exactly one i such that $1 \leq i \leq n$.

Proof. At most one of the A_i's can be in $\mathcal{F}$, since if $A_i, A_j \in \mathcal{F}$ when $i \neq j$ then we'd have $A_i \cap A_j = \emptyset \in \mathcal{F}$, a contradiction.

Assume for a contradiction that $A_i \notin \mathcal{F}$ for each i. Then $A_i^c \in \mathcal{F}$ for each i, and so since $\mathcal{F}$ is closed under finite intersections we find

$$\bigcap_{i=1}^{n} A_i^c \in \mathcal{F}.$$

But then we find

$$\bigcup_{i=1}^{n} A_i = \left(\bigcap_{i=1}^{n} A_i^c \right)^c \notin \mathcal{F},$$

a contradiction. □

1.2 The Ultrapower

Let $\mathbb{R}^{\mathbb{N}}$ denote the set of sequences in $\mathbb{R}$. We will denote a member $r = \langle r_1, r_2, r_3, \ldots \rangle$ of $\mathbb{R}^{\mathbb{N}}$ by $\langle r_n \rangle$. We can define operations termwise addition $\oplus$ and termwise multiplication $\odot$ on $\mathbb{R}^{\mathbb{N}}$ by $\langle r_n \rangle \oplus \langle s_n \rangle = \langle r_n + s_n \rangle$ and $\langle r_n \rangle \odot \langle s_n \rangle = \langle r_n \cdot s_n \rangle$, giving us a commutative ring $\left(\mathbb{R}^{\mathbb{N}}, \oplus, \odot\right)$.

Now, let $\equiv$ denote the relation such that $\langle r_n \rangle \equiv \langle s_n \rangle$ iff $\{n \in \mathbb{N} \,|\, r_n = s_n\} \in \mathcal{F}$. We will write $[[r_n = s_n]]$ to denote $\{n \in \mathbb{N} \,|\, r_n = s_n\}$. When $\langle r_n \rangle = \langle s_n \rangle$, we will write that $r_n = s_n$ $\mathcal{F}$-almost everywhere.

Theorem 1.3 ([1, Exercise 3.3(1)]). $\equiv$ *is an equiavlence relation on* $\mathbb{R}^{\mathbb{N}}$

Proof. Clearly $\equiv$ is reflexive, since $[[r_n = r_n]] = \mathbb{N} \in \mathcal{F}$ so $\langle r_n \rangle \equiv \langle r_n \rangle$.

Similarly, $\equiv$ is symmetric because $[[r_n = s_n]] = [[s_n = r_n]]$, and so $[[r_n = s_n]] \in \mathcal{F}$ iff $[[s_n = r_n]] \in \mathcal{F}$, and so $\langle r_n \rangle \equiv \langle s_n \rangle$ iff $\langle s_n \rangle \equiv \langle r_n \rangle$.

Finally, $\equiv$ is transitive. Say $\langle r_n \rangle \equiv \langle s_n \rangle \equiv \langle t_n \rangle$. Then $[[r_n = s_n]] \in \mathcal{F}$ and $[[s_n = t_n]] \in \mathcal{F}$. Whenever $r_n = s_n$ and $s_n = t_n$, we have $r_n = t_n$, and so $[[r_n = s_n]] \cap [[s_n = t_n]] \subseteq [[r_n = t_n]]$. Since $\mathcal{F}$ is a filter, this implies $[[r_n = t_n]] \in \mathcal{F}$, and so $\langle r_n \rangle \equiv \langle s_n \rangle$. □

Now, we will form equivalence classes in $\mathbb{R}^{\mathbb{N}}$ based on this equivalence relation. Note that $[[r_n = s_n]] \in \mathcal{F}$ iff $[[r_n - s_n = 0]] \in \mathcal{F}$, and so if $I = \{\langle r_n \rangle \in \mathbb{R}^{\mathbb{N}} \,|\, [[r_n = 0]] \in \mathcal{F}\}$ then $\langle r_n \rangle \equiv \langle s_n \rangle$ iff $\langle r_n \rangle \ominus \langle s_n \rangle \in I$. I is an ideal, because it is closed under subtraction—$[[r_n - s_n = 0]] \supseteq [[r_n = 0]] \cap [[s_n = 0]]$, so if the latter two are in $\mathcal{F}$ so is the former—and for any $\langle r_n \rangle \in \mathbb{R}^{\mathbb{N}}$ and $\langle s_n \rangle \in I$, we have $[[r_n \cdot s_n = 0]] \supseteq [[s_n = 0]] \in \mathcal{F}$ and so $[[r_n \cdot s_n = 0]] \in \mathcal{F}$ and $\langle r_n \rangle \cdot \langle s_n \rangle \in I$. Then, we define the *hyperreals*

$${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/I.$$

We denote the equivalence class of an element $r = \langle r_n \rangle$ by $[r]$, $[\langle r_n \rangle]$, or just $[r_n]$ (ommitting the angled brackets).

1.3 Extensions

For any function $f : \mathbb{R} \to \mathbb{R}$, we define ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ by ${}^*f([r_n]) = [f(r_n)]$. For example, if $f(x) = x^2$, then $f([\langle 1, 2, 3, \ldots \rangle]) = [\langle 1, 4, 9, 16, \ldots \rangle]$. ${}^*f$ is well defined, as if $[r_n] = [s_n]$ then $[[r_n = s_n]] \in \mathcal{F}$, and so we have $[[r_n = s_n]] \subseteq [[f(r_n) = f(s_n)]] \in \mathcal{F}$ and so $\langle f(r_n) \rangle \equiv \langle f(s_n) \rangle$, i.e. ${}^*f([r_n]) = [f(r_n)] = [f(s_n)] = {}^*f([s_n])$.

Similarly, for any k-ary relation on the reals $R_k \subseteq \mathbb{R}^k$, we define ${}^*R_k \subseteq {}^*\mathbb{R}^k$ by setting ${}^*R_k([r_n^1], [r_n^2], \ldots, [r_n^k])$ if $[[R_k(r_n^1, r_n^2, \ldots, r_n^k)]] \in \mathcal{F}$. This is well defined because if $[s_n^i] = [r_n^i]$ for any $1 \leq i \leq k$, then $[[s_n^i = r_n^i]] \in \mathcal{F}$ for any i, and so we find

$$\left([[R_k(r_n^1, \ldots, r_n^k)]] \cap \bigcap_{i=1}^{k} [[r_n^i = s_n^i]] \right) \subseteq [[R_k(s_n^1, \ldots, s_n^k)]] \in \mathcal{F}$$

since the right side is a finite intersection.

1-ary relations are subsets, and for them we will use $x \in {}^*R$ in place of $R(x)$. For instance, $[r_n] \in {}^*\mathbb{N}$ iff $[[r_n \in \mathbb{N}]] \in \mathcal{F}$. The set ${}^*\mathbb{N}$ is the set of *hypernaturals*.

We will use symbols such as $<$, $\leq$, $+$, $\cdot$, etc. to refer to their own extensions in the hyperreals. For instance, $[r_n] \leq [s_n]$ iff $[[r_n \leq s_n]] \in \mathcal{F}$, and $[r_n] + [s_n] = [r_n + s_n]$.

A technique we will use frequently is choosing a representative of an equivalence class to meet certain conditions. Say, for instance, we know $[r_n] \in {}^*\mathbb{N}$. Then $[[r_n \in \mathbb{N}]] \in \mathcal{F}$. Now, let $s_n = r_n$ for all $n \in [[r_n \in \mathbb{N}]]$, and $s_n = 0$ elsewhere. Then $[[r_n \in \mathbb{N}]] \subseteq [[r_n = s_n]] \in \mathcal{F}$, so $[r_n] = [s_n]$, and we have $s_n \in \mathbb{N}$ for *all* n. In other words, whenver a condition is true $\mathcal{F}$-almost everywhere for every member of an equivalence class, we can (usually) pick a member where it is true *actually* everywhere. In proofs, we will say something like "$[r_n] \in {}^*\mathbb{N}$, and so we can assume $r_n \in \mathbb{N}$ for all n."

1.4 $\mathbb{R}$ in ${}^*\mathbb{R}$

For any real number $b \in \mathbb{R}$, we have $[b] = [\langle b, b, b, \dots \rangle] \in {}^*\mathbb{R}$. If $b, c \in \mathbb{R}$, then $[b] + [c] = [b + c]$, $[b] \leq [c]$ iff $b \leq c$, etc. In these cases we will usually drop the parenthesis and just write $b \in {}^*\mathbb{R}$. For instance, we might write $[r_n] < 5$ to indicate $[r_n] < [5]$.

1.5 Internal Sets & Functions

Not every subset of or function on ${}^*\mathbb{R}$ is represented by the extension of a subset of or function on $\mathbb{R}$.

Theorem 1.4 ([1, Theorem 3.9.1])**.** *Let $S \subseteq \mathbb{R}$. Then ${}^*S - \mathbb{R} \neq \emptyset$.*

Proof by Goldblatt. The theorem here is relying a bit on an abuse of notation, using $\mathbb{R}$ in the first case to mean the real numbers and in the second case to mean the "copy" of the real numbers in ${}^*\mathbb{R}$.

Let $s_1, s_2, \dots \in S$ be a pairwise distinct sequence of elements of S. Then clearly $[s_n] \in {}^*S$ as $[[s_n \in S]] = \mathbb{N} \in \mathcal{F}$. But for any $x \in \mathbb{R}$, we find $[[s_n = x]]$ is either $\emptyset$ or a singleton, since the s_n are pairwise distinct. So $[[s_n = x]] \notin \mathcal{F}$, i.e. $[s_n] \neq x$. □

Theorem 1.5 ([1, Exercise 3.10(1)])**.** *If A is finite, then ${}^*A = A$.*

Proof. Let $[r_n] \in {}^*A$. Then $[[r_n \in A]] \in \mathcal{F}$, and we have

$$\bigcup_{b \in A} [[r_n = b]] = [[r_n \in A]] \in \mathcal{F}.$$

Note that the $[[r_n = b]]$'s are pairwise disjoint. Then, by Theorem 1.2, we know that $[[r_n = b]] \in \mathcal{F}$ for *exactly one* b, and so we conclude $[r_n] = b$. So every element of ${}^*A$ is an element of A. Furthermore, for any $b \in A$, we have $[[b \in A]] = \mathbb{N} \in \mathcal{F}$, and so $b \in {}^*A$. So by double inclusion, ${}^*A = A$. □

So the extension of any finite subset of $\mathbb{R}$ is finite, and the extension of any infinite subset contains "nonstandard" elements. So if we consider $\mathbb{N}$ *as a subset of* ${}^*\mathbb{R}$, we find that it isn't the extension of anything: it's infinite, but it doesn't contain any nonreal elements.

So we can't use extensions to work with every possible subset of (or function on) ${}^*\mathbb{R}$. What we can do, though, is find "well-behaved" subsets or functions that we can extend our methods to. These subsets and functions are called *internal*.

Definition. If $A_1, A_2, A_3, \dots$ is a sequence of subsets of $\mathbb{R}$, we define the *internal subset* $A \subseteq {}^*\mathbb{R}$, denoted $A = [A_n]$, by

$$[r_n] \in [A_n] \text{ iff } [[r_n \in A_n]] \in \mathcal{F}.$$

Internal set membership is well-defined, as $[r_n] \in [A_n]$ implies $[[r_n \in A_n]] \in \mathcal{F}$, so if $[r_n] = [s_n]$ then $[[r_n \in A_n]] \cap [[r_n = s_n]] \subseteq [[s_n \in A_n]] \in \mathcal{F}$ and so $[s_n] \in [A_n]$. Any finite set of hyperreals is internal, as if $X = \{[r_n^1], \dots, [r_n^k]\}$ then $X = [\{r_n^1, \dots, r_n^k\}]$ [1, p. 126]. A finite set of hyperreals is only the extension of a set of reals if it is itself also a set of reals, so any finite set of hyperreals that contains non-real elements is internal but not the extension of a set of reals. Internal functions are defined similarly to internal sets:

Definition. If $f_1, f_2, f_3, \dots$ is a sequence of functions $f_i : \mathbb{R} \to \mathbb{R}$, we define the *internal function* $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$, denoted $f = [f_n]$, by

$$[f_n]([r_n]) = [f_n(r_n)].$$

2 First Order Logic & Transfer

2.1 The Idea

Here is the basic idea behind our project in this chapter. We will define a formal mathematical language, a language of *first-order logic*, that lets us talk about the reals. That language will use the logical symbols $\forall$ (for all), $\exists$ (there exists), $\wedge$ (and), $\vee$ (or), $\rightarrow$ (if-then), and $\neg$ (not), along with a host of symbols used to name the elements of, subsets of, and functions on $\mathbb{R}$. This language will not be nearly as powerful as English, but the upshot is we will show that any sentence we can write of this language is true in $\mathbb{R}$ if and only if an analogous sentence is true in ${}^*\mathbb{R}$. This is the *transfer principle*, and it is the engine that powers infinitesimal calculus. This notably not true of English: "every natural number is finite" is true in $\mathbb{R}$, but "every hypernatural number is finite" is *not* true in ${}^*\mathbb{R}$. Our method, once we have transfer, will be to write down some useful property of $\mathbb{R}$, transfer it to ${}^*\mathbb{R}$, prove that something is true about ${}^*\mathbb{R}$, and then transfer our result back to $\mathbb{R}$. In this way, we hope to prove things about $\mathbb{R}$ by going "through" ${}^*\mathbb{R}$.

2.2 Our language $\mathcal{L}_\mathbb{S}$

We will define our mathematical language $\mathcal{L}_\mathbb{S}$ on an arbitrary set S. In practice, we will mosty be using $\mathcal{L}_\mathfrak{R}$, our language on the real numbers $\mathbb{R}$, and $\mathcal{L}_{{}^*\mathfrak{R}}$, our language on the hyperreal numbers ${}^*\mathbb{R}$.

2.3 Symbols

Our language will, formally, contain the following symbols:

- The logical connectives $\wedge$, $\vee$, and $\neg$.
- Parentheses (and).
- The universal quantifier symbol $\forall$.
- The set membership symbol $\in$.
- A constant symbol $\dot{x}$ for each $x \in S$.
- A countable number of variables $v_1, v_2, v_3, \ldots$.
- A relation symbol $\dot{R}_n$ for any n-ary relation R_n on S, that is, any $R_n \subseteq S^n$. A 1-ary relation symbol is also called a *set symbol.*
- A function symbol $\dot{f}_n$ for any n-place function f_n on S, that is, any $f_n : S^n \rightarrow S$.

2.4 Terms

The *terms* of our language will be the symbols or sequences that refer to an element of S. They are defined recursively:

- Any single constant symbol $\dot{x}$ is a term.
- Any single variable v_n is a term.
- If $\dot{f}_n$ is an n-place function symbol, and $t_1, t_2, \ldots, t_n$ are terms, then $\dot{f}_n(t_1, t_2, \ldots, t_n)$ is a term.

A *closed* term is a term that contains no variables—$\dot{f}_1(v_1)$ is a term, and $\dot{f}_1(\dot{5})$ is a closed term. The former refers to some element of S, the same way that in common mathematical parlance x might refer to an element of $\mathbb{R}$, but the latter refers to a *specific* moment of S, the same way π refers to a *specific* element of $\mathbb{R}$.

2.5 Well-Formed Formulae

Strictly speaking, a *formula* is just a sequence of symbols. Here's a formula: $\dot{f}_5(\wedge, \forall, \dot{17}) \rightarrow\rightarrow \wedge(()(($. That formula doesn't mean anything. The formulae that *do* mean something are called *well-formed formulae*, or wffs (sometimes pronounced 'wiffs') for short. The wffs of our language are defined recursively, as follows:

- If $\dot{R}_n$ is an n-ary relation symbol, and $t_1, \ldots, t_n$ are terms, then $\dot{R}_n(t_1, \ldots, t_n)$ is a wff. This type of wff is called an *atomic formula.*

- If φ and ψ are wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $\neg\varphi$.
- If φ is a wff and v_n is a variable, then $(\forall v_n)\varphi$ is a wff.

Examples of wffs include "$\dot{f}(\dot{5}) = \dot{7}$", "$\dot{\leq}(\dot{2}, \dot{3})$", and "$\dot{\geq}(v_1, \dot{+}(\dot{2}, \dot{v}_{54}))$". Note here that we are writing $\dot{\leq}(a, b)$ in place of $a\dot{\leq}b$ because our language doesn't allow for infix notation.

2.6 Bound and Free Variables

Consider the wff $(\forall v_1)P(v_1, v_2)$. Is this *true*? Well we can't rightly say—we don't know what v_2 is! We need to know what v_2 means before we determine whether it's true. We *don't* need any more information what v_1 is, because we know it's a stand-in for "any element" due to the quantifier $\forall v_1$. Variables like v_1 we call "bound" (by the quantifier), and variables like v_2 we call *free*.

Formally, an *occurance* of a variable v_n that occurs in a wff φ is called boundif it is in the scope of a quantifier $\forall v_n$, and free otherwise. As another example, in the wff $(P(v_1) \vee (\forall v_1)(Q(v_1) \wedge Q(v_2)))$, the first occurance of v_1 is free, while the second is bound. The occurance of v_2 is also free, since it's not in the scope of $(\forall v_2)$. A wff with no free variables is called a *sentence*—and since a sentence has no free variables, then we can determine whether it is true false, as will be discussed in subsection 2.9.

2.7 Replacement Notation

If φ is a wff that is not a sentence, we write $\varphi(v_1, v_2, \ldots, v_n)$ to indicate that $v_1, \ldots, v_n$ are all the variables that occur freely in φ. Then $\varphi(\dot{x}_1, \ldots, \dot{x}_n)$ is the sentence we get by replacing each free occurance of v_i by the constant symbol $\dot{x}_i$. Similarly, if t is a term, we write $t(v_1, \ldots, v_n)$ to indicate t and $t(\dot{x}_1, \ldots, \dot{x}_n)$ to indicate the term you get by replacing each occurance of v_i with $\dot{x}_i$.

Similarly, we write $\varphi(v_1, \ldots, v_n, \dot{A}_1, \ldots, \dot{A}_m)$ to indicate that $v_1, \ldots, v_n$ are all the variables that occur freely in φ, and $A_1, \ldots, A_m$ are all the set symbols that occur in φ. We then write $\varphi(\dot{x}_1, \ldots, \dot{x}_n, \dot{B}_1, \ldots, \dot{B}_m)$ to mean the sentence we get by replacing v_i by $\dot{x}_i$ and $\dot{A}_i$ by $\dot{B}_i$.

So if $\varphi(v_1, v_2, \dot{A}_1) = \dot{P}(v_1) \wedge (\forall v_1 \in A_1)(v_1 < v_2)$, then $\varphi(\dot{x}, \dot{y}, \dot{T}) = \dot{P}(\dot{x}) \wedge (\forall v_1 \in T)(v_1 < \dot{y})$.

2.8 An Aside: Abbreviations

While we will treat this language, rigorously defined, exactly as it is for the purposes of proofs and the like, when actually writing down sentence of $\mathcal{L}_S$ we will make use of a number of handy abbreviations and shortcuts:

- We will omit outer parentheses and other parentheses not necessary to understand a wff, as in $\dot{P}(v_1) \vee \dot{Q}(v_1)$. We will also sometimes use [and] in place of (and), respectively.
- We will write $x \in \dot{A}$ to mean $\dot{A}(x)$.
- We will write $\varphi \to \psi$ to mean $\neg\varphi \vee \psi$. Intuitively, this *means* the same thing: if φ implies ψ, then either φ is false or φ, and thus ψ, are true. Similarly, if $\neg\varphi \vee \psi$, and φ is true, then ψ must be true.
- We will write $(\exists v_i)\varphi$ to mean $\neg(\forall v_i)\neg\varphi$. An example for why this makes sense: there exists a number greater than 5 iff it's *not* true that *every* number is no greater than 5. If it's not the case that $\neg\varphi$ for every v_i, then there must exist some counterexample for which φ.
- We will write $(\forall v_i \in \dot{A})\varphi$ to mean $(\forall x)(x \in \dot{A} \to \varphi)$, and similarly for $\exists$.
- We will use infix notation for common relations and functions that usually do so, e.g. $\leq$, $\geq$, $=$, $+$, $\cdot$, etc.
- We will usually use a symbol other than v_i as a variable, for instance in $(\forall r \in \mathbb{N})(0\dot{\leq}r)$.
- We will frequently omit the dots from constant, function, relation, and set symbols, e.g. using x to denote the constant symbol associated with $x \in S$.

Here's an example of a sentence of $\mathcal{L}_S$ expressing the *Archimedian property* of $\mathbb{R}$, that for any real number r there is a natural number N such that r is less than N.

$$(\forall r)(\exists N \in \mathbb{N})(r < N)$$

This can be read in English as follows.

$$\underbrace{(\forall r)}_{\text{for any real number } r,} \quad \underbrace{(\exists N \in \mathbb{N})}_{\text{there is a natural number } N} \quad \underbrace{r < N}_{\text{(such that } r \text{ is less than } N)}$$

Without abbreviations, this reads

$$(\forall v_1)\neg(\forall v_2)\neg(\dot{\mathbb{N}}(v_2) \wedge \dot{<}(v_1, v_2)),$$

which means the same thing but is much harder to parse.

2.9 Truth in $\mathcal{L}_{\mathbb{S}}$

Intuitively, it's very easy to see how a sentence might be true or false. A sentence is true just in case it accurately describes the things it is referring to. In order to prove transfer, though, we need a rigorous *mathematical* description of what makes a senetence true or false.

Here's our basic approach. We'll define an s function to be an assignment of all the variables $v_1, v_2, \ldots$ to elements of S. We'll then recursively construct a function $\bar{s}$ that takes any term as its input and outputs what that term corresponds to in S. So if $s(v_1) = 2$ and $s(v_2) = 3$, then $\bar{s}(\text{“}v_1 \dot{+} v_2 \dot{+} \dot{4}\text{”}) = 9$. We'll then say that an s function *satisfies* a wff if that wff correctly describes S when you interpret the terms using $\bar{s}$. So if $s_1(v_1) = 2$ and $s_2(v_1) = 100$, then s_1 satisfies the wff "$v_1 < \dot{50}$" but s_2 does not. A sentence, since it doesn't have any free variables for s to assign, is either satisfied by *every* s or by *no* s, in which case we say it is true or false, respectively.

Now, let's do this rigorously. Let $s : V \to S$, where V is the set of variables. Such a function is called an s function. We then recursively define $\bar{s} : T \to S$, where T is the set of terms, by:

- $\bar{s}(v_n) = s(v_n)$ for any variable v_n.
- $\bar{s}(\dot{x}) = x$ for any constant symbol $\dot{x}$.
- $\bar{s}(\dot{f}_n(t_1, \ldots, t_n)) = f_n(\bar{s}(t_1), \ldots, \bar{s}(t_n))$.

So if $s(v_1) = 3$ and $f(x) = x^2$, then $\bar{s}(\text{“}\dot{f}(v_1 \dot{+} \dot{2})\text{”}) = 25$. Note that $\bar{s}$ outputs 25 the number, not "$\dot{25}$" the symbol. We say that $s : V \to S$ either satisfies or doesn't satisfy any wff φ, defined recursively by:

- s satisfies an atomic formula $\dot{R}_n(t_1, \ldots, t_n)$ if $(\bar{s}(t_1), \ldots, \bar{s}(t_n)) \in R_n$.
- s satisfies $\neg\varphi$ if it doesn't satisfy φ.
- s satisfies $\varphi \vee \psi$ if it satisfies φ or ψ.
- s satisfies $\varphi \wedge \psi$ if it satisfies φ and ψ.
- s satisfies $(\forall v_n)\varphi$ if every $t : V \to S$ such that $t(v_i) = s(v_i)$ for all $i \neq n$ satisfies φ.

So in our example above where $s(v_1) = 3$, s satisfies the wff "$\dot{f}(v_1 \dot{+} \dot{2}) \dot{\geq} \dot{20}$", but if $s'(v_1) = -2$ then s' does not satisfy that wff.

That last bullet deserves some attention. What we're saying is that s satisfies $(\forall v_n)\varphi$ if you can change $s(v_n)$ to be whatever you'd like in s and still satisfy φ no matter what. In other words, s satisfies $(\forall v_n)\varphi$ if for every $x \in S$, you can send v_n to x and still have s satisfy φ.

If φ is a sentence, it is either satisfied by every $s : V \to S$ or not satisfied by every s. Intuitively, this is because every variable v_i in φ is bound by a quantifier, and so the statisfaction or not of φ will depend on whether it is satisfied when v_i takes *any* value—the "original" value of $s(v_i)$ won't matter. We'll skip the details of the proof, which is by induciton on wffs. If a sentence φ is satisfied by every s, we say that it is *true*, and otherwise we say it is *false*.

2.10 The Transfer Principle

Let $\mathcal{L}_{\mathfrak{R}}$ denote our language as defined on $\mathbb{R}$, with a constant symbol for every real number, a relation symbol for every k-ary relation in $\mathbb{R}^k$, and a function symbol for every k-place function $f : \mathbb{R}^k \to \mathbb{R}$. Let $\mathcal{L}_{^*\mathfrak{R}}$ likewise indicate our language as defined on $^*\mathbb{R}$.

Our goal is to show that for any sentence φ of $\mathcal{L}_{\mathfrak{R}}$, the "corresponding" sentence of $\mathcal{L}_{^*\mathfrak{R}}$ has the same truth value. First, we need to define the idea of "corresponding" sentence.

Let $*$ be a function from the set of terms in $\mathcal{L}_{\mathfrak{R}}$ to the set of terms in $\mathcal{L}_{^*\mathfrak{R}}$—the hope being that $*(t)$ will be the term "corresponding" to t. We will denote $*(t)$ by $^*t$. We define $*$ recursively on terms by:

- If v_n is a variable, ${}^*v_n = v_n$.
- If x is a constant symbol, ${}^*x = [(x, x, x, \ldots)]$.
- ${}^*(f_k(t_1, t_2, \ldots, t_k)) = {}^*f_k({}^*t_k, {}^*t_k, \ldots, {}^*t_k)$, where ${}^*f_k$ is the extension of f_k.

Now, for any sentence φ of $\mathcal{L}_{\mathfrak{R}}$, the corresponding sentence ${}^*\varphi$ of $\mathcal{L}_{{}^*\mathfrak{R}}$ is obtained by replacing every term t in φ by ${}^*t$ and every relation symbol R_k in φ by ${}^*R_k$ (the symbol corresponding to the extension of the relation R_k). For instance, if $\varphi = (\forall r \in \mathbb{R})(\exists N \in \mathbb{N})(f(r) < N)$, then ${}^*\varphi = (\forall r \in {}^*\mathbb{R})(\exists N \in {}^*\mathbb{N})({}^*f(r) < N)$.

We can now state and prove the theorem that will imply the transfer principle.

Theorem 2.1 (Łoś's Theorem). *Let $\varphi(v_1, \ldots, v_k, A_1, \ldots, A_m)$ be a wff of $\mathcal{L}_{\mathfrak{R}}$. Then*

$$ {}^*\varphi([x_n^1], [x_n^2], \ldots, [x_n^k], [A_n^1], \ldots, [A_n^m]) \iff [[\varphi(x_n^1, \ldots, x_n^k, A_n^1, \ldots, A_n^m)]] \in \mathcal{F} $$

Where $[[P(n)]] = \{n \in \mathbb{N} \mid P(n)\}$.

Proof heavily adapated from [2, Appendix B]. Every closed term in ${}^*\varphi([x_n^1], [x_n^2], \ldots, [x_n^k], [A_n^1], \ldots, [A_n^m])$ takes the form ${}^*t([x_n^1], \ldots, [x_n^k])$ for some term $t(v_{i_1}, \ldots, v_{i_k})$ in φ. We will show by induction on terms that ${}^*t([x_n^1], \ldots, [x_n^k]) = [t(x_n^1, x_n^2, \ldots, x_n^k)]$.

- If t is a constant term x, then ${}^*t = x = [x, x, \ldots]$.
- If t is a variable v_i, then $t(c) = c$ for any constant symbol c, and so ${}^*t([x_n^1]) = [x_n^1] = [t(x_n^1)]$.
- If t is of the form $f_j(t_1, \ldots, t_j)$, then ${}^*t([x_n^1], \ldots, [x_n^k]) = {}^*f_j({}^*t_1([x_n^1], \ldots, [x_n^k]), \ldots, {}^*t_j([x_n^1], \ldots, [x_n^k])) = {}^*f_j([t_1(x_n^1, \ldots, x_n^k)], \ldots, [t_j(x_n^1, \ldots, x_n^k)]) = [f_j(t_1(x_n^1, \ldots, x_n^k), \ldots, t_j(x_n^1, \ldots, x_n^k))] = [t(x_n^1, \ldots, x_n^k)]$.

We will simplify this by writing ${}^*t = [t_n]$ for any closed term ${}^*t$ of ${}^*\varphi$, ommitting the plugging in of constant symbols. Similarly, we will write $\varphi_n = \varphi(x_n^1, \ldots, x_n^k, A_n^1, \ldots, A_n^m)$ and ${}^*\varphi = {}^*\varphi([x_n^1], [x_n^2], \ldots, [x_n^k], [A_n^1], \ldots, [A_n^m])$. Now we proceed to prove the theorem, that ${}^*\varphi$ iff $[[\varphi_n]] \in \mathcal{F}$, by induction on wffs.

- In the case of atmoic formulas involving internal sets, ${}^*t \in [A_n]$ iff $[t_n] \in [A_n]$ iff $[[t_n \in A_n]] \in \mathcal{F}$ by the definition of $[A_n]$.
- In the case of other atomic formulas, ${}^*R_k({}^*t^1, \ldots, {}^*t^k)$ iff $({}^*t^1, \ldots, {}^*t^k) \in {}^*R_k$ iff $([t_n^1], \ldots, [t_n^k]) \in {}^*R_k$ (by our result about ${}^*t$) iff $[[(t_n^1, \ldots, t_n^k) \in R_k]] \in \mathcal{F}$ iff $[[R_k(t_n^1, \ldots, t_n^k)]] \in \mathcal{F}$.
- If φ_n is of the form $\psi_n^1 \wedge \psi_n^2$, then ${}^*\varphi = {}^*\psi^1 \wedge {}^*\psi^2$. Then ${}^*\varphi$ iff ${}^*\psi^1 \wedge {}^*\psi^2$ iff $([[\psi_n^1]] \in \mathcal{F}$ and $[[\psi_n^2]] \in \mathcal{F})$ iff $[[\psi_n^1]] \cap [[\psi_n^2]] \in \mathcal{F}$ iff $[[\psi_n^1 \wedge \psi_n^2]] = [[\varphi_n]] \in \mathcal{F}$.
- Similarly, if φ_n is of the form $\psi_n^1 \vee \psi_n^2$, then ${}^*\varphi = {}^*\psi^1 \vee {}^*\psi^2$. Hence ${}^*\varphi$ iff ${}^*\psi^1 \vee {}^*\psi^2$ iff $([[\psi_n^1]] \in \mathcal{F}$ or $[[\psi_n^2]] \in \mathcal{F})$ iff $[[\psi_n^1]] \cup [[\psi_n^2]] \in \mathcal{F}$ iff $[[\psi_n^1 \vee \psi_n^2]] = [[\varphi_n]] \in \mathcal{F}$. Note $[[\psi_n^1]] \cup [[\psi_n^2]] \in \mathcal{F}$ implies $[[\psi_n^1]]$ or $[[\psi_n^2]] \in \mathcal{F}$ because, in general, if $A \cup B \in \mathcal{F}$ and $A \notin \mathcal{F}$ then $A^c \in \mathcal{F}$ so $A^c \cap (A \cup B) \subseteq B \in \mathcal{F}$.
- If $\varphi_n = \neg\psi_n$ then ${}^*\varphi = \neg\, {}^*\psi$. So ${}^*\varphi$ iff $\neg\, {}^*\psi$ iff $[[\psi_n]] \notin \mathcal{F}$ iff $[[\psi_n]]^c = [[\varphi_n]] \in \mathcal{F}$.
- If $\varphi_n = (\forall v_i)\psi_n(v_i)$, then ${}^*\varphi = (\forall v_i)\, {}^*\psi(v_i)$. Then ${}^*\varphi$ iff $(\forall v_i)\, {}^*\psi(v_i)$ iff ${}^*\psi([x_n])$ for every $[x_n]$ iff $[[\psi_n(x_n)]] \in \mathcal{F}$ for every sequence $\langle x_n \rangle$ (by the inductive hypothesis). We want to show that this is true iff $[[\varphi_n]] = [[(\forall v_i)\psi_n(v_i)]] \in \mathcal{F}$. First, if $[[(\forall v_i)\psi_n(v_i)]] \in \mathcal{F}$ then for any sequence $\langle x_n \rangle$ we have $[[(\forall v_i)\psi_n(v_i)]] \subseteq [[\psi_n(x_n)]]$ and so $[[\psi_n(x_n)]] \in \mathcal{F}$. Conversely, if $[[(\forall v_i)\psi_n(v_i)]] \notin \mathcal{F}$, then for each $n \notin [[(\forall v_i)\psi_n(v_i)]]$ we can find x_n such that $\neg\psi_n(x_n)$. If we make a sequence out of these x_n, letting x_n take an arbitrary value for $n \in [[(\forall v_i)\psi_n(v_i)]]$, then for any $n \notin [[(\forall v_i)\psi_n(v_i)]]$ we have $\neg\psi_n(x_n)$. So, we get $[[\psi_n(x_n)]] = [[(\forall v_i)\psi_n(v_i)]]$ and so $[[\psi_n(x_n)]] \notin \mathcal{F}$ for some sequence $\langle x_n \rangle$. So $[[\psi_n(x_n)]] \in \mathcal{F}$ for every sequence $\langle x_n \rangle$ iff $[[(\forall v_i)\psi_n(v_i)]] \in \mathcal{F}$. We conclude ${}^*\varphi = (\forall v_i)\, {}^*\psi(v_i)$ iff $[[(\forall v_i)\psi_n(v_i)]] = [[\varphi_n]] \in \mathcal{F}$.

□

Corollary 2.2 (Transfer Principle). *If φ is a sentence of $\mathcal{L}_{\mathfrak{R}}$, then*

$$ \varphi \iff {}^*\varphi. $$

Łoś's theorem is, in actuality, quite a bit more general than 2.1, applying to any ultrapower rather than just ${}^*\mathbb{R}$.

References

[1] Robert Goldblatt. *Lectures on the Hyperreals*. Vol. 188. Graduate Texts in Mathematics. New York, NY: Springer, 1998. ISBN: 978-1-4612-6841-3 978-1-4612-0615-6. DOI: 10.1007/978-1-4612-0615-6.

[2] James M. Henle and Eugene M. Kleinberg. *Infinitesimal Calculus*. Cambridge, Mass: MIT Press, 1979. 135 pp. ISBN: 978-0-262-08097-2.