Nonstandard Analysis

Jeremy Wu

May 2018

Contents

Abstract

Infinitesimals are an attractive mathematical fiction particularly in the beginnings of calculus as Leibniz developed it [1]. Use of this class of numbers helped motivate the main concepts of calculus. However, the rigorous treatment of these ideal numbers was postponed until 1960 when Robinson developed the framework for a more general theory which we now call Nonstandard Analysis [2]. In this essay, we give a brief introduction to Nonstandard Analysis, presenting the logical framework up to and including the celebrated transfer principle. Informally, this result states that true statements in the usual reals are also true when extended to the hyperreals.

Chapter 1

Introduction

1.1 Motivation

To motivate the need for the concept of filters (and, more importantly, ultrafilters) in chapter 2.1, let us discuss what an 'infinitesimal' should be. That is, we want a number ϵ which satisfies the following properties:

- $\epsilon > 0$
- $\forall n \in \mathbb{N}, \epsilon < \frac{1}{n}$

Obviously, looking for such an ϵ in the usual real numbers will not be successful. We will instead search for such an ϵ in a space of real sequences, which we will denote by angular brackets $\langle \cdot \rangle$. Consider the following sequence:

$$\omega = \left\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots \right\rangle$$

In chapter 2.1, we will actually rigorously adopt ω as a representative infinitesimal. Observe that, for any given $n \in \mathbb{N}$, we can always find an index i such that every term of ω indexed beyond i is less than $\frac{1}{n}$. As well, by definition of ω, it is clear that ω is not identically the 0 sequence. This establishes the existence of a 'large' set of indices I such that $\omega_i < \frac{1}{n}$ and $\omega_i \neq 0$ when $i \in I$. In this index set, the two desired properties of an infinitesimal number hold. This is the key intuition for filters (and ultrafilters). We will construct the hyperreal numbers as (equivalences classes of) sequences of real numbers. Concepts like infinitesimality and infiniteness (both of which will become clear later on) should be interpreted in the sense that they hold on a 'sufficiently large' index set. Our definition of filters (and utlrafilters) will assign rigorous meaning to what we mean by 'large' sets.

We will closely follow the spectacular book by Hurd and Loeb [1]. This essay is a strict subset of chapter 1 of Hurd and Loeb with some stylisitc adjustments based on this author's personal preference. In particular, this essay attempts to serve a more active role in the explanation of results. One major difference is the ordering of topics. In this essay, the proof of the transfer principle in section 3.3 is given directly after its statement. Of course, we would be remiss in not pointing out the first treatment of nonstandard analysis by Abraham Robinson [2].

1.2 Outline

Chapter 2 is dedicated to developing the framework for working with the hyperreal numbers. We start this chapter with a brief but important presentation on filters in section 2.1. In the next section 2.2, we apply some of the abstract theory of filters to formally construct and prove some basic properties of the hyperreals. This section also covers a first definition 2.9 of the *-transform which is vital to the transfer principle. The last section 2.3 provides some basic properties of the *-transform as well as its extension to relations on $\mathbb{R}$. Every effort has been made to balance the abstract notions in this chapter with the grounded application of the hyperreal construction so that familiarity is developed before turning towards a more general framework in the next chapter.

Chapter 3 focuses on reformulating the earlier topics of chapter 2 in a more abstract and general setting. While this is presented in much more generality than the setting of the (hyper)reals, we give examples exclusively in $\mathbb{R}$ to better aid the understanding. We discuss fundamental concepts in logic such as relational systems, languages, and interpretation of sentences in the first sections 3.1 and 3.2. In section 3.3, we generalize the *-transform to suit this abstract setting and discuss how it acts on simple sentences. At this point, sufficient terminology has been developed to properly state and prove the transfer principle. The proof is more notationally cumbersome than it is conceptually difficult, as such it deserves an entire section 3.4.

Chapter 2

The Hyperreal Number System

In this chapter we construct the hyperreal number system as the set (of equivalence classes) of real sequences. We will do so following the remarks made in the motivation 1.1 about the importance of looking at sufficiently large index sets. Here, I will denote our index set and while we will only apply these concepts when $I = \mathbb{N}$, we present the definitions in greater generality.

2.1 Ultrafilters

Definition 2.1 (Filter)**.** Let I be a nonempty set. We say that a nonempty collection $\mathcal{U}$ of subsets of I is a *filter* on I if it has the following properties:

1. $\emptyset \notin \mathcal{U}$
2. $A, B \in \mathcal{U} \implies A \cap B \in \mathcal{U}$
3. $A \in \mathcal{U},\, A \subset B \subset I \implies B \in \mathcal{U}$

Example 2.1.1 (Filter)**.** A trivial example of a filter is $\mathcal{U} = \{I\}$.

Definition 2.2 (Cofinite/Fréchet filter)**.** For I an infinite set, we define the *cofinite* or *Fréchet* filter on I as:

$$\mathcal{F}_I := \{A \subset I \mid A^{\complement} \text{ is finite}\}$$

Definition 2.3 (Ultrafilter)**.** We say that a filter $\mathcal{U}$ on I is an *ultrafilter* if:

$$\forall A \subset I, \text{ either } A \in \mathcal{U} \text{ or } A^{\complement} \in \mathcal{U} \text{ (but not both by items 1 and 2 of definition 2.1).}$$

Definition 2.4 (Free ultrafilter)**.** An ultrafilter $\mathcal{U}$ on I is *free* if it contains $\mathcal{F}_I$ as in the above definition 2.2.

Definition 2.5 (Fixed ultrafilter)**.** For $x \in I$, the ultrafilter which *fixes* x defined as:

$$\mathcal{U}_x := \{A \subset I \mid x \in A\}$$

Free ultrafilters are essential in the construction of the hyperreals. Furthermore, we will make logical statements such as 'hyperreal r satisfies property $*P$ (where $*P$ is the hyperreal analogoue of a real property P)' to mean that 'the set of indices I for which r_i satisfies property P for $i \in I$ belongs to $\mathcal{U}$.' All of this will be made precise.

Intuitively, free ultrafilters $\mathcal{U}$ are composed of the 'sufficiently large' index subsets as mentioned in the motivation 1.1. Observe that finite sets F cannot be elements of $\mathcal{U}$ for otherwise $(F^{\complement})^{\complement}$ is finite and hence $F^{\complement}$ also belongs in $\mathcal{U}$. But then by item 2 of filters in definition 2.1, we must have $F \cap F^{\complement} = \emptyset \in \mathcal{U}$ which contradicts item 1. By a similar argument, $\mathcal{U}$ cannot contain disjoint sets. By definition 2.3 and item 1 for filters, $I \in \mathcal{U}$.

At this stage, it is not clear why some ultrafilters might be prefered over others. Later on, in the construction of the hyperreals, we will show that using fixed ultrafilters would actually build an object that is isomorphic to the familar reals (proposition 2.2.1). On the other hand, free ultrafilters remedy this trivial reduction and provide the hyperreals with the desired properties. The following proposition characterizes the differences between free and fixed ultrafilters.

Proposition 2.1.1 (Free iff not fixed)**.** *For a non-empty set I and an ultrafilter $\mathcal{U}$ on it, we have the following equivalence:*

$$\mathcal{U} \textit{ is free} \iff \mathcal{U} \textit{ is not fixed at any } x \in I$$

Proof.
($\Longrightarrow$) Assume for a contradiction that there is a fixed element $x \in I$, that is, $\mathcal{U}$ contains all the subsets of I which contain x. In particular, $\mathcal{U}$ certainly admits $\{x\}$ as an element. Note, however, that $(\{x\}^{\complement})^{\complement} = \{x\}$ is finite. By the definition 2.2 of the Fréchet filter, this means that $\mathcal{U}$ also admits $\{x\}^{\complement}$ as an element. Appealing to items 1 and 2 of the filter definition 2.1, we arrive at a contradiction.
($\Longleftarrow$) Fix $A \subset I$ such that $A^{\complement}$ is finite. The finiteness assumption means that we can express this complement set as $A^{\complement} = \{b_i\}_{i=1}^n$ for some $n \in \mathbb{N}$ and $b_i \in I$, $\forall i = 1, \dots, n$. By hypothesis, $\mathcal{U}$ cannot fix these b_i so we must be able to find $B_i \subset I$ such that $B_i \notin \mathcal{U}$ with $b_i \in B_i$ for every $i = 1, \dots, n$. By definition 2.3 of an ultrafilter, since $B_i \notin \mathcal{U}$, we must have $B_i^{\complement} \in \mathcal{U}$, $\forall i = 1, \dots, n$. Consider now the following sequence of set arithmetic:

$$A^{\complement} = \{b_i\}_{i=1}^n \subset \bigcup_{i=1}^n B_i \implies A \supset \left(\bigcup_{i=1}^n B_i\right)^{\complement} = \bigcap_{i=1}^n B_i^{\complement}$$

Remembering items 2 and 3 for filters when looking at the expression to the right of the implication symbol gives precisely $A \in \mathcal{U}$. □

It is not clear that (free) ultrafilters exist. The existence proof relies on Zorn's lemma which we relegate to appendix A.

Theorem 2.1.1 (Ultrafilter Axiom)**.** *Let I be a non-empty set and $\mathcal{F}$ be a filter on I. Then there is an ultrafilter $\mathcal{U}$ on I which contains $\mathcal{F}$.*

Proof. Contained in appendix at theorem A.0.1 of the same result. □

2.2 Constructing the Hyperreal numbers

Despite the brevity of the previous section 2.1, we have the foundation to formulate the hyperreal numbers system. In what follows, we will take our index set I to be $\mathbb{N}$ and denote a free ultrafilter on it by $\mathcal{U}$. We use the notation $\hat{\mathbb{R}}$ for the set of real sequences.

Definition 2.6 (Almost everywhere)**.** Let $r = \langle r_i \rangle$ and $s = \langle s_i \rangle$ belong to $\hat{\mathbb{R}}$. We say that r is equal to s *almost everywhere (a.e.)* written $r \equiv s$ if the following holds:

$$\{i \in \mathbb{N} \,|\, r_i = s_i\} \in \mathcal{U}$$

Lemma 2.2.1 ($\equiv$ is an equivalence relation)**.** *The relation $\equiv$ as defined above in definition 2.6 is an equivalence relation on $\hat{\mathbb{R}}$.*

Proof.
Reflexive: Clearly $\{i \in \mathbb{N} \,|\, r_i = r_i\} = \mathbb{N} \in \mathcal{U}$.
Symmetric: If $r \equiv s$, then symmetry of $=$ on $\mathbb{R}$ gives the result. In more detail:

$$\mathcal{U} \ni \{i \in \mathbb{N} \,|\, r_i = s_i\} = \{i \in \mathbb{N} \,|\, s_i = r_i\}$$

Transitive: Let $r, s, t \in \hat{\mathbb{R}}$ be such that $r \equiv s, s \equiv t$. Denote $A = \{i \in \mathbb{N} \,|\, r_i = s_i\}$ and $B = \{i \in \mathbb{N} \,|\, s_i = t_i\}$. Clearly $A \cap B \subset C := \{i \in \mathbb{N} \,|\, r_i = t_i\}$. By remembering item 2 of filters (giving $A \cap B \in \mathcal{U}$) and then applying item 3 (as $C \supset A \cap B$), we see that $C \in \mathcal{U}$ hence $r \equiv t$. □

From this equivalence relation, we see that $\omega = \left\langle \frac{1}{n} \right\rangle$ as defined in the motivation 1.1 is an example of an 'arbitrarily small' but non-zero number. Many other such numbers can be easily constructed.

Definition 2.7 (Nonstandard/Hyperreal numbers)**.** Let $\mathbf{R}$ denote the set of equivalence classes of $\hat{\mathbb{R}}$ induced by the equivalence relation defined by $\equiv$ in definition 2.6. $s = \langle s_i \rangle \in \hat{\mathbb{R}}$ will belong to the equivalence class denoted by $[s]$ or $\mathbf{s}$. With this notation, $r \equiv s$ in $\hat{\mathbb{R}}$ is equivalent to $\mathbf{r} = [r] = [s] = \mathbf{s}$. Elements in $\mathbf{R}$ such as those used just now are called *nonstandard* or *hyperreal* numbers.

As with the familiar number systems of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, we would like to endow $\mathbf{R}$ with operations corresponding to addition and multiplication. In what follows, we will abuse notation and use the normal addition and multiplication symbols on sequences to mean 'element-by-element addition and multiplication,' respectively. We will also extend the ordering given by 'less than' (and 'greater than').

Definition 2.8 (Addition, Multiplication, and Ordering)**.** Let $\mathbf{r} = [\langle r_i \rangle]$ and $\mathbf{s} = [\langle s_i \rangle]$ belong to $\mathbf{R}$. We define the following:

1. $\mathbf{r} + \mathbf{s} = [\langle r_i + s_i \rangle]$ so that $[r] + [s] = [r + s]$
2. $\mathbf{r} \cdot \mathbf{s} = [\langle r_i \cdot s_i \rangle]$ so that $[r] \cdot [s] = [r \cdot s]$
3. $\mathbf{r} < \mathbf{s}$ (and $\mathbf{s} > \mathbf{r}$) to mean $\{i \in \mathbb{N} \,|\, r_i < s_i\} \in \mathcal{U}$. Similarly, $\mathbf{r} \leq \mathbf{s}$ ($\mathbf{s} \geq \mathbf{r}$) to mean $\mathbf{r} < \mathbf{s}$ or $\mathbf{r} = \mathbf{s}$.

The structure $(\mathbf{R}, +, \cdot, <)$ will be denoted by $\boldsymbol{\mathcal{R}}$. Although we have not mentioned it previously, we denote the structure $(\mathbb{R}, +, \cdot, <)$ (the usual operations and ordering on $\mathbb{R}$) as $\mathcal{R}$.

Of course, we must make sure that these definitions are independent of representatives chosen from the equivalence classes. Very briefly, if $r \equiv \bar{r}$, $s \equiv \bar{s}$, $r < s$, $\bar{r} < \bar{s}$ with $r = \langle r_i \rangle, \bar{r} = \langle \bar{r}_i \rangle, s = \langle s_i \rangle, \bar{s} = \langle \bar{s}_i \rangle$ we have that A, B, C, $D := \{i \in \mathbb{N} \,|\, r_i = \bar{r}_i\}$, $\{i \in \mathbb{N} \,|\, s_i = \bar{s}_i\}$, $\{i \in \mathbb{N} \,|\, r_i < s_i\}$, $\{i \in \mathbb{N} \,|\, \bar{r}_i < \bar{s}_i\}$ are all in $\mathcal{U}$. The following set inclusions are basic:

$$\begin{gathered}
\{i \in \mathbb{N} \,|\, r_i + s_i = \bar{r}_i + \bar{s}_i\} \supset A \cap B \\
\{i \in \mathbb{N} \,|\, r_i \cdot s_i = \bar{r}_i \cdot \bar{s}_i\} \supset A \cap B \\
\{i \in \mathbb{N} \,|r_i < \bar{s}_i\} \cap \{i \in \mathbb{N} \,|\bar{r}_i < s_i\} \supset A \cap B \cap C \cap D
\end{gathered}$$

Using the relations above and the familiar combination of items 2 and 3 of definition 2.1 gives the desired independence of definitions 2.8 with representatives of the equivalence classes of elements of $\mathbf{R}$.

With definitions 2.8 established, we assert that $\boldsymbol{\mathcal{R}}$ is a linearly ordered field.

Theorem 2.2.1. $\boldsymbol{\mathcal{R}}$ *is a linearly ordered field.*

Proof. The zero and unit are $\mathbf{0} = [\langle 0, 0, \dots \rangle]$ and $\mathbf{1} = [\langle 1, 1, \dots \rangle]$, respectively. One easily establishes the commutative ring axioms by applying them term-by-term in the sequences.

We need to establish the existence of multiplicative inverses for non-zero elements. Let $\mathbf{r} = [\langle r_i \rangle] \neq \mathbf{0}$. This gives $\{i \in \mathbb{N} \,|\, r_i \neq 0\} \in \mathcal{U}$. Define $\mathbf{r}^{-1}$, the inverse of $\mathbf{r}$ by:

$$(r^{-1})_i := \begin{cases} r_i^{-1}, & r_i \neq 0, \\ 0, & r_i = 0 \end{cases}$$

It is easy now to see that $\mathbf{r} \cdot \mathbf{r}^{-1} = \mathbf{1}$.

To complete the final field axiom, we need to establish the integral domain property, i.e. that $\boldsymbol{\mathcal{R}}$ has no (non-zero) zero-divisors. Let $\mathbf{r} \cdot \mathbf{s} = \mathbf{0}$. If, say, $\{i \in \mathbb{N} \,|\, r_i = 0\} \in \mathcal{U}$, then we're done as this means $\mathbf{r} = \mathbf{0}$. If $\{i \in \mathbb{N} \,|\, r_i = 0\} \notin \mathcal{U}$ (and equivalently $\{i \in \mathbb{N} \,|\, r_i \neq 0\} \in \mathcal{U}$), in order for the product to be zero $\mathbf{r} \cdot \mathbf{s} = \mathbf{0}$, we need $\{i \in \mathbb{N} \,|\, r_i \neq 0\} \subset \{i \in \mathbb{N} \,|\, s_i = 0\}$. By item 3 of filters, this shows that $\mathbf{s} = \mathbf{0}$.

To demonstrate linear ordering, we have three things to show:

1. The sum of positive elements is positive.
2. The product of positive elements is positive.

3. (Law of Trichotomy) Every element is either strictly positive, identically $\mathbf{0}$, or is strictly negative.

Showing items 1 and 2 is done in the same way as showing that the field operations are independent of representation in the equivalence classes. Given positive elements $\mathbf{r}$ and $\mathbf{s}$, we have:

$$\{i \in \mathbb{N} \,|\, r_i + s_i \geq 0\}, \{i \in \mathbb{N} \,|\, r_i s_i \geq 0\} \supset \{i \in \mathbb{N} \,|\, r_i \geq 0\} \cap \{i \in \mathbb{N} \,|\, s_i \geq 0\} \in \mathcal{U}$$

For item 3, fix any $\mathbf{r}$ and define $A, B, C := \{i \in \mathbb{N} \,|\, r_i > 0\}, \{i \in \mathbb{N} \,|\, r_i = 0\}, \{i \in \mathbb{N} \,|\, r_i < 0\}$. By the law of trichotomy for $\mathbb{R}$, we have $A \cup B \cup C = \mathbb{N} \in \mathcal{U}$. If none of A, B, C are in $\mathcal{U}$ then their complements are all in $\mathcal{U}$ by the ultrafilter definition 2.3. However, this yields: $\emptyset = (A \cup B \cup C)^{\complement} = (A^{\complement} \cap B^{\complement} \cap C^{\complement}) \in \mathcal{U}$ which is impossible by the familiar combination of items 1 and 2 in the filter definition 2.1. So, at least one of A, B, C are in $\mathcal{U}$. It is easy to see that only one of them can be in $\mathcal{U}$ because disjoint sets cannot simultaneously be elements of $\mathcal{U}$ (again by combining items 1 and 2 of filters). □

We now want to embed $\mathcal{R}$ as a linearly ordered subfield into $\boldsymbol{\mathcal{R}}$. This leads to the following definition of the *-transform, $* : \mathbb{R} \to \boldsymbol{R}$.

Definition 2.9 (*-transform)**.** For $r \in \mathbb{R}$, define $*(r) = {}^*r = [\langle r, r, \dots \rangle] \in \boldsymbol{R}$, the **-transform* of r.

Theorem 2.2.2. *The mapping $*$ as defined above in definition 2.9 is an order-preserving ring homomorphism of $\mathbb{R}$ into $\boldsymbol{R}$.*

Proof. Clearly, $*$ is injective for ${}^*r = {}^*s$ gives $\{i \in \mathbb{N} \,|\, r = s\} \in \mathcal{U}$ so that $r = s$ for $\{i \in \mathbb{N} \,|\, r = s\}$ is either $\mathbb{N}$ or $\emptyset$. By construction of $*$ and theorem 2.2.1, it is easy to see that $*$ preserves the field and ordering properties of $\mathbb{R}$. □

Definition 2.10 (Standard numbers)**.** For $A \subset \mathbb{R}$, we define $(A)_* := \{{}^*a \,|\, a \in A\}$. We refer to $(\mathbb{R})_*$ as the set of *standard* numbers of $\boldsymbol{R}$.

Let us now return to the importance of using *free* ultrafilters. We mentioned this in the previous section 2.1 before proposition 2.1.1 which asserted that an ultrafilter is free iff it is not fixed. To confirm that $\boldsymbol{R}$ actually is different from the standard numbers, consider $\mathbf{1}/\boldsymbol{\omega} = [\langle i \rangle]$ (ω as in the motivation section 1.1). For every $r \in \mathbb{R}$, the set $\{i \in \mathbb{N} \,|\, r = i\}$ is either empty or only has one natural number. This is where the free property comes in, the complement set $\{i \in \mathbb{N} \,|\, r \neq i\} \in \mathcal{U}$ as $\{i \in \mathbb{N} \,|\, r = i\}$ is finite by the previous sentence. Therefore, $\mathbf{1}/\boldsymbol{\omega} \neq {}^*r, \forall r \in \mathbb{R}$. This shows that $\boldsymbol{R}$ is a genuinely different object than $\mathbb{R}$ and so there is potential for interesting theory to be developed in studying $\boldsymbol{R}$.

On the other hand, it was also promised that if $\boldsymbol{R}$ was constructed using fixed ultrafilters instead, we would actually see that $\boldsymbol{R}$ is isomorphic to $\mathbb{R}$.

Proposition 2.2.1 (Triviality of $\boldsymbol{R}$ using fixed ultrafilters)**.** *Let $\mathcal{U}$ be an ultrafilter on $\mathbb{N}$ which fixes some $n \in \mathbb{N}$. Taking this ultrafilter in the construction of $\boldsymbol{R}$ (and its linearly ordered structure $\boldsymbol{\mathcal{R}}$) actually recovers the familiar reals $\mathbb{R}$ (and its linearly ordered structure $\mathcal{R}$).*

Proof. Let $n \in \mathbb{N}$ be the natural number that is fixed by the ultrafilter $\mathcal{U}$. This specifies the almost everywhere definition 2.6 to saying '$r = \langle r_i \rangle$ is almost everywhere equal to $s = \langle s_i \rangle$' if and only if $r_n = s_n$. The 'if' direction follows as $\{n\} \in \mathcal{U}$. The 'only if' direction follows because $\{i \in \mathbb{N} | r_i = s_i\}$ is a set which must contain n as $\mathcal{U}$ is a fixed ultrafilter. This trivializes the equivalence classes constructed by lemma 2.2.1 as an equivalence class, $\mathbf{r}$, containing r can simply be identified with r_n. Thus, the addition, multiplication, and ordering defined by definition 2.8 collapse to those familiar operations and ordering on $\mathbb{R}$ focusing on the nth element of the sequence. Continuing in the construction of $\boldsymbol{\mathcal{R}}$ from this obviously produces an object which is isomorphic to $\mathcal{R}$. □

Theorem 2.2.1 is just one example where we show that $\boldsymbol{\mathcal{R}}$ inherits a property (linearly ordered field) from $\mathcal{R}$. This is generalized by the famous 'Transfer Principle' which we shall explore later on in section 3.3. In the proof of theorem 2.2.1, we needed to make sure that the addition, multiplication, and ordering operations and relation we defined in definition 2.8 made $\boldsymbol{\mathcal{R}}$ a linearly ordered field analogous to how those operations make the familiar $\mathcal{R}$ a linearly ordered field. The proof of theorem 2.2.1 was conceptually simple, however it involved a great deal of tedious manipulation. The transfer principle offers a universal 'extension' of properties enjoyed by $\mathcal{R}$, thereby simplifying proofs for results such as theorem 2.2.1. Informally, it says that 'If property P is satisfied by $\mathcal{R}$, then "an appropriately modified and analogous property" ${}^*P$ is also satisfied by $\boldsymbol{\mathcal{R}}$.' The notation ${}^*P$ is intentional as the *-transform defined in definition 2.9 is precisely the key to extending properties satisfied in $\mathcal{R}$ to analogous properties satisfied in $\boldsymbol{\mathcal{R}}$.

To properly state the transfer principle, we need to make the notion of 'property satisfied by $\mathcal{R}$ (and $\boldsymbol{\mathcal{R}}$)' precise. As well, the *-transform in definition 2.9 is defined for real numbers. We need to understand what it means to take the *-transform of more general objects.

2.3 *-transform of Relations

In this section, we will extend relations from $\mathcal{R}$ to $\boldsymbol{\mathcal{R}}$. To begin, we will review some basic definitions which the reader will already be familiar with (modulo notational conventions).

Definition 2.11 (Basics of Set Theory)**.**
Given a set S, the set $S^n = S \times S \times S \cdots \times S$ (n times) consists of the ordered *n-tuples* $\langle a^1, a^2, \dots, a^n \rangle$ where each $a^i \in S$.
We call $P \subset S^n$ an *n-ary relation* on S. If $\langle a^1, \dots, a^n \rangle \in P$ we will equivalently express this with $P\langle a^1, \dots, a^n \rangle$.
We call the *complement* of an n-ary relation P the relation $P^{\complement} = S^n \setminus P$.
For a relation P, its *domain* is a subset of S^{n-1} given by:

$$\mathrm{dom}P := \{\langle a^1, \dots, a^{n-1} \rangle \in S^{n-1} \,|\, \exists a \in S \text{ s.t. } P\langle a^1, \dots, a^{n-1}, a \rangle\}$$

For a relation P, its *range* is a subset of S given by:

$$\mathrm{range}P := \{a \in S \,|\, \exists \langle a^1, \dots, a^{n-1} \rangle \in S^{n-1} \text{ s.t. } P\langle a^1, \dots, a^{n-1}, a \rangle\}$$

An *S-valued function* f of n variables on S is an $(n+1)$-ary relation such that:

$$f\langle a^1, \dots, a^n, a \rangle = f\langle a^1, \dots, a^n, b \rangle \implies a = b$$

Such an $a \in S$ for which $f\langle a^1, \dots, a^n, a \rangle$ is called the *image* of $\langle a^1, \dots, a^n, a \rangle$ under f. We will also alternate with $f(a^1, \dots, a^n) = a \iff f\langle a^1, \dots, a^n, a \rangle$.

Easy examples of binary relations are $=$ and $<$. In the previous sections involving these relations, we have abused notation, however, the reader will appreciate the common practice of writing, for example, $a = b$ instead of $=\langle a, b \rangle$. Similarly, $+$ and $\cdot$ are functions of two variables (or ternary relations) which we will write $a + b = c$ instead of $+(a, b) = c$ or $+\langle a, b, c \rangle$ with $\cdot$ following in the same way. As we progress, the familiar operations may suppress the formal symbolic manipulation for convenience.

Definition 2.12 (*-transform)**.** Let P be an n-ary relation on $\mathbb{R}$. The **-transform* ${}^*P$ of P is the following set:

$$\{\langle \mathbf{r}^1, \dots, \mathbf{r}^n \rangle \in \boldsymbol{R}^n \,|\, \text{For } \mathbf{r}^k = [\langle r^k_1, r^k_2, \dots \rangle],\ P\langle r^1_i, \dots, r^n_i \rangle \text{ holds a.e.}\}$$

Note that ${}^*P$ is well-defined because if $\mathbf{r}^k = [\langle r^k_i \rangle] = [\langle \bar{r}^k_i \rangle]$ for $k = 1, \dots, n$, let $A = \{i \in \mathbb{N} \,|\, P\langle r^1_i, \dots, r^n_i \rangle\}$ which is in $\mathcal{U}$, say. But as well, $B = \bigcap_{k=1}^n \{i \in \mathbb{N} \,|\, r^k_i = \bar{r}^k_i\} \in \mathcal{U}$ by the all too familiar combination of items 2 and 3 of filters. Hence $\{i \in \mathbb{N} \,|\, P\langle \bar{r}^1_i, \dots, \bar{r}^n_i \rangle\} \supset A \cap B \in \mathcal{U}$, we have that definition 2.12 is well-defined for the equivalence classes defined by 2.6.

The addition, multiplication, and ordering relations defined in 2.8 are readily seen to be special cases of this definition 2.12.

Example 2.12.1. For a subset $A \subset \mathbb{R}$, definition 2.12 gives ${}^*A = \{[\langle s_i \rangle] \in \boldsymbol{R} \,|\, \{i \in \mathbb{N} \,|\, s_i \in A\} \in \mathcal{U}\}$. Easily, ${}^*A \supset (A)_*$; elements of the right-hand side are 'constant sequences' $\langle a, a, \dots \rangle$ whereas the left-hand side contains these elements (and in general, many more). In particular, ${}^*\mathbb{R} = \boldsymbol{R} \supsetneq (\mathbb{R})_*$.

Example 2.12.2. As another example, consider $A = [a, b] = \{x \in \mathbb{R} \,|\, a \leq x \leq b\}$. We calculate ${}^*A$ by definition 2.12 and the just-mentioned comment on subsets of $\mathbb{R}$:

$$\begin{aligned} {}^*A &= \{[\langle s_i \rangle] \in \boldsymbol{R} \,|\, \{i \in \mathbb{N} \,|\, s_i \in A\} \in \mathcal{U}\} \\ &= \{[\langle s_i \rangle] \in \boldsymbol{R} \,|\, \{i \in \mathbb{N} \,|\, a \leq s_i \leq b\} \in \mathcal{U}\} \\ &= \{[\langle s_i \rangle] \in \boldsymbol{R} \,|\, \langle a \rangle \leq \langle s_i \rangle \leq \langle b \rangle \text{ holds a.e.}\} \\ &= \{[\langle s_i \rangle] \in \boldsymbol{R} \,|\, {}^*a \leq [\langle s_i \rangle] \leq {}^*b\} \\ &= \{\mathbf{x} \in \boldsymbol{R} \,|\, {}^*a \leq \mathbf{x} \leq {}^*b\} \end{aligned}$$

We thus have the following ${}^*[a, b] = [{}^*a, {}^*b]$ (forgiving the abuse of notation on the right-hand side).

Example 2.12.3. Let us turn to f a function of n variables so that, for $\mathbf{r}^k = [\langle r_i^k \rangle], k = 1, \ldots, n$ and $\mathbf{s} = [\langle s_i \rangle]$, definition 2.12 gives:

$${}^*f\langle \mathbf{r}^1, \ldots, \mathbf{r}^n, \mathbf{s} \rangle \text{ holds in } \boldsymbol{R}^{n+1} \iff f\langle r_i^1, \ldots, r_i^n, s_i \rangle \text{ holds a.e. } i \in \mathbb{N}$$

So ${}^*f$ is a function of n variables in $\boldsymbol{R}$ and ${}^*f(\mathbf{r}^1, \ldots, \mathbf{r}^n)$ is defined iff $f(r_i^1, \ldots, r_i^n)$ is defined a.e. $i \in \mathbb{N}$. Furthermore, where they are defined, ${}^*f(\mathbf{r}^1, \ldots, \mathbf{r}^n) = \mathbf{s}$ whenever $f(r_i^1, \ldots, r_i^n) = s_i$ a.e. $i \in \mathbb{N}$. In particular, it is true that ${}^*f({}^*r^1, \ldots, {}^*r^n) = {}^*s$. Similar to the previous example on the *-transform of closed and bounded intervals, the *-transform of these relations could be characterized by passing the *-transform onto the constants involved. This is the intuitive rule behind of the transfer principle. We content ourselves at present with the following theorem.

Theorem 2.3.1 (Extension of Relations)**.** *For an n-ary relation P on $\mathbb{R}$ with $P\langle r^1, \ldots, r^n \rangle$ for $r^k \in \mathbb{R}$, we have ${}^*P\langle {}^*r^1, \ldots, {}^*r^n \rangle$.*

Proof. We prove the result by starting from definition 2.12.

$$\begin{aligned} {}^*P\langle {}^*r^1, \ldots, {}^*r^n \rangle &\iff P\left\langle \left({}^*r^1\right)_i, \ldots, \left({}^*r^n\right)_i \right\rangle \text{ holds a.e. } i \in \mathbb{N} \\ &\iff P\langle r^1, \ldots, r^n \rangle \text{ holds a.e. } i \in \mathbb{N} \\ &\iff P\langle r^1, \ldots, r^n \rangle \end{aligned}$$

□

Remark 2.3.1. *Theorem 2.3.1 gives an interpretation of the *-transform of relations P as extensions of those same relations to $\boldsymbol{R}$ when one thinks of identifying r^k with ${}^*r^k$ (i.e. embedding $\mathbb{R}$ in $\boldsymbol{R}$ by the injection given by definition 2.9).*

Theorem 2.3.1 has an obvious application to functions of n variables. We may also like to characterize n-ary relations P by their *characteristic function*:

$$\chi_P(x^1, \ldots, x^n) := \begin{cases} 1 & \langle x^1, \ldots, x^n \rangle \in P \\ 0 & \langle x^1, \ldots, x^n \rangle \notin P \end{cases}$$

Proposition 2.3.1. *For an n-ary relation P on $\mathbb{R}$, we have ${}^*\chi_P = \chi_{{}^*P}$. Here, we abuse notation by identifying $1 \in \mathbb{R}$ with ${}^*1 \in \boldsymbol{R}$ as well with 0 and ^{*0} as alluded to in remark 2.3.1.*

Proof. For any $\mathbf{x}^k = [\langle x_i^k \rangle]$, $k = 1, \ldots, n$, we have:

$$\begin{aligned} \chi_{{}^*P}(\mathbf{x}^1, \ldots, \mathbf{x}^n) &= \begin{cases} 1 & \langle \mathbf{x}^1, \ldots, \mathbf{x}^n \rangle \in {}^*P \\ 0 & \langle \mathbf{x}^1, \ldots, \mathbf{x}^n \rangle \notin {}^*P \end{cases} \\ &= \begin{cases} 1 & \langle x_i^1, \ldots, x_i^n \rangle \in P \text{ a.e. } i \in \mathbb{N} \\ 0 & \langle x_i^1, \ldots, x_i^n \rangle \notin P \text{ a.e. } i \in \mathbb{N} \end{cases} \\ &= \chi_P(x_i^1, \ldots, x_i^n) \text{ a.e. } i \in \mathbb{N} \\ &= {}^*\chi_P(\mathbf{x}^1, \ldots, \mathbf{x}^n) \end{aligned}$$

□

We are now comfortable enough with the *-transform to develop a more abstract framework. This will involve elements of symbolic logic for relational systems which we develop in the next chapter 3. In particular, we will build a coherent theory to state and prove the transfer principle. As stated earlier, the transfer principle says something of the form 'If a statement is true in $\mathbb{R}$, then its extension into $\boldsymbol{R}$ (which we can interpret now as the *-transform) is also true in $\boldsymbol{R}$.' We are closer to making this phrase precise however we will need to incorporate a more abstract theory in order to make sense of this.

Chapter 3

Transfer Principle

In this chapter, we develop the terminology and formalism from symbolic logic to properly state and prove the transfer principle. Towards the end of the last chapter 2, we proved very basic extension results for relations and functions defined on the reals to their *-transforms in the hyperreals. This chapter is devoted to improving these results in a more general and abstract setting. Starting from section 3.1 we collect the fundamental building blocks of this theory. Section 3.3 properly states the transfer principle and gives some applications of it. We defer the proof of the transfer principle to section 3.4.

Following the convention by Hurd and Loeb [1] to unify notation in the same way as Robinson [2], we will change a few symbols as follows:

1. $\boldsymbol{R}$ and $\boldsymbol{\mathcal{R}}$ will be denoted by ${}^*\mathbb{R}$ and ${}^*\mathcal{R}$, respectively.

2. As stated in remark 2.3.1, we will think of $\mathbb{R} \subset {}^*\mathbb{R}$. We will also identify $\mathbb{R}$ with $(\mathbb{R})_*$.

3. On top of the linearly ordered structure, $\mathcal{R}$ will denote the structure consisting of $\mathbb{R}$ together with all its relations and functions. ${}^*\mathcal{R}$ will similarly denote the structure consisting of ${}^*\mathbb{R}$ together with all *extended* relations and functions on $\mathbb{R}$. In other words, every relation and function in ${}^*\mathcal{R}$ is an extension (in the *-transform sense) of a relation or a function in $\mathcal{R}$.

The significance of this is that the theory we develop is general enough so that we are not limited to $\mathbb{R}$ or its *-transform.

3.1 Simple Languages and Simple Sentences

Definition 3.1 (Relational System)**.** For a set S, a collection of relations on it denoted by $\{P_i\}_{i\in I}$, and a collection of functions on it denoted by $\{f_j\}_{j\in J}$ where I and J are (non-empty) index sets, we call the triple $\mathcal{S} = (S, \{P_i\}_{i\in I}, \{f_j\}_{j\in J})$ a *relational system.*

We are working towards formalizing the statement 'If a statement P is true in $\mathbb{R}$, then its analogous extension to ${}^*\mathbb{R}$, ${}^*P$ is also true.' More generally, we need to make the notions of statements of structures $\mathcal{S}$ (not necessarily $\mathbb{R}$ or ${}^*\mathbb{R}$) precise. To wit, we introduce simple languages.

'Definition' 3.1 (Simple Language)**.** For a structure $\mathcal{S}$, we associate to it a *symbolic language* $L_{\mathcal{S}}$ which consists of a set of *basic symbols* and specific combinations of thses basic symbols (which we will call *simple sentences* defined later in definition 3.3). The basic symbols fall into two categories:

Logical Symbols: These symbols are universal to any simple language and do not depend on $\mathcal{S}$.

- **Logical Connectives**: These symbols are $\wedge$ and $\implies$ which will be interpreted later as the usual 'and' and 'implies.'

- **Quantifier Symbol**: The symbol is $\forall$ which will be interpreted as the usual 'for all.'

- **Parenthese**: The symbols $[,], (,), \langle, \rangle$ which will be used for the usual bracketing.

- **Variable Symbols**: This is a countable collection of symbols such as x, y, z, x_1, x_2, m, and n which will be used as 'variables.' One can think of an alphabet of any ordinary national 'language' (or unions thereof).

Parameters: These symbols depend on $\mathcal{S}$.

- **Constant Symbols**: These symbols are $\underline{s}$ calling the name of $s \in S$.
- **Relation Symbols**: These symbols are $\underline{P}$ calling the name of the relation P on S.
- **Function Symbols**: These symbols are $\underline{f}$ calling the name of the relation f on S.

We have used quotation marks for definition 3.1 because we haven't precisely specified what the variable symbols can consist of. However, the mathematically mature reader may content themselves that the variable symbols consist of all the symbols they regularly use to denote variables in their mathematical workings.

For common constant, relation, and function symbols, we will omit underlining them so that, for example, π names the familiar $\pi \in \mathbb{R}$. We may also have more than one name for each of these symbols.

The symbols in simple languages are themselves meaningless. We need to show how to combine them to build meaningful expressions like $\underline{f}(x) + \sin(\pi + 1/2)$.

Definition 3.2 (Term)**.** We define *terms* inductively by the following:

1. Every constant and variable symbol is a term.
2. If $\underline{f}$ names a function of n variables and $\tau^1, \ldots, \tau^n$ are terms, then $\underline{f}(\tau^1, \ldots, \tau^n)$ is also a term.

A *constant term* is a term that contains no variables in the outline provided above.

Example 3.2.1. In $L_{\mathcal{R}}$ an expression like $5x^2 + y$ is a term which we can write formally as $\underline{S}(\underline{P}(5, \underline{Sq}(x)), y)$ where $S(a, b) = a + b, P(a, b) = ab, Sq(a) = a^2$ for $a, b \in \mathbb{R}$.

Definition 3.3 (Simple Sentence)**.** A *simple sentence* is a string of symbols in $L_{\mathcal{S}}$ which takes either of the following forms:

1. *Atomic sentences*: These sentences are of the form $\underline{P}(\tau^1, \ldots, \tau^n)$ where $\underline{P}$ names an n-ary relation and τ^i for $i = 1, \ldots, n$ are constant terms.
2. *Compound sentences*: These sentences are of the form:

$$(\forall x_1) \ldots (\forall x_n) \left[\bigwedge_{i=1}^{k} \underline{P}_i \langle \tilde{\tau}_i \rangle \implies \bigwedge_{j=1}^{l} \underline{Q}_j \langle \tilde{\sigma}_j \rangle \right]$$

Here, $\tilde{\tau}_i$ is an n_i-tuple of terms ($\tilde{\tau}_i = \left\langle \tau_i^1, \ldots, \tau_i^{n_i} \right\rangle$) involving only the variables $x_1, \ldots, x_n$. n_i is the order of the relation P_i named by $\underline{P}_i$ (i.e. P_i is an n_i-ary relation). Similarly for $\tilde{\sigma}_j$ and $\underline{Q}_j$.

This compound sentence will eventually have an 'interpretation' (definition 3.5) as: For every $x_1, \ldots, x_n$ if each $\tilde{\tau}_i$ (which only involve the variables $x_1, \ldots, x_n$) belong to P_i respectively, then each $\tilde{\sigma}_j$ (which again only involve the variables $x_1, \ldots, x_n$) also belongs to Q_j, respectively where i and j run over their appropriate indices.

Example 3.3.1 (Atomic sentence)**.** Consider $\underline{I}$ the name of the inequality relation I so that $I\langle a, b \rangle \iff a < b$. The expression $\underline{I}\langle 3, 19 \rangle$ is an atomic sentence because 3 and 19 are constant terms as constant symbols of $L_{\mathcal{R}}$.

Example 3.3.2 (Compound sentence)**.** Let $\underline{\mathbb{R}}$ be the name for the 1-ary relation $\mathbb{R}$ so that $\mathbb{R}\langle a \rangle \iff a \in \mathbb{R}$. We will write down, under the symbols we have just presented with definitions 3.1 and 3.3, the statement 'A real number added with a positive real number is greater than the original number.' We will mix using relations and normal inequality signs to ease the transition of familiar symbolic manipulation and our formal representation:

$$(\forall x)(\forall y) \left[\underline{\mathbb{R}}\langle x \rangle \wedge \underline{\mathbb{R}}\langle y \rangle \wedge y > 0 \implies \underline{I}\langle x, x + y \rangle \right]$$

The above is a compound sentence. We invite the reader to practice writing basic rules using this framework.

Example 3.3.3 (Non-example)**.** Consider the same sentence in the previous example 3.3.2. We remove the '$(\forall y)$' part just before the square brackets giving:

$$(\forall x)\,[\underline{\mathbb{R}}\langle x\rangle \wedge \underline{\mathbb{R}}\langle y\rangle \wedge y > 0 \implies \underline{I}\langle x, x+y\rangle]$$

This is not a simple sentence because the variable y within the square brackets has no corresponding $(\forall y)$ outside.

We will use the $\iff$ symbol to shorten pairs of compound sentences for which the order of $\bigwedge_{i=1}^{k} \underline{P}_i$ and $\bigwedge_{j=1}^{l} \underline{Q}_j$ are swapped.

3.2 Interpretation of Simple Sentences

Having built the framework for mathematical sentences (statements) in the previous section 3.1, we want to be able to *interpret* these statements within some relational system $\mathcal{S} = (S, \{P_i\}_{i\in I}, \{f_j\}_{j\in J})$. That is, we formalize whether a sentence is true or false.

Clearly, given the reader's assumed familiarity with symbols like $\forall$, $\implies$, and $\wedge$, these should be interpreted in the ordinary way as 'for all, implies,' and 'and.' Implicitly, we have also begun this interpretation process by underlining symbols in $L_{\mathcal{S}}$ ($\underline{P}$) to denote those same objects in $\mathcal{S}$ (P).

As with definition 3.2 for terms, we define interpretability starting from constant terms and build up to functions.

Definition 3.4 (Interpretability)**.** A constant term is *interpretable* in $\mathcal{S}$ if it is:

- a constant symbol $\underline{s}$ which names an element $s \in S$. In this case, it is to be interpreted as s. Or it is
- of the form $\underline{f}(\tau^1, \dots, \tau^n)$ where $\tau^1, \dots, \tau^n$ are interpretable in $\mathcal{S}$ by the above item with interpretations as $s^1, \dots, s^n \in S$. Furthermore, we require that the n-tuple $\langle s^1, \dots, s^n\rangle$ is in the domain of the function f which is named by $\underline{f}$. In this case we interpret $\underline{f}(\tau^1, \dots, \tau^n)$ as $f(s^1, \dots, s^n)$.

Importantly, from this definition, terms which are interpretable must be constant (contain no variable symbol). Another important point is that we cannot interpret expressions involving functions if they appear with an argument outside of their domain.

Example 3.4.1 (Interpretable expression)**.** The following term $3^2 + 9\pi\sqrt{e}$ is interpretable because 3 is interpretable as a constant in $\mathbb{R}$, thus 3^2 is also interpretable as it can be written as $\underline{f}(3)$ where $\underline{f}$ names the obvious squaring function. Similarly, $9, \pi$, and e are interpretable and so their product is interpretable. Combining this with 3^2 by addition gives interpretability of the entire term.

Example 3.4.2 (Not interpretable expression)**.** The term $5 + \log(0)$ is not interpretable because 0 is not in the domain of the logarithm.

Definition 3.5 (True sentences)**.** We will define what it means for sentences to be *true* (or to *hold*) in $\mathcal{S}$ by the following procedure:

1. The atomic sentence $\underline{P}\langle\tau^1, \dots, \tau^n\rangle$ is *true* (or *holds*) in $\mathcal{S}$ if both the following are true:
   - each of the terms τ^i, $i = 1, \dots, n$ are interpretable in $\mathcal{S}$ as elements s^i, $i = 1, \dots, n$ in S respectively and
   - $\langle s^1, \dots, s^n\rangle \in P$ (or equivalently $P\langle s^1, \dots, s^n\rangle$) where P is the relation on S named by $\underline{P}$.
2. The sentence:
$$(\forall x_1)(\forall x_2)\dots(\forall x_n)\left[\bigwedge_{i=1}^{k} \underline{P}_i\langle\tilde{\tau}_i\rangle \implies \bigwedge_{j=1}^{l} \underline{Q}_j\langle\tilde{\sigma}_j\rangle\right]$$
   is *true* (or *holds*) in $\mathcal{S}$ if upon replacing each variable symbol $x_1, \dots, x_n$ with constant symbols $\underline{s}^1, \dots, \underline{s}^n$, when all $\underline{P}_i\langle\tau_i^1, \dots, \tau_i^{n_i}\rangle$ are true for all $i = 1, \dots, k$ then all $\underline{Q}_j\langle\sigma_j^1, \dots, \sigma_j^{n_j}\rangle$ are true for all $j = 1, \dots, l$.

Here, we use the same convention as in definition 3.3 i.e. $\tilde{\tau}_i = \langle \tau_i^1, \ldots, \tau_i^{n_i} \rangle$.

This definition for interpreting sentences in $\mathcal{S}$ as true or not is simply a formalization of what the reader is familiar with.

Example 3.5.1 (Product of positive reals is positive)**.**

$$(\forall x)(\forall y)[x > 0 \wedge y > 0 \implies xy > 0]$$

Example 3.5.2 (Importance of interpretability)**.** Consider the following sentence:

$$(\forall x)[\underline{\mathbb{R}}\langle x \rangle \implies \sqrt{x} \geq 0]$$

This sentence is not true in $\mathcal{R}$ because $\sqrt{x}$ is not defined for all real numbers x.

One immediate concern in the language we are dealing with is treating informal statements involving 'not', 'or', or 'there exists.' Our language can translate 'for all', 'and', and 'implies' very simply.

Incorporating negation (statements involving 'not') attached to an n-ary relation P on S is handled by using the complement $P^{\complement}$ of P. For example, the true statement '1 is not equal to 0' can be expressed by the sentence $\underline{E^{\complement}}\langle 1, 0 \rangle$ where E is the relation $\{\langle x, y \rangle \in \mathbb{R}^2 \mid x = y\}$ and it has complement $E^{\complement} = \{\langle x, y \rangle \in \mathbb{R}^2 \mid x \neq y\}$. Interpretability is crucial here because if a term in $\underline{P}\langle \tau^1, \ldots, \tau^n \rangle$ is not interpretable in $\mathcal{S}$, then neither $\underline{P}\langle \tau^1, \ldots, \tau^n \rangle$ nor $\underline{P^{\complement}}\langle \tau^1, \ldots, \tau^n \rangle$ are interpretable and hence, not true in $\mathcal{S}$.

Incorporating 'or' can be tedious. However, 'or' can be thought of as the negation of 'and.' Since the previous paragraph addressed how to deal with 'not' and we already have a way of translating 'and,' this follows by taking appropriate sequences of negation and appealing to De Morgan's rules.

Incorporating 'there exists' is done by the use of so-called *Skolem functions.* We will not explore the intricacies of such functions, but the curious reader may find more information in Robinson [2] Consider as an example the true statement 'For each positive x in $\mathbb{R}$, there exists a y in $\mathbb{R}$ such that $y^2 = x$.' This statement asserts the existence of a special function ψ of the variable x whose domain is the positive reals such that $\psi(x)^2 = x$. ψ is a Skolem function for this statement and a contracted mathematical statement using it would be 'For all $x \geq 0$, $x = \psi(x)^2$,' and a translation to a simple sentence looks:

$$(\forall x)[\underline{\mathbb{R}}\langle x \rangle \wedge x \geq 0 \implies x = \underline{\psi}(x)^2]$$

We now have all the tools to formally state the transfer principle for simple sentences in $L_{\mathcal{R}}$.

3.3 The Transfer Principle

The previous section 3.2 built the formalisms required for the statement of the transfer principle. As remarked in earlier sections 2.3, we are working towards *-transforming sentences in $L_{\mathcal{R}}$. We adopt a few conventions as follows analogously to the beginning of this chapter:

1. For $\underline{r}$ a name in $L_{\mathcal{R}}$ of $r \in \mathbb{R}$, $\underline{r}$ is also a name in $L_{^*\mathbb{R}}$ of $^*r \in {}^*\mathbb{R}$ (recalling that we identify r and $^*r$ in $^*\mathbb{R}$).

2. For $\underline{P}$ a name in $L_{\mathbb{R}}$ of the relation P on $\mathbb{R}$, $^*\underline{P}$ is also a name in $L_{^*\mathbb{R}}$ of the relation $^*P$ on $^*\mathbb{R}$.

3. For $\underline{f}$ a name in $L_{\mathbb{R}}$ of the function f on $\mathbb{R}$, $^*\underline{f}$ is also a name in $L_{^*\mathbb{R}}$ of the function $^*f$ on $^*\mathbb{R}$.

Recall that, for an n-ary relation P, its *-transform is given in definition 2.12 as:

$${}^*P = \{\langle \mathbf{r}^1, \ldots, \mathbf{r}^n \rangle \in ({}^*\mathbb{R})^n \mid \text{For } \mathbf{r}^k = [\langle r_1^k, r_2^k, \ldots \rangle],\ P\langle r_i^1, \ldots, r_i^n \rangle \text{ holds a.e.}\}$$

We now work towards defining the *-transform of sentences for which we start with terms:

Definition 3.6 (*-transform of terms)**.** The **-transform of terms* is defined successively by the following:

1. For τ a constant or variable symbol, its *-transform is $^*\tau = \tau$.

2. For constants $\tau^1, \ldots, \tau^n$ and a function f with n arguments, the *-transform of $\tau = \underline{f}(\tau^1, \ldots, \tau^n)$ is ${}^*\tau = {}^*\underline{f}({}^*\tau^1, \ldots, {}^*\tau^n)$ $(= {}^*\underline{f}(\tau^1, \ldots, \tau^n)$ by item 1).

Example 3.6.1. The term $\underline{f}(x, \underline{g}(3y + e))$ has *-transform ${}^*\underline{f}(x, {}^*\underline{g}(3y + e))$.

Definition 3.7 (*-transform of simple sentences)**.** For Φ a simple sentence in $L_{\mathcal{R}}$, we define the **-transform of* Φ as ${}^*\Phi$ given by the following rules:

1. For Φ an atomic sentence $\underline{P}\langle\tau^1, \ldots, \tau^n\rangle$, the *-transform ${}^*\Phi$ is ${}^*\underline{P}\langle{}^*\tau^1, \ldots, {}^*\tau^n\rangle$ (which is the same as ${}^*\underline{P}\langle\tau^1, \ldots, \tau^n\rangle$ by item 1 of definition 3.6 above).

2. For Φ a sentence given by

$$(\forall x_1) \ldots (\forall x_n) \left[\bigwedge_{i=1}^{k} \underline{P}_i\langle\tilde{\tau}_i\rangle \implies \bigwedge_{j=1}^{l} \underline{Q}_j\langle\tilde{\sigma}_j\rangle \right]$$

   the *-transform ${}^*\Phi$ is the sentence

$$(\forall x_1) \ldots (\forall x_n) \left[\bigwedge_{i=1}^{k} {}^*\underline{P}_i\langle{}^*\tilde{\tau}_i\rangle \implies \bigwedge_{j=1}^{l} {}^*\underline{Q}_j\langle{}^*\tilde{\sigma}_j\rangle \right]$$

   Here, we use ${}^*\tilde{\tau} = \langle{}^*\tau^1, \ldots, {}^*\tau^n\rangle$ when $\tilde{\tau} = \langle\tau^1, \ldots, \tau^n\rangle$. Of course, as repeatedly remarked, item 1 of definition 3.6 insists that we identify ${}^*\tau^i = \tau^i$ so we may suppress the use of * in later symbolic manipulations.

With these definitions in tow, the formal statement and proof of the celebrated transfer principle can be made:

Theorem 3.3.1 (Transfer Principle)**.** *If Φ is a simple sentence in $L_{\mathcal{R}}$ which is true in $\mathcal{R}$, then ${}^*\Phi$ is also true in ${}^*\mathcal{R}$.*

Remark 3.3.1. *For the rest of this document, we will say that a mathematical statement ${}^*\Phi$ about a nonstandard structure ${}^*\mathcal{S}$* follows by transfer *from a sentence Φ in $L_{\mathcal{S}}$ to indicate the truth of ${}^*\Phi$ follows by the transfer principle in the above theorem 3.3.1 from the fact that Φ is true in $\mathcal{S}$.*

3.4 Proof of Transfer Principle

This section is devoted to the proof of the transfer principle in theorem 3.3.1. Recall that the relations and functions in ${}^*\mathcal{R}$ are the extensions of relations and functions in $\mathcal{R}$. We will start with denoting, for each constant term τ (interpretable or not) in $L_{\mathcal{R}}$, a sequence $\langle T_\tau(n)\rangle$ or T_τ defined by the following temporary definition.

'Definition' 3.2 (T_τ for constant terms τ)**.** For a constant term τ (interpretable or not in $L_{\mathcal{R}}$, define $\langle T_\tau(n)\rangle$ or T_τ by:

1. If τ is a constant term in $L_{{}^*\mathcal{R}}$ which names $r \in {}^*\mathbb{R}$, choose a sequence $\langle r_n\rangle \subset \mathbb{R}$ such that $r = [\langle r_n\rangle]$ (if $r \in \mathbb{R}$, choose the canonical sequence $r_n = r$). Let $\underline{r_n}$ be names in $L_{\mathcal{R}}$ for $r_n \in \mathbb{R}$ for each $n \in \mathbb{N}$. We set:

$$T_\tau(n) = \begin{cases} \underline{r_n} & r \in {}^*\mathbb{R} \setminus \mathbb{R} \\ \tau & r \in \mathbb{R} \end{cases}$$

2. If $\tau = {}^*\underline{f}(\tau^1, \ldots \tau^k)$ where $\underline{f}$ is a name of the function f on $\mathbb{R}$ of k variables and the τ^i are constant terms in $L_{{}^*\mathcal{R}}$ for each $i = 1, \ldots, k$, then we set:

$$T_\tau(n) = \underline{f}\left(T_{\tau^1}(n), \ldots, T_{\tau^k}(n)\right)$$

The conditions given by items 1 and 2 in the above temporary definition 3.2 successively define T_τ for all constant terms τ in $L_{^*\mathcal{R}}$. We now prove a simple form of Lŏs' theorem (for which a more general statement can be found in Robinson [2] or Hurd and Loeb [1]) with this object:

Theorem 3.4.1 (Simple form of Lŏs). *Let $\tau, \tau^1, \ldots, \tau^k$ be constant terms in $L_{^*\mathcal{R}}$, $\langle r_n \rangle$ be a sequence of real numbers, and $^*\underline{P}\langle \tau^1, \ldots, \tau^k \rangle$ be an atomic sentence in $L_{^*\mathcal{R}}$:*

1. *τ is interpretable in $^*\mathcal{R}$ and names $[\langle r_n \rangle] \iff T_\tau(n)$ is a.e. interpretable in $\mathcal{R}$ and names r_n a.e. ($\exists U \in \mathcal{U}$ s.t. $\forall n \in U$, $T_\tau(n)$ is interpretable and names r_n).*
2. *$^*\underline{P}\langle \tau^1, \ldots, \tau^k \rangle$ holds in $^*\mathcal{R} \iff \underline{P}\langle T_{\tau^1}(n), \ldots, T_{\tau^k}(n) \rangle$ holds a.e. in $\mathcal{R}$.*

Proof. 1. We prove item 1 by inducting on the complexity of terms recalling definitions 3.2 and 3.6 which defined terms and their *-transforms, respectively.

   A) For $\tau = \underline{c}$ a constant naming an element of $^*\mathbb{R}$, we have that $\tau = \underline{c}$ names $[\langle r_n \rangle] \iff T_\tau(n)$ names r_n a.e. (indeed, everywhere) by definition of T_τ in item 1 of definition 3.2 above.

   B) Let $\tau = {}^*\underline{f}(\tau^1, \ldots, \tau^k)$ with $\underline{f}$ naming the function f of k variables. By the above item 1A, we know that τ^j is interpretable in $^*\mathcal{R}$ and names $[\langle r_n^j \rangle]$ (for k sequences $\langle r_n^j \rangle \subset \mathbb{R}$, $j = 1, \ldots, k$) $\iff T_{\tau^j}(n)$ is a.e. interpretable and names r_n^j a.e.

   Let $\langle s_n \rangle \subset \mathbb{R}$ be such that τ is interpretable in $^*\mathcal{R}$ and names $[\langle s_n \rangle]$. We need to show item 1 of the theorem which asserts that this happens if and only $T_\tau(n)$ is a.e. interpretable in $\mathcal{R}$ and names s_n. This is done by the following sequence of equivalences starting from the beginning of this paragraph:

      a) $\tau = {}^*\underline{f}(\tau^1, \ldots, \tau^k)$ is interpretable in $^*\mathcal{R}$ and names $[\langle s_n \rangle]$.

      b) There exists, for each $j = 1, \ldots, k$, $[\langle r_n^j \rangle] \in {}^*\mathbb{R}$ such that τ^j is interpretable in $^*\mathcal{R}$ as $[\langle r_n^j \rangle]$. Furthermore, the k-tuple $\big\langle [\langle r_n^1 \rangle], \ldots, [\langle r_n^k \rangle] \big\rangle$ is in the domain of $^*f$ and $^*f\left([\langle r_n^1 \rangle], \ldots, [\langle r_n^k \rangle]\right) = [\langle s_n \rangle]$. The equivalence from the previous step to this follows from defintion 3.4 of interpretability.

      c) There exists, for each $j = 1, \ldots, k$, $\langle r_n^j \rangle \subset \mathbb{R}$ and a set $U \in \mathcal{U}$ such that, for every $m \in U$, $T_{\tau^j}(m)$ is interpretable as r_m^j. Furthermore, the k-tuple $\langle r_m^1, \ldots, r_m^k \rangle$ is in the domain of f and $f(r_m^1, \ldots, r_m^k) = s_m$. The equivalence from the previous step to this follows from definition 2.7 which defined the notion of a.e. equality applied to $^*f\left([\langle r_n^1 \rangle], \ldots, [\langle r_n^k \rangle]\right) = [\langle s_n \rangle]$ and we have used item 1 of definition 3.2 for each $\langle r_m^j \rangle$ to define T_{τ^j}.

      d) $\underline{f}(T_{\tau^1}(n), \ldots, T_{\tau^k}(n))$ is a.e. interpretable as s_n in $\mathcal{R}$. The equivalence from the previous step to this follows trivially as $f(r_m^1, \ldots, r_m^k) = s_m$ and $T_{\tau^j}(m)$ is a.e. interpretable as r_m^j for every $j = 1, \ldots, k$.

      e) $T_\tau(n)$ is a.e. interpretable as s_n in $\mathcal{R}$. The equivalence from the previous step to this follows by item 2 of definition 3.2 as we set $T_\tau(n) = \underline{f}(T_{\tau^1}(n), \ldots, T_{\tau^k}(n))$.

   Hence, item 1 of the theorem is true by induction.

2. We prove item 2 of the theorem similarly as follows. Let $\underline{P}$ name the k-ary relation P on $\mathbb{R}$. We give a sequence of equivalences starting from the left-hand side of item 2 of the theorem:

   a) $^*\underline{P}\langle \tau^1, \ldots, \tau^k \rangle$ holds in $^*\mathcal{R}$.

   b) There exists, for each $j = 1, \ldots, k$, $[\langle r_n^j \rangle] \in {}^*\mathbb{R}$ such that τ^j is interpretable in $^*\mathcal{R}$ as $[\langle r_n^j \rangle]$. Furthermore, the k-tuple $\big\langle [\langle r_n^1 \rangle], \ldots, [\langle r_n^k \rangle] \big\rangle$ is in $^*P$. As with item 1(B)b above, this step follows equivalently from the previous by definition 3.4 of interpretability.

   c) There exists, for each $j = 1, \ldots, k$, $\langle r_n^j \rangle \subset \mathbb{R}$ and a set $U \in \mathcal{U}$ such that, for every $m \in U$, $T_{\tau^j}(m)$ is interpretable as r_m^j. Furthermore, the k-tuple $\langle r_m^1, \ldots, r_m^k \rangle$ is in P. As with item 1(B)c above, this step follows equivalently from the previous by item 1 of definition 3.2 applied on each $\langle r_m^j \rangle$ to define T_{τ^j}.

   d) $\underline{P}\langle T_{\tau^1}(n), \ldots, T_{\tau^k}(n) \rangle$ holds a.e. in $\mathcal{R}$. The equivalence of the previous step to this follows trivially as each $T_{\tau^j}(m)$ is a.e. interpretable as r_m^j and $P\langle r_m^1, \ldots, r_m^k \rangle$ (meaning $\langle r_m^1, \ldots, r_m^k \rangle \in P$) a.e.

Hence, item 2 of the theorem holds as well. This completes the proof. □

With the simple Lŏs theorem 3.4.1 in tow, we are now able to prove the transfer principle stated in theorem 3.3.1.

Proof (Transfer Principle). Suppose Φ is an atomic sentence which holds in $\mathcal{R}$. Theorem 3.4.1 item 2 immediately shows that ${}^*\Phi$ holds in ${}^*\mathcal{R}$. Suppose now that Φ is a simple sentence in $\mathcal{R}$, so it will have the form:

$$(\forall x_1)\dots(\forall x_n)\left[\bigwedge_{i=1}^{k}\underline{P}_i\langle\tau_1^i,\dots,\tau_{p_i}^i\rangle \implies \bigwedge_{j=1}^{l}\underline{Q}_j\langle\sigma_1^j,\dots,\sigma_{q_j}^j\rangle\right]$$

Let ${}^*\tau_s^t$ and ${}^*\sigma_s^t$ be the respective $*$-transforms of τ_s^t and σ_s^t where s and t run over the appropriate indices. Now, replace the variables $x_1,\dots,x_n$ in ${}^*\tau_s^t$ and ${}^*\sigma_s^t$ with constant symbols $\underline{r}_1,\dots,\underline{r}_n$ from $L_{{}^*\mathcal{R}}$. We seek to show that the truth of ${}^*\underline{P}_i\langle{}^*\tau_1^i,\dots,{}^*\tau_{p_i}^i\rangle$ in ${}^*\mathcal{R}$ for each $i=1,\dots,k$ guarantees the truth of ${}^*\underline{Q}_j\langle{}^*\sigma_1^j,\dots,{}^*\sigma_{q_j}^i\rangle$ in ${}^*\mathcal{R}$ for each $j=1,\dots,l$.

Assume that ${}^*\underline{P}_i\langle{}^*\tau_1^i,\dots,{}^*\tau_{p_i}^i\rangle$ holds in ${}^*\mathcal{R}$ for each $i=1,\dots,k$. Theorem 3.4.1 item 2 guarantees the existence of a set $U\in\mathcal{U}$ such that $\underline{P}_i\langle T_{\tau_1^i}(n),\dots,T_{\tau_{p_i}^i}(n)\rangle$ holds in $\mathcal{R}$ for each $i=1,\dots,k$ when $n\in U$ (apply item 2 of theorem 3.4.1 to each atomic sentence to get $U_i\in\mathcal{U}$ then take the intersection to obtain $U\in\mathcal{U}$ by item 2 from the definition 2.1 of filters). By assumption, since Φ holds in $\mathcal{R}$, we have that $\underline{Q}_j\langle T_{\sigma_1^j}(n),\dots,T_{\sigma_{q_j}^j}(n)\rangle$ also holds in $\mathcal{R}$ for each $j=1,\dots,l$ when $n\in U$. Applying theorem 3.4.1 item 2 again, ${}^*\underline{Q}_j\langle{}^*\sigma_1^j,\dots,{}^*\sigma_{q_j}^j\rangle$ holds in ${}^*\mathcal{R}$ for each $j=1,\dots,l$. This is what we wanted to show and so the proof is complete. □

Chapter 4

Transition to Real Analytic Applications

The transfer principle in theorem 3.3.1 is a powerful tool as it condenses many proofs of standard analytic results. On the one hand, the transfer principle makes the flimsy arguments of Leibniz in the early days of (differential) calculus rigorous. On the other hand, one can keep the spirit of these arguments and appeal to the transfer principle to make logically sound conclusions.

Before we get into infinite numbers, infinitesimals, and nonstandard proofs of standard analytic results, we stay within the framework of set theory to establish some preliminaries in section 4.1. The next section 4.2 lays the foundations for real analysis using nonstandard terminology. We will not explore these topics but point the interested reader to Hurd and Loeb [1].

4.1 Preliminary applications of the Transfer Principle

Proposition 4.1.1 (Transfer of relations)**.** *Let P be an n-ary relation on $\mathbb{R}$ with χ_P its associated characteristic function on $\mathbb{R}^n$. Then ${}^*P$ is an extension of P (in the sense that $P\langle r\rangle \implies {}^*P\langle {}^*r\rangle$). Furthermore, ${}^*\chi_P = \chi_{{}^*P}$ and ${}^*\left(P^{\complement}\right) = ({}^*P)^{\complement}$.*

Remark 4.1.1. *Proposition 2.3.1 already established ${}^*\chi_P = \chi_{{}^*P}$ however we will reprove this using the transfer principle.*

Proof. To show the extension result, suppose $\langle c^1, \ldots, c^n\rangle \in P$. This is equivalent to the following atomic sentence being true in $\mathcal{R}$: $\underline{P}\langle c^1, \ldots, c^n\rangle$. By transfer (taking $*$-transforms) ${}^*\underline{P}\langle c^1, \ldots, c^n\rangle$ is true in ${}^*\mathcal{R}$ (remembering that we identify constant and variable symbols without using the $*$ symbol [definition 3.6 item 1]). So ${}^*P$ is an extension of P.

Consider now the following four sentences and their transforms:

$$(\forall x_1)\ldots(\forall x_n)\left[\underline{\mathbb{R}}\langle x_1\rangle \wedge \cdots \wedge \underline{\mathbb{R}}\langle x_n\rangle \implies \underline{\chi_P}(x_1, \ldots, x_n) = \underline{\chi_P}(x_1, \ldots, x_n)\right] \tag{4.1}$$

$$(\forall x_1)\ldots(\forall x_n)\left[\underline{\chi_P}(x_1, \ldots, x_n) \neq 1 \implies \underline{\chi_P}(x_1, \ldots, x_n) = 0\right] \tag{4.2}$$

$$(\forall x_1)\ldots(\forall x_n)\left[\underline{P}\langle x_1, \ldots, x_n\rangle \iff \underline{\chi_P}(x_1, \ldots, x_n) = 1\right] \tag{4.3}$$

$$(\forall x_1)\ldots(\forall x_n)\left[\underline{P^{\complement}}\langle x_1, \ldots, x_n \iff \underline{\chi_P}(x_1, \ldots, x_n) = 0\right] \tag{4.4}$$

The transform of sentence 4.1 establishes that ${}^*\chi_P$ is a well-defined function (on ${}^*\mathbb{R}^n$). The transform of sentence 4.2 establishes that ${}^*\chi_P$ takes only the values 0 or 1. Finally, the transform of sentence 4.3 establishes that ${}^*\chi_P = \chi_{{}^*P}$ because we have:

$$(\forall x_1)\ldots(\forall x_n)\left[{}^*\underline{P}\langle x_1, \ldots, x_n\rangle \iff {}^*\underline{\chi_P}(x_1, \ldots, x_n) = 1\right]$$

But the left-hand side is equivalent to $\underline{\chi_{{}^*P}}(x_1, \ldots, x_n) = 1$. Combining this with the transform of sentence 4.2 gives equality again when they take on the value 0.

By the same reasoning, the transforms of the last three sentences gives ${}^*\left(P^{\mathsf{C}}\right) = ({}^*P)^{\mathsf{C}}$. $\square$

Proposition 4.1.2 (Transfer of functions)**.** *If f is a function of n variables on $\mathbb{R}$, then ${}^*f$ is also a function of n variables on ${}^*\mathbb{R}^n$ and extends f. Furthermore, ${}^*(domf) = dom{}^*f$ and ${}^*(rangef) = range{}^*f$.*

Proof. ${}^*f$ is a function by transforming the following sentence (recalling functions are just relations of a certain type):

$$(\forall x_1)\dots(\forall x_n)(\forall y)(\forall z)\left[\underline{f}\langle x_1,\dots,x_n,y\rangle \wedge \underline{f}\langle x_1,\dots,x_n,z\rangle \implies y = z\right]$$

${}^*f$ extends f by proposition 4.1.1 (again, thinking of functions as relations of a certain form).

Consider the following sentence:

$$(\forall x_1)\dots(\forall x_n)\left[\underline{\mathrm{dom}f}\langle x_1,\dots,x_n\rangle \iff \underline{\mathbb{R}}\langle f(x_1,\dots,x_n)\rangle\right]$$

By transfer, we establish the following chain of equivalence within the square brackets:

$${}^*\mathrm{dom}f\langle x_1,\dots,x_n\rangle \iff {}^*f(x_1,\dots,x_n) \in {}^*\mathbb{R} \iff \mathrm{dom}{}^*f\langle x_1,\dots,x_n\rangle$$

Hence ${}^*(\mathrm{dom}f) = \mathrm{dom}{}^*f$.

To complete the proof and establish commutativity of * and range, we need to use a Skolem function ψ defined on range$f$:

$$(\forall x)(\forall y)\left[\underline{f}(x) = y \implies \underline{\mathrm{range}f}\langle y\rangle\right] \tag{4.5}$$

$$(\forall y)\left[\underline{\mathrm{range}f}\langle y\rangle \implies \underline{f}(\psi(y)) = y\right] \tag{4.6}$$

By transfer, sentence 4.5 gives $\mathrm{range}{}^*f \subset {}^*(\mathrm{range}f)$ while sentence 4.6 gives the reverse inclusion. Here, we abuse notation slightly by treating $x \in \mathbb{R}^n$. $\square$

Proposition 4.1.3 (Transfer of set operations)**.** *We have the following statements for A, A_i, B sets in $\mathbb{R}^n$ where $i \in I$ some index set.*

1. ${}^*\emptyset = \emptyset$
2. ${}^*(A \cup B) = {}^*A \cup {}^*B$ *and* ${}^*(A \cap B) = {}^*A \cap {}^*B$
3. $\bigcup_{i\in I} {}^*A_i \subset {}^*\left[\bigcup_{i\in I} A_i\right]$ *and* $\bigcap_{i\in I} {}^*A_i \supset {}^*\left[\bigcap_{i\in I} A_i\right]$

Proof.

1. $\chi_\emptyset$ is identically 0, so by proposition 4.1.1, we have that ${}^*\chi_\emptyset = \chi_{{}^*\emptyset}$ so ${}^*\emptyset = \emptyset$ establishing item 1 of the result.
2. To establish item 2, we show first the result for intersection. The transfer of the following sentence establishes the result for finite intersection:

$$(\forall x)\left[\underline{(A\cap B)}\langle x\rangle \iff \underline{A}\langle x\rangle \wedge \underline{B}\langle x\rangle\right]$$

   For finite unions, we appeal to De Morgan's laws by transferring the following sentence:

$$(\forall x)\left[\underline{(A\cup B)^{\mathsf{C}}}\langle x\rangle \iff \underline{A^{\mathsf{C}}}\langle x\rangle \wedge \underline{B^{\mathsf{C}}}\langle x\rangle\right]$$

3. We have that $\forall j \in I$:

$$(\forall x)\left[x \in \underline{A_j} \implies x \in \underline{\left[\bigcup_{i\in I} A_i\right]}\right]$$

   By transfer of this sentence, the result for unions is shown. Intersection follows similarly.

$\square$

Proposition 4.1.4 (Transfer of operations and absolute value)**.** *Let f and g be functions of n variables on $\mathbb{R}$. For $x = \langle x_1, \ldots, x_n \rangle$, we have:*

1. *If $x \in dom^* f \cap dom^* g$, then:*

$$^*(f+g)(x) = {}^*f(x) + {}^*g(x), \quad \textit{and} \ {}^*(f \cdot g)(x) = {}^*f(x) {}^*g(x)$$

2. *If $x \in domf$, then:*

$$^*|f(x)| = |^*f(x)|$$

Proof.

1. Addition follows by transfer of the following sentence:

$$(\forall x_1) \ldots (\forall x_n) \left[\underline{\text{dom}f \cap \text{dom}g} \langle x_1, \ldots, x_n \rangle \implies \underline{(f+g)}(x_1, \ldots, x_n) = \underline{f}(x_1, \ldots, x_n) + \underline{g}(x_1, \ldots, x_n) \right]$$

We have used $^*(\text{dom}f \cap \text{dom}g) = \text{dom}^* f \cap \text{dom}^* g$ by the previous proposition 4.1.3 item 2 and proposition 4.1.2. Multiplication follows by the obvious adjustment to the above sentence.

2. Simply take the transfer of the following sentences:

$$(\forall y) \left[y \geq 0 \implies |y| = y \right]$$

$$(\forall y) \left[y < 0 \implies |y| = -y \right]$$

Here, y takes the place of $f(x)$ for $x \in \text{dom}f$ and so the above is sufficient given the universal quantifier.

$\square$

4.2 Infinitesimals, Infinite Numbers, and the Standard Part Map

In this section, we explore some of the algebraic properties of the hyperreals. As a motivation, we want to make the following statement rigorous: If $\epsilon > 0$ is an infinitesimal number, then $\forall x$:

$$x + \epsilon \simeq x$$

for some meaningful interpretation of infinitesimal and $\simeq$.

Definition 4.1 ((Non) standard numbers)**.** A hyperreal number is called *standard* if the number is in $\mathbb{R}$ and *nonstandard* if it is not standard. An n-tuple $\langle a^1, \ldots, a^n \rangle$ is standard if each component is standard and nonstandard otherwise.

Definition 4.2 (Finite, infinite, and infinitesimal)**.** A number $s \in {}^*\mathbb{R}$ is:

1. *finite* if there is a standard natural number n such that $|s| < n$
2. *infinite* if for all standard natural numbers n, we have $|s| > n$
3. *infinitesimal* if for all standard natural numbers n, we have $|s| < \frac{1}{n}$

Example 4.2.1. Recall in the motivation in section 1.1, we had (one member of the equivalence class of) the following nonstandard number:

$$\omega = \left[\left\langle 1, \frac{1}{2}, \dots, \frac{1}{n}, \dots \right\rangle\right]$$

As mentioned then, this number satisfies an intuitive and imprecise notion of infinitesimality. According to the precise definition of infinitesimality given above in item 3 of definition 4.2, ω is indeed infinitesimal.

Similarly, consider the reciprocal of ω:

$$\frac{1}{\omega} = [\langle 1, 2, \dots, n, \dots \rangle]$$

This is an infinite number. More generally, one can easily show from the definition that the reciprocal of any infinite number is infinitesimal.

Theorem 4.2.1 (Ring theoretic properties of ${}^*\mathbb{R}$)**.** *Certain subset of ${}^*\mathbb{R}$ possess ring theoretic properties which we explain below:*

1. *The finite and infinitesimal numbers in ${}^*\mathbb{R}$ are subrings of ${}^*\mathbb{R}$. That is, they are closed under the usual ring operations of ${}^*\mathbb{R}$.*
2. *The infinitesimals are an ideal of the finite numbers. That is, the product of an infinitesimal number and a finite number is infinitesimal.*

Proof. In what follows, let ϵ and δ be infinitesimals and $n \in \mathbb{N}$.

1. By definition, we have both $|\epsilon| < \frac{1}{2n}$ and $|\delta| < \frac{1}{2n}$. Easy applications of the triangle inequality yield $|\epsilon + \delta| < \frac{1}{n}$. Similarly, both $|\epsilon| < \frac{1}{\sqrt{n}}$ and $|\delta| < \frac{1}{\sqrt{n}}$ so their product $|\epsilon\delta| < \frac{1}{n}$. We have implicitly used the familiar properties of the absolute value which hold by transfer. So the infinitesimals are closed under the ring operations. A similar argument shows the same result for finite numbers.
2. Fix x finite. By definition, there is some natural number m such that $|x| < m$. We have that $|\epsilon| < \frac{1}{mn}$. Thus their product has magnitude $|\epsilon x| < \frac{1}{n}$ and since $n \in \mathbb{N}$ was arbitrary, ϵx is infinitesimal.

$\square$

Next, we give two important equivalence relations which are crucial to the nonstandard generalizations of convergence and continuity.

Definition 4.3. Let x and y be elements of ${}^*\mathbb{R}$. They are said to be:

1. *near* or *infinitesimally close* if $x - y$ is infinitesimal. We write $x \simeq y$ if x and y are near and $x \not\simeq y$ otherwise. The *monad* of x is the set:

$$m(x) := \{y \in {}^*\mathbb{R} | x \simeq y\}$$

2. *finitely close* if $x - y$ is finite. We write $x \sim y$ if x and y are finitely close and $x \nsim y$ otherwise. The *galaxy* of x is the set:

$$G(x) := \{y \in {}^*\mathbb{R} | x \sim y\}$$

Remark 4.2.1. *In other literature, 'halo' is used instead of monad. As Goldblatt [3] remarks, this is the popular preference of terminology for the French school of Nonstandard Analysis founded by George Reeb.*

$m(0)$ is the set of infinitesimals and $G(0)$ is the set of finite numbers. It is easy to show that two monads are either equal or disjoint and thus $\simeq$ is an equivalence relation on ${}^*\mathbb{R}$. Similarly, two galaxies are either equal or disjoint so that $\sim$ is also an equivalence relation. Furthermore, for any $x \in {}^*\mathbb{R}$, the monad and galaxy of x are simply translates by x of the monad and galaxy of 0.

Theorem 4.2.2 (Standard Part map)**.** *Every finite number is infinitesimally close to a unique standard number. This unique standard number is called the* standard part *of ρ written $st(\rho)$ or $^\circ\rho$. So the standard part is a map:*

$$st : G(0) \to \mathbb{R}$$

Proof. For fixed $\rho \in {}^*\mathbb{R}$, define the following sets:

$$A := \{x \in \mathbb{R} | \rho \leq x\}, \quad B := \{x \in \mathbb{R} | \rho > x\}$$

Since ρ is finite, there is a standard natural numbers n such that $-n < \rho < n$. Thus, the set B is non-empty (contains $-n$) and has an upper bound. By the completeness axiom, let r be the supremum of B. Now, fix $\epsilon > 0$ in $\mathbb{R}$. By definition of r, we have that $r + \epsilon \in A$ and $r - \epsilon \in B$. This shows that $|r - \rho| \leq \epsilon$. Since $\epsilon > 0$ was arbitrary, this shows that $r \simeq \rho$. This is the existence claim.

For uniqueness, let $r_1 \simeq \rho$ where r_1 is another standard real number. Now, by the triangle inequality, for any arbitrary $\epsilon > 0$:

$$|r_1 - r| \leq |r_1 - \rho| + |\rho - r| \leq 2\epsilon$$

Since both r_1 and r are standard real numbers and $\epsilon > 0$ was arbitrary, this shows that $r_1 = r$. □

Remark 4.2.2. *In other literature, 'shadow' is used instead of standard part. This is remarked by Goldblatt [3] as another popular preference for the French school of Nonstandard Analysis.*

Clearly, st is surjective because $^\circ r = r$ for every $r \in \mathbb{R}$. It also preserves algebraic structure by the following theorem:

Theorem 4.2.3. *For $x, y \in G(0)$, we have that st is an order-preserving ring homomorphism:*

1. $^\circ(x \pm y) = {}^\circ x \pm {}^\circ y$
2. $^\circ(xy) = ({}^\circ x)({}^\circ y)$
3. *When* $^\circ y \neq 0$, $^\circ\left(\frac{x}{y}\right) = \frac{^\circ x}{^\circ y}$
4. $x \leq y \implies {}^\circ x \leq {}^\circ y$

Proof. Fix $x, y \in G(0)$. Firstly, by definition of st, we have the existence of infinitesimals ϵ, δ such that $x = {}^\circ x + \epsilon, y = {}^\circ y + \delta$. We prove the result item by item with this.

1. $x \pm y = ({}^\circ x \pm {}^\circ y) + (\epsilon \pm \delta)$. By thereom 4.2.1 item 1, $\epsilon + \delta$ is infinitesimal so item 1 of the theorem is established.
2. Argue as with item 1 above.
3. Argue as with item 1 above by establishing $^\circ y {}^\circ\left(\frac{x}{y}\right) = {}^\circ x$.
4. Suppose $x \leq y$, this means that $^\circ x + \epsilon \leq {}^\circ y + \delta$. Thus, isolating $^\circ x$ on the left-hand side gives:

   $$^\circ x \leq {}^\circ y + (\delta - \epsilon) < {}^\circ y + \frac{1}{n}$$

   for any $n \in \mathbb{N}$, because $\delta - \epsilon$ is infinitesimal by theorem 4.2.1. Since the above is true for arbitrary $n \in \mathbb{N}$, we have that $^\circ x \leq {}^\circ y$.

□

Corollary 4.2.3.1. *The quotient field $G(0)/m(0)$ is isomorphic to the standard field* $\mathbb{R}$.

Proof. $m(0)$ is the kernel of the standard part map. The previous theorem 4.2.3 established that the standard part map is a ring homomorphism and one simply applies the first isomorphism theorem because st is surjective. □

Corollary 4.2.3.2. *For finite numbers x, x', y, y' such that $x \simeq x', y \simeq y'$, we have:*

1. $x \pm y \simeq x' \pm y'$
2. $xy \simeq x'y'$
3. $y \not\simeq 0$ *(and hence $y' \not\simeq 0$ as well)* $\implies \frac{x}{y} \simeq \frac{x'}{y'}$

Proof. Follows immediately by the previous results. $\square$

4.3 Hyperintegers

Before proceeding to the nonstandard proofs of standard analytic results, we formulate some notions about the *hyperintegers* and *hypernaturals* $^*\mathbb{Z}$ and $^*\mathbb{N}$, respectively. For motivation, consider the most basic definition of convergence of a sequence in analysis. This involves a statement along the lines of 'for any arbitrary $\epsilon > 0$, one can find a natural number n (which may be large) such that...'. The nonstandard treatment of convergence will condense statements such as this by looking only at infinite natural numbers.

Definition 4.4 (Infinite numbers of a set)**.** For a set $A \subset \mathbb{R}$, the *set of infinite numbers in* $^*A$ is the set:

$$^*A_\infty := {}^*A \cap ({}^*\mathbb{R} \setminus G(0))$$

This definition is completely consistent with that of infinite numbers given in item 2 of definition 4.2.

Theorem 4.3.1. *For an infinite set $A \subset \mathbb{N}$, $^*A$ contains natural numbers that are infinite. i.e. $^*A \cap {}^*\mathbb{N}_\infty \neq \emptyset$.*

Proof. Fix $n \in \mathbb{N}$, since A is an infinite subset of $\mathbb{N}$, there exists $a \in A$ such that $a \geq n$. This allows us to define a Skolem function $\psi : \mathbb{N} \to A$ such that $\psi(n) \geq n$ (ψ takes the place of a which depends on n). Thus, the following sentence is true in $\mathbb{R}$:

$$(\forall n)[\underline{\mathbb{N}}\langle n \rangle \implies \underline{A}\left\langle \underline{\psi}(n) \right\rangle \wedge \underline{\psi}(n) \geq n]$$

By transfer of this sentence, $^*\psi(n) \in {}^*A$ and $^*\psi(n) \geq n$ for all $n \in {}^*\mathbb{N}$. In particular, this is true for $n = \frac{1}{\omega} = [\langle 1, 2, 3, \dots \rangle] \in {}^*\mathbb{N}_\infty$. Hence, $^*\psi\left(\frac{1}{\omega}\right) \in {}^*A \cap {}^*\mathbb{N}_\infty$. $\square$

We now give some basic properties of the hyperintegers which follow from the transfer principle.

Proposition 4.3.1. *$^*\mathbb{Z}$ is a linearly ordered subring of $^*\mathbb{R}$.*

Proof. We need to check closure under addition and multiplication as well as the linear ordering. The closure properties follow by transfer of the following sentence stating that $\mathbb{Z}$ is a subring of $\mathbb{R}$:

$$(\forall x)(\forall y)[\underline{\mathbb{Z}}\langle x \rangle \wedge \underline{\mathbb{Z}}\langle y \rangle \implies \underline{\mathbb{Z}}\langle x + y \rangle \wedge \underline{\mathbb{Z}}\langle xy \rangle]$$

The linear ordering of $^*\mathbb{Z}$ is inherited from $^*\mathbb{R}$. $\square$

Proposition 4.3.2. *For every $x \in {}^*\mathbb{R}$ there exists $k \in {}^*\mathbb{Z}$ s.t. $k \leq x < k + 1$.*

Proof. In $\mathcal{R}$ there is a greatest integer (or floor) function $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$ which satisfies:

$$\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1, \forall x \in \mathbb{R}$$

Thus, the *-transform of the function function $^*\lfloor \cdot \rfloor : {}^*\mathbb{R} \to {}^*\mathbb{Z}$ satisfies, by the transfer principle:

$$^*\lfloor x \rfloor \leq x < {}^*\lfloor x \rfloor + 1, \forall x \in {}^*\mathbb{R}$$

$\square$

Corollary 4.3.1.1. *There are positive and negative infinite hyperintegers.*

Proof. This result could have been proved without the directly previous results by looking at $\pm\frac{1}{\omega} = \pm[\langle 1, 2, 3, \dots \rangle]$. Alternatively, looking at $x = [\langle e^j \rangle_{j\in\mathbb{N}}]$ which is positive infinite, find $k \in {}^*\mathbb{Z}$ from proposition 4.3.2 and see that $k + 1$ is guaranteed to be positive infinite by the linear ordering proven in proposition 4.3.1. Take $-(k + 1)$ to get a negative infinite hyperinteger. ☐

Corollary 4.3.1.2 (Archimedean property)**.** $\forall x \in {}^*\mathbb{R}, \exists n \in {}^*\mathbb{N}$ *s.t.* $|x| < n$

Proof. Take the $k \in {}^*\mathbb{Z}$ from proposition 4.3.2 and define $n := |k + 1|$. This hypernatural number is the desired number. ☐

Proposition 4.3.3. *For every $n \in {}^*\mathbb{Z}, n + 1$ is the smallest hyperinteger greater than n.*

Proof. Note, that the proposition is true if the '*' and 'hyper' were omitted in ${}^*\mathbb{Z}$ and hyperintegers of the statement. Hence apply the transfer principle to this statement made formal by the following sentence:

$$(\forall x)(\forall y)[\underline{\mathbb{Z}}\langle x\rangle \wedge \underline{\mathbb{Z}}\langle y\rangle \wedge (x \leq y \leq x + 1) \wedge (y \neq x) \implies (y = x + 1)]$$

☐

Corollary 4.3.1.3. ${}^*\mathbb{Z} \cap G(0) = \mathbb{Z}$. *Finite hyperintegers are ordinary integers. Here, we are abusing the notation with '=' to mean 'isomorphic to.'*

Proof. Fix k a finite hyperinteger. Then $^\circ k$ is a real integer. Furthermore, by the standard version of propsition 4.3.2, there exists $n \in \mathbb{Z}$ such that $n \leq {}^\circ k < n + 1$. Now, since $k = {}^\circ k + \epsilon$ for some infinitesimal ϵ, we have that:

$$0 \leq |n - k| = |n - {}^\circ k - \epsilon| \leq |n - {}^\circ k| + |\epsilon| < 1$$

Now, since $|n - k| \in {}^*\mathbb{Z}$, we must have $n = k$ by proposition 4.3.3 because the above shows $|n - k| \neq 1$. ☐

Corollary 4.3.1.4. *For $x \in {}^*\mathbb{Z}$ we have ${}^*\mathbb{Z} \cap m(x) = \{x\}$.*

Proof.
$\supset$: By assumption and the fact that 0 is infinitesimal, this inclusion is clear.
$\subset$: Fix $y \in {}^*\mathbb{Z} \cap m(x)$. By virtue of y being in the monad of x, we have that $y = x + \epsilon$ for some infinitesimal ϵ. Now $0 \leq |y - x| = |\epsilon| < 1$ by making a particular choice of $n = 1$ in $|\epsilon| < \frac{1}{n}$. Again, since y and x are hyperintegers, we must have that $y = x$. ☐

Chapter 5

Conclusion

In this essay, we have presented a rigorous treatment of Nonstandard Analysis, a general theory originally due to Robinson [2]. Under this framework, the concept of infinitesimal numbers can be made precise and one can expect it to satisfy the usual arithmetic of the reals. As well, the transfer principle gives a powerful tool that extends logical statements of the reals to the hyperreals.

This is of course not the be all end all of Nonstandard Analysis. As mentioned in the introduction chapter 1, this essay very closely follows only a strict subset of chapter 1 of Hurd and Loeb [1]. Indeed, in their book, they present deeper insights of Nonstandard Analysis both for logic and analysis. In chapter 2 of Hurd and Loeb, they discuss Nonstandard Analysis on Superstructures. In this essay, the *-transform (and the transfer principle) were generalized only as far as relations of $\mathbb{R}$ (themselves subsets of $\mathbb{R}^n$). The results presented in this essay are insufficient in discussing the *-transform or transfer of statements involving, for example, the set of intervals of $\mathbb{R}$ (a subset of a subset of $\mathbb{R}^n$). More precisely, relations are examples *first order structures* whereas the set of intervals is an example of a *second order structure.* The intuitive idea of Nonstandard Analysis on Superstructures in this context is to be able to discuss the *-transform and transfer principle for structures of arbitrary high order (subsets of subets of ...). In chapters 3 and 4 of Hurd and Loeb, they discuss the Nonstandard Theory of Topology and Integration theory. Clearly, advanced topics in analysis can be couched in the framework of Nonstandard Analysis.

We have presented what some refer to as (an introduction of) the ultrapower method of Nonstandard Analysis. A related but subtly different theory is the so-called synthetic differential geometry approach to Nonstandard Analysis. Bell [5] provides an exposition of such a framework. Under this approach, the focus is on constructing infinitesimals as nilpotent objects from the reals. The familiar reals only admit 0 as a nilpotent element, but Bell's presentation constructs the reals from first principles *without the law of excluded middle* which roughly states that logical statements are either true or false. In the construction of the reals, the law of excluded middle leads to the familiar law of trichotomy. By removing this law, the search for nonzero nilpotent reals is meaningful. There is an even more fundamental consequence of this difference. While not given here, most nonstandard characterizations of basic real analytic notions involve proofs by contradiction. By removing the law of excluded middle, proof by contradiction is no longer a valid technique to prove results.

There is a variety of resources on Nonstandard Analysis. We hope that this essay has given the reader a brief but informative education and appreciation to the theory. We have only presented one fantastic result of Nonstandard Analysis here, namely the transfer principle, but further study in the area will quickly reveal many more interesting and mathematically significant results.

Appendix A

Existence of (free) ultrafilters

Here, we revise and apply Zorn's lemma to show the existence of (free) ultrafilters as claimed in the Ultrafilter axiom 2.1.1

Definition A.1 (Partially ordered set)**.** For X a non-empty set and $\leq$ a binary relation on X, we say that the pair $(X, \leq)$ is a *partially ordered* set if the following properties are satisfied $\forall x, y, z \in X$:

1. reflexivity - $x \leq x$
2. antisymmetry - $x \leq y$ and $y \leq x \implies x = y$
3. transitivity - $x \leq y$ and $y \leq z \implies x \leq z$

Definition A.2 (Chain)**.** Given a partially ordered set $(X, \leq)$ and a subset $C \supset X$, we say that C is a *chain* if all elements of C can be ordered by $\leq$ i.e.

$$\forall x, y \in C, \text{ either } x \leq y \text{ or } y \leq x$$

Definition A.3 (Upper bound)**.** Given a partially ordered set $(X, \leq)$, a subset $B \subset X$ and an element $x \in B$, we say x is an *upper bound* of B if:

$$\forall b \in B, \, b \leq x$$

Definition A.4 (Maximal)**.** Given a partially ordered set $(X, \leq)$ and an element $m \in X$, we say m is *maximal* if:

$$x \in X \text{ s.t. } m \leq x \implies x = m$$

We now have all the necessary definitions appearing in Zorn's lemma.

Lemma A.0.1 (Zorn's lemma)**.** *Let $(X, \leq)$ be a partially ordered set. If every chain in X has an upper bound, then X has at least one maximal element.*

The application of Zorn's lemma provides the existence of some maximal element under some partial ordering. To apply it in our context of filters and ultrafilters, we must show that ultrafilters are maximal with respect to some partial ordering (which will be set inclusion).

Lemma A.0.2 (Ultrafilter maximality)**.** *A filter $\mathcal{F}$ on a non-empty set I is an ultrafilter iff whenever $\mathcal{G}$ is a filter on I such that $\mathcal{F} \subset \mathcal{G}$, we actually have $\mathcal{F} = \mathcal{G}$.*

Proof.
($\implies$) Suppose $\mathcal{F}$ is an ultrafilter so that $\forall A \subset I$, either $A \in \mathcal{F}$ or $A^{\complement} \in \mathcal{F}$ by definition 2.3. Let $\mathcal{G}$ be a filter on I which includes $\mathcal{F}$, i.e. $\mathcal{F} \subset \mathcal{G}$. To establish set equality, we show the reverse inclusion: $\mathcal{G} \subset \mathcal{F}$. Assume, for a contradiction, that the reverse inclusion does not hold so that $\exists B \in \mathcal{G}$ and $B \notin \mathcal{F}$. By the definition of $\mathcal{F}$ being an ultrafilter, we must have that $B^{\complement} \in \mathcal{F} \subset \mathcal{G}$. Again, by combination of items 1 and 2 of filters, we have $B \cap B^{\complement} \in \mathcal{G}$ (using definition 2.1 item 2) which contradicts item 1 because $B \cap B^{\complement} = \emptyset \notin \mathcal{G}$.

This establishes equality of $\mathcal{F}$ and $\mathcal{G}$.
($\Longleftarrow$) Suppose $\mathcal{F}$ is maximal in the sense of set inclusion i.e. any other filter $\mathcal{G}$ on I containing $\mathcal{F}$ has to be equal to $\mathcal{G}$. We wish to show the ultrafilter property 2.3, that is subsets of I are either in $\mathcal{F}$ or their complements are. By symmetry of complementation, suppose $A \notin \mathcal{F}$ and we will show $A^{\complement} \in \mathcal{F}$. Define the set

$$\mathcal{G} := \{X \subset I \,|\, \exists F \in \mathcal{F} \text{ s.t. } A \cap F \subset X\}$$

We have that $\mathcal{F} \subsetneq \mathcal{G}$ because for any $F \in \mathcal{F}$, $A \cap F$ is obviously included in F which, in turn, belongs to $\mathcal{G}$ by its definition. By assumption, $\mathcal{G}$ cannot be a filter because it is a set containing, but not equal to, $\mathcal{F}$. $\mathcal{G}$ can fail to be a filter by violating any of the three items in definition 2.1. We now show that it satisfies items 2 and 3. Fix $B, C \in \mathcal{G}$ and $C \subset D \subset I$. By definition, $\exists F_1, F_2 \in \mathcal{F}$ s.t. $A \cap F_1 \subset B$ and $A \cap F_2 \subset C$. Since $\mathcal{F}$ is a filter, we have $F_1 \cap F_2 \in \mathcal{F}$ and thus:

$$A \cap (F_1 \cap F_2) = (A \cap F_1) \cap (A \cap F_2) \subset B \cap C$$

Hence $B \cap C \in \mathcal{G}$. Trivially, we have the inclusion $A \cap F_2 \subset C \subset D$ which shows $D \in \mathcal{G}$. We have just shown that $\mathcal{G}$ satisfies items 2 and 3 of filters in 2.1. Since $\mathcal{G}$ is not a filter, we must have that item 1 is violated so that $\emptyset \in \mathcal{G}$. By definition, this means $\exists F \in \mathcal{F}$ s.t. $A \cap F \subset \emptyset$ which implies $F \subset A^{\complement}$. In the definition of filters 2.1, item 3 yields $A^{\complement} \in \mathcal{F}$ and the proof is complete. $\square$

We are now ready to prove the existence of ultrafilters in the 'Ultrafilter Axiom.' Observe that the previous lemma gave an equivalent characterization of ultrafilters, namely they are maximal with respect to the partial ordering of set inclusion on the subsets of the set of filters on I.

Theorem A.0.1 (Ultrafilter Axiom)**.** *Let I be a non-empty set and $\mathcal{F}$ be a filter on I. Then there is an ultrafilter $\mathcal{U}$ on I which contains $\mathcal{F}$.*

Proof. Let $\hat{\mathcal{F}}$ be the set of all filters which contain $\mathcal{F}$. $\hat{\mathcal{F}}$ is non-empty because $\mathcal{F}$ is trivially an element of this set. We equip $\hat{\mathcal{F}}$ with the binary relation set inclusion, thus making $(\hat{\mathcal{F}}, \subset)$ a partially ordered set.

To apply Zorn's lemma A.0.1, we need to show that any chain has an upper bound. Thus fix $\tilde{\mathcal{C}}$ a chain in $\hat{\mathcal{F}}$ and consider $\tilde{\mathcal{F}} := \bigcup_{\mathcal{C} \in \tilde{\mathcal{C}}} \mathcal{C}$. We claim that $\tilde{\mathcal{F}}$ is an upper bound of $\tilde{\mathcal{C}}$. We need to show that $\tilde{\mathcal{F}}$ is a filter for it is obvious that it includes every element of $\tilde{\mathcal{C}}$. Since all $\mathcal{C} \in \tilde{\mathcal{C}}$ are themselves filters, we have $\emptyset \notin \mathcal{C}$ for every such $\mathcal{C}$. Taking the union over all $\mathcal{C}$ in $\tilde{\mathcal{C}}$ preserves this exclusion of $\emptyset$ so item 1 of filters 2.1 is established. Let $A, B \in \tilde{\mathcal{F}}$ so that $A \in \mathcal{C}_1$, $B \in \mathcal{C}_2$ for some $\mathcal{C}_1, \mathcal{C}_2 \in \tilde{\mathcal{C}}$. Now, since $\tilde{\mathcal{C}}$ is a chain, one of $\mathcal{C}_1, \mathcal{C}_2$ is included in the other. Without loss of generality, say $\mathcal{C}_1 \subset \mathcal{C}_2$. Then, $A \in \mathcal{C}_2$ so applying item 2 for $\mathcal{C}_2$ gives $A \cap B \in \mathcal{C}_2 \subset \tilde{\mathcal{F}}$ establishing item 2 for $\tilde{\mathcal{F}}$. To establish the final item, assume $A \in \tilde{\mathcal{F}}$ and $A \subset B \subset I$. A has to be an element of some filter $\mathcal{C}$ of $\tilde{\mathcal{C}}$ so $B \in \mathcal{C} \subset \tilde{\mathcal{F}}$ and we have established item 3. Hence $\tilde{\mathcal{F}}$ is a filter so it is an upper bound for $\tilde{\mathcal{C}}$.

Having established that every chain within the set of filters, $\hat{\mathcal{F}}$, containing $\mathcal{F}$ has an upper bound, Zorn's lemma gives the existence of a maximal element of $\hat{\mathcal{F}}$. By the previous lemma A.0.2, this maximal element is an ultrafilter containing $\mathcal{F}$. $\square$

While the ultrafilter axiom guarantees the existence of ultrafilters, we would like them to enjoy some non-trivial properties. In particular, we would like the ultrafilter guaranteed by the ultrafilter axiom A.0.1 to be free as in definition 2.4.

Theorem A.0.2. *Free ultrafilters exist on infinite sets.*

Proof. Let I be an infinite set. The collection $\mathcal{F}_1 = \{F \subset I \,|\, F^{\complement} \text{ is finite}\}$ (as defined in definition 2.2) is a filter on I. The empty set cannot be in $\mathcal{F}_1$ for I is not finite. If A and B are elements of $\mathcal{F}_1$, then their intersection is also in $\mathcal{F}_1$ because $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$ is finite as a union of two finite sets using one of De Morgan's Laws. If A is an element of $\mathcal{F}_1$ and $A \subset B \subset I$, then we have the inclusion $B^{\complement} \subset A^{\complement}$ where the latter is finite, so $B^{\complement}$ is also finite and hence B belongs to $\mathcal{F}_1$. Thus, all three items of filters in definition 2.1 are established for $\mathcal{F}_1$.

By appealing to the Ultrafilter Axiom (theorem A.0.1), there is an ultrafilter $\mathcal{U}$ on I containing $\mathcal{F}_1$. Exactly by definition 2.4, $\mathcal{U}$ is free. $\square$

Bibliography

[1] Hurd, A. E., Loeb, P. A. *An Introduction to Nonstandard Real Analysis*. London: Academic Press Inc. (London) Ltd; 1985.

[2] Robinson A. *Non-Standard Analysis*. Amsterdam: North-Holland Publishing Company; 1961.

[3] Goldblatt R. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. New York: Springer-Verlag Inc.; 1998.

[4] Keisler H. J. *Elementary Calculus*. Boston: Prindle, Weber and Schmidt; 1967.

[5] Bell J. L. *A Primer of Infinitesimal Analysis*. Cambridge: Cambridge University Press; 2008.