Introduction to

Nonstandard Analysis

by

Bashir Abdel-Fattah
bashir.abdelfattah@gmail.com

Abstract The basic idea of nonstandard analysis is to extend the reals to a field that includes "infinite" and "infinitesimal" elements in order to simplify proofs and concepts by replacing limits and epsilon-delta proofs by expressions involving infinitesimals. This paper discusses the construction and properties of the hyperreals and some basic results results in differential calculus before moving on to logic and culminating in the Łóŝ Theorem/transfer principle. The contents are largely based on the papers (Davis, 2009) and (Rayo, 2015), with lesser contributions from (Fletcher et al., 2017), (Claassens, 2016), (Marker, 2010), (Keef and Guichard, 2015), and (Murnaghan, 2015). Also note that this paper assumes little background knowledge outside of analysis (the first section does begin with a brief high-level overview in terms of abstract algebra, but if the reader is unfamiliar with that material they can safely proceed knowing that everything else they need will be introduced throughout the paper).

Contents

1

Basics of Nonstandard Analysis

1.1 Constructing Infinitesimals

Although there are several different schemes of varying degrees of sophistication and utility in rigorously introducing infinite and infinitesimal elements to analysis, in this paper we will focus on the ultrafilter construction developed by Abraham Robinson in the 1960s. Robinson's method, being both one of the oldest nonstandard schemes and relatively straightforward, is well developed in the mathematical literature, and the basic idea of the method is as follows: for any set X and ring R, the set R^X of all function $f : X \to R$ is a ring under pointwise addition and multiplication

$$(f+g)(x) = f(x) + g(x) \qquad (f \cdot g)(x) = f(x) \cdot g(x)$$

However, even if R is an integral domain or a field, R^X won't be an integral domain in general due to the presence of zero divisors. For example, let X be any set and let R be any ring. Then, for any subset $S \subseteq X$, let $\chi_S(x)$ denote the characteristic function

$$\chi_S(x) = \begin{cases} 1_R & \text{if } x \in S \\ 0_R & \text{if } x \notin S \end{cases}$$

Then, letting A and B are two disjoint nonempty subsets of X, neither $\chi_A(x)$ nor $\chi_B(x)$ are uniformly zero, yet $(\chi_A \cdot \chi_B)(x) = \chi_A(x) \cdot \chi_B(x) = \chi_{A \cap B}(x) = \chi_\emptyset(x)$ is uniformly zero. In the case of the hyperreals ${}^*\mathbb{R}$, our aim is the quotient the ring $\mathbb{R}^\mathbb{N}$ of real-valued sequences by a maximal ideal MAX so that $\mathbb{R}^\mathbb{N}/\text{MAX}$ is a field. In a heuristic sense, this amounts to finding an equivalence relation $=_{\text{MAX}}$ that declares two sequences $\langle a_n \rangle$ and $\langle b_n \rangle$ equal if the set $\{n \in \mathbb{N} : a_n = b_n\}$ is "too large", so that if $\langle a_n \rangle \cdot \langle b_n \rangle = \langle a_n b_n \rangle =_{\text{MAX}} \langle 0 \rangle$, then at least one of $\langle a_n \rangle$ and $\langle b_n \rangle$ contained "enough" zeros to be equivalent to zero to begin with. This scheme of eliminating zero divisors is accomplished with the notion of filters.

1.1.1 Definitions and Properties of Filters

We begin with a definition of filters in terms of their properties. For many of the properties below, there are several equivalent definitions, many of which are more general or more sophisticated than those stated here, but we shall proceed with following in the interests of simplicity.

Definition 1.1.1. *A **filter** on a set X is a subset $\mathcal{F}$ of the power set $\mathcal{P}(X)$ satisfying the following properties:*

1. ***Proper Filter:*** *$\emptyset \notin \mathcal{F}$*

2. ***Finite Intersection Property:*** *If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$*

3. ***Superset Property:*** *If $A \in \mathcal{F}$ and $B \supset A$, then $B \in \mathcal{F}$*

Additionally, $\mathcal{F}$ is said to be a ***ultrafilter*** *if it also satisfies:*

4. ***Maximality:*** *For any $A \subseteq X$, either $A \in \mathcal{F}$ or $X \setminus A \in \mathcal{F}$*

$\mathcal{F}$ is further said to be a ***free*** *ultrafilter if it satisfies:*

5. ***Freeness:*** *$\mathcal{F}$ contains no finite subsets of X*

Note that ultrafilters actually obey a stronger version of the maximality property (of which the above definition is a special case), which we prove below in order to familiarize the reader with the style of reasoning associated with filters:

Lemma 1.1.1. *Let $\mathcal{U}$ be an ultrafilter on a set X, and let $\{A_1, A_2, \ldots, A_n\}$ be a finite collection of disjoint subsets of S such that $\bigcup_{i=1}^{n} A_i = X$. Then there is exactly one set $A_j \in \{A_1, A_2, \ldots, A_n\}$ such that $A_j \in \mathcal{U}$.*

Proof. First, suppose that $\mathcal{U}$ contains none of the sets $A_1, A_2, \ldots A_n$. Then, by the maximality property of ultrafilters, $(X \setminus A_1), (X \setminus A_2), \ldots, (X \setminus A_n) \in \mathcal{U}$, so by the finite intersection property

$$\bigcap_{m=1}^{n} (X \setminus A_m) = X \setminus (\bigcup_{m=1}^{n} A_m) = X \setminus X = \emptyset$$

is an element of $\mathcal{U}$. This is a contradiction since $\mathcal{U}$ must be a proper filter, so we must have that at least one of the sets $A_1, A_2, \ldots A_n$ is in $\mathcal{U}$. Now suppose that $\mathcal{U}$ contains more than one of the sets $A_1, A_2, \ldots, A_n$. Let A_i and A_j be two distinct sets such that $A_i, A_j \in \mathcal{U}$. Then, by the intersection property, we have

$$A_i \cap A_j = \emptyset \in \mathcal{U}$$

This is also a contradiction, so we must have that $\mathcal{U}$ contains at most one of the sets $A_1, A_2, \ldots, A_n$. Therefore $\mathcal{U}$ contains exactly one of the sets $A_1, A_2, \ldots, A_n$. ■

Examples of filters on a set X include:

1. The **trivial filter** $\mathcal{F} = X$.

2. The **principle filter** $\mathcal{F}_A$, which is defined in terms of any set $A \subset X$ as

$$\mathcal{F}_A = \{Y \subset X : Y \supset A\}$$

   In the case that $A = \{a\}$ contains only a single element, then the principle filter $\mathcal{F}_a$ is in fact an ultrafilter.

3. The **cofinite/Fréchet filter** $\mathcal{F}^{co} = \{Y \subset X : X \setminus Y \text{ is finite}\}$. Fréchet filters are non-principal.

All of the above examples are relatively simple and uninteresting for our purposes, and it turns out that every free ultrafilter is non-principle, and in fact cannot be explicitly constructed. One might then doubt whether any collection of sets that obeys all five of the given properties even exists, but we can prove the existence of free ultrafilters on any infinite set using Zorn's Lemma. We will proceed to that proof in short order, but first we need the following helpful lemma.

Lemma 1.1.2. *Let $\mathcal{F}$ be a filter on a set X, and let $A \subset X$ be a set such that $A \notin \mathcal{F}$ and $(X \setminus A) \notin \mathcal{F}$. Then $\mathcal{F}$ can be extended to a filter $\mathcal{F}'$ on X containing both $\mathcal{F}$ and A.*

Proof. To begin with, assume $\mathcal{F} \neq \emptyset$ (if $\mathcal{F} = \emptyset$, then we can just take $\mathcal{F}' = \mathcal{F}_A$, the principle filter generated by the set A, and the result follows trivially), and consider the collection

$$\mathcal{F}' = \{Y' \subset X : \text{there exists some } Y \in \mathcal{F} \text{ such that } Y' \supset (Y \cap A)\}$$

We will now verify that $\mathcal{F}'$ satisfies each of the properties of a filter in order.

1. **Proper Filter:** First, note that $X \in \mathcal{F}$ by the superset property (given any $Y \in \mathcal{F}$, then $X \supset Y$, so $X \in \mathcal{F}$). This means that $A \neq \emptyset$, because in the case that $A = \emptyset$ we have $(X \setminus A) = X \in \mathcal{F}$, which contradicts the hypotheses of the lemma. Furthermore, for any $Y \in \mathcal{F}$ we have $Y \cap A \neq \emptyset$, because if $Y \cap A = \emptyset$ then $(X \setminus A) \supset Y$, which implies that $(X \setminus A) \in \mathcal{F}$ by the superset property of $\mathcal{F}$, a contradiction. Therefore, given any $Y' \in \mathcal{F}'$, then there is some $Y \in \mathcal{F}$ such that $Y' \supset (Y \cap A) \neq \emptyset$, thus $Y' \neq \emptyset$.

2. **Finite Intersection Property:** Suppose $Y_1', Y_2' \in \mathcal{F}'$. Then there exist some sets $Y_1, Y_2 \in \mathcal{F}$ such that $Y_1' \supset (Y_1 \cap A)$ and $Y_2' \supset (Y_2 \cap A)$. Since $Y_1 \cap Y_2 \in \mathcal{F}$ by the finite intersection property of $\mathcal{F}$, then $Y_1' \cap Y_2' \supset (Y_1 \cap Y_2) \cap A$ is an element of $\mathcal{F}'$ by definition.

3. **Superset Property:** This follows fairly trivially. Suppose $Y_1' \in \mathcal{F}'$ and that $Y_2' \supset Y_1'$. Since there exists some $Y_1 \in \mathcal{F}$ such that $Y_1' \supset (Y_1 \cap A)$, in which case $Y_2' \supset (Y_1 \cap A)$ as well, so $Y_2' \in \mathcal{F}'$ by definition.

Therefore $\mathcal{F}'$ is a filter, and we can readily verify that $A \in \mathcal{F}'$ and $\mathcal{F} \subset \mathcal{F}'$: for any $Y \in \mathcal{F}$ we have that $Y, A \supset (Y \cap A)$, hence $Y, A \in \mathcal{F}'$ by definition. ■

Note that the above lemma also justifies the choice of terminology for that maximality property: any set that does not obey the maximality property of ultrafilters is not maximal in the sense that there is a strictly larger filter containing it. We will prove in the following theorem that a maximal filter (in the sense that there is no strictly larger filter containing it) obeys the maximality property of ultrafilters, and thus the two notions are equivalent.

Theorem 1.1.3 (Ultrafilter Lemma)**.** *Let $\mathcal{F}$ be a filter on the set X. Then $\mathcal{F}$ can be extended to an ultrafilter $\mathcal{U}$ on X.*

Proof. In this proof, we will first show that there is a maximal filter $\mathcal{U}$ on X containing $\mathcal{F}$, then demonstrate that $\mathcal{U}$ is an ultrafilter. The first step will be accomplished using **Zorn's Lemma**, which is equivalent to the Axiom of Choice:

> Let $\mathcal{S}$ be a family of sets. If for each chain $\mathcal{C} \subset \mathcal{S}$ there exists a member of $\mathcal{S}$ that contains all members of $\mathcal{C}$, then $\mathcal{S}$ contains a maximal member.

Recall that a **chain** is a collection of sets $\mathcal{C}$ such for any pair of distinct sets $A, B \in \mathcal{C}$, either $A \subset B$ or $B \subset A$ (thus a chain can be written $\ldots \subset A \subset B \subset C \subset \ldots$, justifying the terminology). Then let Φ be the set of filters on X containing $\mathcal{F}$, and let $\mathcal{C}$ be any chain of filters $\mathcal{F}_0 = \mathcal{F} \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ in Φ. Then we claim that

$$\mathcal{G} = \bigcup_{n=0}^{\infty} \mathcal{F}_n$$

is a filter in Φ that contains every member $\mathcal{C}$. We will now verify the properties of filters in order:

1. **Proper Filter:** This follows trivially, given that $\emptyset \notin \mathcal{F}_n$ for all $n \geq 0$ (since each filter $\mathcal{F}_n$ is proper), and hence $\emptyset \notin \mathcal{G}$.

2. **Finite Intersection Property:** Given any $Y_1, Y_2 \in \mathcal{G}$, then there must exist some filters $\mathcal{F}_{n_1}, \mathcal{F}_{n_2} \in \mathcal{C}$ such that $Y_1 \in \mathcal{F}_{n_1}$ and $Y_2 \in \mathcal{F}_{n_2}$. Suppose without loss of generality that $n_2 \leq n_1$. Then $\mathcal{F}_{n_2} \subset \mathcal{F}_{n_1}$, so both Y_1 and Y_2 are elements of $\mathcal{F}_{n_1}$. By the finite intersection property of $\mathcal{F}_{n_1}$, we have that $(Y_1 \cap Y_2) \in \mathcal{F}_{n_1} \subset \mathcal{G}$.

3. **Superset Property:** Suppose that $Y_1 \in \mathcal{G}$ and that $Y_2 \supset Y_1$. Since $Y_1 \in \mathcal{F}_n$ for some $\mathcal{F}_n \in \mathcal{C}$, then $Y_2 \in \mathcal{F}_n \subset \mathcal{G}$ by the superset property of $\mathcal{F}_n$.

Thus $\mathcal{G}$ is a indeed a filter, and $\mathcal{G}$ contains both $\mathcal{F}$ and every element of $\mathcal{C}$, so $\mathcal{G}$ is an element of Φ that contains every element of $\mathcal{C}$. Since such a filter can be constructed for every chain in Φ, then by Zorn's Lemma Φ must contain a maximal member, which we denote $\mathcal{U}$. Since $\mathcal{U}$ is an element of Φ, then $\mathcal{U}$ is a filter containing $\mathcal{F}$, and it is also clear that $\mathcal{U}$ satisfies the finite intersection property: for any set $A \subset X$, if $A \notin \mathcal{U}$ and $(X \setminus A) \notin \mathcal{U}$ then $\mathcal{U}$ can be extented to a filter $\mathcal{U}' \supset (\mathcal{U} \cup \{A\})$ by Lemma 1.1.2, contradicting the maximality of $\mathcal{U}$. Therefore $\mathcal{U}$ must contain either A of $(X \setminus A)$ for any $A \subset X$ (it cannot contain both because then by the finite intersection property $\mathcal{U}$ would also contain the empty set, contradicting the fact that $\mathcal{U}$ is a proper filter), thus $\mathcal{U}$ is an ultrafilter on X. ■

Given that any filter can be extended to an ultrafilter, we can also demonstrate the existence of a free ultrafilter $\mathcal{U}$ on an infinite set X by extending the cofinite filter $\mathcal{F}^{co}$ on X. For any finite set $A \subset X$, $\mathcal{F}^{co}$ contains $(X \setminus A)$ because $(X \setminus A)$ is infinite (X is infinite and A is finite), and hence $(X \setminus A) \in \mathcal{U}$ because $\mathcal{U} \supset \mathcal{F}^{co}$. In particular, $\mathcal{U}$ does not contain A (as before, it cannot contain both X and $(X \setminus A)$ because then $\mathcal{U}$ would also contain the empty set, a contradiction), so $\mathcal{U}$ cannot contain any finite sets.

1.1.2 Ultrafilter Construction of the Hyperreals

Definition 1.1.2 (Equivalence Modulo an Ultrafilter). *Given an ultrafilter $\mathcal{U}$ on $\mathbb{N}$, then define an equivalence relation $\equiv_{\mathcal{U}}$ on $\mathbb{R}^{\mathbb{N}}$ as follows: given any real-valued sequences $\langle a_n \rangle, \langle b_n \rangle \in \mathbb{R}^{\mathbb{N}}$,*

$$\langle a_n \rangle \equiv_{\mathcal{U}} \langle b_n \rangle \text{ if and only if } \{n \in \mathbb{N} : a_n = b_n\} \in \mathcal{U}$$

Proof. Recall that a relation $\sim$ on a set X is an equivalence relation if it satisfies the following three properties:

1. Reflexivity: $(\forall x \in X) : x \sim x$
2. Symmetry: $(\forall x, y \in X) : x \sim y \Longrightarrow y \sim x$
3. Transitivity: $(\forall x, y, z \in X) : (x \sim y \text{ and } y \sim z) \Longrightarrow x \sim z$

One can readily verify the reflexivity, symmetry, and transitivity of the relation $\equiv_{\mathcal{U}}$ on $\mathbb{R}^{\mathbb{N}}$. Given any sequence $\langle a_n \rangle \in \mathbb{R}^{\mathbb{N}}$, then the set $\{n \in \mathbb{N} : a_n = a_n\} = \mathbb{N}$ is an element of the ultrafilter $\mathcal{U}$ by the maximality principle (either $\mathbb{N}$ or $\mathbb{N} \setminus \mathbb{N} = \emptyset$ is an element of $\mathcal{U}$, but $\emptyset \notin \mathcal{U}$ because $\mathcal{U}$ is a proper filter), so $\langle a_n \rangle \equiv_{\mathcal{U}} \langle a_n \rangle$. For symmetry, if $\langle a_n \rangle \equiv_{\mathcal{U}} \langle b_n \rangle$, then $\{n \in \mathbb{N} : a_n = b_n\} = \{n \in \mathbb{N} : b_n = a_n\} \in \mathcal{U}$, so $\langle b_n \rangle \equiv_{\mathcal{U}} \langle a_n \rangle$ trivially. Finally, suppose $\langle a_n \rangle \equiv_{\mathcal{U}} \langle b_n \rangle$ and $\langle b_n \rangle \equiv_{\mathcal{U}} \langle c_n \rangle$. Then $\{n \in \mathbb{N} : a_n = b_n\}$ and $\{n \in \mathbb{N} : b_n = c_n\}$ are both elements of the ultrafilter $\mathcal{U}$, so $\{n \in \mathbb{N} : a_n = b_n \text{ and } b_n = c_n\} = \{n \in \mathbb{N} : a_n = b_n\} \cap \{n \in \mathbb{N} : b_n = c_n\} \in \mathcal{U}$ by the finite intersection property. Since $\{n \in \mathbb{N} : a_n = c_n\} \supset \{n \in \mathbb{N} : a_n = b_n \text{ and } b_n = c_n\}$, then $\{n \in \mathbb{N} : a_n = c_n\} \in \mathcal{U}$ by the superset property, so $\langle a_n \rangle \equiv_{\mathcal{U}} \langle c_n \rangle$. Therefore the given relation is indeed a valid equivalence relation. ■

Recall that the **equivalence class** of an element a of a set X equipped under an equivalence relation $\sim$ is the set of all elements $b \in X$ such that $b \sim a$. Then going forwards we will use $[\langle a_n \rangle]$ to denote the equivalence class of the sequence $\langle a_n \rangle \in \mathbb{R}^{\mathbb{N}}$ under ultrafilter equivalence, where $[\langle a_n \rangle] = [\langle b_n \rangle]$ if and only if $\langle a_n \rangle \equiv_{\mathcal{U}} \langle b_n \rangle$ (i.e., if and only if $\{n \in \mathbb{N} : a_n = n_n\} \in \mathcal{U}$). We will also denote the set of all such equivalence classes by $\mathbb{R}^{\mathbb{N}} / \equiv_{\mathcal{U}}$.

Lemma 1.1.4. *The operations of pointwise addition and multiplication*

$$[\langle a_n \rangle] + [\langle b_n \rangle] = [\langle a_n + b_n \rangle]$$

$$[\langle a_n \rangle] * [\langle b_n \rangle] = [\langle a_n b_n \rangle]$$

are well-defined binary operations on $\mathbb{R}^{\mathbb{N}} / \equiv_{\mathcal{U}}$.

Proof. Let $\langle a_n \rangle, \langle \alpha_n \rangle, \langle b_n \rangle, \langle \beta_n \rangle \in \mathbb{R}^{\mathbb{N}}$ be real valued sequences such that $[\langle a_n \rangle] = [\langle \alpha_n \rangle]$ and $[\langle b_n \rangle] = [\langle \beta_n \rangle]$. Then $\{n \in \mathbb{N} : a_n = \alpha_n\}$ and $\{n \in \mathbb{N} : b_n = \beta_n\}$ are both elements of $\mathcal{U}$, so $\{n \in \mathbb{N} : a_n = \alpha_n \text{ and } b_n = \beta_n\} = \{n \in \mathbb{N} : a_n = \alpha_n\} \cap \{n \in \mathbb{N} : b_n = \beta_n\} \in \mathcal{U}$ by the finite intersection property. The sets $\{n \in \mathbb{N} : a_n + b_n = \alpha_n + \beta_n\}$ and $\{n \in \mathbb{N} : a_n b_n = \alpha_n \beta_n\}$ are both supersets of $\{n \in \mathbb{N} : a_n = \alpha_n \text{ and } b_n = \beta_n\}$, so both are also elements of $\mathcal{U}$ by the superset property. Therefore

$$[\langle a_n \rangle] + [\langle b_n \rangle] = [\langle a_n + b_n \rangle] = [\langle \alpha_n + \beta_n \rangle] = [\langle \alpha_n \rangle] + [\langle \beta_n \rangle]$$

$$[\langle a_n \rangle] * [\langle b_n \rangle] = [\langle a_n b_n \rangle] = [\langle \alpha_n \beta_n \rangle] = [\langle \alpha_n \rangle] * [\langle \beta_n \rangle]$$

Thus the results of addition and multiplication as defined are independent of which representatives are chosen for each equivalence class, and hence the given operations are well defined. ■

Theorem 1.1.5. *Given an ultrafilter* $\mathcal{U}$ *on* $\mathbb{N}$, *then* $\mathbb{R}^{\mathbb{N}} / \equiv_{\mathcal{U}}$ *is a field under pointwise addition and multiplication.*

Proof. Perhaps the most concrete definition of a field is the following: a field is a set F equipped with binary operations $+$ and $\cdot$ such that for all $x, y, z \in F$

1. $x + y = y + x$ (commutativity of addition)
2. $(x + y) + z = x + (y + z)$ (associativity of addition)
3. There exists some element $0 \in F$ such that $x + 0 = x$ for all $x \in F$ (existence of an additive identity)
4. For each $x \in F$ there is some element $-x \in F$ such that $x + (-x) = 0$ (existence of additive inverses)
5. $x \cdot y = y \cdot x$ (commutativity of multiplication)
6. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ (associativity of multiplication)
7. There esists some element $1 \in F$ such that $x \cdot 1 = x$ for all $x \in F$ (existence of a multiplicative identity)
8. For each nonzero element $x \in F$ there exists some element $x^{-1} \in F$ such that $x \cdot x^{-1} = 1$ (existence of multiplicative inverses)
9. $(x + y) \cdot z = x \cdot z + y \cdot z$ (distributivity)

Although most of the above properties are evident in the case of $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$ (for instance, commutativity and associativity of addition and multiplication are inherited directly from the corresponding properties of the field $\mathbb{R}$), we will nonetheless show each property in order for definiteness.

1. $[\langle a_n\rangle] + [\langle b_n\rangle] = [\langle a_n + b_n\rangle] = [\langle b_n + a_n\rangle] = [\langle b_n\rangle] + [\langle a_n\rangle]$
2. $$\begin{aligned}\Big([\langle a_n\rangle] + [\langle b_n\rangle]\Big) + [\langle c_n\rangle] &= [\langle a_n + b_n\rangle] + [\langle c_n\rangle] = [\langle (a_n + b_n) + c_n\rangle] \\ &= [\langle a_n + (b_n + c_n)\rangle] = [\langle a_n\rangle] + [\langle b_n + c_n\rangle] \\ &= [\langle a_n\rangle] + \Big([\langle b_n\rangle] + [\langle c_n\rangle]\Big)\end{aligned}$$
3. Consider the element $[\langle 0\rangle]$, the equivalence class of the sequence that is identically zero. Then for any $[\langle a_n\rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$,
$$[\langle a_n\rangle] + [\langle 0\rangle] = [\langle a_n + 0\rangle] = [\langle a_n\rangle]$$
Thus $[\langle 0\rangle]$ is the additive identity in $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$.
4. For each $[\langle a_n\rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$, define $-[\langle a_n\rangle] = [\langle -a_n\rangle]$. Then
$$[\langle a_n\rangle] + \Big(-[\langle a_n\rangle]\Big) = [\langle a_n\rangle] + [\langle -a_n\rangle] = [\langle a_n - a_n\rangle] = [\langle 0\rangle]$$
5. $[\langle a_n\rangle] * [\langle b_n\rangle] = [\langle a_n b_n\rangle] = [\langle b_n a_n\rangle] = [\langle b_n\rangle] * [\langle a_n\rangle]$
6. $$\begin{aligned}\Big([\langle a_n\rangle] * [\langle b_n\rangle]\Big) * [\langle c_n\rangle] &= [\langle a_n b_n\rangle] * [\langle c_n\rangle] = [\langle (a_n b_n) c_n\rangle] \\ &= [\langle a_n (b_n c_n)\rangle] = [\langle a_n\rangle] * [\langle b_n c_n\rangle] \\ &= [\langle a_n\rangle] * \Big([\langle b_n\rangle] * [\langle c_n\rangle]\Big)\end{aligned}$$

7. Consider the element $[\langle 1\rangle]$, the equivalence class of the sequence of all ones. Then for any $[\langle a_n\rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$,

$$[\langle a_n\rangle] * [\langle 1\rangle] = [\langle a_n \cdot 1\rangle] = [\langle a_n\rangle]$$

Thus $[\langle 1\rangle]$ is the multiplicative identity in $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$.

8. For any equivalence class $[\langle a_n\rangle] \neq [\langle 0\rangle]$, since we would have $[\langle a_n\rangle] = [\langle 0\rangle]$ if $\{n \in \mathbb{N} : a_n = 0\} \in \mathcal{U}$, then it must be true that $\{n \in \mathbb{N} : a_n = 0\} \notin \mathcal{U}$. By the maximality property, this means that $\{n \in \mathbb{N} : a_n \neq 0\} = \mathbb{N} \setminus \{n \in \mathbb{N} : a_n = 0\} \in \mathcal{U}$. Then consider, for instance, the real-valued sequence defined by

$$\alpha_n = \begin{cases} a_n & \text{if } a_n \neq 0 \\ 1 & \text{if } a_n = 0 \end{cases}$$

Since $\{n \in \mathbb{N} : \alpha_n = a_n\} = \{n \in \mathbb{N} : a_n \neq 0\} \in \mathcal{U}$, then $[\langle \alpha_n\rangle] = [\langle a_n\rangle]$. Because $\alpha_n \neq 0$ for all $n \in \mathbb{N}$, then we can define the inverse

$$[\langle \alpha_n\rangle]^{-1} = [\langle \alpha_n^{-1}\rangle]$$

with the property

$$[\langle a_n\rangle] * [\langle \alpha_n\rangle]^{-1} = [\langle \alpha_n\rangle] * [\langle \alpha_n\rangle]^{-1} = [\langle \alpha_n\rangle] * [\langle \alpha_n^{-1}\rangle] = [\langle \alpha_n \alpha_n^{-1}\rangle] = [\langle 1\rangle]$$

9. $$\begin{aligned}\Big([\langle a_n\rangle] + [\langle b_n\rangle]\Big) * [\langle c_n\rangle] &= [\langle a_n + b_n\rangle] * [\langle c_n\rangle] = [\langle (a_n + b_n)c_n\rangle] \\ &= [\langle a_n c_n + b_n c_n\rangle] = [\langle a_n c_n\rangle] + [\langle b_n c_n\rangle] \\ &= [\langle a_n\rangle] * [\langle c_n\rangle] + [\langle b_n\rangle] * [\langle c_n\rangle]\end{aligned}$$

■

1.1.3 Properties of the Hyperreals

In order to show that this newly-constructed field contain "infinite" and "infinitesimal" numbers, we first need some notion of size on $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$. We accomplish this in analogy to the definition of equality on $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$:

Definition 1.1.3 (Inequality Modulo an Ultrafilter)**.** *Given an ultrafilter $\mathcal{U}$ on $\mathbb{N}$, define the relation $\leq_{\mathcal{U}}$ on $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$ as follows: given any two elements $[\langle a_n\rangle], [\langle b_n\rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$,*

$$[\langle a_n\rangle] \leq_{\mathcal{U}} [\langle b_n\rangle] \text{ if and only if } \{n \in \mathbb{N} : a_n \leq b_n\} \in \mathcal{U}$$

Then $\leq_{\mathcal{U}}$ is a total ordering on $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$.

Proof. Recall that a relation $\preceq$ on a set X is a **(partial) ordering** of X if it is

1. Reflexive: $(\forall x \in X) : x \preceq x$

2. Anti-symmetric: $(\forall x, y \in X) : (x \preceq y \text{ and } y \preceq x) \implies x = y$

3. Transitive: $(\forall x, y, z \in X) : (x \preceq y \text{ and } y \preceq z) \implies x \preceq z$

Additionally, $\preceq$ is a **total ordering** on X if it has the additional property that for all $x, y \in X$, either $x \preceq y$ or $y \preceq x$.

The proof that $\leq_{\mathcal{U}}$ is a partial ordering proceeds analogously to the case of equality modulo and ultrafilter. Given element $[\langle a_n \rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$, then the set $\{n \in \mathbb{N} : a_n \leq a_n\} = \mathbb{N}$ is an element of the ultrafilter $\mathcal{U}$ by the maximality principle (either $\mathbb{N}$ or $\mathbb{N} \setminus \mathbb{N} = \emptyset$ is an element of $\mathcal{U}$, but $\emptyset \notin \mathcal{U}$ because $\mathcal{U}$ is a proper filter), so $[\langle a_n \rangle] \leq_{\mathcal{U}} [\langle a_n \rangle]$. For transitivity, suppose $[\langle a_n \rangle] \leq_{\mathcal{U}} [\langle b_n \rangle]$ and $[\langle b_n \rangle] \leq_{\mathcal{U}} [\langle c_n \rangle]$. Then $\{n \in \mathbb{N} : a_n \leq b_n\}$ and $\{n \in \mathbb{N} : b_n \leq c_n\}$ are both elements of the ultrafilter $\mathcal{U}$, so $\{n \in \mathbb{N} : a_n \leq b_n \text{ and } b_n \leq c_n\} = \{n \in \mathbb{N} : a_n \leq b_n\} \cap \{n \in \mathbb{N} : b_n \leq c_n\} \in \mathcal{U}$ by the finite intersection property. Since $\{n \in \mathbb{N} : a_n \leq c_n\} \supset \{n \in \mathbb{N} : a_n \leq b_n \text{ and } b_n \leq c_n\}$, then $\{n \in \mathbb{N} : a_n \leq c_n\} \in \mathcal{U}$ by the superset property, so $[\langle a_n \rangle] \leq_{\mathcal{U}} [\langle c_n \rangle]$. To demonstrate that $\leq_{\mathcal{U}}$ is anti-symmetric, suppose $[\langle a_n \rangle] \leq_{\mathcal{U}} [\langle b_n \rangle]$ and $[\langle b_n \rangle] \leq_{\mathcal{U}} [\langle a_n \rangle]$. Then $\{n \in \mathbb{N} : a_n \leq b_n\}$ and $\{n \in \mathbb{N} : b_n \leq a_n\}$ are both elements of $\mathcal{U}$, so $\{n \in \mathbb{N} : a_n = b_n\} = \{n \in \mathbb{N} : a_n \leq b_n \text{ and } b_n \leq a_n\} = \{n \in \mathbb{N} : a_n \leq b_n\} \cap \{n \in \mathbb{N} : b_n \leq a_n\} \in \mathcal{U}$ by the finite intersection property. Therefore $[\langle a_n \rangle] = [\langle b_n \rangle]$.

Finally, to show that $\leq_{\mathcal{U}}$ is a total ordering, consider any two elements $[\langle a_n \rangle]$ and $[\langle b_n \rangle]$ in $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$. If $\{n \in \mathbb{N} : a_n \leq b_n\} \in \mathcal{U}$, then $[\langle a_n \rangle] \leq_{\mathcal{U}} [\langle b_n \rangle]$ and we are done. If $\{n \in \mathbb{N} : a_n \leq b_n\} \notin \mathcal{U}$, then $\{n \in \mathbb{N} : a_n > b_n\} = \mathbb{N} \setminus \{n \in \mathbb{N} : a_n \leq b_n\} \in \mathcal{U}$ by the maximality property, and $\{n \in \mathbb{N} : a_n \geq b_n\} \supset \{n \in \mathbb{N} : a_n > b_n\}$ is an element of $\mathcal{U}$ by the superset property. Therefore $[\langle b_n \rangle] \leq_{\mathcal{U}} [\langle a_n \rangle]$. ■

Now that we know that $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$ is an ordered field, we will suppress the details of its construction when not necessary going forward. We define ${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$ to be the set of **hyperreals**, and we will write, for instance,

$$(\exists\, a, b \in {}^*\mathbb{R}) : \; a \leq b$$

instead of

$$(\exists\, [\langle a_n \rangle], [\langle b_n \rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}) : \; [\langle a_n \rangle] \leq_{\mathcal{U}} [\langle b_n \rangle]$$

Note that we now take $\mathcal{U}$ to be a *free* ultrafilter, although none of the properties of ${}^*\mathbb{R}$ that we have demonstrated thus far rely on freeness. We also introduce the notation ${}^\sigma\mathbb{R}$ for the set of equivalence classes of constant-valued sequences in $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$, which is an embedding of the standard real numbers $\mathbb{R}$ into the hyperreals ${}^*\mathbb{R}$ that is naturally isomorphic to $\mathbb{R}$:

$$a \in \mathbb{R} \quad \longleftrightarrow \quad [\langle a \rangle] = [\langle a, a, a, \ldots \rangle] \in \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$$

We can now proceed to state the definition of infinite and infinitesimal numbers in ${}^*\mathbb{R}$ in terms of the function

$$|a| = \max(a, -a) = \begin{cases} a & \text{if } a \geq -a \\ -a & \text{if } a \leq -a \end{cases}$$

which is well-defined on ${}^*\mathbb{R}$ (if $a \leq -a$ and $-a \leq a$, then $\{n \in \mathbb{N} : a_n \leq -a_n\} \in \mathcal{U}$ and $\{n \in \mathbb{N} : -a_n \leq a_n\} \in \mathcal{U}$, so $\{n \in \mathbb{N} : a_n = -a_n\} = \{n \in \mathbb{N} : a_n = 0\} = \{n \in \mathbb{N} : a_n \leq -a_n\} \cap \{n \in \mathbb{N} : -a_n \leq a_n\}$ is an element of $\mathcal{U}$ by the finite intersection property, and hence $a = -a = 0$).

Definition 1.1.4 (Infinite and Infinitesimal Numbers). *Let ${}^\sigma\mathbb{R}^+$ denote the embedding of $\mathbb{R}^+ = \{r \in \mathbb{R} : r > 0\}$ into ${}^*\mathbb{R}$. Then a hyperreal number $a \in {}^*\mathbb{R}$ is* ***finite***

*if there exists some $r \in {}^\sigma\mathbb{R}^+$ such that $|a| \leq r$, **infinitesimal** if $|a| \leq r$ for every $r \in {}^\sigma\mathbb{R}^+$, and **infinite** if $|a| \geq r$ for every $r \in {}^\sigma\mathbb{R}^+$. We denote the set of all finite numbers in ${}^*\mathbb{R}$ by $\mathcal{O}$, and the set of all infinitesimal numbers of ${}^*\mathbb{R}$ by ϑ (which is a proper subset of $\mathcal{O}$).*

Note that the only infinitesimal in the standard reals ${}^\sigma\mathbb{R} \cong \mathbb{R}$ is 0, and ${}^\sigma\mathbb{R}$ doesn't contain any infinite elements. In order to demonstrate the existence of nonzero infinitesimals and infinite numbers in the hyperreals, consider the elements $[\langle n\rangle] = [\langle 1, 2, 3, 4, \ldots\rangle]$ and $[\langle \frac{1}{n}\rangle] = [\langle 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\rangle]$. Then for any $[\langle r\rangle] \in {}^\sigma\mathbb{R}^+$, there are only finitely many terms such that $n < r$ and finitely many terms such that $\frac{1}{n} > r$, so $\{n \in \mathbb{N} : n < r\}$ and $\{n \in \mathbb{N} : \frac{1}{n} > r\}$ are not elements of $\mathcal{U}$ because $\mathcal{U}$ is free (that is, it contains no finite sets). Therefore $\{n \in \mathbb{N} : n \geq r\}$ and $\{n \in \mathbb{N} : \frac{1}{n} \leq r\}$ are elements of $\mathcal{U}$ by the maximality property of ultrafilters, and hence $[\langle n\rangle] \geq_{\mathcal{U}} [\langle r\rangle]$ and $[\langle \frac{1}{n}\rangle] \leq_{\mathcal{U}} [\langle r\rangle]$ for every $[\langle r\rangle] \in {}^*\mathbb{R}^+$. Thus $[\langle \frac{1}{n}\rangle]$ is infinitesimal and $[\langle n\rangle]$ is infinite.

Definition 1.1.5 (Infinitesimal Closeness)**.** *Two finite hyperreals $a, b \in \mathcal{O}$ are said to be infinitesimally close if $a - b \in \vartheta$, which we denote $a \approx b$. The relation $\approx$ is an equivalence relation.*

Proof. The reflexivity of $\approx$ is clear: for all $a \in \mathcal{O}$, $a - a = 0 \in \vartheta$. That the relation is symmetric and transitive is also clear from the definition of an infinitesimal, since the negative of an infinitesimal is also infinitesimal and the sum of two infinitesimals is also infinitesimal (these assertions follow trivially from the definition of an infinitesimal). For any $a, b \in \mathcal{O}$, if $a - b \in \vartheta$, then $b - a = -(a - b) \in \vartheta$, so $a \approx b \implies b \approx a$. Similarly, for any $a, b, c \in \mathcal{O}$, if $a - b \in \vartheta$ and $b - c \in \vartheta$, then $a - c = (a - b) + (b - c) \in \vartheta$, so if $a \approx b$ and $b \approx c$, then $a \approx c$. ■

Definition 1.1.6. *Every finite hyperreal $a \in \mathcal{O}$ is infinitesimally close to a unique standard real number in ${}^\sigma\mathbb{R}$, which we call the **standard part of a** and denote $st(a)$.*

Proof. The existence of such a standard real number is perhaps intuitive, but the proof requires results from abstract algebra and hence is beyond the scope of this paper. Instead, we shall demonstrate the uniqueness of such a number. Suppose that $c, c' \in {}^\sigma\mathbb{R}$ are such that $c \approx a$ and $c' \approx a$. Then, by transitivity, $c \approx c'$ and hence $c - c' \in \vartheta$. Since c and c' are both standard real numbers, then their difference is also a standard real number, and because there is only one standard real number in ϑ, it follows that $c - c' = 0$. ■

Note that there are some drawbacks to the ultrafilter construction of the hyperreals that later methods have sought to rectify, the foremost issue being that it is not possible to determine any free ultrafilters on $\mathbb{N}$ (the best we can do is demonstrate their existence). As a result, the order relation on ${}^*\mathbb{R}$ is not explicitly known, and one might puzzle over trying to order sequences such as the following:

$$\langle 1, 0, 1, 0, 1, 0, \ldots\rangle$$

$$\langle 0, 1, 0, 1, 0, 1, \ldots\rangle$$

$$\langle 1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}, 7, \frac{1}{8}, \ldots\rangle$$

For the first two sequences above, one must be equivalent to $\langle 1 \rangle$ modulo $\mathcal{U}$ and the other must be equivalent to $\langle 0 \rangle$ modulo $\mathcal{U}$, but it is impossible to tell which one is which. For the third sequence, one might wonder whether the sequence represents an infinite number, an infinitesimal, or something in between. Also of concern is that different ultrafilters result in distinct fields, and it is an open problem whether or not these field will turn out to be isomorphic. However, in the section on transfer, we will find that the results that hold in these fields don't depend on the ultrafilter, because any two such fields are so-called *elementary equivalent*.

1.1.4 Enriching Sets and Functions to the Hyperreal Setting

For every set $A \subset \mathbb{R}$, we can associate the **natural extension** ${}^*A \subset \mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$ by

$$[\langle x_n \rangle] \in {}^*A \text{ if and only if } \{n \in \mathbb{N} : x_n \in A\} \in \mathcal{U}$$

The above definition can be readily extended to Cartesian products: given sequences $x_n^1, x_n^2, \ldots, x_n^m$ (where it is understood that the superscripts index the sequence in this case and do not denote exponentiation) and a set $A \in \mathbb{R}^m$, we define

$$\big([\langle x_n^1 \rangle], [\langle x_n^2 \rangle], \ldots, [\langle x_n^m \rangle]\big) \in {}^*A \text{ if and only if } \{n \in \mathbb{N} : (x_n^1, x_n^2, \ldots, x_n^m) \in A\} \in \mathcal{U}$$

This allows us to readily extend m-ary functions $f(x_1, x_2, \ldots, x_m)$ and predicates/relations $P(x_1, x_2, \ldots, x_m)$ to the nonstandard domain (note that a predicate is essentially a Boolean-valued function). One can view a m-ary relation P on a set X as a subset $\widetilde{P}$ of the Cartesian product space X^m, where we define

$$P(x_1, x_2, \ldots, x_m) \text{ if and only if } (x_1, x_2, \ldots, x_m) \in \widetilde{P}$$

and a function $f : X^m \to X^k$ can be viewed as a subsetset Γ_f of X^{m+k}, where

$$f(x_1, \ldots, x_m) = (y_1, \ldots, y_k) \text{ if and only if } (x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_k) \in \Gamma_f$$

In the case of $X = \mathbb{R}$, the natural extensions of $\widetilde{P}$ and Γ_f thus give us the corresponding function ${}^*f$ and relation ${}^*P$ in the nonstandard domain:

$${}^*P\big([\langle x_n^1 \rangle], [\langle x_n^2 \rangle], \ldots, [\langle x_n^m \rangle]\big) \iff \{n \in \mathbb{N} : P(x_n^1, x_n^2, \ldots, x_n^m)\} \in \mathcal{U}$$

$${}^*f\big([\langle x_n^1 \rangle], \ldots, [\langle x_n^m \rangle]\big) = \big([\langle y_n^1 \rangle], \ldots, [\langle y_n^k \rangle]\big) \iff \{n \in \mathbb{N} : f(x_n^1, \ldots, x_n^m) = (y_n^1, \ldots, y_n^k)\} \in \mathcal{U}$$

The extensions of a function $f : \mathbb{R}^m \to \mathbb{R}^k$ can be written more naturally in terms of the extension of the equality relation of $\mathbb{R}^k$:

$${}^*f([\langle x_n^1 \rangle], [\langle x_n^2 \rangle], \ldots, [\langle x_n^m \rangle]) = [\langle f(x_n^1, x_n^2, \ldots, x_n^m) \rangle]$$

However, we also will want more sophisticated sets in nonstandard analysis than just copies of subsets of $\mathbb{R}$, such as iterating the idea of a power set. In order to transfer the idea of the power set $\mathcal{P}(\mathbb{R})$ to the nonstandard domain, we introduce the idea of the **ultrapower set of** $\mathbb{R}$, which is the set of equivalence classes of sequences of subsets of $\mathbb{R}$ under ultrafilter equivalence (in analogy to $\mathbb{R}^{\mathbb{N}}/\equiv_{\mathcal{U}}$, we now have $\mathcal{P}(\mathbb{R})^{\mathbb{N}}/\equiv_{\mathcal{U}}$, if you will)

Definition 1.1.7. *Given two sequences $\langle A_n \rangle$ and $\langle B_n \rangle$ of subsets A_n, $B_n \subset \mathbb{R}$, we define*

$$\langle A_n \rangle \equiv_{\mathcal{U}} \langle B_n \rangle \text{ if and only if } \{n \in \mathbb{N} : A_n = B_n \in \mathcal{U}\}$$

The proof that this constitutes a valid equivalence relation is essentially identical to the proof in the case of equivalence modulo and ultrafilter for sequences of real numbers, and hence is omitted. We can also define a notion of containment for the equivalence classes in the ultrapower of $\mathcal{P}(\mathbb{R})$:

Definition 1.1.8. *Given an element $[\langle a_n \rangle] \in \mathbb{R}^{\mathbb{N}} / \equiv_{\mathcal{U}}$ and an element $[\langle A_n \rangle]$ of the ultrapower of $\mathcal{P}(\mathbb{R})$, we define a relation ${}^*\!\in$ by*

$$[\langle a_n \rangle] \ {}^*\!\in [\langle A_n \rangle] \text{ if and only if } \{n \in \mathbb{N} : a_n \in A_n\} \in \mathcal{U}$$

Although elements of the ultrapower of $\mathcal{P}(\mathbb{R})$ are not subsets of $\mathbb{R}^{\mathbb{N}} / \equiv_{\mathcal{U}}$ and ${}^*\!\in$ is not the usual definition of membership, we can define a subset ${}^*A \subset {}^*\mathbb{R}$ corresponding to each sequence $[\langle A_n \rangle]$ in the ultrapower of $\mathcal{P}(\mathbb{R})$ by

$$[\langle x_n \rangle] \in {}^*A \text{ if and only if } [\langle x_n \rangle] \ {}^*\!\in [\langle A_n \rangle]$$

Subsets of ${}^*\mathbb{R}$ that are association with members of the ultrapower of $\mathcal{P}(\mathbb{R})$ in this way are called **internal sets**. The collection of all internal sets in ${}^*\mathbb{R}$ is denoted ${}^*\mathcal{P}(\mathbb{R})$, which is a proper subset of the power set $\mathcal{P}({}^*\mathbb{R})$. One immediate consequence of this is that the natural extension of $A \subset \mathbb{R}$ from before is the internal set associated to the constant sequence $A_n = A$ for all $n \in \mathbb{N}$.

These ideas will be revisited in a more general setting in the section on the transfer principle.

1.2 Basic Analysis using the Hyperreals

In this section, we discuss some basic results from standard real analysis and differential calculus (of course, integrals and other more sophisticated mathematical objects can be extended to the nonstandard domain, but that is unfortunately beyond the scope of this paper). This section is only meant to illustrate how analysis can function in the nonstandard domain and the facility with which results can be proved. We will not expend much effort in proving such results, however, because the transfer principle discussed in later sections will make it unnecessary to re-prove results in the nonstandard domain.

Definition 1.2.1 (Infinitesimal Continuity)**.** *The function ${}^*f : {}^*A \to {}^*\mathbb{R}$ is continuous at a point $a \in {}^*A$ if ${}^*f(a + \epsilon) \approx {}^*f(a)$ for every infinitesimal $\epsilon \in \vartheta$.*

Example 1.2.1. Let ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ be the extension ${}^*f(x) = x^2$. Then for any finite hyperreal $a \in \mathcal{O}$, we have for any infinitesimal $\epsilon \in \vartheta$

$${}^*f(a + \epsilon) = (a + \epsilon)^2 = a^2 + 2a\epsilon + \epsilon^2$$

Since a is finite and ϵ is infinitesimal, then $2a\epsilon$ is infinitesimal (and clearly ϵ^2 is also infinitesimal, so

$${}^*f(a + \epsilon) = a^2 + 2a\epsilon + \epsilon^2 \approx a^2 = {}^*f(a)$$

Since this holds for any $\epsilon \in \vartheta$ for each $a \in \mathcal{O}$, then ${}^*f$ is continuous for all finite x by definition. However, ${}^*f$ is not necessary continuous for infinite values of x. For instance, given any infinite number ω, then for the infinitesimal $1/\omega$ we have

$$^*f(\omega + \frac{1}{\omega}) = (\omega + \frac{1}{\omega})^2 = \omega^2 + 2\omega \cdot \frac{1}{\omega} + \frac{1}{\omega^2} = \omega^2 + 2 + \frac{1}{\omega^2} \not\approx \omega^2 = {}^*f(\omega)$$

so ${}^*f$ is not continuous at any infinite ω.

Definition 1.2.2 (Infinitesimal Differentiability)**.** *The function ${}^*f : {}^*A \to {}^*\mathbb{R}$ is differentiable at a point $a \in {}^*A$ if there exists a finite $b \in {}^\sigma\mathbb{R}$ such that*

$$\frac{^*f(a+\epsilon) - {}^*f(a)}{\epsilon} \approx b$$

for every nonzero infinitesimal $\epsilon \in \vartheta$. In this case, we define ${}^*f'(a) = b$.

Theorem 1.2.1. *If a function ${}^*f : {}^*A \to {}^*\mathbb{R}$ is differentiable at $a \in {}^*A$, then ${}^*f$ is continuous at a.*

Proof. Since ${}^*f$ is differentiable at a, then for any infinitesimal $\epsilon \in \vartheta$ we have

$$^*f(a+\epsilon) - {}^*f(a) \approx \epsilon \cdot b$$

Since b is finite and ϵ is infinitesimal, then $\epsilon \cdot b$ is also infinitesimal, so

$$^*f(a+\epsilon) - {}^*f(a) \approx \epsilon \cdot b \approx 0$$

$$\Downarrow$$

$$^*f(a+\epsilon) \approx {}^*f(a)$$

for any $\epsilon \in \vartheta$. Therefore ${}^*f$ is continuous at a by definition. ∎

Theorem 1.2.2 (Chain Rule)**.** *Let $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ and $g : {}^*\mathbb{R} \to {}^*\mathbb{R}$ be two functions such that g is differentiable at $a \in {}^*\mathbb{R}$ and f is differentiable at $g(a)$. Then the function $f \circ g :^* \mathbb{R} \to {}^*\mathbb{R}$ is differentiable at a, and $(f \circ g)'(a) = f'(g(a))g'(a)$.*

Proof. For any $x \neq a$ such that $x \approx a$, consider the expression

$$\frac{f(g(x)) - f(g(a))}{x - a}$$

If $g(x) = g(a)$, then

$$\frac{g(x) - g(a)}{x - a} = 0 \qquad \text{and} \qquad \frac{f(g(x)) - f(g(a))}{x - a} = 0$$

so $(f \circ g)'(a) = 0 = f'(g(a))g'(a)$. If $g(x) \neq g(a)$, then we can write

$$\frac{f(g(x)) - f(g(a))}{x - a} = \frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a}$$

Since $g(x)$ is differentiable at a, then

$$\frac{g(x) - g(a)}{x - a} \approx g'(a)$$

and since $g(x)$ is continuous at a, then $g(x) \approx g(a)$, so

$$\frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \approx f'(g(a))$$

Therefore

$$(f \circ g)'(a) \approx \frac{f(g(x)) - f(g(a))}{x - a} = \frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a} \approx f'(g(a))g'(a)$$

whenever $x \approx a$, so $f \circ g$ is differentiable at a with derivative $(f \circ g)'(a) = f'(g(a))g'(a)$ by definition. ■

Note that this proof if the Chain Rule is essentially trivial, and follows from a small bit of algebra that one might naively, albeit incorrectly, attempt when first learning calculus. We can also find incredible utility in the introduction of infinite numbers when modeling probabilistic behaviors such as the physics of a large collection of particles, for example, where we often want to consider the limit where the number of particles is extremely large. It is also possible to introduce exotic mathematical objects in the nonstandard domain that can (and have) been used to prove unsolved problems in analysis.

2

The Transfer Principle

The basic idea of this section is that, rather than developing two distinct systems of analysis using either the reals or the hyperreals and having to manually prove whether or not results derived in one system hold in the other, we want a principle that guarantees that statements of certain forms hold in standard analysis if and only if they hold in the nonstandard setting. This is somewhat analogous to, for instance, the **principle of permanence of functional equations** in complex analysis, which, if one forgives the imprecise statement, roughly gives the following:

> Given two functions $g, h : \mathbb{R} \to \mathbb{R}$ and some relationship $F : \mathbb{R}^2 \to \mathbb{R}$ such that
> $$F(g(x), h(x)) = 0 \qquad \text{for all } x \in \mathbb{R}$$
> then if g, h, and F permit extensions $g^*, h^* : \mathbb{C} \to \mathbb{C}$ and $F^* : \mathbb{C}^2 \to \mathbb{C}$ such that $g^*(z)$ and $h^*(z)$ are analytic and $F^*(z, w)$ is analytic in z for each fixed w and analytic in w for each fixed z, then $g^*(z)$ and $h^*(z)$ satisfy
> $$F^*(g^*(z), h^*(z)) = 0 \qquad \text{for all } z \in \mathbb{C}$$

Although more sophisticated results have since been developed that allow for the transfer of more complicated statements, this section is intended to discuss the Łóŝ Theorem on the transfer of first-order logical formulae.

2.1 Formal Language and First-Order Logic

To begin with, we first want to define a language with which we construct logical formulae. This language will contain the following logical symbols

$\neg$	not	$\exists$	existential quantifier (there exists)
$\wedge$	and	$\mathbf{x_1}, \mathbf{x_2}, \ldots$	variables
$\vee$	or	(	left bracket
$\Rightarrow$	implies	)	right bracket
$\Leftrightarrow$	is equivalent to (if and only if)	,	comma
$\forall$	universal quantifier (for all)		

In addition to logical symbols, there are also constants, function symbols, and predicates.

Example 2.1.1. Let $P(x)$ denote the predicate $(x \equiv 0 \mod 2\pi)$, and define the functions $f_1(x) = \cos(x)$ and $f_2(x) = 2 - \cos(x)$. Then we can form the statement

$$(\forall x \in \mathbb{R})(P(x) \implies f_1(x) = f_2(x))$$

Definition 2.1.1. *A* ***language*** $\mathcal{L}$ *is a set of symbols containing*

1. *The logical symbols*
2. *A set of constants symbols* $\mathcal{C}$
3. *A set of functions* $\mathcal{F}$ *and positive integers* n_f *for each* $f \in \mathcal{F}$ *indicating that* f *is a function of* n_f *variables.*
4. *A set of predicates/relations* $\mathcal{R}$ *and positive integers* n_R *for each* $R \in \mathcal{R}$ *indicating that* R *is an* n_R*-ary relation.*

Technically, the inclusion of constants above is superfluous, because they can be regarded as *nullary functions*, functions with no inputs and one output. Since the logical symbols are a part of every language, we usually omit them when defining and discussing languages. For instance, we would denote language that only contains the logical symbols by $\mathcal{L} = \emptyset$.

Given a language, one can construct **formulas**, which are combinations of symbols arranged with the proper syntax (of course, the reader should already be familiar with the syntax of logical and mathematical expressions).

Definition 2.1.2. *Given a language* $\mathcal{L}$*, let* M *be a nonempty set and* $V \subset \mathcal{L}$ *be the set of all variables in* $\mathcal{L}$*. A* ***variable assignment*** *is a mapping* $\beta : V \to M$ *which assigns elements of* M *to all variables in* $\mathcal{V}$*.*

Definition 2.1.3. *Let* $\mathcal{L}$ *be a language. An* $\mathcal{L}$***-structure*** $\mathcal{M}$ *is given by the following data:*

1. *A nonempty set* M *called the* ***universe/domain/underlying set*** *of* $\mathcal{M}$
2. *A function* $f^{\mathcal{M}} : M^{n_f} \to M$ *for each* $f \in \mathcal{F}$
3. *A set* $R^{\mathcal{M}} \subset M^{n_R}$ *for each* $R \in \mathcal{R}$*, where we interpret* $R(x_1, x_2, \ldots, x_{n_R})$ *as "true" if and only if* $(x_1, x_2, \ldots, x_{n_R}) \in R^{\mathcal{M}}$
4. *An element* $c^{\mathcal{M}}$ *for every* $c \in \mathcal{C}$

We refer to $f^{\mathcal{M}}$*,* $R^{\mathcal{M}}$*, and* $c^{\mathcal{M}}$ *as* ***interpretations*** *of the symbols* f*,* R*, and* c*. We often write the structure as* $\mathcal{M} = (M, f^{\mathcal{M}}, R^{\mathcal{M}}, c^{\mathcal{M}} : f \in \mathcal{F}, R \in \mathcal{R}$*, and* $c \in \mathcal{C})$*, or we sometimes denote it* $\mathcal{M} = (M, I, \beta)$*, where* β *is a variable assignment function and* I *is an* ***interpretation function*** *with domain the set of all constants, symbols, and functions in* $\mathcal{L}$ *(i.e.,* $I(c) = c^{\mathcal{M}}$*,* $I(f) = f^{\mathcal{M}}$*, and* $I(R) = R^{\mathcal{M}}$ *for all* $c \in \mathcal{C}$*,* $f \in \mathcal{F}$*,* $R \in \mathcal{R}$*).*

Closely related is the concept of a relational structure:

Definition 2.1.4. *A* ***relational structure*** $\mathcal{S} - \{M, \mathcal{R}, \mathcal{F}\}$ *consists of a set* M*, a set* $\mathcal{R}$ *or finitary relations on* M*, and a set* $\mathcal{F}$ *of functions on* M*.*

Of special interest will be the relational structures $\mathfrak{R}$, consisting of $\mathbb{R}$ and all possible relations and functions $\mathbb{R}$, and ${}^*\mathfrak{R}$, which consists of ${}^*\mathbb{R}$ and the extensions of the relations and functions in $\mathfrak{R}$.

Definition 2.1.5. *A **term** is a string of symbols from a language $\mathcal{L}$ that is defined recursively as follow:*

- *Every constant and every variable of $\mathcal{L}$ is a term.*
- *If $\tau_1, \tau_2, \ldots, \tau_n$ are terms and f is an n_f-ary function, then $f(\tau_1, \tau_2, \ldots, \tau_n)$ is a term.*
- *A string of symbols is a term if it can be constructed by the application of finitely many of the above steps.*

Definition 2.1.6. *Let $\mathcal{L}$ be a language and $\mathcal{M} = (M, I, \beta)$ be an $\mathcal{L}$-structure of $\mathcal{L}$. Then the **interpretation** $(\tau)^{I,\beta}$ of any term τ of symbols in $\mathcal{L}$ is defined as follows:*

- *If $\tau = c$ for some constant $c \in \mathcal{C}$, then $(\tau)^{I,\beta} = I(c) = c^{\mathcal{M}}$*
- *If $\tau = x$ for some variable x, then $(\tau)^{I,\beta} = \beta(x)$*
- *If $\tau = f(\tau_1, \tau_2, \ldots, \tau_{n_f})$ for some n_f-ary function $f \in \mathcal{F}$, then $(\tau)^{I,\beta} = I(f)((\tau_1)^{I,\beta}, (\tau_2)^{I,\beta}, \ldots, (\tau_2)^{I,\beta}) = f^{\mathcal{M}}((\tau_1)^{I,\beta}, \ldots, (\tau_2)^{I,\beta})$*

Definition 2.1.7. *A **formula** is a string of symbols from a language $\mathcal{L}$ that is defined recursively as follows:*

1. *If τ_1 and τ_2 are terms, then $(\tau_1 = \tau_2)$ is an formula*
2. *If R is a n_R-ary relation and $\tau_1, \tau_2, \ldots, \tau_{n_R}$ are terms, then $R(\tau_1, \tau_2, \ldots, \tau_{n_R})$ is formula*
3. *If φ is a formula then so is $(\neg\varphi)$*
4. *If φ and ψ are formulas, then so is $(\varphi \wedge \psi)$*
5. *If φ is a formula and x is a variable, then $(\exists x)(\varphi)$ is a formula*
6. *A string of symbols is a formula if it can be constructed by the application of finitely many of the above steps.*

*Formulas given in the form of #1 and #2 above are called **atomic formulae**.*

Note that the inclusion of #1 above is not strictly necessary to the definition, since it is just a special case of #2. Also note that it is only necessary to consider the symbols $\wedge$, $\neg$, and $\exists$ both here and in proofs later on because all other logical symbols can be written in terms of these three. For instance, $(\varphi \vee \psi) = \neg((\neg\varphi) \wedge (\neg\psi))$, $(\forall x)(\varphi) = \neg((\exists x)(\neg\varphi))$, and $(\varphi \Rightarrow \psi) = \neg(\varphi \wedge (\neg\psi))$.

Definition 2.1.8. *Let $\mathcal{L}$ be a language and $\mathcal{M} = (M, I, \beta)$ be an $\mathcal{L}$-structure for $\mathcal{L}$, and let φ be a formula in $\mathcal{L}$. Then we say that $\mathcal{M}$ **satisfies** φ and write $\mathcal{M} \models \varphi$ if:*

- *If* $\varphi = R(\tau_1, \ldots, \tau_{n_R})$ *for some* n_R*-ary relation* $R \in \mathcal{R}$*, meaning that* φ *is atomic, then* $\mathcal{M} \models \varphi$ *if* $((\tau_1)^{I,\beta}, \ldots, (\tau_{n_R})^{I,\beta}) \in I(R) = R^{\mathcal{M}}$
- *If* $\varphi = \neg\psi$ *for some atomic formula* ψ*, then* $\mathcal{M} \models \varphi$ *if* $\mathcal{M}$ *does not satisfy* ψ *(that is, if* $\mathcal{M} \not\models \varphi$*)*
- *If* $\varphi = (\mu \wedge \nu)$ *for some atomic formulae* μ *and* ν*, then* $\mathcal{M} \models \varphi$ *if* $\mathcal{M} \models \mu$ *and* $\mathcal{M} \models \nu$*.*
- *If* $\varphi = (\exists x)(\psi)$ *for some atomic formula* ψ*, then* $\mathcal{M} = (M, I, \beta) \models \varphi$ *if there exists some* $c \in M$ *such that* $(M, I, \beta[x, c]) \models \psi$*, where*

$$\beta[x, c](y) = \begin{cases} c & \text{if } y = x \\ \beta(y) & \text{if } y \neq x \end{cases}$$

Definition 2.1.9. *Given a language* $\mathcal{L}$*, a set* T *of non-empty formulas in* $\mathcal{L}$ *is called a* ***theory****. We say that an* $\mathcal{L}$*-structure* $\mathcal{M}$ *is a* ***model*** *of* T *if* $\mathcal{M} \models \varphi$ *for all* $\varphi \in T$*, and write* $\mathcal{M} \models T$*. The* ***theory*** *of* $\mathcal{M}$*, denoted* $Th(\mathcal{M})$*, is the set of all sentences* φ *of* $\mathcal{L}$ *such that* $\mathcal{M} \models \varphi$*.*

Definition 2.1.10. *Let* φ *be a formula. In the expressions* $(\forall x)(\varphi)$ *and* $(\exists x)(\varphi)$*, we call* φ *the* ***range of the quantifier****.*

Definition 2.1.11. *An occurrence of a variable* x *is called* ***bound*** *if it lies in the range of a universal or existential quantifier, and* ***free*** *otherwise. A formula in which all variables are bound is called a* ***sentence****, and is said to be* ***closed****.*

Thus concludes the basic overview of the relevant topics in first-order logic. In the next section, we discuss ultraproducts and conclude and Łóŝ' Theorem.

2.2 Basics of Transfer

2.2.1 Ultraproducts and Ultrapowers

Definition 2.2.1. *Let* $\mathcal{U}$ *be a free ultrafilter on some (infinite) indexing set* J*, and let* $\{M_j\}_{j\in J}$ *be a collection of nonempty sets. Then we define the* ***arbitrary product*** *of the collection as*

$$\prod_{j\in J} M_j = \{f : J \to \bigcup_{j\in J} M_j \mid f(j) \in M_j \text{ for all } j \in J\}$$

Two functions $f, g \in \prod_{j\in J} M_j$ *are* ***modulo equivalent*** *if* $\{j \in J : f(j) = g(j)\} \in \mathcal{U}$*, in which case we write* $f =_{\mathcal{U}} g$ *or, in terms of equivalence classes,* $[f]_{\mathcal{U}} = [g]_{\mathcal{U}}$*.*

Note that the proof that modulo equivalence is an equivalence relation is entirely analogous to the proof that $\equiv_{\mathcal{U}}$ is an equivalence relation on $\mathbb{R}^{\mathbb{N}}$.

Definition 2.2.2. *The* ***ultraproduct*** *of* $\{M_j\}_{j\in J}$ *modulo* $\mathcal{U}$ *is the set of equivalence classes*

$$\Big(\prod_{j\in J} M_j\Big)\Big/\mathcal{U} = \{[f]_{\mathcal{U}} : f \in \prod_{j\in J} M_j\}$$

The ultrapower of a set M is the ultraproduct of the constant sequence $M_j = M$ for all $j \in J$:

$$(\prod_{j\in J} M)\Big/\mathcal{U} = \{[f]_\mathcal{U} : f \in \prod_{j\in J} M\}$$

Using the above notation, we would denote the hyperreals by ${}^*\mathbb{R} = \prod_{n\in\mathbb{N}}\mathbb{R}/\mathcal{U}$.

Definition 2.2.3. *Let J be an index set with some ultrafilter $\mathcal{U}$ on J, and let $\mathcal{M}_j = (M_j, I_j, \beta_j)$ be a $\mathcal{L}$-structure for some language $\mathcal{L}$ for all $j \in J$. Then the ultraproduct ${}^*\mathcal{M} = (\prod_{j\in J} M_j/\mathcal{U}, {}^*I, {}^*\beta)$ is a model of $\mathcal{L}$ with an interpretation function ${}^*I$ and a variable assignment function ${}^*\beta$ defined as follows:*

- *If x is a variable in $\mathcal{L}$, then ${}^*\beta(x) = [\beta_j(x)]_\mathcal{U}$*
- *If $c \in \mathcal{C}$ is a constant in $\mathcal{L}$, then ${}^*I(c) = [I_j(c)]_\mathcal{U}$*
- *If $f \in \mathcal{F}$ is an n_f-ary function, then*

$${}^*I(f)([g_1]_\mathcal{U}, [g_2]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) = [I_j(f)(g_1(j), g_2(j), \ldots, g_{n_f}(j))]_\mathcal{U}$$

- *If R is a n_R-ary relation, then*

$$([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) \in {}^*I(R) \iff \{j \in J : (g_1(j), \ldots, g_{n_f}(j)) \in I_j(R)\} \in \mathcal{U}$$

The above definitions are well-defined.

Proof. To demonstrate that the above is well-defined, we want to show that the definitions of ${}^*I(f)$ and ${}^*I(R)$ for each given argument $([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U})$ is independent of the choice of representative of each equivalence class $[g_m]_\mathcal{U}$

In case of the relation R, fix $g_1, g_2, \ldots, g_{n_R}, g'_1, g'_2, \ldots g'_{n_R} \in \prod_{j\in J} M_j$ such that $g_m =_\mathcal{U} g'_m$ for all $1 \leq m \leq n_R$. Then $\{j \in J : g_m(j) = g'_m(j)\} \in \mathcal{U}$ for all $1 \leq m \leq n_R$, so by the finite intersection property

$$\{j \in J : (g_1(j), \ldots, g_{n_R}(j)) = (g'_1(j), \ldots, g'_{n_R}(j))\} = \bigcap_{m=1}^{n_R} \{j \in J : g_m(j) = g'_m(j)\} \in \mathcal{U}$$

If $([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) \in {}^*I(R) \implies \{j \in J : (g_1(j), \ldots, g_{n_f}(j)) \in I_j(R)\} \in \mathcal{U}$, then by the finite intersection property again the set below, which is the intersection of the previous two sets, is also an element of $\mathcal{U}$.

$$\{j \in J : (g_1(j), \ldots, g_{n_R}(j)) = (g'_1(j), \ldots, g'_{n_R}(j)) \text{ and } (g_1(j), \ldots, g_{n_R}(j)) \in I_j(R)\}$$

Then by the superset property, the set below, which is a superset of the set above, is an element of $\mathcal{U}$.

$$\{j \in J : (g'_1(j), \ldots, g'_{n_R}(j)) \in I_j(R)\}$$

Therefore $([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) \in {}^*I(R) \implies ([g'_1]_\mathcal{U}, \ldots, [g'_{n_f}]_\mathcal{U}) \in {}^*I(R)$, and the same reasoning with each g_m and g'_m switched will demonstrate $([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) \in {}^*I(R) \impliedby ([g'_1]_\mathcal{U}, \ldots, [g'_{n_f}]_\mathcal{U}) \in {}^*I(R)$, so

$$([g_1]_\mathcal{U}, \ldots, [g_{n_f}]_\mathcal{U}) \in {}^*I(R) \iff ([g'_1]_\mathcal{U}, \ldots, [g'_{n_f}]_\mathcal{U}) \in {}^*I(R)$$

and hence ${}^*I(R)$ is well-defined.

The case of functions proceeds similarly. Given $g_1, g_2, \ldots, g_{n_f}, g'_1, g'_2, \ldots g'_{n_f} \in \prod_{j \in J} M_j$ such that $g_m =_{\mathcal{U}} g'_m$ for all $1 \leq m \leq n_f$, then

$$\{j \in J : (g_1(j), \ldots, g_{n_f}(j)) = (g'_1(j), \ldots, g'_{n_f}(j))\} = \bigcap_{m=1}^{n_f} \{j \in J : g_m(j) = g'_m(j)\} \in \mathcal{U}$$

as before. By the superset property, the set below, which is a superset of the set above set, is also an element of $\mathcal{U}$

$$\{j \in J : I_j(f)(g_1(j), \ldots, g_{n_f}(j)) = I_j(f)(g'_1(j), \ldots, g'_{n_f}(j)\}$$

Therefore $[I_j(f)(g_1(j), \ldots, g_{n_f}(j))]_{\mathcal{U}} = [I_j(f)(g'_1(j), \ldots, g'_{n_f}(j))]_{\mathcal{U}}$, so ${}^*I(f)$ is well-defined. ■

2.2.2 Łóŝ' Theorem

Theorem 2.2.1 (Łóŝ' Theorem)**.** *Let $\mathcal{L}$ be a language, J be a set with some ultrafilter $\mathcal{U}$ on J, and $\mathcal{M} = (M_j, I_j, \beta_j)$ be an $\mathcal{L}$-structure for all $j \in J$. Then for any formula φ of $\mathcal{L}$ we have that ${}^*\mathcal{M} = (\prod_{j \in J} M_j/\mathcal{U}, {}^*I, {}^*\beta) \models \varphi$ if and only if $\{j \in J : \mathcal{M}_j \models \varphi\} \in \mathcal{U}$*

Proof. Since every formula is defined recursively, we proceed with the proof by induction on the complexity of the formula. First, as the base case, we consider atomic formulae

- If φ is an atomic formula (i.e., a relation $R(\tau_1, \ldots, \tau_{n_R})$ between the terms $\tau_1, \ldots, \tau_{n_R}$) then the statement holds by definition of relations on the ultraproduct (revisit Definition 2.2.3 if this is unclear)

Now, given any formula φ, suppose as an induction hypothesis that the result holds for any formula that can be constructed in strictly fewer steps than φ.

- If $\varphi = (\mu \wedge \nu)$ for some formulas μ and ν, then suppose for the forwards direction of the proof that ${}^*\mathcal{M} \models (\mu \wedge \nu)$. Then ${}^*\mathcal{M} \models \mu$ and $\mathcal{M} \models \nu$ by definition. Since μ and ν can be constructed in one fewer step than φ, then by the induction hypothesis ${}^*\mathcal{M} \models \mu \implies \{j \in J : \mathcal{M}_j \models \mu\} \in \mathcal{U}$ and ${}^*\mathcal{M} \models \nu \Rightarrow \{j \in J : \mathcal{M}_j \models \nu\} \in \mathcal{U}$. Then, by the finite intersection property,

  $$\begin{aligned} \{j \in J : \mathcal{M}_j \models (\mu \wedge \nu)\} &= \{j \in J : \mathcal{M}_j \models \mu \text{ and } \mathcal{M}_j \models \nu\} \\ &= \{j \in J : \mathcal{M}_j \models \mu\} \cap \{j \in J : \mathcal{M}_j \models \nu\} \end{aligned}$$

  is an element of $\mathcal{U}$.

  For the reverse direction, suppose that $\{j \in J : \mathcal{M}_j \models (\mu \wedge \nu)\} \in \mathcal{U}$. Then, since $\{j \in J : \mathcal{M}_j \models (\mu \wedge \nu)\} = \{j \in J : \mathcal{M}_j \models \mu \text{ and } \mathcal{M}_j \models \nu)\}$, we have that

  $$\{j \in J : \mathcal{M}_j \models \mu\} \supset \{j \in J : \mathcal{M}_j \models \mu \text{ and } \mathcal{M}_j \models \nu)\}$$
  $$\{j \in J : \mathcal{M}_j \models \nu\} \supset \{j \in J : \mathcal{M}_j \models \mu \text{ and } \mathcal{M}_j \models \nu)\}$$

  are both elements of $\mathcal{U}$ by the superset property. Therefore ${}^*\mathcal{M} \models \mu$ and ${}^*\mathcal{M} \models \nu$ by the induction hypothesis, so ${}^*\mathcal{M} \models (\mu \wedge \nu)$ by definition.

- If $\varphi = (\neg\psi)$ for some formula ψ, then ${}^*\mathcal{M} \models \varphi$ if and only if ${}^*\mathcal{M} \not\models \psi$ by definition. Since ψ can be constructed in one fewer step that φ, it follows from the induction hypothesis that ${}^*\mathcal{M} \not\models \psi \iff \{j \in J : \mathcal{M}_j \models \psi\} \notin \mathcal{U}$. By the maximality property, $\{j \in J : \mathcal{M}_j \models \psi\} \notin \mathcal{U}$ if and only if

$$J \setminus \{j \in J : \mathcal{M}_j \models \psi\} = \{j \in J : \mathcal{M}_j \not\models \psi\} = \{j \in J : \mathcal{M}_j \models \varphi\} \in \mathcal{U}$$

Therefore ${}^*\mathcal{M} \models \varphi \iff \{j \in J : \mathcal{M}_j \models \varphi\} \in \mathcal{U}$ in this case.

- If $\varphi = (\exists x)(\psi)$ for some formula ψ with a free variable x, then suppose for the forwards direction that ${}^*\mathcal{M} \models (\exists x)(\psi)$. By definition, this means that there exists some $[g]_\mathcal{U} \in \Pi_{j\in J} {}^{M_j}/_\mathcal{U}$ such that $(\Pi_{j\in J} {}^{M_j}/_\mathcal{U}, {}^*I, {}^*\beta[x, [g]_\mathcal{U}]) \models \psi$. Since ψ can be constructed in one fewer step than φ, then by the induction hypothesis this implies that $\mathcal{U}$ contains $\{j \in J : (M_j, I_j, \beta_j[x, g(j)]) \models \psi\}$, and hence by the superset property contains

$$\{j \in J : \mathcal{M}_j = (M_j, I_j, \beta_j) \models (\exists x)(\psi)\} \supset \{j \in J : (M_j, I_j, \beta_j[x, g(j)]) \models \psi\}$$

For the reverse direction, suppose $\{j \in J : \mathcal{M}_j \models (\exists x)(\psi)\} \in \mathcal{U}$. Then, using the Axiom of Choice, define a function $g \in \prod_{j\in J} M_j$ such that for all $j \in \{j \in J : \mathcal{M}_j = (M_j, I_j, \beta_j) \models (\exists x)(\psi)\}$ we have that $g(j)$ is chosen so that $(M_j, I_j, \beta_j[x, g(j)]) \models \psi$ (that is, $g(j)$ is an element in M_j that satisfies ψ for each j where such an element exists). Then by the induction hypothesis we have that $\{j \in J : (M_j, I_j, \beta_j[x, g(j)]) \models \psi\} \in \mathcal{U} \implies (\Pi_{j\in J} {}^{M_j}/_\mathcal{U}, {}^*I, {}^*\beta[x, [g]_\mathcal{U}]) \models \psi$, so ${}^*\mathcal{M} = (\Pi_{j\in J} {}^{M_j}/_\mathcal{U}, {}^*I, {}^*\beta) \models (\exists x)(\psi)$ by definition.

Since every formula in $\mathcal{L}$ can be obtained by the application of finitely many of the above steps, then the result holds by induction. ■

If one returns to the notion that a filter roughly represents a notion of "largeness" on a set, then in a heuristic sense Łóŝ' Theorem essentially says that a statement holds in the ultraproduct ${}^*\mathcal{M}$ if and only if it holds in a "large enough" proportion of the $\mathcal{L}$-structures $\mathcal{M}_j$.

Corollary 2.2.1.1 (Transfer Principle)**.** *Let $\mathcal{L}$ be a language, J be a set with some ultrafilter $\mathcal{U}$ on J, and $\mathcal{M} = (M, I, \beta)$ be an $\mathcal{L}$-structure. Then for all φ of $\mathcal{L}$ we have that ${}^*\mathcal{M} = (\Pi_{j\in J} {}^M/_\mathcal{U}, {}^*I, {}^*\beta) \models \varphi$ if and only if $\mathcal{M} = (M, I, \beta) \models \varphi$. That is, ${}^*\mathcal{M} \models Th(M)$.*

Proof. If ${}^*\mathcal{M} \models \varphi$, then $\{j \in J : \mathcal{M}_j \models \varphi\} \in \mathcal{U}$ by Łóŝ' Theorem. However, ${}^*\mathcal{M}_j = \mathcal{M}$ for all $j \in J$, then either $\{j \in J : \mathcal{M}_j \models \varphi\} = J$ or $\{j \in J : \mathcal{M}_j \models \varphi\} = \emptyset$. Since $\emptyset \notin \mathcal{U}$, then it follows that $\{j \in J : \mathcal{M}_j \models \varphi\} = \{j \in J : \mathcal{M} \models \varphi\} = J$ and hence $\mathcal{M} \models \varphi$. On the other hand, if $\mathcal{M} \models \varphi$, then $\{j \in J : \mathcal{M}_j \models \varphi\} = \{j \in J : \mathcal{M} \models \varphi\} = J \in \mathcal{U}$, so ${}^*\mathcal{M} \models \varphi$ by Łóŝ' Theorem. ■

In the case of the hyperreals ${}^*\mathbb{R}$, we accomplish transfer by way of the $*$-transform:

Definition 2.2.4 ($*$-Transform)**.** *If a term τ is a variable or a constant, then ${}^*\tau$ is the embedding of τ into ${}^\sigma\mathbb{R}$. If τ has the form $f(\tau_1, \ldots, \tau_{n_f})$ for an n_f-ary function f and terms $(\tau_1, \ldots, \tau_{n_f})$, then ${}^*\tau = {}^*f({}^*\tau_1, \ldots, {}^*\tau_1)$, where ${}^*f$ is the extension of f to the hyperreals. In the case of formulas, we define the $*$-transform inductively on the construction of the formula:*

- *For atomic formulae,* ${}^*(R(\tau_1, \ldots, \tau_{n_R})) = {}^*R({}^*\tau_1, \ldots, {}^*\tau_n)$
- ${}^*(\neg\varphi) = \neg({}^*\varphi)$
- ${}^*(\varphi \wedge \psi) = ({}^*\varphi) \wedge ({}^*\psi)$
- ${}^*((\exists x \in A)(\varphi)) = (\exists x \in {}^*A)({}^*\varphi)$

The $*$-transform essentially consists of putting a $*$ on every term, relation symbol, function symbol, and set acting as a bound on a variable.

Also note that the $*$-transforms of the relations $=$ and $\leq$ are $\equiv_{\mathcal{U}}$ and $\leq_{\mathcal{U}}$, respectively.

Example 2.2.1. The $*$-transform of

$$(\forall \epsilon \in \mathbb{R}^+)(\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})(n, m \geq N \Longrightarrow |s_n - s_m| < \epsilon)$$

is given by

$$(\forall \epsilon \in {}^*\mathbb{R}^+)(\exists N \in {}^*\mathbb{N})(\forall n, m \in {}^*\mathbb{N})(n, m \geq_{\mathcal{U}} N \Longrightarrow |{}^*s_n - {}^*s_m| <_{\mathcal{U}} \epsilon)$$

It is readily apparent that the $*$-transform coincides with ${}^*I$ and ${}^*\beta$ by verifying that the definitions of the embedding of constants and extensions of sets, relations, and functions into ${}^*\mathbb{R}$ from Section 1.1.4 are simply special cases of Definition 2.2.3. Therefore we have from the transfer principle that $\varphi \iff {}^*\varphi$ for any sentence φ. To finish the paper, we will prove as examples the equivalence between the standard and infinitesimal notions of continuity and differentiability using the transfer principle.

Theorem 2.2.2. *A function $f : \mathbb{R} \to \mathbb{R}$ satisfies the standard epsilon-delta definition of continuity at some point $a \in \mathbb{R}$ if and only if its extension ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ satisfies the infinitesimal definition of continuity (Definition 1.2.1) at $a \in {}^\sigma\mathbb{R}$.*

Proof. First, suppose that f satisfies the epsilon-delta definition of continuity at a. Then for any fixed $\epsilon \in \mathbb{R}^+$, there exists a $\delta \in \mathbb{R}^+$ such that

$$(\forall x \in \mathbb{R})(|x - a| < \delta \Longrightarrow |f(x) - f(a)| < \epsilon)$$

Taking the $*$-transform of the above sentence, we have

$$(\forall x \in {}^*\mathbb{R})(|x - a| <_{\mathcal{U}} \delta \Longrightarrow |{}^*f(x) - {}^*f(a)| <_{\mathcal{U}} \epsilon)$$

Then for any $x \approx a$, we have that $x - a$ is infinitesimal and hence less than the standard positive real number δ, so $|{}^*f(x) - {}^*f(a)| <_{\mathcal{U}} \epsilon$. Since this is true for any $\epsilon \in \mathbb{R}^+$, then $|{}^*f(x) - {}^*f(a)| < \epsilon$ for every $\epsilon \in \mathbb{R}^+$ and hence ${}^*f(x) - {}^*f(a)$ is infinitesimal by definition. Therefore for all $x \approx a$ it follows that ${}^*f(x) \approx {}^*f(a)$, so ${}^*f$ satisfies infinitesimal continuity by definition.

Now, suppose that ${}^*f$ satisfies infinitesimal continuity at a, and fix some $\epsilon \in \mathbb{R}^+$. Then for any infinitesimal $\delta \in {}^*\mathbb{R}^+$, $|x - a| <_{\mathcal{U}} \delta$ means that $x \approx a$, which implies that ${}^*f(x) \approx {}^*f(a)$ by infinitesimal continuity. Therefore $|{}^*f(x) - {}^*f(a)|$ is infinitesimal and hence less than the positive real number ϵ, so we have demonstrated

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*\mathbb{R})(|x - a| <_{\mathcal{U}} \delta \Longrightarrow |{}^*f(x) - {}^*f(a)| <_{\mathcal{U}} \epsilon)$$

and hence, because the sentence above is the $*$-transform of the sentence below, it follows from the transfer principle that

$$(\exists \delta \in \mathbb{R}^+)(\forall x \in \mathbb{R})(|x-a| < \delta \Longrightarrow |f(x)-f(a)| < \epsilon)$$

Since the above holds for any $\epsilon > 0$, we have demonstrated the standard continuity of f at a. ■

Theorem 2.2.3. *A function $f : \mathbb{R} \to \mathbb{R}$ is (standard) differentiable at $a \in \mathbb{R}$ with derivative $f'(a)$ if and only if its extension ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ satisfies*

$$\frac{{}^*f(a+h) - {}^*f(h)}{h} \approx f'(a)$$

for every infinitesimal $h \in \vartheta$.

Proof. First, suppose that f is (standard) differentiable at a with deritive $f'(a)$, so

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$

$$\Downarrow$$

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall h \in \mathbb{R})\Big(|h| < \delta \Longrightarrow \Big|\frac{f(a+h)-f(a)}{h} - f'(a)\Big| < \epsilon\Big)$$

Then, fixing some $\epsilon > 0$ and the corresponding δ, we have by taking the $*$-transform that

$$(\forall h \in {}^*\mathbb{R})\Big(|h| <_{\mathcal{U}} \delta \Longrightarrow \Big|\frac{{}^*f(a+h) - {}^*f(a)}{h} - f'(a)\Big| <_{\mathcal{U}} \epsilon\Big)$$

For any infinitesimal value of h it follows that $|h|$ is less than the standard positive real number δ, and hence

$$\Big|\frac{{}^*f(a+h) - {}^*f(a)}{h} - f'(a)\Big| <_{\mathcal{U}} \epsilon$$

Since this is true for any standard $\epsilon > 0$, we must have that $\big|\frac{{}^*f(a+h)-{}^*f(a)}{h} - f'(a)\big|$ is infinitesimal, and hence

$$\frac{{}^*f(a+h) - {}^*f(a)}{h} \approx f'(a)$$

for any infinitesimal h.

For the reverse direction, suppose that ${}^*f$ satisfies the infinitesimal differentiability property at a, and fix some $\epsilon \in \mathbb{R}^+$. Then for any infinitesimal δ, whenever $|h| \leq_{\mathcal{U}} \delta$ we have that

$$\frac{{}^*f(a+h) - {}^*f(a)}{h} \approx f'(a)$$

and hence

$$\Big|\frac{{}^*f(a+h) - {}^*f(a)}{h} - f'(a)\Big| \leq_{\mathcal{U}} \epsilon$$

Therefore we have demonstrated

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*\mathbb{R})\Big(|h| <_{\mathcal{U}} \delta \Longrightarrow \Big|\frac{{}^*f(a+h) - {}^*f(a)}{h} - f'(a)\Big| <_{\mathcal{U}} \epsilon\Big)$$

and hence, since the above statement is the $*$-transform of the statement below, we must have

$$(\exists \delta \in \mathbb{R}^+)(\forall x \in \mathbb{R})\Big(|h| < \delta \Longrightarrow \big|\frac{f(a+h) - f(a)}{h} - f'(a)\big| < \epsilon\Big)$$

by the transfer principle. Since this holds for any $\epsilon \in \mathbb{R}^+$, it follows from the epsilon-delta definition of a limit that

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a)$$

thus f is (standard) differentiable at a. ■

Bibliography

F. C. Claassens. Non-standard analysis. Web page, 2016. URL https://www.universiteitleiden.nl/binaries/content/assets/science/mi/scripties/masterclaassens.pdf. Accessed 2020/06/04.

I. Davis. An introduction to nonstandard analysis. Web page, 2009. URL https://www.math.uchicago.edu/~may/VIGRE/VIGRE2009/REUPapers/Davis.pdf. Accessed 2020/06/04.

P. Fletcher, K. Hrbacek, V. Kanovei, M. G. Katz, C. Lobry, and S. Sanders. Approaches to analysis with infinitesimals following robinson, nelson, and others, 2017. URL https://arxiv.org/pdf/1703.00425.pdf.

P. Keef and D. Guichard. Introduction to higher mathematics. Web page, 2015. URL https://www.whitman.edu/mathematics/higher_math_online/. Accessed 2020/06/04.

D. Marker. Model theory for algebra and algebraic geometry. Web page, 2010. URL http://homepages.math.uic.edu/~marker/orsay/orsay1.pdf. Accessed 2020/06/04.

F. Murnaghan. Mat 240 - algebra i: Fields. Web page, 2015. URL http://www.math.toronto.edu/fiona/courses/mat240/field.pdf. Accessed 2020/06/04.

D. A. B. Rayo. Introduction to non-standard analysis. Web page, 2015. URL http://math.uchicago.edu/~may/REU2015/REUPapers/Rayo.pdf. Accessed 2020/06/04.