INTRODUCTION TO NON-STANDARD ANALYSIS

DIEGO ANDRES BEJARANO RAYO

Abstract. In this paper we introduce the basic concepts of non-standard analysis. First, we provide an overview of first order logic and the theory of filters to successfully construct the set of hyperreal numbers, which we will use as our object of study. Using the transfer principle, a simple corollary of Łoś's Theorem, we not only introduce the non-standard notions of continuity, differentiability and Riemann integrability for functions of one variable, but also show that they are equivalent to the standard concepts taught in a first-year calculus course.

Contents

1. First Order Logic.

Before making an introductory exposition of the ideas behind non-standard analysis and the related construction, it is important that the reader familiarizes himself with the theory of first order logic. This section presents an introduction to these ideas, with a few examples. The system of first order logic is defined inductively, starting by defining the *alphabet* or the collection of symbols that will be used.[1]

Definition 1.1. The *alphabet* of first order logic is a set containing the following elements:

- An infinite list of constant symbols: $a, b, c, a_1, b_1, c_1, \ldots$
- An infinite list of variable symbols: $x, y, z, x_1, y_1, z_1, \ldots$
- An infinite list of function symbols: $f, g, h, f_1, g_1, h_1, \ldots$
- An infinite list of relation symbols: $P, R, Q, P_1, R_1, Q_1, \ldots$
- Logical connectors: $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$, with their usual interpretation.
- Logical quantifiers: $\forall, \exists$.
- The equality symbol: $=$.
- The two parentheses "(" and ")".

Note that every function and relation symbol is an n-placed function or relation symbol. This number is commonly referred to as the *arity* of the function or relation. For example, a function with arity 1 is called unary, with arity 2 is called binary, so on and so forth. We will denote the arity of a function as $ar(f)$, so that for a binary function f, for example, we have $ar(f) = 2$.

Definition 1.2. A *term* is a string of symbols from the alphabet that is defined recursively as follows:

[1]The material presented in this section is based off of [1] and [2].

(1) Every constant is a term.
(2) Every variable is a term.
(3) If f is a function with arity n and $t_1, \dots, t_n$ are terms then $f(t_1, \dots, t_n)$ is also a term.
(4) A string of symbols is a term if it can be constructed applying the previous steps finitely many times.

At this point, it becomes important to point out that only the $\wedge, \neg$ and $\exists$ logical symbols are worth mentioning, since all the other logical connectors can be derived from these three. Note that $(\psi \vee \varphi) = \neg((\neg\psi) \wedge (\neg\varphi))$, $(\psi \to \varphi) = \neg(\psi \wedge (\neg\varphi))$, $(\psi \leftrightarrow \varphi) = \neg(\psi \wedge (\neg\varphi)) \wedge \neg(\varphi \wedge (\neg\psi))$ and $(\forall x)\varphi = \neg(\exists x \neg\varphi)$. Therefore, from now on, we will only mention the three necessary connectors, and assume that the rest follow from these equalities.

Definition 1.3. A *formula* is a string of symbols from the alphabet that is defined recursively as follows:

(1) If t_1 and t_2 are terms then $(t_1 = t_2)$ is a formula.
(2) If R is a relation with arity n and $t_1, \dots, t_n$ are terms then $(R(t_1, \dots, t_n))$ is a 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 also a formula.
(6) A string of symbols is a formula if it can be constructed by finitely many applications of the previous steps.

This also implies that, for example, $(\psi \vee \varphi)$ is a formula if, and only if, ψ and φ are formulas by using the fact that $(\psi \vee \varphi) = \neg((\neg\psi) \wedge (\neg\varphi))$ and the definition above.

Remark 1.4. The formulas described in 1.3.1 and 1.3.2 are called *atomic formulas.* Also, note that (1) can be understood as a binary relation $R_{=}(t_1, t_2)$ which is true if, and only if, $t_1 = t_2$.

One of the limitations of first order logic is that it does not allow quantification over relations, only variables. This is the key difference between first and higher order logic. While the definitions in this section will focus on developing first order logic, this distinction between first order and higher logic statements will be key in the development of non-standard analysis.

Next, we will discuss the difference between *free* and *bound* variables. In not so rigorous terms, we say that a variable is *free* if it doesn't appear next to a quantifier and *bound* otherwise.

Definition 1.5. Let φ be a formula. We define the set of *free variables of* φ, denoted as $FV(\varphi)$, inductively as follow:

(1) If $\varphi = (t_1 = t_2)$, then $FV(\psi) = \{x \mid x \text{ appears in } t_1 \text{ or } t_2\}$
(2) If $\varphi = (R(t_1, \dots, t_n))$ for some relation of arity n, then $FV(\varphi) = \{x \mid x \text{ appears in } t_i \text{ for some } 1 \leq i \leq n\}$.
(3) If $\varphi = (\neg\psi)$, where ψ is a formula, then $FV(\varphi) = FV(\psi)$.
(4) If $\varphi = (\mu \wedge \nu)$, where μ and ν are formulas, then $FV(\varphi) = FV(\mu) \cup FV(\nu)$.
(5) If $\varphi = (\exists x)\psi$, where ψ is a formula, then $FV(\varphi) = FV(\psi) \setminus \{x\}$.

Definition 1.6. A formula φ is called a *sentence* if it has no free variables, meaning $FV(\varphi) = \emptyset$.

Definition 1.7. A *language* $\mathcal{L}$ is a set containing all logical symbols and quantifiers (including the equality sign and the parenthesis) and some arbitrary number of constants, variables, function symbols and relation symbols.

It is understood that all formulas made from any language $\mathcal{L}$ follow the previous rules.

Definition 1.8. Let A be some nonempty set and $V \subset \mathcal{L}$ be the set of all variables in $\mathcal{L}$. A *variable assignment* is a mapping $\beta : V \to A$, which assigns elements of A to all variables in V. Particularly, for some element $k \in A$, some variable $x \in V$ and some variable assignment function β, there is a function $\beta[x, v]$ defined as

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

Definition 1.9. A *model or structure* M for some language $\mathcal{L}$ is an ordered triple $M = (A, I, \beta)$, where A is a nonempty set, β is a variable assignment function and I is an interpretation function with domain the set of all constants, relations and function symbols in $\mathcal{L}$ such that:

(1) For every constant symbol $c \in \mathcal{L}$, we have that $I(c) \in A$.
(2) For every function symbol $f \in \mathcal{L}$ with arity n, we have that $I(f) \in A^n \times A$. Meaning $I(f)$ is a function of arity n defined on A.
(3) For every relation symbol $R \in \mathcal{L}$ with arity n, we have that $I(R) \subset A^n$. Meaning $I(R)$ is the set of all n-tuples that satisfy R under I.

A is frequently called the *universe* of M. Note that most authors define models as the pair (A, I) and leave the variable assignment function out, even when some have defined such a function. The reason why we decide to define models as a triple is simple; it makes the following definition easy to understand and does not cause any real change on what we consider a model to be.

Definition 1.10. Let $\mathcal{L}$ be a language and $M = (A, I, \beta)$ a model of $\mathcal{L}$. Then the *interpretation* of any term t, denoted as $(t)^{I,\beta}$, of symbols in $\mathcal{L}$ is defined as follows:

- If $t = c$ for some constant c, then $(t)^{I,\beta} = I(c)$.
- If $t = x$ for some variable x, then $(t)^{I,\beta} = \beta(x)$.
- if $t = f(t_1, \ldots, t_n)$ for some function f of arity n, then $(t)^{I,\beta} = I(f)((t_1)^{I,\beta}, \ldots, (t_n)^{I,\beta})$.

Definition 1.11. Let $\mathcal{L}$ be a language, $M = (A, I, \beta)$ a model for $\mathcal{L}$ and φ some formula in $\mathcal{L}$. Then, we say that M *satisfies* φ and write $M \models \varphi$ or $(A, I, \beta) \models \varphi$ whenever:

- If $\varphi = R(t_1, \ldots, t_n)$ for some relation $R \in \mathcal{L}$ of arity n, meaning φ is atomic, then $(A, I, \beta) \models \varphi$ if $((t_1)^{I,\beta}, \ldots, (t_n)^{I,\beta}) \in I(R)$.
- If $\varphi = \neg\psi$ for some atomic formula ψ, then $(A, I, \beta) \models \varphi$ if (A, I, β) does not satisfy ψ.
- If $\varphi = (\mu \wedge \nu)$ for some atomic formulas μ and ν, then $(A, I, \beta) \models \varphi$ if $(A, I, \beta) \models \mu$ and $(A, I, \beta) \models \nu$.
- If $\varphi = (\exists x)\psi$ for some atomic formula ψ, then $(A, I, \beta) \models \varphi$ if there exists some $k \in A$ such that $(A, I, \beta[x, k]) \models \psi$. Note that in this case we assume that x is a free variable on ψ.

This concludes the exposition of first order logic. While the reader might feel like the next section is quite disconnected with the previous one, both are essential in constructing the basic framework of non-standard analysis.

2. Filters, Ultrafilters, Ultraproducts and Ultrapowers

Filters are a way of formalizing the notion of "big" within set theory. The following definitions set the groundwork for the study of filters, ultrafilters, ultraproducts, ultrapowers and their application within logic.[2]

Definition 2.1. A *filter* F on a set I is a set $F \subset \mathcal{P}(I)$ such that

i. $I \in F$
ii. If $X \in F$ and $X \subset Y$, then $Y \in F$ for all $X, Y \in \mathcal{P}(I)$.
iii. If $X, Y \in F$, then $X \cap Y \in F$ for all $X, Y \in \mathcal{P}(I)$.

Intuitively, we are stating that the sets in F are considered to be "big," while those not in it are not. While this notion can be quite arbitrary, as we'll see in the next few examples, we are setting a few rules to this notion of "big": the entire set must be big, or else nothing in that set should be; sets which are bigger than "big" sets are also "big" and finite intersections of "big" sets are "big." While the final condition might not be quite as intuitive as the first two, it proves to be quite useful and can be explained by stating that sets are actually "big" if their intersections are also "big."

Remark 2.2. $\{I\}$ and $\mathcal{P}(I)$ are filters of I for any set I.

The proof of this remark should be straightforward and is thus left to the reader. Note that the filters in 2.2 are called the *trivial* filters of I. A non-trivial filter is called *proper*.

Definition 2.3. Let $x \in A$, then $F_x = \{Y \in \mathcal{P}(I) \mid x \in Y\}$ is called the *principal filter of* x over I.

[2]The material presented in Sections 2 and 3 is based off of [3], [4], [5] and the Logic lectures by Denis Hirschfeldt and Maryanthe Malliaris during the 2015 Math REU at the University of Chicago.

Lemma 2.4. *For any set $x \in A$, the principal filter of x over I is a filter of I.*

Proof. The proof of this lemma is quite simple. Since $x \in I$, then $I \in F_x$. If $x \in X \in F_x$ and $X \subset Y$, it must follows that $x \in Y$, so $Y \in F_x$. Finally F_x is closed under finite intersection since $x \in X$ and $x \in Y$ implies $X \in X \cap Y$, so F_x is a filter. $\square$

Definition 2.5. A non-trivial and non-principal filter of I is called a *free filter.*

Theorem 2.6. *A filter F over some set I is free if, and only if,* $\bigcap\limits_{A \in F} A = \emptyset$

Proof. $\Rightarrow$ Take the contrapositive. Assume that $\bigcap_{A \in F} A \neq \emptyset$. Therefore, there exists some set $x \in I$ such that $x \in Y$ for all $Y \in F$. Thus, $F \subset \{Y \in \mathcal{P}(I) \mid x \in Y\}$. Furthermore, since F is a filter, it is upwardly closed, so it must contain *every* set containing x and, hence, $F = \{Y \in \mathcal{P}(I) \mid x \in Y\}$. So F is principal.
$\Leftarrow$ Let $\bigcap_{A \in F} A = \emptyset$ and assume towards contradiction that F is principal. Then there is some $x \in I$ such that $F = \{Y \in \mathcal{P}(I) \mid x \in Y\}$. This implies that $\bigcap_{A \in F} A = \{x\}$, but we assume that $\bigcap_{A \in F} A = \emptyset$. This gives us that $\{x\} = \emptyset$. Contradiction. $\square$

Remark 2.7. Since filters are closed under finite intersection, 2.6 implies there are no free filters on finite sets.

Definition 2.8. The *Fréchet Filter* on some set I is defined as $\mathcal{F}_I = \{X \subset I \mid X \text{ is cofinite}\}$.

Lemma 2.9. *Let I be an infinite set. Then the Fréchet Filter over I is a free filter.*

Proof. Assume towards contradiction that the Fréchet filter over I is principal. Then, by 2.6 we have that $\bigcap_{A \in \mathcal{F}} A \neq \emptyset$. Fix some $k \in \bigcap_{A \in \mathcal{F}} A$ and some $X \in \mathcal{F}$. This implies that $k \in X$. However, note that $k \notin X \setminus \{k\}$. Since X is cofinite, $X \setminus \{k\}$ is also cofinite, which gives us that $X \setminus \{k\} \in \mathcal{F}$. Therefore, $k \notin \bigcap_{A \in \mathcal{F}} A$. Contradiction. $\square$

Lemma 2.10. *Every free filter contains the Fréchet filter.*

Proof. Let $\mathscr{F}$ be a free filter over some infinite set I. Let $\mathcal{F}_I$ denote the Frechet filter over I. Fix some $Y \in \mathcal{F}_I$, then it follows that $I \setminus Y$ is a finite set. Since $\mathscr{F}$ is a free filter, for every $x \in I \setminus Y$ there exists some set $K_x \in \mathscr{F}$ such that $x \notin K_x$. Since $\mathscr{F}$ is closed under finite intersection, it follows that $K = \bigcap_{x \in I \setminus Y} K_x \in \mathscr{F}$ and $K \subset Y$. This implies that $Y \in \mathscr{F}$ since $\mathscr{F}$ is upwardly closed. Thus, $\mathcal{F}_i \subset \mathscr{F}$. $\square$

Definition 2.11. A filter $\mathcal{U}$ on a set I is an *ultrafilter* of I if for all $X \subset I$ either $X \in \mathcal{U}$ or $I \setminus X \in \mathcal{U}$ but not both.

Theorem 2.12. *Let $x \in A$, then the principal filter of x over I is an ultrafilter of I.*

Proof. Let $\mathcal{F}_x$ be the principal filter of x over I. By previous observations, $\mathcal{F}_x$ is a filter. Next, let Y be an arbitrary subset of I. Then $x \in Y$ or $x \notin Y$; if $x \in Y$ then $x \notin I \setminus Y$ so $Y \in \mathcal{F}_x$ and $I \setminus Y \notin \mathcal{F}_x$. Conversely, if $x \notin Y$, then $x \in I \setminus Y$ so $Y \notin \mathcal{F}_x$ and $I \setminus Y \in \mathcal{F}_x$. Either way, only one of the sets is in the filter. Thus, $\mathcal{F}_x$ is an ultrafilter by 2.11. $\square$

Note that it is consistent with ZF set theory that there are not non-principal ultrafilters. In this paper, we require the Axiom of Choice to prove the existence of non-principal ultrafilters.

Definition 2.13. A set $G \subset \mathcal{P}(I)$ has the *finite intersection property* (FIP) if the intersection of any finite number of elements of G is nonempty.

Remark 2.14. Note that every filter has the finite intersection property.

Theorem 2.15. *Every $S \subset \mathcal{P}(I)$ with FIP has a proper filter containing it. We call this filter the filter generated by S.*

Proof. Let $\mathscr{F} = \bigcap\{F \subset \mathcal{P}(I) \mid S \subset F \text{ and } F \text{ is a proper filter on } I\}$. Note that $\emptyset \notin \mathscr{F}$ and $I \in \mathscr{F}$ since $\mathscr{F}$ is the intersection of proper filters and S has the FIP. Let $X \in \mathscr{F}$ and $X \subset Y$. It follows that X is an element of every filter and, therefore, so is Y; this implies that $Y \in \mathscr{F}$. Finally, let $X, Y \in \mathscr{F}$; by similar argument it follows that $X \cap Y \in \mathscr{F}$. Thus, $\mathscr{F}$ is a filter on I and $S \in \mathscr{F}$ by definition. $\square$

Lemma 2.16. *(Ultrafilter) Every proper filter $E \subset \mathcal{P}(I)$ is contained in some ultrafilter $\mathcal{U} \subset \mathcal{P}(I)$.*

Proof. Let $\mathscr{F} = \{F \subset \mathcal{P}(I) \mid F \supset E$ is a proper filter of $I\}$. Note then that $\mathscr{F}$ is a partially ordered set with $\subset$ as a partial order relationship. Let $C \subset \mathscr{F}$ be an arbitrary chain (totally ordered subset) of $\mathscr{F}$; we want to show that C is bounded. Consider $\bigcup C$, it is clear that every element in C is a subset of $\bigcup C$, so $\bigcup C$ is a bound of C. Thus, it suffices to show that $\bigcup C \in \mathscr{F}$ to show that $\mathscr{F}$ has a maximum element by *Zorn's lemma*.
Since every element in C is a subset of $\mathcal{P}(I)$, it follows that $\bigcup C \subset P(I)$. Note also that $I \in \bigcup C$. Next consider some $X \in \bigcup C$ and let $X \subset Y \subset \mathcal{P}(I)$; since $X \in \bigcup C$ there exists some filter F such that $X \in F$. Therefore, $Y \in F$ as F is upwardly closed and, hence, $Y \in \bigcup C$. Finally, let $X, Y \in \bigcup C$, then there are filters F, F' such that $X \in F$ and $Y \in F'$; since C is a chain we can assume without loss of generality that $F \subset F'$, which implies that $X, Y \in F'$. This gives us that $X \cap Y \in F'$ and, hence, $X \cap Y \in \bigcup C$. Hence $\bigcup C$ is a filter, so $\bigcup C \in \mathscr{F}$. Thus, by *Zorn's lemma*, $\mathscr{F}$ contains a maximal element, denote this element by $\mathcal{U}$.
Finally, we want to show that $\mathcal{U}$ is an ultrafilter; it is clear that $\emptyset \notin \mathcal{U}$ so $\mathcal{U}$ cannot contain both a subset of $\mathcal{P}(I)$ and its complement. Assume for contradiction that $\mathcal{U}$ is not an ultrafilter; then there is some set $A \subset I$ such that $A \notin \mathcal{U}$ and $I \setminus A \notin U$. Consider then $\mathcal{U} \cup \{A\}$, it follows that $\mathcal{U} \cup \{A\}$ has a finite intersection property since $\mathcal{U}$ has a finite intersection property and $I \setminus A \notin \mathcal{U}$ (so no subset of $I \setminus A$ is an element of $\mathcal{U}$). Let S be the filter generated by $\mathcal{U} \cup \{A\}$. It is clear that $\mathcal{U} \subset S$, which implies that $\mathcal{U}$ is not maximal in $\mathscr{F}$. Contradiction. □

Corollary 2.17. *There exists some free ultrafilter on* $\mathbb{N}$.

Proof. Let $\mathcal{F}_{\mathbb{N}}$ be the Frechet filter on $\mathbb{N}$. Since $\mathbb{N}$ is infinite such a filter exists. By 2.16 there exists some ultrafilter of $\mathbb{N}$, denoted by $\mathcal{U}$, containing $\mathcal{F}_{\mathbb{N}}$. Thus, $\bigcap_{A \in \mathcal{U}} A \subset \bigcap_{A \in \mathcal{F}_{\mathbb{N}}} A$. This implies that $\bigcap_{A \in \mathcal{U}} A = \emptyset$ and, hence, $\mathcal{U}$ is a free ultrafilter. □

From this point on we assume I is an infinite set with some free ultrafilter $\mathcal{U}$ and $\{A_i\}_{i \in I}$ is a collection of nonempty sets.

Definition 2.18. Let $\mathcal{U}$ be a free ultrafilter on some indexing set I and let $\{A_i\}_{i \in I}$ be a collection of nonempty sets. Then the *arbitrary product* of the collection is defined as

$$\prod_{i \in I} A_i := \{f \mid f \text{ has domain } I \text{ and } f(i) \in A_i \text{ for all } i \in I\}$$

Definition 2.19. Two functions $f, g \in \prod_{i \in I} A_i$ are *modulo* $\mathcal{U}$ *equivalent* if $\{i \in I \mid f(i) = g(i)\} \in \mathcal{U}$. We write $f =_{\mathcal{U}} g$ to indicate this relationship.

Lemma 2.20. *Modulo* $\mathcal{U}$ *equivalence is an equivalence relation.*

Proof. It suffices to show that $=_{\mathcal{U}}$ is reflexive, symmetric and transitive.

- Consider some function $f \in \prod_{i \in I} A_i$. Note that
$$\{i \in I \mid f(i) = f(i)\} = I \in \mathcal{U},$$
which implies that $f =_{\mathcal{U}} f$.
- Let $f, g \in \prod_{i \in I} A_i$. Then if $f =_{\mathcal{U}} g$, we have $\{i \in I \mid f(i) = g(i)\} \in \mathcal{U}$, but notice that
$$\{i \in I \mid f(i) = g(i)\} = \{i \in I \mid g(i) = f(i)\} \in \mathcal{U},$$
so $g =_{\mathcal{U}} f$, which implies $=_{\mathcal{U}}$ is reflexive.
- Finally, let $f, g, h \in \prod_{i \in I} A_i$ be such that $f =_{\mathcal{U}} g$ and $g =_{\mathcal{U}} h$. Then $K = \{i \in I \mid f(i) = g(i)\} \in \mathcal{U}$ and $L = \{i \in I \mid g(i) = h(i)\} \in \mathcal{U}$. Since $\mathcal{U}$ is closed under finite intersection, we have that $K \cap L = \{i \in I \mid f(i) = g(i) = h(i)\} \in \mathcal{U}$, which gives us that $f =_{\mathcal{U}} h$.

Thus, $=_{\mathcal{U}}$ is an equivalence relation. □

Definition 2.21. We define $[f]_{\mathcal{U}}$ as the equivalence class of all functions $g \in \prod_{i \in I} A_i$ such that $g =_{\mathcal{U}} f$ for some function $f \in \prod_{i \in I} A_i$.

Definition 2.22. The *ultraproduct of* $\{A_i\}_{i\in I}$ *modulo* $\mathcal{U}$ is

$$\prod_{i\in I} A_i \,/\mathcal{U} := \{[f]_{\mathcal{U}} \mid f \in \prod_{i\in I} A_i\}$$

Definition 2.23. If we let $A_i = A$ for some set, then *ultrapower of* A *modulo* $\mathcal{U}$ is

$$\prod_{i\in I} A \,/\mathcal{U} = \{[f]_{\mathcal{U}} \mid f \in \prod_{i\in I} A\}$$

Definition 2.24. Let $\mathcal{U}$ be a free filter of $\mathbb{N}$, then the set of *hyperreal numbers*, denoted as ${}^*\mathbb{R}$, is defined as the ultrapower of $\mathbb{R}$ modulo $\mathcal{U}$. In other words:

$${}^*\mathbb{R} = \prod_{n\in\mathbb{N}} \mathbb{R} \,/\mathcal{U}$$

This definition implies that the elements of the Hyperreals are the equivalence classes of sequences under modulo $\mathcal{U}$ equivalence.

3. The Structure of the Ultraproduct

One of the immediate questions that arises from the ultraproduct construction is "How is the Ultraproduct of a collection of sets similar, or different, to the sets in that collection?" This section focuses on introducing the structure of the Ultraproduct, using the tools introduced in Section 1, and provides a proof for Łoś's Theorem, which will be key to the development of non-standard analysis.

Definition 3.1. Let $\mathcal{L}$ be a language, then a *theory* of $\mathcal{L}$ is a set of sentences of $\mathcal{L}$.

Definition 3.2. Let $\mathcal{L}$ be a language, M a model for $\mathcal{L}$ and T a theory. We say that M satisfies T, and write $M \models T$ if $M \models \varphi$ for all $\varphi \in T$.

Definition 3.3. Let $\mathcal{L}$ be a language, M a model for $\mathcal{L}$. The *theory* of M, denoted as $Th(M)$, is the set of all sentences φ of $\mathcal{L}$ such that $M \models \varphi$.

Definition 3.4. Let I be an index set with some ultrafilter $\mathcal{U}$ on I and let $M_i = (A_i, I_i, \beta_i)$ be a model for some language $\mathcal{L}$ for all $i \in I$. Then the ultraproduct ${}^*M = \left(\prod_{i\in I} A_i \,/\mathcal{U}, {}^*I, {}^*\beta\right)$ is a model of $\mathcal{L}$ with an interpretation function ${}^*I$ and variable assignment function ${}^*\beta$ defined as follows:

- If x is a variable in $\mathcal{L}$, then ${}^*\beta(x) = [(\beta_i(x))]_{\mathcal{U}}$.
- If c is a constant in $\mathcal{L}$, then ${}^*I(c) = [(I_i(c))]_{\mathcal{U}}$.
- If f is a function symbol of arity n, then ${}^*I(f)([g_1]_{\mathcal{U}}, \dots, [g_n]_{\mathcal{U}}) = [(I_i(f)(g_1(i), \dots, g_n(i)))]_{\mathcal{U}}$
- If R is a relation symbol of arity n, then $([g_1]_{\mathcal{U}}, \dots, [g_n]_{\mathcal{U}}) \in ({}^*I)(R)$ if, and only if, $\{i \in I \mid (g_1(i), \dots, g_n(i)) \in I_i(R)\} \in \mathcal{U}$.

It is worth noting that in the definition above we make use of sequence notation.

Remark 3.5. The previous definition is well defined, meaning that it does not depend on the choices of the $[g_i]_{\mathcal{U}}$.

Proof. Fix $g_1, \dots g_n, g_1, \dots g_n \prod_{i\in I} A_i$ such that $g_1 =_{\mathcal{U}} g_1', \dots, g_n =_{\mathcal{U}} g_n'$ and $([g_1]_{\mathcal{U}}, \dots, [g_n]_{\mathcal{U}}) \in ({}^*I)(R)$ for some relation symbol R of arity n. Notice then that $S = \{i \in I \mid g_1(i) = g_1'(i) \wedge \dots \wedge g_n(i) = g_n'(i)\} \in \mathcal{U}$ since $\mathcal{U}$ is closed under finite intersections. This implies that

$$(\forall i \in S)((g_1(i), \dots, g_n(i)) \in I_i(R) \Leftrightarrow (g_1'(i), \dots, g_n'(i)) \in I_i(R))$$

which gives us that $\{i \in I \mid (g_1'(i), \dots, g_n'(i)) \in I_i(R)\} \in \mathcal{U}$ since it is a superset of S. This implies that $([g_1']_{\mathcal{U}}, \dots, [g_n']_{\mathcal{U}}) \in ({}^*I)(R)$. So R is well-defined.

In the case of functions, we get a similar result. Consider again some functions $g_1, \dots g_n, g_1', \dots g_n' \in \prod_{i\in I} A_i$ as above and some function symbol f of arity n. Define the set S as above and notice that $S \in \mathcal{U}$, then

$$(\forall i \in S)(I_i(f)(g_1(i), \dots, g_n(i)) = I_i(f)(g_1'(i), \dots, g_n'(i))$$

since f is a function and all the inputs are the same. Therefore, the sequence $(I_i(f)(g_1(i), \ldots, g_n(i)))$ is modulo $\mathcal{U}$ equivalent to the sequence $(I_i(f)(g'_1(i), \ldots, g'_n(i))$ so both belong to the same equivalence class. This implies that f is well defined. $\square$

Theorem 3.6. *(Łoś's Theorem) Let $\mathcal{L}$ be a language, I be a set with some ultrafilter $\mathcal{U}$ on I and $M_i = (A_i, I_i, \beta_i)$ be a model for $\mathcal{L}$ for all $i \in I$. Then for all φ of $\mathcal{L}$ we have that ${}^*M = \left(\prod_{i \in I} A_i /_{\mathcal{U}}, {}^*I, {}^*\beta\right) \models \varphi$ if, and only if, $\{i \in I \mid M_i \models \varphi\} \in \mathcal{U}$.*

Proof. We proceed inductively on the complexity of φ:

(1) If φ is an atomic formulae, then the statement holds by the previous definition.
(2) If $\varphi = (\mu \wedge \nu)$, where μ and ν are atomic formulas, then:
$\Rightarrow$ If ${}^*M \models \mu \wedge \nu$, then ${}^*M \models \mu$ and ${}^*M \models \nu$ by definition. Therefore, since μ and ν are atomic, it follows that $\{i \in I \mid M_i \models \mu\} \in \mathcal{U}$ and $\{i \in I \mid M_i \models \nu\} \in \mathcal{U}$. Since $\mathcal{U}$ is closed under finite intersection, we have that $\{i \in I \mid M_i \models \mu \wedge M_i \models \nu\} = \{i \in I \mid M_i \models (\mu \wedge \nu)\} \in \mathcal{U}$.
$\Leftarrow$ If $\{i \in I \mid M_i \models (\mu \wedge \nu)\} \in \mathcal{U}$, note that $\{i \in I \mid M_i \models (\mu \wedge \nu)\} \subset \{i \in I \mid M_i \models \mu\}$ and $\{i \in I \mid M_i \models (\mu \wedge \nu)\} \subset \{i \in I \mid M_i \models \nu\}$. Since $\mathcal{U}$ is upwardly closed, it follows that both sets are in $\mathcal{U}$, so ${}^*M \models \mu$ and ${}^*M \models \nu$, which gives us that ${}^*M \models \mu \wedge \nu$.
(3) If $\varphi = (\neg\psi)$, where ψ is an atomic formula, then if ${}^*M \models \varphi$, we have that ${}^*M$ does not model ψ, which implies that $\{i \in I \mid M_i \models \psi\} \notin \mathcal{U}$ since ψ is atomic. This gives us that $\{i \in I \mid M_i$ does not model $\psi\} \in \mathcal{U}$ since $\mathcal{U}$ is an ultrafilter. However, note that $\{i \in I \mid M_i$ does not model $\psi\} = \{i \in I \mid M_i \models \varphi\} \in \mathcal{U}$. Note that all the steps in this part of the proof are reversible, so this proves the biconditional.
(4) If $\varphi = (\exists x)\psi$, where ψ is an atomic formula and x a free variable in ψ, then:
$\Rightarrow$ If ${}^*M \models (\exists x)\psi$, we have that there is a $[g]_{\mathcal{U}} \in \prod_{i \in I} A_i /_{\mathcal{U}}$ such that $(\prod_{i \in I} A_i /_{\mathcal{U}}, {}^*I, {}^*\beta[x, [g]_{\mathcal{U}}]) \models \psi$. Since ψ is atomic, this implies that $\{i \in I \mid (A_i, I_i, \beta_i[x, g(i)]) \models \psi\} \in \mathcal{U}$. By definition, it follows that $\{i \in I \mid (A_i, I_i, \beta_i[x, g(i)] \models \psi\} = \{i \in I \mid M_i \models (\exists x)\psi\} \in \mathcal{U}$.
$\Leftarrow$ If $\{i \in I \mid M_i \models (\exists x)\psi\} \in \mathcal{U}$, define a function $g\colon I \to \bigcup_{i \in I} A_i$ such that for all $i \in \{i \in I \mid M_i \models (\exists x)\psi\}$ we have that $g(i)$ is such that $(A_i, I_i, \beta_i[x, g(i)]) \models \psi$ and $g(i) \in A_i$ otherwise. Note that such $g(i)$ exist by assumption but this step also requires the axiom of choice. Furthermore, it is clear that $[g]_{\mathcal{U}} \in \prod_{i \in I} A_i /_{\mathcal{U}}$ by the definition of the ultraproduct. Therefore, $(\prod_{i \in I} A_i /_{\mathcal{U}}, {}^*I, {}^*\beta[x, [g]_{\mathcal{U}}]) \models \psi$ by the definition of g and this in turn gives us that ${}^*M \models (\exists x)\psi$.

Since all formulas of $\mathcal{L}$ are obtained by finite application of the steps above, the proof is complete $\square$

An important corollary of Łoś's Theorem is known as the *Transfer Principle*, and it has to do with how the first order Theory of Ultrapowers relates to the first order theory of the set that makes up the ultrapower.

Corollary 3.7. *(Transfer Principle) Let $\mathcal{L}$ be a language, I be a set with some ultrafilter $\mathcal{U}$ on I and $M = (A, I, \beta)$ be a model for $\mathcal{L}$. Then for all φ of $\mathcal{L}$ we have that ${}^*M = \left(\prod_{i \in I} A /_{\mathcal{U}}, {}^*I, {}^*\beta\right) \models \varphi$ if, and only if, $M = (A, I, \beta) \models \varphi$. In other words, ${}^*M \models Th(M)$.*

Proof. This statement follow directly from Łoś's theorem. If ${}^*M \models \varphi$, then $\{i \in I \mid M_i \models \varphi\} \in \mathcal{U}$, but since all the $A_i = A$, it follows that $\{i \in I \mid M_i \models \varphi\} = I$ since otherwise it would equal the empty set. This implies that $M \models \varphi$. Conversely, if $M \models \varphi$, then $\{i \in I \mid M_i \models \varphi\} = I \in \mathcal{U}$, so ${}^*M \models \varphi$ by Łoś's theorem. $\square$

4. The Structure of the Hyperreals

It follows from 3.7 that ${}^*\mathbb{R}$ satisfies the same first order theory that $\mathbb{R}$ satisfies. This implies first that if we define addition and multiplication as determined by 3.4, then ${}^*\mathbb{R}$ is a field. While this last statement can be proven without Łoś's theorem (or in this case 3.7), we will approach it using this particular result. Note that it suffices to show that all field axioms can be expressed as first order logic statements, meaning using only the rules introduced in Section 1 of this paper, since 3.4 and 3.7 prove that ${}^*\mathbb{R}$ must also satisfy these statements.[3]

[3]From this point on the paper relies heavily on the works of [5] and [6].

Theorem 4.1. ${}^*\mathbb{R}$ *is a field.*

Proof. For the remainder of this proof let $x, y, z \in {}^*\mathbb{R}$. This implies that $x = [(x_n)]_{\mathcal{U}}, y = [(y_n)]_{\mathcal{U}}$ and $z = [(z_n)]_{\mathcal{U}}$ for some sequences $(x_n), (y_n), (z_n)$. Furthermore, following 3.4 we define addition and multiplication point-wise, so that $x + y = [(x_n + y_n)]_{\mathcal{U}}$ and $xy = [(x_n y_n)]_{\mathcal{U}}$. These operations are well-defined as shown in 3.5. Also, note that addition and multiplication are closed binary operations on ${}^*\mathbb{R}$ since the sum or product of two sequences is also a sequence. Thus, it suffices to show that the field axioms can be written as first order statements:

(1) Commutativity of Addition: $(\forall x, y)(x + y = y + x)$
(2) Associativity of Addition: $(\forall x, y, z)(x + (y + z) = (x + y) + z)$
(3) Existence of an Additive Identity: $(\exists 0)(\forall x)(x + 0 = 0 = 0 + x)$
(4) Existence of an Additive Inverse: $(\forall x)(\exists y)(x + y = 0 = y + x)$
(5) Commutativity of Multiplication: $(\forall x, y)(x \cdot y = y \cdot x)$
(6) Associativity of Multiplication: $(\forall x, y, z)(x \cdot (y \cdot z) = (x \cdot y) \cdot z)$
(7) Existence of a Multiplicative Identity: $(\exists 1)(\forall x)(x \cdot 1 = 1 = 1 \cdot x)$
(8) Existence of a Multiplicative Inverse: $(\forall x)(\exists y)(\neg(x = 0) \to (x \cdot y = 1 = y \cdot x))$
(9) Distributivity of Multiplication over Addition: $(\forall x, y, z)(x \cdot (y + z) = (x \cdot y) + (x \cdot z))$
(10) Distinctness of Identities: $\neg(0 = 1)$

While the reader might already be familiar with these statements and with the fact that these are indeed first order logic statements, this should provide sufficient proof of this claim. Note that we did not use the notation $+(a, b)$ and instead chose to write $(a + b)$; this is done mostly due to convention and does not affect the validity of the statements. Thus, by 3.7, it follows that ${}^*\mathbb{R}$ satisfies all the above and it is a field. $\square$

Furthermore, we can also show that ${}^*\mathbb{R}$ is totally ordered by a relation $\leq_{\mathcal{U}}$. To define this relation, we again appeal to 3.4 and say that $x \leq y$ if, and only if, $\{n \in \mathbb{N} \mid x_n \leq y_n\} \in \mathcal{U}$. It should be clear that the fact that ${}^*\mathbb{R}$ is totally ordered follows from the fact that $\mathbb{R}$ is totally ordered and that $\mathcal{U}$ is an ultrafilter. However, we can also use 3.7 by showing that the order axioms can be written using first order logic. Thus, we have that ${}^*\mathbb{R}$ is a totally ordered field. However, we will show that ${}^*\mathbb{R}$ is not complete, as it does not satisfy the least-upper-bound property. To do so, we will first introduce some important relationships between $\mathbb{R}$ and ${}^*\mathbb{R}$.

Theorem 4.2. *There exists an embedding of* $\mathbb{R}$ *into* ${}^*\mathbb{R}$. *That is, there exists a function* $i\colon \mathbb{R} \to {}^*\mathbb{R}$ *that is injective and preserves addition, multiplication, and ordering.*

Proof. Define $i\colon \mathbb{R} \to {}^*\mathbb{R}$ as $i(r) = [(r, r, r, ...)]_{\mathcal{U}}$. To show that i is injective, consider the contrapositive of injectivity. If $r \neq r'$, then the sequences $(r, r, r, ...)$ and $(r', r', r', ...)$ are not modulo $\mathcal{U}$ equivalent as they differ in all entries. To show that i respects addition and multiplication, meaning that $i(r) + i(s) = i(r + s)$ and that $i(r) \cdot i(s) = i(r \cdot s)$ (where addition and multiplication are done in the respective sets), it suffices to see that

$$i(r) + i(s) = [(r, r, r, ...)]_{\mathcal{U}} + [(s, s, s, ...)]_{\mathcal{U}} = [(r + s, r + s, r + s, ...)]_{\mathcal{U}} = i(r + s)$$

and that the corresponding statement for multiplication is the same. To show that i respects ordering, meaning that if $r \leq s$ then $i(r) \leq i(s)$, let $r \leq s$ and then note that $[(r, r, r, ...)]_{\mathcal{U}} \leq [(s, s, s, ...)]_{\mathcal{U}}$ since $\{n \in \mathbb{N} \mid r \leq s\} = \mathbb{N} \in \mathcal{U}$. $\square$

This embedding will be crucial in our development of nonstandard analysis since it will allow us to represent real numbers as hyperreals. Also, note that i also follows directly from 3.4, which is also a crucial insight. Next, we will prove the existence of two kinds of hyperreals that will make nonstandard analysis possible: unlimited and infinitesimal numbers.

Theorem 4.3. *There exists a hyperreal number* ω *such that* $|\omega| > i(r)$ *for all* $r \in \mathbb{R}^+$. *Similarly, there exists a hyperreal number* ε *such that* $0 < |\varepsilon| < i(r)$ *for all* $r \in \mathbb{R}^+$.

Note that we define absolute value in ${}^*\mathbb{R}$ using 3.4. Furthermore, the absolute value function holds all of the same properties it has in $\mathbb{R}$, including the triangle inequality. The proof of this comes from applying this property to every entry in the sequence and then using the properties of $\mathcal{U}$ as an ultrafilter.

Proof. Let $\omega = [(a_n = n)]_{\mathcal{U}}$, so $\omega = [(1, 2, 3, ...)]_{\mathcal{U}}$. Fix some $r \in \mathbb{R}$, note that since $\mathbb{R}$ is Archimedean, we have that $r \leq m$ for some $m \in \mathbb{N}$. Let $i(m) = [(m_n = m)]_{\mathcal{U}}$; it follows that for all $n \leq m$ we have $a_n \leq m$. However, this implies that $\{n \in \mathbb{N} \mid a_n \leq m_n\}$ is finite, so $\{n \in \mathbb{N} \mid a_n > m_n\} \in \mathcal{U}$ since $\mathcal{U}$ contains all the cofinite sets. Thus, $\omega > i(m) \geq i(r)$. Since r was chosen arbitrarily, his holds for all $r \in \mathbb{R}^+$. Next, let $\varepsilon = [(1, \frac{1}{2}, \frac{1}{3}, ...)]_{\mathcal{U}} = \omega^{-1}$. Since ${}^*\mathbb{R}$ is a field it follows that $\varepsilon = \omega^{-1} < i(r)^{-1}$, but since this holds for all $r \in \mathbb{R}^+$, we also get that $\varepsilon < i(r)$ for all $r \in \mathbb{R}^+$. □

This motivates the following classification for hyperreal numbers:

Definition 4.4. A hyperreal number ω is said to be *unlimited* if $|\omega| > i(r)$ for all $r \in \mathbb{R}^+$.

Definition 4.5. A hyperreal number ε is said to be *infinitesimal* if $|\varepsilon| < i(r)$ for all $r \in \mathbb{R}^+$.

Definition 4.6. A hyperreal number n is said to be *finite* if there exists $r, s \in \mathbb{R}^+$ such that $i(r) \leq |n| \leq i(s)$.

Definition 4.7. A hyperreal number x is said to be *limited* if it is not unlimited.

For example, the ω presented in 4.3 is *unlimited* and the ε presented in 4.3 is *infinitesimal.* Also of note is that $0 \in \mathbb{R}$ is the only real number whose image under i is infinitesimal.

It is important to consider then how addition and multiplication of hyperreals plays with these definitions. As a general rule, the addition of any hyperreal to an unlimited hyperreal is also unlimited, the addition of any limited hyperreal to a finite hyperreal is also finite, and only the addition of two infinitesimals is infinitesimal. As for the products, the product of a finite hyperreal with and infinitesimal is infinitesimal; the product of two finite hyperreals is finite, and the product of an unlimited hyperreal and an unlimited or a finite hyperreal is also unlimited. Finally, the product of an unlimited hyperreal and an infinitesimal follows no general rules and depends on the numbers themselves. The proofs for all of these claims follow from the definitions given above. With this in mind, we can show that ${}^*\mathbb{R}$ is not complete:

Theorem 4.8. *$i(\mathbb{R}) \subset {}^*\mathbb{R}$ is a nonempty bounded set without supremum.*

Proof. Note that $i(\mathbb{R})$ is bounded above by any *unlimited* hyperreal, and it is of course nonempty. However, assume for contradiction that $i(\mathbb{R})$ has a supremum and set $\omega = \sup i(\mathbb{R})$. Then ω must be unlimited, as otherwise it wouldn't be an upper bound, but this also implies that $\omega - i(1)$ is unlimited and, hence, an upper bound of $i(\mathbb{R})$. However, it is clear that $\omega - i(1) <_{\mathcal{U}} \omega$. Therefore, ω is not the supremum of $i(\mathbb{R})$. Contradiction. □

Now that we have defined what it means for a hyperreal to be infinitesimal, we can introduce an important notion between hyperreal numbers:

Definition 4.9. Two hyperreal numbers are said to be *close*, denoted as $x \simeq y$, if their difference is infinitesimal.

Theorem 4.10. *Closeness, as defined in 4.9, is an equivalence relation.*

Proof. The relation is clearly reflexive since for any hyperreal x, we have that $x - x = i(0)$ and we know that $i(0)$ is infinitesimal. Closeness is also reflexive since if $x \simeq y$ then $|x - y| < i(r)$ for all $r \in \mathbb{R}^+$, but this also implies that $|y - x| < i(r)$ for all $r \in \mathbb{R}^+$ due to the properties of the absolute value; thus $y \simeq x$. To show that closeness is transitive, set $x \simeq y$ and $y \simeq z$, then fix any real number $r \in \mathbb{R}^+$; it holds that $|x - y| < \frac{i(r)}{2}$ and $|y - z| < \frac{i(r)}{2}$, so by triangle inequality we obtain $|x - z| < i(r)$, so $x \simeq z$ since this holds for all $r \in \mathbb{R}^+$. □

From this notion of closeness, we can introduce some new concepts about the hyperreals.

Definition 4.11. For any hyperreal number x_0, we define the *halo*, sometimes called the *monad*, of x_0 as the set of all hyperreals that are close to x_0. More formally, $hal(x_0) = \{x \in {}^*\mathbb{R} \mid x \simeq x_0\}$.

Definition 4.12. For any hyperreal number x_0, we define the *galaxy* of x_0 as the set of all hyperreals x such that $x - x_0$ is limited. More formally, $gal(x_0) = \{x \in {}^*\mathbb{R} \mid x - x_0 \text{ is limited}\}$.

A few important remarks come from this last definition. Note that the set of all infinitesimals is precisely $hal(0)$ and, thus, we will normally denote a number as infinitesimal by writing $x \simeq 0$. It is clear that for 0 we mean $i(0)$, but to ease notation we will omit the i whenever it is clear that the number should be hyperreal. Also, we will denote unlimited numbers using the notation $\omega \simeq \infty$, which is meant to capture the notion that unlimited numbers are greater than all real numbers.

It is worth noting that we have a way to relate real numbers to specific hyperreal numbers (via the embedding defined above), but so far we have no way of relating arbitrary hyperreal numbers to some specific real number. The next theorem is meant to resolve that issue:

Theorem 4.13. *Given any limited hyperreal x, there exists a unique real number r, called the shadow or standard part of x, such that $i(r) \simeq x$. We denote this relation by writing $sh(x) = r$.*

Proof. Fix some hyperreal x and let $A = \{r \in \mathbb{R} \mid i(r) < x\}$. Since x is limited, it follows that A is nonempty and bounded above; thus, A attains supremum. Let $a = \sup A$; we claim that $a = sh(x)$. To show this, fix $\varepsilon \in \mathbb{R}^+$ and consider $k - i(a)$; since $a = \sup A$, it follows that $0 \leq k - i(a)$, which implies that $i(\varepsilon) < k - i(a)$. Notice that since $a = \sup A$ and $\varepsilon > 0$, it follows that $k < i(a + \varepsilon)$, which gives us that $k - i(a) < i(\varepsilon)$. Thus, $|k - a| < i(\varepsilon)$. Since this holds for all $\varepsilon \in \mathbb{R}^+$, we have that $k \simeq i(a)$.
To prove that a is indeed unique, assume that $k \simeq i(a)$ and $k \simeq i(a')$. This implies that $i(a) \simeq i(a')$. Thus, $i(a - a')$ is infinitesimal, but note that this is only possible if $a - a' = 0$. Thus, $a = a'$. □

Theorem 4.14. *The shadow operator respects addition, multiplication and ordering.*

Proof. First we will show that the shadow operator respects addition. Fix any $a, b \in {}^*\mathbb{R}$, we want to show that $sh(a) + sh(b) = sh(a + b)$. Let $sh(a) = a$ and $sh(b) = d$; thus, $a \simeq i(c)$ and $b \simeq i(d)$. It suffices to show that $a + b \simeq i(c) + i(d)$. Fix $\varepsilon \in \mathbb{R}^+$ and notice that $|a - i(c)| < \frac{i(\varepsilon)}{2}$ and $|b - i(d)| < \frac{i(\varepsilon)}{2}$, by triangle inequality we get that $|(a + b) - (i(c) + i(d)| < i(\varepsilon)$, which shows that $a + b \simeq i(c + d)$ and thus completes the proof. To show that $sh(a)sh(b) = sh(ab)$ we again let $a \simeq i(c)$ and $b \simeq i(d)$ and fix $\varepsilon \in \mathbb{R}^+$ and let

$$|a - i(c)| < \min\Big(1, \frac{i(\varepsilon)}{2(|i(d)| + 1)}\Big) \quad \text{and} \quad |b - i(d)| < \frac{i(\varepsilon)}{2(|i(c)| + 1)}$$

This implies that $|ab - i(c)i(d)| < i(\varepsilon)$, so $ab \simeq i(c)i(d)$. The proof of this statement follows similarly from the proof of this same statement in $\mathbb{R}$, which should not be foreign to the reader, and applying the properties of $\mathcal{U}$ as an ultrafilter.
Finally, we want to show that if $a < b$, then $sh(a) \leq sh(b)$. We have two cases, either $a \simeq b$ or $a \not\simeq b$. If $a \simeq b$ and $c = sh(a)$, then $c = sh(b)$ since closeness is transitive; thus, the order cannot be strict. Now, assume that $a < b$ but $a \not\simeq b$; fix $a \simeq i(c)$ and $b \simeq i(d)$. This implies that $hal(a) \cap hal(b) = \varnothing$, but since ${}^*\mathbb{R}$ is totally ordered, it follows that every point in one set must be less that every point in the other set. Since $a < b$, this gives us that $i(c) < i(d)$, so $c < d$, which implies that $sh(a) < sh(b)$. □

Finally, we introduce the concept of extensions. So far, we are able to relate numbers between the reals and the hyperreals. However, for the rest of the paper our focus will be on functions, so it is important to formalize the concept of transferring a function from $\mathbb{R}$ to ${}^*\mathbb{R}$. Although 3.4 deals with this matter, the introduction of these new concepts will ease notation and facilitate understanding.

Definition 4.15. Given some set $A \subset \mathbb{R}$, the *natural extension* of A to ${}^*\mathbb{R}$ is a set ${}^*A \subset {}^*\mathbb{R}$ defined as follows:

$$[(r_n)]_{\mathcal{U}} \in {}^*A \quad \text{if, and only if,} \quad \{n \in \mathbb{N} \mid r_n \in A\} \in \mathcal{U}$$

Again, we appeal to the intuition behind ultrafilters to claim this extension is 'natural,' by saying that a hyperreal should belong to the set if the set of entries in the sequence that belong to the original set is 'big.' Note also that $i(A) \subset {}^*A$; however, there are also new hyperreals that belong to ${}^*A$, a property that will actually prove to be useful. Next, we consider the natural extension of a function.

Definition 4.16. Given some function $f\colon A \to \mathbb{R}$, the *natural extension* of f to ${}^*\mathbb{R}$ is a function ${}^*f\colon {}^*A \to {}^*\mathbb{R}$ defined as follows:

$${}^*f([(r_n)]_{\mathcal{U}}) = [(f(r_n))]_{\mathcal{U}}$$

Now that we have defined these notions it is important to clear up what we mean for "transfer." We know from the transfer principle that a first order statement is satisfied by $\mathbb{R}$ if, and only if, it is satisfied by ${}^*\mathbb{R}$. However, note that we have not explicitly said how this would happen, although 3.4 makes it somewhat clear. However, in more explicit terms here is how the rule follows: we take constants in $\mathbb{R}$ to their image under i, we take functions in $\mathbb{R}$ to their natural extension and relations in $\mathbb{R}$ to their natural extension in ${}^*\mathbb{R}$. For variables, we extend the sets to which they belong and revert any of these changes when transferring back to $\mathbb{R}$. We will perform transfers in the next section so the reader will be able to see some examples of these transfers then. With this last concept cleared up, we can move on to consider the concepts of non-standard analysis.

5. Continuity and Differentiability

The first concept that we will introduce is the idea of continuity. Intuitively, continuity implies that the function takes points which are 'close' to each other to point that are also 'close' to each other. The problem with this notion in $\mathbb{R}$ is how close is 'close.' A solution to this problem is the currently accepted $\varepsilon - \delta$ definition. However, in the hyperreals we have a formalized notion of closeness, which we will use to formalize the intuitive concept of continuity.[4]

Definition 5.1. Let $A \subset {}^*\mathbb{R}$, a function $f\colon A \to {}^*\mathbb{R}$ is said to be *micro-continuous* at $x_0 \in A$ if $x \simeq x_0$ implies $f(x) \simeq f(x_0)$ for all $x \in A$.

Note that if $hal(x_0) \in A$, then f is micro-continuous at x_0 if, and only if, $f(hal(x_0)) \subset hal(f(x_0))$. The first important result of non-standard analysis follows:

Theorem 5.2. *A function $f\colon A \to \mathbb{R}$ is continuous at $c \in \mathbb{R}$ if, and only if, the natural extension of f is micro-continuous at $i(c)$.*

Proof. $\Rightarrow$ Assume that f is continuous at $c \in \mathbb{R}$ and let $x_0 \simeq i(c)$. This implies that

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall x \in A)(|x - c| < \delta \to |f(x) - f(c)| < \varepsilon)$$

Next, fix an arbitrary $\varepsilon \in \mathbb{R}^+$, so there is a $\delta \in \mathbb{R}^+$ that satisfies the condition above. Applying the transfer principle we get that

$$(\forall x \in {}^*A)(|x - i(c)| <_{\mathcal{U}} i(\delta) \to |{}^*f(x) - {}^*f(i(c))| <_{\mathcal{U}} i(\varepsilon))$$

Since $x_0 \simeq i(c)$, we have that $|x_0 - i(c)| <_{\mathcal{U}} i(\delta)$. Therefore, it follows that $|f(x_0) - f(i(c))| <_{\mathcal{U}} i(\varepsilon)$. Since ε was arbitrary, this holds for all $\varepsilon \in \mathbb{R}^+$, which implies that ${}^*f(x_0) \simeq {}^*f(i(c))$.
$\Leftarrow$ Assume that ${}^*f$ is micro-continuous at $i(c)$. Fix an arbitrary $\varepsilon \in \mathbb{R}^+$ and consider a $\delta \in {}^*\mathbb{R}^+$ with $\delta \simeq 0$. Thus, for all $x \in {}^*A$, if $|x - i(c)| <_{\mathcal{U}} \delta$, then $x \simeq i(c)$. This implies that ${}^*f(x) \simeq {}^*f(i(c))$, but this also gives us that $|{}^*f(x) - {}^*f(i(c))| <_{\mathcal{U}} i(\varepsilon)$. Hence we get that

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*A)(|x - i(c)| <_{\mathcal{U}} \delta \Rightarrow |{}^*f(x) - {}^*f(i(c))| <_{\mathcal{U}} i(\varepsilon))$$

and applying the transfer principle we get

$$(\exists \delta \in \mathbb{R}^+)(\forall x \in A)(|x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon)$$

which is exactly the desired statement. Since this holds of any $\varepsilon \in \mathbb{R}^+$, we have that f is continuous at c. $\square$

With notion of micro-continuity at hand, it is possible to prove two important theorems regarding continuous functions.

Theorem 5.3. *(Intermediate Value Theorem) Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then, for every point d between $f(a)$ and $f(b)$ there exists a $c \in (a, b)$ such that $f(c) = d$.*

Proof. The theorem is trivial if $f(a)$ or $f(b)$ are equal to d, so assume without loss of generality that $f(a) < d < f(b)$. Next consider the following statement:

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists n \in \mathbb{N}^0)(f(a + n\varepsilon) \leq d \wedge ((\forall m \in \mathbb{N}^0)(m > n \to f(a + m\varepsilon) > d)))$$

[4]The paper assumes that the reader is comfortable with the basic notions of standard single variable calculus.

This statement tells us that given a positive real number, there is a maximal natural number, which might be zero, such that $f(a+n\varepsilon) \leq d$. By transfer we get that

$$(\forall \varepsilon \in {}^*\mathbb{R}^+)(\exists n \in {}^*\mathbb{N}^0)({}^*f(i(a)+n\varepsilon) \leq_{\mathfrak{U}} i(d) \wedge ((\forall m \in {}^*\mathbb{N}^0)(m >_{\mathfrak{U}} n \to {}^*f(i(a)+m\varepsilon) >_{\mathfrak{U}} i(d))))$$

Given this, set $\varepsilon = \dfrac{i(b-a)}{\omega}$ for some $\omega \simeq \infty$ and note that $\varepsilon \simeq 0$. Thus, there exists some $n \in {}^*\mathbb{N}^0$ satisfying the condition above. Note that in this case n is unlimited but $n < \omega$. Let $c = sh(i(a)+n\varepsilon)$ and note that $a < c < b$; we claim that $f(c) = d$. To show this, notice that $i(a)+n\varepsilon \simeq i(a)+(n+1)\varepsilon$ since their difference is just ε, which implies that $c = sh((a)+(n+1)\varepsilon)$ since closeness is transitive. Thus, since ${}^*f$ is micro-continuous, we have that ${}^*f(i(c)) \simeq {}^*f(i(a)+n\varepsilon) \simeq {}^*f(i(a)+(n+1)\varepsilon)$, but we also have that

$${}^*f(i(a)+n\varepsilon) \leq_{\mathfrak{U}} i(d) < {}^*f(i(a)+(n+1)\varepsilon)$$

However, this also implies that ${}^*f(i(a)+n\varepsilon) \simeq i(d)$. By transitivity we get that ${}^*f(i(c)) \simeq i(d)$ and, thus, $f(c) = d$ since ${}^*f$ is f applied pointwise. ☐

Theorem 5.4. *(Extreme Value Theorem) Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Then there exist $c, d \in [a,b]$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [a,b]$.*

Proof. We will only show that f attains a maximum, since the proof that it attains a minimum is similar. Consider the following statement:

$$(\forall n \in \mathbb{N})(\exists m \in \mathbb{N} \wedge m \leq n)(\forall k \in \mathbb{N})\left(k \leq n \to f\left(a + k\frac{b-a}{n}\right) \leq f\left(a + m\frac{b-a}{n}\right)\right)$$

This implies that any partition of $[a,b]$ into equal subintervals has a partition point that is maximum. By transfer we have that

$$(\forall n \in {}^*\mathbb{N})(\exists m \in {}^*\mathbb{N} \wedge m \leq_{\mathfrak{U}} n)(\forall k \in {}^*\mathbb{N})\left(k \leq_{\mathfrak{U}} n \to {}^*f\left(i(a) + k\frac{i(b-a)}{n}\right) \leq_{\mathfrak{U}} {}^*f\left(i(a) + m\frac{i(b-a)}{n}\right)\right)$$

Fix $\omega \in {}^*\mathbb{N}$, $\omega \simeq \infty$; then there exists some $m \in {}^*\mathbb{N}$ satisfying the conditions above. Define $d = sh\left(a + m\frac{b-a}{n}\right)$, we claim that this is the desired point. Notice that since f is continuous we have that ${}^*f\left(a + m\frac{b-a}{n}\right) \simeq {}^*f(i(d))$. Now, we must show that $f(d)$ is actually maximal; fix some $x \in [a,b]$ and notice that

$$(\exists n_0 \in {}^*\mathbb{N}^0)\left(a + n_0\frac{b-a}{\omega} \leq_{\mathfrak{U}} i(x) \wedge \left((\forall m \in {}^*\mathbb{N}^0)\left(n_0 <_{\mathfrak{U}} m \leq_{\mathfrak{U}} \omega \to i(x) <_{\mathfrak{U}} a + m\frac{b-a}{\omega} \leq_{\mathfrak{U}} i(b)\right)\right)\right)$$

This statement follows from transfer. Note that it only states that x exists between two points of the partition, which is clearly true in $\mathbb{R}$ and it transfers into ${}^*\mathbb{R}$. Note that this implies that $i(x) \simeq a + n_0\frac{b-a}{\omega}$, which also gives us that ${}^*f(i(x)) \simeq {}^*f\left(a + n_0\frac{b-a}{\omega}\right)$ since f is continuous. However, note that $n_0 \leq_{\mathfrak{U}} \omega$, so

$${}^*f(i(x)) \simeq {}^*f\left(a + n_0\frac{b-a}{\omega}\right) \leq_{\mathfrak{U}} {}^*f\left(a + m\frac{b-a}{n}\right) \simeq {}^*f(i(d))$$

This implies that $f(x) \leq f(d)$ since both x and d are real. Since this holds for all $x \in [a,b]$, the proof is complete. ☐

Next, we will give a non-standard definition of the derivative and prove some properties using non-standard analysis. Intuitively, the derivative is supposed to be the slope of the tangent line at some point x_0. However, the problem is that to define such a line we require at least two points, while we only have one. In standard analysis, this is resolved by letting a second point x approach the first point using limits and looking at the slope of the line when the two points are arbitrarily close together. In non-standard analysis, we use a similar approach but instead make use of infinitesimals to define the second point.

Definition 5.5. Let $f\colon A \to \mathbb{R}$ be a function. We say that f is differentiable at $x_0 \in A$ if there exists a unique $L \in {}^*\mathbb{R}$ such that for every nonzero infinitesimal ε we have

$$\frac{{}^*f(i(x_0)+\varepsilon) - {}^*f(i(x_0))}{\varepsilon} \simeq L$$

If so, we define the derivative of f at x_0 to be the shadow of L, meaning $f'(x_0) = sh(L)$.

While the definition should appear intuitive to the reader, we still must show that it is equivalent to the standard definition using limits.

Theorem 5.6. *A function f is differentiable at x_0 as defined in 5.5 if, and only if, it is differentiable at x_0 in the usual sense. Furthermore, if it is differentiable, the derivatives are the same.*

Proof. $\Rightarrow$ Assume f is differentiable at $x_0 \in A$ as defined in 5.5 such that $f'(x_0) = sh(L)$ for some $L \in {}^*\mathbb{R}$. Fix some $\varepsilon \in \mathbb{R}$ and consider some $\delta \in {}^*\mathbb{R}^+$ with $\delta \simeq 0$, so if $i(0) \leq |\mu| < \delta$, then $\mu \simeq 0$. Since f is differentiable we get that

$$\left|\frac{{}^*f(i(x_0)+\mu) - {}^*f(i(x_0))}{\mu} - sh(L)\right| <_{\mathfrak{U}} i(\varepsilon)$$

since the difference is infinitesimal (closeness is transitive). Therefore, we have

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall \mu \in {}^*\mathbb{R})\left(i(0) <_{\mathfrak{U}} |\mu| <_{\mathfrak{U}} \delta \rightarrow \left|\frac{{}^*f(i(x_0)+\mu) - {}^*f(i(x_0))}{\mu} - sh(L)\right| <_{\mathfrak{U}} i(\varepsilon)\right)$$

which by transfer gives us that

$$(\exists \delta \in \mathbb{R}^+)(\forall \mu \in \mathbb{R})\left(0 < |\mu| < \delta \rightarrow \left|\frac{f(x_0+\mu) - f(x_0)}{\mu} - sh(L)\right| < \varepsilon\right)$$

Since this holds for all $\varepsilon \in \mathbb{R}^+$ we have that f is differentiable at x_0 and its derivative is $sh(L) \in \mathbb{R}$.
$\Leftarrow$ Assume that f is differentiable in the usual sense with $f'(x_0) = L$ for some $L \in \mathbb{R}$. Fix an $\varepsilon \in \mathbb{R}^+$ and by differentiability we get that there is a $\delta \in \mathbb{R}^+$ such that

$$(\forall h \in \mathbb{R})\left(0 < |h| < \delta \rightarrow \left|\frac{f(x_0+h) - f(x_0)}{h} - L\right| < \varepsilon\right)$$

which by transfer gives us that

$$(\forall h \in {}^*\mathbb{R})\left(0 <_{\mathfrak{U}} |h| <_{\mathfrak{U}} i(\delta) \rightarrow \left|\frac{{}^*f(i(x_0)+h) - {}^*f(i(x_0))}{h} - i(L)\right| <_{\mathfrak{U}} i(\varepsilon)\right)$$

Now, assume that $h \simeq 0$, then the statement above follows. Since this holds for all $\varepsilon \in \mathbb{R}^+$, we get that

$$\frac{{}^*f(i(x_0)+h) - {}^*f(i(x_0))}{h} \simeq i(L) \quad \text{for all } h \simeq 0$$

This implies that f differentiable in the non-standard sense and $f'(x_0) = sh(i(L)) = L$. □

As in standard analysis, we can redefine the derivative by looking at points infinitely close to x_0.

Definition 5.7. Let $f\colon A \to \mathbb{R}$ be a function. We say that f is differentiable at $x_0 \in A$ if there exists a unique $L \in {}^*\mathbb{R}$ such that for every $y \in hal(x_0) \setminus \{x_0\}$

$$\frac{{}^*f(y) - {}^*f(i(x_0))}{y - i(x_0)} \simeq L$$

If so $f'(x_0) = sh(L)$.

The fact that these two definitions are in fact equivalent should be evident, since the fact that $y \in hal(x_0) \setminus \{x_0\}$ implies that $i(x) - y$ is a nonzero infinitesimal.

With this new definition of the derivative we can easily motivate Leibniz's notation for derivatives, albeit with some abuse of notation. Let dx be a nonzero infinitesimal, then we can define Δf as the increment of f at some point x by setting $\Delta f = {}^*f(i(x)+dx) - {}^*f(i(x))$. Notice that it follows that

$$\frac{\Delta f}{dx} \simeq f'(x) \quad \text{for all nonzero infinitesimals } dx.$$

Furthermore, notice that since $f'(x)$ is limited and $\Delta f = \frac{\Delta f}{dx}dx$, we have that $\Delta f \simeq 0$ since the product of a limited number and an infinitesimal is infinitesimal. Note that this proves an important fact.

Theorem 5.8. *If f is differentiable at x_0, then f is continuous at x_0.*

And by setting $df = f'(x)dx$, we see that

$$\frac{\Delta f}{dx} \simeq \frac{df}{dx} = f'(x)$$

Finally, of note is the fact that this definition of the derivative allows us to give proofs of all the known differentiation rules without appealing to the limits definition. However, the proofs are left to the reader.

6. Integration

When integration is first introduced in a standard calculus course the intuition behind it generally proceeds as follows: we consider a bounded function f over a certain interval $[a, b]$; then we see that we can approximate the value of the area under the curve by means of rectangles normally by partitioning $[a, b]$ into smaller intervals of the same length and then using the image under f of the first point or last point of this subintervals as the length of the rectangle. We quickly notice that as we make the subintervals smaller, the better our approximation of the total area is. Thus, when the subintervals have infinitesimal length, we should obtain the actual value of the area under the curve.
However, there are some issues with this notion, which we will regard as the "partition problem." To introduce this issue we will first present the common notions of integration in the standard sense. From here on out, we assume f to be a bounded function defined on some interval $[a, b] \subset \mathbb{R}$.

Definition 6.1. A *Partition* of an interval $[a, b]$ is a finite set $P = \{x_0, x_1, \dots x_n\} \subset [a, b]$ such that $x_0 = a$, $x_n = b$ and $x_{i-1} < x_i$ for all $0 \le i \le n$.

Let M_i and m_i be the least upper bound and greatest lower bound of f in the interval $[x_{i-1}, x_i]$ then we define

$$L(f, P) = \sum_{i=0}^{n} m_i(x_i - x_{i-1}) \qquad U(f, P) = \sum_{i=0}^{n} M_i(x_i - x_{i-1})$$

$L(f, P)$ is known as the lower sum and $U(f, P)$ is known as the upper sum. Note that $L(f, P) \le U(f, P)$ for all partitions P. Graphically, $L(f, P)$ takes the smallest value of f in each subinterval as the length of the rectangle while $U(f, P)$ takes the highest value. Note that for any bounded function we have

$$L(f, P) \le U(f, Q) \qquad \text{for any two partitions } P \text{ and } Q \text{ of } [a, b]$$

We can use these values to define integration as follows:

Definition 6.2. We say that a function is *Riemann Integrable* on an interval $[a, b]$ if given any $\varepsilon > 0$, there exits a partition P such that $L(f, P) - U(f, P) < \varepsilon$. This is sometimes know as the *Cauchy Criterion* for integration.

Intuitively, it means that we can make the subintervals so small that the difference between the lower and upper sums is less than any value we desire. Note that this definition resembles the limit definition. This also means that over all partitions of $[a, b]$, $L(f, P)$ is bounded above and $U(f, P)$ is bounded below; furthermore, it implies that these bounds are equal and we call this value the integral of f from a to b. The proof of this fact should be familiar to the reader.

Now, if we want to transfer this notion to nonstandard analysis, we might be tempted to make it so that the difference between any two points in the partition is infinitesimal, maybe by dividing $b-a$ by un unlimited number ω. This would imply that P would not only be infinite but also uncountable, which brings up the question of how do you sum over uncountably many elements? This is precisely the partition problem, that it is difficult to define a rigorous way of summing over a partition which has as many elements as $\mathbb{R}$ itself. One solution, presented by Goldblatt in [6], is to consider only partitions where all consecutive points are equidistant (meaning $x_i - x_{i-1} = \Delta x$) except for the last two points whose difference might be less that Δx. Then, we can think of $L(f, P)$ and $U(f, P)$ as functions defined on $\mathbb{R}^+$ determined by this particular value Δx and extend this functions to $^*\mathbb{R}$ in the natural way. However, this only masks the problem; it doesn't solve it since it doesn't really tell us how these values are computed or what their meaning is. Thus, we will take a slight detour, which is less intuitive but more rigorous, by introducing the notions of *hyperfinite sets* and *hyperfinite sums*.

Definition 6.3. Given a sequence (A_n) of subsets of $\mathbb{R}$, we define a set $[(A_n)]_{\mathcal{U}} \subset {}^*\mathbb{R}$ by the following rule

$$[(r_n)]_{\mathcal{U}} \in [(A_n)]_{\mathcal{U}} \Leftrightarrow \{n \in \mathbb{N} \mid r_n \in A_n\} \in \mathcal{U}$$

A set $A \subset {}^*\mathbb{R}$ is called *internal* if it can be defined in this way. Otherwise, we say that A is *external*. Furthermore, note that if $A_n = A \subset \mathbb{R}$ for all $n \in \mathbb{N}$, then ${}^*A = [(A_n)]_{\mathcal{U}}$.

Definition 6.4. An internal set $[(A_n)]_{\mathcal{U}} \subset {}^*\mathbb{R}$ is called *hyperfinite* if $\{n \in \mathbb{N} \mid A_n \text{ is finite}\} \in \mathcal{U}$. In this case, we define the cardinality of $[(A_n)]_{\mathcal{U}}$ as follows: $|[(A_n)]_{\mathcal{U}}| = [(|A_n|)]_{\mathcal{U}} \in {}^*\mathbb{N}$.

Definition 6.5. Given an internal set $A = [(A_n)]_{\mathcal{U}} \subset {}^*\mathbb{R}$ and a function f such that A_n is in the domain of f for all $n \in \mathbb{N}$, we can define the *hyperfinite sum* of f on A as follows

$$\sum_{x\in A} {}^*f(x) = \sum_{[(x_n)]_{\mathcal{U}}\in A} {}^*f([(x_n)]_{\mathcal{U}}) = \left[\left(\sum_{x_n\in A_n} f(x_n)\right)\right]_{\mathcal{U}}$$

Note that this definition makes sense since A_n is finite for most $n \in \mathbb{N}$. Also note that this sum respects most of the properties normal sums do in $\mathbb{R}$. This definition also follows the same idea we've been using all throughout nonstandard analysis of defining concepts in ${}^*\mathbb{R}$ as the pointwise application of its analogoes in $\mathbb{R}$. With this notion, we can begin to define non-standard integration.

Definition 6.6. Given a positive infinitesimal $dx = [(\varepsilon_n)]_{\mathcal{U}}$, for each ε_n we define a partition P_n by noting that

$$(\forall\varepsilon \in \mathbb{R}^+)(\exists k \in \mathbb{N})(a + k\varepsilon \le b \wedge ((\forall m \in \mathbb{N})(m > k \to a + m\varepsilon > b)))$$

so $P_n = \{a, a + \varepsilon_n, a + 2\varepsilon_n, \ldots, a + k_n\varepsilon_n, b\}$, where k_n is the corresponding k, as defined above, for ε_n. This gives us a *hyperfinite partition* $P = [(P_n)]_{\mathcal{U}}$ of $[a, b]$

Note that if $\varepsilon_n > a - b$, then $P_n = \{a, b\}$; similarly, if $\varepsilon_n < 0$, we also define $P_n = \{a, b\}$. Furthermore, for all $0 \le i \le k_n - 1$ we have that $x_i - x_{i-1} = \varepsilon_n$ and $x_n - x_{n-1} \le \varepsilon_n$.
This will allow us to sum in the usual sense. First let's redefine $L(f, P)$ and $U(f, P)$ as functions of $\varepsilon \in \mathbb{R}^+$ by saying that $L(f, \varepsilon) = L(f, P)$ where P is the partition defined in 6.6, same for $U(f, \varepsilon)$. Therefore, we can extend this functions as follows:

Definition 6.7. Given a positive hyperreal $dx = [(\varepsilon_n)]$ we define the *extensions of the upper and lower sums* of f on $[a, b]$ as

$${}^*U(f, dx) = \left[\left(\sum_{i_n=0}^{k_n} M_{i_n}\varepsilon_n\right)\right]_{\mathcal{U}} = [(U(f, \varepsilon_n))]_{\mathcal{U}}$$

$${}^*L(f, dx) = \left[\left(\sum_{i_n=0}^{k_n} m_{i_n}\varepsilon_n\right)\right]_{\mathcal{U}} = [(L(f, \varepsilon_n))]_{\mathcal{U}}$$

Note that we would like ${}^*U(f, dx)$ to be defined as $\displaystyle\sum_{i=0}^{\omega} M_i\ dx$ where $dx = \dfrac{b-a}{\omega}$. This definition captures that sense by letting A_n, as defined in 6.5, be the set containing all the M_i for the partition P_n. Now, we can define integrability:

Definition 6.8. We say that a function f is **Riemann integrable* on $[a, b]$ if given any positive infinitesimals dx, we have that ${}^*U(f, dx) \simeq {}^*L(f, dx)$. Furthermore, we say that

$$\int_a^b f = sh\left({}^*U(f, dx)\right) = sh\left({}^*L(f, dx)\right) \text{ for any nonzero infinitesimal } dx.$$

It might not be evident to the reader why this should be the value of the integral, or at least it does not seem to follow intuitively. First of all, note that since $dx \simeq 0$, it follows that the sequence (ε_n) converges to 0; this implies that given any positive real number δ, there are infinitely many elements of the sequence which are less than δ. In that sense, we are making the partitions infinitesimals. Nonetheless, a proof might be more convincing to the reader. But first, it is convenient to redefine the *Cauchy Criterion* in terms of positive real numbers.

Definition 6.9. We say that a function is *Riemann integrable* on an interval $[a,b]$ if

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall \Delta x \in \mathbb{R}^+)(\Delta x < \delta \to U(f,\Delta x) - L(f,\Delta x) < \varepsilon)$$

The proof that this is equivalent to the *Cauchy Criterion* should be easy for the reader to follow, so we will only give an overview. If 6.9 is satisfied then the *Cauchy Criterion* is satisfied by creating a partition using 6.6. Conversely, if the *Cauchy Criterion* is satisfied, then let $\delta = min(x_i - x_{i-1})$ over all possible i in the partition P that the criterion guarantees. It is left to the reader to show that this δ is sufficient for 6.9.

Theorem 6.10. *A function f satisfies the Cauchy Criterion for integration if, and only if, f satisfies 6.8.*

Proof. $\Rightarrow$ Fix $\varepsilon \in \mathbb{R}^+$ then there is a $\delta \in \mathbb{R}^+$ such that

$$(\forall \Delta x \in \mathbb{R}^+)(\Delta x < \delta \to U(f,\Delta x) - L(f,\Delta x) < \varepsilon)$$

. By transfer we get that

$$(\forall dx \in {}^*\mathbb{R}^+)(dx <_{\mathfrak{U}} i(\delta) \to {}^*U(f,dx) - {}^*L(f,dx) <_{\mathfrak{U}} i(\varepsilon))$$

Note that this implies that for all $dx \in {}^*\mathbb{R}^+$, $dx \simeq 0$, we have that ${}^*U(f,dx) \simeq {}^*L(f,dx)$ since this statement holds for all $\varepsilon \in \mathbb{R}^+$. We can ignore the absolute value since ${}^*U(f,dx) \leq {}^*L(f,dx)$ for all dx. This implies that f satisfies 6.8.

$\Leftarrow$ Fix $\varepsilon \in \mathbb{R}^+$ and put $\delta \in {}^*\mathbb{R}^+$ with $\delta \simeq i(0)$. Therefore, if $dx < \delta$, then ${}^*U(f,dx) - {}^*L(f,dx) < i(\varepsilon)$. Thus we get that

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall dx \in {}^*\mathbb{R}^+)(dx <_{\mathfrak{U}} i(\delta) \to {}^*U(f,dx) - {}^*L(f,dx) <_{\mathfrak{U}} i(\varepsilon))$$

. By transfer we have

$$(\exists \delta \in \mathbb{R}^+)(\forall \Delta x \in \mathbb{R}^+)(\Delta x < \delta \to U(f,\Delta x) - L(f,\Delta x) < \varepsilon)$$

which is the desired result and completes the proof. $\square$

Thus, it only remains to show that the value of the integral is indeed the same. To do this, we will characterize that value of the integral as follows:

Lemma 6.11. *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. Given some $I \in \mathbb{R}$ we have that $I = \int_a^b f$ if, and only if,*

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall \Delta x \in \mathbb{R}^+)(\Delta x < \delta \to ((U(f,\Delta x) - I < \varepsilon) \wedge (I - L(f,\Delta x) < \varepsilon)))$$

The proof of this should follow easily by noting that this implies that I is the least upper bound of the lower sums and the greatest lower bound of the upper sums (which we defined to be that value of the integral). However, the formal proof is assumed to be familiar to the reader and it is thus omitted.

Theorem 6.12. *Let $f\colon [a,b] \to \mathbb{R}$ be an integrable function. Then the value of the integral, as defined in 6.8, is equal to the value of the integral as defined in 6.11.*

Proof. $\Rightarrow$ Let $I = \int_a^b f$ by means of 6.11. Fix $\varepsilon \in \mathbb{R}^+$; thus there is a $\delta \in \mathbb{R}^+$ such that by transfer we get that

$$(\forall dx \in {}^*\mathbb{R}^+)(dx <_{\mathfrak{U}} i(\delta) \to ({}^*U(f,dx) - i(I) <_{\mathfrak{U}} i(\varepsilon) \wedge i(I) - {}^*L(f,dx) <_{\mathfrak{U}} i(\varepsilon)))$$

This implies that for all positive infinitesimals ${}^*U(f,dx) \simeq i(I)$ and ${}^*L(f,dx) \simeq i(I)$. This implies that $sh({}^*U(f,dx)) = I$, so I is also the value of the integral by 6.8.

$\Leftarrow$ Let $\int_a^b f = sh({}^*U(f,dx))$ for some positive infinitesimal $dx \in {}^*\mathbb{R}^+$. Fix $\varepsilon \in \mathbb{R}^+$ and let $\delta \in {}^*\mathbb{R}^+$ be a positive infinitesimal. This implies that

$$(\forall dy \in {}^*\mathbb{R}^+)(dy <_{\mathfrak{U}} \delta \to ({}^*U(f,dy) - sh({}^*U(f,dx)) <_{\mathfrak{U}} i(\varepsilon) \wedge sh({}^*U(f,dx)) - {}^*L(f,dy) <_{\mathfrak{U}} i(\varepsilon))$$

since $dy < \delta$, it follows that dy is a positive infinitesimal, so $U(f,dy) \simeq {}^*U(f,dx) \simeq sh({}^*U(f,dx))$ and similarly for the lower sum. Thus,

$$(\exists \delta \in {}^*\mathbb{R}^+)(\forall dy \in {}^*\mathbb{R}^+)(dy <_{\mathfrak{U}} \delta \to ({}^*U(f,dy) - sh({}^*U(f,dx)) <_{\mathfrak{U}} i(\varepsilon) \wedge sh({}^*U(f,dx)) - {}^*L(f,dy) <_{\mathfrak{U}} i(\varepsilon))$$

and applying transfer we get that

$$(\exists \delta \in \mathbb{R}^+)(\forall \Delta y \in \mathbb{R}^+)(\Delta y < \delta \to (U(f,\Delta y) - sh({}^*U(f,dx)) < \varepsilon \wedge sh({}^*U(f,dx)) - L(f,\Delta y) < \varepsilon)$$

since this holds for any $\varepsilon \in \mathbb{R}^+$, we get that $\int_a^b f = sh({}^*U(f, dx))$ by 6.11, completing the proof. $\square$

This new definition, combined with the commonly known properties of partitions that are taught in a first-year calculus course, can be used to prove most of the commonly known theorems of calculus including both fundamental theorems of calculus.

7. Conclusion

Non-standard analysis provides a new take on the usual concepts of analysis. While we only explored the basics of single variable calculus, most of analysis can be developed in this matter; Robinson and Goldblatt provide a more thorough study of non-standard analysis, including more advanced concepts like measure theory and the Lebesgue integral, and are a must read if the reader wants to become more familiar with these concepts.

Acknowledgments. I would like to thank my mentor Jacob Perlman for his help and advice in the construction of this paper. I would also like to thank Professor Peter May for making the 2015 Math REU possible and all the lecturers, mentors and students involved for making the REU such an enriching experience.

References

[1] William Weiss and Cherie D'Mello, *Fundamentals of Model Theory*, University of Toronto, 1997, http://www.math.toronto.edu/weiss/model_theory.pdf.
[2] William DeMeo, *First-Order Logic (Predicate Calculus)*, https://williamdemeo.files.wordpress.com/2013/01/cs-firstorder.pdf.
[3] Alex Kruckman, *Notes on Ultrafilters*, Toolbox Seminar 11/7/12, https://math.berkeley.edu/ kruckman/ultrafilters.pdf.
[4] Klaus Kaiser, *Ultraproducts*, University of Houston, http://www.math.uh.edu/ klaus/ultraproducts.pdf.
[5] Sabine Koppelberg, *Logik und Mengenlehre (Chapter 2)*, Freien Universitt Berlin, http://www.mi.fu-berlin.de/math/groups/ag-logik/Lehre/UST-chapter02.pdf.
[6] Robert Goldblatt, *Lectures on the Hyperreals, An introduction to Nonstandard Analysis*, Springer, 1998.
[7] Robert A. Herrman, *Nonstandard Analysis - A Simplified Approach* -, http://arxiv.org/pdf/math/0310351.pdf, 2003.
[8] Abraham Robinson, *Non-standard Analysis*, Princeton University Press, 1974.