The Appeal of Nonstandard Analysis

Eli Baur

February 21, 2022

Abstract

Nonstandard Analysis emerged in the early 1960's as an attempt to make rigorous the notions of infinitesimal and infinite that pervade calculus. The construction of "Hyperreal" numbers provides a means to this end, with infinite (unlimited) and infinitesimal (infinitely small) numbers joining the familiar reals to form dense sub-orders around each real number so as to contain only one real value. The effectiveness of this treatment is captured by the transfer principle, a corollary of Łoś's Theorem from mathematical logic, which contends that sentences (for our purposes, theorems) which are true in the Hyperreals are true for the Reals as well. Thus, the rigor of nonstandard analysis serves to demonstrate the breadth of mathematical logic's applications. This paper discusses mathematical logic and its notions of what constitutes truth insofar as it is required to prove the transfer principle. We then give a construction of the Hyperreals and turn to some familiar and crucial theorems in calculus to prove them using the methods of nonstandard analysis.

Contents

1 Elementary Logic

We begin with an overview of some important concepts in the field of first order logic, defined inductively from a collection of symbols[1]. In short, our goal here is to give a more formal structure to what a mathematical statement *is*, per se. That is, when we say things like "the function f evaluated at x comes infinitely close to the value y as x comes infinitely close to a," we can also write "$(\forall\epsilon > 0)(\exists\delta > 0)\big((-\delta < x - a < \delta) \rightarrow (-\epsilon < f(x) - y < \epsilon)\big)$". First order logic considers this sentence to be a formal structure, built from more atomic structures like $<$,$-$, and y, which we recognize as relations, functions, and variables respectively. As in most mathematical fields, the logician has license to combine these ideas ad-infinitum to build increasingly complex structures; we will build towards the conception of a formal sentence as we have presented here.

The other concept we introduce here is some (extremely) elementary model theory. In the context of "the function f evaluated at x comes infinitely close to the value y as x comes infinitely close to a," the model theorist asks "how else could we 'encode' this information?" and, more importantly, "how does our choice of how to encode this information illuminate our understanding of it?" This consideration will motivate our construction of the hyperreals, which will be shown to be a model equivalent to the real numbers.

Definition 1.1. A **language** $\mathscr{L}$ is a collection of symbols, each belonging to one of three groups: **relations** P_i, **functions** F_i, and **constants** c_i.
For standard relations like order and common functions like addition and multiplication, we use the usual symbols of $\leq, +$, and $\cdot$ in place of some P_i and F_i.

Definition 1.2. A **model** of a language $\mathscr{L}$ is a tuple of a set of symbols A, called a **universe**, and a function $\mathscr{F}$ mapping $\mathscr{L}$ to A called an **interpretation**.
It is important to note that if a symbol a in $\mathscr{L}$ is a relation, we refer to the symbol sends a to, call it b, in A is treated as a relation. The same is true of functions and constants. We write this model as $\mathfrak{U} = \langle A, \mathscr{F}\rangle$, where A is the universe of the model, and $\mathscr{F}$ is its interpretation.

Remark 1.3. The following logical symbols are used in the formalizing of each language $\mathscr{L}$:
parentheses (,), variables v_i for $i \in \mathbb{N}$, connectives $\wedge$ (better known as "and") and $\neg$ (better known as "not"), and the identity relation $=$.

Definition 1.4. We use strings of the symbols in Definition 1.3 to construct **terms** of $\mathscr{L}$ inductively as follows:

i. Any variable is a term

ii. Any constant of $\mathscr{L}$ is a term

iii. For an n-arity function F and n terms $t_1, \cdots t_n$, $F(t_1, \cdots, t_n)$ is a term

iv. A string of symbols is a term if and only if it can be shown to be so through finite invocations of (i)-(iii).

Definition 1.5. We inductively define an **atomic formula** of $\mathscr{L}$ as a string of symbols so that

i. $t_1 = t_2$ when t_1 and t_2 are terms of $\mathscr{L}$ is an atomic formula of $\mathscr{L}$

ii. For an m-arity relation P and m terms $t_1 \cdots t_m$ of $\mathscr{L}$, $P(t_1, \cdots t_m)$ is an atomic formula of $\mathscr{L}$

Definition 1.6. We inductively define a **formula** of $\mathscr{L}$ as a string of symbols so that

i. An atomic formula is a formula

ii. If φ and ψ are formulas, then $\varphi \wedge \psi$ is a formula

iii. If v is a variable and φ is a formula, then $(\exists v)\varphi$ is a formula

[1]The material presented here is based on chapter 1.3 of Chang and Kiesler's Model Theory [1]

iv. A string of symbols is a formula if and only if it can be shown to be so through finite invocations of (i)-(iii).

Remark 1.7. We introduce some additional relation symbols as abbreviations for our established symbols. Specifically, for formulas φ, ψ,

$$\varphi \to \psi = \neg(\varphi \wedge (\neg\psi)), \quad \psi \leftrightarrow \psi = (\varphi \to \psi) \wedge (\psi \to \varphi),$$
$$(\forall v)\varphi = \neg((\exists v)(\neg\varphi)), \quad \varphi \vee \psi = \neg((\varphi \wedge \psi) \wedge (\neg\varphi \wedge \neg\psi))$$

Definition 1.8. A variable is **free** if the formula in which it occurs is not prefixed by a quantifier ($\forall$ or $\exists$) which binds that variable. A **sentence** is a formula with no free variables.

Remark 1.9. We denote a term t whose variables form a subset of $\{v_0, \cdots, v_n\}$ by $t(v_0, \cdots, v_n)$. Similarly, we denote a formula φ whose free variables form a subset of $\{v_0, \cdots, v_n\}$ by $\varphi(v_0, \cdots, v_n)$.

Definition 1.10. The **value** of a term $t(v_0, \cdots, v_n)$ for constants $\{x_0, \cdots, x_n\}$ is denoted by $t[x_0, \cdots, x_n]$ and is defined as follows:

i. For $t = v_i$, $t[x_0, \cdots, x_n] = x_i$

ii. If t is the constant c, then $t[x_0, \cdots, x_n] = b$ where b is the interpretation of c in $\mathfrak{U}$

iii. For $t = F(t_1, \cdots, t_m)$, where F is an m-arity function with interpretation G in $\mathfrak{U}$, $t[x_0, \cdots, x_n] = G(t_1[x_0, \cdots, x_n], \cdots, t_m[x_0, \cdots, x_n])$

Definition 1.11. We write $\mathfrak{U} \models \varphi$ to denote that a model $\mathfrak{U}$ **satisfies** a formula φ. Satisfaction plays a key role in devising truth in a model, and is defined as follows for atomic formulas:

i. $\mathfrak{U} \models (t_1 = t_2)[x_0, \cdots, x_n] \iff t_1[x_0, \cdots, x_n] = t_2[x_0, \cdots, x_n]$

ii. $\mathfrak{U} \models P(t_1, \cdots, t_m)[x_0, \cdots, x_n] \iff R(t_1[x_0, \cdots, x_n], \cdots, t_m[x_0, \cdots, x_n])$

For a formula $\varphi(v_1, \cdots, v_m)$ of $\mathscr{L}$, satisfaction is defined as follows:

i. When $\varphi = \psi_1 \wedge \psi_2$, $\mathfrak{U} \models \varphi[x_0, \cdots, x_n] \iff \mathfrak{U} \models \psi_1[x_0, \cdots, x_n]$ and $\mathfrak{U} \models \psi_2[x_0, \cdots, x_n]$

ii. When $\varphi = \neg\psi$, then $\mathfrak{U} \models \varphi[x_0, \cdots, x_n] \iff \mathfrak{U} \not\models \psi[x_0, \cdots, x_n]$

iii. When $\varphi = (\exists v_i)\psi$, $\mathfrak{U} \models \varphi[x_0, \cdots, x_n] \iff$ there is some $y \in A$ s.t., $\mathfrak{U} \models \psi[x_0, \cdots, x_{i-1}, y, x_{i+1}, \cdots, x_n]$

Definition 1.12. A sentence σ is **true** in a model $\mathfrak{U}$ if, and only if, $\mathfrak{U} \models \sigma[\mathbf{x_1}, \cdots, \mathbf{x_n}]$ for some $\{x_1, \cdots, x_n\} \subset A$

Definition 1.13. Two models $\mathfrak{U}$ and $\mathfrak{B}$ are **elementary equivalent** if every sentence that is true in $\mathfrak{U}$ is true in $\mathfrak{B}$. We denote this with $\mathfrak{U} \equiv \mathfrak{B}$.

To sum all of this up, a statement like "For any $x \in \mathbb{Z}$, there is a $y \in \mathbb{Z}$ such that $y < x$" is a sentence which makes an "existentially quantified" claim about the relation $<$ between two symbols x and y. This claim is true because of how the relation $<$ is defined in the model $\mathbb{Z}$, and that sense of truth is captured by our definition of satisfaction in the model. Not all sentences are true in all models (the sentence we just described is false in $\mathbb{N}$ for example), but certain models exhibit the ability to satisfy sentences whenever they are satisfied by another model, and this characteristic of elementary equivalence is what we are after between the $\mathbb{R}$ and the hyperreals.

2 Filters

In order to construct the hyperreals and use them to prove things about $\mathbb{R}$, we will first want to make rigorous the idea that something is true "almost everywhere" in something infinite, like a sequence. We will do this with filters and U-equivalence, as we will see that each ultrafilter over the naturals will contain every cofinite set, along with some other "large" sets [2].

[2]The material in this section is based on chapter 4.1 of Chang and Kiesler's "*M*odel Theory" [1]

Definition 2.1. Given a nonempty set I, a **filter** F is any subset of the power set of I (henceforth denoted by $P(I)$) with the following properties:

i. $I \in F$

ii. For any $X, Y \in F$, $X \cap Y \in F$

iii. For any $X \in F$ and $Z \in P(I)$, if $X \subset Z \subset I$, then $Z \in F$

A filter over I is **proper** if $F \neq P(I)$

Lemma 2.2. *A filter F is not proper $\iff$ $\emptyset \in F$*

Proof.
$\Rightarrow$: $F = P(I)$, so $\emptyset \in F$
$\Leftarrow$: For $X \in P(I)$, as F is a filter, $\emptyset \subset X$, by the third property of filters $X \in F$. $\square$

Definition 2.3. A proper filter F is **principal** if there is some finite $E \in P(I)$ such that

$$F = \{Y \in P(I) : E \subseteq Y\}$$

A filter is **non-principal** if it is not principal.

Below are some examples of filters, both principal and non-principal: Let $E \subset \mathbb{N}$, then $F_E = \{X \in P(\mathbb{N}) : E \subset X\}$ is a principal filter, as $E \in X \cap Y \in F_E$ for $X, Y \in P(\mathbb{N})$ and $E \subset X \subset Z \implies E \subset Z \implies Z \in F_E$ and $E \subset \mathbb{N} \in F_E$.

Let $F_{cof} = \{X \in P(\mathbb{N}) : {}^cX \text{is finite}\}$. This is a filter as $\mathbb{N}$ has finite complement $\emptyset$, ${}^cX \cap Y = {}^cX \cup {}^cY$ is finite, and ${}^cX \supset {}^cZ$ when $X \subset Z$, so the relevant sets $X \cap Y$ and Z are in F_{cof}.

Definition 2.4. A filter F is called an **ultrafilter** if

$$X \in F \iff I \setminus X \notin F$$

Lemma 2.5. *F is an ultrafilter which contains a finite set E $\iff$ F is a principal filter generated by a singleton set $\{x\} \subseteq E$*

Proof.
$\Rightarrow$: Suppose for contradiction that for every $x_j \in E = \{x_1, \cdots x_n\}$, $\{x_j\} \notin F$. By the ultrafilter property, $I \setminus \{x_j\} \in F$ for all x_j, so $E \cap \left(\bigcap_{j=1}^{n} I \setminus \{x_j\}\right) = E \cap \left(I \setminus E\right) = \emptyset \in F$. By Lemma 2.2 then, F is not proper, and this is a contradiction.
Thus we know that for some x_j, $\{x_j\} \in F$, and as F is proper (since $P(I)$ does not satisfy the ultrafilter property, so $F \neq P(I)$), we have $\{x_j\} \cap Y \neq \emptyset$ for all $Y \in F$. Thus, $\{x_j\} \subset Y$ for all $Y \in F$, so F is principal.

$\Leftarrow$: For any $Y \in P(I)$, $Y \in F \iff \{x\} \subseteq Y \iff \{x\} \not\subseteq {}^cY \iff {}^cY \notin F$, so $Y \in F \iff I \setminus Y \notin F$, hence F is an ultrafilter, and as $\{x\} \subseteq E$, $E \in F$. $\square$

Lemma 2.6. *(**The Ultrafilter Lemma**) A filter F is an ultrafilter $\iff$ F is a maximal proper filter*

Proof.
$\Rightarrow$: Suppose there is some proper filter F' with $F \subset F'$ but $F' \neq F$. It must be the case then that $\exists\, x \in F'$ with $x \notin F$, but as F is an ultrafilter, ${}^cx \in F \subset F'$. As F' is a filter, we have then that $\emptyset = x \cap {}^cx \in F'$, so F' is not a proper filter by Lemma 2.2, and this is a contradiction.

$\Leftarrow$: Let $X \in P(I)$.
If $X \in F$, then it cannot be the case that ${}^cX \in F$ as then $X \cap {}^cX = \emptyset \in F$ so F would not be proper.

If $X \notin F$ and $X \neq \emptyset$ (as $I \setminus \emptyset = I \in F \implies \emptyset \notin F$ by the case above), then as the filter $F' = F \cup \{X\}$ [3] exhibits $F \subset F'$, it must be that F' is not proper since F is maximal. By Lemma 2.2 then,

$$\emptyset \in F' \setminus F \subset \{Y \cap X' \in P(I) : Y \in F, X \subset X' \subset I\}$$

Since $\emptyset \notin F$ and $\emptyset \neq X$, there is some set $X' \in P(I)$ with $X \subset X'$ such that $X' = {}^cY$ for some $Y \in F$. As $X \subset X'$, $Y \subseteq I \setminus X$, so since F is a filter $I \setminus X \in F$. □

Definition 2.7. A set $S \subset P(I)$ is said to have **the finite intersection property** if for any finite intersection of elements of S is itself an element of S, i.e.

$$\{s_1, \cdots, s_n\} \subset S \implies \bigcap_{i=1}^{n} s_i \neq \emptyset$$

Proposition 2.8. ***(Existence of Ultrafilters)*** *For a set $S \subset P(I)$ with the finite intersection property, there is an ultrafiler $M \supset S$*

Proof.
This follows from Zorn's lemma.

Consider the set W of all proper filters F on I which exhibit $S \subset F$.
We note that W is nonempty since we can construct

$$F_0 = \{X \cap Y : X, Y \subset I, \exists\ A_1, \cdots, A_n, B_1, \cdots, B_n \in S \text{such that} \bigcap_{i=1}^{n} A_i \subset X, \bigcap_{i=1}^{n} B_i \subset Y\}$$

and F_0 is a proper filter containing S as:

i. $I = I \cap I$ and for any $A_1, B_1 \in S$, $A_1, B_1 \subset I$

ii. If $P, Q \in F_0$, $P = W \cap X$ and $Q = Y \cap Z$ with $\bigcap_{i=1}^{n} A_i \subset W$, $\bigcap_{i=1}^{n} B_i \subset X$, $\bigcap_{i=1}^{n} C_i \subset Y$, and $\bigcap_{i=1}^{n} D_i \subset Z$ (we can make all the n values the same by adding random sets in S to any interesection with less that n the maximum number of sets being interesected as the intersection of the larger collection of said sets will still be a subset). Then $(\bigcap_{i=1}^{n} A_i) \cap (\bigcap_{i=1}^{n} B_i) \subset W \cap X$ and $(\bigcap_{i=1}^{n} C_i) \cap (\bigcap_{i=1}^{n} D_i) \subset Y \cap Z$, so $(W \cap X) \cap (Y \cap Z) \in F_0$

iii. If $P \subset R$ and $P \in F_0$, then the relevant X and Y which exhibit $\bigcap_{i=1}^{n} A_i \subset X$ and $\bigcap_{i=1}^{n} B_i \subset Y$ will yield $\bigcap_{i=1}^{n} A_i \subset X \cup R$ and $\bigcap_{i=1}^{n} B_i \subset Y \cup R$, so $(X \cup R) \cap (Y \cup R) = (X \cap Y) \cup R = R$ as $X \cap Y = P \subset R$. Thus, $R \in F_0$.

iv. $\emptyset \notin F_0$ as it is not possible for $X \cap Y = \emptyset$ when $\bigcap_{i=1}^{n} A_i \subset X$ and $\bigcap_{i=1}^{n} B_i \subset Y$ as then $\emptyset \neq (\bigcap_{i=1}^{n} A_i) \cap (\bigcap_{i=1}^{n} B_i) \subset X \cap Y$ as S has the finite intersection property and $A_i, B_i \in S$ for all i.

v. For any $s \in S$, $s \neq \emptyset$ as S has the finite intersection property, so $S = S \cap S$ with $A_1 = S = B_1$ yielding $A_1 \subset S$ and $B_1 \subset S$, so $S \cap S = S \in F_0$.

Our set W is partially ordered by subset inclusion, since $X \subset X$, $X \subset Y$ and $Y \subset X \implies X = Y$, and $X \subset Y \subset Z \implies X \subset Z$ for all $X, Y, Z \in P\big(P(I)\big)$, and $W \subset P\big(P(I)\big)$.

For any chain $C \subset W$, the union of all filters in the chain, $\bigcup C = \bigcup_{F \in C} F$, is a proper filter since

i. For any $f \in C$, $I \in f$, so $I \in \bigcup C$

ii. If $A, B \in \bigcup C$, there are filters $F_A, F_B \in C$ with $A \in F_A$ and $B \in F_B$. As C is a chain and hence totally ordered, without loss of generality we can say $F_A \subseteq F_B$, so $A \in F_B$, hence as F_B is a filter, $A \cap B \in F_B \subseteq \bigcup C$.

iii. If $A \in \bigcup C$ and $A \subset B \in P(I)$, then as there is some $F_A \ni A$, $B \in F_A \subset \bigcup C$.

[3] that is, the filter created by closing the set $F \cup \{X\}$ under finite intersections and including any supersets of those sets.

iv. For every filter $F \in C$, F is proper by our construction of W, so $\emptyset \notin F$, hence $\emptyset \notin \bigcup_{F \in C} F = \bigcup C$.

So as every $F \in C$ exhibits $F \subseteq \bigcup C \in W$ by our construction, the chain C is bounded above.

By Zorn's Lemma then, there is a proper filter $M \in W$ which is maximal in W.
If a filter F on I exhibits $M \subsetneq F$, then $E \subset F$, so $F \in W$, and hence as M is maximal in W, $F = M$. We can see then that M is maximal in the set of ultrafilters over I, so by the Ultrafilter Lemma, M is an ultrafilter on I. □

Corollary 2.9. *Every proper filter F is contained in an ultrafilter*

Proof.
Every proper filter has the finite intersection property since $\emptyset \notin F$, so $\bigcap_{i=1}^{n} x_i \neq \emptyset$ for any $\{x_1, \cdots, x_n\} \subset F$ as $\bigcap_{i=1}^{n} x_i \in F$. □

Definition 2.10. Given a nonempty I and sets A_i for all $i \in I$, we can construct the Cartesian product of all such A_i,

$$\prod_{i \in I} A_i = \{(\cdots, x_i, \cdots) : x_i \in A_i\}$$

Given two elements $f, g \in \prod_{i \in I} A_i$, we write that the element at the index i of f and g as $f(i)$ and $g(i)$ respectively. For a filter U over I, We say that f and g are **U-equivalent** if

$$\{i \in I : f(i) = g(i)\} \in U$$

and we denote this by $f =_U g$

Remark 2.11. If we let $I = \mathbb{N}$, we can also think of $\prod_{i \in I} A_i$ as the set of all sequences S where $S_n \in A_n$ for each natural n. This characterization is closer to how we will construct the hyperreals in a moment, but not necessarily accurate for any given product as in theory we could have $I = \mathbb{Z}$ or $I = \mathbb{R}$.

Lemma 2.12. $=_U$ forms an equivalence relation

Proof.

i. Reflexive: $\{i \in I : f(i) = f(i)\} = I \in U$

ii. Symmetric: As equality is symmetric, $\{i \in I : f(i) = g(i)\} = \{i \in I : g(i) = f(i)\} \in U$

iii. Transitive: If $\{i \in I : f(i) = g(i)\} \in U$ and $\{i \in I : g(i) = h(i)\} \in U$, then
$\{i \in I : f(i) = g(i)\} \cap \{i \in I : g(i) = h(i)\} = \{i \in I : f(i) = g(i) = h(i)\} \in U$, so
$\{i \in I : f(i) = g(i) = h(i)\} \subseteq \{i \in I : f(i) = h(i)\} \in U$

□

Definition 2.13. The **equivalence classes** of $\prod_{i \in I} A_i$ are denoted by $[f]$ for some $f \in \prod_{i \in I} A_i$, so

$$[f] = \left\{ g \in \prod_{i \in I} A_i : f =_U g \right\}$$

Definition 2.14. Given a nonempty set I, a proper filter U over I, and sets A_i for all $i \in I$, we can construct a **reduced product** of A_i's modulo U as the set of all equivalence classes of the Cartesian product of our A_i's, so

$$\prod_U A_i = \left\{ [f] : f \in \prod_{i \in I} A_i \right\}$$

If U is an ultrafilter, we call this an **ultraproduct** of A_i modulo U.
If an ultraproduct has $A_i = A$ for all $i \in I$, it is an **ultrapower** of A modulo U, denoted by $\prod_U U$.

Remark 2.15. We now come to the theorem which motivates our entire enquiry, Łoś's Theorem. This is an exceedingly technical result arising from our investigation of truth in models, and it says that a sentence which is true in "almost every model" (specifically, in a collection of models for which the indices form a set in an ultrafilter) for an infinite collection of models is also true in the ultraproduct of those models (constructed using the same ultraproduct as in determining what is true "almost everywhere". Perhaps the more surprising fact of Łoś's Theorem is that this relationship is *b*idirectional, so our understanding of the collection of models as captured in the ultraproduct construction reveals information about the majority of models themselves. The corollary to this, the Transfer Principle, will be essential in our investiagtion of the hyperreal numbers.

The intuition behind Łoś's Theorem is that we can leverage our ultrafilter in interpreting statements in the ultraproduct to necessitate that their truth implies truth in a majority of the models involved in the ultraproduct's construction. Below, we describe this interpretation:

For a nonempty set I and a proper filter U over I, consider an ultraproduct of models of $\mathscr{L}$, $\prod_U \mathfrak{U}_i$. For an arbitrary relation P, function F, and constant c, each $\mathfrak{U}_i = \langle A_i, \mathscr{F} \rangle$ interprets them as R_i, G_i, and b_i respectively. This ultraproduct models $\mathscr{L}$ as follows:

i. The universe of $\prod_U \mathfrak{U}_i$ is $\prod_U A_i$

ii. An n-arity relation P is interpreted as the relation S where

$$S([f_1], \cdots [f_n]) \iff \{i \in I : R_i\big(f_1(i), \cdots, f_n(i)\big)\} \in U$$

iii. An m-arity function F is interpreted as the function H given by

$$H([f_1], \cdots, [f_m]) = \Big[\prod_{i \in I} G_i\big(f_1(i), \cdots, f_m(i)\big)\Big]$$

iv. A constant c is interpreted as

$$b = \Big[\prod_{i \in I} b_i\Big]$$

Theorem 2.16. ***(*Łoś's Theorem*)*** *Given a nonempty set I, an ultrafilter U over I, and models $\mathfrak{U}_i = \langle A_i, \mathscr{F}_i \rangle$ of $\mathscr{L}$, we can construct the ultraproduct $\mathfrak{B} = \langle A, \mathscr{F} \rangle = \prod_U \mathfrak{U}$, where $A = \prod_U A_i$ and $\mathscr{F}$ is the interpretation described above so that:*

i. For any term $t(x_1, \cdots, x_n)$ of $\mathscr{L}$ and elements $\{[f_1], \cdots, [f_n]\} \subset V$, we have

$$t_{\mathfrak{B}}\big[[f_1] \cdots [f_n]\big] = \Big[\prod_{i \in I} t_{\mathfrak{U}_i}[f_1(i), \cdots, f_n(i)]\Big]$$

ii. For any formula $\varphi(x_1, \cdots, x_n)$ of $\mathscr{L}$ and $[f_1], \cdots, [f_n] \in V$, we have

$$\mathfrak{B} \models \varphi\big[[f_1], \cdots, [f_n]\big] \iff \{i \in I : \mathfrak{U}_i \models \varphi[f_1(i) \cdots f_n(i)]\} \in U$$

iii. For any sentence φ of $\mathscr{L}$,

$$\mathfrak{B} \models \varphi \iff \{i \in I : \mathfrak{U}_i \models \phi\} \in U$$

Note that the subscript on a term indicates which model we are interpreting it in.

Proof. For each relation P of $\mathscr{L}$, we denote the interpretation of P in $\mathfrak{U}_i$ by R_i. A similar convention carries F to G_i, and c to b_i for each $\mathfrak{U}_i$.

i. Recall that a term has three possible forms:

(a) t is a variable v_j: By Definition 1.10, $t_{\mathfrak{U}_i}[f_1(i)\cdots f_n(i)] = f_j(i)$. Hence, by Remark 2.15

$$t_{\mathfrak{B}}\big[[f_1]\cdots[f_n]\big] = [f_j] = \Big[\prod_{i\in I} f_j(i)\Big] = \Big[\prod_{i\in I} t_{\mathfrak{U}_i}[f_1(i)\cdots f_n(i)]\Big]$$

(b) t is a constant c: For any $i \in [1,n]$, $t_{\mathfrak{U}_i}[f_1(i)\cdots f_n(i)] = b_i$. So, by Remark 2.15(4),

$$t_{\mathfrak{B}}\big[[f_1]\cdots[f_n]\big] = b = \Big[\prod_{i\in I} b_i\Big] = \Big[\prod_{i\in I} t_{\mathfrak{U}_i}[f_1(i)\cdots f_n(i)]\Big]$$

(c) t is a function F of terms $t_1,\cdots,t_m$: Suppose that (1) holds for each t_j, then by def. 1.10

$$t_{j,\mathfrak{B}}\big[[f_1],\cdots,[f_n]\big] = \Big[\prod_{i\in I} t_{j,\mathfrak{U}_i}[f_1(i),\cdots,f_n(i)]\Big] \qquad (a)$$

$$t_{\mathfrak{U}_i}\big[f_1(i)\cdots f_n(i)\big] = G_i\Big(t_{1,\mathfrak{U}_i}\big[f_1(i),\cdots,f_n(i)\big],\cdots,t_{m,\mathfrak{U}_i}\big[f_1(i),\cdots,f_n(i)\big]\Big) \qquad (b)$$

Hence,

$$\begin{aligned} t_{\mathfrak{B}}\big[[f_1]\cdots[f_n]\big] &= H(t_{1,\mathfrak{B}}\big[[f_1],\cdots,[f_n]\big],\cdots,t_{m,\mathfrak{B}}\big[[f_1],\cdots,[f_n]\big]) && \text{by 1.10}\\ &= H(\Big[\prod_{i\in I} t_{1,\mathfrak{U}_i}[f_1(i),\cdots,f_n(i)]\Big],\cdots,\Big[\prod_{i\in I} t_{m,\mathfrak{U}_i}[f_1(i),\cdots,f_n(i)]\Big]) && \text{by (a)}\\ &= \Big[\prod_{i\in I} G_i\big(t_{1,\mathfrak{U}_i}[f_1(i),\cdots,f_n(i)],\cdots,t_{m,\mathfrak{U}_i}[f_1(i),\cdots,f_n(i)]\big)\Big] && \text{by 2.15}\\ &= \Big[\prod_{i\in I} t_{\mathfrak{U}_i}\big[f_1(i)\cdots f_n(i)\big]\Big] && \text{by (b)} \end{aligned}$$

By induction, every term exhibits the equality in (1)

ii. Recall that a formula has four possible forms

(a) φ is an atomic formula, and is an m-arity relation P of term $t_1,\cdots,t_m$:

$$\begin{aligned} \mathfrak{B} \models \varphi\big[[f_1],\cdots,[f_n]\big] &\iff R\Big(t_{1,\mathfrak{B}}\big[[f_1],\cdots,[f_n]\big],\cdots,t_{m,\mathfrak{B}}\big[[f_1],\cdots,[f_n]\big]\Big) && \text{by 1.11}\\ &\iff \{i\in I : R_i\big(f_1(i),\cdots,f_n(i)\big)\} \in U && \text{by 2.15}\\ &\iff \{i\in I : \mathfrak{U}_i \models \varphi\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{by 1.11} \end{aligned}$$

Note that this holds for the two-arity relation $=$.

(b) $\varphi = \psi \wedge \theta$ where ψ and θ satisfy (2):

$$\begin{aligned} &\mathfrak{B} \models \varphi\big[[f_1],\cdots,[f_n]\big]\\ &\iff \mathfrak{B} \models \psi\big[[f_1],\cdots,[f_n]\big] \text{ and } \mathfrak{B} \models \theta\big[[f_1],\cdots,[f_n]\big] && \text{by 1.11}\\ &\iff \{i\in I : \mathfrak{U}_i \models \psi\big[f_1(i),\cdots,f_n(i)\big]\} \in U, \{i\in I : \mathfrak{U}_i \models \theta\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{by (2)}\\ &\iff \{i\in I : \mathfrak{U}_i \models \psi\big[f_1(i),\cdots,f_n(i)\big]\} \cap \{i\in I : \mathfrak{U}_i \models \theta\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{by 2.1}\\ &\iff \{i\in I : \mathfrak{U}_i \models \psi\big[f_1(i),\cdots,f_n(i)\big] \text{ and } \mathfrak{U}_i \models \theta\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{see [4]}\\ &\iff \{i\in I : \mathfrak{U}_i \models \varphi\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{by 1.11} \end{aligned}$$

(c) $\varphi = \neg\psi$ for ψ which satisfies (2)

$$\begin{aligned} &\mathfrak{B} \models \varphi\big[[f_1],\cdots,[f_n]\big]\\ &\iff \mathfrak{B} \not\models \psi\big[[f_1],\cdots,[f_n]\big] && \text{by 1.11}\\ &\iff \{i\in I : \mathfrak{U}_i \models \psi\big[f_1(i),\cdots,f_n(i)\big]\} \notin U && \text{by (2)}\\ &\iff \{i\in I : \mathfrak{U}_i \not\models \psi\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{as } U \text{ is an ultrafilter}\\ &\iff \{i\in I : \mathfrak{U}_i \models \varphi\big[f_1(i),\cdots,f_n(i)\big]\} \in U && \text{by 1.11} \end{aligned}$$

[4] The equivalence here is given by the fact that the sets in question are equal

(d) $\varphi = (\exists v_1)\psi$ for some ψ which satisfies (2):

$$\begin{aligned}
&\mathfrak{B} \models \varphi\big[[f_1], \cdots, [f_n]\big] \\
&\iff \text{ there is some } [f] \in A \text{ s.t. } \mathfrak{B} \models \psi\big[[f], [f_2] \cdots, [f_n]\big] && \text{by 1.11 on } \varphi \\
&\iff \exists\ [f] \in A \text{ s.t. } \{i \in I : \mathfrak{U}_i \models \psi\big[f(i), f_2(i) \cdots, f_n(i)\big]\} \in U && \text{by (2) on } \psi \\
&\iff \{i \in I : \exists\ x \in A_i \text{ s.t. } \mathfrak{U}_i \models \psi\big[x, f_2(i), \cdots, f_n(i)\big]\} \in U && \text{for } x = f(i) \\
&\iff \{i \in I : \mathfrak{U}_i \models \varphi\big[f_1(i), f_2(i), \cdots, f_n(i)\big]\} \in U && \text{by 1.11}
\end{aligned}$$

By induction, (2) holds for any formula.

iii. This follows directly from (2) and the definition of truth in the model, Definition 1.12.

□

Corollary 2.17. *(The Transfer Principle) Any model $\mathfrak{U}$ is elementary equivalent to its ultrapower $\prod_U \mathfrak{U}$.*

$$\prod_U \mathfrak{U} \equiv \mathfrak{U}$$

Proof. If a formula φ exhibits $\mathfrak{U} \models \varphi$, then at every index i of I, $\mathfrak{U}_i = \mathfrak{U}$ and hence $\{i \in I : \mathfrak{U}_i \models \varphi\} = I \in U$, so by Łoś's Theorem, $\prod_U \mathfrak{U} \models \varphi$.
Conversely, by Łoś's Theorem if $\prod_U \mathfrak{U} \models \varphi$ then $\{i \in I : \mathfrak{U}_i \models \varphi\} \in U$. As D is proper, $\{i \in I : \mathfrak{U}_i \models \varphi\} \neq \emptyset$, so for some $i \in I$, $\mathfrak{U}_i \models \varphi$, and as every $\mathfrak{U}_i = \mathfrak{U}$, we have $\mathfrak{U} \models \varphi$. □

Remark 2.18. What the transfer principle is formalizing is the idea that every sentence which is true in the ultrapower of a model (such as the model $\mathbb{R}$) is true in that model itself, and vice versa. The process of transferring a sentence preserves variables and relations, but changes the universe of the sentence, hence

$$(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})(x < y)$$

becomes

$$(\forall x \in \prod_U \mathbb{R})(\exists y \in \prod_U \mathbb{R})(x < y)$$

where $<$ represents the same relation of $\mathscr{L}$ in each case. The reverse is also possible by transfer. Not every sentence which we will transfer is written in its most fundamental expression, using only the symbols outlined in Definition 1.6 and Remark 1.7, but the reader is encouraged to consider how each symbol might be expressed using the fundamental ones outlined.

3 Hyperreals

We turn now to the structure of the constants, relations, and functions of the hyperreals.[5]

Definition 3.1. We define the hyperreals to be the set of all equivalence classes of real-valued sequences for any nonprincipal ultrafilter U on $\mathbb{N}$. [6] We represent this symbolically with $^*\mathbb{R}$, so

$$^*\mathbb{R} = \{[r] : r \in \mathbb{R}^{\mathbb{N}}\}$$

where $\mathbb{R}^{\mathbb{N}} = \{\langle r_1, r_2, r_3, \cdots\rangle : r_i \in \mathbb{R}\}$ is the set of all real valued sequences. Notice that $^*\mathbb{R}$ as we have defined it here is an ultrapower of $\mathbb{R}$ using $I = \mathbb{N}$, so the transfer principle is applicable to any sentence of $^*\mathbb{R}$ or $\mathbb{R}$ respectively.

[5] The material presented in the following sections is based on Goldblatt's " *Lectures on the Hyperreals"* [2]

[6] The choice of ultrafilter is arbitrary under the continuum hypothesis; for a more in depth explanation, see page 33 of Goldblatt's "*L*ectures on the Hyperreals" [2].

Definition 3.2. Given $\mathbf{r}, \mathbf{s} \in \mathbb{R}^{\mathbb{N}}$, we define $\oplus$ and $\odot$ as addition and multiplication along the indices of each sequence, so

$$\mathbf{r} \oplus \mathbf{s} = \langle r_n : n \in \mathbb{N}\rangle \oplus \langle s_n : n \in \mathbb{N}\rangle = \langle r_n + s_n : n \in \mathbb{N}\rangle$$

and

$$\mathbf{r} \odot \mathbf{s} = \langle r_n : n \in \mathbb{N}\rangle \odot \langle s_n : n \in \mathbb{N}\rangle = \langle r_n \cdot s_n : n \in \mathbb{N}\rangle$$

Definition 3.3. For $[\mathbf{r}], [\mathbf{s}] \in^* \mathbb{R}$, we define the following operations and relations:

i. $[\mathbf{r}] + [\mathbf{s}] = [\mathbf{r} \oplus \mathbf{s}]$

ii. $[\mathbf{r}] \cdot [\mathbf{s}] = [\mathbf{r} \odot \mathbf{s}]$

iii. $[\mathbf{r}] = [\mathbf{s}] \iff \mathbf{r} =_U \mathbf{s}$

iv. $[\mathbf{r}] < [\mathbf{s}] \iff \{n \in \mathbb{N} : r_n < s_n\} \in U$

v. $[\mathbf{r}] \leq [\mathbf{s}] \iff \{n \in \mathbb{N} : r_n \leq s_n\} \in U$

Remark 3.4. For each $x \in \mathbb{R}$, we identify the corresponding equivalence class for the constant sequence $(x, x, x, \cdots)$. So the map $^* : \mathbb{R} \to {}^*\mathbb{R}$ given by $^*(r) = [r]$ is the identity function and is injective. Notice that $^*(\mathbb{R}) \subset {}^*\mathbb{R}$.
As a convention, when we wish to denote an equivalence class of a constant sequence, we will first specify an $x \in \mathbb{R}$ and then write $\mathbf{x}$ for said equivalence class. If we wish to take the equivalence class of a non constant sequence, we will usually write $\mathbf{r} \in \mathbb{R}^{\mathbb{N}}$, and r_n for its nth term. So, for example, $[\mathbf{0}] = [(0, 0, 0, \cdots)]$ and $\mathbf{r} = [(r_1, r_2, r_3, \cdots)]$.
Unbolded letters are reserved for elements of $^*\mathbb{R}$ or $\mathbb{R}$ in general (we will usually specify).

Theorem 3.5. *The function* * :$\mathbb{R} \to {}^*\mathbb{R}$ *described in Remark 3.3 is an order preserving field isomorphism.*

Proof. Let $r, s \in \mathbb{R}$, $^*(r + s) = [\mathbf{r} + \mathbf{s}] = [\mathbf{r} \oplus \mathbf{s}] = [r] + [s]$, and $^*(r \cdot s) = \big[\mathbf{r} \cdot \mathbf{s}\big] = \big[\mathbf{r} \odot \mathbf{s}\big] = [\mathbf{r}] \cdot [\mathbf{s}]$ by 3.3 and 3.4. By 3.4, $[\mathbf{r}] < [\mathbf{s}] \implies \{n \in \mathbb{N} : r_n < s_n\} \in U \implies \exists\ n \in \mathbb{N}$ s.t. $r_n < s_n$ since U is proper. As $r_n = r$ and $s_n = s$, we have $[\mathbf{r}] < [\mathbf{s}] \implies r < s$. The reverse of this is given by the fact that $\{i \in I : r < s\} = \mathbb{N} \in U$, hence $[\mathbf{r}] < [\mathbf{s}] \iff r < s$. The same process yields $[\mathbf{r}] = [\mathbf{s}] \iff r = s$ $\square$

Remark 3.6. In light of Theorem 3.5, we allow comparison between a real value r and a hyperreal value h by adopting $r < h$ to mean $[\mathbf{r}] < h$

Theorem 3.5 also ensures that any real constant is preserved in the transfer of a sentence via the transfer principle.

Theorem 3.7. *The structure* $\langle {}^*\mathbb{R}, +, \cdot, < \rangle$ *is an ordered field with zero* $[\mathbf{0}]$ *and unity* $[\mathbf{1}]$.

Proof. For any $\mathbf{r} \in \mathbb{R}^{\mathbb{N}}$,

$$[\mathbf{r}] + [\mathbf{0}] = [\mathbf{r} \oplus \mathbf{0}] = [\langle r_n + 0 : n \in \mathbb{N}\rangle] = [\mathbf{r}]$$

$$[\mathbf{r}] \cdot [\mathbf{1}] = [\mathbf{r} \odot \mathbf{1}] = [\langle r_n \cdot 1 : n \in \mathbb{N}\rangle] = [\mathbf{r}]$$

We can quickly construct an additive inverse $-[\mathbf{r}]$ as $[-\mathbf{r}]$, as

$$[\mathbf{r}] + [-\mathbf{r}] = [\mathbf{r} \oplus -\mathbf{r}] = [\langle r_n - r_n : n \in \mathbb{N}\rangle] = [\mathbf{0}]$$

For $[\mathbf{r}] \neq [\mathbf{0}]$, $\{n \in \mathbb{N} : r_n = 0\} \notin U$, so as U is an ultrafilter we have $\{n \in \mathbb{N} : r_n \neq 0\} \in U$. Consider the sequence $\mathbf{s}$ given by

$$s_n = \begin{cases} \frac{1}{r_n} & r_n \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

So as the set of indices where $\mathbf{r} \odot \mathbf{s}$ equals 1 is $\{n \in \mathbb{N} : r_n \neq 0\} \in U$, $\mathbf{r} \odot \mathbf{s} =_U \mathbf{1}$, so $[\mathbf{r}] \cdot [\mathbf{s}] = [\mathbf{1}]$, thus the multiplicative inverse of $\mathbf{r}$ is $\mathbf{s}$.
It follows directly from the fact that $\mathbb{R}$ is closed and our definition of $+$ and $*$ on $^*\mathbb{R}$ that $^*\mathbb{R}$ is closed under addition and multiplication, and the remaining axioms follow directly from the associativity, commutivity, and distributivity of $+$ and $\cdot$ on $\mathbb{R}$; formal proofs of these are omitted. $\square$

Definition 3.8.

i. A hyperreal ϵ is **infinitesimal** if for all positive $x \in \mathbb{R}$, we have $-x < \epsilon < x$.

ii. An hyperreal ω is **unlimited** if for all $x \in \mathbb{R}$, we have $x < \omega$ [7], or if for all $x \in \mathbb{R}$, we have $\omega < x$ [8].

iii. A hyperreal α is **limited** if it is not unlimited, that is: $\exists x, y \in \mathbb{R}$ s.t. $x < \alpha < y$

iv. An hyperreal β is **appreciable** if it is limited but not infinitesimal, so $\exists x \in \mathbb{R}$ s.t. $0 < x < |\beta|$.

Example 3.9. A common example of an unlimited hyperreal is $[\omega] = [(1, 2, 3, 4, 5, \cdots)]$, since any real number r is bounded above by some natural number n and hence there are finitely many indices where it can be the case that $\omega_i < r$. A classic example of an infinitesimal is $[(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots)]$ for reasoning similar to that outlined above.

Definition 3.10. Two hyperreals $a, b \in {}^*\mathbb{R}$ are **infinitely close** if $a - b$ is infinitesimal. We denote this with $\simeq$.
The **halo** of a hyperreal $x \in {}^*\mathbb{R}$ is the set of all hyperreals infinitely close to x. We denote this with

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

Two hyperreals $a, b \in {}^*\mathbb{R}$ are a **limited distance apart** if $a - b$ is limited. We denote this with $\sim$.
The **galaxy** of a hyperreal $x \in {}^*\mathbb{R}$ is the set of all hyperreals which are a limited distance apart from x. We denote this with

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

Proposition 3.11. *$\simeq$ forms an equivalence relation*

Proof.

i. Reflexive: As 0 is infinitesimal, $x \simeq x$

ii. symmetric: As $|x - y| = |y - x| < \epsilon$, $a \simeq b \implies b \simeq a$

iii. Transitive: For any positive $\epsilon \in \mathbb{R}$, $x \simeq y$ and $y \simeq z \implies |x - y| + |y - z| < \frac{\epsilon}{2} + \frac{\epsilon}{2} \implies |x - z| < \epsilon$, so $x \simeq z$

□

Theorem 3.12. *Every limited hyperreal x is within the halo of exactly one real number*

Proof. Consider the set $A = \{y \in \mathbb{R} : y < x\}$. As x is limited, there is some $r \in \mathbb{R}$ s.t. $x < r$, so A is bounded above. As $\mathbb{R}$ is complete, A has a least upper bound c. For any positive real ϵ, $b \leq c + \epsilon$ by virtue of c being the least upper bound, and in fact we have $b < c + \epsilon$ since no equality could hold for some value of ϵ with $\leq$ also holding for $\frac{\epsilon}{2}$. Since $c - \epsilon$ cannot be an upper bound of A we have $c - \epsilon < b$, so as $-\epsilon < b - c < \epsilon$ we have $b \simeq c$.
For any $d \in \mathbb{R}$ s.t. $b \simeq d$, d is infinitely close to c by transitivity of $\simeq$, which would necessitate that $d = c$ as 0 is the only real infinitesimal and $d - c$ is real, and this proves the uniqueness of the halo. □

Definition 3.13. The **shadow** of a hyperreal x is the real number r for which $x \in \text{hal}(r)$. The shadow of x is denoted by $\text{sh}(x)$.

Lemma 3.14. *Let $x, y \in {}^*\mathbb{R}$ be limited*

ii. $sh(x + y) = sh(x) + sh(y)$

ii. *If $sh(x) \neq 0$, then* $sh(\frac{1}{x}) = \frac{1}{sh(x)}$

iii. $sh(x \cdot y) = sh(x) \cdot sh(y)$

[7] Numbers with this property are sometimes called "Positive Unlimited"
[8] Numbers with this property are sometimes called "Negative Unlimited"

iv. $sh(|x|) = |sh(x)|$

v. $b \leq c \implies sh(b) \leq sh(c)$

Proof.

i. Let $a = \text{sh}(x)$ and $b = \text{sh}(y)$. We can find infinitesimals x', y' s.t. $x = a + x'$ and $y = b + y'$, so

$$|(x + y) - (a + b)| = |x' + y'|$$

which is infinitesimal.

ii. Let $z = \text{sh}(x)$. As $\text{sh}(x) \neq 0$, neither x nor z can be infinitesimal, and is limited by our overall assumption. For any $\epsilon \in \mathbb{R}$, $|x - z| < \epsilon|xz|$ as $|xz|$ is not infinitesimal and is limited. Hence,

$$\epsilon > \frac{|x - z|}{|xz|} = |\frac{1}{z} - \frac{1}{x}|$$

So $\frac{1}{x} \simeq \frac{1}{z}$, and by the uniqueness of shadows we have our result.

iii. Let $a = \text{sh}(x)$ and $b = \text{sh}(y)$. We can find infinitesimals x', y' s.t. $x = a + x'$ and $y = b + y'$, so

$$|xy - ab| = |(a + x')(b + y') - ab| = |ay' + x'b + x'y'|$$

As $ay', x'b$, and $x'y'$ are all infinitesimal, so too is their sum, so we have that $xy \simeq ab$, and by the uniqueness of shadows we have our result.

iv. Let $a = \text{sh}(x)$. We can find an infinitesimal x' s.t. $x = a + x'$, so

$$||x| - |a|| \leq |x - a| = |x'|$$

hence $\text{sh}(|x|) \simeq |a| = |\text{sh}(a)|$, so $\text{sh}(|x|) \simeq |\text{sh}(a)|$

v. The cases of $b = c$ and $b \simeq c$ are trivial by the uniqueness of shadows, so we let $b < c$. Suppose that $\text{sh}(c) < \text{sh}(b)$, then $b < c \simeq \text{sh}(c) < \text{sh}(b)$ and hence $|\text{sh}(b) - b| \not< \text{sh}(b) - \text{sh}(c)$. This would contradict b belonging to the halo of $\text{sh}(b)$, so it must be the case that $\text{sh}(c) \leq \text{sh}(b)$

$\square$

Definition 3.15. A subset $A \subset \mathbb{R}$ can be "enlarged" to a subset ${}^*A \in {}^*\mathbb{R}$ defined as the set of all hyperreals which are U-equivalent to a sequence whose range belongs to A, so

$${}^*A = \{[\mathbf{r}] : \mathbf{r} \in \mathbb{R}^{\mathbb{N}}, \{n \in \mathbb{N} : r_n \in A\} \in U\}$$

Definition 3.16. A function $f : \mathbb{R} \to \mathbb{R}$ can be extended to a function ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ given by

$${}^*f([\mathbf{r}]) = [(f(r_1), f(r_2), f(r_3), \cdots)]$$

for any $[\mathbf{r}] \in {}^*\mathbb{R}$.
When $\mathbf{r} =_U \mathbf{r}'$, $\{n \in \mathbb{N} : r_n = r'_n\} \subseteq \{n \in \mathbb{N} : f(r_n) = f(r'_n)\}$, so $\{n \in \mathbb{N} : f(r_n) = f(r'_n)\} \in U$ and ${}^*f$ is well defined.
Similarly, for a function $f : A \to \mathbb{R}$, we can extend to a function ${}^*f : {}^*A \to {}^*\mathbb{R}$ given by ${}^*f([\mathbf{r}]) = [\mathbf{s}]$, where $\mathbf{s}$ is the sequence given by

$$s_n = \begin{cases} f(r_n) & \text{if } n \in \{n \in \mathbb{N} : r_n \in A\} \\ 0 & \text{otherwise} \end{cases}$$

4 Sequences

Definition 4.1. For a real valued sequence (r_n), we consider the function $s : \mathbb{N} \to \mathbb{R}$ given by $s(n) = r_n$. We can extend the sequence as we extended functions in 3.16 to a function ${}^*S :^* \mathbb{N} \to^* \mathbb{R}$, so

$$^*S([\mathbf{n}]) = \Big[\Big(S(n_1), S(n_2), \cdots\Big)\Big]$$

.

An **extended term** of a sequence is the a value at an unlimited index N, and the **extended tail** of a sequence is the set of all of the extended terms of that sequence, i.e.

$$\{s_n : n \in^* \mathbb{N},\ \text{n is unlimited}\}$$

Theorem 4.2. *A real-valued sequence converges to $L \in \mathbb{R} \iff$ The members of the extended tail are all infinitely close to L*

Proof.
$\Rightarrow$: For any positive $\epsilon \in \mathbb{R}$, we have some $m_\epsilon \in \mathbb{N}$, the sentence

$$(\forall n \in \mathbb{N})(m_\epsilon < n \to |s_n - L| < \epsilon)$$

So by transfer, we have the sentence

$$(\forall n \in {}^*\mathbb{N})(m_\epsilon < n \to |s_n - L| < \epsilon)$$

An unlimited $N \in {}^*\mathbb{N}$ exhibits $m_\epsilon < N$ for all real m_ϵ, so $|s_N - L| < \epsilon$ for all positive ϵ, hence $S_N \simeq L$.

$\Leftarrow$: For any positive $\epsilon \in \mathbb{R}$ and an unlimited hypernatural $N \in {}^*\mathbb{N}$, if $n > N$, then n is unlimited and $s_n \simeq L$, so $|s_n - L| < \epsilon$. For each positive real ϵ, we have the sentence

$$(\exists N \in {}^*\mathbb{N})(\forall n \in {}^*\mathbb{N})(n > N \to |s_n - L| < \epsilon)$$

Then, by transfer, we have the sentence

$$(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \to |s_n - L| < \epsilon)$$

Since we can find such a sentence for every $0 < \epsilon \in \mathbb{R}$, our sequence converges. □

Lemma 4.3. *A real-valued sequence (s_n) is bounded in $\mathbb{R} \iff$ The elements of its extended tail are all limited.*

Proof.
$\Rightarrow$: We have some $b \in \mathbb{R}$ s.t. $(\forall n \in \mathbb{N})(|s_n| < b)$, so by transfer $(\forall n \in {}^*\mathbb{N})(|s_n| < b)$, and this applies to every unlimited $N \in {}^*\mathbb{N}$.

$\Leftarrow$: For any limited index n, a sequence (s_n) has $s_n \in \mathbb{R}$ bounded. As we also have s_n limited for every unlimited n, for any unlimited $r \in {}^*\mathbb{R}$ every index of (s_n) is less than r by transitivity, so $(\exists r \in {}^*\mathbb{R})(\forall n \in {}^*\mathbb{N})(|s_n| < r)$, hence by transfer $(\exists r \in \mathbb{R})(\forall n \in \mathbb{N})(|s_n| < r)$ □

Theorem 4.4. ***(Nonstandard Cauchy Convergence Criterion)*** *A real valued sequence (s_n) converges $\iff$ any two extended terms are infinitely close to each other.*

Proof.
$\Rightarrow$: For any unlimited $N, M \in {}^*\mathbb{N}$, $s_N \simeq L$ and $L \simeq s_M$ by Theorem 4.2, so by transitivity of $\simeq$, $s_N \simeq s_M$.

$\Leftarrow$: For any $\epsilon \in \mathbb{R}$, we have the sentence $(\exists N \in {}^*\mathbb{N})(n > N \to |s_n - s_N| < \epsilon)$, specifically for any unlimited $N \in {}^*\mathbb{N}$, so by transfer our sequence is Cauchy. This implies that (S_n) is bounded[9], so by Theorem 4.3 for an unlimited $m \in {}^*\mathbb{N}$ we find that s_m is limited. By Theorem 3.12, s_m has a shadow L, so by the transitivity of $\simeq$ with our assumption we have that the extended tail of (S_n) is a subset of the halo of L, hence by Theorem 4.2 we have our result. □

[9] as for $\epsilon = 1$ we have that $(s_n) \subset (\min(s_1, s_2, \cdots s_{N_\epsilon} - \epsilon), \max(s_1, s_2, \cdots s_{N_\epsilon} + \epsilon))$

5 Continuity

Theorem 5.1. *A function f is continuous at a point $c \iff f(x) \simeq f(c)$ whenever $x \simeq c$*

Proof.
$\Rightarrow$: For a positive $\epsilon \in \mathbb{R}$ we have $\delta \in \mathbb{R}$ such that $(\forall x \in \mathbb{R})(|x - c| < \delta \to |f(x) - f(c)| < \epsilon)$ is true. By transfer, $(\forall x \in {}^*\mathbb{R})(|x - c| < \delta \to |f(x) - f(c)| < \epsilon)$ is true, so when $x \simeq c$, $|x - c| < \delta$ for all $\delta \in \mathbb{R}$, including our δ in question, hence $|f(x) - f(c)| < \epsilon$. As this holds for all ϵ, $f(x) \simeq f(c)$.

$\Leftarrow$: For an $\epsilon \in \mathbb{R}$, consider an infinitesimal $\delta \in {}^*\mathbb{R}$. Whenever $|x - c| < \delta$, $x \simeq c$, so $f(x) \simeq f(c)$ and hence $|f(x) - f(c)| < \epsilon$. By transfer, the sentence $(\exists \delta \in {}^*\mathbb{R}^+)(\forall x \in {}^*\mathbb{R})(|x - c| < \delta \to |f(x) - f(c)| < \epsilon)$ yields $(\exists \delta \in \mathbb{R}^+)(\forall x \in \mathbb{R})(|x - c| < \delta \to |f(x) - f(c)| < \epsilon)$, and since we can find such a sentence for any ϵ, we have that f is continuous. □

Corollary 5.2. *The following follow almost immediately from Theorem 5.1:*

i. $\lim_{x \to c} f(x) = L \iff f(x) \simeq L$ *for all* $x \in {}^*A$ *with* $x \simeq c$ *and* $x \neq c$.

ii. $\lim_{x \to c} f(x) = \pm\infty \iff f(x)$ *is unlimited for all* $x \in {}^*A$ *with* $x \simeq c$ *and* $x \neq c$.

iii. $\lim_{x \to \pm\infty} f(x) = L \iff f(x) \simeq L$ *for all positive/negative unlimited* $x \in {}^*A$ *(and such* x *elements exist)*

Formal proofs of these are omitted.

Theorem 5.3. ***(The Intermediate Value Theorem)*** *If a real function f is continuous over $[a, b]$, then for any d between $f(a)$ and $f(b)$ there is some $c \in (a, b)$ with $f(c) = d$*

Proof. For each $n \in \mathbb{N}$, consider the partition $\{p_k = a + \frac{k(b-a)}{n} : k \in \mathbb{N}, k \leq n\}$. We let s_n denote the largest p_k such that $f(p_k) < d$. It must be the case that $a \leq s_n < b$ and $f(s_n) < d \leq f(s_n + \frac{b-a}{n})$. Since we can find such an s_n for any $n \in \mathbb{N}$, by transfer we can find such an s_N for an unlimited $N \in {}^*\mathbb{N}$. As s_N is limited, we let $c \in \mathbb{R}$ denote its shadow. As $\frac{b-a}{N}$ is infinitesimal, $s_N \simeq s_N + \frac{b-a}{N}$, so $s_N + \frac{b-a}{N} \simeq c$. By Theorem 5.1 then, $f(s_N) \simeq f(c) \simeq f(s_N + \frac{b-a}{N})$, and as $f(s_N) < d \leq f(s_N + \frac{b-a}{n})$, $d \simeq f(c)$. Since both of these are real values, we have $d = f(c)$. □

Theorem 5.4. ***(Extreme Value Theorem)*** *If a f is continuous on $[a, b]$ then f attains an absolute maximum and minimum on the interval.*

Proof. For each $n \in \mathbb{N}$, consider the partition $\{p_k = a + \frac{k(b-a)}{n} : k \in \mathbb{N}, k \leq n\}$. We let s_n denote the p_k such that $f(p_k) \leq f(s_n)$ for $0 < k \leq n$. By transfer, for an unlimited hypernatural N we can find s_N with $a \leq s_N \leq b$ and $f(a + \frac{k(b-a)}{N}) \leq f(s_N)$ for all hypernaturals $k \leq N$. We let d be the shadow of s_N, and by continuity note that $f(s_N) \simeq f(d)$. To show that this is a maximum, consider that for any natural n, every real $x \in [a, b]$ lies between two partition points for some k, i.e. $a + \frac{k(b-a)}{n} \leq x \leq a + \frac{(k+1)(b-a)}{n}$. By transfer, this holds for our unlimited hypernatural N, so $\frac{K(b-a)}{N} \leq x \leq \frac{(K+1)(b-a)}{N}$ for some hypernatural K, hence $x \simeq a + \frac{K(b-a)}{N}$ as $x - \frac{K(b-a)}{N} \leq \frac{(K+1)(b-a)}{N} - \frac{K(b-a)}{N} = \frac{(b-a)}{N} < \epsilon$ for all real epsilon since N is unlimited. It follows then that $f(x) \simeq f(a + \frac{K(b-a)}{N}) \leq f(s_N) \simeq d$, so d is indeed the max on the interval as hyperreal inequalities are preserved by shadows (Lemma 3.14(5)).
The proof of the existence of a minimum is similar. □

Theorem 5.5. *A function f is uniformly continuous on $A \iff f(x) \simeq f(y)$ for any $x, y \in {}^*A$ s.t. $x \simeq y$*

Proof.
$\Rightarrow$: We have that for any positive $\epsilon \in \mathbb{R}$ there is a positive $\delta \in \mathbb{R}$ s.t. $(\forall x, y \in A)(|x - y| < \delta \to |f(x) - f(y)| < \epsilon)$, so by transfer we have $(\forall x, y \in {}^*A)(|x - y| < \delta \to |f(x) - f(y)| < \epsilon)$. It follows then that when $x \simeq y$, $|x - y| < \delta$ for all positive real delta, hence $|f(x) - f(y)| < \epsilon$ for all positive real epsilon, so $f(x) \simeq f(y)$.

$\Leftarrow$: Let $d \in {}^*\mathbb{R}$ be a positive infinitesimal and let ϵ be any positive real number. For $x, y \in {}^*A$ exhibiting $|x-y| < d$, $x \simeq y$ by transitivity, so $f(x) \simeq f(y)$ and hence $|f(x) - f(y)| < \epsilon$. We can write the sentence $(\exists d \in {}^*\mathbb{R})(\forall x, y \in {}^*A)(|x-y| < d \to |f(x) - f(y)| < \epsilon)$, and so by transfer we have $(\exists d \in \mathbb{R})(\forall x, y \in A)(|x-y| < d \to |f(x) - f(y)| < \epsilon)$, and this is the standard definition for uniform continuity on a set. □

Theorem 5.6. *If a real function f is continuous on a closed interval $[a, b] \subset \mathbb{R}$, then f is uniformly continuous on $[a, b]$*

Proof. For $x, y \in {}^*[a, b]$ s.t. $x \simeq y$, we let $c = \mathrm{sh}(x)$. By transitivity, $y \simeq c$. As f is continuous, by Theorem 5.1, $f(x) \simeq f(c)$ and $f(c) \simeq f(y)$, so $f(x) \simeq f(y)$ and by Theorem 5.5 we have our result. □

Definition 5.7. Consider a sequence of functions (f_n) which all map A to $\mathbb{R}$. We can extend this sequence to hypernatural indices by identifying it with the function $F : \mathbb{N} \times A \to \mathbb{R}$ given by $F(n, x) = f_n(x)$. The extension of F, ${}^*F : {}^*\mathbb{N} \times {}^*A \to {}^*\mathbb{R}$, maps along the same pattern as before for each index, so for $\mathbf{N} = [n_1, n_2, n_3, \cdots] \in {}^*\mathbb{N}$ and a hyper real $\mathbf{a} = [a_1, a_2, a_3, \cdots] \in {}^*A$, we have

$${}^*F(\mathbf{N}, \mathbf{a}) = \big[(f_{n_1}(a_1), f_{n_2}(a_2), f_{n_3}(a_3), \cdots)\big]$$

Proposition 5.8. *A sequence of real valued functions (f_n) defined on $A \subset \mathbb{R}$ converges pointwise to a function $f : A \to \mathbb{R} \iff$ for each $x \in A$ and each unlimited $N \in {}^*\mathbb{N}$ we have $f_n(x) \simeq f(x)$*

Proof.
$\Rightarrow$: For any $x \in A$, $\lim_{n\to\infty} f_n(x) = f(x)$, so by Theorem 4.2 each unlimited $N \in {}^*\mathbb{N}$ exhibits $f_N(x) \simeq f(x)$

$\Leftarrow$: Let $x \in A$. As every unlimited $N \in {}^*\mathbb{N}$ has $f_N(x) \simeq f(x)$, by Theorem 4.2 $\lim_{n\to\infty} f_n(x) = f(x)$. As this holds for all $x \in A$, the sequence converges pointwise to f. □

Proposition 5.9. *A sequence of real valued functions (f_n) defined on $A \subset \mathbb{R}$ converges uniformly to a function $f : A \to \mathbb{R} \iff$ for each $x \in {}^*A$ and each unlimited $N \in {}^*\mathbb{N}$ we have $f_n(x) \simeq f(x)$*

Proof.
$\Rightarrow$: For any positive real ϵ, we can find $N_\epsilon \in \mathbb{N}$ s.t. the sentence $(\forall n \in \mathbb{N})(\forall x \in A)(n > N_\epsilon \to |f_n(x) - f(x)| < \epsilon)$ is true. By transfer then, $(\forall n \in {}^*\mathbb{N})(\forall x \in {}^*A)(n > N \to |f_n(x) - f(x)| < \epsilon)$. For any unlimited hypernatural m, $m > N_\epsilon$ for any N_ϵ, hence for any $x \in {}^*A$ we have $f_m(x) \simeq f(x)$.

$\Leftarrow$: Let $\epsilon \in \mathbb{R}$ be positive and let $N \in {}^*\mathbb{N}$ be unlimited. For any $n > N$, as n is unlimited we have $f_n(x) \simeq f(x)$ for any $x \in {}^*A$. We can write then that $(\exists N \in {}^*\mathbb{N})(\forall n \in {}^*\mathbb{N})(\forall x \in {}^*A)(n > N \to |f_n(x) - f(x)| < \epsilon)$, so by transfer we have $(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(\forall x \in A)(n > N \to |f_n(x) - f(x)| < \epsilon)$. Since we can write such a sentence for any positive real ϵ, our sequence converges uniformly. □

Theorem 5.10. *For a sequence of functions (f_n), if each f_n is continuous on $A \subset \mathbb{R}$, and (f_n) converges uniformly to a function $f : A \to \mathbb{R}$, then f is continous on A*

Proof. Let $c \in A$ and let $x \in {}^*A$ s.t. $x \simeq c$. We need to prove that $|f(x) - f(c)| < \epsilon$ for any real ϵ.
Notice first that for any $n \in \mathbb{N}$, $|f(x) - f(c)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|$. By uniform convergence of (f_n), $\exists\ m \in \mathbb{N}$ s.t. for $n > m$, $|f_n(x) - f(x)| < \frac{\epsilon}{3}$. For $n = m+1$, f_n is continuous, so $|f_n(x) - f_n(c)|$ is infinitesimal by Theorem 5.1. Hence,

$$|f(x) - f(c)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon$$

As we can say this for arbitrary ϵ, $f(x) \simeq f(c)$, so f is continuous at c and hence continuous on A. □

6 Differentiation

Theorem 6.1. *For f defined at $x \in \mathbb{R}$, the Derivative of f at x is some $L \in \mathbb{R} \iff$ for every nonzero infinitesimal δ, $f(x+\delta)$ is defined and $\frac{f(x+\delta)-f(x)}{\delta} \simeq L$*

Proof. Let $g(h) = \frac{f(x+h)-f(x)}{h}$. f is differentiable at $x \iff \lim_{h\to 0} g(h) = L$, and by Corollary 5.2(1) we have our result. ☐

Definition 6.2. When f is differentiable at x, we write that $f'(x) = \mathrm{sh}(\frac{f(x+\epsilon)-f(x)}{\epsilon})$ for all nonzero infinitesimals ϵ.

Definition 6.3. We let Δx denote an *increment of* x, a nonzero infinitesimal. For a function of x, we denote the corresponding *increment of* f with $\Delta f = f(x+\Delta x) - f(x)$.
We sometimes write $\Delta f(x, \Delta x)$ for Δf to specify the x and Δx used in the function increment's creation.

Theorem 6.4. *If a function f is differentiable at x, then f is continuous at x*

Proof. As f is differentiable, by Theorem 6.1 $\frac{\Delta f}{\Delta x} \simeq f'(x)$ for an arbitrary Δx, so we see that $\frac{\Delta f}{\Delta x}$ is limited. As $\Delta f = \frac{\Delta f}{\Delta x}\Delta x \simeq f'(x)\Delta x$, Δf is infinitely close to the product of an infinitesimal and a limited number, so it is an infinitesimal and is less than every positive real. Hence, $f(x+\Delta x) \simeq f(x)$, and since this holds for any infinitesimal Δx, by Theorem 5.1 f is continuous. ☐

Definition 6.5. The *differential of f at x for an increment* Δx is denoted by $df = f'(x)\Delta x$

Theorem 6.6. ***(The Incremental Equation)*** *If $f'(x)$ exists at a real x and $\Delta x = dx$ is infinitesimal, then Δf and df are infinitesimal, and there is an infinitesimal ϵ for x and Δx s.t.*

$$\Delta f = f'(x)\Delta x + \epsilon\Delta x = df + \epsilon\Delta x$$

and hence

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \epsilon\Delta x$$

Proof. As $dx = \Delta x$, Definition 6.5 yields $f'(x) = \frac{df}{dx}$ as $dx = \Delta x$, and since $f'(x)$ is limited and dx is infinitesimal, df must be infinitesimal. By Theorem 6.4, f is continuous, so $\Delta f = f(x+\Delta x) - f(x)$ is infinitesimal.
As we saw in Theorem 6.1, $\frac{\Delta f}{\Delta x} \simeq f'(x)$, so for

$$\epsilon = \frac{\Delta f}{\Delta x} - f'(x)$$

we see that ϵ is infinitesimal. Notice then that

$$\Delta f - df = \Delta f - f'(x)\Delta x = \epsilon\Delta x$$

Hence we have $\Delta f = df + \epsilon\Delta x$.

☐

Theorem 6.7. ***(Addition, Multiplication, and Quotient rules)*** *If f and g are differentiable at $x \in \mathbb{R}$, then $f+g$, $f \cdot g$, and $\frac{f}{g}$ (for $g(x) \neq 0$) are differentiable at x with*

i. $(f+g)'(x) = f'(x) + g'(x)$

ii. $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$

iii. $(\frac{f}{g})'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$

Proof.

i. For any infinitesimal $\Delta x \neq 0$, by Theorem 6.1 $(f+g)(x+\Delta x) = f(x+\Delta x)+g(x+\Delta x)$ is defined.

$$\Delta(f+g) = \big(f(x+\Delta x)+g(x+\Delta x)\big) - \big(f(x)+g(x)\big) = \Delta f + \Delta g$$

From here, we notice that as in Theorem 6.4 $\frac{\Delta f}{\Delta x} \simeq f'(x)$ and $\frac{\Delta g}{\Delta x} \simeq g'(x)$, we notice that

$$\frac{\Delta(f+g)}{\Delta x} = \frac{\Delta f}{\Delta x} + \frac{\Delta g}{\Delta x} \simeq f'(x)+g'(x)$$

So by Theorem 6.1 we have $(f+g)'(x) = f'(x)+g'(x)$.

ii. For any infinitesimal $\Delta x \neq 0$, by Theorem 6.1 $(fg)(x+\Delta x) = f(x+\Delta x)g(x+\Delta x)$ is defined. Then,

$$\begin{aligned}\Delta fg &= f(x+\Delta x)g(x+\Delta x) - f(x)g(x)\\ &= \big(f(x)+\Delta f\big)\big(g(x)+\Delta g\big) - f(x)g(x)\\ &= \Delta f \cdot g(x) + \Delta g \cdot f(x) + \Delta f \cdot \Delta g\end{aligned}$$

The fact that Δg is infinitesimal yields $\Delta g \simeq 0$, so by Theorem 6.1 and Lemma 3.14,

$$\frac{\Delta fg}{\Delta x} = \frac{\Delta f}{\Delta x}\cdot g(x)+f(x)\cdot\frac{\Delta g}{\Delta x}+\frac{\Delta f}{\Delta x}\cdot\Delta g \simeq f'(x)\cdot g(x)+f(x)\cdot g'(x)+f'(x)\cdot 0 = f'(x)\cdot g(x)+f(x)\cdot g'(x)$$

Hence by Theorem 6.1, $(fg)'(x) = f'(x)g(x)+f(x)g'(x)$.

iii. By Theorem 6.4, g is continuous at x, so as $0 \neq g(x) \in \mathbb{R}$, $g(x)$ is not infinitesimal. For any infinitesimal Δx we have $g(x+\Delta x) \simeq g(x)$, so $g(x+\Delta x) \not\simeq 0$ and thus $g(x+\Delta x) \neq 0$. By Theorem 6.1 then, $(\frac{f}{g})(x+\Delta x) = \frac{f(x+\Delta x)}{g(x+\Delta x)}$ is defined. We also note that by Lemma 3.14, $\frac{1}{g(x+\Delta x)} \simeq \frac{1}{g(x)}$. From here we see that

$$\begin{aligned}\Delta(\frac{f}{g})(x) &= \frac{f(x+\Delta x)}{g(x+\Delta x)} - \frac{f(x)}{g(x)}\\ &= \frac{f(x+\Delta x)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+\Delta x)}{g(x+\Delta x)g(x)}\\ &= \frac{\Delta f\cdot g(x) - f(x)\cdot\Delta g}{g(x+\Delta x)g(x)}\end{aligned}$$

Then, by Lemma 3.14,

$$\frac{\Delta(\frac{f}{g})(x)}{\Delta x} = \frac{\frac{\Delta f}{\Delta x}\cdot g(x) - f(x)\cdot\frac{\Delta g}{\Delta x}}{g(x+\Delta x)g(x)} \simeq \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{g(x)^2}$$

So by Theorem 6.1, $(\frac{f}{g})'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}$

□

Theorem 6.8. ***(Chain Rule)*** *For a function f differentiable at $x \in \mathbb{R}$ and a function g differentiable at $f(x)$, $g \circ f$ is differentiable at x and $(g\circ f)'(x) = g'(f(x))f'(x)$*

Proof. As $f'(x)$ exists, for any infinitesimal Δx, $f(x+\Delta x)$ exists and $f(x+\Delta x) \simeq f(x)$ by Theorems 6.1 and 6.4 respectively. Similarly then, $g(f(x+\Delta x))$ and $g(f(x)) \simeq g(f(x+\Delta x))$ since $g'(f(x))$ exists. For $\Delta f = f(x+\Delta x) - f(x)$, notice that

$$\Delta(g\circ f)(x,\Delta x) = g(f(x+\Delta x)) - g(f(x)) = g(\Delta f + f(x)) - g(f(x)) = \Delta g(f(x),\Delta f)$$

so by the incremental equation for g, $\Delta(g\circ f) = g'(f(x))\Delta f + \epsilon\Delta f$, and thus

$$\frac{\Delta(g\circ f)}{\Delta x} = g'(f(x))\frac{\Delta f}{\Delta x} + \epsilon\frac{\Delta f}{\Delta x} \simeq g'(f(x))f'(x) + 0f'(x)$$

and so by Theorem 6.1 we have $(g\circ f)'(x) = g'(f(x))f'(x)$. □

Theorem 6.9. ***(Critical Point Theorem)*** *If f has a maximum or minimum at $x \in (a,b) \subset \mathbb{R}$, and f is differentiable at x, $f'(x) = 0$.*

Proof. Take the case that $f(x)$ is a maximum. We have the sentence $(\forall y \in \mathbb{R})(a < y < b \to f(y) \leq f(x))$, so by transfer for any infinitesimal Δx we have $f(x + \Delta x) \leq f(x)$. We let ϵ and δ be a positive and negative infinitesimal respectively, so by Theorem 6.1

$$f'(x) \simeq \frac{f(x+\epsilon) - f(x)}{\epsilon} \leq 0 \leq \frac{f(x+\delta) - f(x)}{\delta} \simeq f'(x)$$

It is a consequence of Theorem 3.12 that since each hyperreal has a unique shadow and $f'(x)$ is real, $0 \leq f'(x) \leq 0$ [10].
The proof for $f(x)$ as a minimum is similar. □

Theorem 6.10. ***(Inverse Function Theorem)*** *For a continuous and strictly monotonic function f with inverse g, if f is differentiable at $x \in (a,b)$ with $f'(x) \neq 0$, then g is differentiable at $y = f(x)$ with $g'(y) = \frac{1}{f'(x)}$*

Proof. Let Δy be a nonzero infinitesimal. We want to show that $\frac{g(y+\Delta y) - g(y)}{\Delta y} \simeq \frac{1}{f'(x)}$.
Suppose that $g(y + \Delta y) \not\simeq g(y)$, so we can find a real value r between them. By the monotonicity of f, $f(r)$ lies between y and $y + \Delta y$. This implies that y and $y + \Delta y$ are an appreciable distance apart, but as Δy is infinitesimal this is not true, hence

$$\Delta x = g(y + \Delta y) - g(y) = \Delta g(y, \Delta y)$$

is infinitesimal, and since g is strictly monotonic, Δx is nonzero.
From here, as $g(y) = x$, we have $\Delta x + x = g(y + \Delta y)$, so $f(\Delta x + x) = f(g(y + \Delta y)) = y + \Delta y$, and hence

$$\Delta y = f(\Delta x + x) - f(x) = \Delta f(x, \Delta x)$$

So, as $\frac{\Delta g}{\Delta y} = \frac{\Delta x}{\Delta f} = \frac{1}{\frac{\Delta f}{\Delta x}}$ and $\frac{\Delta f}{\Delta x} \simeq f'(x)$, by Lemma 3.14 $\frac{\Delta g}{\Delta y} \simeq \frac{1}{f'(x)}$, so by Theorem 6.1 $g'(y) = \frac{1}{f'(x)}$. □

7 Integration

Definition 7.1. A **partition** of $[a,b]$ is a set of points $\{a_0, \cdots, x_n\} \subset [a,b]$ with $a = x_0 < x_1 < \cdots < x_n = b$. A **refinement** of a partition P is another partition P' s.t. $P \subset P'$.

Definition 7.2. For a function f bounded over $[a,b]$ and a partition P of $[a,b]$, we define $M_i = \sup\{f(x) : x \in [x_{i-1}, x_i]\}$ and $m_i = \inf\{f(x) : x \in [x_{i-1}, x_i]\}$ for all i s.t. $x_i \in P \setminus \{x_0\}$. We then define the following

i. **Upper Riemann Sum**: $U_a^b(f,P) = \sum_{i=1}^n M_i \cdot (x_i - x_{i-1})$

ii. **Lower Riemann Sum**: $L_a^b(f,P) = \sum_{i=1}^n m_i \cdot (x_i - x_{i-1})$

iii. **Ordinary Riemann Sum**: $S_a^b(f,P) = \sum_{i=1}^n f(x_{i-1}) \cdot (x_i - x_{i-1})$

Remark 7.3. For $m = \inf\{f(x) : x \in [a,b]\}$ and $M = \sup\{f(x) : x \in [a,b]\}$, notice that for any function f over $[a,b]$ and any partition P of $[a,b]$ we have

$$m(b,a) \leq L_a^b(f,P) \leq S_a^b(f,P) \leq U_a^b(f,P) \leq M(b-a)$$

Lemma 7.4. *For partitions P_1, P_2 and their common refinement P', it can be shown that*

$$L_a^b(f,P_1) \leq L_a^b(f,P') \leq U_a^b(f,P') \leq U_a^b(f,P_2)$$

A formal proof of this is omitted, but we notice that in general, any upper sum over an interval $[a,b]$ is greater than any lower sum over $[a,b]$.

[10] If $f'(x) < 0$, then as $\frac{f(x+\delta)-f(x)}{\delta}$ is "closer" to $f'(x)$ than 0 is, hence $\frac{f(x+\delta)-f(x)}{\delta} < 0$

Definition 7.5. A function f is **Riemann Integrable** on $[a,b]$ if the supremum of the lower sums is equal to the infimum of the upper sums. We denote this value by $\int_a^b f(x)dx$, so

$$\int_a^b f(x)dx = \sup\{L_a^b(f,P) : P \text{ is a partition of } [a,b]\} = \inf\{U_a^b(f,P) : P \text{ is a partition of } [a,b]\}$$

Remark 7.6. Our definition makes it clear that a value $\int_a^b f(x)dx$ is the integral of f over $[a,b]$ if, and only if, $L_a^b(f,P) \leq \int_a^b f(x)dx \leq U_a^b(f,P)$ for all partitions P of $[a,b]$, and if for any real $\epsilon > 0$, we can find a partition P s.t. $U_a^b(f,P) - L_a^b(f,P) < \epsilon$

Remark 7.7. For any $\Delta x \in \mathbb{R}$, we let $P_{\Delta x}$ be a partition with the maximum number of subintervals of width Δx. So, for any $i \in [1, n-1]$ we have $x_i - x_{i-1} = \Delta x$, and for $i = n$ we have $x_i - x_{i-1} \leq \Delta x$

We write that $L_a^b(f,\Delta x), S_a^b(f,\Delta x)$, and $U_a^b(f,\Delta x)$ to represent $L_a^b(f,P_{\Delta x}), S_a^b(f,P_{\Delta x})$, and $U_a^b(f,P_{\Delta x})$. Since $L_a^b(f,\Delta x), S_a^b(f,\Delta x)$, and $U_a^b(f,\Delta x)$ are all real valued functions, we extend them to functions as in Definition 3.16.

Theorem 7.8. *If a function f is continuous on the interval $[a,b]$, then for any positive infinitesimal Δx we have $L_a^b(f,\Delta x) \simeq U_a^b(f,\Delta x)$*

Proof. For any positive real Δx, we can find $\mu(\Delta x)$ equal to the maximum difference between M_i and m_i for $i \in [1,n]$ in the partition $P_{\Delta x}$. For j s.t. $M_j - m_j = \mu(\Delta x)$, by the extreme value Theorem we can find values $c(\Delta x), d(\Delta x) \in [x_{j-1}, x_j]$ with $f(c(\Delta x)) = M_j$ and $f(d(\Delta x)) = m_j$. Hence,

$$U_a^b(f,\Delta x) - L_a^b(f,\Delta x) = \sum_{i=1}^{n}(M_i - m_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{n}\mu(\Delta x)(x_i - x_{i-1}) = \mu(\Delta x)\sum_{i=1}^{n}(x_i - x_{i-1}) = \mu(\Delta x)(b-a)$$

So we have

$$|c(\Delta x) - d(\Delta x)| \leq \Delta x \text{ and } U_a^b(f,\Delta x) - L_a^b(f,\Delta x) \leq \Big[f\big(c(\Delta x)\big) - f\big(d(\Delta x)\big)\Big](b-a)$$

for any real Δx, and by transfer we then have this for any hyperreal Δx. If we let Δx be a positive infinitesimal, we see that $c(\Delta x) \simeq d(\Delta x)$ by Lemma 3.14(5) as the endpoints of their subinterval are infinitely close, so by the continuity of f we have $f(c(\Delta x)) \simeq f(d(\Delta x))$. Consequently, $U_a^b(f,\Delta x) - L_a^b(f,\Delta x)$ is infinitesimal. □

Theorem 7.9. *Any function f which is continuous over a closed interval is Riemann Integrable over that interval. Additionally, $\int_a^b f(x)dx = sh\big(S_a^b(f,\Delta x)\big)$ for any infinitesimal $\Delta x \in {}^*\mathbb{R}$*

Proof. Take $\Delta x_1, \Delta x_2 \in {}^*\mathbb{R}$ to be infinitesimals and let L_1, L_2 represent their respective lower sums and U_1, U_2 represent their upper sums. Recall that in Remark 7.4 we saw that $L_a^b(f,\Delta y) \leq U_a^b(f,\Delta z)$ for any real $\Delta y, \Delta z$, so by transfer we have $L_a^b(f,\Delta y) \leq U_a^b(f,\Delta z)$ for any hyperreals $\Delta y, \Delta z$. As such, one of the following must be true

$$L_1 \leq L_2 \leq U_1 \leq U_2, \text{ or } L_2 \leq L_1 \leq U_1 \leq U_2, \text{ or}$$
$$L_1 \leq L_2 \leq U_2 \leq U_1, \text{ or } L_2 \leq L_1 \leq U_2 \leq U_1$$

By Theorem 7.8 though, $L_1 \simeq U_1$ and $L_2 \simeq U_2$, and by the inequality above it must then be the case that $L_1 \simeq L_2 \simeq U_1$ or $L_2 \simeq L_1 \simeq U_2$, hence $L_1 \simeq L_2 \simeq U_1 \simeq U_2$.
By transfer of Remark 7.3, it must be the case that the ordinary sums lie between the upper and lower sums, so then $S_a^b(f,\Delta x_1) \simeq S_a^b(f,\Delta x_2)$ are both bounded by $m(b-a)$ and $M(b-a)$, hence we use I to denote the shadow $\mathrm{sh}\big(S_a^b(f,\Delta x_1)\big)$, shared by all infinitesimals Δx.

To show that I is $\int_a^b f(x)dx$, recall the criteria listed in Remark 7.6.
For any partition P, we can choose any infinitesimal Δx and see that

$$L_a^b(f,P) \leq U_a^b(f,\Delta x) \simeq I \simeq L_a^b(f,\Delta x) \leq U_a^b(f,P)$$

and as $L_a^b(f,P)$ and $U_a^b(f,P)$ are real this yields $L_a^b(f,P) \leq I \leq U_a^b(f,P)$.
For any positive $\epsilon \in \mathbb{R}$, by Theorem 7.8 we can take any positive infinitesimal Δx and have

$$U_a^b(f,\Delta x) - L_a^b(f,\Delta x) < \epsilon$$

and by transfer this yields the existence of a real Δx which satisfies this property. □

Theorem 7.10. ***(Standard Properties of Integration)***

i. $\int_a^b cf(x)dx = c\int_a^b f(x)dx$

ii. $\int_a^b (f(x)+g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx$

iii. $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$ **if** $a \leq c \leq b$

iv. $\int_a^b f(x)dx \leq \int_a^b g(x)$ **if** $f(x) \leq g(x)$ **for all** $x \in [a,b]$

v. $m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$ **if** $m \leq f(x) \leq M$ **for all** $x \in [a,b]$

Proof.

i. For any real Δx, our understanding of summations shows that

$$S_a^b(cf,\Delta x) = \sum_{i=1}^{n} cf(x_{i-1})\cdot(x_i - x_{i-1}) = c\cdot\sum_{i=1}^{n} f(x_{i-1})\cdot(x_i - x_{i-1}) = c\cdot S_a^b(f,\Delta x)$$

By transfer, this holds for any hyperreal Δx, including any infinitesimal. Taking the shadow of each end of this equality, by Lemma 3.14 and Theorem 7.9 we have

$$\int_a^b cf(x)dx = \mathrm{sh}(S_a^b(cf,\Delta x)) = \mathrm{sh}(cS_a^b(f,\Delta x)) = c\cdot\mathrm{sh}(S_a^b(f,\Delta x)) = c\int_a^b f(x)dx$$

ii. For any real Δx, $S_a^b(f+g,\Delta x) = S_a^b(f,\Delta x) + S_a^b(g,\Delta x)$, so by transfer this holds for any infinitesimal Δx. Taking the shadow of each side of this, by Lemma 3.14 and 7.9 we have our result.

iii. For any real Δx, $S_a^b(f,\Delta x) = S_a^c(f,\Delta x) + S_c^b(f,\Delta x)$, so by transfer this holds for any infinitesimal Δx. Taking the shadow of each side of this, by Lemma 3.14 and 7.9 we have our result.

iv. For any real Δx, $S_a^b(f,\Delta x) \leq S_a^b(g,\Delta x)$, so by transfer this holds for any infinitesimal Δx. Taking the shadow of each side of this, by Lemma 3.14 and 7.9 we have our result.

v. For any real Δx, $m(b-a) \leq S_a^b(f,\Delta x) \leq M(b-a)$, so by transfer this holds for any infinitesimal Δx. Taking the shadow of each component of this, by Lemma 3.14 and 7.9 we have our result.

□

Theorem 7.11. ***(The Fundamental Theorem of Calculus)*** *For a continuous f over $[a,b]$ and $x \in [a,b]$, a function $F(x) = \int_a^x f(t)dt$ is differentiable on $[a,b]$ with $F'(x) = f(x)$*

Proof. Take $\Delta x \in \mathbb{R}$ s.t. $\Delta x < b - x$. By the standard properties of integration,

$$F(x+\Delta x) - F(x) = \int_x^{x+\Delta x} f(t)dt$$

By the extreme value Theorem, $f([x, x+\Delta])$ attains a minimum and maximum at $x_1, x_2 \in [x, x+\Delta]$, so

$$f(x_1)\Delta x \leq \int_x^{x+\Delta x} f(t)dt \leq f(x_2)\Delta x$$

Combining these, $f(x_1) \leq \frac{F(x+\Delta x)-F(x)}{\Delta x} \leq f(x_2)$. Since we have this for any real with $0 < \Delta x < b - x$, by transfer we can say the same for any positive infinitesimal Δx. Then, as $x_1, x_2 \in [x, x + \Delta x]$, $0 \leq (x_1 - x) + (x_2 - x) \leq 2(x + \Delta x - x)$, so $x_1 \simeq x \simeq x_2$, and by the continuity of f we have

$$f(x) \simeq f(x_1) \leq \frac{F(x + \Delta x) - F(x)}{\Delta x} \leq f(x_2) \simeq f(x)$$

so $\frac{F(x+\Delta x)-F(x)}{\Delta x} \simeq f(x)$. A similar proof works for a negative Δx, so by Theorem 6.1, $F'(x) = f(x)$. $\square$

Acknowledgements: I would like Dr. Peter May for his work in organizing the 2021 UChicago Math REU and the wealth of lecturers involved therein for all of their stimulating talks. I would also like to thank my mentor Connor Lockhart for his help in comprehending, outlining, and drafting this paper.

References

[1] C. C. Chang and H. Jerome Keisler. *Model Theory.* third edition. Dover, 2012.

[2] Robert Goldblatt. *Lectures on the Hyperreals.* Springer, 1998.