An Exceedingly Brief Overview of Infinitesimal Calculus

Paul Schulze

December 16th, 2024

Introduction

The idea of the *infinitesimal*, an infinitely small number that is still distinct from 0, has been around since the very beginning of calculus. Not only did Leibniz formulate his calculus in terms of infinitesimals, but they were used calculus education long after his death as well. Of course at the time infinitesimal theories were ill-formed, which made it difficult for mathematicians or students to go beyond simple derivatives and integrals [2]. It wasn't until the early 1960's when Abraham Robinson managed to formalize infinitesimal calculus, creating the field of non-standard analysis. Today non-standard analysis goes beyond simple calculus on the reals, but this paper will focus on the rigorous development of infinitesimal calculus, assuming no prior knowledge. Abraham Robinson's formulation of the infinitesimal calculus relied on the *compactness theorem* in mathematical logic—it is taught in MATH 499. We will instead use the more popular *ultrapower* formulation, which has the benefit of requiring less mathematical background. So much so that Henle & Kleinberg, in [2], attempted to write a book aimed at those learning calculus for the first time.

In section 1, we will go through the development of the *hyperreals* ${}^*\mathbb{R}$ as an *ultrapower* of $\mathbb{R}$. The basic idea is to take the ring of countable sequences of real numbers, $\mathbb{R}^{\mathbb{N}}$, and form equivalence classes. Two sequences will be in the same equivalence class if they are the same "almost everywhere"—a concept we will define rigorously using an *ultrafilter* on $\mathbb{N}$. ${}^*\mathbb{R}$ will end up being a set that contains all the real numbers within it, as well as a plethora of infinitesimal elements.

In section 2, we will develop a formal mathematical language. This language will be able to say much less than English, but we will prove the *transfer* principle, that any sentence of our formal language is true in $\mathbb{R}$ if and only if it is true in ${}^*\mathbb{R}$. This will give us a powerful tool to learn about ${}^*\mathbb{R}$.

In section 3, we will explore the structure of ${}^*\mathbb{R}$—what hyperreals are there, how do they relate to the real numbers, what can we say about subsets and functions on ${}^*\mathbb{R}$, etc. The concepts we cover in this section will be largely utilitarian, enabling our discussion in the final two sections.

In section 4 and section 5, we will develop differential and integral calculus. Our basic approach will be to take some useful fact about $\mathbb{R}$, transfer it to ${}^*\mathbb{R}$, use that fact to prove some result we want in ${}^*\mathbb{R}$, and then transfer that result back to $\mathbb{R}$. In this way, we will prove things about $\mathbb{R}$ by going "through" ${}^*\mathbb{R}$. Additionally, we will develop definitions of the derivative and integral that hew more closely to our intuition about what they are, as opposed to unwieldy ϵ-δ definitions of standard calculus. subsection 4.5, about the exponential function exp, was developed independently without the aid of any texts (although it has been done before).

Theorems, lemmas, and corollaries have all been credited to Goldblatt [1] or Henle & Kleinberg [2], the two texts referenced for this paper, where appropriate. If a theorem, lemma, or corollary is not cited, it to my knowledge did not appear in either text. (There is at least one example in the paper of a theorem that is not stated in the most relevant section of [1], but is a consequence of a later theorem, so there may be more instances of that that I am unaware of.) Definitions are from [1], with two exceptions. Firstly, the language used in section 2 has been modified for simplicity (as we will not be doing non-standard analysis on anything but $\mathbb{R}$), and the treatment of set symbols has been made rigorous. Secondly, the definition of $\exp = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ is not listed in either text, although it is hardly original. Proofs are frequently listed as "adapted from" one of the texts, meaning that they were modified in some way to be done with the tools already defined in the paper or otherwise streamlined. The proof of a restricted Łoś's Theorem in section 2 got the core approach of inducting on wffs from [2], but has been heavily modified to work with the language presented and give a formal treatment of set symbols, which in turn enables easy proofs of several theorems that appear in [1].

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

1.1 The Ultrafilter

The idea behind the *ultrafilter* $\mathcal{F} \subseteq \mathcal{P}(\mathbb{N})$ is that it should include all of the "very large" subsets of $\mathbb{N}$. This is so that we can define an equivalence relation $\equiv$ on sequences of real numbers, so that $\langle r_n \rangle \equiv \langle s_n \rangle$ when they are the same "almost everywhere," i.e. when $\{n \in \mathbb{N} \mid r_n = s_n\} \in \mathcal{F}$. We will use the following definition.

Definition. $\mathcal{F} \subseteq \mathbb{N}$ is a (non-principal) ultrafilter on $\mathbb{N}$:

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

A subset of $\mathcal{P}(\mathbb{N})$ that satisfies 1 and 2 is a *filter*. For instance, $\mathcal{P}(\mathbb{N})$ is a filter on $\mathbb{N}$, as is $\emptyset$. A *proper* filter is one that does not contain $\emptyset$ (as if $\emptyset \in \mathcal{F}$, then $\mathcal{P}(\mathbb{N}) = \mathcal{F}$ by 2). Strictly speaking, any proper filter that satisfies 3 is an ultrafilter, but we will use "ultrafilter" to refer only to filters that satisfy 4 as well.

Call a set *cofinite* if its complement is finite. Every cofinite subset of $\mathbb{N}$ is in $\mathcal{F}$, by 3 and 4.

If we think of $\mathcal{F}$ as being the subsets of $\mathbb{N}$ that include "almost all" of the natural numbers, then these make intuitive sense. If A and B both include "almost all" of the natural numbers, then surely $A \cap B$ does too—"basically no" elements are in $A - B$ or $B - A$ since "basically no" elements are outside A or B. If A includes "almost all" of $\mathbb{N}$, and $A \subseteq B$, then surely B includes "almost all" of $\mathbb{N}$ too. Etc. Note also that by a simple induction on 1, $\mathcal{F}$ is closed under finite intersections.

Of course, there are some unintuitive things about $\mathcal{F}$. By 3, it contains either the set of even numbers $2\mathbb{N}$ or the set of odd numbers $2\mathbb{N} + 1$, but not both.

Theorem 1.1. *There is at least one ultrafilter $\mathcal{F}$ on $\mathbb{N}$.*

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

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

So any totally ordered subset $T \subset P$ has a maximal element. By Zorn's Lemma, P has a maximal element—call it $\mathcal{F}$. We want to show $\mathcal{F}$ is an ultrafilter (this is [1, Exercise 2.5(6)]). By the definition of P, $\mathcal{F}$ is a proper filter. Since $\mathcal{F}^{co} \subseteq \mathcal{F}$, if we can show that for every $A \subseteq \mathbb{N}$ either A or A^c is in $\mathcal{F}$ then we will be done.

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

Let $\mathcal{F}'$ be the filter obtained by adding to $\mathcal{F}$ the the intersection $A \cap B$ for any $B \in \mathcal{F}$, and any superset of $A \cap B$ (this technique adapted from [2, Appendix A]). Note that $A = A \cap \mathbb{N}$ and $\mathbb{N} \in \mathcal{F}$, so $A \in \mathcal{F}'$. We will show that $\mathcal{F}' \in P$. First, for any $B \in \mathcal{F}$, we have $B \nsubseteq A^c$ (as $A^c \notin \mathcal{F}$) and so $A \cap B \neq \emptyset$, showing $\mathcal{F}'$ is proper. Next, if $X, Y \in \mathcal{F}' - \mathcal{F}$, where $A \cap B \subseteq X$ and $A \cap C \subseteq Y$, then $(A \cap B) \cap (A \cap C) = A \cap (B \cap C) \subseteq X \cap Y$, with $B \cap C \in \mathcal{F}$, so $X \cap Y \in \mathcal{F}'$. A similar proof shows that for any $X \in \mathcal{F}$ and $Y \in \mathcal{F}' - \mathcal{F}$, $X \cap Y = Y \cap X \in \mathcal{F}'$. Similarly, if $X \in \mathcal{F}' - \mathcal{F}$ and $X \subseteq Y$, where $A \cap B \subseteq X$, then $A \cap B \subseteq Y$ and so $Y \in \mathcal{F}'$. So $\mathcal{F}'$ is a proper filter, hence $\mathcal{F} \subset \mathcal{F}' \in P$, violating our assumption that $\mathcal{F}$ is maximal in $(P, \subseteq)$. So we have our contradiction and we are done. $\square$

Theorem 1.2. *Let $\mathcal{F}$ be an ultrafilter and $\{A_1, \dots, A_n\}$ a finite collection of pairwise disjoint sets such that*

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

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

Proof [1, Exercise 2.5(4)]. At most one of the A_i's can be in $\mathcal{F}$, since if $A_i, A_j \in \mathcal{F}$ when $i \neq j$ then we'd have $A_i \cap A_j = \emptyset \in \mathcal{F}$, a contradiction.

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

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

But then we find

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

a contradiction. $\square$

1.2 The Ultrapower

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

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

Theorem 1.3. *$\equiv$ is an equiavlence relation on $\mathbb{R}^{\mathbb{N}}$*

Proof [1, Exercise 3.3(1)]. Clearly $\equiv$ is reflexive, since $[[r_n = r_n]] = \mathbb{N} \in \mathcal{F}$ so $\langle r_n \rangle \equiv \langle r_n \rangle$.

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

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

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

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

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

One might reasonably worry that we chose no particular ultrafilter $\mathcal{F}$ on $\mathbb{N}$, and so ${}^*\mathbb{R}$ might not be unique. We can prove that ${}^*\mathbb{R}$ is unique if we assume the continuum hypothesis—otherwise, the uniqueness of ${}^*\mathbb{R}$ is indeterminate in ZFC [1, p. 33].

1.3 Extensions

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

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

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

since the right side is a finite intersection.

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

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

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

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

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

1.5 Internal Sets & Functions

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

Theorem 1.4. *Let $S \subseteq \mathbb{R}$. Then ${}^*S - \mathbb{R} \neq \emptyset$.*

Proof from [1, Theorem 3.9.1]. The theorem here is relying a bit on an abuse of notation, using $\mathbb{R}$ in the first case to mean the real numbers and in the second case to mean the "copy" of the real numbers in ${}^*\mathbb{R}$.

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

Theorem 1.5. *If A is finite, then ${}^*A = A$.*

Proof [1, Exercise 3.10(1)]. Let $[r_n] \in {}^*A$. Then $[[r_n \in A]] \in \mathcal{F}$, and we have

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

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

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

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

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

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

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

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

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

2 First Order Logic & Transfer

2.1 The Idea

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

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

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

2.3 Symbols

Our language will, formally, contain the following symbols:

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

2.4 Terms

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

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

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

2.5 Well-Formed Formulae

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

- If $\dot{R}_k$ is an k-ary relation symbol, and $t_1, \ldots, t_k$ are terms, then $\dot{R}_k(t_1, \ldots, t_k)$ is a wff. This type of wff is called an *atomic formula*.
- If φ and ψ are wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $\neg\varphi$.

- If φ is a wff and v_n is a variable, then $(\forall v_n)\varphi$ is a wff.

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

2.6 Bound and Free Variables

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

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

2.7 Replacement Notation

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

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

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

2.8 An Aside: Abbreviations

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

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

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

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

This can be read in English as follows.

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

Without abbreviations, this reads

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

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

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

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

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

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

- $\bar{s}(v_i) = s(v_i)$ for any variable v_i.
- $\bar{s}(\dot{x}) = x$ for any constant symbol $\dot{x}$.
- $\bar{s}(\dot{f}_k(t_1, \dots, t_k)) = f_k(\bar{s}(t_1), \dots, \bar{s}(t_k))$.

So if $s(v_1) = 3$ and $f(x) = x^2$, then $\bar{s}(\text{“}f(v_1 + 2)\text{”}) = 25$. Note $\bar{s}$ outputs the number 25, not $\dot{25}$ the symbol of $\mathcal{L}_{\mathbb{S}}$.

We say that $s : V \to S$ either satisfies or doesn't satisfy any wff φ, defined recursively by:

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

That last bullet deserves some attention. What we're saying is that s satisfies $(\forall v_n)\varphi$ if you can change $s(v_i)$ to be whatever you'd like and still satisfy φ. In other words, s satisfies $(\forall v_i)\varphi$ if for every $x \in S$, you can put $s'(v_i) = x$ and $s'(v_j) = s(v_j)$ for any $j \neq i$ and always have s' satisfying φ.

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

2.10 The Transfer Principle

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

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

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

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

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

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

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

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

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

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

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

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

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

□

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

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

Two final notes. The Transfer Principle is stated but not proved on [1, p. 44]. Also, Łoś's Theorem is actually much more general than what we've proved, working on any ultrapower, not just ${}^*\mathbb{R}$.

3 Structure of ${}^*\mathbb{R}$

The goal of this section is to explore what ${}^*\mathbb{R}$ is "like." To do this, we will use both work directly with the ultrapower definition of ${}^*\mathbb{R}$ and use the transfer principle.

To start with, all of the field axioms can be written down in our language $\mathcal{L}_{\mathfrak{R}}$. For instance, the fact that addition is commutative is

$$(\forall x)(\forall y)(x + y = y + x),$$

and the fact that multiplicative inverses exist is

$$(\forall x)(x \neq 0 \to (\exists y)(x \cdot y = 1)).$$

So, we transfer all of the field axioms to ${}^*\mathbb{R}$ and conclude ${}^*\mathbb{R}$ is a field. Throughout this paper we will make use of high-school-level algebraic properties to rearrange equations and such, even beyond the field axioms—unless otherwise noted, the relevant properties of ${}^*\mathbb{R}$ can be justified by a simple use of transfer.

3.1 Infinitesimal and Unlimited Hyperreals

Definition. A hyperreal $x \in {}^*\mathbb{R}$ is called

- *infinitesimal* if $|x| < b$ for any $b \in \mathbb{R}^+$.
- *unlimited* if $b < |x|$ for any $b \in \mathbb{R}^+$, and *limited* otherwise.
- *appreciable* if it is both limited and non-infinitesimal, i.e. if there are $b, c \in \mathbb{R}^+$ such that $b \leq |x| \leq c$.

There is exactly one real infinitesimal, 0. In ${}^*\mathbb{R}$, there are many more. Consider for a start $\delta = [\delta_n] = [\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots \rangle]$. For any $r \in \mathbb{R}^+$, there is some $m \in \mathbb{N}$ such that $\frac{1}{m} < r$. So the cofinite set $\{n \in \mathbb{N} \,|\, n \geq m\}$ is a subset of $[[\delta_n < r]]$, so $[[\delta_n < r]] \in \mathcal{F}$, and so $\delta < r$. However, $[[0 < \delta_n]] = \mathbb{N} \in \mathcal{F}$, and so $0 < \delta$. So $0 < \delta < r$ for any $r \in \mathbb{R}^+$, and so δ is infinitesimal.

Let $b \in \mathbb{R} - \{0\}$. Now, in $\mathbb{R}$, it's true that

$$(\forall x)(\forall r)\left(|x| < \frac{r}{|b|} \to |b \cdot x| < r\right).$$

We can then transfer this to ${}^*\mathbb{R}$. With our abbreviations, we would write the sentence the exact same way, so we can say that this sentence is also true in ${}^*\mathbb{R}$. Now, if it's true for all x, then it must be true for δ.

$$(\forall r)\left(|\delta| < \frac{r}{|b|} \to |b \cdot \delta| < r\right)$$

Let r be any positive real number. Then $\frac{r}{|b|} \in \mathbb{R}^+$, so $|\delta| < \frac{r}{|b|}$ (because δ is infinitesimal). Then, by applying our sentence, we know that $|b \cdot \delta| < r$. And this holds for any $r \in \mathbb{R}^+$, so $b \cdot \delta$ is infinitesimal. Similarly, we know that

$$(\forall x)(0 \neq x \to 0 \neq (b \cdot x))$$

in $\mathbb{R}$. So we transfer this ${}^*\mathbb{R}$, plug in δ for x, and since $0 \neq \delta$ we find $0 \neq (b \cdot x)$. Throughout this reasoning, we used the fact that δ was infinitesimal, but nothing else about it, so we conclude that if δ is any nonzero infinitesimal and $b \in \mathbb{R} - \{0\}$, then $b \cdot \delta$ is a nonzero infinitesimal. Generalizing, we can conclude (with a little bit more work) that for any limited hyperreal b and any infinitesimal δ, we have $b \cdot \delta$ infinitesimal [1, Chapter 5.2].

Now consider unlimited hyperreals. There are no unlimited reals. The hyperreal $H = [n] = [\langle 1, 2, 3, 4, \ldots \rangle]$ is unlimited, however. For any $r \in \mathbb{R}^+$, there is $m \in \mathbb{N}$ such that $r < n$ for all $n > m$, and so the cofinite set $[[r < n]] \in \mathcal{F}$ meaning $r < [n] = H$. This is true for every $r \in \mathbb{R}^+$, so H is unlimited.

Theorem 3.1 (Hyperreal Arithmetic [1, Chapter 5.2])**.** *If ϵ, δ are infinitesimal, b, c are appreciable, and H, K are unlimited, we have:*

- $-\epsilon$, $\epsilon + \delta$, $\epsilon \cdot \delta$, $\epsilon \cdot b$, *and* $\frac{1}{H}$ *infinitesimal.*
- $-b$, $b + \epsilon$, $b \cdot c$, *and* $\frac{1}{b}$ *appreciable.*
- $b + c$ *limited (possibly infinitesimal, as in* $(1 + \delta) + (-1) = \delta$*).*
- $-H$, $H + \epsilon$, $H + b$, $b \cdot H$, $H \cdot K$, *and* $\frac{1}{\epsilon}$ *(if* $\epsilon \neq 0$*) unlimited.*

3.2 Infinite Closeness & Standard Parts

Definition. We say that two hyperreals x, y are *infinitely close*, and write $x \simeq y$, if $x - y$ is infinitesimal.

Note that $\simeq$ is an equivalence relation. $x - x = 0$ is infinitesimal, $x - y$ is infinitesimal iff $-(x - y) = y - x$ is, and whenever $x - y$ and $y - z$ are infinitesimal their sum $(x - y) + (y - z) = x - z$ is as well. Note also that x is infinitesimal iff $x \simeq 0$, and $x \simeq y$ iff $|x - y| < r$ for any $r \in \mathbb{R}^+$.

Theorem 3.2 ([1, Theorem 5.6.1])**.** *Every limited hyperreal b is infinitely close to exactly one real number, called the standard part of b, denoted by* $\operatorname{st}(b)$*.*

So the limited hyperreals are broken up into disjoint "halos" around each real number. The standard part function also has the nice property of respecting most arithmetic functions. For instance, $\operatorname{st}(b + c) \simeq b + c$ (by definition), and $b + c - (\operatorname{st}(b) + \operatorname{st}(c)) = (b - \operatorname{st}(b)) + (c - \operatorname{st}(c))$, which is infinitesimal (as the sum of two infinitesimals). So $b + c \simeq \operatorname{st}(b) + \operatorname{st}(c)$, and so by transitivity of $\simeq$ we have $\operatorname{st}(b + c) \simeq \operatorname{st}(b) + \operatorname{st}(c)$. Since both $\operatorname{st}(b + c)$ and $\operatorname{st}(b) + \operatorname{st}(c)$ are real, their difference $\operatorname{st}(b + c) - (\operatorname{st}(b) + \operatorname{st}(c))$ must be real, and since they are infinitely close, their difference must be infinitesimal. Since the only real infinitesimal is 0, that means $\operatorname{st}(b + c) = \operatorname{st}(b) + \operatorname{st}(c)$.

That last move will be used frequently enough that it is worth stating as a theorem.

Theorem 3.3 ([1])**.** *If $a, b \in \mathbb{R}$, and $a \simeq b$, then $a = b$.*

Also, since $\simeq$ is transitive, if $x, y \in {}^*\mathbb{R}$ are limited and $x \simeq y$, we have $\operatorname{st}(x) \simeq x \simeq y \simeq \operatorname{st}(y)$ and therefore $\operatorname{st}(x) = \operatorname{st}(y)$. Similarly, if $\operatorname{st}(x) = \operatorname{st}(y)$, then $x \simeq \operatorname{st}(x) = \operatorname{st}(y) \simeq y$ and so $x \simeq y$.

Corollary 3.4 ([1])**.** *If $x, y \in {}^*\mathbb{R}$ are both limited, then $x \simeq y$ if and only if* $\operatorname{st}(x) = \operatorname{st}(y)$*.*

We can write similar proofs that st distributes over subtraction, multiplication, division, even absolute value ($\operatorname{st}(|x|) = |\operatorname{st}(x)|$) for limited values [1, Theorem 5.6.2].

Theorem 3.5. *Let $\mathbb{L}$ denote the ring of limited hyperreals, and $\mathbb{I}$ the ring of infinitesimals. Show that $\mathbb{R}$ is isomorphic to the ring of limited hyperrationals ${}^*\mathbb{Q} \cap \mathbb{L}$ factored by its ideal ${}^*\mathbb{Q} \cap \mathbb{I}$.*

Proof [1, Exercise 5.7(4)]. Consider $\operatorname{st} : {}^*\mathbb{Q} \cap \mathbb{L} \to \mathbb{R}$. We want to use the first isomorphism theorem on rings, so we need to show that st is a ring homomorphism, its kernel is ${}^*\mathbb{Q} \cap \mathbb{I}$, and its image is $\mathbb{R}$. That st is a ring homomorphism follows from the facts that $\operatorname{st}(a \pm b) = \operatorname{st}(a) \pm \operatorname{st}(b)$ and $\operatorname{st}(a \cdot b) = \operatorname{st}(a) \cdot \operatorname{st}(b)$ for any $a, b \in {}^*\mathbb{Q} \cap \mathbb{L}$ because such a, b are limited. $\operatorname{st}(a) = 0$ iff $a \simeq 0$ iff a is infinitesimal iff $a \in {}^*\mathbb{Q} \cap \mathbb{I}$, so we conclude the kernel of st is ${}^*\mathbb{Q} \cap \mathbb{I}$.

Finally, take any $x \in \mathbb{R}$. We have, by the density of $\mathbb{Q}$ in $\mathbb{R}$,

$$(\forall r \in \mathbb{R}^+)(\exists q \in \mathbb{Q})(|x - q| < r).$$

If we transfer this, we get

$$(\forall r \in {}^*\mathbb{R}^+)(\exists q \in {}^*\mathbb{Q})(|x - q| < r). \tag{$\star$}$$

What is ${}^*\mathbb{R}^+$? Well $(\forall y)(y \in \mathbb{R}^+ \leftrightarrow 0 < y)$, so by transfer $(\forall y)(y \in {}^*\mathbb{R}^+ \leftrightarrow 0 < y)$. So ${}^*\mathbb{R}^+$ is the set of positive hyperreals. Take $\star$ and plug in some positive infinitesimal δ for r. Then we have

$$(\exists q \in {}^*\mathbb{Q})(|x - q| < \delta).$$

But if $|x - q| < \delta$, then $|x - q| < r$ for any $r \in \mathbb{R}^+$ (since δ is infinitesimal), and so $x \simeq q$. Since $x \in \mathbb{R}$, this implies $\operatorname{st}(q) = x$, and thus that x is in the image of st.

This is true for any $x \in \mathbb{R}$, so the image of st must be all of $\mathbb{R}$. Thus, we can use the first isomorphism theorem on rings to conclude ${}^*\mathbb{Q} \cap \mathbb{L} / {}^*\mathbb{Q} \cap \mathbb{I} \cong \mathbb{R}$. $\square$

3.3 Internal Structure of ${}^*\mathbb{R}$

If we want to know about the subsets of ${}^*\mathbb{R}$, we might hope that we can generalize Theorem 2.1 to even more powerful logics, such as second-order logic, adding quantifiers on sets to get sentences like $(\forall A)(A \neq \emptyset \to (\exists x)(x \in A))$. This is impossible. $\mathbb{R}$ is Dedekind complete—every set that has an upper bound has a least upper bound. With second-order quantification, we can write

$$(\forall A)\big((\exists x)(\forall y \in A)(y \leq x) \to (\exists x)[(\forall y \in A)(y \leq x) \wedge (\forall z)(z < x \to (\exists y \in A)(z < y))]\big).$$

The transfer of this is not true in ${}^*\mathbb{R}$. The following theorem is stated on [1, p. 140], but proven differently.

Theorem 3.6. *The set $\mathbb{I}$ of infinitesimals has an upper bound, but no least upper bound.*

Proof. Any $r \in \mathbb{R}^+$ is an upper bound on $\mathbb{I}$. Assume for contradiction L is the least upper bound of $\mathbb{I}$. Surely $0 < L$, as there are infinitesimals greater than 0. If L is appreciable (or unlimited), then $\frac{L}{2}$ is appreciable (or unlimited), meaning it is an upper bound for $\mathbb{I}$ and thus that L is not the least upper bound, contradiction. If L is infinitesimal, though, then $2L$ is an infinitesimal with $L < 2L$, so L is not an upper bound on $\mathbb{I}$, contradiction. So L cannot be infinitesimal or non-infinitesimal, a contradiction. □

So, $\mathbb{R}$ is Dedekind complete and ${}^*\mathbb{R}$ is not. We do get the following theorem, stated on [1, p. 133] but not proven, as an easy consequence of Theorem 2.1.

Theorem 3.7. *If φ is a sentence of $\mathcal{L}_{\mathfrak{R}}$ and $\varphi(A)$ for every $A \subseteq \mathbb{R}$, then ${}^*\varphi(X)$ for every internal subset $X \subseteq {}^*\mathbb{R}$.*

So, for any *internal* subset $X \subseteq {}^*\mathbb{R}$, if X has an upper bound, it has a least upper bound. This is proven as [1, Theorem 11.5.1] without using the above theorem. Goldblatt proves the above theorem in [1, Theorem 13.12.1], although he's thinking about *nonstandard frameworks*, a generalization of what we're doing here. Intuitively, the idea is that we're making a mistake when we try to transfer second-order sentences. In first order logic, if we write $(\forall v)\varphi$, we mean the same thing as $(\forall v \in \mathbb{R})\varphi$, and the appropriate transfer is $(\forall v \in {}^*\mathbb{R})\varphi$. Similarly, if we write $(\forall A)\varphi$, this is equivalent to $(\forall A \in \mathcal{P}(\mathbb{R}))\varphi$, which transfers as $(\forall A \in {}^*\mathcal{P}(\mathbb{R}))\varphi$, and ${}^*\mathcal{P}(\mathbb{R}) \neq \mathcal{P}({}^*\mathbb{R})$. Intuitively, the $*$-transform of $\mathcal{P}(\mathbb{R})$ should be all the sets $A = [A_n]$ such that $[[A_n \in \mathcal{P}(\mathbb{R})]] \in \mathcal{F}$, the internal sets. This is a reckless extension of concepts into undefined territory, and simplified to the point of falsehood—the rigorous details can be found in [1, Chapter 13], specifically Theorem 13.12.1.

Theorem 3.8. *[1, Exercise 12.2(1)] If $f = [f_n]$ is an internal function, and $A = [A_n]$ is any internal subset of the domain* $\operatorname{dom} f$*, then the image set $f(A) = \{f(a) \mid a \in A\}$ is the internal set $[f_n(A_n)]$*

Proof. Say $y \in f(A)$. Then there exists $x \in A$ such that $f(x) = y$. Say $x = [x_n]$. Then $y = f(x) = [f_n(x_n)]$, i.e. $[[y = f_n(x_n)]] \in \mathcal{F}$. Since $f_n(x_n) \in f_n(A_n)$ for every n, this means $[[y \in f_n(A_n)]] \in \mathcal{F}$, i.e. $y \in [f_n(A_n)]$.

Conversely, if $y \in [f_n(A_n)]$, then $[[y \in f_n(A_n)]] \in \mathcal{F}$. For each $n \in [[y \in f_n(A_n)]]$, there is an $x_n \in A_n$ such that $y_n = f_n(x_n)$, so we can define a sequence with all the x_n's (taking any value we'd like where $n \notin [[y \in f_n(A_n)]]$) and get $[[y_n = f_n(x_n)]] \in \mathcal{F}$, i.e. $y = f([x_n])$. Since $[[x_n \in A_n]] \in \mathcal{F}$, this means $[x_n] \in A$ and $y \in f(A)$. □

Theorem 3.9 ([1, Exercise 12.2(7)])**.** *Let f be an internal function such that $f(x) \simeq 0$ whenever f is defined. Show that the range of of f has an infinitesimal least upper bound.*

Proof. First, we will show that the range of f is internal, and then we will apply the Dedekind completeness. $\operatorname{dom} f$ is defined as $[\operatorname{dom} f_n]$, and so is internal. So f is an internal function and $\operatorname{dom} f$ is an internal set, so the range $f(\operatorname{dom} f)$ is internal by Theorem 3.8. Now, $f(x) \simeq 0$ for all $x \in \operatorname{dom} f$, so 1 is an upper bound for the range of f. Hence, by Dedekind completeness, there is a least upper bound L on the range of f. If L were appreciable (or unlimited), then it would be positive (as any negative appreciable number is less than every infinitesimal) and $\frac{L}{2}$ would be a lesser upper bound, contradicting L being the least upper bound. Hence L is infinitesimal. □

A particularly weird consequence of this is that any countable sequence $s^1, s^2, \ldots$ of infinitesimals has an infinitesimal upper bound, despite there being no largest infinitesimal. If each $s^k = [s_n^k]$, we can define the internal function $f : {}^*\mathbb{N} \to {}^*\mathbb{R}$ by $f = [f_n]$ where $f_n(k) = s_n^k$ for any $k \in \mathbb{N}$. Then $f(k) = [f_n(k)] = [s_n^k] = s^k$. So by the theorem, the range of f has an infinitesimal upper bound, which is also a bound for the sequence $s^k = f(\mathbb{N})$.

4 Differentiation

4.1 Definition

One of the primary motivations for infinitesimal calculus is that it allows us to access more intuitive, less roundabout conceptualizations of derivatives and integrals. Unlike integrals, which still take some work to define nonstandardly, the nonstandard derivative is almost exactly what we would first guess it to be.

Let Δx be a nonzero infinitesimal, and let $\Delta f(x, \Delta x) = f(x + \Delta x) - f(x)$. Intuitively, Δx is an infinitesimal change in x, and Δf is the corresponding change in f caused by "moving" Δx along the x-axis. Then we have:

$$f'(x) = \frac{\Delta f(x, \Delta x)}{\Delta x}.$$

One small problem: we'd like f' to be a real-valued function on the reals. Luckily, we have a tool to do that:

$$f'(x) = \mathrm{st}\left(\frac{\Delta f(x, \Delta x)}{\Delta x}\right).$$

And we have our definition. Well, this might also not be well-defined: we get around *that* problem by definition. We only consider $f'(x)$ to exist when $\mathrm{st}\left(\frac{\Delta f(x,\Delta x)}{\Delta x}\right)$ is the same for *any* nonzero infinitesimal Δx. In this case, we say that f is *differentiable* at x. Note that this definition of differentiability and the derivative is equivalent to the standard definition.

Theorem 4.1 ([1, Theorem 8.1.1])**.** *If f is defined at $x \in \mathbb{R}$, then $\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = L$ iff for every nonzero infinitesimal ϵ, $f(x+\epsilon)$ is defined and $\frac{f(x+\epsilon) - f(x)}{\epsilon} \simeq L$.*

Definition. Given a function $f : \mathbb{R} \to \mathbb{R}$ and a real number x, we say that f is *differentiable* at x if there is some constant $f'(x)$ such that, for any nonzero infinitesimal Δx,

$$f'(x) = \mathrm{st}\left(\frac{f(x + \Delta x) - f(x)}{\Delta x}\right).$$

4.2 Simple Proofs Using Infinitesimals

A number of proofs of basic calculus results can be easily accomplished by infinitesimals. Perhaps most striking is the chain rule.

Theorem 4.2 (Chain Rule)**.** *Given differentiable functions $f, g : \mathbb{R} \to \mathbb{R}$,*

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x).$$

Proof adapted from [1, Chapter 8.4]. Let Δx be any nonzero infinitesimal, and let $\Delta g = g(x+\Delta x) - g(x)$. If $\Delta g = 0$, then $g(x + \Delta x) = g(x)$, so $(f \circ g)'(x) = \mathrm{st}\left(\frac{f(g(x + \Delta x)) - f(g(x))}{\Delta x}\right) = 0$ and clearly $g'(x) = \mathrm{st}\left(\frac{\Delta g}{\Delta x}\right) = 0$, so we are done. If $\Delta g \neq 0$, then since $g'(x) = \mathrm{st}\left(\frac{\Delta g}{\Delta x}\right)$ is defined, we conclude $\frac{\Delta g}{\Delta x}$ is limited and (since Δx is infinitesimal) that Δg is infinitesimal. Thus

$$\begin{aligned}
(f \circ g)'(x) &= \mathrm{st}\left(\frac{f(g(x + \Delta x)) - f(g(x))}{\Delta x}\right) \\
&= \mathrm{st}\left(\frac{f(g(x) + \Delta g) - f(g(x))}{\Delta g} \cdot \frac{\Delta g}{\Delta x}\right) \\
&= \mathrm{st}\left(\frac{f(g(x) + \Delta g) - f(g(x))}{\Delta g}\right) \cdot \mathrm{st}\left(\frac{\Delta g}{\Delta x}\right) \\
&= f'(g(x)) \cdot g'(x).
\end{aligned}$$

□

4.3 Partial Derivatives

Definition. If $f : \mathbb{R}^n \to \mathbb{R}$, then we define

$$f_{x_k}(b_1, b_2, \ldots, b_n) = \mathrm{st}\left(\frac{f(b_1, \ldots, b_k + \Delta x, \ldots, b_n) - f(b_1, \ldots, b_k, \ldots, b_n)}{\Delta x}\right)$$

for any infinitesimal Δx. This is only defined when it doesn't depend on our choice of Δx.

This gets complicated when we want to deal with repeated partial derivatives. Say $f : \mathbb{R}^2 \to \mathbb{R}$, and denote the inputs of f by $f(x, y)$. Then we might want to write

$$\begin{aligned}
f_{yx}(a, b) &= \mathrm{st}\left(\frac{f_y(a + \Delta x, b) - f_y(a, b)}{\Delta x}\right) \\
&= \mathrm{st}\left(\frac{\mathrm{st}\left(\frac{f(a+\Delta x, b+\Delta y) - f(a+\Delta x, b)}{\Delta y}\right) - \mathrm{st}\left(\frac{f(a, b+\Delta y) - f(a, b)}{\Delta y}\right)}{\Delta x}\right),
\end{aligned}$$

which would allow us to easily get $f_{yx}(a,b) = f_{xy}(a,b)$ after more algebra. However, this isn't right. Firstly, the numerator is the difference of two real numbers, and so is real, which would mean $f_{yx}(a,b)$ is either 0 or undefined (as a real divided by an infinitesimal is either 0 or unlimited). The mistake here is that $f_y(a+\Delta x, b) \neq \mathrm{st}\left(\frac{f(a+\Delta x, b+\Delta y)-f(a+\Delta x)}{\Delta y}\right)$. Since in this case f_y is taking a nonreal input, we have to use the extension ${}^*f_y(a+\Delta x, b)$. If $\Delta x = [\Delta x_n]$, then this is $[f_y(a+\Delta x_n, b)] = \left[\mathrm{st}\left(\frac{f(a+\Delta x_n, b+\Delta y)-f(a+\Delta x_n, b)}{\Delta y}\right)\right]$. This is the equivalence class of a sequence of real numbers, but it needn't be a real number itself.

Theorem 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Let $\hat{f}_{x_k}$ denote the standard definition of the partial derivative, namely*

$$\hat{f}_{x_k}(b_1, \dots, b_n) = \lim_{h \to 0} \frac{f(b_1, \dots, b_k + h, \dots, b_n) - f(b_1, \dots, b_k, \dots, b_n)}{h}.$$

Then $\hat{f}_{x_k}(b_1, \dots, b_n)$ exists iff $f_{x_k}(b_1, \dots, b_n)$ does. If they both exist, then $\hat{f}_{x_k}(b_1, \dots, b_n) = f_{x_k}(b_1, \dots, b_n)$.

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = f(b_1, \dots, b_{k-1}, x, b_{k+1}, \dots, b_n)$. Let $\hat{g}'(x)$ denote the standard derivative of g and $g'(x)$ denote the nonstandard derivative of g. Clearly $\hat{g}'(x) = \hat{f}_{x_k}(x)$ and $g'(x) = f_{x_k}(x)$, and by Theorem 4.1 $\hat{g}'(b_k)$ is defined iff $g'(b_k)$, and when both exist $\hat{g}'(b_k) = g'(b_k)$, so we are done. □

4.4 Series

Let $a_n : \mathbb{N} \to \mathbb{R}$. We can define $s_n : \mathbb{N} \to \mathbb{R}$ by

$$s_n = \sum_{i=0}^{n} a_i.$$

Now, for any unlimited $N \in {}^*\mathbb{N}$, we can define

$$\sum_{i=0}^{N} a_i = {}^*s_N.$$

Finally, if for any unlimited $N, M \in {}^*\mathbb{N}$ we have $\sum_{i=0}^{N} a_i \simeq \sum_{i=0}^{M} a_i$ with both limited, then we say $\sum_{i=0}^{\infty} a_i$ *converges* and define

$$\sum_{i=0}^{\infty} a_i = \mathrm{st}\left(\sum_{i=0}^{N} a_i\right).$$

When $n > m$, we write $\sum_{i=m}^{n} a_i$ to mean $\sum_{i=0}^{n} a_i - \sum_{i=0}^{m-1} a_i$. This gives the expected result when using finite naturals, but also extends to unlimited hypernaturals. Then $\sum_{i=0}^{\infty} a_i$ converges when $\sum_{i=M+1}^{N} a_i \simeq 0$ for any unlimited $N, M \in {}^*\mathbb{N}$ such that $N > M$.

Theorem 4.4 (Cauchy Convergence Criterion for Series)**.** *If for any unlimited $N, M \in {}^*\mathbb{N}$, $\sum_{i=0}^{N} a_i \simeq \sum_{i=0}^{M} a_i$, then $\sum_{i=0}^{\infty} a_i$ converges.*

Proof adapted from [1, Theorem 6.5.2]. The issue here is that $\sum_{i=0}^{N} a_i$ might be unlimited for all unlimited $N \in {}^*\mathbb{N}$. Let $s(n) = \sum_{i=0}^{N} a_i$.

First, assume that (in $\mathbb{R}$) $\neg(\exists r \in \mathbb{R})(\forall n \in \mathbb{N})(|s(n)| < r)$. This says the sequence $s(n)$ is not bounded. By transfer, this is true in ${}^*\mathbb{R}$ as well. Take an unlimited $N \in {}^*\mathbb{N}$. Since $|s(n)|$ has no upper bound in ${}^*\mathbb{R}$, there is some $M \in {}^*\mathbb{N}$ such that $|s(N)| \leq 2 \cdot |s(N)| < |s(M)|$. Since $|s(N)| \simeq |s(M)|$, this implies $|s(N)| \simeq 2 \cdot |s(N)|$, implying $|s(N)| \simeq 0$, and so for any unlimited $M \in {}^*\mathbb{N}$ we have $|s(M)| \simeq |s(N)| \simeq 0$ and so $\sum_{i=0}^{\infty} a_i = 0$.

Now, say $(\exists r \in \mathbb{R})(\forall n \in \mathbb{N})(|s(n)| < r)$. Whatever this r is, call it b. Then $(\forall n \in \mathbb{N})(|s(n)| < b)$, and so by transfer $(\forall n \in {}^*\mathbb{N})(|s(n)| < b)$. So if we take an unlimited $N \in {}^*\mathbb{N}$, we know $s(N)$ is bounded and hence $\sum_{i=0}^{\infty} a_i = \mathrm{st}\,(s(N))$. □

Theorem 4.5 (Geometric Series)**.** *Let $0 < r < 1$. Then*

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}.$$

Proof [1, Exercise 6.7]. By difference of powers, we know that $1 - r^{n+1} = (1 + r + r^2 + \cdots + r^n)(1 - r)$. So, we conclude that $\sum_{i=0}^{n} r^i = \dfrac{1 - r^{n+1}}{1 - r}$. Hence, by transfer, for any unlimited $N \in {}^*\mathbb{N}$ we have $\sum_{i=0}^{N} r^i = \dfrac{1 - r^{N+1}}{1 - r}$.

Consider r^N. Assume for a contradiction $\mathrm{st}\left(r^N\right) > 0$. Then as $0 < r < 1$, we have $r^{N+1} = r \cdot r^N \simeq r \cdot \mathrm{st}\left(r^N\right) < \mathrm{st}\left(r^N\right)$. Hence, in ${}^*\mathbb{R}$, $(\exists M \in {}^*\mathbb{N})(r^M < \mathrm{st}\left(r^N\right))$. By transfer, $(\exists m \in \mathbb{N})(r^m < \mathrm{st}\left(r^N\right))$. But r^m is real, so $r^m < r^N$, even though $0 < r < 1$ and $m < N$. However, $\frac{r^N}{r^m} = r^{N-m}$ (by a simple transfer), and since $N - m > 0$ we find $r^{N-m} < 1$. So $\frac{r^N}{r^m} < 1$, so $r^N < r^m$. This is a contradiction, so we find $\mathrm{st}\left(r^N\right) \leq 0$, and since $0 < r < 1$ we conclude $\mathrm{st}\left(r^N\right) = 0$. Hence,

$$\sum_{i=0}^{\infty} r^i = \mathrm{st}\left(\sum_{i=0}^{N} r^i\right) = \mathrm{st}\left(\frac{1 - r^{N+1}}{1 - r}\right) = \frac{\mathrm{st}\,(1) - \mathrm{st}\left(r^{N+1}\right)}{\mathrm{st}\,(1 - r)} = \frac{1}{1 - r}.$$

$\square$

Theorem 4.6 (Absolute Convergence Implies Convergence)**.** *If $\sum_{i=0}^{\infty} |a_i|$ converges, then $\sum_{i=0}^{\infty} a_i$ converges.*

Proof. Take any two unlimited $N, M \in {}^*\mathbb{N}$, say $N > M$. In $\mathbb{R}$, the triangle inequality implies

$$(\forall n \in \mathbb{N})(\forall m \in \mathbb{N})\left(n > m \to \left|\sum_{i=m+1}^{n} a_i\right| \leq \sum_{i=m+1}^{n} |a_i|\right),$$

and so this is true in ${}^*\mathbb{R}$ too. So, we have

$$0 < \left|\sum_{i=M+1}^{N} a_i\right| \leq \sum_{i=M+1}^{N} |a_i| \simeq 0,$$

since $\sum_{i=0}^{\infty} |a_i|$ converges. This implies $\sum_{i=M+1}^{N} a_i \simeq 0$, so $\sum_{i=0}^{\infty} a_i$ converges. $\square$

Theorem 4.7 (Ratio Test)**.** *Let $a_i : \mathbb{N} \to \mathbb{R}$ be a sequence. If for every unlimited $M \in {}^*\mathbb{N}$ we have $\left|\mathrm{st}\left(\frac{a_{M+1}}{a_M}\right)\right| = L$ for some $L < 1$, then $\sum_{i=0}^{\infty} a_i$ converges.*

Proof [1, Exercise 6.8]. We wil prove the theorem assuming that $a_i \geq 0$. For other sequences, we can then apply our newly-proven theorem to $|a_i|$ and use Theorem 4.6. Now, take $r \in \mathbb{R}$ such that $L < r < 1$, so that $\left|\frac{a_{M+1}}{a_M}\right| < r < 1$ for any unlimited hypernatural M.

Take any unlimited $N \in {}^*\mathbb{N}$. For any $M \in {}^*\mathbb{N}$, if $N \leq M$, then M is also unlimited, and so $\frac{a_{M+1}}{a_M} < r$. So we have the sentence

$$(\exists k \in {}^*\mathbb{N})(\forall m \in {}^*\mathbb{N})\left(k \leq m \to \frac{a_{m+1}}{a_m} < r\right).$$

If we transfer this to the reals, we find a $k \in \mathbb{N}$ such that for any $m \geq k$ we have $\frac{a_{m+1}}{a_m} < r$. We will show by a simple induction that for any $n \in \mathbb{N}$, $a_{k+n} \leq r^n a_k$. As our base case, $a_k = r^0 a_k$. Next, if $a_{k+n} \leq r^n a_k$, then since $k + n \geq k$ we know $\dfrac{a_{k+n+1}}{a_{k+n}} < r$ and so $a_{k+n+1} < r \cdot a_{k+n} < r \cdot r^n a_k = r^{n+1} a_k$.

Now, in the reals, for any $n, m \in \mathbb{N}$ with $n \geq m$, we have

$$\sum_{i=k+m}^{k+n} a_i \leq \sum_{i=k+m}^{k+n} r^{i-k} a_k = \left(\sum_{i=m}^{n} r^i\right) a_k.$$

Transferring this to the hyperreals, if we have two unlimited $N, M \in {}^*\mathbb{N}$ with $N > M$, we get

$$\sum_{i=M+1}^{N} a_i = \sum_{i=k+(M+1-k)}^{k+(N-k)} a_i \ \leq \left(\sum_{i=M+1-k}^{N-k} r^i\right) \cdot a_k.$$

Note that $M + 1 - k$ and $N - k$ are both unlimited hyperreals, and so $\sum_{i=M+1-k}^{N-k} r^i$ is infinitesimal by Theorem 4.5 (since $\sum_{i=0}^{N-k} r^i \simeq \sum_{i=0}^{M+1-k} \simeq \frac{1}{1-r}$). Since a_k is appreciable, this means the product is infinitesimal, hence $\sum_{i=M+1}^{N} a_i$ is infinitesimal, hence $\sum_{i=0}^{N} a_i - \sum_{i=0}^{M} a_i$ is infinitesimal, hence $\sum_{i=0}^{N} a_i \simeq \sum_{i=0}^{M} a_i$. So by Theorem 4.4$\sum_{i=0}^{\infty} a_i$ converges. $\square$

4.5 The exp function

This entire section is individual work without the guidance of any texts. We define the exponential function

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}.$$

For convenience, we will write $e(x,k) = \sum_{i=0}^{k} \frac{x^i}{i!}$. For any unlimited $N \in {}^*\mathbb{N}$, we have $\exp(x) = \mathrm{st}\,(e(x,N))$.

Theorem 4.8. *For any real x, $\exp(x)$ exists.*

Proof. We use the Ratio Test (Theorem 4.7) to prove that $\sum_{i=0}^{\infty} \frac{x^i}{i!}$ converges. Let M be an unlimited hypernatural. Then

$$\frac{x^{M+1}}{(M+1)!} \div \frac{x^M}{M!} = \frac{x^{M+1}M!}{x^M(M+1)!} = \frac{x}{M+1}.$$

Since x is real and $M+1$ is unlimited, $\frac{x}{M+1}$ is infinitesimal, and hence has standard part 0. Since $0 < 1$, and since this holds for any unlimited hypernnatural, the conditions of Theorem 4.7 are met and we are done. $\square$

We now want to prove that $\exp'(x) = \exp(x)$. This will involve the following lemma.

Lemma 4.9. *Let $R_k = e(x+d,k) - e(x,k) - d \cdot e(x,k-1)$ for some $x, d \in \mathbb{R}$ with $|d| < 1$ and $k \in \mathbb{N}$. Then*

$$|R_k| < \frac{|d|^2}{1-|d|} \cdot \exp(|x|).$$

Proof. First, we will find a more explicit formula for R_k. We have

$$e(x+d,k) = 1 + (x+d) + \frac{x^2+2xd+d^2}{2!} + \cdots + \frac{x^k + kx^{k-1}d + \cdots + d^k}{k!},$$

and so

$$\begin{aligned} e(x+d,k) - e(x,k) &= d + \frac{2xd+d^2}{2!} + \frac{3x^2d+3xd^2+d^3}{3!} + \cdots + \frac{kx^{k-1}d + \cdots + d^k}{k!} \\ &= d\left(1 + \frac{2x}{2!} + \frac{3x^2}{3!} + \cdots + \frac{kx^{k-1}}{k!}\right) + \frac{d^2}{2!} + \frac{3xd^2+d^3}{3!} + \cdots + \frac{\binom{k}{2}x^{k-2}d^2 + \cdots + d^k}{k!} \\ &= d \cdot e(x,k-1) + \frac{d^2}{2!} + \frac{3xd^2+d^3}{3!} + \cdots + \frac{\binom{k}{2}x^{k-2}d^2 + \cdots + d^k}{k!}. \end{aligned}$$

So we find that

$$\begin{aligned} R_k &= \frac{d^2}{2!} + \frac{3xd^2+d^3}{3!} + \cdots + \frac{\binom{k}{2}x^{k-2}d^2 + \cdots + d^k}{k!} \\ &= \sum_{i=1}^{k} \sum_{q=0}^{i-2} \binom{i}{q} \frac{x^q d^{(i-q)}}{i!}. \end{aligned}$$

Using the fact that $\displaystyle\binom{i}{q} = \frac{i!}{q!(i-q!)}$, we have

$$\begin{aligned} R_k &= \sum_{i=2}^{k} \sum_{q=0}^{i-2} \binom{i}{q} \frac{x^q d^{(i-q)}}{i!} \\ &= \sum_{i=2}^{k} \sum_{q=0}^{i-2} \frac{x^q d^{(i-q)}}{(i-q)!q!} \\ &= \sum_{q=0}^{k-2} \sum_{i=q+2}^{k} \frac{x^q d^{(i-q)}}{(i-q)!q!}. \end{aligned}$$

So

$$|R_k| = \sum_{q=0}^{k-2} \sum_{i=q+2}^{k} \frac{|x|^q \cdot |d|^{(i-q)}}{(i-q)!q!}.$$

Now, $i - q \geq 2$ in every term, and so $(i-q)! \geq 2! \geq 1$ and so $\dfrac{|x|^q \cdot |d|^{(i-q)}}{(i-q)!q!} \leq \dfrac{|x|^q \cdot |d|^{(i-q)}}{q!}$. Combining this move with a change of index, setting $p = i - q$, we get

$$|R_k| \leq \sum_{q=0}^{k-2} \sum_{i=q+2}^{k} \frac{|x|^q \cdot |d|^{(i-q)}}{q!} = \sum_{q=0}^{k-2} \sum_{p=2}^{k-q} \frac{|x|^q \cdot |d|^p}{q!} = \sum_{q=0}^{k-2} \left(\frac{|x|^q}{q!} \cdot \sum_{p=2}^{k-q} |d|^p \right).$$

Since $|d| < 1$, Theorem 4.5 tells us that for any unlimited hypernatural N, $\sum_{p=0}^{N} |d|^p \simeq \frac{1}{1-|d|}$. Since each term of $|d|^p$ is positive, we know $\sum_{p=0}^{k-q} |d|^p < \sum_{p=0}^{N} |d|^p$, so

$$\sum_{p=2}^{k-q} |d|^p = \sum_{p=0}^{k-q} |d|^p - \sum_{p=0}^{1} |d|^p \leq \frac{1}{1-|d|} - (1+|d|) = \frac{1 - 1(1-|d|) - |d|(1-|d|)}{1-|d|} = \frac{1 - 1 + |d| - |d| + |d|^2}{1-|d|} = \frac{|d|^2}{1-|d|}.$$

Then

$$|R_k| \leq \sum_{q=0}^{k-2} \left(\frac{|x|^q}{q!} \cdot \frac{|d|^2}{1-|d|} \right) = \frac{|d|^2}{1-|d|} \cdot \sum_{q=0}^{k-2} \frac{|x|^q}{q!} \leq \frac{|d|^2}{1-|d|} \cdot \exp(|x|).$$

□

Theorem 4.10. *For any real x, $\exp'(x) = \exp(x)$.*

Proof. By Lemma 4.9, for any $k \in \mathbb{N}$ and $x, d \in \mathbb{R}$ with $0 < |d| < 1$, we have

$$|e(x+d,k) - e(x,k) - d \cdot e(x,k-1)| \leq \frac{|d|^2}{1-|d|} \cdot \exp(|x|),$$

and so

$$(\forall k \in \mathbb{N}) \left(\left| \frac{e(x+d,k) - e(x,k)}{d} - e(x,k-1) \right| \leq \frac{|d|}{1-|d|} \cdot \exp(|x|) \right).$$

We then transfer this statement to ${}^*\mathbb{R}$, pick an unlimited $N \in {}^*\mathbb{N}$ to plug in for k, and we get

$$\left| \frac{e(x+d,N) - e(x,N)}{d} - e(x,N-1) \right| \leq \frac{|d|}{1-|d|} \cdot \exp(|x|).$$

Taking the standard part of both sides (which does nothing to the right side since it's real), we get

$$\begin{aligned} \operatorname{st}\left(\left| \frac{e(x+d,N) - e(x,N)}{d} - e(x,N-1) \right| \right) &= \left| \frac{\operatorname{st}\left(e(x+d,N)\right) - \operatorname{st}\left(e(x,N)\right)}{d} - \operatorname{st}\left(e(x,N-1)\right) \right| \\ &= \left| \frac{\exp(x+d) - \exp(x)}{d} - \exp(x) \right| \leq \frac{|d|}{1-|d|} \cdot \exp(|x|). \end{aligned}$$

So we know the following sentence is true in $\mathbb{R}$

$$(\forall d \in \mathbb{R}) \left(\left(|d| < 1 \wedge d \neq 0\right) \to \left| \frac{\exp(x+d) - \exp(x)}{d} - \exp(x) \right| \leq \frac{|d|}{1-|d|} \cdot \exp(|x|) \right).$$

Now, transfer this to the hyperreals and plug in any nonzero infinitesimal δ in place of d. Now we know

$$\left| \frac{\exp(x+\delta) - \exp(x)}{\delta} - \exp(x) \right| \leq \frac{|\delta|}{1-|\delta|} \cdot \exp(|x|).$$

The right side of this is infinitesimal, since $|\delta|$ is infinitesimal while $1 - |\delta|$ and $\exp(|x|)$ are appreciable (the latter by Theorem 4.8). Hence $\left| \frac{\exp(x+\delta)-\exp(x)}{\delta} - \exp(x) \right|$ is infinitesimal. Hence $\dfrac{\exp(x+\delta) - \exp(x)}{\delta} \simeq \exp(x)$. Since $\exp(x)$ is real, this implies

$$\operatorname{st}\left(\frac{\exp(x+\delta) - \exp(x)}{\delta} \right) = \exp'(x) = \exp(x).$$

□

5 Integration

5.1 Integrability & Integrals

If f is a function bounded on $[a,b]$ and $P = \{x_0 = a, x_1, \ldots, x_n = b\}$ a partition on $[a,b]$, let M_i and m_i be the supremum and infimum respectively of f on $[x_{i-1}, x_i]$, and $\Delta x_i = x_i - x_{i-1}$. We then define:

- The upper Riemann sum $U_a^b(f,P) = \sum_{i=1}^n M_i \Delta x_i$.
- The lower Riemann sum $L_a^b(f,P) = \sum_{i=1}^n m_i \Delta x_i$.
- The ordinary Riemann sum $S_a^b(f,P) = \sum_{i=1}^n f(x_{i-1}) \Delta x_i$.

For any positive real Δx, let $P_{\Delta x} = \{a, a + \frac{b-a}{\Delta x}, a + 2\frac{b-a}{\Delta x}, \ldots, a + n\frac{b-a}{\Delta x}, b\}$. Now, for any positive real Δx, let $U_a^b(f, \Delta x) = U_a^b(f, P_{\Delta x})$. Similarly for $L_a^b(f, \Delta x)$ and $S_a^b(f, \Delta x)$.

A function $f : \mathbb{R} \to \mathbb{R}$ is *integrable* on $[a,b]$ if $L_a^b(f, \Delta x) \simeq U_a^b(f, \Delta x)$ for every infinitesimal Δx. This is equivalent to the standard definition of integrability [1, p. 110].

Definition. If f is integrable on $[a,b]$, then we define $\int_a^b f(x)\ dx = \mathrm{st}\left(S_a^b(f, \Delta x)\right)$ for some positive infinitesimal Δx.

Note that we can't properly speaking extend S_a^b, as S_a^b has a function as an argument and we can only extend functions that take in real numbers as arguments. Really, we're extending the function $g(x) = S_a^b(f, x)$, and writing $S_a^b(f, \Delta x) = {}^*g(\Delta x)$. To show this is well-defined, we need a lemma.

Lemma 5.1 ([1, p. 106])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ and $[a,b] \subseteq \mathbb{R}$. Given any two partitions P_1 and P_2 of $[a,b]$, we find $L_a^b(f, P_1) \leq U_a^b(f, P_2)$.*

Corollary 5.2. *If $0 < \Delta x_1, \Delta x_2$, then $L_a^b(f, \Delta x_1) \leq U_a^b(f, \Delta x_2)$.*

Theorem 5.3. *If f is integrable on $[a,b]$, then $\int_a^b f(x)dx$ is well-defined.*

Proof adapted from [1, Chapter 9.2]. The issue here is that $\int_a^b f(x)\ dx = \mathrm{st}\left(S_a^b(f, \Delta x)\right)$ might depend on our choice of positive infinitesimal Δx. Let Δx_1 and Δx_2 be two positive infinitesimals. We want to show that $S_a^b(f, \Delta x_1) \simeq S_a^b(f, \Delta x_2)$.

Note that in the reals, $(\forall \Delta x)(S_a^b(f, \Delta x) \leq U_a^b(f, \Delta x))$. This is because $\sum_{i=0}^n f(a + (i-1)\frac{b-a}{\Delta x})\Delta x \leq \sum_{i=0}^n M_i \Delta x$, as M_i is a maximum of f on $[a + (i-1)\frac{b-a}{\Delta x}, +i\frac{b-a}{\Delta x}]$ and so $f(a + (i-1)\frac{b-a}{\Delta x}) \leq M_i$. We can transfer this statement to the hyperreals to conclude that

$$S_a^b(f, \Delta x_1) \leq U_a^b(f, \Delta x_1).$$

By a similar line of reasoning, we conclude $L_a^b(f, \Delta x_1) \leq S_a^b(f, \Delta x_1)$. Since $L_a^b(f, \Delta x_1) \simeq U_a^b(f, \Delta x_1)$, this implies $S_a^b(f, \Delta x_1) \simeq U_a^b(f, \Delta x_1)$ as well. All of this equally applies to Δx_2, of course.

Let $L_1 = L_a^b(f, \Delta x_1)$, $L_2 = L_a^b(f, \Delta x_2)$, and similarly for U_1 and U_2. We know by Corollary 5.2 that $L_1, L_2 \leq U_1, U_2$, and so the possible orderings are $L_1 \leq L_2 \leq U_1 \leq U_2$, $L_1 \leq L_2 \leq U_2 \leq U_1$, or either of those with the indices swapped. In any case, the fact that $L_1 \simeq U_1$ and $L_2 \simeq U_2$ implies $L_1 \simeq L_2 \simeq U_1 \simeq U_2$. But we have $S_a^b(f, \Delta x_1) \simeq U_1 \simeq U_2 \simeq S_a^b(f, \Delta x_2)$, so we are done. $\square$

We now state the Fundamental Theorem of Calculus, proven for nonstandard objects in [1, pp. 111–112].

Theorem 5.4 (Fundamental Theorem of Calculus, [1, Theorem 9.4.2])**.** *If a function G has a continuous derivative f on $[a,b]$, then $\int_a^b f(x)dx = G(b) - G(a)$.*

5.2 Improper Integrals

Say $f : [a,b) \to \mathbb{R}$ is integrable on every interval $[a,c]$ for $a < c < b$. Standardly, we take the *improper integral* (where defined) to be

$$\int_a^b f(x)dx := \lim_{c \to b} \int_a^c f(x)dx.$$

Nonstandardly, we instead take

$$\int_a^b f(x)dx = \mathrm{st}\left({}^*\!\!\int_a^\gamma f(x)dx\right),$$

where $b \simeq \gamma < b$ and ${}^*\!\int_a^t f(x)dx$ indicates the extension ${}^*g(t)$ of $g(t) = \int_a^t f(x)dx$. Similarly, we have

$$\int_a^\infty f(x)dx = \operatorname{st}\left({}^*\!\int_a^N f(x)dx\right)$$

Where N is a positive unlimited hyperreal. To be clear, there is no guarantee that these standard parts exist, or that they are the same across all potential γ's or N's—in those cases, the improper integral is undefined.

Say we want to take $\int_0^1 \frac{1}{\sqrt{x}}\, dx$. Since $\frac{1}{\sqrt{x}}$ isn't bounded on $[0,1]$, we can't take a proper integral. So we instead venture to take $\int_\delta^1 \frac{1}{\sqrt{x}}\, dx$ for some positive infinitesimal δ.

By 5.4, we have the sentence

$$(\forall a \in \mathbb{R})\left(0 < a < 1 \to \int_a^1 \frac{1}{\sqrt{x}}\, dx = 2 - 2\sqrt{a}\right).$$

This transfers to ${}^*\mathbb{R}$, so we conclude $\int_\delta^1 \frac{1}{\sqrt{x}}\, dx = 2 - 2\sqrt{\delta}$. Then $\int_0^1 \frac{1}{\sqrt{x}}\, dx = \operatorname{st}\left(2 - 2\sqrt{\delta}\right) = 2 - 2 \cdot \operatorname{st}\left(\sqrt{\delta}\right) = 2$.

5.3 Hyperfinite Sets & Sums

Definition. An internal set $A = [A_n]$ is *hyperfinite* if every A_n is finite.

The hyperfinite sets are "internally finite." They share a lot of properties with finite sets.

Theorem 5.5. *If $\varphi(A)$ holds for every finite $A \subseteq \mathbb{R}$, then ${}^*\varphi(X)$ holds for every hyperfinite $X \subseteq {}^*\mathbb{R}$.*

Proof. Say $X = [A_n]$, with each A_n finite. Then $[[\varphi(A_n)]] = \mathbb{N}$, and so by Łoś's Theorem ${}^*\varphi(X)$. $\square$

This lets us easily get a lot of nice properties about hyperfinite sets. For instance, $(\exists x \in A)(\forall y \in A)(x \geq y)$ ensures that every hyperfinite set has a maximum element.

Now, for any finite set A_n and function $f_n : \mathbb{R} \to \mathbb{R}$, we can easily define the sum $\sum_{x \in A_n} f_n(x)$. This is, after all, just a sum of a finite collection of numbers. Using this, however, we can easily extend our summation to hyperfinite sets:

Definition. If $A = [A_n]$ is a hyperfinite set, and $f = [f_n]$ is an internal function, we define the *hyperfinite sum*

$$\sum_{x \in [A_n]} f(x) = \left[\sum_{x \in A_n} f_n(x)\right].$$

5.4 Integrals as Hyperfinite Sums

Now, let $f : [a,b] \to \mathbb{R}$ be an integrable function. To take $\int_a^b f(x)dx$, we want to divide $[a,b]$ into infinitely many segments of infinitesimal width, and then add up the area of the rectangles above or below those segments. Hyperfinite sums give us a way to do this.

To divide $[a,b]$ into infinitely many segments of infinitesimal width, we will construct a hyperfinite partition where the segments are of infinitesimal width $dx > 0$. Say $dx = [\langle \Delta x_1, \Delta x_2, \ldots \rangle]$. Let $P_n \cup \{b\}$ be partition of $[a,b]$ into segments of width Δx_n (plus a final"remainder" segment of length $\leq \Delta x_n$). So

$$P_n = \left\{a + k\Delta x_n \,|\, 0 \leq k < \frac{b-a}{\Delta x_n},\ k \in \mathbb{N}\right\}.$$

Let c_n denote the greatest element of P_n, the second-to-last element of our partition (the last element is b). So when Δx_n doesn't "evenly divide" $b - a$, we have $c_n = a + \lfloor \frac{b-a}{\Delta x_n} \rfloor \cdot \Delta x_n$. Let r_n denote the length of the "remainder" segment $r_n = b - c_n$. Notice $r_n \leq \Delta x_n$. Then, the ordinary Riemann sum of f on $[a,b]$ with partition $P_n \cup \{b\}$ is

$$S_a^b(f, \Delta x_n) = \sum_{x \in P_n} f(x)\Delta x_n + f(c_n) \cdot (r_n - \Delta x_n).$$

This is because $\sum_{x \in P_n} f(x)\Delta x_n$ has the term $f(c_n)\Delta x_n$, while $S_a^b(f, \Delta x_n)$ has the term $f(c_n) \cdot r_n$. This is because $S_a^b(f, \Delta x_n) = \sum_{i=1}^n f(x_{i-1})\Delta_i$, where Δ_i is the width of the i^{th} interval, and the width of the last interval is r_n.

Now, let $P = [P_n]$ be our "hyperfinite partition" of $[a, b]$ into hyperfinitely many intervals of infinitesimal width dx. Intuitively, we would hope that $\int_a^b f(x)dx$ is equal to the hyperfinite "Riemann sum" over this partition, i.e.

$$\operatorname{st}\left(\sum_{x\in P} f(x)\cdot dx\right) = \int_a^b f(x)dx = \operatorname{st}\left(S_a^b(f, dx)\right).$$

(Note that we're taking the hyperfinite sum over the internal function $g(x) = f(x) \cdot dx$. This is internal because we let $g = [g_n]$, where $g_n(x) = f(x) \cdot \Delta x_n$. Then $g(x) = [f(x) \cdot \Delta x_n] = [f(x)] \cdot [\Delta x_n] = f(x) \cdot dx$.) To show this, we can write (adapted from [1, Chapter 12.7]):

$$\operatorname{st}\left(\sum_{x\in P} f(x)\cdot dx\right) = \operatorname{st}\left(\left[\sum_{x\in P_n} f(x)\cdot \Delta x_n\right]\right).$$

Now, letting N_n be such that $a + N_n \frac{b-a}{\Delta x_n} = c_n$ and letting $x_i = a + i \cdot \frac{b-a}{\Delta x_n}$, we have:

$$\begin{aligned}
\operatorname{st}\left(\left[\sum_{x\in P_n} f(x)\cdot \Delta x_n\right]\right) &= \operatorname{st}\left(\left[\sum_{i=0}^{N_n+1} f(x_{i-1})\cdot \Delta x_n\right]\right)\\
&= \operatorname{st}\left(\left[S_a^b(f, \Delta x_n)\right]\right)\\
&= \operatorname{st}\left({}^*S_a^b(f, dx)\right) = \int_a^b f(x)dx.
\end{aligned}$$

Intuitively, this is enough to see that $\sum_{x\in P} f(x) \cdot dx$ and ${}^*S_a^b(f, dx)$ really are doing the same thing, in that they're taking the hyperreal corresponding to a sequence of real numbers that are closer and closer approximations of $\int_a^b f(x)dx$ by Riemann sums.

The issue here is that $\sum_{i=0}^{N_n+1} f(x_{i-1}) \cdot \Delta x_n$ isn't necessarily equal to $S_a^b(f, \Delta x_n)$, due to our concerns about the width of the final segment of our partition r_n. In order to make this reasoning rigorous, we need to deal with this "last segment" problem (not addressed in [1]):

Theorem 5.6.

$$\sum_{x\in P} f(x)\cdot dx \simeq S_a^b(f, dx)$$

Proof. We have that

$$\sum_{x\in P} f(x)\cdot dx - S_a^b(f, dx) = \left[\sum_{x\in P_n} (f(x)\cdot \Delta x_n) - S_a^b(f, \Delta x)\right]$$

and our earlier observation that

$$S_a^b(f, \Delta x_n) = \sum_{x\in P_n} f(x)\Delta x_n + f(c_n)\cdot(r_n - \Delta x_n),$$

and so

$$\sum_{x\in P} f(x)\cdot dx - S_a^b(f, dx) = [f(c_n)\cdot(\Delta x_n - r_n)] = [f(c_n)]\cdot[\Delta x_n - r_n].$$

We want to show that this is infinitesimal. Now, $0 < r_n \leq \Delta x_n$, so $0 \leq \Delta x_n - r_n < \Delta x_n$ and $0 \leq [\Delta x_n - r_n] < [\Delta x_n] = dx \simeq 0$. Hence $[\Delta x_n - r_n] \simeq 0$. Since f is integrable on $[a, b]$, it is (by definition) bounded on $[a, b]$, and so $[f(c_n)]$ is limited. So since a limited number times an infinitesimal is infinitesimal, $[f(c_n)] \cdot [\Delta x_n - r_n]$ is infinitesimal, and we're done. ∎

With this lemma, we can write:

$$\int_a^b f(x)dx = \operatorname{st}\left(S_a^b(f, dx)\right) = \operatorname{st}\left(\sum_{x\in P} f(x)\cdot dx\right).$$

Acknowledgements

I'd like to thank Professor Kelly Isham for her invaluable guidance during the writing of this paper and the semester as a whole. I'd also like to thank Professor Dan Saracino for suggesting the topic and sources, and being an excellent resource on the subject matter. Any mistakes are, of course, entirely my own.

References

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

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