# AN INTRODUCTION TO NONSTANDARD ANALYSIS

ISAAC DAVIS

Abstract. In this paper we give an introduction to nonstandard analysis, starting with an ultrapower construction of the hyperreals. We then demonstrate how theorems in standard analysis "transfer over" to nonstandard analysis, and how theorems in standard analysis can be proven using theorems in nonstandard analysis.

## 1. Introduction

For many centuries, early mathematicians and physicists would solve problems by considering infinitesimally small pieces of a shape, or movement along a path by an infinitesimal amount. Archimedes derived the formula for the area of a circle by thinking of a circle as a polygon with infinitely many infinitesimal sides [1]. In particular, the construction of calculus was first motivated by this intuitive notion of infinitesimal change. G.W. Leibniz's derivation of calculus made extensive use of "infinitesimal" numbers, which were both nonzero but small enough to add to any real number without changing it noticeably. Although intuitively clear, infinitesimals were ultimately rejected as mathematically unsound, and were replaced with the common $\epsilon$-$\delta$ method of computing limits and derivatives. However, in 1960 Abraham Robinson developed nonstandard analysis, in which the reals are rigorously extended to include infinitesimal numbers and infinite numbers; this new extended field is called the field of hyperreal numbers. The goal was to create a system of analysis that was more intuitively appealing than standard analysis but without losing any of the rigor of standard analysis.

In this paper, we will explore the construction and various uses of nonstandard analysis. In section 2 we will introduce the notion of an ultrafilter, which will allow us to do a typical ultrapower construction of the hyperreal numbers. We then demonstrate that these hyperreals do in fact satisfy the axioms of a totally ordered field, and consider the relationship between the field of real numbers and the field of hyperreals.

In section 4 we introduce the main theorem of nonstandard analysis, the transfer principle, which allows us to transfer first-order sentences back and forth between the reals and the hyperreals. This is an incredibly powerful tool, and most of the rest of the paper will be spent exploring the many uses of the transfer principle. We will show how the transfer principle can be used to easily demonstrate that the hyperreals satisfy all the properties we wish them to satisfy, and to give simple, intuitively clear proofs of theorems in standard calculus which would be impossible without resorting to calculations in the hyperreals. In the last section, we will touch on how the transfer principle might be extended to include certain higher-order sentences.

*Date*: August 14, 2009.

## 2. Constructing the Hyperreals

The basic idea behind constructing the hyperreal numbers is to create a field of real-valued sequences, in which every standard real number is embedded as the corresponding constant sequence. We begin by considering the set of real-valued sequences, which we denote $\mathbb{R}^{\mathbb{N}}$, under pointwise addition and multiplication. Clearly if we add or multiply two real-valued sequences pointwise we get another real-valued sequence, so $\mathbb{R}^{\mathbb{N}}$ is closed under pointwise addition and multiplication. Similarly, for any real-valued sequence we can construct its additive inverse by taking the additive inverse of each term in the sequence. Thus, under pointwise addition and multiplication, $\mathbb{R}^{\mathbb{N}}$ forms a commutative ring with identity. It falls short, however, of satisfying all the properties of a field. One problem we would run into is the presence of zero-divisors. Consider, for example, the two sequences

$$a = 0, 1, 0, 1, \dots \quad b = 1, 0, 1, 0 \dots$$

Neither of these sequences is equal to the zero sequence $e_n = 0$. However, pointwise multiplication would give us

$$a \cdot b = 0 \cdot 1, 1 \cdot 0, 0 \cdot 1, 1 \cdot 0, \dots = 0, 0, 0, 0, \dots = e$$

Thus we have two nonzero elements whose product is zero, which violates the field axioms. To avoid this problem, we must first introduce the notion of a free ultrafilter on $\mathbb{N}$.

### 2.1. Free Ultrafilters.

**Definition 2.1.** (Free Ultrafilter) A *filter* $\mathcal{U}$ on a set $J$ is a subset of $\mathcal{P}(J)$, the power set of $J$, satisfying the following properties:

(1) Proper filter: $\emptyset \notin \mathcal{U}$,
(2) Finite intersection property: If $A, B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$,
(3) Superset property: If $A \in \mathcal{U}$ and $A \subseteq B$, then $B \in \mathcal{U}$.
$\mathcal{U}$ is said to be an *ultrafilter* if it also satisfies:
(4) Maximality: For all $A \subseteq J$, either $A \in \mathcal{U}$ or $J \backslash A \in \mathcal{U}$
$\mathcal{U}$ is said to be *free ultrafilter* if it also satisfies:
(5) Freeness: $\mathcal{U}$ contains no finite subsets of $J$.

Ultrafilters actually satisfy a slightly stronger property than 4; this property states that, given any ultrafilter $\mathcal{U}$ on $\mathbb{N}$ and any finite collection of disjoint subsets of $\mathbb{N}$ whose union is $\mathbb{N}$, exactly one of these subsets must be in $\mathcal{U}$. This result will come in handy later in the paper, so we will prove it here.

**Lemma 2.2.** *Let $\mathcal{U}$ be an ultrafilter on $\mathbb{N}$, and let $\{A_1, \dots, A_n\}$ be a finite collection of disjoint subsets such that $\bigcup_{j=1}^{n} A_j = \mathbb{N}$. Then $A_i \in \mathcal{U}$ for exactly one $i \in \{1, \dots, n\}$.*

*Proof.* First we will prove that $\mathcal{U}$ contains at least one such subset, then we will prove that it can contain only one. Suppose, first, that $\mathcal{U}$ does not contain any of the subsets $A_1, \dots, A_n$. Then, by property 4, $\mathcal{U}$ must contain the complement of each subset $\mathbb{N} \backslash A_i$. However, $\mathcal{U}$ must also contain the intersection of these complements, which is

$$\bigcap_{j=1}^{n} \left(A_j^c\right) = \left(\bigcup_{j=1}^{n} A_j\right)^c = (\mathbb{N})^c = \emptyset$$

However, $\mathcal{U}$ cannot contain the empty set. Therefore $\mathcal{U}$ must contain at least one of $A_1, \ldots, A_n$.

Now, suppose that $\mathcal{U}$ contains $A_i$ and $A_j$ for some $i \neq j$. Then $\mathcal{U}$ must also contain $A_i \cap A_j$. However, $A_i$ and $A_j$ are disjoint, so $A_i \cap A_j = \emptyset$ and, again, $\mathcal{U}$ cannot contain the empty set. Therefore $\mathcal{U}$ can contain only one of $A_1, \ldots, A_n$. □

It is not immediately obvious that such an object exists. In fact the only ultrafilters which can be explicitly constructed are not free. These are the fixed, or *principal* ultrafilters, which are concentrated on a finite number of points. Such an ultrafilter would take the form $\mathcal{U} = \{S \in \mathcal{P}(\mathbb{N}) | x_1, \ldots, x_n \in S\}$, for some finite number of elements $x_1, \ldots, x_n \in \mathbb{N}$. But these principal ultrafilters are not very interesting, and are not sufficient for our purposes. We specifically need a free, or *nonprincipal* ultrafilter. And although we cannot explicitly construct one, we can use Zorn's Lemma to prove their existence. In fact we will prove something even stronger: that any filter can be extended to a maximal filter, and that this maximal filter satisfies the properties of an ultrafilter. Applying this extension to the right filter will give us a free ultrafilter.

**Lemma 2.3.** *(Ultrafilter Lemma) Let $A$ be a set and $F_0 \subset \mathcal{P}(A)$ be a filter on $A$. Then $F_0$ can be extended to an ultrafilter $\mathcal{F}$ on $A$.*

*Proof.* There are two steps to this proof. First, we apply Zorn's Lemma to demonstrate the existence of a maximal filter on $A$ containing $F_0$. Then we demonstrate that this maximal filter satisfies property 4 of ultrafilters.

- Consider the set $\Phi$ of filters on $A$ which contain $F_0$. This forms a partially ordered set under the relation $\subseteq$. Now, consider any chain $F_0 \subset F_1 \subset F_2 \ldots$ of filters in $\Phi$. Our claim is that $(\bigcup_{n=0}^{\infty} F_n) = G$ is a filter in $\Phi$ which is an upper bound of the chain. Clearly if $\emptyset \notin F_n$ for any $n \geq 0$, then $\emptyset \notin G$. Similarly, for any $a \in G$, $a \in F_n$ for some $n$, so for any $b \supseteq a$, $b \in F_n \subset G$. Therefore $G$ satisfies the superset property. Lastly, suppose $a, b \in G$. Then $a \in F_m$ and $b \in F_n$ for some $m, n \geq 0$. Suppose without loss of generality that $m \leq n$. Then $F_m \subseteq F_n$, so $a, b \in F_n$, which by the finite intersection property means that $a \cap b \in F_n \subset G$. Therefore $G$ satisfies the finite intersection property as well. This means that $G$ is both a filter in $\Phi$ and an upper bound for $\{F_0, F_1, \ldots\}$. So $\Phi$ satisfies the hypothesis of Zorn's Lemma, which means that $\Phi$ must contain some maximal element, which we denote $\mathcal{F}$.
- Pick any subset $X \subseteq A$, and suppose that $\mathcal{F}$ contains neither $X$ nor $A \backslash X$. Then $\mathcal{F}$ must contain some some $a \in \mathcal{F}$ such that $a \cap X = \phi$. If not, then $\mathcal{F} \cup \{X\}$ would also be a filter, which violates the maximality of $\mathcal{F}$. Similarly, $\mathcal{F}$ must contain some set $b$ such that $b \cap (A \backslash X) \neq \emptyset$. However, by the finite intersection property we must have $a \cap b \neq \emptyset$. But this is impossible, as $a$ exists entirely outside $X$ and $b$ exists entirely outside $A \backslash X$. Therefore $\mathcal{F}$ must contain either $X$ or $A \backslash X$.

□

The existence of a free ultrafilter (which contains no finite sets) follows immediately from this theorem, by simply taking the filter consisting of all cofinite sets and extending it to an ultrafilter. Since this filter contains no finite sets, and the resulting ultrafilter contains every set or its complement, it is clear that this ultrafilter contains no finite sets, and thus satisfies freeness.

The last tool we need before we can rigorously construct the field of hyperreals is an equivalence relation on the set of real-valued sequences.

**Definition 2.4.** (Equivalence Modulo an Ultrafilter) Given a free ultrafilter $\mathcal{U}$ on $\mathbb{N}$, and real-valued sequences $a, b \in \mathbb{R}^{\mathbb{N}}$, we define the relation $=_{\mathcal{U}}$ by $a =_{\mathcal{U}} b$ if $\{j \in \mathbb{N} | a_j = b_j\} \in \mathcal{U}$

We now show that this is an equivalence relation. For reflexivity, take any sequence $a$. We have $\{j \in \mathbb{N} | a_j = a_j\} = \mathbb{N}$, and by properties 1 and 4 of a free ultrafilter, $\mathbb{N} \in \mathcal{U}$. Therefore $a =_{\mathcal{U}} a$. Similarly, for any sequences $a$ and $b$ we have $\{j \in N | a_j = b_j\} = \{j \in \mathbb{N} | b_j = a_j\}$, so $a =_{\mathcal{U}} b$ implies $b =_{\mathcal{U}} a$, which proves commutativity. Now, suppose $a =_{\mathcal{U}} b$ and $b =_{\mathcal{U}} c$. Then the sets $\{j \in \mathbb{N} | a_j = b_j\}$ and $\{j \in \mathbb{N} | b_j = c_j\}$ are both elements of $\mathcal{U}$. But the set on which $a_j = b_j$ and $b_j = c_j$ is $\{j \in \mathbb{N} | a_j = b_j \ \& \ b_j = c_j\} = \{j \in \mathbb{N} | a_j = b_j\} \cap \{j \in \mathbb{N} | b_j = c_j\}$, which by property 2 of a free ultrafilter must also be an element of $\mathcal{U}$. Therefore $a =_{\mathcal{U}} b$ and $b =_{\mathcal{U}} c$ implies $a =_{\mathcal{U}} c$, which proves transitivity. We will use $[a]$ to denote the equivalence class of the sequence $a$.

So $=_{\mathcal{U}}$ defines an equivalence relation on the set of real-valued sequences, which we call *equivalence modulo* $\mathcal{U}$. But how will this fix the problems mentioned before? Consider once again the sequences $a = 0, 1, 0, 1, \ldots$ and $b = 1, 0, 1, 0, \ldots$. Multiplying pointwise still gives us the zero sequence. However, by property 4 of free ultrafilters, an ultrafilter $\mathcal{U}$ must contain either the set of even numbers or the set of odd numbers, but not both. Therefore one of the sequences $a$ and $b$ must be equivalent to 1 modulo $\mathcal{U}$, and the other must be equivalent to 0 modulo $\mathcal{U}$.

2.2. **An Ultrapower Construction of the Hyperreals.** Consider $\mathbb{R}^{\mathbb{N}}$, the set of real-valued sequences. As previously observed, this set does not form a field; the best we can do with $\mathbb{R}^{\mathbb{N}}$ is a ring. However, if we take the quotient of $\mathbb{R}^{\mathbb{N}}$ by the equivalence relation described above, the resulting set of equivalence classes will form a field. We must proceed carefully though, as it is not obvious that addition and multiplication of these equivalence classes are well-defined. To prove this, we must prove that if $a_1 =_{\mathcal{U}} a_2$ and $b_1 =_{\mathcal{U}} b_2$, then $a_1 + b_1 =_{\mathcal{U}} a_2 + b_2$. That is, if we pick any elements from two different equivalence classes and add them together pointwise, we should end up in the same equivalence class regardless of which elements we originally chose. To do so, we shall introduce a notational convention, borrowed from R. Goldblatt [1], which we will continue to use throughout this paper. Let $a$ and $b$ be real-valued sequences. We denote their "agreement set," that is $\{j \in \mathbb{N} | a_j = b_j\}$ by $[[a = b]]$. Thus we have $a =_{\mathcal{U}} b$ if and only if $[[a = b]] \in \mathcal{U}$.

**Lemma 2.5.** *Pointwise addition and multiplication are well-defined binary operations on the set of real-valued sequences under ultrafilter equivalence.*

*Proof.* Suppose that $a_1 =_{\mathcal{U}} a_2$ and $b_1 =_{\mathcal{U}} b_2$. This means that $[[a_1 = a_2]] \in \mathcal{U}$ and $[[b_1 = b_2]] \in \mathcal{U}$, so by the finite intersection property we know that $[[a_1 = a_2]] \cap [[b_1 = b_2]] = \{j \in N | a_1 = a_2 \ \& \ b_1 = b_2\} \in \mathcal{U}$. But the set on which both $a_1$ and $a_2$ agree and $b_1$ and $b_2$ agree is the same set on which $a_1 + b_1$ and $a_2 + b_2$ agree. That is, $\{j \in N | a_1 = a_2 \ \& \ b_1 = b_2\} = [[a_1 + b_1 = a_2 + b_2]] \in \mathcal{U}$. Therefore $a_1 + b_1 =_{\mathcal{U}} a_2 + b_2$, so addition is well-defined. An analagous proof follows for multiplication. $\square$

Formally, the hyperreal numbers are these equivalence classes of $\mathbb{R}^{\mathbb{N}}$ under ultrafilter equivalence, which we denote $^*\mathbb{R}$. Now that we know our operations are well defined, we have all the necessary tools to prove that $^*\mathbb{R}$ is a field.

**Theorem 2.6.** *The set* $^*\mathbb{R}$ *with pointwise addition and multiplication is a field.*

*Proof.* Commutativity, associativity, and distributivity follow directly from the corresponding properties of the real numbers. We will prove distributivity to illustrate this, then move on to the existence and uniqueness of identies and inverses.

- Distributivity: Let $a, b, c, d \in^* \mathbb{R}$. By distributivity of real numbers, we have $[[a_j(b_j + c_j) = d_j]] = [[a_j b_j + a_j c_j = d_j]]$. Therefore if $a(b+c) =_{\mathcal{U}} d$, we must also have $ab + ac =_{\mathcal{U}} d$.
- Identities: Let 1 and 0 denote their corresponding constant sequences in $^*\mathbb{R}$. We clearly have $1 \cdot a =_{\mathcal{U}} a$ and $0 + a =_{\mathcal{U}} a$ for all $a \in^* \mathbb{R}$. We need only prove the uniqueness of these identities. Suppose there exists some $e \in^* \mathbb{R}$ such that for all $a \in {}^*\mathbb{R}$, $ea =_{\mathcal{U}} a$. Then we have $[[e_j a_j = a_j]] \in \mathcal{U}$. But by the uniqueness of the identity in $\mathbb{R}$, we must have $[[e_j a_j = a_j]] = [[e_j = 1] \in \mathcal{U}$. Therefore $e =_{\mathcal{U}} 1$. A similar proof follows for the additive identity.
- Additive Inverse: For any hyperreal $[a] \in {}^*\mathbb{R}$, we can define its additive inverse $-[a]$ pointwise; i.e, if $a = (a_n)_{n\in\mathbb{N}}$, then $-[a] = [-a] = [(-a_n)_{n\in\mathbb{N}}]$. Uniqueness then follows from the laws of addition: suppose there exist $x, y \in^* \mathbb{R}$ such that $x + a =_{\mathcal{U}} a + x =_{\mathcal{U}} 0$ and $y + a =_{\mathcal{U}} a + y =_{\mathcal{U}} = 0$. Then by associativity we have $y + (a + x) =_{\mathcal{U}} (y + a) + x$. But $a + x =_{\mathcal{U}} 0$ and $y + a =_{\mathcal{U}} 0$, so this implies that $y =_{\mathcal{U}} x$. Therefore additive identities are unique modulo an ultrafilter.
- Multiplicative Inverse: The multiplicative inverse is slightly trickier than the additive inverse. For any hyperreal $a = [(a_n)_{n\in\mathbb{N}}]$, consider the set $X = [[a_j = 0]]$. If $X \in \mathcal{U}$, then $a =_{\mathcal{U}} 0$, so it has no multiplicative inverse. Otherwise, by property 4 of the ultrafilter, $\mathbb{N}\backslash X$ must be an element of $\mathcal{U}$, so we have $a =_{\mathcal{U}} a'$, where $a'$ is the real-valued sequence defined by

  $$a'_n = \begin{cases} a_n & \text{if } n \in \mathbb{N}\backslash X \\ 1 & \text{if } n \in X \end{cases}$$

  Since no term in the sequence $(a'_n)_{n\in\mathbb{N}}$ is 0, there is no problem defining the inverse of the sequence pointwise. We therefore define $a^{-1}$ by

  $$a^{-1} = [((a'_n)^{-1})_{n\in\mathbb{N}}]$$

  We can then use commutativity and associativity of multiplication to prove the uniqueness (modulo an ultrafilter) of these inverses, using the same technique employed with the additive inverses.

□

Knowing that $^*\mathbb{R}$ is a field is all well and good, but it is not yet enough to understand how infinitesimal and infinite numbers can exist. We need further to define an order on $^*\mathbb{R}$. We do so in a manner similar to our definition of equivalence.

**Definition 2.7.** (Inequality Modulo an Ultrafilter) Given two hyperreals $a = [(a_n)_{n\in\mathbb{N}}]$ and $b = [(b_n)_{n\in\mathbb{N}}]$, and an ultrafilter $\mathcal{U}$ on $\mathbb{N}$, we define a relation $\leq_{\mathcal{U}}$ by $a \leq_{\mathcal{U}} b$ if $\{j \in \mathbb{N} | a_j \leq b_j\} \in \mathcal{U}$.

The fact that this relation imposes an order on ${}^*\mathbb{R}$ follows from property 2 of a free ultrafilter (the proof of transitivity of $\leq_{\mathcal{U}}$ is the same as the proof of transitivity for $=_{\mathcal{U}}$). To prove that $\leq_{\mathcal{U}}$ totally orders ${}^*\mathbb{R}$, take any two hyperreals $a, b \in {}^*\mathbb{R}$, and let $X = \{j \in \mathbb{N} | a_j \leq b_j\}$. By the maximality property either $X$ or $\mathbb{N} \backslash X$ must be an element of $\mathcal{U}$. If $X \in \mathcal{U}$, then $a \leq_{\mathcal{U}} b$. If $X \notin \mathcal{U}$, then $\mathbb{N} \backslash X = \{j \in \mathbb{N} | a_j > b_j\} \in \mathcal{U}$, therefore $b_j \leq_{\mathcal{U}} a_j$. This implies that $\leq_{\mathcal{U}}$ is a total ordering on ${}^*\mathbb{R}$.

2.3. **Infinitesimal and Infinite Numbers.** We now have our totally ordered field ${}^*\mathbb{R}$ of hyperreal numbers, but we have yet to rigorously define or demonstrate the existence of infinitesimal and infinite numbers. To do so, we will first introduce notation for the "standard" hyperreal numbers, those constant sequences corresponding to real numbers. We let ${}^\sigma\mathbb{R}$ denote the set of constant-valued sequences in ${}^*\mathbb{R}$ (similarly, ${}^\sigma\mathbb{N}$ denotes the set of constant sequences with natural number values). Using this notation, we now define infinite and infinitesimal numbers.

**Definition 2.8.** (Infinite and Infinitesimal Numbers) A hyperreal number $a \in {}^*\mathbb{R}$ is said to be *infinitesimal* if $a \leq_{\mathcal{U}} n$ for every $n \in {}^\sigma\mathbb{N}$, and *infinite* if $n \leq_{\mathcal{U}} a$ for every $n \in {}^\sigma\mathbb{N}$.

We now demonstrate the existence of an infinite number and an infinitesimal number. Let $\omega$ be the hyperreal number defined by $\omega_n = n$, and let $j$ be any standard natural number $j \in {}^\sigma\mathbb{N}$. Then $\omega_n \leq j$ for all $n \leq j$, but $\omega_n \geq j$ for all $n > j$. Therefore the set of indices on which $\omega_n$ is less than $j$ is finite. But as we have already seen, any free ultrafilter $\mathcal{U}$ must contain all cofinite subsets of $\mathbb{N}$, so $\{n \in \mathbb{N} | \omega_n > j\} \in \mathcal{U}$. So for any standard natural number $j$, $j \leq_{\mathcal{U}} \omega$, which makes $\omega$ an infinite number.

Similarly, let us now consider the hyperreal number $\frac{1}{\omega}_n = \frac{1}{n}$. This time, for any standard natural number $j$, we know that $\frac{1}{\omega}_n$ can only be greater than $j$ for a finite number of indices, which, as demonstrated above, means that $\frac{1}{\omega} \leq_{\mathcal{U}} j$ for any standard natural $j$. Therefore $\frac{1}{\omega}$ is an infinitesimal number.

2.4. **Properties of the Field ${}^*\mathbb{R}$.** We might ask what other properties ${}^*\mathbb{R}$ shares with $\mathbb{R}$, and what properties of $\mathbb{R}$ are lost in our construction of ${}^*\mathbb{R}$. In doing so, we must be careful in distinguishing between standard and nonstandard, limited and unlimited numbers. For the purpose of our discussion, the set of standard hyperreals ${}^\sigma\mathbb{R}$ is the image of the embedding $f : \mathbb{R} \to {}^*\mathbb{R}$ defined by $f(a) = [\langle a, a, a, a, \ldots \rangle]$. All other hyperreals are nonstandard. A hyperreal $a$ is finite, or *limited*, if there exists some $p, q \in {}^\sigma\mathbb{R}$ such that $p \leq_{\mathcal{U}} a \leq_{\mathcal{U}} n$, and unlimited (infinite) otherwise.

With our new vocabulary in mind, let us consider the Archimedean property of $\mathbb{R}$, which states that for every real number $a$, there exists some natural number $n$ such that $a < n$. Is this true of ${}^*\mathbb{R}$? Well, the statement "for every hyperreal number $a$, there is some standard natural number $n$ such that $a < n$" is clearly false; just consider the infinite number $\omega$ we constructed in the previous section. If, however, we allow $n$ to range over all hypernaturals, that is, all sequences with natural number values, then the statement is true: if we take some hyperreal $a$, then for each term in the sequence we can find a natural number greater than that term, and the hypernatural $n$ we construct in this manner will be greater than $a$ on all of $\mathbb{N}$, proving that $a <_{\mathcal{U}} n$. As we will see later, such considerations are essential to determining which properties of the real numbers (and hence which theorems in standard analysis) "transfer" over to the hyperreals and nonstandard analysis. In fact, once we have determined the criteria for a property to be transferable, we will revisit the Archimedean property and discuss it in this context to illustrate how such a transfer works.

Even though we are so far ill-equipped to discuss such properties of ${}^*\mathbb{R}$ thoroughly, we can still consider other properties relating to the structure of ${}^*\mathbb{R}$, and see exactly where and how all of the information in $\mathbb{R}$ is hidden in ${}^*\mathbb{R}$. To do so, we first introduce the following notation: we let $\mathcal{O}$ denote the subset of ${}^*\mathbb{R}$ containing all the finite numbers. That is, $\mathcal{O} = \{a \in {}^*\mathbb{R} | p \leq_{\mathcal{U}} a \leq_{\mathcal{U}} q \text{ for some } p, q \in {}^\sigma\mathbb{R}\}$. We then define a subset $\vartheta$ of $\mathcal{O}$ containing all infinitesimal numbers.

It should be clear intuitively that $\mathcal{O}$ is an ordered subring of ${}^*\mathbb{R}$: the sum of two finite numbers is finite, the product of two finite numbers is finite, and the additive inverse of a finite number is finite. The order is simply inherited from the order on ${}^*\mathbb{R}$. Perhaps less clear is the fact that $\vartheta$ is a proper ideal of $\mathcal{O}$. We will not prove this here, but it should not be hard for the reader to convince him or herself that the sum of two infinitesimals is infinitesimal, the product of two infinitesimals is an infinitesimal, the additive inverse of an infinitesimal is infinitesimal, and the multiplicative inverse of an infinitesimal is never infinitesimal (nor for that matter is it finite).

In fact, $\vartheta$ is a maximal proper ideal of $\mathcal{O}$. It is maximal because $\vartheta$ consists of all the nonregular elements of $\mathcal{O}$ (those finite numbers without finite multiplicative inverses). Since a proper ideal cannot contain any regular elements of the ring, $\vartheta$ must be maximal. Furthermore, suppose we have $b \in \vartheta$. Then $b$ is infinitesimal, so $0 \leq_{\mathcal{U}} a \leq_{\mathcal{U}} b$ implies that $a$ is also infinitesimal, and $a \in \vartheta$. Because $\vartheta$ satisfies this property, it is called an *order ideal* of $\mathcal{O}$. We will now use these facts to prove that $\mathcal{O}/\vartheta$ is in fact order isomorphic to $\mathbb{R}$. To do so, we will use a few elementary results from basic algebra on rings and fields without proof, which we list here.

**Lemma 2.9.** *Let $A$ be a commutative ring with identity, and let $I$ be a proper ideal of $A$.*

*(1) The quotient ring $A/I$ is a field if and only if $I$ is a maximal proper ideal of $A$.*
*(2) If $A$ is ordered and $I$ is an order ideal, then $A/I$ is a totally ordered ring*

We will also use without proof the fact that a totally ordered field is isomorphic to a subfield of $\mathbb{R}$ if and only if it is Archimedean, which can be proven using arithmetic on Dedekind cuts.

**Theorem 2.10.** *The quotient ring $\mathcal{O}/\vartheta$ is order isomorphic to $\mathbb{R}$.*

*Proof.* As we just demonstrated, $\vartheta$ is a maximal proper ideal of $\mathcal{O}$, so by lemma 2.9, part 1, $\mathcal{O}/\vartheta$ is a field. Furthermore, since $\vartheta$ is an order ideal, $\mathcal{O}/\vartheta$ is a totally ordered field by lemma 2.9 part 2. All that's left to show is that $\mathcal{O}/\vartheta$ is Archimedean, and hence isomorphic to a subfield of $\mathbb{R}$, and that $\mathcal{O}/\vartheta$ contains a copy of $\mathbb{R}$. Since $\mathbb{R}$ contains no proper subfields isomorphic to itself, this will prove that $\mathcal{O}/\vartheta$ is isomorphic to all of $\mathbb{R}$. To prove that $\mathcal{O}/\vartheta$ is Archimedean, observe that it contains a copy of the standard natural numbers, namely $\{n + \vartheta | n \in {}^\sigma\mathbb{N}\}$. Since $\mathcal{O}$ consists only of the finite hyperreals, we know that for any $a \in \mathcal{O}$, there exists some $n \in {}^\sigma\mathbb{R}$ such that $|a| \leq_{\mathcal{U}} n$ (we define $|a| = \max\{a, -a\}$). But the canonical homomorphism of $\mathcal{O}$ into $\mathcal{O}/\vartheta$ preserves order, so this means that $|a + \vartheta| \leq |n + \vartheta|$ in the quotient ring. Therefore $\mathcal{O}/\vartheta$ is Archimedean.

Now, take any $a \neq b \in {}^\sigma\mathbb{R}$, which is contained in $\mathcal{O}$. Since $b - a$ is finite but not infinitesimal, we have $a + \vartheta \neq b + \vartheta$. So the image of the canonical homomorphism of $\mathcal{O}$ into $\mathcal{O}/\vartheta$, restricted to ${}^\sigma\mathbb{R}$, is a natural copy of $\mathbb{R}$. Therefore $\mathcal{O}/\vartheta$ is a totally ordered Archimedean field containing a natural copy of $\mathbb{R}$, and is order isomorphic to all of $\mathbb{R}$. □

The isomorphism between $\mathcal{O}/\vartheta$ and $\mathbb{R}$ suggests an equivalent order homomorphism from $\mathcal{O}$ into $\mathbb{R}$ with kernel $\vartheta$. We define this to be the "standard part" homomorphism $st : \mathcal{O} \to \mathbb{R}$ (this notation comes from K.D. Stroyan and W.A.J. Luxemburg [2]. In some textbooks this function is called the "shadow" of a finite hyperreal). This standard part homomorphism induces an equivalence relation on $\mathcal{O}$ which we will use extensively in our definitions of calculus. We will introduce that relation here:

**Definition 2.11.** (Infinitesimal Closeness) Two finite hyperreals $a, b \in \mathcal{O}$ are said to be *infinitesimally close*, denoted $a \approx b$, if and only if $a - b \in \vartheta$.

We will check that this is an equivalence relation. For any finite hyperreal $a$, we have $a - a = 0 \in \vartheta$ (note that 0 is the only standard infinitesimal number). It is also clear that $a - b = b - a$, so $a \approx b$ implies $b \approx a$. Lastly, if $a \approx b$ and $b \approx c$, then $a - b \in \vartheta$ and $b - c \in \vartheta$, and the sum of two infinitesimals is infinitesimal, so $a - c = (a - b) + (b - c) \in \vartheta$. Therefore $a \approx c$.

Theorem 2.10, along with the notion of this "standard part" homomorphism, imply a very useful result which we will need in later sections of this paper. We state and prove this result here:

**Corollary 2.12.** *Every finite hyperreal $a \in \mathcal{O}$ is infinitesimally close to a unique real $c \in {}^\sigma\mathbb{R}$.*

*Proof.* Let $st : \mathcal{O} \to \mathbb{R}$ denote the canonical homomorphism from $\mathcal{O}$ into $\mathbb{R}$ with kernel $\vartheta$, resulting from the isomorphism demonstrated in the proof of theorem 2.10, and let $g : \mathbb{R} \to {}^\sigma\mathbb{R}$ denote the standard embedding of $\mathbb{R}$ in ${}^*\mathbb{R}$. Then for any real number $x \in \mathbb{R}$, the preimage under $st$ of $x$ is the set of hyperreals infinitesimally close to $g(x)$; that is, $st^{-1}(x) = \{b \in \mathcal{O} | b \approx g(x)\}$. But the preimage of all of $\mathbb{R}$ is all of $\mathcal{O}$, so we have

$$\mathcal{O} = st^{-1}(\mathbb{R}) = \bigcup_{x \in \mathbb{R}} st^{-1}(x)$$

So, for any $a \in \mathcal{O}$, we must have $a \in st^{-1}(x)$ for some $x \in \mathbb{R}$. But this means that $a \approx g(x)$, and $g(x) \in {}^\sigma\mathbb{R}$. Therefore $a$ is infinitesimally close to some real number.

To prove the uniqueness of this real number, suppose $a \approx c$ and $a \approx c'$, where $c, c' \in {}^\sigma\mathbb{R}$. Then by transitivity we have $c \approx c'$, which by definition means that $c - c' \in o$. But if $c$ and $c'$ are both real numbers, we must also have $c - c' \in {}^\sigma\mathbb{R}$. Therefore $c - c' \in (\vartheta \cap {}^\sigma\mathbb{R}) = \{0\}$, so $c = c'$. □

## 3. Some Basic Calculus with Hyperreals

We now have the tools necessary to extend sequences and functions to the hyperreals, and to do some basic calculus on the hyperreals. We will not, however, go too deep, for as we will see in the next section, the logical construction of a nonstandard model for analysis will allow us to "transfer" all of the necessary theorems from analysis into nonstandard analysis, without having to explicitly prove them using hyperreal calculus. We include this section simply to get the reader more comfortable working with the hyperreal system.

We can extend real-valued functions to hyperreal-valued functions in the following way: let $f : \mathbb{R} \to \mathbb{R}$ be a function. Then for every real-valued sequence $a \in \mathbb{R}^{\mathbb{N}}$, we let $f(a) = \langle f(a_1), f(a_2), \ldots \rangle$. We then define the extended function ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ by ${}^*f([a]) = [f(a)]$ for any hyperreal number $[a] \in {}^*\mathbb{R}$, where $[a]$ represents the equivalence class of some real-valued sequence $a \in \mathbb{R}^{\mathbb{N}}$.

This extension, however, is insufficient when we consider functions whose domain is a proper subset of $\mathbb{R}$. That is, if we have a function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, and we wish to extend $f$ to the hyperreals, we have to define the extension of its domain $A$ to a subset ${}^*A$ of the hyperreals. We define the extension $A^*$ of a subset $A \subset \mathbb{R}$ to be the set

$$ {}^*A = \{[r] \in {}^*\mathbb{R} | r_n \in A \text{ for all } n \in \mathbb{N}\} $$

So ${}^*A$ is the set of equivalence classes of sequences whose values range over the elements of $A$. This allows us to extend a function $f : A \to \mathbb{R}$ to a function ${}^*f : {}^*A \to {}^*\mathbb{R}$ in the same manner as above.

Now that we can extend real-valued functions to hyperreal-valued functions, let's look at some desireable properties of these functions (we will simply be giving definitions of these properties here; in later sections, once we have the transfer principle, we will prove that these definitions are equivalent to the familiar definitions we know for calculus).

**Definition 3.1.** (Infinitesimal Continuity) A function ${}^*f : {}^*A \to {}^*\mathbb{R}$ is continuous at a point $a \in {}^*A$ if for every infinitesimal $\epsilon \in \vartheta$, we have ${}^*f(a+\epsilon) \approx {}^*f(a)$. The function $f$ is said to be continuous if $f$ is continuous at $a$ for all $a \in {}^*A$.

Let's see what a continuous hyperreal function looks like. Consider, for example, the function ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ defined by ${}^*f(x) = x^2$. Clearly this is the hyperreal extension of the real-valued function $f(x) = x^2$, which is continuous on all of $\mathbb{R}$. Now, take any finite hyperreal $a \in \mathcal{O}$. For any infinitesimal $\epsilon \in \vartheta$, we have $f(a+\epsilon) = a^2 + 2a\epsilon + \epsilon^2$. An infinitesimal times itself is infinitesimal, an infinitesimal times a finite number is infinitesimal, and the sum of two infinitesimals is infinitesimal. Therefore ${}^*f(a+\epsilon) = a^2 + 2a\epsilon + \epsilon^2 \approx a^2 = {}^*f(a)$. So, by our definition, $f$ is continuous at $a$. But does this hold at infinite numbers? Consider as an example (again borrowed from [2]) the infinite number $\omega$ and the infinitesimal change $\frac{1}{\omega}$. We have ${}^*f(\omega + \frac{1}{\omega}) = \omega^2 + 2 + \frac{1}{\omega^2} \approx \omega^2 + 2$. Therefore ${}^*f(\omega + \frac{1}{\omega}) \not\approx {}^*f(\omega)$, so $f$ is not continuous at $\omega$, or for that matter any infinite number.

We can define differentiability in a similar fashion, which we will also prove is equivalent to our standard definition of differentiability.

**Definition 3.2.** (Infinitesimal Differentiability) A function ${}^*f : {}^*A \to {}^*\mathbb{R}$ is differentiable at a point $a \in {}^*A$ if there exists a finite $b \in {}^*\mathbb{R}$ such that for every $\epsilon \in \vartheta$, we have

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

In this case, $b$ is the derivative of ${}^*f$ at $a$.

Now, let's prove something familiar using these definitions, to convince ourselves that the same theorems we are used to in standard calculus will hold in this nonstandard calculus. Let's prove, for example, that differentiability implies continuity:

**Theorem 3.3.** *If a function $^*f : {}^*A \to {}^*\mathbb{R}$ is differentiable at $a \in {}^*A$, then $^*f$ is continuous at $a$.*

*Proof.* Suppose $^*f$ is differentiable at $a$. Then for any infinitesimal $\epsilon$ we have

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

for some $b \in \mathcal{O}$. But any infinitesimal times a finite number also infinitessimal, so $^*f(a+\epsilon) - {}^*f(a) \in o$. Therefore $^*f(a+\epsilon) \approx {}^*f(a)$, which means that $^*f$ is continuous at $a$. $\square$

This was a good exercise in convincing ourselves that this new calculus in at least some way resembles the standard calculus with which we were familiar. However, the proof itself differed very little from the usual proof one might give for this theorem in standard calculus. Let us now prove something slightly more complicated, the chain rule, to illustrate how working with the hyperreals may simplify things.

**Theorem 3.4.** *(Chain Rule) Let $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ and $g : {}^*\mathbb{R} \to {}^*\mathbb{R}$ be functions, with $g$ differentiable at $a \in {}^*\mathbb{R}$ and $f$ differentiable at $f(g(a))$. Then the function $f \circ g : {}^*\mathbb{R} \to {}^*\mathbb{R}$ is differentiable at $a$, and $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$.*

*Proof.* Let $x \approx a$, but with $x \neq a$. Consider the difference quotient

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

If $g(x) = g(a)$, then this difference quotient is 0, which means that $(f(g(a)))' = 0 = f'(g(a)) \cdot g'(a)$, since $g'(a)$ must also be zero. In the case that $g(x) \neq g(a)$, we can rewrite the difference quotient as follows:

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

However, we just proved that if $g$ is differentiable at $a$ then it must be continuous at $a$, and $x \approx a$, so $g(x) \approx g(a)$. So the above expression is really just the product of the difference quotients of $f$ at $g(a)$ and $g$ at $a$. And since $f$ is differentiable at $g(a)$ and $g$ is differentiable at $a$, it follows that

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

Therefore $f \circ g$ is differentiable at $a$ and $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$. $\square$

The above proof is deceptively simple. In fact, it is the exact proof that a slightly misguided high school math student might give; we say "misguided" here because it doesn't quite work in the context of standard calculus, since dividing by $x - a$ is invalid, as $x$ approaches $a$. In the nonstandard context, however, we are not taking a limit, so $x$ and $a$ are static numbers (neither one is approaching the other) which just happen to be very close to one another. There is no problem with dividing an expression by $x - a$. So this proof, which in standard calculus is intuitively appealing but invalid, is perfectly sound in nonstandard calculus.

In the next section, we will introduce the transfer principle, and use it not only to prove facts about $^*\mathbb{R}$ using $\mathbb{R}$, but to prove facts about $\mathbb{R}$ using $^*\mathbb{R}$, which is the main acheivement of nonstandard analysis.

## 4. The Transfer Principle

In this section, we will construct a method of transforming formulae in $\mathbb{R}$ to formulae in ${}^*\mathbb{R}$. We will then introduce the main theorem for this section, called Łóŝ Theorem, which will allow us to prove that a first-order formula in $\mathbb{R}$ is true if and only if it is true in ${}^*\mathbb{R}$. Granted, this is perhaps more limited than we want: there are plenty of theorems in analysis which cannot be formulated as first order formulae. We will first explore how far we can get by staying within first-order sentences (which is actually very far) before investigating a method of expanding this "transfer" to higher-order statements.

4.1. **The Formal Language and Relational Structure.** If we want to talk about transforming formulae, we must first specify the language in which we are constructing these formulae. Our language will consist of:

- The usual variables $x, y, z, \ldots$,
- the logical connectives

| | |
|---|---|
| and | $\bigwedge$ |
| or | $\bigvee$ |
| not | $\neg$ |
| implies | $\rightarrow$ |
| if and only if | $\leftrightarrow$, |

- quantifiers

| | |
|---|---|
| for all | $\forall$ |
| there exists | $\exists$, |

- operation, relation, and function symbols, $P$, $f$, $*$, et cetera,
- parentheses and brackets $(,)$ and $[,]$.

We also have constants in our language, but it is unnecessary to mention them in addition to the above list, since constants can be represented as nullary functions (functions with no inputs and one output). For example, our language could contain a nullary function which we call 1. If we were to interpret our language into a field or multiplicative group, this 1 function would be mapped to the multiplicative identity of the structure.

Furthermore, we must know exactly what structure we will be describing in this language. For this, we introduce the concept of a relational structure.

**Definition 4.1.** (Relational Structure) A *relational structure* $\mathcal{S} = \{S, R, F\}$ consists of a set $S$, a set $R$ of finitary relations on $S$, and a set $F$ of function relations on $S$.

Note that an $n$-ary relation as described above is a set of ordered $n$-tuples satisfying that relation (for example, the relation $<$ on $\mathbb{R}$ is the set of ordered pairs $\{(a, b) | a < b\}$). A unary relation $P$ is simply a subset of the whole set $S$, so the statement $P(\tau)$, where $\tau \in S$ simply says that $\tau$ belongs to the set $P$. For this reason we will write these unary relations as $\tau \in P$ (i.e., the sentence $(\forall x \in P)(\phi)$ is really shorthand for $(\forall x)(P(x) \rightarrow \phi)$). Similarly, a function $f$ which accepts $k$ elements as an input and outputs $l$ elements will be the set of ordered pairs of $k$-tuples and $l$-tuples $\{((a_1, \ldots, a_k), (b_1, \ldots, b_l)) \,|\, f(a_1, \ldots, a_k) = (b_1, \ldots, b_l)\}$. So all of these relations and functions are actually just sets of $n$-tuples satisfying those relations or functions. A statement like $P(\tau_1, \ldots, \tau_n)$, where $P$ is a relation symbol, simply means that the ordered $n$-tuple $(\tau_1, \ldots, \tau_n)$ is an element of the set $P$ of ordered $n$-tuples.

Writing all functions as ordered $k+l$-tuples can be very tedious, so whenever there is no chance for confusion, the familiar operations and functions (addition, multiplication, exponential function, etc.) will be written simply as $a+b=c$, $a \cdot b = d$, $e^a = b$,, and so on, rather than the sentences $+((a,b),(c))$, $\cdot((a,b),(d))$, and $\exp((a),(b))$.

We will be interested in two relational structures in this paper. The first is $\Re$, which consists of $\mathbb{R}$, and all possible relations (equality, inequality, etc.) and functions on $\mathbb{R}$. The second we will denote ${}^*\Re$, which consists of ${}^*\mathbb{R}$, and the "extensions" of the relations and functions from $\Re$. Here, however, we must explain what the extension of a relation is. Let $P$ be an $n$-ary relation on $\mathbb{R}$. For $b_1, \dots, b_n \in \mathbb{R}^{\mathbb{N}}$, we define

$$ {}^*P(b_1, \dots, b_n) \leftrightarrow \{j \in \mathbb{N} | P(b_1(j), \dots, b_n(j)\} \in \mathcal{U} $$

For the purpose of discussing relational structures, functions can be extended in a similar fashion. Namely, for $a_1, \dots, a_k, b_1, \dots, b_l \in {}^*\mathbb{R}$, we define

$$ {}^*f(a_1, \dots, a_k) = (b_1, \dots, b_l) \leftrightarrow \{j \in \mathbb{N} | f(a_1(j), \dots, a_k(j)) = (b_1(j), \dots, b_l(j))\} \in \mathcal{U} $$

We will define sentences in our language in the usual inductive fashion (note that sentences are formulae with no free variables). Any variable or constant (nullary function) is a term, and if $\tau_1, \dots, \tau_n$ are terms and $f$ is an n-ary operation or function, then $f(\tau_1, \dots, \tau_n)$ is a term. The atomic sentences are of the form $P(\tau_1, \dots, \tau_n)$, where $P$ is an $n$-ary relation and $\tau_1, \dots, \tau_n$ are terms. We define all other sentences inductively. If $\phi$ and $\psi$ are sentences, then the following are also sentences:

- $\neg\phi$
- $\phi \bigwedge \psi$
- $\phi \bigvee \psi$
- $\phi \rightarrow \psi$
- $\phi \leftrightarrow \psi$
- $(\forall x)(\phi)$
- $(\exists x)(\phi)$

Now that we have a general schematic for writing first-order sentences, we can construct a method of transforming sentences in $\mathbb{R}$ to sentences in ${}^*\mathbb{R}$.

4.2. **$*$-Transforms of First-Order Sentences.** Given a sentence $\phi$ in $\mathbb{R}$, how do we transform it into a corresponding sentence ${}^*\phi$ in ${}^*\mathbb{R}$ in such a way that preserves the "meaning"? We will define such a $*$-transform in the same inductive fashion we employed to define formulae. We start with terms. If $\tau$ is a variable or constant, its $*$-transform is just $\tau$ (we can think of the $*$-transform as simply mapping real constants to the corresponding hyperreal constants in ${}^\sigma\mathbb{R}$). If $\tau$ is a term of the form $f(\tau_1, \dots, \tau_n)$, where $f$ is an $n$-ary function, then ${}^*\tau = {}^*f({}^*\tau_1, \dots, {}^*\tau_n)$. We define all other $*$-transforms of sentences explicitly on the construction of the sentence:

$$\begin{aligned}
{}^*(P(\tau_1,\ldots,\tau_n)) &:= {}^*P({}^*\tau_1,\ldots,{}^*\tau_n)\\
{}^*(\neg\phi) &:= \neg{}^*\phi\\
{}^*\left(\phi\bigwedge\psi\right) &:= {}^*\phi\bigwedge{}^*\psi\\
{}^*\left(\phi\bigvee\psi\right) &:= {}^*\phi\bigvee{}^*\psi\\
{}^*\left(\phi\to\psi\right) &:= {}^*\phi\to{}^*\psi\\
{}^*\left(\phi\leftrightarrow\psi\right) &:= {}^*\phi\leftrightarrow{}^*\psi\\
{}^*(\forall x\in A)(\phi) &:= (\forall x\in{}^*A)({}^*\phi)\\
{}^*(\exists x\in A)(\phi) &:= (\exists x\in{}^*A)({}^*\phi)
\end{aligned}$$

Thus, constructing a $*$-transform of a sentence $\phi$ really just consists of putting a $*$ on every term in $\phi$, putting a $*$ on any relation symbol in $\phi$, and putting a $*$ on every set in $\phi$ acting as a bound on a variable. Note that in our case the $*$-transforms of the familiar relations $=$ and $\leq$ are $=_{\mathcal{U}}$ and $\leq_{\mathcal{U}}$, respectively.

As an example, consider once again the Archimedean property mentioned in the first section. The first-order sentence expressing the Archimedean property of the real numbers is the following:

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

As we mentioned before, this statement does not hold in ${}^*\mathbb{R}$ if $n$ is only allowed to vary over the standard hypernaturals ${}^\sigma\mathbb{N}$. However, the $*$-transform of this sentence is

$$(\forall x\in{}^*\mathbb{R})(\exists n\in{}^*N)(x<n)$$

which *is* true in ${}^*\mathbb{R}$: as we demonstrated before, as long as $n$ can take on the value of any hypernatural number, we can always find an $n$ such that $x<n$.

4.3. **Łóŝ' Theorem and the Transfer Principle.** Our hope in constructing this $*$-transform was that a sentence $\phi$ would be true if and only if ${}^*\phi$ is true. This idea is called the *transfer principle*, and this is the main result of nonstandard analysis. It not only allows us to show that ${}^*\mathbb{R}$ has all the properties of $\mathbb{R}$, but it allows us to prove theorems about $\mathbb{R}$ by first proving them in ${}^*\mathbb{R}$ (which is often simpler and more intuitively clear) and then transfering them back to $\mathbb{R}$.

To prove the transfer principle, we will need a theorem from the famous Polish logician Jerzy Łóŝ about first order formulae on ultraproducts of structures that share the same language. To discuss this, however, we will have to formally define an ultraproduct, because although we implicity used one in our construction of ${}^*\mathbb{R}$, an ultraproduct is a more general mathematical structure, and we only used a special case of an ultraproduct called a countable ultrapower.

**Definition 4.2.** (Ultraproduct) Let $\{A_\alpha|\alpha\in I\}$ be a family of structures with the same signature indexed by an infinite set $I$, and let $\mathcal{U}$ be a free ultrafilter on $I$. We take the cartesian product of these structures over the set $I$ modulo the ultrafilter equivalence relation defined in the first section. The resulting set of equivalence classes, which we denote

$$\mathcal{B}=\prod_{\alpha\in I}A_\alpha/\mathcal{U}$$

is the *ultraproduct* of the family $A_\alpha$ of structures with respect to the ultrafilter $\mathcal{U}$. If $A_\alpha = A_\beta = A$ for all $\alpha, \beta \in I$, then $\mathcal{B}$ is called an *ultrapower* of the structure $A$.

Note that, in general, the ultraproduct $\mathcal{B}$ that we obtain depends on the ultrafilter we chose beforehand. Consider, for example, an ultraproduct of a family of finite fields of increasing prime order. Some of the fields will have order congruent to $1 \mod 4$, and others will have order congruent to $3 \mod 4$. We could therefore choose one ultrafilter concentrated on the indices of fields with order congruent to 1 mod 4, and another ultrafilter concentrated on the indices of fields with order congruent to $3 \mod 4$. These two ultraproducts, with respect to these two ultrafilters, would not be isomorphic.

It would seem important for the purpose of nonstandard analysis to know that all fields ${}^*\mathbb{R}$ are isomorphic, regardless of which ultrafilter we choose. As it turns out, this is an unknown result (it can only be proven assuming the continuum hypothesis). However, as a result of this miracle of ultraproducts known as Łóŝ' Theorem, it is irrelevant whether or not the resulting fields are isomorphic. Instead they are called *elementary equivalent*, which means that any first-order sentence which holds in one will hold in all others.

We will now state this "miracle," demonstrate how the transfer principle we need is simply a special case, and prove this special case.

**Theorem 4.3.** *(Łóŝ' Theorem) Let $\{A_\alpha | \alpha \in I\}$ be a family of structures with the same signature, and let $\mathcal{B}$ be an ultraproduct of these structures with respect to some free ultrafilter $\mathcal{U}$ on $I$. Then for any first order formula $\phi(x_1, \ldots, x_n)$, with free variables $x_1, \ldots, x_n$, and any $[b_1], \ldots, [b_n] \in \mathcal{B}$,*

$$\mathcal{B} \models \phi([b_1], \ldots, [b_n]) \leftrightarrow \{\alpha \in I | A_\alpha \models \phi(b_1(\alpha), \ldots, b_n(\alpha)\} \in \mathcal{U}$$

The intuitive way to think about Łóŝ' Theorem is that a first-order formula holds in the ultraproduct if and only if it holds for "almost all" of the original structures $A_\alpha$, where "almost all" is defined by the ultrafilter $\mathcal{U}$.

Now, our transfer principle is only concerned with sentences, so we will not have to worry about these free variables. Observe, however, that the $*$-transform is defined inductively on the construction of a formula, and all formulae can be broken down into combinations of atomic relations and terms, under conjunction, negation, quantification, and so on. Furthermore, we defined the $*$-transform of an atomic relation $P$ to be true if $P(r_1(j), \ldots, r_n(j))$ is true for "almost all" $j \in \mathbb{N}$. So, when we consider only atomic formulae, our definition coincides exactly with the predictions of Łóŝ' Theorem. And because all other formulae are just combinations of atomic formulae, it is easy to show that the transfer principle is carried through to all other formulae. Our transfer principle is therefore a special case of Łóŝ' Theorem, which we will prove here.

**Lemma 4.4.** *(Transfer Principle) A sentence $\phi$ is true if and only if ${}^*\phi$ is true.*

*Proof.* We will prove this by induction on sentences. For the base case, suppose the atomic formula $P(\tau_1, \ldots, \tau_n)$ is true for chosen values of $\tau_1, \ldots, \tau_n$. The $*$-transform of this sentence is ${}^*P({}^*\tau_1, \ldots, {}^*\tau_n)$, which by definition is true if and only if the set of indices $j$ such that $P({}^*\tau_1(j), \ldots, {}^*\tau_n(j))$ is in our ultrafilter. But the $*$-transforms of constants in $\mathbb{R}$ are just the corresponding constant sequences, so by the very definition of ${}^*\tau$ it follows that the set of indices $j$ such that $P({}^*\tau_1(j), \ldots, {}^*\tau_n(j))$ is all of $\mathbb{N}$, which must be in our ultrafilter. Therefore $P(\tau_1, \ldots, \tau_n) \rightarrow {}^*(P(\tau_1, \ldots, \tau_n))$.

Conversely, suppose $\neg P(\tau_1, \dots, \tau_n)$. Then, by the same argument, set of indices $j$ such that $P({}^*\tau_1(j), \dots, {}^*\tau_n(j))$ is the empty set, which cannot be in our ultrafilter. Therefore $\neg P(\tau_1, \dots, \tau_n) \to \neg{}^*(P(\tau_1, \dots, \tau_n))$, so $P(\tau_1, \dots, \tau_n) \leftrightarrow {}^*(P(\tau_1, \dots, \tau_n))$.

Now, for the induction step, let $\phi$ and $\psi$ be sentences, and suppose $\phi \leftrightarrow {}^*\phi$ and $\psi \leftrightarrow {}^*\psi$. We will prove that the same property holds for all sentences which can be constructed out of $\phi$ and $\psi$.

- Consider the sentence $\phi \bigwedge \psi$, and its $*$-transform ${}^*(\phi \bigwedge \psi) = {}^*\phi \bigwedge {}^*\psi$. If $\phi \bigwedge \psi$ is true, then both $\phi$ and $\psi$ are true, which by induction means that ${}^*\phi$ and ${}^*\psi$ are true. Therefore ${}^*\phi \bigwedge {}^*\psi$ is true, so $\phi \bigwedge \psi \to^* (\phi \bigwedge \psi)$. Conversely, if $\phi \bigwedge \psi$ is false, then one or both of $\phi$ and $\psi$ is false, which means that one or both of ${}^*\phi$ and ${}^*\psi$ is false, so ${}^*\phi \bigwedge {}^*\psi$ must be false. Therefore $\neg(\phi \bigwedge \psi) \to \neg^*(\phi \bigwedge \psi)$, so $\phi \bigwedge \psi \leftrightarrow^* (\phi \bigwedge \psi)$.
- Consider the sentence $\phi \bigvee \psi$, and its $*$-transform ${}^*(\phi \bigvee \psi) = {}^*\phi \bigvee {}^*\psi$. If $\phi \bigvee \psi$ is true, then at least one of $\phi$ and $\psi$ is true, which by induction means that at least one of ${}^*\phi$ and ${}^*\psi$ is true. Therefore ${}^*\phi \bigvee {}^*\psi$ is true, so $\phi \bigvee \psi \to^* (\phi \bigvee \psi)$. Conversely, if $\phi \bigvee \psi$ is false, then both $\phi$ and $\psi$ are false, which means that both ${}^*\phi$ and ${}^*\psi$ are false, so ${}^*\phi \bigvee {}^*\psi$ must be false. Therefore $\neg(\phi \bigvee \psi) \to \neg^*(\phi \bigvee \psi)$, so $\phi \bigvee \psi \leftrightarrow {}^*(\phi \bigvee \psi)$.
- Consider the sentence $\neg\phi$ and its $*$-transform ${}^*(\neg\phi) = \neg{}^*\phi$. If $\neg\phi$ is true, then $\phi$ is false, which by induction means that ${}^*\phi$ is false. Therefore ${}^*(\neg\phi)$ is true, so $\neg\phi \to {}^*(\neg\phi)$. Conversely, if $\neg\phi$ is false, then $\phi$ is true, so ${}^*\phi$ is true, therefore ${}^*(\neg\phi)$ is false. So $\neg(\neg\phi) \to \neg^*(\neg\phi)$. Therefore $\neg\phi \leftrightarrow \neg{}^*\phi$.
- Consider the sentence $\phi \to \psi$. This is equivalent to $\neg(\phi \bigwedge \neg\psi)$, so by combining the previous three parts of this proof, we get $\neg(\phi \bigwedge \neg\psi) \leftrightarrow {}^*(\neg(\phi \bigwedge \neg\psi))$, which is the same as $(\phi \to \psi) \leftrightarrow {}^*(\phi \to \psi)$
- Consider the sentence $\phi \leftrightarrow \psi$. This is equivalent to $(\phi \to \psi) \bigwedge (\psi \to \phi)$, so by the previous four parts of this proof, we can conclude $(\phi \leftrightarrow \psi) \leftrightarrow {}^*(\phi \leftrightarrow \psi)$.
- For the quantifiers, observe that $(\exists x)(\phi)$ is equivalent to $\neg(\forall x)(\neg\phi)$. Because of this, we will prove the transfer principle for both quantifiers together, by proving the existential transfer from $\mathbb{R}$ to ${}^*\mathbb{R}$, the universal transfer from ${}^*\mathbb{R}$ to $\mathbb{R}$, and combining the results and the above fact to complete the proof.
  - Let $A$ be a subset of $\mathbb{R}$, and consider the sentence $(\exists x \in A)(\phi)$ and its $*$-transform $(\exists x \in {}^*A)({}^*\phi)$. If $\phi$ does not contain $x$ as a free variable, then $(\exists x \in A)(\phi)$ is true whenever $\phi$ itself is true, so the transfer principle clearly holds. If $\phi$ contains $x$ as a free variable, then if $(\exists x \in A)(\phi)$ is true, there is some constant $s \in A$ such that $\phi(s)$ is true, therefore ${}^*(\phi(s)) = {}^*\phi({}^*s) = {}^*\phi(s)$ is true. But if $s \in A$, then the $*$-transform of $s$ maps to the equivalence class of the constant sequence $s, s, s, s, \dots$, which by the definition of subset extension is contained in ${}^*A$. Therefore $(\exists x \in A)(\phi) \to (\exists x \in {}^*A)({}^*\phi)$.
  - Let ${}^*A$ be the extension of some subset $A$ of $\mathbb{R}$, and consider the sentence $(\forall x \in {}^*A)({}^*\psi)$, where ${}^*\psi$ is the $*$-transform of some first order sentence $\psi$ in $\mathbb{R}$. Then extended set ${}^*A$ contains the $*$-transforms of all elements of the original set $A$ as constant sequences. And because $\psi(x)$ is true for all elements $x$ of ${}^*A$, $\psi(x)$ is true in particular for those elements $x$ which are $*$-transforms of elements of $A$. But for

these elements, ${}^*\psi(x) \to \psi(x)$, so this implies $(\forall x \in {}^*A)({}^*\psi) \to (\forall x \in A)(\psi)$.
- Consider the previous result applied to the sentence $\neg\phi$. Then $(\forall x \in {}^*A)(\neg{}^*\phi) \to (\forall x \in A)(\neg\phi)$. We've already shown that if ${}^*\psi \to \psi$, then $\neg{}^*\psi \to \neg\psi$. Applying this to the above sentence gives us $\neg(\forall x \in {}^*A)(\neg{}^*\phi) \to \neg(\forall x \in A)(\neg\phi)$. However, $\neg(\forall x \in A)(\neg\phi)$ is equivalent to $(\exists x \in A)(\phi)$. Therefore we have $(\exists x \in {}^*A)({}^*\phi) \to (\exists x \in A)(\phi)$. Combining this with the previous result on the existential quantifier proves $(\exists x \in {}^*A)({}^*\phi) \leftrightarrow (\exists x \in A)(\phi)$
- Consider the previous results applied to the sentence $\neg\phi$. Then we have $(\exists x \in A)(\neg\phi) \to (\exists x \in {}^*A)(\neg{}^*\phi)$. This implies $\neg(\exists x \in A)(\neg\phi) \to \neg(\exists x \in {}^*A)({}^*\neg\phi)$. But $\neg(\exists x \in A)(\neg\phi)$ is equivalent to $(\forall x \in A)(\phi)$, so this proves that $(\forall x \in A)(\phi) \to (\forall x \in {}^*A)({}^*\phi)$. Combining this with our previous result on the universal quantifier gives us $(\forall x \in A)(\phi) \leftrightarrow (\forall x \in {}^*A)({}^*\phi)$

□

4.4. **Using the Transfer Principle.** Once we familiarize ourselves with the transfer principle, we will see that it is mostly unnecessary to do any ultrafilter calculations on ${}^*\mathbb{R}$, or even think of ${}^*\mathbb{R}$ in terms of its ultrapower construction. Consider, for example, our proof that ${}^*\mathbb{R}$ is a field. To do so, we considered each element of ${}^*\mathbb{R}$ as a sequence in $\mathbb{R}^{\mathbb{N}}$, and rigorously tested and confirmed each field property. But observe that the axioms of a totally ordered field are all first order definable:

- Associativity:
  $(\forall x)(\forall y)(\forall z)(x + (y + z) = (x + y) + z)$
  $(\forall x)(\forall y(\forall z)(x \cdot (yz) = (xy) \cdot (z))$
- Commutativity:
  $(\forall x)(\forall y)(x + y = y + x)$
  $(\forall x)(\forall y)(x \cdot y = y \cdot x)$
- Distributivity
  $(\forall x)(\forall y)(\forall z)(x \cdot (y + z) = x \cdot y + x \cdot z)$
- Existence of Identities
  $(\exists x)\left((0 = x) \bigwedge (\forall y)(x + y = y)\right)$
  $(\exists x)\left((1 = x) \bigwedge ((\forall y)(x \cdot y = y)\right)$
- Existence of Inverses
  $(\forall x)(\exists y)(x + y = 0))$
  $(\forall x)(x \neq 0 \to \exists y)(x \cdot y = 1)$
- Total Ordering
  $(\forall x)(\forall y)(((x \leq y) \bigwedge (y \leq x)) \to (x = y))$
  $(\forall x)(\forall y)(\forall z)(((x \leq y) \bigwedge (y \leq z)) \to (x \leq z))$
  $(\forall x)(\forall y)((x \leq y) \bigvee (y \leq x))$

So, instead of explicitly proving distributivity, constructing inverses, and so on, we can simply take the $*$-transforms of the above list of first-order sentences, which we now know to be true by the transfer principle. We have thus proven that ${}^*\mathbb{R}$ is a totally ordered field without ever considering ${}^*\mathbb{R}$ as an ultrapower of $\mathbb{R}$, nor even doing a single ultrafilter calculation.

This method of transferring properties from $\mathbb{R}$ to ${}^*\mathbb{R}$ is very useful for proving that the familiar theorems of standard analysis hold in a nonstandard setting. But remember that the transfer principle goes both ways: we can prove theorems about

$\mathbb{R}$ by first proving some fact about ${}^*\mathbb{R}$, formulating this fact as the $*$-transform of some first-order sentence in $\mathbb{R}$, and "un-transforming" the sentence in ${}^*\mathbb{R}$ to get the desired result in $\mathbb{R}$. This is the great strength of nonstandard analysis, because very often a proof will be more intuitive and easier to construct in ${}^*\mathbb{R}$. To illustrate this, let's prove something strong about $\mathbb{R}$, but for which the proof using nonstandard techniques is much more intuitively clear. A good example is the Cauchy completeness of $\mathbb{R}$, which states that every Cauchy sequence converges. We will now use corollary 2.12, which states that every finite hyperreal is infinitesimally close to a unique real number, to give a very intuitively clear proof of the Cauchy completeness of $\mathbb{R}$.

**Theorem 4.5.** *Every Cauchy sequence in $\mathbb{R}$ converges.*

*Proof.* Let $s : \mathbb{N} \to \mathbb{R}$ be a real-valued Cauchy sequence. Observe that the property of being Cauchy is first-order defineable:

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

Since $s$ is Cauchy this sentence is true, which means that its $*$-transform,

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

is also true (we define the $*$-extension of a sequence the same way we defined the $*$-extension of a function whose domain is a proper subset of $\mathbb{R}$). We know that there exists some $k \in {}^\sigma\mathbb{N}$ such that for all $n, m \in {}^*\mathbb{N}$, $|{}^*s_n - {}^*s_m| <_{\mathcal{U}} 1$. Now, let $N \in {}^*N \backslash {}^\sigma\mathbb{N}$ be any infinite hypernatural. Clearly, $N, k \geq_{\mathcal{U}} k$, so by the $*$-transform of the fact that $s$ is Cauchy, we have

$$|{}^*s_k - {}^*s_N| <_{\mathcal{U}} 1$$

Since ${}^*s_k$ is finite (because $k$ is a standard natural number), this implies that ${}^*s_N$ is a finite hyperreal, and by lemma 2.12 must be infinitesimally close to a unique real number $L \in {}^\sigma\mathbb{R}$. Our claim is that the sequence $s$ converges to $L$ (note that while $s$ converges to $L$, ${}^*s$ converges to the constant sequence corresponding to $L$). To prove this, take any $\epsilon > 0$. We know that there exists some $j_\epsilon \in \mathbb{N}$ such that for all $m, n \geq j_\epsilon$, $|s_n - s_m| < \epsilon$. But by transfer this gives us $|{}^*s_n - {}^*s_N| <_{\mathcal{U}} \epsilon$ whenever $n \geq_{\mathcal{U}} j_\epsilon$, since we always have $N \geq_{\mathcal{U}} j_\epsilon$ for infinite $N$. So, let $n$ be any standard hypernatural greater than $j_\epsilon$. Then we have

$$\begin{aligned}
|{}^*s_n - L| &= |({}^*s_n - {}^*s_N) + ({}^*s_N - L)| \\
|{}^*s_n - L| &\leq_{\mathcal{U}} |{}^*s_n - {}^*s_N| + |{}^*s_N - L| \\
|{}^*s_n - L| &<_{\mathcal{U}} \epsilon + \delta
\end{aligned}$$

Here, $\delta$ is some infinitesimal number, since ${}^*s_N \approx L$. However, $|{}^*s_n - L|$ must be a standard real number, and since $\delta$ is infinitesimal, this implies that $|{}^*s_n - L| <_{\mathcal{U}} \epsilon$ (since a real number $x$ can't be less than a real number $a$ plus an infinitesimal $\delta$ unless it is also less than $a$ itself). Thus we have proven that, given any positive $\epsilon$, there exists a $j_\epsilon$ such that the sentence

$$(\forall n \in {}^*\mathbb{N})[(n \geq_{\mathcal{U}} j_\epsilon) \to \ (|{}^*s_n - L| <_{\mathcal{U}} \epsilon)]$$

is true. If we "de-transform" each such sentence back to $\mathbb{R}$, we have proven that $s_n$ converges to $L$. $\square$

Thus, by employing the transfer principle, we were able to prove something important about $\mathbb{R}$ using a relation which holds only in ${}^*\mathbb{R}$. In the next section, we will use the transfer principle extensively to discuss various formulations of continuity, differentiability, and convergence, and to prove the main theorems of single-variable calculus.

## 5. Continuity, Differentiability and Convergence

In this section, we will consider various definitions of continuity, differentiability, and convergence and prove their equivalence. We will then use these new definitions to prove some of the main theorems of single-variable calculus: the intermediate value theorem, critical point theorem, extreme value theorem, and Bolzano-Weierstrass theorem.

### 5.1. **Continuity.**

In the first section of this paper, we gave a definition of "infinitesimal continuity" which we claimed was equivalent to the standard $\epsilon$-$\delta$ definition of continuity. Now that we have the transfer principle, we can prove their equivalence. First, however, we will introduce some notation. We will refer to $\epsilon$-$\delta$ continuity at a point $c \in \mathbb{R}$ as $SC(c)$ (for standard continuity). Formally, a function $f : \mathbb{R} \to \mathbb{R}$ satisfies $SC(c)$ if and only if the sentence

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

is true. Similarly, we will refer to infinitesimal continuity at a point $c \in {}^*\mathbb{R}$ as $IC(c)$ (this $SC/IC$ notation comes from [2]). Recall that a function ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ satisfies $IC(c)$ if and only if ${}^*f(c + \epsilon) \approx {}^*f(c)$ for all infinitesimal $\epsilon \in \vartheta$. We will now prove the equivalence of these definitions of continuity.

**Theorem 5.1.** *A function $f : \mathbb{R} \to \mathbb{R}$ satisfies $SC(c)$ for some point $c \in \mathbb{R}$ if and only if its extension ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ satisfies $IC(c)$.*

*Proof.* Suppose first that $f$ satisfies $SC(c)$. Then for any positive $\epsilon \in \mathbb{R}^+$, there exists a $\delta \in \mathbb{R}^+$ such that

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

is true. Taking the $*$-transform of this sentence gives us

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

But if we pick any $x \approx c$, then we must have $|x - c| <_{\mathcal{U}} \delta$, since $\delta \in {}^\sigma\mathbb{R}$, which means that $|{}^*f(x) - {}^*f(c)| <_{\mathcal{U}} \epsilon$ for every $\epsilon \in {}^\sigma\mathbb{R}^+$. Therefore $|{}^*f(x) - {}^*f(c)|$ must be infinitesimal, so $x \approx c$ implies ${}^*f(x) \approx {}^*f(c)$. Therefore ${}^*f$ satisfies $IC(c)$.

Conversely, suppose that ${}^*f$ satisfies $IC(c)$, and pick any real $\epsilon \in \mathbb{R}^+$. Then if $\delta$ is some infinitesimal number, $|x - c| <_{\mathcal{U}} \delta$ implies $x \approx c$, so ${}^*f(x) \approx {}^*f(c)$ by $IC(c)$. But since $\epsilon$ is real, this implies that $|{}^*f(x) - {}^*f(c)| <_{\mathcal{U}} \epsilon$. So, given this $\epsilon$, the sentence

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

is true. But this is the $*$-transform of

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

which must therefore also be true. Thus, for any positive real $\epsilon$, we have demonstrated the existence of a positive real $\delta$ such that $|x - c| < \delta \to |f(x) - f(c)| < \epsilon$. This means that $f$ satisfies $SC(c)$. $\square$

We will now use this to give a very intuitively appealing proof of the intermediate value theorem, by dividing the interval into infinitely many infinitesimal subintervals.

**Theorem 5.2.** *(Intermediate Value Theorem) Let $f : [a, b] \to \mathbb{R}$ be continuous, and suppose $f(a) < f(b)$. Then for any $f(a) < c < f(b)$, there exists some $x \in [a, b]$ such that $f(x) = c$.*

*Proof.* For all $n \in \mathbb{N}$, consider a partition of the interval $[a, b]$ into $n$ subintervals of equal width. Thus each subinterval is of the form

$$[p_{k-1}, p_k] = \left[a + \frac{k(b-a)}{n}, a + \frac{(k+1)(b-a)}{n}\right]$$

for $0 \leq k \leq n-1$. Consider the set of points $p_k$ such that $f(p_k) < c$. This set is finite and nonempty, so for each $n$ we can let $s_n$ be the maximal element of $\{p_k | f(p_k) < c\}$. Then, treating $s : \mathbb{N} \to [a, b]$ as a function embedded in our relational structure, we have the sentence

$$(\forall n \in \mathbb{N})[(a \leq s_n \leq b) \bigwedge (f(s_n) < c \leq f(s_n + (b-a)/n))]$$

We can then take the $*$-transform of this sentence and evaluate the function ${}^*s$ at some infinite hypernatural $N \in {}^*\mathbb{N} \backslash {}^\sigma\mathbb{N}$. Then $a \leq_{\mathcal{U}} s_N \leq_{\mathcal{U}} b$, so $s_N$ is limited, and is therefore infinitesimally close to a unique real number $L$. However, $N$ is infinite, so $\frac{b-a}{N}$ is infinitesimal, which means that $s_N \approx s_N + \frac{b-a}{N}$. And by theorem 5.1, this means that ${}^*f(s_N + (b-a)/N) \approx {}^*f(s_N) \approx f(L)$, since $f$ is continuous. But by transfer we still have

$$f(s_N) <_{\mathcal{U}} c \leq_{\mathcal{U}} f\left(s_N + \frac{b-a}{N}\right)$$

Since $c$ is sandwiched between $f(s_N)$ and $f(s_N+(b-a)/N)$, and $f(s_N+(b-a)/N) \approx f(s_N) \approx f(L)$, we must have $c \approx f(L)$. But $c$ and $f(L)$ are both real, which means that $c = f(L)$. Therefore $L \in [a, b]$ satisfies $f(L) = c$. □

Partitioning the interval into subintervals of equal size is a powerful trick in nonstandard calculus, since transferring the partition to ${}^*\mathbb{R}$ allows us to construct a partition consisting of infinitely many infinitesimally small subintervals. We will now use the same trick to prove the extreme value theorem.

**Theorem 5.3.** *(Extreme Value Theorem) If a function $f : [a, b] \to \mathbb{R}$ is continuous, then $f$ attains both a maximum and a minimum on $[a, b]$.*

*Proof.* As in the proof of the intermediate value theorem, for any $n \in \mathbb{N}$, partition the interval $[a, b]$ into $n$ subintervals of the form

$$[p_k, p_{k+1}] = \left[a + \frac{k(b-a)}{n}, a + \frac{(k+1)(b-a)}{n}\right]$$

for $0 \leq k \leq n-1$. The set $\{f(p_k) | 0 \leq k \leq n\}$ is finite and nonempty, so it has some maximum element $M$. We therefore let $s_n$ denote the element of $\{p_k | 0 \leq k \leq n\}$ such that $f(s_n) = M$. So, treating $s : \mathbb{N} \to [a, b]$ as a function embedded in our relational structure, we have the sentence

$$(\forall n \in \mathbb{N}) \Big[(a \leq s_n \leq b) \bigwedge ((\forall k \in \mathbb{N})(k \leq n \to f(a + k(b-a)/n) \leq f(s_n)))\Big]$$

We can then take the $*$-transform of this sentence and evaluate the function ${}^*s$ at some infinite hypernatural $N \in {}^*\mathbb{N} \backslash {}^\sigma\mathbb{N}$. By transfer, we have $a \leq_{\mathcal{U}} s_N \leq_{\mathcal{U}} b$, so $s_N$ is limited and hence infinitesimally close to a unique real number $d \in {}^\sigma\mathbb{R}$. Since $f$ is continuous, it follows that ${}^*f(s_N) \approx {}^*f(d)$.

Now, we claim that the infinite partition

$$P = \left\{ a + \frac{K(b-a)}{N} | K \in {}^*N \text{ and } K \leq_{\mathcal{U}} N \right\}$$

contains points infinitesimally close to all real numbers in ${}^*[a,b]$. To prove this, first observe that, in the standard context, for any $x \in [a,b]$ the definition of partition gives us

$$(\forall n \in \mathbb{N})(\exists k \in \mathbb{N}) \left[ (k \leq n) \bigwedge \left( a + \frac{k(b-a)}{n} \leq x \leq a + \frac{(k+1)(b-a)}{n} \right) \right]$$

If we take the $*$-transform of this sentence and consider our infinitesimal partition into $N$ subintervals, there must exist some hypernatural $K \in {}^*N$ such that $K <_{\mathcal{U}} N$ and

$$x \in \left[ a + \frac{K(b-a)}{N}, a + \frac{(K+1)(b-a)}{N} \right]$$

Since this interval has infinitesimal width $(b-a)/N$, it follows that $x \approx a + K(b-a)/N$, which is in the infinite partition described above. And again, applying the continuity of $f$ gives us

$${}^*f(x) \approx {}^*f\left( a + \frac{K(b-a)}{N} \right)$$

However, by transfer we must also have

$${}^*f\left( a + \frac{K(b-a)}{N} \right) \leq_{\mathcal{U}} {}^*f(s_N) \approx d$$

so combining these two expressions gives us

$${}^*f(x) \approx {}^*f\left( a + \frac{K(b-a)}{N} \right) \leq_{\mathcal{U}} {}^*f(s_N) \approx f(d)$$

Since ${}^*f(x)$ and $f(d)$ are real, it follows that for all $x \in [a,b]$, $f(x) \leq f(d)$. Therefore $f$ attains a maximum at $d \in [a,b]$. A similar proof follows for the minimum of $f$ on $[a,b]$. $\square$

5.2. **Differentiability.** The standard definition of differentiability is as follows: the function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if the limit

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

exists. If $f$ is differentiable, we call $f'(a)$ its derivative at $a$. This property, since it is defined by the existence of a limit, is first-order defineable:

$$(\forall \epsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+) \left[ (\forall h \in \mathbb{R}) \left( |h| < \delta \to \left| \frac{f(a+h) - f(a)}{h} - f'(a) \right| < \epsilon \right) \right]$$

We can now use this first-order definition of the derivative to prove the equivalence of our alternate definition of derivative with the standard definition given above.

**Theorem 5.4.** *The function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ with derivative $f'(a)$ if and only if its extension function ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ satisfies*

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

*for every infinitesimal value of $h$.*

*Proof.* Suppose that $f$ is differentiable at $a$ with derivative $f'(a)$, and suppose we fix some positive $\epsilon$ and the corresponding $\delta$ which satisfies the condition of the limit. Then we have the sentence

$$(\forall h \in \mathbb{R}) \left( |h| < \delta \to \left| \frac{f(a+h) - f(a)}{h} - f'(a) \right| < \epsilon \right)$$

Taking the $*$-transform of this sentence gives us

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

Since $\delta$ is a standard real, any infinitesimal value of $h$ satisfies $|h| <_{\mathcal{U}} \delta$, therefore

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

for every infinitesimal value of $h$. But this holds for all real values of $\epsilon$. So whenever $h$ is an infinitesimal, we have

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

Conversely, suppose ${}^*f$ satisfies the above property for every infinitesimal value of $h$. Then for any positive real value of $\epsilon$, the sentence

$$(\exists \delta \in {}^*\mathbb{R}^+) \left[ (\forall h \in {}^*\mathbb{R}) \left( |h| <_{\mathcal{U}} \delta \to \left| \frac{{}^*f(a+h) - {}^*f(a)}{h} - f'(a) \right| <_{\mathcal{U}} \epsilon \right) \right]$$

is true. But this is the $*$-transform of the sentence

$$(\exists \delta \in \mathbb{R}) \left( (|h| < \delta \to \left| \frac{f(a+h) - f(a)}{h} - f'(a) \right| < \epsilon \right)$$

which must therefore also be true. Thus, for every positive real $\epsilon$ we have demonstrated the existence of a corresponding $\delta$ satisfying the condition of the limit. Therefore

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

so $f$ is differentiable at $a$ with derivative $f'(a)$. □

The equivalence of these two definitions of derivate allows us to introduce a powerful notation (again borrowed from [1]), which had been used for centuries before the development of nonstandard analysis due to its intuitive appeal. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and let ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ be its extension. For any real $x$ and any infinitesimal $\Delta x$, we let

$$\Delta f(x, \Delta x) = {}^*f(x + \Delta x) - {}^*f(x)$$

In general we will omit the $(x, \Delta x)$ and write only $\Delta f$, whenever there is no chance for confusion. The strength of this notation is that, by theorem 5.4, a function $f$ is differentiable at $a \in \mathbb{R}$ if and only if

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

To demonstrate the power of this notation, we will now use it to give a very simple proof of an important theorem in single-variable calculus, the critical point theorem.

**Theorem 5.5.** *(Critical Point Theorem) Suppose $f : (a, b) \to \mathbb{R}$ is differentiable. If $f$ achieves a maximum or minimum at some point $c \in (a, b)$, then $f'(c) = 0$.*

*Proof.* Suppose $f$ achieves a maximum at $x \in \mathbb{R}$. Then for some fixed $\epsilon \in \mathbb{R}^+$, the sentence

$$(\forall c \in \mathbb{R})(|c - x| < \epsilon \to f(c) \leq f(x))$$

is true. Taking its $*$-transform gives us

$$(\forall c \in {}^*\mathbb{R})(|c - x| <_{\mathcal{U}} \epsilon \to {}^*f(c) \leq_{\mathcal{U}} {}^*f(x))$$

But $\epsilon$ is real, so for every infinitesimal $\Delta x$, we must have $|(x + \Delta x) - x| <_{\mathcal{U}} \epsilon$. Therefore

$${}^*f(x + \Delta x) \leq_{\mathcal{U}} {}^*f(x) \;\Rightarrow\; \Delta f \leq_U 0$$

for all infinitesimal $\Delta x$. Now, consider a positive infinitesimal $\Delta x_1$ and a negative infinitesimal $\Delta x_2$. Since $f$ is differentiable at $c$, we must have $\frac{\Delta f}{\Delta x} \approx f'(x)$ for any infinitesimal $\Delta x$. Combining this with the above inequality gives us

$$f'(x) \approx \frac{\Delta f}{\Delta x_1} \leq_{\mathcal{U}} 0 \leq_{\mathcal{U}} \frac{\Delta f}{\Delta x_2} \approx f'(x)$$

Since $f'(x)$ is real, this implies $f'(x) = 0$. □

These three theorems form a good basis for the rest of differential calculus, as many other important theorems in calculus follow as corollaries of these theorems; Rolle's Theorem follows from the critical point theorem and the extreme value theorem, the mean value theorem is a corollary of Rolle's Theorem, and so on. Because the proofs of these corollaries follow so directly and do not require any further use of the transfer principle, we omit them in this paper.

5.3. **Sequences and Convergence.** In this last section, we give an alternate definition of convergence for sequences and prove its equivalence with the standard definition. We will then use a similar definition of limit point to give a simple proof of the Bolzano-Weierstrass theorem in $\mathbb{R}$. We begin by stating this new definition of convergence.

**Theorem 5.6.** *A sequence $s : \mathbb{N} \to \mathbb{R}$ converges to some real number $L$ if and only if ${}^*s_N \approx L$ for all infinite hypernaturals $N \in {}^*\mathbb{N} \backslash {}^\sigma\mathbb{N}$.*

*Proof.* Suppose first that $s$ converges to $L$. For every $\epsilon > 0$, there exists some corresponding $m \in \mathbb{N}$ such that

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

We can then take the $*$-transform of this sentence. For any infinite $N$, we clearly have $N >_{\mathcal{U}} m$, so by transfer $|{}^*s_N - L| <_{\mathcal{U}} \epsilon$ holds. But this holds for every real $\epsilon$, therefore $s_N \approx L$ for any infinite $N$.

Conversely, suppose that ${}^*s_N \approx L$ for any infinite hypernatural $N$, and fix some real $\epsilon \in {}^\sigma\mathbb{R}^+$. Then the sentence

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

is true, since any infinite hypernatural $N$ will satisfy the given property. But this is the $*$-transform of the sentence

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

which by transfer must also be true. Since this holds for every $\epsilon \in \mathbb{R}^+$, $s$ converges to $L$. □

To prove the Bolzano-Weierstrass, we will first need the first-order definition of a limit point. Recall that $L \in \mathbb{R}$ is a limit point of a sequence $s : \mathbb{N} \to \mathbb{R}$ if every $\epsilon$-neighborhood of $L$ contains infinitely many points of $s$. We can express this by the sentence

$$(\forall \epsilon \in \mathbb{R}^+)(\forall m \in \mathbb{N})(\exists n \in \mathbb{N}) \left(n > m \bigwedge |s_n - L| < \epsilon\right)$$

We will now use this to prove a theorem of which Bolzano-Weierstrass is an immediate corollary.

**Theorem 5.7.** *Let $s : \mathbb{N} \to \mathbb{R}$ be a real valued sequence, and let $L$ be a real number. Then $L$ is a limit point of $s$ if and only if ${}^*s_N \approx L$ for some infinite $N \in {}^*N \backslash {}^\sigma\mathbb{N}$.*

*Proof.* Suppose first that ${}^*s_N \approx L$ for some infinite $N$. Then for every standard $\epsilon \in {}^\sigma\mathbb{R}^+$ and $m \in {}^\sigma\mathbb{N}$, there exists an $n \in {}^*\mathbb{N}$ such that $n >_{\mathcal{U}} m$ and $|{}^*s_n - L| <_{\mathcal{U}} \epsilon$-namely, our chosen infinite $N$. Therefore the sentence

$$(\exists n \in {}^*N) \left(n >_{\mathcal{U}} m \bigwedge |{}^*s_n - L| <_{\mathcal{U}} \epsilon\right)$$

is true. But this is the $*$-transform of

$$(\exists n \in N) \left(n > m \bigwedge |s_n - L| < \epsilon\right)$$

which by transfer is also true. Thus for every positive $\epsilon$ and natural number $m$ we have demonstrated the existence of a natural number $n > m$ such that $|s_n - L| < \epsilon$. Therefore $L$ is a limit point of $s$.

Conversely, suppose that $L$ is a limit point of $s$, so the sentence

$$(\forall \epsilon \in \mathbb{R}^+)(\forall m \in \mathbb{N})(\exists n \in \mathbb{N}) \left(n > m \bigwedge |s_n - L| < \epsilon\right)$$

is true. We can take the $*$-transform of this sentence and consider some infinite $m \in {}^*\mathbb{N} \backslash {}^\sigma\mathbb{N}$ and some infinitesimal $\epsilon \in \vartheta$. By the transfer principle, there must exist some $N \in {}^*\mathbb{N}$ such that $N > m$ and $|{}^*s_N - L| <_{\mathcal{U}} \epsilon$. Since $m$ is infinite, $N$ must also be infinite, and since $\epsilon$ is infinitesimal, it follows that

$${}^*s_N \approx L$$

for an infinite hypernatural $N$. □

To see why the Bolzano-Weierstrass theorem follows immediately from this, observe that, if a sequence $s$ is bounded, then ${}^*s_N$ must be finite, even for infinite values of $N$. But by lemma 2.12, this means that for every infinte hypernatural $N$, ${}^*s_N \approx L$ for a unique value of $L$. Thus, we have

**Corollary 5.8.** *(Bolzano-Weierstrass Theorem) Every bounded, real-valued sequence has a limit point.*

## 6. Conclusion

To conclude, we will briefly consider how one might extend the transfer principle to higher order sentences. As we saw, we were able to do quite a bit dealing only with first-order sentences. Some important properties, however, are not first-order defineable. Consider, for example, the property of Dedekind completeness, which states that every bounded, nonempty subset of $\mathbb{R}$ has a least upper bound. We cannot yet transfer this property to $^*\mathbb{R}$, as it is not possible to formulate a first-order sentence with a variable ranging over $\mathcal{P}(\mathbb{R})$, the power set of $\mathbb{R}$.

We will need to work on a structure even larger than our relational structure- in fact one that contains our relational structure. Our structure must contain all of analysis, from the natural numbers to the reals, complex numbers, functions, relations, Euclidean spaces- anything that can be described in real analysis. We will construct this structure as follows. Let $X_0 = \mathbb{N}$ denote the base set. We then inductively define $X_n = \mathcal{P}(X_{n-1})$; that is, each set is the power set of the previous set. Our *superstructure* will be the collection

$$\mathcal{X} = \bigcup_{n=0}^{\infty} X_n$$

along with notions of equality and membership among its elements (set elements of $\mathcal{X}$ are called *entities*, while elements of the base set $X_0$ are called individuals). The important characteristic of the superstructure is that it contains the set theory of its own elements [2]. Although we will not prove these properties here, it is not difficult to show that if $x, x_1, \ldots, x_n$ are entities of $\mathcal{X}$, then

- $y$ is an entity whenever $y \in x$,
- $y$ is an entity whenever $y \subseteq x$,
- $\mathcal{P}(x)$ is an entity,
- $\{x_1, \ldots, x_n\}$ is an entity,
- $(x_1, \ldots, x_n)$ is an entity.

It is not necessarily obvious that $\mathcal{X}$ does in fact contain all of analysis. We cannot explicitly prove that "every object in analysis" is contained in $\mathcal{X}$ here, but we can motivate it with the following observations. First, $\mathcal{X}$ contains the natural numbers as the base set. It also contains all ordered pairs of natural numbers, from which we can form the integers. It contains all ordered pairs of integers, from which we can form the rationals. It must also contain all subsets of the rationals, from which we can form Dedekind cuts to construct the reals, and we can take ordered pairs of reals to form the complex numbers. We can also form functions and relations as ordered $n$-tuples of reals. For this reason, the superstructure can be thought of as a "standard" model of analysis [2].

The idea behind a higher-order transfer is to embed this standard model $\mathcal{X}$ in a larger superstructure $\mathcal{Y}$. We call this embedding the $*$-transform (which looks very familiar), and our wish is for the $*$-transform to preserve all finitary set operations- union, intersection, complementation, etc. An embedding satisfying these properties is called a superstructure monomorphism. And, much in the same way we formed $*$-transforms of first order sentences by first taking atomic formulae in $\mathfrak{R}$ and extending them to $^*\mathfrak{R}$, we can create a method of transforming higher order sentences whose variables range over entities in $\mathcal{X}$ to equivalent sentences in $\mathcal{Y}$. The full version of the transfer principle then states that a sentence $\phi$ in $\mathcal{X}$ with bounded formalization (i.e. with variables ranging over entities in $\mathcal{X}$) is true if and only if its $*$-transform ${}^*\phi$ in $\mathcal{Y}$ is true. This is the full version of the transfer principle, and it is extraordinarily powerful in nonstandard analysis. With higher order transfer at our disposal, we can do nonstandard complex analysis, nonstandard fourier series, nonstandard p-adic analysis- anything contained in the superstructure $\mathcal{X}$ can be extended to a nonstandard version, and this nonstandard version can be used extensively to prove theorems in the standard version, as we saw in this paper.

**Acknowledgments.** I would like to thank my mentors Asaf Hadari and Rita Jiminez Rolland for all their help in writing and editing this paper.

## References

[1] Goldblatt, R.: *Lectures on the Hyperreals- An Introduction to Nonstandard Analaysis*, Springer (1998)

[2] Stroyan, K.D. and Luxemburg, W.A.J.: *Introduction to the Theory of Infinitesimals*, Academic Press (1976)