UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2014:13

Limitless Analysis

Daniel Ahlsén

Examensarbete i matematik, 15 hp
Handledare och examinator: Vera Koponen
Maj 2014

Department of Mathematics
Uppsala University

Limitless Analysis

Daniel Ahlsén

May 2014

Abstract

We formulate a theory of analysis without limit concepts, using a model theoretic construction known as the ultraproduct. It is exemplified by proofs of some classic results in analysis. Topics covered are sequences, series, limits, continuity and sequences of functions.

Contents

1 Introduction

1.1 The Modern Approach to Analysis

The modern view of analytical concepts is that they are *limiting processes*: processes which have no well-defined endpoint or stop. The process of adding one term after another from some infinite sequence is one example, the derivative of a function at a point is another; it is the limit of average velocity of a function over successively smaller intervals around it.

Many will agree that the heart of analysis still lies in functions from a subset of $\mathbb{R}$ to $\mathbb{R}$. The modern concepts in analysis, such as differential forms and Banach spaces, are generalizations of concepts from real analysis. Derivative is rate of change; integrals correspond to area under the graph, and so on. The common denominator is the use of limits.

The reason for taking a limit often is that the standard operations in $\mathbb{R}$ does not quite suffice. As an example, take a series $\sum_i a_i$ and suppose we wish assign to it a value. If the sequence is finite, there is no problem since the expression $\sum_i^n a_i$ is computable. However, if the sequence is infinite no straightforward computation exists which solves our problem directly, since addition is only defined for finitely many terms.

Limits solves this problem by treating addition of any finite number of terms as an approximation to the complete summation. If the approximations comes arbitarily close to some value as the number of terms grows arbitarily high, we say that it converges and assign a value the infinite sum. Otherwise it diverges, and we cannot assign to it a real number.

The limit approach to analysis is popular for several reasons. One of them that it does not demand the existence of other mathematical systems than $\mathbb{R}$. In this formulation, analysis with limits only uses real numbers and functions from the reals to itself.

Normally, when mathematicians discover that the number systems they work in does not encompass everything which they are interested in, they find a larger system which contains the original system as a substructure. For example, when $\mathbb{N}$ did not contain solutions to the equation $x + 1 = 0$, they invented $\mathbb{Z}$. When $\mathbb{Z}$ could not solve $2x - 1 = 0$, $\mathbb{Q}$ was invented, and so on.

What keeps mathematicians from doing so when it comes to analysis, seems to be the implausability of an enlargement existing. For such a system to be of interest, it must contain numbers ”infinitely close” to each other, as well as numbers larger than any existing real number. Many view these properties as contradictory, or at least not intuitive.

Historically, this was not the case. The foundations of calculus laid down by Newton and Leibniz involved infinitesimals, at least implicitly, and for 150 years the modern concept of limits did not exist. In those days, analysis was done exclusively using infinitesimals. Even today some mathematicians have not given up the tradition of using infinitesimals in analysis, especially in mathematics connected to physics.

Traces of it is found in the traditional notation dy/dx for the derivative of y with respect to x. Here, dy stands for the change in y over some infinitesimally small distance dx, and their quotient is the derivative.

Still, most of todays mathematicians thinks of analytical concepts in terms of limits, and not as infinitesimals. Those who utilize infinitesimal, use them as a tool and try to replace them with limits at some point. Many feel that infinitesimals are a "bad" way of doing analysis, a rigorous proof must involve a ε-δ argument at some point.

Advances in mathematical logic proves that this is false. Abraham Robinsons work in the 1960's exhibits an enlargement of $\mathbb{R}$ containing infinitesimals, using a model theoretic construction called the ultraproduct, and shows how limits and related concepts in $\mathbb{R}$ naturally translates into his new system [1].

The number system he developed was named *the hyperreals* (here denoted $\mathcal{R}$, in other places $\mathbb{R}^*$ or ${}^*\mathbb{R}$), and the associated theory was called *nonstandard analysis*.

1.2 Outline and Prerequisites

This thesis is divided into two chapters. First, we cover the various set-theoretic and model theoretic constructions (such as filters, boolean extensions, ultraproducts and so on) which form the foundations of the hyperreals. We use this to construct the hyperreals and formulate the Transfer Principle. This becomes our main tool for proving that nonstandard and standard real analysis are equivalent. At the end of the first chapter we take a look at the structure of the hyperreals, and how it relates to the reals.

The second chapter investigates a few topics of analysis, and see how infinitesimals can be used to formulate them without limits. We cover sequences, series, limits, continuity and function sequences. Due to size constraints, we do not go into differentiation, integration and other subject normally found in a book on calculus. The reader interested in this are invited to read the first half of Goldblatt's book *Lectures on the Hyperreals* [2], from which the majority of proofs in Chapter 3 are derived.

We assume that the reader has basic knowledge of the model theory of first order logic. In particular, syntax, semantics, substructures and embeddings are important. All this material is found in Rothmaler's *Introduction to Model Theory* [3], mostly in chapters

1-3. For the reader completely new to logic, Hedman's excellent book *A First Course in Logic* is recommended [4].

It is convenient to have knowledge of basic real analysis, most importantly the role of the Dedekind completeness property. The construction of $\mathbb{R}$ as equivalence classes of Cauchy sequences of $\mathbb{Q}$ is similar to the ultraproduct, and is recommended for those wishing to get a firmer grasp of the ultraproduct.

1.3 Terminology and Notation

For models of first order logic, we use letters in calligraphic style, such as $\mathcal{M}$ and $\mathcal{N}$. Their universes are denoted by the same letter in latin style, in our example M and N. First order languages are considered triples $L = (C_L, F_L, R_L)$ of sets where C_L are the constant symbols, F_L are the function symbols and R_L are the relation symbols. Every language is assumed to contain at least one binary relation symbol =, which is interpreted as identity.

The letters $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$ and $\mathbb{N}$ represents the reals, rational numbers, integers and natural numbers, while the letter $\mathcal{R}$ denotes the hyperreals. The letter $\mathbb{R}_+$ denotes the set of strictly positive reals. To separate $\mathbb{R}$ and $\mathcal{R}$, we use greek letters α and β for arbitary hyperreals and latin letters a and b for arbitary reals. The only exception from this rule is the use of ε and δ to represent small real quantities, which is traditional.

The reader must be attentive to what a variable represents, since in many proofs real numbers are first considered as reals and then as hyperreals. In this case, the variable refering to the real number will not change.

Finally, some set theory is needed. If A is a set, the *power set* $\mathcal{P}(A)$ of A is the set of subsets of A, i.e

$$\mathcal{P}(A) = \{X : X \subset A\}.$$

Also, if $A \subset U$, the *complement* A^c of A is the set

$$A^c = U \setminus A = \{x \in U : x \notin A\}.$$

Usually, what U is will be clear from the context.

2 Construction of the Hyperreals

One of the first conclusions drawn from the Compactness Theorem of first order logic is that there exists models of the (first order) theory of the reals which contains *infinitesimals*; elements which are smaller than $1/n$ for any $n \in \mathbb{N}$ yet different from 0.

Unfortunately, the compactness theorem gives us no clue to what such a model looks like, since it only proves its existence. Thankfully, a model theoretic construction called the *ultraproduct* gives us an explicit model displaying these characteristics (in fact, it can be used to prove the Compactness Theorem, as Rothmaler does in [3], Chapter 4). Here, we detail this construction and investigates its structure.

2.1 Filters and Ultrafilters

The ultraproduct is a product of models. That in itself is not very interesting. Our aim is to find an equivalence relation on the product, and define a "quotient model" which behaves nicely with regards to satisfiability.

Elements of the product are thought of as I-tuples of elements from the factors (where I is some set indexing them). It is natural to say that they are equal if they agree in a "large" subset of I. What we consider as a large subset is described by the following structure.

Definition 2.1.1. A *filter (on U)* is a set $\mathcal{F} \subset \mathcal{P}(U)$ such that

(i) $\emptyset \notin \mathcal{F}$

(ii) If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

(iii) If $A \in \mathcal{F}$ and $A \subset B \subset U$, then $B \in \mathcal{F}$.

Remark. The points (i) and (iii) in this definition are pretty clear. Obviously, $\emptyset$ is not a large subset and subsets of U which contain large subsets are themselves large. Condition (ii) might seem unintuitive, but closure under finite intersections guarantees that the relation we later define is an equivalence relation.

Definition 2.1.2. A subset $A \subset U$ is called *cofinite* if A^c is finite. The set of cofinite subsets of U is denoted U^{co}. When $U = \mathbb{N}$, the filter is called the *Frechét filter*.

Lemma 2.1.1. *For any infinite set U, the set U^{co} constitutes a filter on U.*

Proof. (i) $\emptyset^c = U$ is infinite, so $\emptyset$ is not cofinite.

(ii) Let A and B be cofinite. Then $(A \cap B)^c = A^c \cup B^c$ is finite. Hence $A \cap B$ is cofinite.

(iii) Let A be cofinite, and $A \subset B \subset U$. Then $B^c \subset A^c$, so B^c is finite and B is cofinite.

□

Definition 2.1.3. An *ultrafilter* on U is a filter $\mathcal{F}$ on U such that for any $A \subset U$, either $A \in \mathcal{F}$ or $A^c \in \mathcal{F}$ (but not both).

Ultrafilters are maximal filters, in the sense that we cannot add any sets to them without destroying the filter structure.

Definition 2.1.4. If $x \in U$, we call $\mathcal{F}_x = \{A \subset U : x \in A\}$ the *principal filter (on U) generated by x.*

Lemma 2.1.2. *For any $x \in U$, the principal filter $\mathcal{F}_x$ is an ultrafilter.*

Proof. (i) $x \notin \emptyset$, so $\emptyset \notin \mathcal{F}_x$.

(ii) If $A, B \in \mathcal{F}_x$, then $x \in A \cap B$. Therefore $A \cap B \in \mathcal{F}_x$.

(iii) If $A \in \mathcal{F}_x$ and $A \subset B \subset \mathcal{F}_x$, we have $x \in A \subset B$. Thus $x \in B$, so $B \in \mathcal{F}_x$.

(iv) For any $A \subset U$, we have $x \in A$ or $x \in A^c$. Thus $A \in \mathcal{F}_x$ or $A^c \in \mathcal{F}_x$.

□

Sets of subsets in which finite intersections are nonempty play an important roles in mathematics, for example in general topology and logic.

Definition 2.1.5. A collection of subsets $\mathcal{H}$ has the *finite intersection property* if for any $A_1, \ldots, A_n \in \mathcal{H}$, the intersection $A_1 \cap \cdots \cap A_n$ is nonempty.

Lemma 2.1.3. *Any filter $\mathcal{F}$, and in particular U^{co}, has the finite intersection property.*

Proof. A simple induction on the number of sets in the intersection yields that if $A_1, \ldots, A_n \in \mathcal{F}$ then $A_1 \cap \cdots \cap A_n \in \mathcal{F}$. It follows that their intersection is nonempty, since $\emptyset \notin \mathcal{F}$. □

So far, we have only encountered principal ultrafilters, which are rather mundane. The following construction is used to generate other, more interesting ultrafilters.

Definition 2.1.6. Suppose $\mathcal{H} \subset \mathcal{P}(U)$ has the finite intersection property. The set

$$\mathcal{F}^{\mathcal{H}} = \{A \subset U : B_1 \cap \cdots \cap B_n \subset A \text{ where } B_i \in \mathcal{H}\}$$

is called the *filter (on U) generated by $\mathcal{H}$.*

Lemma 2.1.4. *For any $\mathcal{H} \subset \mathcal{P}(U)$ with the finite intersection property, $\mathcal{F}^{\mathcal{H}}$ is a filter.*

Proof. (i) Let $A \in \mathcal{F}^{\mathcal{H}}$. By assumption, there exists sets $A_1, \ldots, A_n \in \mathcal{H}$ with nonempty intersection, such that $A_1 \cap \cdots \cap A_n \subset A$. Thus A is nonempty as well and $\emptyset \notin \mathcal{F}^{\mathcal{H}}$.

(ii) Let $A, B \in \mathcal{F}^{\mathcal{H}}$. By assumption, there exists $A_1, \ldots, A_n \in \mathcal{H}$ so that $A_1 \cap \cdots \cap A_n \subset A$ and $B_1, \ldots, B_m \in \mathcal{H}$ such that $B_1 \cap \cdots \cap B_m \subset B$ from $\mathcal{H}$. Then the intersection

$$A_1 \cap \cdots \cap A_n \cap B_1 \cap \cdots \cap B_m \subset A \cap B.$$

Consequently, $A \cap B \in \mathcal{F}^{\mathcal{H}}$.

(iii) Suppose $A \in \mathcal{F}^{\mathcal{H}}$, and $A \subset B \subset U$. Then there exists $A_1, \ldots, A_n \in \mathcal{H}$ so that $A_1 \cap \cdots \cap A_n \subset A$. Hence $A_1 \cap \cdots \cap A_n \subset B$ and $B \in \mathcal{F}^{\mathcal{H}}$.

□

Our next theorem on ultrafilters requires a result known as *Zorn's lemma*, which is equivalent to the Axiom of Choice. Consequently, it is independent of the other current axioms of set theory. To state Zorn's lemma, we make a few definitions.

Definition 2.1.7. A *partial order* is a set U with a binary relation $\leq$ such that

(i) $x \leq x$ for all $x \in U$,

(ii) if $x \leq y$ and $y \leq z$, then $x \leq z$.

(iii) if $x \leq y$ and $y \leq x$, then $x = y$.

If, in addition to this, $x \leq y$ or $y \leq x$ for every $x, y \in U$ we call U a *total order*.

Usually, total orders are seen as lying on a line, while the structure of partial orders are more treelike. Classic examples of total orders are $\mathbb{R}$ and $\mathbb{Q}$, and an example of a partial order is $\mathcal{P}(\mathbb{N})$ under set-inclusion.

Definition 2.1.8. Suppose $\mathcal{P}$ is a partial order. A *chain* $\mathcal{C}$ in $\mathcal{P}$ is a subset of $\mathcal{P}$ which is a total order.

Definition 2.1.9. A *maximal element* of a partial order $\mathcal{P}$ is an element $x \in \mathcal{P}$ such that for no $y \in \mathcal{P}$ different from x, we have $y \geq x$. An *upper bound* of a set $\mathcal{S} \subset \mathcal{P}$ is an element $B \in \mathcal{P}$ so that $x \leq B$ for every $x \in \mathcal{S}$.

Zorn's lemma connects all these concepts into one theorem, which is surprisingly useful.

Theorem 2.1.1. Zorn's lemma: *If every chain in a partial order $\mathcal{P}$ has an upper bound, then $\mathcal{P}$ contains a maximal element.*

The equivalence of the Axiom of Choice and Zorn's lemma is a standard result in axiomatic set theory and is found in every textbook on the subject, for example Moschovakis' *Notes on Set Theory* [5].

Theorem 2.1.2. *Every $\mathcal{H} \subset \mathcal{P}(U)$ with the finite intersection property is extendable to an ultrafilter.*

Proof. Consider the set $\mathcal{S}$ of filters $\mathcal{F}$ such that $\mathcal{F}^{\mathcal{H}} \subset \mathcal{F}$, partially ordered by set inclusion. Suppose $\mathcal{C} = (\mathcal{F}_i : i \in I)$ is a chain in $\mathcal{S}$. We prove that $\bigcup \mathcal{C} = \bigcup_{i \in I} \mathcal{F}_i$ is a filter, and thus an upper bound on $\mathcal{C}$.

(i) Clearly, $\emptyset \notin \mathcal{F}_i$ for any $i \in I$, so $\emptyset \notin \bigcup \mathcal{C}$.

(ii) Let $A, B \in \bigcup \mathcal{C}$. Then there are $i, j \in I$ such that $A \in \mathcal{F}_i$ and $B \in \mathcal{F}_j$, and since $\bigcup \mathcal{C}$ is a total order either $\mathcal{F}_i \subset \mathcal{F}_j$ or $\mathcal{F}_j \subset \mathcal{F}_i$. Without loss of generality, we can assume $\mathcal{F}_i \subset \mathcal{F}_j$. But then $A, B \in \mathcal{F}_j$, and since $\mathcal{F}_j$ is a filter, $A \cap B \in \mathcal{F}_j$. Hence $A \cap B \in \bigcup \mathcal{C}$.

(iii) Let $A \in \bigcup \mathcal{C}$, and $A \subset B \subset U$. Then $A \in \mathcal{F}_i$ for some $i \in I$, and thus $B \in \mathcal{F}_i$. Hence $B \in \bigcup \mathcal{C}$.

Thus $\bigcup \mathcal{C}$ is an upper bound for $\mathcal{C}$, and since $\mathcal{C}$ is arbitrary every chain in $\mathcal{S}$ has an upper bound. By Zorn's lemma, $\mathcal{S}$ has a maximal element $\mathcal{U}$. It remains to prove that $\mathcal{U}$ is an ultrafilter. The conditions (i)-(iii) of **Definition 2.1.3** hold by construction.

To prove that $\mathcal{U}$ is an ultrafilter, suppose $A \subset U$, and consider $\mathcal{U} \cup \{A\}$. If there exists $C \in \mathcal{U}$ such that $C \cap A = \emptyset$, then $C \subset U \setminus A$ and (iii) implies that $U \setminus A \in \mathcal{U}$. Otherwise, $\mathcal{U} \cup \{A\}$ has the finite intersection property and hence $A \in \mathcal{U}$ by the maximality of $\mathcal{U}$.

Since A is arbitary, $\mathcal{U}$ is an ultrafilter. □

Nonprincipal ultrafilters are crucial in the construction of the hyperreals, and we are happy to conclude this section with a proof of their existence.

Theorem 2.1.3. *There exists a nonprincipal ultrafilter on every infinite set U.*

Proof. Consider the set U^{co}. By **Lemma 2.1.3**, this set has the finite intersection property and is extendable, by **Theorem 2.1.2**, to an ultrafilter $\mathcal{U}$. Suppose that $\mathcal{U}$ is principal. Then for some $i \in U$, we have $\{i\} \in \mathcal{U}$. But $U \setminus \{i\}$ is cofinite and consequently belongs to $\mathcal{U}$. Thus $\{i\} \cap (U \setminus \{i\}) = \emptyset \in \mathcal{U}$, a contradiction. Hence, $\mathcal{U}$ is a nonprincipal ultrafilter on U. □

2.2 The Ultraproduct

In this section, $L = (C_L, R_L, F_L)$ is a fixed first-order language.

Definition 2.2.1. Suppose $(\mathcal{M}_i : i \in I)$ is a family of L-structures, and $\mathcal{F}$ a filter on I. Let $M = \prod_{i \in I} M_i$. For any L-formula $\varphi(x_1, \ldots, x_n)$ and tuple $a = (a_1, \ldots, a_n) \in M^n$, the *boolean extension of* $\varphi(a_1, \ldots, a_n)$ is defined by

$$\|\varphi(a)\| = \|\varphi(a_1, \ldots, a_n)\| = \{i \in I : \mathcal{M}_i \models \varphi(a_1(i), \ldots, a_n(i))\}.$$

The following consequences are immediate.

Lemma 2.2.1. *For any formulas φ and ψ and $a \in M^n$, the following holds:*

(i) $\|\varphi(a) \wedge \psi(a)\| = \|\varphi(a)\| \cap \|\psi(a)\|$

(ii) $\|\varphi(a) \vee \psi(a)\| = \|\varphi(a)\| \cup \|\psi(a)\|$

(iii) $\|\neg\varphi(a)\| = I \setminus \|\varphi(a)\|$

(iv) *For all $a \in M^{n-1}$ and $b \in M$, we have $\|\varphi(a,b)\| \subset \|\exists x\varphi(a,x)\|$ and there exists $c \in M$ such that $\|\varphi(a,c)\| = \|\exists x\varphi(a,x)\|$.*

Proof. The proof is immediate, except for the second part of (iv) where we define $c \in \prod_{i\in I} M_i$ as follows. For every $i \in I$, let $c(i) \in M_i$ be such that $\mathcal{M}_i \models \varphi(a, c(i))$ if $\mathcal{M}_i \models \exists x\varphi(a,x)$. Otherwise, pick $c(i)$ arbitarily from M_i. By the definition of c, we must have that $\|\exists x\varphi(a,x)\| \subset \|\varphi(a,c)\|$, and hence $\|\exists x\varphi(a,x)\| = \|\varphi(a,c)\|$. □

A filter on I naturally induces an equivalence relation on $M = \prod_{i\in I} M_i$, in the following way.

Definition 2.2.2. Suppose $(\mathcal{M}_i : i \in I)$ is a family of models, $M = \prod_{i\in I} M_i$ and $\mathcal{F}$ a filter on I. Suppose $a, b \in M$. Then we say that a and b are *equivalent modulo $\mathcal{F}$*, written $a \sim_{\mathcal{F}} b$, if $\|a = b\| \in \mathcal{F}$. Moreover, if $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ belong to M^n, we write $a \sim_{\mathcal{F}} b$ if $a_i \sim_{\mathcal{F}} b_i$ for each $1 \leq i \leq n$.

Intuitively, we regard a subset of I as "large" if it lies in $\mathcal{F}$. Two elements $a, b \in M$ are equal if their coordinates are equal in a "large" subset of I. It turns out that this is an equivalence relation.

Lemma 2.2.2. *$\sim_{\mathcal{F}}$ is an equivalence relation.*

Proof. $\|a = a\| = I \in \mathcal{F}$ for every filter $\mathcal{F}$. Hence $a \sim_{\mathcal{F}} a$. Furthermore, $\mathcal{M}_i \models a(i) = b(i)$ if and only if $\mathcal{M}_i \models b(i) = a(i)$ so $\|a = b\|$ equals $\|b = a\|$ and thus $a \sim_{\mathcal{F}} b$ implies $b \sim_{\mathcal{F}} a$.

Finally, let $a \sim_{\mathcal{F}} b$ and $b \sim_{\mathcal{F}} c$. Then $\|a = b\|, \|b = c\| \in \mathcal{F}$, and hence

$$\|a = b\| \cap \|b = c\| = \|(a = b) \wedge (b = c)\| \in \mathcal{F}.$$

Also, $\mathcal{M}_i \models ([a(i) = b(i)] \wedge [b(i) = c(i)])$ implies $\mathcal{M}_i \models a(i) = c(i)$, for every $i \in I$. Thus

$$\|(a = b) \wedge (b = c)\| \subset \|a = c\| \in \mathcal{F},$$

so $a \sim_{\mathcal{F}} c$, and $\sim_{\mathcal{F}}$ is an equivalence relation. □

Since $\sim_{\mathcal{F}}$ is an equivalence relation, it partitions $M = \prod_{i\in I} M_i$ into a set of equivalence classes. The set of equivalence classes is denoted by $M/\mathcal{F}$, and for any $a \in M$ we denote the corresponding equivalence class by $a/\mathcal{F}$. If $a = (a_1, \ldots, a_n) \in M^n$, we can write $a/\mathcal{U}$ for $(a_1/\mathcal{U}, \ldots, a_n/\mathcal{U})$. This is helpful in the following definition.

Definition 2.2.3. Suppose $(\mathcal{M}_i : i \in I)$ is a family of models, and $\mathcal{F}$ a filter on I. The *reduced product (modulo $\mathcal{F}$)* of $(\mathcal{M}_i : i \in I)$, denoted $\mathcal{M}/\mathcal{F}$, is the model defined as follows:

(i) The universe of $M/\mathcal{F}$ is the set of equivalence classes under $\sim_{\mathcal{F}}$.

(ii) For each $c \in C_L$, we let $c^{\mathcal{M}/\mathcal{F}} = (c^{\mathcal{M}_i} : i \in I)/\mathcal{F}$.

(iii) For each $R \in R_L$ and $a \in M^n$, we let $\mathcal{M}/\mathcal{F} \models R(a/\mathcal{F})$ if and only if $\|R(a)\| \in \mathcal{F}$.

(iv) For each $f \in F_L$ and $a \in M^n$, we let $f^{\mathcal{M}/\mathcal{F}}(a/\mathcal{F}) = (f^{\mathcal{M}_i}(a(i)) : i \in I)/\mathcal{F}$

Furthermore, if $\mathcal{F}$ is an ultrafilter call $\mathcal{M}/\mathcal{F}$ the *ultraproduct of* $(\mathcal{M}_i : i \in I)$ *(modulo $\mathcal{F}$)*.

It is not clear that this model is well-defined. The constant symbols are fine, but it is not obvious that for all $a, b \in M^n$ so that $a \sim_{\mathcal{F}} b$, we have $\mathcal{M} \models R(a)$ if and only if $\mathcal{M} \models R(b)$ and $f(a) \sim_{\mathcal{F}} f(b)$, where $R \in R_L$ and $f \in F_L$. Fortunately, we have the following lemma.

Lemma 2.2.3. *The model $\mathcal{M}/\mathcal{F}$ is well-defined.*

Proof. Suppose $R \in R_L$ is n-ary, and $a, b \in M^n$ with $a \sim_{\mathcal{F}} b$. Suppose $\mathcal{M} \models R(a)$. Since $a \sim_{\mathcal{F}} b$, we know that $\|a_j = b_j\| \in \mathcal{F}$ for any $0 \le j \le n$. Since $\mathcal{F}$ is a filter, we have that

$$\|a_1 = b_1\| \cap \cdots \cap \|a_n = b_n\| \in \mathcal{F}.$$

Hence **Lemma 2.2.1** implies that $\left\|\bigwedge_{j=0}^{n}(a_j = b_j)\right\| \in \mathcal{F}$. Moreover, $\|R(a)\| \in \mathcal{F}$ by definition and hence

$$\|R(a)\| \cap \left\|\bigwedge_{j=0}^{n}(a_j = b_j)\right\| \in \mathcal{F}.$$

However, if $\mathcal{M}_i \models \bigwedge_{j=0}^{n} a_j = b_j$, then $\mathcal{M}_i \models R(a)$ implies that $\mathcal{M}_i \models R(b)$. Hence

$$\|R(a)\| \cap \left\|\bigwedge_{j=0}^{n} a_j = b_j\right\| \subset \|R(b)\| \in \mathcal{F}.$$

So $\mathcal{M} \models R(b)$. The case for function symbols is done similarly. □

Remark. Note that we assumed from the outset that R_L contained an identity relation '='. This is included in our definition of the reduced product $\mathcal{M}/\mathcal{F}$, which means we must confirm that the induced relation '=' on $\mathcal{M}/\mathcal{F}$ is the actual identity relation.

But this is immediate, since for any $a, b \in M = \prod_{i \in I} M_i$, we have $\mathcal{M}/\mathcal{F} \models a/\mathcal{F} = b/\mathcal{F}$ if and only if $\|a = b\| \in \mathcal{F}$ if and only if $a \sim_{\mathcal{F}} b$ if and only if $a/\mathcal{F}$ and $b/\mathcal{F}$ are the same equivalence class (and hence the same element in $\mathcal{M}/\mathcal{F}$).

If $\mathcal{M}_i$ equals some L-structure $\mathcal{N}$ for each $i \in I$, we call the model $\mathcal{M}/\mathcal{F} = \mathcal{N}^I/\mathcal{F}$ a *reduced power of $\mathcal{N}$ (modulo $\mathcal{F}$)*. If $\mathcal{F}$ is an ultrafilter, $\mathcal{M}$ is called an *ultrapower* of $\mathcal{N}$.

The following theorem shows that ultraproducts behave nicely with regards to satisfiability of formulas. It is named after Polish mathematician Jerzy Loś, who proved it in 1955.

Theorem 2.2.1. Loś Theorem: *Suppose $(\mathcal{M}_i : i \in I)$ is family of models, $\mathcal{U}$ an ultrafilter on I and $M = \prod_{i \in I} M_i$. Then, for every n-tuple $(a_1, \ldots, a_n) \in M^n$ and formula $\varphi(x_1, \ldots, x_n)$, we have $\mathcal{M}/\mathcal{U} \models \varphi(a_1/\mathcal{U}, \ldots, a_n/\mathcal{U})$ if and only if $\|\varphi(a_1, \ldots, a_n)\| \in \mathcal{U}$.*

Proof. Note that throughout this proof, if $a = (a_1, \ldots, a_n) \in M^n = (\prod_{i \in I} M_i)^n$, then $a(i)$ refers to $(a_1(i), \ldots, a_n(i))$. We begin by proving the following claim.

Claim. *For every n-ary term $t(x)$ and tuple $a \in M^n$,*

$$t^{\mathcal{M}/\mathcal{U}}(a/\mathcal{U}) = \left(t^{\mathcal{M}_i}(a(i)) : i \in I\right)/\mathcal{U}.$$

Proof. We proceed by induction on the complexity of the term t.

- **Base case:** If t is a constant it holds by definition, and if t is a variable, for example if $t = x$, then for every $a \in M$

$$t^{\mathcal{M}/\mathcal{U}}(a) = a/\mathcal{U} = (a(i) : i \in I)/\mathcal{U} = \left(t^{\mathcal{M}_i}(a(i)) : i \in I\right)/\mathcal{U}.$$

- **Induction step:** Suppose that $t(x) = f(t_1(x_1), \ldots, t_n(x_n))$, where $t_j(x_j)$ are m_j-ary terms such that for every $b \in M^{m_j}$, we have

$$t_j^{\mathcal{M}/\mathcal{U}}(b/\mathcal{U}) = \left(t_j^{\mathcal{M}_i}(b(i)) : i \in I\right)/\mathcal{U}.$$

  Then for any $a = (a_1, \ldots, a_n)$, where $a_j \in M^{m_j}$ for all $1 \le j \le n$, we have

$$\begin{aligned} t^{\mathcal{M}/\mathcal{U}}(a/\mathcal{U}) &= f^{\mathcal{M}/\mathcal{U}}(t_1(a_1/\mathcal{U}), \ldots, t_n(a_n/\mathcal{U})) = \\ &= f^{\mathcal{M}/\mathcal{U}}\left(\left(t_1^{\mathcal{M}_i}(a_1(i)) : i \in I\right)/\mathcal{U}, \ldots, \left(t_n^{\mathcal{M}_i}(a_n(i)) : i \in I\right)/\mathcal{U}\right) = \\ &= \left(f^{\mathcal{M}_i}\left(t_1^{\mathcal{M}_i}(a_1(i)), \ldots, t_n^{\mathcal{M}_i}(a_n(i))\right) : i \in I\right)/\mathcal{U} = \\ &= \left(t^{\mathcal{M}_i}(a(i)) : i \in I\right)/\mathcal{U} \end{aligned}$$

□

We use induction on the complexity of formulas.

- **Base case:** Suppose $\varphi(x)$ is atomic, and that $\varphi(x,y) = (t_1(x) = t_2(y))$. Furthermore, suppose $a \in M^n$ and $b \in M^m$. If $\mathcal{M}/\mathcal{U} \models \varphi(a,b)$, the claim implies that

$$\left(t_1^{\mathcal{M}_i}(a(i) : i \in I\right)/\mathcal{U} = \left(t_2^{\mathcal{M}_i}(b(i)) : i \in I\right)/\mathcal{U}.$$

In other words, $t_1^{\mathcal{M}/\mathcal{U}}(a) \sim_{\mathcal{U}} t_2^{\mathcal{M}/\mathcal{U}}(b)$ and therefore $\left\|t_1^{\mathcal{M}/\mathcal{U}}(a) = t_2^{\mathcal{M}/\mathcal{U}}(b)\right\| \in \mathcal{U}$. The converse is proved similarly.

On the other hand, suppose that $\varphi(x_1,\dots,x_n) = R(t_1(x_1),\dots,t_n(x_n))$ where $R \in L_R$, $t_j(x_j)$ are m_j-ary terms and $a_j \in M^{m_j}$ for every $1 \le j \le n$. Let $a = (a_1,\dots,a_n)$. Then

$$\begin{aligned}
& \mathcal{M}/\mathcal{U} \models \varphi(a_1/\mathcal{U},\dots,a_n/\mathcal{U}) \\
\Leftrightarrow\quad & \mathcal{M}/\mathcal{U} \models R(t_1(a_1/\mathcal{U}),\dots,t_n(a_n/\mathcal{U})) \\
\Leftrightarrow\quad & \mathcal{M}/\mathcal{U} \models R\left(\left(t_1^{\mathcal{M}_i}(a_1(i)) : i \in I\right)/\mathcal{U},\dots,\left(t_n^{\mathcal{M}_i}(a_n(i)) : i \in I\right)/\mathcal{U}\right) \\
\Leftrightarrow\quad & \left\|R\left(\left(t_1^{\mathcal{M}_i}(a_1(i)) : i \in I\right),\dots,\left(t_n^{\mathcal{M}_i}(a_n(i)) : i \in I\right)\right)\right\| \in \mathcal{U} \\
\Leftrightarrow\quad & \left\{i \in I : \mathcal{M}_i \models R\left(t_1^{\mathcal{M}_i}(a_1(i)),\dots,t_n^{\mathcal{M}_i}(a_n(i))\right)\right\} \in \mathcal{U} \\
\Leftrightarrow\quad & \{i \in I : \mathcal{M}_i \models \varphi(a_1(i),\dots,a_n(i))\} \in \mathcal{U} \\
\Leftrightarrow\quad & \|\varphi(a_1,\dots,a_n)\| \in \mathcal{U}.
\end{aligned}$$

- **Induction step:** We consider three cases. Let $a \in M^n$.
  - Suppose φ is a formula for which the assertion in the theorem holds. Then $\mathcal{M}/\mathcal{U} \models \neg\varphi(a/\mathcal{U})$ if and only if $\mathcal{M}/\mathcal{U} \not\models \varphi(a/\mathcal{U})$ if and only if $\|\varphi(a)\| \notin \mathcal{U}$ if and only if $I \setminus \|\varphi(a)\| \in \mathcal{U}$, since $\mathcal{U}$ is an ultrafilter. **Lemma 2.2.1** implies that $\|\neg\varphi(a)\| \in \mathcal{U}$ if and only if $\mathcal{M}/\mathcal{U} \models \neg\varphi(a/\mathcal{U})$.
  - Suppose φ and ψ are formulas for which the above assertion hold. Then $\mathcal{M}/\mathcal{U} \models \varphi(a/\mathcal{U}) \wedge \psi(a/\mathcal{U})$ if and only if $\mathcal{M}/\mathcal{U} \models \varphi(a/\mathcal{U})$ and $\mathcal{M}/\mathcal{U} \models \psi(a/\mathcal{U})$. By assumption, this is equivalent with $\|\varphi(a)\| \in \mathcal{U}$ and $\|\psi(a)\| \in \mathcal{U}$.

    If $\|\varphi(a)\| \in \mathcal{U}$ and $\|\psi(a)\| \in \mathcal{U}$, then $\|\varphi(a)\| \cap \|\psi(a)\| \in \mathcal{U}$, and by **Lemma 2.2.1** we have $\|\varphi(a) \wedge \psi(a)\| \in \mathcal{U}$. On the other hand, $\|\varphi(a) \wedge \psi(a)\|$ is contained in $\|\varphi(a)\|$ and $\|\psi(a)\|$, so if $\|\varphi(a) \wedge \psi(a)\| \in \mathcal{U}$, we must have $\|\varphi(a)\| \in \mathcal{U}$ and $\|\psi(a)\| \in \mathcal{U}$.

    Thus $\mathcal{M}/\mathcal{U} \models \varphi(a/\mathcal{U}) \wedge \psi(a/\mathcal{U})$ if and only if $\|\varphi(a) \wedge \psi(a)\| \in \mathcal{U}$.
  - Suppose φ is a formula for which the above assertion holds. We have that $\mathcal{M}/\mathcal{U} \models \exists x\varphi(x, a/\mathcal{U})$ if and only if $\mathcal{M}/\mathcal{U} \models \varphi(b/\mathcal{U}, a/\mathcal{U})$ for some $b \in M = \prod i \in IM_i$, which is equivalent to $\|\varphi(b,a)\| \in \mathcal{U}$. By **Lemma 2.2.1**, this is equivalent to the fact that $\|\exists x\varphi(x,a)\| \in \mathcal{U}$.

□

Note that is only when dealing with negations that we need $\mathcal{U}$ to be an ultrafilter, in all other cases if suffices for $\mathcal{U}$ to be a regular filter. Also, recall that $\mathcal{M}/\mathcal{U}$ is an L-structure: hence Łoś theorem holds for all formulas stated in the original language of the factors and not for some new we might define on $\mathcal{M}/\mathcal{U}$. This is important when $\mathcal{M}/\mathcal{U}$ are the hyperreals and the factors are the reals.

For any ultrapower $\mathcal{M}/\mathcal{U}$ of $\mathcal{N}$, the function $\mu : N \to M = \prod_{i\in I} M_i$ sending $a \in N$ to the constant sequence $(a : i \in I) \in M$ is an embedding. By Łoś Theorem, this embedding preserves formulas.

Corollary 2.2.1. *Suppose μ is as above, and that $\mathcal{M}/\mathcal{U}$ is an ultrapower of $\mathcal{N}$. Then for every $(a_1,\dots,a_n) \in N^n$ and formula $\varphi(x_1,\dots,x_n)$ we have $\mathcal{N} \models \varphi(a_1,\dots,a_n)$ if and only if $\mathcal{M}/\mathcal{U} \models \varphi(\mu(a_1),\dots,\mu(a_n))$.*

In other words, μ is an elementary embedding of $\mathcal{N}$ into $\mathcal{M}/\mathcal{U}$ and we can regard $\mathcal{N}$ as an elementary substructure of $\mathcal{M}/\mathcal{U}$.

Proof. For every $a = (a_1,\dots,a_n) \in N^n$ and formula $\varphi(x_1,\dots,x_n)$ we have, by Łoś Theorem, that $\mathcal{M} \models \varphi(\mu(a_1),\dots,\mu(a_n))$ if and only if $\left\|\varphi(\mu(a_1),\dots,\mu(a_n))\right\| \in \mathcal{U}$.

Suppose that $\mathcal{M} \models \varphi(\mu(a_1),\dots,\mu(a_n))$. Then $\left\|\varphi(\mu(a_1),\dots,\mu(a_n))\right\|$ is nonempty, and consequently $\mathcal{N} \models \varphi(a_1(i),\dots,a_n(i))$ for some $i \in I$. Hence $\mathcal{N} \models \varphi(a_1,\dots,a_n)$. On the other hand, if $\mathcal{N} \models \varphi(a)$, then $\left\|\varphi(\mu(a_1),\dots,\mu(a_n))\right\| = I \in \mathcal{U}$ and $\mathcal{M} \models \varphi(\mu(a_1),\dots,\mu(a_n))$. □

Note that μ is surjective if and only if every $x \in \mathcal{M}$ is equivalent modulo $\mathcal{U}$ to some constant sequence. This observation explains all the work to prove the existence of nonprincipal ultrafilters.

Corollary 2.2.2. *Suppose $\mathcal{U}$ is a principal ultrafilter on I. Then $\mathcal{M}^I/\mathcal{U}$ is isomorphic to $\mathcal{M}$.*

Proof. Let $a \in \mathcal{M}^I/\mathcal{U}$. Since $\mathcal{U}$ is principal, there exists $i \in I$ such that $\{i\} \in \mathcal{U}$. Take b as the constant sequence with value $a(i) \in \mathcal{M}$. Then $a \sim_{\mathcal{U}} b$ (since $\{i\} \subset \|a = b\| \in \mathcal{U}$), which implies that the embedding detailed above is surjective and an isomorphism. □

2.3 The Hyperreals as an Ultrapower

In the previous chapter, we noted that an ultrapower is a model over same language as the original model. In order to use Łoś theorem to its full extent, we make the language contain as much information as we can.

Definition 2.3.1. By $L_{\mathbb{R}}$ we denote the language with

(i) constant symbols c_r for every $r \in \mathbb{R}$, and

(ii) n-ary relation symbols R_A for every $A \subset \mathbb{R}^n$, for every $n \in \mathbb{N}$.

Note that the language does not contain any function symbols. The reason for this is that if we were to add a function symbol for every function defined on a subset of $\mathbb{R}$, we would run into an issue with having undefined terms. Not that this is insurmountable by any means, this approach is used in [2].

However, no such terms are needed. We can reduce each statement we make to $L_{\mathbb{R}}$, since the language is quite powerful. A case in point, if f_i $(0 \leq i \leq n)$ are n_i-ary functions and R an n-ary relation, there exists a $\sum n_i$-ary relation $\tilde{R}$ so that the relation $R(f_1(x_1),\dots,f_n(x_n))$ holds if and only $\tilde{R}(x_1,\dots,x_n)$ holds.

As an example, consider the sentence $\varphi = \forall x \forall y(\exp(x+y) = \exp x \exp y)$. If R is the 4-ary relation such that $R(a,b,c,d)$ holds if and only if $\exp(a+b) = \exp c \exp d$, then $\forall x \forall y R(x,y,x,y)$ is an $L_{\mathbb{R}}$-sentence equivalent to φ.

In general, we are not entirely explicit with reducing every first order-expression to $L_{\mathbb{R}}$. Instead, we write more intelligible statements, safe in the knowledge that every such expression is reducible to one with only relation and constant symbols.

The following abbreviations are used for clarity. Suppose $\varphi(x_1,\dots,x_n)$ is a formula, $A \subset \mathbb{R}^m$ and $x_1,\dots,x_n$ are m-tuples of variables. Then

- $(\forall x_1,\dots,x_n \in A)\varphi(x_1,\dots,x_n)$ abbreviates

$$\forall x_1 \dots \forall x_n (R_A(x_1) \wedge \dots \wedge R_A(x_n) \to \varphi(x_1,\dots,x_n)),$$

  and

- $(\exists x_1,\dots,x_n \in A)\varphi(x_1,\dots,x_n)$ abbreviates

$$\exists x_1 \dots \exists x_n (R_A(x_1) \wedge \dots \wedge R_A(x_n) \wedge \varphi(x_1,\dots,x_n)).$$

We usually use r instead of its related constant symbol c_r, for the same reason. Having decided upon a language to use, we can explicitly state what we consider as our "base" model of $\mathbb{R}$.

Definition 2.3.2. By $\mathcal{M}_{\mathbb{R}}$ we denote the $L_{\mathbb{R}}$-model with universe $\mathbb{R}$, and

(i) $c_r^{\mathcal{M}_{\mathbb{R}}} = r$ for every $r \in \mathbb{R}$.

(ii) $R_A^{\mathcal{M}_{\mathbb{R}}} = A$ for every $A \subset \mathbb{R}^n$, for every $n \in \mathbb{N}$.

Note that if $A \subset \mathbb{R}$ and $x \in \mathbb{R}$, then $\mathcal{M}_{\mathbb{R}} \models R_A(x)$ if and only if $x \in A$, and $\mathcal{M}_{\mathbb{R}} \models x = c_r$ if and only if $r = x$.

Definition 2.3.3. Suppose $\mathcal{U}$ is a nonprincipal ultrafilter on $\mathbb{N}$ containing $\mathbb{N}^{co}$. Then

$$\mathcal{M}_{\mathcal{R}} = \prod_{n \in \mathbb{N}} \mathcal{M}_{\mathbb{R}} \Big/ \mathcal{U}$$

is *the (model of the) hyperreals*. The universe of $\mathcal{M}_{\mathcal{R}}$ is denoted $\mathcal{R}$.

Most of the time we identify, as per mathematical custom, $\mathcal{M}_{\mathbb{R}}$ with $\mathbb{R}$ and $\mathcal{M}_{\mathcal{R}}$ with $\mathcal{R}$. Hence for some $L_{\mathbb{R}}$-formula φ we write $\mathbb{R} \models \varphi$ and $\mathcal{R} \models \varphi$ when we mean $\mathcal{M}_{\mathbb{R}} \models \varphi$ and $\mathcal{M}_{\mathcal{R}} \models \varphi$.

At the moment, $\mathbb{R}$ and $\mathcal{R}$ are models of the same language, namely $L_{\mathbb{R}}$. This is a consequence of the ultrapower construction, but is cumbersome to work with. Thus, when we interpret $L_{\mathbb{R}}$-sentences over $\mathcal{R}$ we append a $*$-symbol as to make it clear over which model we are interpreting it.

Definition 2.3.4. Suppose φ is an $L_{\mathbb{R}}$-sentence. Then φ^* is the sentence such that every nonlogical symbol S is replaced by the symbols S^*, unless S is one of the following:

- A constant symbol.
- One of the relations $=, <, >, \leq$ or $\geq$.
- One of the functions $|\cdot|, \cdot, +, \div$ or $-$.

φ^* is called the *$*$-transform* of φ. The resulting language is named $L_{\mathcal{R}}$, and as a result $\mathcal{R}$ is considered an $L_{\mathcal{R}}$-model.

At the end of the previous section (more precisely, in **Corollary 2.2.1**), we proved that there is an elementary embedding of $\mathcal{N}$ into $\mathcal{M}$ whenever $\mathcal{M}$ is an ultrapower of $\mathcal{N}$. Hence we can regard $\mathbb{R}$ as a substructure of $\mathcal{R}$. The fact that the embedding is elementary implies the following principle.

Theorem 2.3.1. The Transfer Principle*: Suppose* $(r_1, \dots, r_n) \in \mathbb{R}^n$ *and* $\varphi(x_1, \dots, x_n)$ *is an* $L_{\mathbb{R}}$*-formula. Then* $\mathbb{R} \models \varphi(r_1, \dots, r_n)$ *if and only if* $\mathcal{R} \models \varphi^*(r_1, \dots, r_n)$.

Given a subset A of $\mathbb{R}^n$, there exists a corresponding $R_A \in L_{\mathbb{R}}$ so that $\mathbb{R} \models R_A(x)$ if and only if $x \in A$. This naturally yields an extension $A^* \subset \mathcal{R}$.

Definition 2.3.5. Suppose A is a subset of $\mathbb{R}^n$. The *extension of A* is the set

$$A^* = \{x \in \mathcal{R}^n : \mathcal{R} \models R_A^*(x)\} \subset \mathcal{R}^n.$$

Many important subsets of $\mathcal{R}$ constructed this way, for example the set of *hypernaturals* $\mathbb{N}^*$ is the extension of the natural numbers $\mathbb{N}$. It has many interesting properties in common with $\mathbb{N}$, but contains a host of unlimited numbers in addition to $\mathbb{N}$.

If $A \subset \mathbb{R}$, any function $f : A \to \mathbb{R}$ can be regarded as a relation on $\mathbb{R}^2$ in a natural way.

Definition 2.3.6. Suppose $f : A \to \mathbb{R}$ is a function where $A \subset \mathbb{R}$. The *graph* of f is defined as the set $G_f = \{(x, f(x)) : x \in A\} \subset \mathbb{R}^2$.

The relation G_f can be extended to a relation G_f^* on $\mathcal{R}$, which is used to define the extension of a function $f : A \to \mathbb{R}$.

Definition 2.3.7. Suppose $A \subset \mathbb{R}$ and $f : A \to \mathbb{R}$ is a function, and G_f is the graph of f, Furthermore, suppose that $x \in A^*$. The *extension of* f is the function $f^*(x) : A^* \to \mathcal{R}$ defined by

$$f^*(x) = \text{ the unique y such that } \mathcal{R} \models G_f^*(x, y).$$

Remark. This is well-defined, by transfer of the sentence

$$(\forall x \in A)(\exists y)(G_f(x, y) \wedge (\forall z G_f(x, z) \to y = z)).$$

Furthermore, if $x \in A \subset \mathcal{R}$ we have $f^*(x) = f(x)$.

Note that this includes all the standard operations and functions, such as the absolute value, addition, multiplication and their inverses. Also, since a sequence $s = (s_n : n \in \mathbb{N})$ is regarded as a function $s : \mathbb{N} \to \mathbb{R}$ it has an extension $s^* : \mathbb{N}^* \to \mathcal{R}$, which is called the *hypersequence*.

Often, transfer is used to investigate properties of extensions, as in the following example.

Example 2.3.1. For any $x, y \in \mathcal{R}$, we have $|x + y| \leq |x| + |y|$.

Proof. Transfer of the sentence $\forall x \forall y\, (|x + y| \leq |x| + |y|)$. □

Another example is the exponential function $\exp : \mathbb{R} \to \mathbb{R}$ which has an extension $\exp^* : \mathcal{R} \to \mathcal{R}$.

Example 2.3.2. The exponential function $\exp^* : \mathcal{R} \to \mathcal{R}$ is strictly increasing, and satisfy the property that $\exp^*(x + y) = \exp^* x \exp^* y$. In addition, it is strictly positive.

Proof. Transfer of the sentence $\forall x \forall y (x < y \to \exp(x) < \exp(y))$ shows that $\exp^*$ is strictly increasing, transfer of the sentence $\forall x \forall y\, (\exp(x + y) = \exp x \exp y)$ yields the second part. Positivity follows from transfer of $\forall x (\exp x > 0)$. □

2.4 Hyperreals, Great and Small

We begin by examining some of the additions to $\mathbb{R}$ that $\mathcal{R}$ provide: infinitesimals and unlimited numbers. First, note that there exists an elementary embedding of $\mathbb{R}$ into $\mathcal{R}$ given by the constant sequences. Hence $\mathbb{R}$ is an elementary substructure of $\mathcal{R}$, which implies that $\mathbb{R}$ and $\mathcal{R}$ are elementarily equivalent.

Recall that every hyperreal is a sequence (or rather, an equivalence class of sequences) and vice versa. In particular, the sequence $\alpha = (\alpha_n = n^{-1} : n \in \mathbb{N})$ is a hyperreal. Consider the formula $\varphi_\varepsilon(x) = x \geq \varepsilon$. For each $\varepsilon \in \mathbb{R}_+$, we have $\|\varphi_\varepsilon(\alpha)\| = \|\alpha \geq \varepsilon\|$, and since α converges to 0 there is some $N \in \mathbb{N}$ so that $n > N$ implies $\alpha_n < \varepsilon$.

Hence the set $\|\alpha \geq \varepsilon\|$ is finite, and does not belong to the ultrafilter. Loś Theorem implies that $\mathcal{R} \models \alpha < \varepsilon$, for all $\varepsilon \in \mathbb{R}_+$ (since it was arbitary). Also, $\alpha_n \neq 0$ for all n so the set $\|\alpha = 0\|$ is empty, meaning $\mathcal{R} \models \alpha \neq 0$.

So α is a number different from 0, yet smaller than any real quantity: we have proven the existence of infinitesimals in $\mathcal{R}$, if we regard infinitesimals as those hyperreals which are smaller than any real numbers. The same argument can be used to prove that some hyperreals are greater than any real number, by letting $\alpha = (n : n \in \mathbb{N})$.

Definition 2.4.1. A nonzero hyperreal α such that $|\alpha| < \varepsilon$ for every $\varepsilon \in \mathbb{R}_+$ is an *infinitesimal*. The set of infinitesimals is denoted $\mathbb{I}$.

Definition 2.4.2. A hyperreal α such that $|\alpha| > R$ for every $R \in \mathbb{R}$ is called *unlimited*.

Definition 2.4.3. A hyperreal α such that $|\alpha| < R$ for some $R \in \mathbb{R}$ is called *limited*. The set of limited numbers are denoted $\mathbb{L}$.

Lemma 2.4.1. *The set* $\mathbb{I}$ *is closed under addition and multiplication.*

Proof. Suppose α and β are infinitesimals. Then $|\alpha| < \frac{\varepsilon}{2}$ and $|\beta| < \frac{\varepsilon}{2}$ for every $\varepsilon \in \mathbb{R}_+$. **Example 2.3.1** yields

$$|\alpha + \beta| \leq |\alpha| + |\beta| < \varepsilon$$

for every $\varepsilon \in \mathbb{R}_+$, meaning that α and β are infinitesimal.

The proof for multiplication is similar, using that $|\alpha|$ and $|\beta|$ are less than $\sqrt{\varepsilon}$. □

In the above terminology, we can define precisely what it means when two hyperreals α and β are infinitely close to each other.

Definition 2.4.4. Two hyperreals α and β are *infinitely close*, denoted $\alpha \simeq \beta$, if $|\alpha - \beta|$ is infinitesimal.

Lemma 2.4.2. $\simeq$ *is an equivalence relation.*

Proof. Clearly, $\simeq$ is reflexive and symmetric. To prove that it is transitive, note that the triangle inequality by **Example 2.3.1** holds in $\mathcal{R}$ and let $\varepsilon \in \mathbb{R}_+$. Suppose that $\alpha \simeq \beta \simeq \gamma$. Then $|\alpha - \beta| < \frac{\varepsilon}{2}$ and $|\beta - \gamma| < \frac{\varepsilon}{2}$, and we see that

$$|\alpha - \gamma| = |\alpha - \beta + \beta - \gamma| \leq |\alpha - \beta| + |\beta - \gamma| < \varepsilon.$$

So $\simeq$ is transitive and thus an equivalence relation. □

The concept of infinite closeness is tightly connected with the idea of continuity, as continous functions preserve infinite closeness in various ways.

Example 2.4.1. If $\alpha \simeq 0$, then $\exp^*(\alpha) \simeq 1$.

Proof. Let $\varepsilon \in \mathbb{R}_+$. Then $\alpha < \ln(1+\varepsilon) = \ln^*(1+\varepsilon)$, since $1+\varepsilon \in \mathbb{R}$. Since $\exp^*$ is strictly increasing and positive, by **Example 2.3.2**, we get that

$$|\exp^*(\alpha) - 1| < |\exp^*(\ln^*(1+\varepsilon)) - 1| = |1+\varepsilon-1| = \varepsilon.$$

We conclude that $\exp(\alpha) \simeq 1$. □

Definition 2.4.5. Let $r \in \mathbb{R} \subset \mathcal{R}$. The *halo of r* is the set

$$\text{hal}(r) = \{x \in \mathcal{R} : r \simeq x\}.$$

Definition 2.4.6. Let $r \in \mathbb{R}$ and $\rho \in \mathcal{R}$. If $|r - \rho|$ is infinitesimal, we say that r is a *shadow* of ρ.

When persuing an axiomatic approach to real analysis, a common way to do this is to regard $\mathbb{R}$ as an ordered field imbued with the *Dedekind completeness property*. This property separates $\mathbb{R}$ from other totally ordered fields such as $\mathbb{Q}$.

Definition 2.4.7. A totally ordered field F has the *Dedekind completeness property* if every $A \subset F$ which is nonempty and bounded from above has a *least upper bound*.

It turns out that being Dedekind complete is a strong requirement: any field which is Dedekind-complete is isomorphic to $\mathbb{R}$ (Appendix A in [5]). The following is a nonstandard characterization of Dedekind completeness.

Theorem 2.4.1. $\mathbb{R}$ *has the Dedekind completeness property if and only if each limited* $\alpha \in \mathcal{R}$ *has exactly one shadow* $a \in \mathbb{R}$.

Proof. "⇒": Consider $A = \{r \in \mathbb{R} : r < \alpha\}$. Since α is limited, there exists $r, s \in \mathbb{R}$ such that $r < \alpha < s$, and we conclude that A is nonempty and bounded above. Hence there exists a least upper bound $a \in \mathbb{R}$ of A.

To show $\alpha \simeq a$, let $\varepsilon \in \mathbb{R}_+$. Clearly $\alpha \leq a + \varepsilon$, and since a is the least upper bound of A, must have $\alpha \geq a - \varepsilon$. Thus we conclude that $a - \varepsilon \leq \alpha \leq a + \varepsilon$ for all $\varepsilon \in \mathbb{R}_+$. Hence $\alpha \simeq a$.

For uniqueness, suppose $\alpha \simeq a_1$ and $\alpha \simeq a_2$. By **Lemma 2.4.2** we have $a_1 \simeq a_2$ and since they are real $a_1 = a_2$. In other words, the upper bound is unique.

"⇐": Suppose $A \subset \mathbb{R}$ is nonempty and bounded from above. Clearly, a number a is an upper bound of A if and only if it is an upper bound of $\hat{A} = \bigcup\{(-\infty, a) : a \in A\}$. Hence it suffices to find a least upper bound of $\hat{A}$

By assumption, there exists $p_0 < q_0$ so that $p_0 \in \hat{A}$ and $q_0 \notin \hat{A}$. We inductively define sequences $p = (p_n : n \in \mathbb{N})$ and $q = (q_n : n \in \mathbb{N})$ by

$$p_{n+1} = \begin{cases} p_n & \text{if } (p_n + q_n)/2 \notin \hat{A} \\ (p_n + q_n)/2 & \text{otherwise} \end{cases}$$

$$q_{n+1} = \begin{cases} q_n & \text{if } (p_n + q_n)/2 \in \hat{A} \\ (p_n + q_n)/2 & \text{otherwise.} \end{cases}$$

Clearly, $p_n \in \hat{A}$ and $q_n \notin \hat{A}$ for every n. Moreover, since every hyperreal is a sequence of real numbers, p and q can be regarded as hyperreals.

Consider the sequence $(p_n - q_n : n \in \mathbb{N})$. We prove that this converges to 0. Let $n \in \mathbb{N}$. If $(p_n + q_n)/2 \in \hat{A}$ we have $p_{n+1} = (p_n + q_n)/2$ and $q_{n+1} = q_n$. Then

$$|p_{n+1} - q_{n+1}| = \left|\frac{p_n - q_n}{2}\right| = \frac{1}{2}|p_n - q_n|$$

and similar if $(p_n + q_n)/2 \notin \hat{A}$. By induction we have that

$$|p_n - q_n| = \frac{1}{2^n}|p_0 - q_0|,$$

which approaches 0 as n tends to infinity. Thus $p/\mathcal{U} - q/\mathcal{U} \simeq 0$, and by assumption they have a common shadow $L \in \mathbb{R}$. Clearly, $p_n \leq L \leq q_n$ for all $n \in \mathbb{N}$. Also, both p and q converges to L.

It remains to show that L is an upper bound of $\hat{A}$. Suppose there is $a \in \hat{A}$ so that $L < a$. Then $q_n < a$ for some $n \in \mathbb{N}$. Hence $q_n \in \hat{A}$, which is a contradiction.

Also, L is the least upper bound of $\hat{A}$ since if u is some upper bound of $\hat{A}$, we have $p_n \leq u$ for every $n \in \mathbb{N}$, which implies that $\lim_{n\to\infty} p_n = L \leq u$. Hence L is the supremum of $\hat{A}$, and thus of A. □

Definition 2.4.8. The mapping $\text{sh} : \mathbb{L} \to \mathbb{R}$ which maps any $\alpha \in \mathbb{L}$ to its shadow is called the *shadow mapping*, and a is the *shadow* of α.

Remark. Note that $\mathbb{L}$ is closed under multiplication, addition and subtraction. Readers familiar with abstract algebra notices that $\mathbb{L}$ is a subring of $\mathcal{R}$. Furthermore, if α is infinitesimal and ρ is limited, then $\alpha\rho$ is infinitesimal. Thus $\mathbb{I}$ forms an ideal in the ring $\mathbb{L}$.

It is easy to see that $\mathbb{I}$ is a maximal ideal, since any limited α not in $\mathbb{I}$ must be larger than some $\varepsilon \in \mathbb{R}_+$, in which case $\alpha^{-1} < \varepsilon^{-1}$. Hence α^{-1} is limited, meaning that any ideal containing α also contains $\alpha\alpha^{-1} = 1$, making it trivial.

This implies that the quotient ring $\mathbb{L}/\mathbb{I}$ is a field, which turns out to be isomorphic to $\mathbb{R}$ by the mapping sending each $\alpha \in \mathbb{L}/\mathbb{I}$ to $\text{sh}(\alpha)$.

Thus every limited hyperreal can be written as a real part plus an infinitesimal part. The latter of these plays an important role in nonstandard analysis, concepts such as limits and convergence are formulated using the shadow mapping and halos, as we see in the next chapter.

3 Analysis Without Limits

In this chapter, we utilize the tools furnished in the previous chapter, and find a nonstandard theory of real analysis which is equivalent to the standard theory. Before we do this, recall that every sequence $s = (s_n : n \in \mathbb{N})$ is a function $s : \mathbb{N} \to \mathbb{R}$. This yields an extended function $s^* : \mathbb{N}^* \to \mathbb{R}^*$, called the hypersequence. The hypersequence contains a lot of information about the sequence itself.

3.1 Sequences as Hypersequences

Remember the standard definition of convergence for sequences.

Definition 3.1.1. Suppose $s = (s_n : n \in \mathbb{N})$ is a sequence and $r \in \mathbb{R}$. We say that *s converges to r* if for every $\varepsilon \in \mathbb{R}_+$, there exists $N_\varepsilon \in \mathbb{N}$ such that if $n > N_\varepsilon$, then $|s_n - r| < \varepsilon$.

The following theorem provides a nonstandard characterisation of convergence.

Theorem 3.1.1. *A real sequence $s = (s_n : n \in \mathbb{N})$ converges to the real number r, if and only if $s^*_N \simeq r$ for each unlimited hypernatural N.*

Proof. "⇒": Suppose s converges to r. For every $\varepsilon \in \mathbb{R}^+$ there exists $N_\varepsilon \in \mathbb{N}$ such that

$$\mathbb{R} \models (\forall n \in \mathbb{N})(n > N_\varepsilon \to |s_n - r| < \varepsilon).$$

By transfer, we get

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(n > N_\varepsilon \to |s^*_n - r| < \varepsilon).$$

Suppose M is an unlimited hypernatural. Since M is unlimited, we have $M > N_\varepsilon$ for every $N_\varepsilon \in \mathbb{N}$ and hence $|s^*_N - r| < \varepsilon$ for every $\varepsilon \in \mathbb{R}_+$. Thus $s^*_M \simeq r$.

"⇐": Suppose $s^*_N \simeq r$ for every unlimited hypernatural N, and suppose N is an unlimited hypernatural. Then for all hypernaturals $n > N$, we have $s^*_n \simeq r$ as well. Thus we conclude that

$$\mathcal{R} \models (\exists N \in \mathbb{N}^*)(\forall n \in \mathbb{N}^*)(n > N \to |s^*_n - r| < \varepsilon)$$

for all $\varepsilon \in \mathbb{R}^+$. Transfer yields that

$$\mathbb{R} \models (\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \to |s_n - r| < \varepsilon)$$

Which is exactly the statement that s_n converges to r. □

Sequences which grow arbitarily large are commonplace throughout mathematics, we usually say that they tend towards infinity.

Definition 3.1.2. A sequence $s = (s_n : n \in \mathbb{N})$ diverges to ∞ (or $-\infty$) if for every $R \in \mathbb{R}_+$ there exists $N_R \in \mathbb{N}$ so that $n > N_R$ implies $s_n > R$ (or $-s_n > R$).

The following theorem gives a similar nonstandard characterization of sequences which tend to infinity.

Theorem 3.1.2. *A real sequence* $s = (s_n : n \in \mathbb{N})$

(i) diverges to infinity if and only if s^*_N *is positive unlimited for every unlimited hypernatural* N.

(ii) diverges to negative infinity if and only if s^*_N *is negative unlimited for every unlimited hypernatural* N.

Proof. It suffices to prove (i), since (ii) is identical up to reversal of the order.

"⇒": Since s diverges, for any $R \in \mathbb{R}$ there exists $N_R \in \mathbb{N}$ such that

$$\mathbb{R} \models (\forall n \in \mathbb{N})(n > N_R \rightarrow s_n > R).$$

By transfer we obtain

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(n > N_R \rightarrow s^*_n > R)$$

Suppose M is an unlimited hypernatural. Then $M > N_R$ for all $N_R \in \mathbb{N}$, so $s^*_M > R$ for all $R \in \mathbb{R}$. Hence s^*_M is positive unlimited.

"⇐": Suppose that s^*_N is positive unlimited for all unlimited hypernaturals N. Then we have

$$\mathcal{R} \models (\exists N \in \mathbb{N}^*)(\forall n \in \mathbb{N}^*)(n > N \rightarrow s^*_n > R)$$

for all $R \in \mathbb{R}$. Transfer gives us that

$$\mathbb{R} \models (\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \rightarrow s_n > R)$$

which is exactly the statement that s diverges to positive infinity. □

It is clear that if $s = (s_n : n \in \mathbb{N})$ converges, the difference between successive terms in the sequence must grow arbitarily small. In other words, $|s_{n+1} - s_n|$ must approach 0. Unfortunately, this is not a sufficient condition, as is evidenced by the harmonic series $\sum \frac{1}{n}$. This series tend to infinity, and hence its partial sums also tend to infinity. But

$$|s_{n+1} - s_n| = \left| \sum_{i=1}^{n+1} \frac{1}{i} - \sum_{i=1}^{n} \frac{1}{i} \right| = \left| \frac{1}{n+1} \right|$$

tends to 0. So $|s_{n+1} - s_n|$ approaching 0 does not guarantee that s converges. However, there exists a similar requirement which do imply convergence, due to Cauchy.

Definition 3.1.3. A sequence $s = (s_n : n \in \mathbb{N})$ is a *Cauchy sequence* if for every $\varepsilon \in \mathbb{R}_+$ there exists N_ε so that for every $n, m \in \mathbb{N}$ such that $n, m > N_\varepsilon$ we have $|s_n - s_m| < \varepsilon$.

We show that a sequence converges if and only if it is a Cauchy sequence. First, we prove the following lemmas.

Lemma 3.1.1. *A real sequence $s = (s_n : n \in \mathbb{N})$ is bounded if and only if s^*_N is limited for every unlimited hypernatural N.*

Proof. "⇒": Suppose s is bounded. Then there exists $B \in \mathbb{R}$ such that

$$\mathbb{R} \models (\forall n \in \mathbb{N})(|s_n| < B).$$

By transfer we have

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(|s^*_n| < B).$$

In other words, s^*_n is limited for every $n \in \mathbb{N}^*$ and thus for every unlimited N as well.

"⇐": Suppose s^*_N is limited for every unlimited hypernatural N. Suppose B is any unlimited hyperreal. Then $|s^*_N| < B$ for all unlimited N, and since $s^*_n = s_n$ is limited for any natural number n, we get that $|s^*_n| < B$ for all $n \in \mathbb{N}^*$. Thus

$$\mathcal{R} \models \exists B(\forall n \in \mathbb{N}^*)(|s^*_n| < B)$$

and by transfer we obtain

$$\mathbb{R} \models \exists B(\forall n \in \mathbb{N})(|s_n| < B).$$

Thus, s is bounded. □

Lemma 3.1.2. *A real sequence $s = (s_n : n \in \mathbb{N})$ is a Cauchy sequence if and only if $s^*_N \simeq s^*_M$ for all unlimited hypernaturals N, M.*

Proof. "⇒": Suppose s is a Cauchy sequence. Then for each $\varepsilon \in \mathbb{R}_+$ there exists $N_\varepsilon \in \mathbb{N}$ such that

$$\mathbb{R} \models (\forall n, m \in \mathbb{N})(n > N_\varepsilon \wedge m > N_\varepsilon \rightarrow |s_n - s_m| < \varepsilon).$$

By transfer we obtain

$$\mathcal{R} \models (\forall n, m \in \mathbb{N}^*)(n > N_\varepsilon \wedge m > N_\varepsilon \rightarrow |s^*_n - s^*_m| < \varepsilon).$$

Suppose N and M are unlimited hypernaturals. Then $N, M > N_\varepsilon$ for every $N_\varepsilon \in \mathbb{N}$, and we conclude that $|s^*_N - s^*_M| < \varepsilon$ for every $\varepsilon \in \mathbb{R}_+$. Thus $s^*_N \simeq s^*_M$.

"⇐": Suppose that $s^*_N \simeq s^*_M$ for all unlimited hypernaturals N and M. Since all hypernaturals greater than some unlimited hypernatural are also unlimited, we conclude that

$$\mathcal{R} \models (\exists N \in \mathbb{N}^*)\,(\forall n, m \in \mathbb{N}^*)(n > N \wedge m > N \rightarrow |s^*_n - s^*_m| < \varepsilon).$$

for every $\varepsilon \in \mathbb{R}_+$. By transfer we get

$$\mathbb{R} \models (\exists N \in \mathbb{N})(\forall n, m \in \mathbb{N})(n > N \wedge m > N \rightarrow |s_n - s_m| < \varepsilon),$$

for every $\varepsilon \in \mathbb{R}_+$. Hence s is a Cauchy sequence. □

Theorem 3.1.3. *A sequence $s = (s_n : n \in \mathbb{N})$ converges if and only if it is a Cauchy sequence.*

Proof. "$\Rightarrow$": Suppose s converges to some $L \in \mathbb{R}$. Suppose N and M are two unlimited hypernaturals. Then $s^*_N \simeq L \simeq s^*_M$ and **Lemma 2.4.2** implies that $s^*_N \simeq s^*_M$ for all unlimited hypernaturals N and M. By **Lemma 3.1.2** the sequence s is a Cauchy sequence.

"$\Leftarrow$": Suppose s is a Cauchy sequence. By definition, there exists $N \in \mathbb{N}$ such that $|s_n - s_m| < 1$ for all $n, m > N$. In particular, for all $m > N$, we have

$$|s_m| \leq |s_m - s_{N+1}| + |s_{N+1}| < 1 + |s_{N+1}|.$$

This means that s is bounded from above by $\max\{|s_0|, \ldots, |s_{N+1}|, 1 + |s_{N+1}|\}$.

By **Lemma 3.1.1** s^*_N is bounded for every unlimited hypernatural N, and since s is Cauchy $s^*_N \simeq s^*_M$ for every unlimited hypernaturals N and M. Hence there exists $L \in \mathbb{R}$ such that $s^*_N \simeq L$ for every unlimited hypernatural N. **Theorem 3.1.1** implies that s converges. □

The following theorem is an application of this.

Theorem 3.1.4. *A real sequence $s = (s_n : n \in \mathbb{N})$ converges in $\mathbb{R}$ if any one of the following conditions hold:*

(i) s is bounded above in $\mathbb{R}$ and nondecreasing.

(ii) s is bounded below in $\mathbb{R}$ and nonincreasing.

Proof. Since the proof of (ii) is identical to the proof of (i), up to reversal of the order, it suffices to prove (i). Suppose s is bounded above by B and nondecreasing. Then

$$\mathbb{R} \models (\forall n \in \mathbb{N})(s_1 \leq s_n \leq B)$$

and by transfer we get

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(s^*_1 \leq s^*_n \leq B).$$

Suppose N is an unlimited hypernatural. Then s^*_N is limited, and has a shadow L. We prove that this is the least upper bound of the set $S = \{s_n : n \in \mathbb{N}\}$ in $\mathbb{R}$. It is clearly an upper bound, since

$$\mathbb{R} \models (\forall n, m \in \mathbb{N})(n \leq m \leftrightarrow s_n \leq s_m)$$

and hence we obtain by transfer

$$\mathcal{R} \models (\forall n, m \in \mathbb{N}^*)(n \leq m \leftrightarrow s^*_n \leq s^*_m).$$

Since $n < N$ for every $n \in \mathbb{N}$, we must have $s_n = s^*_n \leq s^*_N \simeq L$. Since s_n and L are real, L is an upper bound in of S. To prove that it is minimal, suppose $r \in \mathbb{R}$ is another upper bound. Then

$$\mathbb{R} \models (\forall n \in \mathbb{N})(s_n \leq r)$$

and by transfer

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(s_n^* \leq r).$$

In particular $s_N^* \leq r$, so $L \leq r$ since $L \simeq s_N^*$. Hence L is minimal, and it is clear that since s is nondecreasing $s_N^* \simeq L$ for every unlimited hypernatural N. Thus s converges to L. □

The above theorem is useful to prove the convergence of a sequence, without knowing the exact limit of it. Another application is calculation of the following limit.

Lemma 3.1.3. *If $0 < c < 1$ and $c \in \mathbb{R}$, then the sequence $s_n = c^n$ converges to 0. In other words, $c^N \simeq 0$ for every unlimited hypernatural N.*

Proof. Since $0 < c < 1$ and $c \in \mathbb{R}$, the sequence $s_n = c^n$ is nonincreasing and bounded from below. Therefore, it converges to some $L \in \mathbb{R}$. In other words, if N is an unlimited hypernatural $c^N \simeq L$ by **Theorem 3.1.1**. Also, note that

$$\mathbb{R} \models (\forall n \in \mathbb{N})(s_{n+1} = cs_n)$$

and by transfer

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(s_{n+1}^* = cs_n^*).$$

Thus

$$L \simeq c^{N+1} = c \cdot c^N \simeq cL$$

and we conclude that $L = cL$ (since both numbers are real), and as $c \neq 1$, we must have $L = 0$. □

Even though there exists sequences which neither diverges to infinity nor converges, it is still possible to break down sequences into more manageble parts, by looking at points to which the sequence comes ”infinitely close”. The following notion makes this intuition precise.

Definition 3.1.4. A *cluster point* of a sequence s is a number $L \in \mathbb{R}$ such that for all $\varepsilon \in \mathbb{R}_+$ there are infinitely many $n \in \mathbb{N}$ such that $|s_n - L| < \varepsilon$.

Closely related concept is the idea of a *subsequence*.

Definition 3.1.5. A *subsequence* $t = (t_n : n \in \mathbb{N})$ of $s = (s_n : n \in \mathbb{N})$ is a sequence for which there exists a strictly increasing sequence of natural numbers $(n_j : j \in \mathbb{N})$ such that $t_m = s_{n_m}$ for all $m \in \mathbb{N}$.

The following connects the two concepts, and show that they coincide.

Lemma 3.1.4. *A sequence $s = (s_n : n \in \mathbb{N})$ has a cluster point at L if and only there exists a subsequence $t = (t_n : n \in \mathbb{N})$ of s which converges to L.*

Proof. "$\Rightarrow$": Suppose that s has a cluster point L. We inductively define a subsequence t which converges to L. Let $t_0 = s_0$. Then set $t_n = s_m$, where m is the least $m \in \mathbb{N}$ such that $|s_m - L| < \frac{1}{n}$ and such that m is greater than any number used in the first n steps (existence of this m is ensured by the definition of a cluster point). By construction, t converges to L.

"$\Leftarrow$": Suppose that t is a subsequence of s converging to L. Then for every $n \in \mathbb{N}$, there exists $m \in \mathbb{N}$ such that $t_n = s_m$. Let $\varepsilon \in \mathbb{R}_+$. By assumption there are infinitely many n such that $|t_n - L| < \varepsilon$. Hence there are infinitely many m such that $|s_m - L| < \varepsilon$. Thus L is a cluster point of s. □

The following is a hyperreal characterization of cluster points, which showcases the intimate connections between cluster points and points of convergence. Note that the first implication uses the characterization of cluster points as points of convergence of subsequences, while the other direction uses **Definition 3.1.4**.

Theorem 3.1.5. *A real sequence $s = (s_n : n \in \mathbb{N})$ has a cluster point at $L \in \mathbb{R}$ if and only if $s^*_N \simeq L$ for some unlimited hypernatural N.*

Proof. "$\Rightarrow$": Suppose s has a cluster point at $L \in \mathbb{R}$. Then there exists a subsequence $t = (t_n : n \in \mathbb{N})$ of s which converges to L. Since

$$\mathbb{R} \models (\forall n \in \mathbb{N})(\exists m \in \mathbb{N})(n \leq m \wedge t_n = s_m),$$

by transfer we obtain

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(\exists m \in \mathbb{N}^*)(n \leq m \wedge t^*_n = s^*_m).$$

By assumption, t converges to L so there is some unlimited N such that $t^*_N \simeq L$. Hence there exists $M \in \mathbb{N}^*$ such that $N \leq M$ (meaning that M is also unlimited) and $s^*_M \simeq L$.

"$\Leftarrow$": Suppose $s^*_N \simeq L \in \mathbb{R}$ for some unlimited hypernatural N. Then we have

$$\mathcal{R} \models (\exists n \in \mathbb{N}^*)(n > m \wedge |s^*_n - L| < \varepsilon)$$

for every $\varepsilon \in \mathbb{R}_+$ and $m \in \mathbb{N}$. By transfer,

$$\mathbb{R} \models (\exists n \in \mathbb{N})(n > m \wedge |s_n - L| < \varepsilon)$$

for every $\varepsilon \in \mathbb{R}_+$. This is exactly the statement that L is a cluster point of s. □

This yields a direct proof of the following famous theorem.

Corollary 3.1.1. Bolzano-Weierstrass: *Every bounded sequence s has at least one cluster point.*

Proof. Since s is bounded, every s^*_N is limited for any unlimited hypernatural N (by **Lemma 3.1.1**). **Theorem 2.4.1** implies that there exists $L \in \mathbb{R}$ such that $s^*_N \simeq L$. Each such L is by **Theorem 3.1.5** a cluster point. □

3.2 Series and Convergence Test

In some circumstances, we wish to consider sums of infinitely many terms, or more precisely a limit of partial sums. Such an expression is called a *series*. Just as before, a series $S = \sum_{i=0}^{\infty} a_i$ gives rise to a *hyperseries* by viewing S a function $S : \mathbb{N} \to \mathbb{R}$ where $S(n)$ is the partial sum of the first $n+1$ terms. In other words, S can be considered a sequence. This yields a suitable concept of convergence.

Definition 3.2.1. We say that a series $\sum_{i=0}^{\infty} a_i$ *converges* if the sequence of partial sums $\left(\sum_{i=0}^{n} a_i : n \in \mathbb{N}\right)$ converges.

For some applications, it is not enough for a series to converge but a stronger criterion is neccessary. For example, a series which converges in the ordinary sense can have its terms arranged so that it diverges, or converges to any other number.

Definition 3.2.2. A series $\sum_{i=0}^{\infty} a_i$ is said to *converge absolutely* if the series $\sum_{i=0}^{\infty} |a_i|$ converge.

Note that the partial sums of $\sum_{i=0}^{\infty} |a_i|$ forms a nondecreasing sequence.

This section is dedicated to proving three standard convergence tests (the comparison, ratio and alternating series tests). First, we prove two lemmas.

Lemma 3.2.1. *For any hypernaturals $n \leq m$, we have*

$$\sum_{i=0}^{m} a_i - \sum_{i=0}^{n} a_i = \sum_{i=n+1}^{m} a_i.$$

Proof. Suppose $S_1 : \mathbb{N}^2 \to \mathbb{R}$ is defined by

$$S_1(n,m) = \sum_{i=0}^{m} a_i - \sum_{i=0}^{n} a_i$$

and $S_2 : \mathbb{N}^2 \to \mathbb{R}$ by

$$S_2(n,m) = \begin{cases} \sum_{i=n}^{m} a_i & \text{if } n \leq m \\ 0 & \text{otherwise} \end{cases}$$

Then we have

$$\mathbb{R} \models (\forall n, m \in \mathbb{N})(n \leq m \to S_1(n,m) = S_2(n,m)).$$

Transfer yields that

$$\mathcal{R} \models (\forall n, m \in \mathbb{N}^*)(n \leq m \to S_1^*(n,m) = S_2^*(n,m))$$

and we conclude that for any hypernaturals $n \leq m$,

$$S_1^*(n,m) = \sum_{i=0}^{m} a_i - \sum_{i=0}^{n} a_i = S_2^*(n,m) = \sum_{i=n}^{m} a_i.$$

□

Lemma 3.2.2. *A series $\sum_{i=0}^{\infty} a_i$ converges in $\mathbb{R}$ if and only if $\sum_{i=N}^{M} \simeq 0$ for every pair $N \geq M$ of unlimited hypernaturals.*

Proof. "⇒": Suppose $N \geq M$ are unlimited hypernaturals, and $\sum_{i=0}^{\infty} a_i$ a series converging to $L \in \mathbb{R}$. Then $\sum_{i=0}^{N} a_i \simeq L \simeq \sum_{i=0}^{M} a_i$ by **Theorem 3.1.1**, so

$$\sum_{i=0}^{N} a_i - \sum_{i=0}^{M} a_i = \sum_{i=M+1}^{N} a_i \simeq 0$$

by definition of $\simeq$ and **Lemma 3.2.1**.

"⇐": We prove that the sequence $s = (s_n = \sum_{i=0}^{n} a_i : n \in \mathbb{N})$ of partial sums is a Cauchy sequence. Suppose $N \leq M$ are unlimited hypernaturals. $s_N^* = \sum_{i=0}^{N} a_i$ and $s_M^* = \sum_{i=0}^{M} a_i$, so we conclude that $s_M^* - s_N^* = \sum_{i=N+1}^{M} a_i \simeq 0$. Thus $s_N^* \simeq s_M^*$, and since N and M were arbitary **Lemma 3.1.2** tells us that s is a Cauchy sequence. □

Now we are ready for the three convergence tests.

Theorem 3.2.1. Comparison Test: *Suppose $\sum_{i=0}^{\infty} a_i$ and $\sum_{i=0}^{\infty} b_i$ are series consisting of nonnegative terms, and suppose that $a_i \leq b_i$ for all $i \in \mathbb{N}$. If $\sum_{i=0}^{\infty} b_i$ converges, then $\sum_{i=0}^{\infty} a_i$ converges.*

Proof. Suppose $\sum_{i=0}^{\infty} a_i$ and $\sum_{i=0}^{\infty} b_i$ are as above. Then there exists $L \in \mathbb{R}$ such that $\sum_{i=0}^{N} b_i \simeq L$ for all unlimited hypernaturals N. Furthermore, the sequences of partial sums are nondecreasing, which implies that L is an upper bound of $\sum_{i=0}^{\infty} b_i$.

Finally

$$\mathbb{R} \models (\forall n \in \mathbb{N}) \left(\sum_{i=0}^{n} a_i \leq \sum_{i=0}^{n} b_i \right)$$

and by transfer

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*) \left(\sum_{i=0}^{n} a_i \leq \sum_{i=0}^{n} b_i \right).$$

In particular, L is an upper bound of $\sum_{i=0}^{n} a_i$ for all $n \in \mathbb{N}^*$. Hence $\sum_{i=0}^{\infty} a_i$ is nondecreasing and bounded from above, so by **Theorem 3.1.4** it converges. □

Theorem 3.2.2. Ratio Test: *Suppose* $\sum_{i=0}^{\infty} a_i$ *is a series, and that the limit*

$$r = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists and is strictly less than 1. *Then* $\sum_{i=0}^{\infty} a_i$ *converges absolutely.*

Proof. Since r exists and is strictly less than 1, we can take $c \in \mathbb{R}$ such that $r < c < 1$. By assumption there exists $N_c \in \mathbb{N}$ such that for all $n > N_c$, we have

$$\left| \frac{a_{n+1}}{a_n} \right| < c.$$

This is implies that

$$|a_{n+1}| < c|a_n|$$

for all $n > N_c$. Generalizing this, we have that $|a_m| < c^{m-N_c} \left|a_{N_c}\right|$. In total, we have

$$\sum_{i=0}^{\infty} |a_i| < \sum_{i=0}^{N_c-1} |a_i| + \left|a_{N_c}\right| \sum_{j=0}^{\infty} c^j.$$

Since $N_c \in \mathbb{N}$, it suffices to prove the convergence of the series $\sum_{j=0}^{\infty} c^j$. Recall the geometric series:

$$\mathbb{R} \models (\forall n \in \mathbb{N})\left(\sum_{i=0}^{n} c_i = \frac{1-c^n}{1-c} \right)$$

By transfer, this yields

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)\left(\sum_{i=0}^{n} c_i = \frac{1-c^n}{1-c} \right)$$

In particular, if N is any unlimited hypernatural we have (by **Lemma 3.1.3**)

$$\sum_{i=0}^{N} c_i = \frac{1-c^N}{1-c} \simeq \frac{1}{1-c}.$$

In other words,

$$\sum_{i=0}^{N} |a_i| < \sum_{i=0}^{N_c-1} |a_i| + \left|a_{N_c}\right| \sum_{j=0}^{N} c^j \simeq \sum_{i=0}^{N_c-1} |a_i| + \left|a_{N_c}\right| \frac{1}{1-c}.$$

Hence $\sum_{i=0}^{N} |a_i|$ is bounded from above, and since it is nondecreasing we have by **Theorem 3.1.4** that it converges. Thus, the series converges absolutely. □

Theorem 3.2.3. Alternating Series Test: *Suppose $s = (s_n : n \in \mathbb{N})$ is a nonincreasing sequence (i.e $s_n \geq s_{n+1}$ for all $n \in \mathbb{N}$) of positive real numbers, converging to 0. Then the alternating series*

$$\sum_{n=0}^{\infty}(-1)^n s_n$$

converges.

Proof. We prove that if $n \geq m$ are natural numbers, then

$$\left|\sum_{i=m}^{n}(-1)^i s_i\right| \leq |s_m|.$$

Without loss of generality, we can assume that m is even, since

$$\left|\sum_{i=m}^{n}(-1)^i s_i\right| = \left|\sum_{i=m}^{n}(-1)^{i+1} s_i\right|$$

Suppose that n is odd. Then

$$\left|\sum_{i=m}^{n}(-1)^i s_i\right| = |s_m - s_{m+1} + s_{m+2} - \ldots - s_{n-2} + s_{n-1} - s_n| \leq$$
$$\leq |s_m - s_{m+1} + s_{m-1} - s_{m+2} \ldots - s_{n+2} + s_{n-2} - s_n| = |s_m - s_n| \leq |s_m|$$

and similarly if n is even.

By transfer of the sentence

$$(\forall n, m \in \mathbb{N})\left(\left|\sum_{i=m}^{n}(-1)^i s_i\right| \leq |s_m|\right)$$

this extends to the hyperseries as well. To finish this proof, it suffices to prove that the sequence of partial sums $s = (s_n = \sum_{i=m}^{n}(-1)^i s_i : n \in \mathbb{N})$ is a Cauchy sequence. Suppose N and M are unlimited hypernaturals. Without loss of generality we can assume $N > M$. Then

$$\left|s_N^* - s_M^*\right| = \left|\sum_{i=M+1}^{N}(-1)^i s_i^*\right| \leq \left|s_{M+1}^*\right|$$

But s converges to 0, so $s_M^* \simeq 0$ for all unlimited hypernaturals M. Thus $s_M^* \simeq s_N^*$. But N and M were arbitary, so **Lemma 3.1.2** implies that s is a Cauchy sequence. Hence the series converges. □

As an application of the comparison test, we prove the following.

Theorem 3.2.4. *If $\sum_{i=0}^{\infty} a_i$ converges absolutely, then $\sum_{i=0}^{\infty} a_i$ converges.*

Proof. Let $\sum_{i=0}^{\infty} a_i$ converge absolutely to $L \in \mathbb{R}$. Clearly, for any $i \in \mathbb{N}$ we have that $0 \le a_i + |a_i| \le 2|a_i|$. Hence, for any $n \in \mathbb{N}$.

$$\sum_{i=0}^{n} (a_i + |a_i|) \le 2 \sum_{i=0}^{n} |a_i| \le L$$

In other words, the sequence $\left(\sum_{i=0}^{n} (a_i + |a_i|) : n \in \mathbb{N}\right)$ is a nondecreasing sequence which is bounded from above. By **Theorem 3.1.4** it converges to some $R \in \mathbb{R}$. Moreover,

$$\mathbb{R} \models (\forall n \in \mathbb{N}) \left(\sum_{i=0}^{n} a_i = \sum_{i=0}^{n} (a_i + |a_i|) - \sum_{i=0}^{n} |a_i| \right)$$

and by transfer

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*) \left(\sum_{i=0}^{n} a_i^* = \sum_{i=0}^{n} \left(a_i^* + |a_i^*|\right) - \sum_{i=0}^{n} |a_i^*| \right).$$

In particular, if N is any unlimited hypernatural, we have that $\sum_{i=0}^{N} |a_i^*| \simeq L$ and $\sum_{i=0}^{N} \left(a_i^* - |a_i^*|\right) \simeq R$, so $\sum_{i=0}^{N} a_i^* \simeq R - L$. Since N is arbitary, the series converges. □

3.3 Limits, Continuity and Halos

From now on, we turn our attentions to functions. Throughout the rest of this thesis, if f (without the star) denotes a function we will assume that it is defined on some subset of $\mathbb{R}$, and not of $\mathcal{R}$. Many functions are analyzed in terms of behaviour near that point. For example, the function $f(x) = \sin x / x$ is not defined at 0, but when we approach $x = 0$, the value of f approaches 1.

Definition 3.3.1. Suppose f is defined in a punctured neighborhood of $a \in \mathbb{R}$ and suppose $L \in \mathbb{R}$. We say that *f approaches L when x goes to a*, or *the limit of f in a is L*, written $\lim_{x \to a} f(x) = L$, if for every $\varepsilon \in \mathbb{R}_+$ there exists $\delta \in \mathbb{R}_+$ so that $|x - a| < \delta$ implies $|f(x) - L| < \varepsilon$.

The following is a nonstandard characterization of limits.

Theorem 3.3.1. *Let $a, L \in \mathbb{R}$. For any function f defined in a punctured neighborhood of a, we have $\lim_{x \to a} f(x) = L$ if and only if $f^*(\alpha) \simeq L$ for each $\alpha \simeq a$.*

Proof. "$\Rightarrow$": Suppose $\lim_{x \to a} f(x) = L$. For every $\varepsilon \in \mathbb{R}_+$ there exists $\delta \in \mathbb{R}_+$ such that

$$\mathbb{R} \models \forall x \left(|x - a| < \delta \rightarrow |f(x) - L| < \varepsilon\right).$$

By transfer we obtain

$$\mathcal{R} \models \forall x(|x-a| < \delta \to |f^*(x) - L| < \varepsilon).$$

Let $\alpha \simeq a$. Then $|\alpha - a| \simeq 0$, so $|\alpha - a| < \delta$ for every $\delta \in \mathbb{R}_+$. Hence $|f^*(\alpha) - L| < \varepsilon$ for every $\varepsilon \in \mathbb{R}_+$, and we conclude that $f^*(\alpha) \simeq L$.

"$\Leftarrow$": Let $a, L \in \mathbb{R}$, and suppose that for every $\alpha \simeq a$, we have $f^*(\alpha) \simeq L$. Then for any infinitesimal δ and $x \in \mathcal{R}$, we have that if $|x-a| < \delta$ then $x \simeq a$. Thus

$$\mathcal{R} \models \exists\delta(|x-a| < \delta \to |f^*(x) - L| < \varepsilon)$$

for all $\varepsilon \in \mathbb{R}_+$. Transfer yields that

$$\mathbb{R} \models \exists\delta(|x-a| < \delta \to |f(x) - L| < \varepsilon)$$

for all $\varepsilon \in \mathbb{R}_+$, which precisely means that $\lim_{x\to a} f(x) = L$. □

Limits gives rise to an exact definition of a continuity, and naturally a nonstandard account of continuity.

Definition 3.3.2. A function f defined in a neighborhood of a point a is continous at a if $\lim_{x\to a} f(x) = f(a)$.

Theorem 3.3.2. *A function f defined in a neighborhood of a is continous at $a \in \mathbb{R}$ if and only if $f^*(\alpha) \simeq f(a)$ for every $\alpha \simeq a$.*

Proof. Let $L = f(a)$ in **Theorem 3.3.1**. □

The following is an application.

Example 3.3.1. The function $\exp : \mathbb{R} \to \mathbb{R}$ is continous at every $a \in \mathbb{R}$.

Proof. Let $\alpha \simeq a \in \mathbb{R}$. Then there exists $\varepsilon \simeq 0$ such that $\alpha = a + \varepsilon$.

By **Example 2.4.1** and **Example 2.3.2**, we get

$$\exp^*(\alpha) = \exp^*(a+\varepsilon) = \exp^*(a)\exp^*(\varepsilon) \simeq \exp^*(a),$$

so exp is continous. □

Recall that if $a, b \in \mathbb{R}$, the notation $[a, b]$ refers to the real, closed interval between a and b, that is $[a, b] = \{r \in \mathbb{R} : a \leq r \leq b\}$ and (a, b) refers to the corresponding open interval.

The Intermediate and Extreme Value Theorems are considered highlights of calculus. In standard analysis, the most common proof involves proving that a continous function defined on a closed interval $[a, b]$ is bounded, and then using the Completeness property and Bolzano-Weierstrass to find a sequence which converges to the number we seek.

In nonstandard analysis there is another way.

Theorem 3.3.3. The Intermediate Value Theorem: *Suppose that* $a, b \in \mathbb{R}$ *and* f *is continous on a closed, nonempty interval* $[a,b]$*. For every* $d \in (f(a), f(b))$*, there exists a number* $c \in (a,b)$ *such that* $f(c) = d$*.*

Proof. The basic idea behind this proof is quite appealing. We partition the interval $[a,b]$ into subintervals of infinitesimal length, and find an interval which endpoints are on either side of d. Then c is the common shadow of these endpoints.

More explicitly, let $n \in \mathbb{N}$, and set

$$p_k = a + k\frac{b-a}{n}$$

for $0 \leq k \leq n$. Since $p_0 = a$ and $p_n = b$, the set $\{p_k : f(p_k) < d\}$ is nonempty and finite. Hence it contains a maximal element. Take s_n as this element. This way we obtain a sequence $(s_n : n \in \mathbb{N})$, for which we have:

$$\mathbb{R} \models (\forall n \in \mathbb{N})\left([a \leq s_n < b] \wedge \left[f(s_n) < d < f\left(s_n + \frac{b-a}{n}\right)\right]\right).$$

Transfer yields

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)\left([a \leq s_n^* < b] \wedge \left[f^*(s_n^*) < d < f^*\left(s_n^* + \frac{b-a}{n}\right)\right]\right).$$

Suppose N is an unlimited natural. Then s_N^* is limited, so it has a shadow $c \in \mathbb{R}$. Moreover, since $\frac{b-a}{N}$ is infinitesimal, we have

$$s_N^* + \frac{b-a}{N} \simeq s_N^* \simeq c.$$

Since f is continous and c is real, it follows that $f^*\left(s_N^* + \frac{b-a}{N}\right) \simeq f^*(c)$ and $f(s_N^*) \simeq f^*(c)$. However, we also have that

$$f^*(s_N^*) < d < f^*\left(s_N^* + \frac{b-a}{N}\right).$$

In other words, $d \simeq f^*(c) = f(c)$, and since c and d both are real $f(c) = d$. $\square$

The other theorem is proved similarly.

Theorem 3.3.4. The Extreme Value Theorem: *Suppose* f *is continous on a closed, nonempty interval* $[a,b]$*. Then there exists* $c, d \in [a,b]$ *such that for all* $x \in [a,b]$*, we have* $f(c) \leq f(x) \leq f(d)$*.*

Proof. Since finding the minimum of f is equivalent to finding the maximum of $-f$, it suffices to prove that f has a maximum. Let $n \in \mathbb{N}$, and set

$$p_k = a + k\frac{b-a}{n}.$$

where $0 \leq k \leq n$. Since n is finite, the set $\{f(p_k) : 0 \leq k \leq n\}$ has a maximal element, and we let s_n equal this p_k.

In this way, we obtain a sequence $(s_n : n \in \mathbb{N})$ which is bounded. By **Lemma 3.1.1** s_n^* is limited for all $n \in \mathbb{N}^*$. Suppose N is any unlimited hypernatural. Then s_N^* has a shadow $d \in \mathbb{R}$. By continuity of f, we have $f^*(s_N^*) \simeq f(d)$.

Thus, we have found a d such that $f(d)$ is maximal among $f(p_k)$ for all k. However, this does not prove in itself that $f(d)$ is maximal for *all* $x \in f([a,b])$. To show this, we prove that

$$P = \left\{a + k\frac{b-a}{N} : k \in \mathbb{N}^*, 0 \leq k \leq N\right\}$$

is an infinitely close approximation of $[a,b]$, in the sense that every $x \in [a,b]$ is infinitely close to a point from P.

Let $x \in [a,b]$. Then we have

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

By transfer we have

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(\exists k \in \mathbb{N}^*)\left(k < n \wedge \left[a + k\frac{b-a}{n} \leq x \leq a + (k+1)\frac{b-a}{n}\right]\right).$$

In particular, if N is an unlimited hypernatural, there exists $K < N$ such that

$$a + K\frac{b-a}{N} \leq x \leq a + (K+1)\frac{b-a}{N}.$$

However, the difference

$$a + K\frac{b-a}{N} - \left(a + (K+1)\frac{b-a}{N}\right) = \frac{b-a}{N}$$

is infinitesimal, and thus $x \simeq a + K\frac{b-a}{N}$. By the continuity of f,

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

But $f^*\left(a + K\frac{b-a}{N}\right) \leq f^*(s_N^*)$ by definition, and thus

$$f^*(x) \simeq f^*\left(a + K\frac{b-a}{N}\right) \leq f^*(s_N^*) \simeq f^*(d).$$

Hence $f(x) \leq f(d)$ for all $x \in [a,b]$, and $f(d)$ is a maximum of f on $[a,b]$. □

The notion of *uniform continuity* is important in real analysis. Unlike regular continuity, which is defined both locally (as in *continuous at a point*) and globally (*continous in a set*), uniform continuity is only defined globally.

Definition 3.3.3. A function $f : A \to \mathbb{R}$ defined on an open set $A \subset \mathbb{R}$ is *uniformly continous* if for every $\varepsilon \in \mathbb{R}_+$, there exists a $\delta \in \mathbb{R}_+$ so that for every $x, y \in A$, $|x - y| < \delta$ implies $|f(x) - f(y)|$.

Remark. Uniform continuity implies regular continuity, the difference being that if we fix a $\varepsilon \in \mathbb{R}_+$, then a function is uniformly continous if there exists a *fixed* $\delta \in \mathbb{R}_+$ such $|x - y| < \delta$ forces $|f(x) - f(y)| < \varepsilon$ for *any* $x, y \in A$.

Regular continuity only demands that for every $x, y \in A$, there should exist such a δ. In other words δ may depend on the points x and y, as well as ε.

Here are two examples.

Example 3.3.2. The function $f(x) = x$ is uniformly continous.

Proof. For every ε, take $\delta = \varepsilon$. Then $|x - y| < \varepsilon$ implies $|f(x) - f(y)| = |x - y| < \varepsilon = \delta$ for any $x, y \in \mathbb{R}$. □

Example 3.3.3. The function $\exp : \mathbb{R} \to \mathbb{R}$ is not uniformly continous.

Proof. Suppose exp is uniformly continous, and let $\varepsilon \in \mathbb{R}_+$. Then there exists $\delta \in \mathbb{R}_+$ such that $|x - y| < \delta$ implies $|\exp x - \exp y| < \varepsilon$ for all $x, y \in \mathbb{R}$.

Let $y = x + \frac{\delta}{2}$. Then $|x - y| = \frac{\delta}{2} < \delta$. However, $|f(x) - f(y)| = \exp x \exp(\frac{\delta}{2} - 1)$ is unbounded, so taking x large enough we have $|f(x) - f(y)| > \varepsilon$, a contradiction. Hence $\exp : \mathbb{R} \to \mathbb{R}$ is not uniformly continous. □

The following theorem provides a nonstandard version of uniform continuity.

Theorem 3.3.5. *A function f is uniformly continous on A if and only if $f^*(\alpha) \simeq f^*(\beta)$ for all hyperreals $\alpha \simeq \beta$ in A^*.*

Proof. "$\Rightarrow$": Suppose that f is uniformly continous on A. Then for all $\varepsilon \in \mathbb{R}_+$ there exists $\delta \in \mathbb{R}_+$ such that

$$\mathbb{R} \models (\forall x, y \in A)(|x - y| < \delta \to |f(x) - f(y)| < \varepsilon).$$

Transfer gives us

$$\mathcal{R} \models (\forall x, y \in A^*)(|x - y| < \delta \to |f^*(x) - f^*(y)| < \varepsilon).$$

Suppose $\alpha \simeq \beta$ are hyperreals in A^*. Then $|\alpha - \beta| < \delta$ for all $\delta \in \mathbb{R}_+$, which implies that $|f^*(\alpha) - f^*(\beta)| < \varepsilon$ for every $\varepsilon \in \mathbb{R}_+$. Hence $f^*(\alpha) \simeq f^*(\beta)$.

"$\Leftarrow$": Suppose that for every hyperreals $\alpha \simeq \beta$ in A^*, we have $f^*(\alpha) \simeq f^*(\beta)$. Then for any infinitesimal δ and $x, y \in A^*$, we have that $|x - y| < \delta$ implies $|f^*(x) - f^*(y)| < \varepsilon$, for any $\varepsilon \in \mathbb{R}_+$. Hence

$$\mathcal{R} \models (\exists\delta)(x, y \in A^*)(|x - y| < \delta \to |f^*(x) - f^*(y)| < \varepsilon)$$

for every $\varepsilon \in \mathbb{R}_+$. Transfer yields that

$$\mathbb{R} \models (\exists\delta)(x, y \in A)(|x - y| < \delta \to |f(x) - f(y)| < \varepsilon).$$

In other words, f is uniformly continous. □

Remark. This theorem highlights the difference between regular and uniform continuity. A function f is regular continous at the real point a if and only if $\alpha \simeq a$ implies that $f^*(a) \simeq f^*(\alpha)$. In other words, regular continuity demands that the relation $\simeq$ is preserved by f^* only if *one of the points is real*. Uniform continuity on the contrary demands that $\simeq$ is preserved at *all* points in the domain of f^*, that is, between arbitary hyperreals in the domain.

Of course, the two notions might coincide.

Corollary 3.3.1. *If a function f is continous on a closed interval $[a, b]$, then f is uniformly continous on $[a, b]$.*

Proof. Let $\alpha, \beta \in [a, b]^*$ and $\alpha \simeq \beta$. Suppose c is the shadow of α. Since $a \leq \alpha \leq b$, we must have $c \in [a, b]$. Thus f is continous at c, and we have $f(\alpha) \simeq f(c)$ and $f(\beta) \simeq f(c)$. Hence $f(\alpha) \simeq f(\beta)$ and by **Theorem 3.3.5**, f is uniformly continous. □

Some uniformly continous functions are better behaved than others, and of special interest are those functions which satisfy the following condition.

Definition 3.3.4. A function $f : \mathbb{R} \to \mathbb{R}$ is a *Lipschitz function* if there exists $c \in \mathbb{R}$ (called a *Lipschitz constant*) such that $|f(x) - f(y)| \leq c|x - y|$ for all $x, y \in \mathbb{R}$. A Lipschitz function such that $c < 1$ is called a *contraction mapping* (or simply a *contraction*).

Here is an example of a Lipschitz function.

Example 3.3.4. The function $f(x) = \frac{x}{3} + 2$ is a Lipschitz function, we can take $c = \frac{1}{3}$.

Remark. Every Lipschitz function is uniformly continous, which we see by taking $\delta = \frac{\varepsilon}{c}$. Then $|x-y| < \delta$ implies that $|f(x)-f(y)| < c|x-y| < \varepsilon$, so f is uniformly continous.

A contraction mapping f acts on any two points $x, y \in \mathbb{R}$ by moving and them closer together, since $|f(x) - f(y)| < |x - y|$. Intuitively, we can see that f contracts $\mathbb{R}$ by "pushing it together". By this reasoning, the function $f(f(x))$ contracts $\mathbb{R}$ more, and $f(f(f(x)))$ takes the points even closer together. This observation is at the core of the following theorem.

Theorem 3.3.6. *Any contraction $f : \mathbb{R} \to \mathbb{R}$ has a unique fixed point.*

Proof. Suppose $c < 1$ is a Lipschitz constant of f. Take $x \in \mathbb{R}$, define a sequence $(s_n : n \in \mathbb{N})$ by

$$s_n = \begin{cases} x & \text{if } n = 0 \\ f(s_{n-1}) & \text{if } n \geq 1 \end{cases}$$

Note that $|s_n - s_{n+1}| \leq c|s_{n-1} - s_n| \leq c^n|s_0 - s_1|$, which implies that

$$|s_0 - s_n| \leq \sum_{i=0}^{n} |s_i - s_{i+1}| \leq \sum_{i=0}^{n} c^i |s_0 - s_1| = \frac{1 - c^n}{1 - c}|s_0 - s_1|.$$

Since $c \geq 0$, we have

$$|s_0 - s_n| \leq \frac{1}{1 - c}|s_0 - s_1|.$$

Let N be some unlimited hypernatural. From the above inequality, we conclude that s is bounded. Hence s^*_N has a shadow $L \in \mathbb{R}$. Since f is continous, $f^*(s^*_N) \simeq f(L)$. But $f(s^*_N) = s^*_{N+1}$ by definition, and hence

$$|s^*_N - s^*_{N+1}| \leq c^N |s_0 - s_1|.$$

Since $c < 1$, we have $c^N \simeq 0$ (by **Lemma 3.1.3**) and thus $s^*_N \simeq s^*_{N+1}$. Putting it all together, we get

$$f(L) \simeq f(s^*_N) = s^*_{N+1} \simeq s^*_N \simeq L.$$

Since both $f(L)$ and L are real, we conclude that $f(L) = L$.

For uniqueness, suppose that S is a fixed point of f and recall that by definition

$$0 \leq |L - S| = |f(L) - f(S)| \leq c|L - S|$$

for some $|c| < 1$. Hence $0 \leq (1 - c)|L - S| \leq 0$, and consequently $L = S$. □

3.4 Sequences of Functions

So far, we have focused on sequences and series of real numbers, and continous functions. In this section, we consider sequences of functions and limits of these. First, we define more precisely what the hypersequence of $(f_n : n \in \mathbb{N})$ is.

Definition 3.4.1. Suppose $(f_n : n \in \mathbb{N})$ is a sequence of functions defined on A. Then the sequence $(f^*_n : n \in \mathbb{N}^*)$ refers to the hypersequence determined by the extension of $F : \mathbb{N} \times A \to \mathbb{R}$ such that $F(n, x) = f_n(x)$. In other words, $f^*_n(x) = F^*(n, x)$.

There are different ways for a sequence $(f_n : n \in \mathbb{N})$ of functions to converge to a function f. We handle the two of the most common: *pointwise convergence* and *uniform convergence*.

Definition 3.4.2. A sequence of functions $(f_n : n \in \mathbb{N})$ defined on A *converges pointwise* to $f : A \to \mathbb{R}$ if for all $x \in \mathbb{R}$ we have $\lim_{n\to\infty} f_n(x) = f(x)$.

Remark. Equivalently, f_n converges to f pointwise if for all $x \in A$ and $\varepsilon \in \mathbb{R}_+$, there exists $N_{\varepsilon,x} \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ for which $n > N_{\varepsilon,x}$, we have $|f_n(x) - f(x)| < \varepsilon$.

Our discussion on limits and convergence of sequences immediately yields the following result.

Theorem 3.4.1. *A sequence $(f_n : n \in \mathbb{N})$ of functions defined on $A \subset \mathbb{R}$ converges pointwise to the function $f : A \to \mathbb{R}$ if and only if for each $x \in A$ and unlimited hypernatural N we have $f_N^*(x) \simeq f(x)$.*

Proof. Suppose $(s_n : n \in \mathbb{N})$ is the sequence defined by $s_n = f_n(x)$. The result then follows from **Theorem 3.1.1**. □

The other convergence which we study is uniform convergence, a stronger condition of which implies pointwise convergence. As with the notion of uniform continuity, a sequence of functions cannot converge uniformly in a single point, but only on subsets of $\mathbb{R}$.

Definition 3.4.3. A sequence $(f_n : n \in \mathbb{N})$ defined on $A \subset \mathbb{R}$ *converges uniformly* to $f : A \to \mathbb{R}$ if for every $\varepsilon \in \mathbb{R}_+$, there exists an $N_\varepsilon \in \mathbb{N}$ such that for all $x \in A$ and $n \in \mathbb{N}$ such that $n > N_\varepsilon$ we have $|f_n(x) - f(x)| < \varepsilon$.

Remark. The attentive reader notices the similarity between uniform convergence and uniform continuity. Uniform continuity demands that there exists a $\delta \in \mathbb{R}_+$ which works on the entire domain of f, while uniform convergence demands the existence of an $N_\varepsilon \in \mathbb{N}$ which works for every x in the domain of f_N.

This connection becomes clear in the following theorem.

Theorem 3.4.2. *A sequence f_n of functions defined on $A \subset \mathbb{R}$ converges uniformly to the function $f : A \to \mathbb{R}$ if and only if for each $x \in A^*$ and unlimited hypernatural N we have $f_N^*(x) \simeq f(x)$.*

Proof. "$\Rightarrow$": Suppose that $(f_n : n \in \mathbb{N})$ converges uniformly to the function $f : A \to \mathbb{R}$. Then, for any $\varepsilon \in \mathbb{R}_+$ there exists $N_\varepsilon \in \mathbb{N}$ so that

$$\mathbb{R} \models (\forall n \in \mathbb{N})(\forall x \in A)(n > N_\varepsilon \to |f_n(x) - f(x)| < \varepsilon).$$

By transfer we have that

$$\mathcal{R} \models (\forall n \in \mathbb{N}^*)(\forall x \in A^*)(n > N_\varepsilon \to |f_n^*(x) - f^*(x)| < \varepsilon)$$

If N is an unlimited hypernatural, then $N > N_\varepsilon$ for any $N_\varepsilon \in \mathbb{N}$. We conclude that $|f_N^*(x) - f(x)| < \varepsilon$ for any $\varepsilon \in \mathbb{R}_+$. Thus $f_N^*(x) \simeq f^*(x)$ for any $x \in A^*$.

"$\Leftarrow$": Suppose that for every $x \in A^*$ and unlimited hypernatural N, the number $f_N^*(x) \simeq f^*(x)$. Then

$$\mathcal{R} \models (\exists n \in \mathbb{N}^*)(\forall x \in A^*)(m > n \to |f_m^*(x) - f^*(x)| < \varepsilon)$$

for every $\varepsilon \in \mathbb{R}_+$. Transfer yields that

$$\mathbb{R} \models (\exists n \in \mathbb{N})(\forall x \in A)(m > n \to |f_m(x) - f(x)| < \varepsilon).$$

In other words, f_n converges uniformly to f. □

The final theorem of this thesis highlights one of the differences between uniform and pointwise convergence; uniform convergence of a sequence assures us that the limit is retains properties of the functions in the sequence.

Theorem 3.4.3. Uniform Convergence Theorem: *If a sequence f_n of continous functions defined on $A \subset \mathbb{R}$ converges uniformly to the function $f : A \to \mathbb{R}$, then f is continous as well.*

Proof. Let $c \in A$, and take any $x \simeq c$. Then, for any $n \in \mathbb{N}$ we have

$$|f^*(x) - f^*(c)| \leq |f^*(x) - f_n^*(x)| + |f_n^*(x) - f_n^*(c)| + |f_n^*(c) - f^*(c)|.$$

Let $\varepsilon \in \mathbb{R}_+$. Since f_n is continous, $f_n^*(x) \simeq f_n^*(c)$, in particular $|f_n^*(x) - f_n^*(c)| < \frac{\varepsilon}{3}$ for all $n \in \mathbb{N}$. Since f_n converges to f uniformly, we can find N_ε so that $m > N_\varepsilon$ implies $|f_m^*(c) - f^*(c)| < \frac{\varepsilon}{3}$ for all $c \in A$. Fix such an m, for example $m = N_\varepsilon + 1$. Then

$$\mathbb{R} \models (\forall y \in A)\left(|f_m(y) - f(y)| < \frac{\varepsilon}{3}\right)$$

and transfer yields

$$\mathcal{R} \models (\forall y \in A^*)\left(|f_m^*(y) - f^*(y)| < \frac{\varepsilon}{3}\right).$$

This implies that $|f_m^*(x) - f^*(x)| < \frac{\varepsilon}{3}$. Putting all this together,

$$|f^*(x) - f^*(c)| \leq |f^*(x) - f_m^*(x)| + |f_m^*(x) - f_m^*(c)| + |f_m^*(c) - f^*(c)| < \varepsilon$$

In other words, f is continous. □

Remark. The converse of the above theorem is false, the sequence $(x^n : n \in \mathbb{N})$ consists of continous functions and converges pointwise (but not uniformly) to the function $f(x) = 0$, which is also continous.

Also, there are sequences of continous functions which converge pointwise to discontinous functions. As an example, take the sequence $(f_n : [0,1] \to \mathbb{R} : n \in \mathbb{N})$ defined by

$$f_n(x) = \begin{cases} 1 - nx & \text{if } 0 \leq x \leq \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

For all $n \in \mathbb{N}$, we have $\frac{1}{n} > 0$ and thus $f_n(0) = 1$. But for all $x \in (0,1]$, there exists n so that $\frac{1}{n} < x$. This implies that f_n converges pointwise to the function

$$f(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

which is discontinous.

Bibliography

[1] Abraham Robinson, *Non-standard Analysis*. Princeton University Press, 1974.

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

[3] Philipp Rothmaler, *Introduction to Model Theory*. Taylor & Francis, 2000.

[4] Shawn Hedman, *A First Course in Logic*. Oxford University Press, 2004.

[5] Yiannis Moschovakis, *Notes on Set Theory*, 2nd edition. Springer Verlag, 2006.