UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

ULTRAPRODUCTS AND THEIR APPLICATION TO NON-STANDARD ANALYSIS

Autor: Miriam Pérez López-Serrano

Director: Dr. Enrique Casanovas Ruiz-Fornells
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 13 de juny de 2022

Abstract

The intuitive notion of infinitesimals has been used in mathematical arguments for many years. However, they are still seen as controversial. The aim of this work is to give an introduction to ultraproducts, bring the reader closer to non-standard analysis and to prove that it is rigorously defined. In order to do that, we introduce the concept of reduced product of structures, as well as we give an ultraproduct version of the compactness theorem of first-order logic. We also build the set of hyperreal numbers and give some known results of calculus using infinitesimals.

2020 Mathematics Subject Classification. 03C07, 03C20, 26E35

Acknowledgements

En primer lugar, me gustaría dar las gracias al Dr. Enrique Casanovas por haberme guiado durante el desarrollo del trabajo y haberme ayudado siempre que lo he necesitado. Gracias por tu infinita paciencia.

Agradecer también a mis amigos y amigas estos años llenos de buenos momentos. Y a mi familia, por estar siempre ahí.

Finalmente, quiero darle las gracias a David, por apoyarme incondicionalmente y confiar en mí como lo hace.

Contents

Introduction

The notion of an infinitely small number has been used in mathematical arguments for hundreds of years. But it was in the seventeenth century when the intuitive concept of these numbers, called infinitesimals, was crucial to the development of calculus, specially in the work of Leibniz and Newton. However, since infinitesimals were not formally defined, in the late nineteenth century they were rejected and replaced by the ε, δ method, which is the one used nowadays, at least for most mathematicians.

In 1960, the mathematician Abraham Robinson, who was interested in infinitesimals, gave a rigorous development of the calculus based on them. He wrote in his book *Non-standard Analysis* [8]:

> [...] the idea of infinitely small or infinitesimal quantities seems to appeal naturally to our intuition. At any rate, the use of infinitesimals was widespread during the formative stages of the Differential and Integral Calculus. As for the objection [...] that the distance between two distinct real numbers cannot be infinitely small, Gottfried Wilhelm Leibniz argued that the theory of infinitesimals implies the introduction of ideal numbers which might be infinitely small or infinitely large compared with the real numbers but which were to possess the same properties as the latter.

Robinson's development was based on model theory, a branch of mathematical logic. His method is called non-standard analysis because it uses a non-standard model of analysis, and the fundamental way of constructing structures in it is the ultrapower.

Years later, some mathematicians started studying this method. One of them is H. Jerome Keisler, who published the first edition of the monograph *Foundations of Infinitesimal Calculus* in 1976. This monograph is a companion of his textbook, which is centered on developing calculus using infinitesimals, instead of the ε, δ approach, and its latest edition [4] is the reference book most used for the elaboration of this project.

Non-standard analysis has been applied to many areas of mathematics and other sciences, such as physics. Nevertheless, this method is still seen as controversial and most mathematicians seem reluctant to use it, since they are unfamiliar to it.

When I started to read about this topic, I found really curious that even though infinitesimals gave the intuition for the original development of the calculus, they do not appear in standard presentations of formal calculus. Therefore, the aim of this work is to bring the reader closer to non-standard analysis, introducing ultraproducts and building the set of hyperreal numbers from the ultrapower of $\mathbb{R}$, and to prove that it is rigorously defined. Usually, we will give alternate proofs of standard results by using infinitesimals. Some results will be left without proof if this one is based on standard methods only, but a reference to check them will always be given.

In the first chapter, we give some definitions and results related to first-order logic that are necessary to define reduced products of structures and, in particular, the ultraproducts.

The second chapter contains concepts such as ultrafilters or ultraproducts in order to characterize the hyperreal numbers in Chapter 3. One of the main results proved in this chapter is the Fundamental Theorem of Ultraproducts, also known as Łoś Theorem, which states the denotation and the satisfaction of terms and formulas, respectively. Furthermore, an ultraproduct version of the well-known compactness theorem of first-order logic is shown.

In the last chapter, we define the set of real numbers as an structure $\mathbb{R}$ and we give an axiomatization of its complete first-order theory. Once this is done, we are ready to define the hyperreal numbers as an elementary extension of $\mathbb{R}$. Finally, we give a rigorous development of non-stardard analysis, including some topological concepts, derivatives, continuous functions and integration.

Chapter 1

Preliminaries

Throughout this project we will work by using first-order logic. So, before getting into the main subject, we need to define some basic notions.

In first-order logic, we have a set of non-logical symbols, called *language*. These symbols can be *constants*, *function symbols* and *relation symbols*. The two last ones have a natural number associated with them $n \geq 1$, which is their *arity*.

We also have *terms*, which are a finite sequence of symbols built according to the following rules:

1. Every variable is a term.

2. Every constant is a term.

3. If $t_1, \ldots, t_n$ are terms and F is an n-ary function, then $F(t_1, \ldots, t_n)$ is a term.

In addition, first order logic has the same logical connectives as propositional logic $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\rightarrow$ (implies), $\leftrightarrow$ (if and only if) and also the equality symbol $\doteq$ and quantifiers $\forall$ (for all) and $\exists$ (exists).

An *equation* is an expression of the form $t_1 \doteq t_2$, where t_1, t_2 are terms. We call *atomic formulas* the equations and the formulas of the form $R(t_1, \ldots, t_n)$, where R is an n-ary relation symbol and $t_1, \ldots, t_n$ are terms.

The *formulas* are expressions defined as follows:

1. Every atomic formula φ is a formula.

2. If φ, ψ are formulas, so are

$$\neg\varphi, \quad (\varphi \wedge \psi), \quad (\varphi \vee \psi), \quad (\varphi \rightarrow \psi), \quad (\varphi \leftrightarrow \psi).$$

3. If φ is a formula and x is a variable, then $\forall x\varphi$ and $\exists x\varphi$ are formulas.

Whenever a quantifier $Q = \forall, \exists$ appears in a formula φ, it is immediately followed by a variable and then comes a subformula ψ of φ. The formula $Qx\psi$ is called the *scope* of that particular occurrence of the quantifier Q in φ. Every occurrence of the variable x that appears in that formula is said to be bound by Q. The *free variables* of a formula φ are those variables x that have at least one occurrence which is not within the scope of any quantifier that binds the variable x. On the other hand, *bound variables* of φ are those that have at least one occurrence which is within such a subformula.

Given a list of different variables $x_1, \ldots, x_n$, we denote a term whose variables are in that list by $t(x_1, \ldots, x_n)$ and we denote a formula whose free variables are in the list by $\varphi(x_1, \ldots, x_n)$. We will write $t = t(x_1, \ldots, x_n)$ and $\varphi = \varphi(x_1, \ldots, x_n)$ if there is no possible confusion. A first-order formula with no free variables is called a *sentence*.

A *structure* of language L, also called L-*structure*, is a pair $\mathcal{M} = (M, I)$, where M is a nonempty set, called the *universe* of the structure, and I is a map, called *interpretation*, with domain L such that:

1. For every constant $c \in L$, $I(c) \in M$.
2. For every n-ary function symbol $F \in L$, $I(F) : M^n \to M$.
3. For every n-ary relation symbol $R \in L$, $I(R) \subseteq M^n$.

If $\mathcal{M} = (M, I)$, we usually use the notation $\xi^{\mathcal{M}}$ instead of $I(\xi)$ for every symbol $\xi \in L$. Therefore, from now on we will write $\mathcal{M} = (M, \xi^{\mathcal{M}})_{\xi \in L}$.

The *cardinal* of a substructure $\mathcal{M}$ is the cardinal of its universe M.

Let $\mathcal{M}$ be an L-structure, t a term and φ a formula of L. An *assignment* s is a function whose domain is a set of variables and whose range is a subset of M. We say that an assignment s is defined for t if its domain contains all variables of t. Similarly, s is defined for φ if its domain contains all free variables of φ. Given an assignment s, a variable x and an element $a \in M$, s^x_a is defined as the assignment such that:

1. $\mathrm{dom}(s^x_a) = \mathrm{dom}(s) \cup \{x\}$
2. $s^x_a(x) = a$
3. $s^x_a(y) = s(y)$ for any other variable $y \neq x$ of the domain of s

The *denotation* of t in $\mathcal{M}$ under an assignment s defined for t, denoted by $t^{\mathcal{M}}[s]$, is defined recursively as follows:

1. $x^{\mathcal{M}}[s] = s(x)$

2. $c^{\mathcal{M}}[s] = c^{\mathcal{M}}$

3. $F^{\mathcal{M}}(t_1, \ldots, t_n)[s] = F^{\mathcal{M}}(t_1^{\mathcal{M}}[s], \ldots, t_n^{\mathcal{M}}[s])$

If $t = t(x_1, \ldots, x_n)$ and $s = \{(x_1, a_1), \ldots, (x_n, a_n)\}$, we will usually write $t^{\mathcal{M}}[a_1, \ldots, a_n]$ instead of $t^{\mathcal{M}}[s]$.

The *satisfaction* $\mathcal{M} \models \varphi[s]$ of φ in $\mathcal{M}$ under an assignment s defined for φ is defined recursively as follows:

1. $\mathcal{M} \models t_1 \doteq t_n$ if and only if $t_1^{\mathcal{M}}[s] = t_2^{\mathcal{M}}[s]$

2. $\mathcal{M} \models R(t_1, \ldots, t_n)[s]$ if and only if $(t_1^{\mathcal{M}}[s], \ldots, t_n^{\mathcal{M}}[s]) \in R^{\mathcal{M}}$

3. $\mathcal{M} \models \neg\varphi[s]$ if and only if $\mathcal{M} \not\models \varphi[s]$

4. $\mathcal{M} \models (\varphi \land \psi)[s]$ if and only if $\mathcal{M} \models \varphi[s]$ and $\mathcal{M} \models \psi[s]$. We define it similarly for the other connectives.

5. $\mathcal{M} \models \exists x\varphi[s]$ if and only if there exists $a \in M$ such that $\mathcal{M} \models \varphi[s_a^x]$

6. $\mathcal{M} \models \forall x\varphi[s]$ if and only if for each $a \in M$, $\mathcal{M} \models \varphi[s_a^x]$

As above, if we have $\varphi = \varphi(x_1, \ldots, x_n)$ and $s = \{(x_1, a_1), \ldots, (x_n, a_n)\}$, we will write $\mathcal{M} \models \varphi[a_1, \ldots, a_n]$ instead of $\mathcal{M} \models \varphi[s]$.

Now we will give the definition of model, which is one of the main concepts. But, first, we need a couple of new definitions.

Let $\mathcal{M}$ be a L-structure, s an assignment in $\mathcal{M}$ and let Σ be a set of formulas of language L. We write $\mathcal{M} \models \Sigma[s]$ if for each $\varphi \in \Sigma$ we have $\mathcal{M} \models \varphi[s]$. The set Σ is said to be *satisfiable* if there exist an L-structure $\mathcal{M}$ and an assignment s in $\mathcal{M}$ defined for the formulas of Σ such that $\mathcal{M} \models \Sigma[s]$. We say that a formula of L φ is a *consequence* of Σ, and we write $\Sigma \models \varphi$, if for each L-structure $\mathcal{M}$ and each assignment s in $\mathcal{M}$ defined for the formulas of Σ and for φ such that $\mathcal{M} \models \Sigma[s]$, we have $\mathcal{M} \models \varphi[s]$.

Given a set of sentences Σ, i.e. a set of formulas without free variables, the satisfiability of Σ can be expressed without assignments, that is there exists a structure $\mathcal{M}$ such that $\mathcal{M} \models \Sigma$. In that case, we say that the structure $\mathcal{M}$ is a *model* of Σ.

We now give some definitions related to structures which will be needed in the next chapters.

Definition 1.1. *Let L, L' be languages such that $L \subseteq L'$ and let $\mathcal{M}$ be an L-structure. Let I' be any interpretation for the symbols of $L' \smallsetminus L$ in $\mathcal{M}$. Then $\mathcal{M}' = (M, I \cup I')$ is an L'-structure and we say that it is an* expansion *of $\mathcal{M}$ to L'. We can also say that $\mathcal{M}$ is the* reduct *of $\mathcal{M}'$ to L, and we write $\mathcal{M} = \mathcal{M}' \restriction_L$.*

Definition 1.2. *We define the theory of a structure $\mathcal{M}$, and we denote it by* $\mathrm{Th}(\mathcal{M})$, *as the set of sentences that hold in $\mathcal{M}$, that is*

$$\mathrm{Th}(\mathcal{M}) = \{\varphi : \mathcal{M} \models \varphi\}.$$

Definition 1.3. *Two structures $\mathcal{M}$ and $\mathcal{N}$ are elementarily equivalent, and we write $\mathcal{M} \equiv \mathcal{N}$, if for every sentence φ,*

$$\mathcal{M} \models \varphi \quad \textit{if and only if} \quad \mathcal{N} \models \varphi.$$

Note that $\mathcal{M}$ and $\mathcal{N}$ are elementarily equivalent if and only if $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{M})$.

Definition 1.4. *A function $f : M \to N$ is called a* homomorphism *if for all elements $a_1, \ldots, a_n \in M$ we have*

$$\begin{array}{rcl} f(c^{\mathcal{M}}) & = & c^{\mathcal{N}} \\ f(F^{\mathcal{M}}(a_1, \ldots, a_n)) & = & F^{\mathcal{N}}(f(a_1), \ldots, f(a_n)) \\ R^{\mathcal{M}}(a_1, \ldots, a_n) & \Longrightarrow & R^{\mathcal{N}}(f(a_1), \ldots, f(a_n)) \end{array}$$

for all constants c, n-ary function symbols F and relation symbols R from L. We denote it by

$$f : \mathcal{M} \to \mathcal{N}.$$

When the third condition is "if and only if" we say that f is a strong homomorphism. *If f is a strong homomorphism and is also injective, then f is called an* embedding. *An* isomorphism *is an exhaustive embedding, and we write $f : \mathcal{M} \cong \mathcal{N}$. If there is an isomorphism between $\mathcal{M}$ and $\mathcal{N}$, the two structures are called* isomorphic *and we write $\mathcal{M} \cong \mathcal{N}$.*

Definition 1.5. *Let $\mathcal{M}$ and $\mathcal{N}$ be L-structures. We call $\mathcal{M}$ a* substructure *of $\mathcal{N}$, and we write $\mathcal{M} \subseteq \mathcal{N}$, if $M \subseteq N$ and if the inclusion map is an embedding from $\mathcal{M}$ to $\mathcal{N}$, i.e.,*

1. *For every constant of L, $c^{\mathcal{M}} = c^{\mathcal{N}}$;*
2. *For every n-ary function symbol F of L and every $a_1, \ldots, a_n \in M$, we have $F^{\mathcal{M}}(a_1, \ldots, a_n) = F^{\mathcal{N}}(a_1, \ldots, a_n)$;*
3. *For every n-ary relation symbol R of L, $R^{\mathcal{M}} = R^{\mathcal{N}} \cap M^n$.*

We say $\mathcal{N}$ is an *extension* of $\mathcal{M}$ if $\mathcal{M}$ is a substructure of $\mathcal{N}$.

We proceed to state some results related to the above definitions, which will be useful later. They can be easily proved, so we omit the proofs. Let $\mathcal{M}$ be a substructure of $\mathcal{N}$.

1. If $f : \mathcal{M} \to \mathcal{N}$ is a homomorphism, $t = t(x_1, \ldots, x_n)$ is a term of L and $a_1, \ldots, a_n \in M$, then
$$f(t^{\mathcal{M}}[a_1, \ldots, a_n]) = t^{\mathcal{N}}[f(a_1), \ldots, f(a_n)].$$
2. If $f : \mathcal{M} \to \mathcal{N}$ is an embedding, then $f(M)$ is the universe of a substructure of $\mathcal{N}$ and $f : \mathcal{M} \cong f(\mathcal{M})$, where $f(\mathcal{M})$ is the substructure that has $f(M)$ as its universe.
3. If $f : \mathcal{M} \to \mathcal{N}$ is an embedding, $\varphi_1 = \varphi_1(x_1, \ldots, x_n)$ is a quantifier-free formula of L and $a_1, \ldots, a_n \in M$, then
$$\mathcal{M} \models \varphi_1[a_1, \ldots, a_n] \text{ iff } \mathcal{N} \models \varphi_1[f(a_1), \ldots, f(a_n)];$$

   If $\varphi_2 = \varphi_2(x_1, \ldots, x_n)$ is a universal formula of L and $a_1, \ldots, a_n \in M$, then
$$\mathcal{N} \models \varphi_2[a_1, \ldots, a_n] \text{ implies } \mathcal{M} \models \varphi_2[f(a_1), \ldots, f(a_n)];$$

   If $\varphi_3 = \varphi_3(x_1, \ldots, x_n)$ is an existential formula of L and $a_1, \ldots, a_n \in M$, then
$$\mathcal{M} \models \varphi_3[a_1, \ldots, a_n] \text{ implies } \mathcal{N} \models \varphi_3[f(a_1), \ldots, f(a_n)].$$
4. If $f : \mathcal{M} \to \mathcal{N}$ is an isomorphism, $\varphi = \varphi(x_1, \ldots, x_n)$ is a formula of L and $a_1, \ldots, a_n \in M$, then
$$\mathcal{M} \models \varphi[a_1, \ldots, a_n] \text{ if and only if } \mathcal{N} \models \varphi[f(a_1), \ldots, f(a_n)].$$

The following lemma, together with a result that will be shown later, will be crucial in the last chapter.

Lemma 1.6. *Let $f : \mathcal{M} \to \mathcal{N}$ be an embedding, where $\mathcal{M}$ and $\mathcal{N}$ are L-structures. Then, there exist $\mathcal{M}' \supseteq \mathcal{M}$ and $f' : \mathcal{M}' \cong \mathcal{N}$ such that f' restricted to M is f.*

Proof. Let X be a set such that $X \cong N \setminus f(M)$ and $X \cap M = \emptyset$, and let the function $h : X \to N \setminus f(M)$ be a bijection. Then we can consider the function $f' = f \cup h$. In order to define the structure $\mathcal{M}'$, we need to give its universe M' and the map interpretation. Considering $M' = M \cup X$ and the following interpretations:

1. For every constant of L, $c^{\mathcal{M}'} = c^{\mathcal{M}}$
2. For every n-ary function symbol F of L and every $a_1, \ldots, a_n \in M'$,
$$F^{\mathcal{M}'}(a_1, \ldots, a_n) = (f')^{-1}(F^{\mathcal{N}}(f'(a_1), \ldots, f'(a_n))$$

3. For every n-ary relation symbol R of L and every $a_1, \ldots, a_n \in M'$,

$$R^{\mathcal{M}'}(a_1, \ldots, a_n) = \{(a_1, \ldots, a_n) \in (M')^n \,:\, (f'(a_1), \ldots, f'(a_n)) \in R^{\mathcal{N}}\}$$

the proof is completed. □

Definition 1.7. *Let $\mathcal{M} \subseteq \mathcal{N}$. The embedding $f : \mathcal{M} \to \mathcal{N}$ is called an* elementary embedding*, and we write $f : \mathcal{M} \lesssim \mathcal{N}$, if for every formula $\varphi = \varphi(x_1, \ldots, x_n)$ and for every $a_1, \ldots, a_n \in M$*

$$\mathcal{M} \models \varphi[a_1, \ldots, a_n] \text{ iff } \mathcal{N} \models \varphi[f(a_1), \ldots, f(a_n)].$$

In the case where f is the identity function, i.e.,

$$\mathcal{M} \models \varphi[a_1, \ldots, a_n] \text{ iff } \mathcal{N} \models \varphi[a_1, \ldots, a_n],$$

for every formula $\varphi = \varphi(x_1, \ldots, x_n)$ and for every $a_1, \ldots, a_n \in M$, we say that $\mathcal{M}$ is an elementary substructure *of $\mathcal{N}$, or that $\mathcal{N}$ is an* elementary extension *of $\mathcal{M}$. We denote this by $\mathcal{M} \preceq \mathcal{N}$.*

Chapter 2

Ultraproducts

In this chapter we define the notion of an ultrafilter and a method of constructing structures: the ultraproduct construction. This one will play a fundamental role in Chapter 3. Furthermore, some important results related to both ultraproducts and first-order logic will be shown.

2.1 Ultrafilters

First of all, before defining an ultrafilter, we shall introduce the notion of a filter over a set I.

Definition 2.1. *Let I be a nonempty set and $\mathcal{P}(I)$ the set of all subsets of I. A* filter *D over I is a set $D \subseteq \mathcal{P}(I)$ such that*

1. *$I \in D$;*
2. *if $X, Y \in D$, then $X \cap Y \in D$;*
3. *if $X \in D$ and $X \subseteq Z \subseteq I$, then $Z \in D$.*

We call the filter $D = \mathcal{P}(I)$ the *improper filter*.

The filter D over I is said to be a *proper filter* if it is not the improper filter.

Definition 2.2. *Let E be a subset of $\mathcal{P}(I)$. The intersection D of all filters over I which include E,*

$$D = \bigcap\{F : E \subseteq F \text{ and } F \text{ is a filter over } I\},$$

is called the filter generated by *E.*

E is said to have the *finite intersection property* if the intersection of any finite number of elements of E is nonempty.

Proposition 2.3. *Let E be any subset of $\mathcal{P}(I)$ and let D be the filter generated by E. Then:*

1. *D is a filter over I.*
2. *D is the set of all $X \in \mathcal{P}(I)$ such that either $X = I$ or $Y_1 \cap \ldots \cap Y_n \subseteq X$, for some $Y_1, \ldots, Y_n \in E$.*
3. *D is a proper filter if and only if E has the finite intersection property.*

Proof. 1. It's clearly true.

2. Let D' be the set of all $X \in \mathcal{P}(I)$ such that either $X = I$ or $Y_1 \cap \ldots \cap Y_n \subseteq X$, for some $Y_1, \ldots, Y_n \in E$. We want to show that $D = D'$. Let $X, X' \in D'$ and, for $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$, let $Y_i, Y_j \in E$ be such that

$$Y_1 \cap \ldots \cap Y_n \subseteq X \text{ and } Y_1' \cap \ldots \cap Y_m' \subseteq X'.$$

By the definition of D', $I \in D'$. Moreover, since

$$Y_1 \cap \ldots \cap Y_n \cap Y_1' \cap \ldots \cap Y_m' \subseteq D',$$

we have that $X \cap X' \in D'$. Now, if we suppose $X \subseteq Z \subseteq I$, then

$$Y_1 \cap \ldots \cap Y_n \subseteq Z,$$

so $Z \in D'$. Therefore, D' is a filter over I and it follows that $D \subseteq D'$, because obviously $E \subseteq D'$.

Let F be any filter over I which includes E. Then $I \in F$. Besides, for any $Y_1, \ldots, Y_n \in E$, we have that $Y_1 \cap \ldots \cap Y_n \in F$ and hence any element X of $\mathcal{P}(I)$ such that $Y_1 \cap \ldots \cap Y_n \subseteq X$ belongs to F. So $D' \subseteq F$ and, consequently, $D' \subseteq D$. Thus, $D = D'$.

3. It follows easily from (2).

□

Now we have given the definition of a filter, we can define the notion of an ultrafilter.

Definition 2.4. *D is said to be an* ultrafilter *over I if D is a filter over I such that for all $X \in \mathcal{P}(I)$,*

$$X \in D \quad \textit{if and only if} \quad (I \setminus X) \notin D.$$

The following proposition characterizes the ultrafilters.

Proposition 2.5. *The following statements are equivalent:*

1. *D is an ultrafilter over I.*
2. *D is a maximal proper filter over I.*

Proof. Firstly, we will prove (2) assuming (1). Suppose that D is an ultrafilter over I. Then $\emptyset \notin D$, as $I \in D$ and $\emptyset = I \setminus I$. Therefore, D is a proper filter. Let F be any proper filter over I such that $D \subseteq F$. If $X \in F$ and $X \notin D$, then $I \setminus X \in D$, so $I \setminus X \in F$. Thus, by the definition of filter, we have

$$\emptyset = X \cap (I \setminus X) \in F.$$

This contradicts the assumption that F is a proper filter. Hence, $D = F$ and (2) holds.

Now, assume (2). Both X and $I \setminus X$ cannot belong to D, because then $\emptyset \in D$. If we prove that $X \in D$ or $I \setminus X \in D$, then the proof will be completed. Suppose $X \notin D$ and $I \setminus X \notin D$.

Claim. $D \cup \{X\}$ has the finite intersection property or $D \cup \{I \setminus X\}$ has it.

Suppose that neither $D \cup \{X\}$ nor $D \cup \{I \setminus X\}$ has the finite intersection property. Then, for any $D_1, \ldots, D_n \in D$ we have

$$X \cap D_1 \cap \ldots \cap D_n = \emptyset \qquad \text{and} \qquad (I \setminus X) \cap D_1 \cap \ldots \cap D_n = \emptyset.$$

Hence,

$$D_1 \cap \ldots \cap D_n = (D_1 \cap \ldots \cap D_n) \cap (X \cup (I \setminus X)) = \emptyset,$$

which contradicts the assumption that D is a proper filter.

Therefore, by Proposition 2.3, either $D \cup \{X\}$ or $D \cup \{I \setminus X\}$ can be extended to a proper filter, which is a contradiction with the fact that D is maximal. □

We now prove the existence of ultrafilters.

Proposition 2.6 (Ultrafilter Theorem)**.** *If $E \subseteq \mathcal{P}(I)$ and E has the finite intersection property, then there exists an ultrafilter D over I such that $E \subseteq D$.*

Proof. By Proposition 2.3, the filter F generated by E is proper. Consider any nonempty chain C of proper filters over I. Then, $\bigcup C$ is a proper filter over I. Moreover, if each $D \in C$ includes E, then $\bigcup C$ also includes E. Using Zorn's Lemma, we have that the set of all proper filters which include E has a maximal element, say D. Then, D contains E and it is an ultrafilter over I, by Proposition 2.5. □

Corollary 2.7. *Any proper filter over I can be extended to an ultrafilter over I.*

Proof. Every proper filter has the finite intersection property. □

Now we give an example of ultrafilter which will be important in the construction of hyperreal numbers in the next chapter.

Definition 2.8. *Given a filter D, if there exists $Y \subseteq I$ such that $D = \{X \subseteq I : Y \subseteq X\}$, then D is said to be a* principal filter. *In that case, we say that the principal filter D is generated by Y. A* principal ultrafilter *is an ultrafilter which is principal. Otherwise, the ultrafilter is called* non-principal.

The following lemmas are some results related to principal and non-principal ultrafilters.

Lemma 2.9. *If D is a principal filter, then D is an ultrafilter if and only if the set that generates D contains just one element.*

Proof. Suppose D is generated by $x \in I$, that is $D = \{X \subseteq I : x \in X\}$. Then for each $X \subseteq I$ either $x \in X$ or $x \in I \smallsetminus X$, whence by definition D is an ultrafilter.

Now suppose that D is a principal ultrafilter and the set $X' = \bigcap\{X : X \in D\}$ contains the two different elements x, y. Since D is an ultrafilter, either $\{x\}$ or $I \smallsetminus \{x\}$ is in D. In the first case $y \notin X'$ and in the second case $x \notin X'$. Therefore X' contains just one element. □

Lemma 2.10. *A non-principal ultrafilter contains no finite sets.*

Proof. Let D be an ultrafilter that contains a finite set and let X be a set of least cardinal in D. Since D is proper, X is not empty. Suppose that x and y are different elements of X. Then, by hypothesis, $\{x\} \notin D$ and so $I \smallsetminus \{x\} \in D$. Therefore $X \cap (I \smallsetminus \{x\}) = X \smallsetminus \{x\} \in D$. But $X \smallsetminus \{x\}$ has fewer elements than X, since X is finite, contradicting our hypothesis. Thus X contains just one element, say x.

Claim $Y \in D$ if and only if $x \in Y$.

Assume $Y \in D$. Then, since D is an ultrafilter, $\{x\} \cap Y \in D$, so $\{x\} \cap Y \neq 0$. Therefore $x \in Y$. Now suppose $x \in Y$. Then $\{x\} \subseteq Y \subseteq I$, so $Y \in D$.

Thus D is the principal ultrafilter generated by $\{x\}$. □

Lemma 2.11. *If I is infinite, then there is a non-principal ultrafilter over I.*

Proof. Let $\mathcal{P}_\omega(I)$ be the set of all cofinite subsets of I. Therefore $\mathcal{P}_\omega(I)$ has the finite intersection property and so can be extended to an ultrafilter D over I. By construction, D does not contain finite sets and consequently it is non-principal. □

2.2 Ultraproducts

We are now ready to introduce the reduced product of structures and, in particular, the notion of ultraproduct. First, we give the definition to sets, and then to structures.

Suppose that I is a nonempty set, D is a proper filter over I and, for each $i \in I$, A_i is a nonempty set. Let

$$C = \prod_{i \in I} A_i$$

be the Cartesian product of these sets. Therefore, C is the set of all functions f with domain I such that for each $i \in I$, $f(i) \in A_i$.

We need the following in order to continue.

Definition 2.12. *Given two functions $f, g \in C$, we say that f and g are D-*equivalent *and we write $f =_D g$ if*

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

Proposition 2.13. *The relation $=_D$ is an equivalence relation over the set C.*

Proof. Reflexivity and symmetry can be proved easily. Let's prove $=_D$ is a transitive relation. Let $f, g, h \in C$ be such that $f =_D g$ and $g =_D h$. Then, by definition,

$$X_{fg} = \{i \in I : f(i) = g(i)\} \in D \text{ and } X_{gh} = \{i \in I : g(i) = h(i)\} \in D.$$

Since $X_{fg} \cap X_{gh} \subseteq X_{fh} := \{i \in I : f(i) = h(i)\}$ and $X_{fg} \cap X_{gh} \in D$, we have that $X_{fh} \in D$. Therefore, f and h are D-equivalent. □

Let I be a nonempty set, let D be a proper filter over I and let f_D be the equivalence class of f, that is:

$$f_D = \{g \in C : f =_D g\}.$$

The *reduced product of sets A_i modulo D*, denoted by $\prod_D A_i$=$\prod_D (A_i : i \in I)$, is defined as the set of all equivalence classes of $=_D$. Thus

$$\prod_D A_i = \{f_D : f \in \prod_{i \in I} A_i\}.$$

We call the reduced product of sets $\prod_D A_i$ an *ultraproduct* if D is an ultrafilter over I. In the case when all the sets A_i are the same, i.e., $A_i = A$, the reduced product is written $\prod_D A$ and it is called the *reduced power of set A modulo D*. In particular, if D is an ultrafilter, then $\prod_D A$ is called the *ultrapower of set A modulo D*.

We now give the definition of reduced product of structures. For each $i \in I$ let $\mathcal{M}_i$ be an L-structure. The *reduced product* $\prod_D \mathcal{M}_i$, which will be denoted by $\mathcal{M}$, is the L-structure described as follows:

1. The universe set of $\mathcal{M}$ is $\prod_D M_i$.
2. Let c be a constant of L. Then
$$c^{\mathcal{M}} = (c^{\mathcal{M}_i} : i \in I)_D.$$
3. Let F be an n-placed function symbol L. Then
$$F^{\mathcal{M}}(f^1_D, \ldots, f^n_D) = (F^{\mathcal{M}_i}(f^1(i), \ldots, f^n(i)) : i \in I)_D.$$
4. Let R be an n-placed relation symbol of L. Then
$$R^{\mathcal{M}}(f^1_D, \ldots, f^n_D) \text{ if and only if } \{i \in I : R^{\mathcal{M}_i}(f^1(i), \ldots, f^n(i))\} \in D.$$

Note: $F^{\mathcal{M}}(f^1_D, \ldots, f^n_D)$ and $R^{\mathcal{M}}(f^1_D, \ldots, f^n_D)$ depend only on the equivalence classes $f^1_D, \ldots, f^n_D$ and do not depend on the representatives $f^1, \ldots, f^n$ of these equivalence classes.

The reduced product $\prod_D \mathcal{M}_i$ is said to be an *ultraproduct* if D is an ultrafilter over I. When all the structures $\mathcal{M}_i$ are the same $\mathcal{M}$, the reduced product is called the *reduced power of $\mathcal{M}$ modulo D* and is written $\prod_D \mathcal{M}$. In particular, if D is an ultrafilter, then $\prod_D \mathcal{M}$ is called the *ultrapower of $\mathcal{M}$ modulo D*.

We now prove an important theorem related to the reduced product of expansions of arbitrary structures.

Theorem 2.14 (Expansion Theorem)**.** *Let L be a subset of the language L'. Let I be a nonempty set and for each $i \in I$ let $\mathcal{M}_i$ be an L-structure and $\mathcal{N}_i$ an expansion of $\mathcal{M}_i$ to L'. Then the reduced product $\prod_D \mathcal{N}_i$ is an expansion of the reduced product $\prod_D \mathcal{M}_i$ to L'.*

Proof. For each $i \in I$, the structures $\mathcal{M}_i$ and $\mathcal{N}_i$ have the same universe M_i, since $\mathcal{N}_i$ is an expansion of $\mathcal{M}_i$. Hence, the reduced products have the same universe, $\prod_D M_i$. Moreover, each symbol of L has the same interpretation in $\mathcal{M}_i$ as in $\mathcal{N}_i$. As the interpretation of a symbol of L in $\prod_D M_i$ depends only on its interpretation in the structures $\mathcal{M}_i$, and on the universe and the filter D, we have that each symbol of L has the same interpretation in $\prod_D \mathcal{M}_i$ as in $\prod_D \mathcal{N}_i$. $\square$

The next theorem is one of the most important, since it states the denotation and the satisfaction of terms and formulas in an ultraproduct, respectively. It is also known as the Łoś theorem.

Theorem 2.15 (The Fundamental Theorem of Ultraproducts)**.** *Let L be a language, for each $i \in I$ let $\mathcal{M}_i$ be a L-structure, where I is the index set. Let $\mathcal{M}$ be the ultraproduct $\prod_D \mathcal{M}_i$. Then:*

1. *For any term $t(x_1, \ldots, x_n)$ of L and elements $f_D^1, \ldots, f_D^n \in \mathcal{M}$, we have*

$$t^{\mathcal{M}}[f_D^1, \ldots, f_D^n] = (t^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I)_D.$$

2. *For any formula $\varphi(x_1, \ldots, x_n)$ of L and $f_D^1, \ldots, f_D^n \in \mathcal{M}$, we have*

$$\mathcal{M} \models \varphi[f_D^1, \ldots, f_D^n] \text{ if and only if } \{i \in I \,:\, \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D.$$

3. *For any sentence φ of L,*

$$\mathcal{M} \models \varphi \text{ if and only if } \{i \in I \,:\, \mathcal{M}_i \models \varphi\} \in D.$$

Proof. The third condition is an immediate consequence of the first one and the second one. We will prove (1) and (2) by induction on the terms and formulas, respectively.

From now on we will call φ the formula $\varphi(x_1, \ldots, x_n)$, t the term $t(x_1, \ldots, x_n)$ and, for each $i \in \{1, \ldots, m\}$, t_i will denote the term $t_i(x_1, \ldots, x_n)$.

1. From the definition of reduced product we see that (1) holds whenever t is a constant symbol or variable. Suppose that

$$t = F(t_1, \ldots, t_m),$$

where F is a function symbol of L and the terms $t_1, \ldots, t_m$ satisfy (1). By the definition of denotation of terms, we have

$$t^{\mathcal{M}}[f_D^1, \ldots, f_D^n] = F^{\mathcal{M}}(t_1^{\mathcal{M}}[f_D^1, \ldots, f_D^n] \;, \ldots, \; t_m^{\mathcal{M}}[f_D^1, \ldots, f_D^n]),$$

Since $t_1, \ldots, t_m$ satisfy (1), we have for $k = 1, \ldots, m$,

$$t_k^{\mathcal{M}}[f_D^1, \ldots, f_D^n] = g_D^k.$$

where

$$g^k = (t_k^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I).$$

Then, by the definition of reduced product, we have

$$F^{\mathcal{M}}(g^1_D, \ldots, g^m_D) = (F^{\mathcal{M}_i}[g^1(i), \ldots, g^m(i)] \,:\, i \in I)_D.$$

Using the definition of denotation of terms again,

$$t^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] = F^{\mathcal{M}_i}(g^1(i), \ldots, g^m(i)).$$

If we now combine these results, we obtain

$$t^{\mathcal{M}}[f^1_D, \ldots, f^n_D] = F^{\mathcal{M}}(g^1_D, \ldots, g^m_D) = (t^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I)_D.$$

Thus, t satisfies (1).

2. First of all, we need to prove that (2) is true for all atomic formulas. Suppose that φ is an equation

$$\varphi = t_1 \doteq t_2,$$

where the terms t_1, t_2 satisfy (1). Then the following are equivalent:

$$\mathcal{M} \models \varphi[f^1_D, \ldots, f^n_D];$$

$$t_1^{\mathcal{M}}[f^1_D, \ldots, f^n_D] = t_2^{\mathcal{M}}[f^1_D, \ldots, f^n_D];$$

$$(t_1^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I)_D = (t_2^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I)_D;$$

$$(t_1^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I) =_D (t_2^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] \,:\, i \in I);$$

$$\{i \in I \,:\, t_1^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)] = t_2^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)]\} \in D;$$

$$\{i \in I \,:\, \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D.$$

Therefore, φ satisfies (2).

Now, assume that

$$\varphi = R(t_1, \ldots, t_m),$$

where R is a symbol function of L and $t_1, \ldots, t_m$ satisfy (1). Then the following statements are equivalent:

$$\mathcal{M} \models \varphi[f^1_D, \ldots, f^n_D];$$

$$(t_1^{\mathcal{M}}[f^1_D, \ldots, f^n_D], \ldots, t_m^{\mathcal{M}}[f^1_D, \ldots, f^n_D]) \in R^{\mathcal{M}};$$

$$\{i \in I \,:\, (t_1^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)], \ldots, t_m^{\mathcal{M}_i}[f^1(i), \ldots, f^n(i)]) \in R^{\mathcal{M}_i}\} \in D;$$

$$\{i \in I : \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D.$$

Thus, φ satisfies (2).

The next step is to prove that if ψ satisfies (2), then so does $\neg\psi$. Suppose that $\varphi = \neg\psi$, where (2) holds for $\psi = \psi(x_1, \ldots, x_n)$. Then the following are equivalent:

$$\mathcal{M} \models \varphi[f^1_D, \ldots, f^n_D];$$

$$\mathcal{M} \not\models \psi[f^1_D, \ldots, f^n_D];$$

$$\{i \in I : \mathcal{M}_i \models \psi[f^1(i), \ldots, f^n(i)]\} \notin D;$$

$$\{i \in I : \mathcal{M}_i \not\models \psi[f^1(i), \ldots, f^n(i)]\} \in D;$$

$$\{i \in I : \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D.$$

The third and fourth statements are equivalent because D is an ultrafilter.

Suppose now that $\varphi = \varphi_1 \wedge \varphi_2$, where φ_1 and φ_2 satisfy (2).

Then we have the following equivalences:

$$\mathcal{M} \models \varphi[f^1_D, \ldots, f^n_D];$$

$$\mathcal{M} \models \varphi_1[f^1_D, \ldots, f^n_D] \text{ and } \mathcal{M} \models \varphi_2[f^1_D, \ldots, f^n_D];$$

$$\{i \in I : \mathcal{M}_i \models \varphi_1[f^1(i), \ldots, f^n(i)]\} \in D \text{ and } \{i \in I : \mathcal{M}_i \models \varphi_2[f^1(i), \ldots, f^n(i)]\} \in D$$

$$\{i \in I : \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D.$$

Let's see the last equivalence. Let X_1 be the set

$$\{i \in I : \mathcal{M}_i \models \varphi_1[f^1(i), \ldots, f^n(i)]\},$$

let X_2 be the set

$$\{i \in I : \mathcal{M}_i \models \varphi_2[f^1(i), \ldots, f^n(i)]\}$$

and let X be $X_1 \cap X_2$, i.e.,

$$X = \{i \in I : \mathcal{M}_i \models \varphi_1 \wedge \varphi_2[f^1(i), \ldots, f^n(i)]\}.$$

If $X_1 \in D$ and $X_2 \in D$, we clearly have that $X \in D$. Now, if $X \in D$, since $X \subseteq X_1$, $X \subseteq X_2$ and D is an ultrafilter, we have that $X_1 \in D$ and $X_2 \in D$.

Finally, suppose that $\varphi = (\exists x_0)\psi(x_0, x_1, \ldots, x_n)$ and that (2) holds for ψ. Then the following statements are equivalent:

$$\mathcal{M} \models \varphi[f^1_D, \ldots, f^n_D];$$

$$\text{there exists } f^0_D \in M \text{ such that } \mathcal{M} \models \psi[f^0_D, f^1_D, \ldots, f^n_D];$$

$$\text{there exists } f^0_D \in M \text{ such that } \{i \in I \,:\, \mathcal{M}_i \models \psi[f^0(i), f^1(i), \ldots, f^n(i)]\} \in D. \tag{2.1}$$

On the one hand, the third statement implies

$$X = \{i \in I \,:\, \mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]\} \in D, \tag{2.2}$$

since $\mathcal{M}_i \models \psi[f^0(i), f^1(i), \ldots, f^n(i)]$ implies $\mathcal{M}_i \models \varphi[f^1(i), \ldots, f^n(i)]$.

Now suppose that (2.2) holds. Then, we can consider the function

$$f^0 : I \to \bigcup M_i$$

such that

1. if $i \in X$, $f^0(i) \in M_i$ satisfies $\mathcal{M}_i \models \psi[f^0(i), f^1(i), \ldots, f^n(i)]$;
2. if $i \notin X$, $f^0(i) \in M_i$ is arbitrary.

Since $X \in D$ and $X \subseteq Y = \{i \in I \,:\, \mathcal{M}_i \models \psi[f^0(i), f^1(i), \ldots, f^n(i)]\}$, we have that $Y \in D$.

So (2.1) and (2.2) are equivalent and, hence, the formula φ satisfies the second condition. We have now completed our induction. □

One important application of the fundamental theorem is an ultraproduct version of the compactness theorem of first-order logic:

Corollary 2.16 (An ultraproduct version of the compactness theorem)**.** *Let Σ be a set of sentences of L, let I be the set of all finite subsets of Σ, and for each $i \in I$ let $\mathcal{M}_i$ be a model of i. Then there exists an ultrafilter D over I such that the ultraproduct $\prod_D \mathcal{M}_i$ is a model of Σ.*

Proof. For each $\sigma \in \Sigma$, let $\hat{\sigma}$ be the set of all $i \in I$ such that $\sigma \in i$. The set

$$E = \{\hat{\sigma} \,:\, \sigma \in \Sigma\}$$

has the finite intersection property because

$$\{\sigma_1, \ldots, \sigma_n\} \in \hat{\sigma}_1 \cap \ldots \cap \hat{\sigma}_n.$$

As E has this property and $E \subseteq \mathcal{P}(I)$, E can be extended to an ultrafilter D over I by the ultrafilter theorem. If $i \in \hat{\sigma}$, then $\sigma \in i$. Thus, since $\mathcal{M}_i \models i$, we have that $\mathcal{M}_i \models \sigma$. Hence for each $\sigma \in \Sigma$

$$\hat{\sigma} \subseteq \{i \in I \,:\, \mathcal{M}_i \models \sigma\} \text{ and } \hat{\sigma} \in D.$$

Therefore, since D is an ultrafilter

$$\{i \in I : \mathcal{M}_i \models \sigma\} \in D.$$

By the fundamental theorem 2.15, $\prod_D \mathcal{M}_i \models \sigma$ for all $\sigma \in \Sigma$. Thus $\prod_D \mathcal{M}_i$ is a model of Σ. □

Corollary 2.17. *The Corollary 2.16 also holds when Σ is a set of formulas of L.*

Proof. All free variables of the set Σ can be replaced by constants. □

Corollary 2.18. *Let Σ be a set of sentences of L. If $\Sigma \models \varphi$, then exists $\Sigma_0 \subseteq \Sigma$ finite such that $\Sigma_0 \models \varphi$.*

Proof. For each Σ_0 finite such that $\Sigma_0 \subseteq \Sigma$, suppose that $\Sigma_0 \not\models \varphi$. Then, for each $\Sigma_0 \subseteq \Sigma$ finite, $\Sigma_0 \cup \{\neg\varphi\}$ is finitely satisfiable. Hence $\Sigma \cup \{\neg\varphi\}$ is finitely satisfiable and by the compactness theorem we have that $\Sigma \cup \{\neg\varphi\}$ is satisfiable. Thus $\Sigma \not\models \varphi$. □

Another application of the Łoś theorem is the following one, which shows that each structure $\mathcal{M}$ is elementarily embeddable in every ultrapower of $\mathcal{M}$ in a natural way. Before that, we need to define the canonical embedding of $\mathcal{M}$ into its ultrapower. This embedding is very important and will be crucial for the construction of hyperreal numbers later.

Definition 2.19. *Let I be a nonempty set, D an ultrafilter over I and $\mathcal{M}$ a structure. The* canonical embedding *is the function*

$$d : \mathcal{M} \to \mathcal{M}^D$$

such that $d(a) = (a : i \in I)_D$, where $\mathcal{M}^D$ denotes the ultrapower $\prod_D \mathcal{M}$.

Corollary 2.20. *Let $\mathcal{M}$ be an L-structure and D an ultrafilter. Then the canonical embedding of $\mathcal{M}$ into the ultrapower $\mathcal{M}^D$ is an elementary embedding.*

Proof. Let $\varphi = \varphi(x_1, \ldots, x_n)$ be a formula of L and $a_1, \ldots, a_n \in M$. Then

$$\begin{aligned} \mathcal{M}^D \models \varphi[d(a_1), \ldots, d(a_n)] &\iff \{i \in I : \mathcal{M} \models \varphi[a_1, \ldots, a_n]\} \in D \\ &\iff \mathcal{M} \models \varphi[a_1, \ldots, a_n] \end{aligned}$$

Note that the set $X = \{i \in I : \mathcal{M} \models \varphi[a_1, \ldots, a_n]\}$ can only be the empty set or I. Since D is an ultrafilter, $X \neq \emptyset$. Therefore, the second equivalence is true. □

Consider the structure $\mathcal{M}$ and the ultrapower $\mathcal{M}^D$. By the above corollary, we know there is an elementary embedding of $\mathcal{M}$ into $\mathcal{M}^D$. Therefore, by Lemma 1.6, we can identify the elements of $\mathcal{M}$ with the corresponding ones under the canonical embedding in its ultrapower. Hence, we may consider that $\mathcal{M}$ is an elementary substructure of $\mathcal{M}^D$.

Chapter 3

Non-standard analysis

3.1 Real closed fields

In this chapter we will apply all we have seen in the previous ones to non-standard analysis. First of all, we introduce the following L-structure

$$\mathbb{R} = (\mathbb{R}, +, -, \cdot, <, 0, 1),$$

where the universe $\mathbb{R}$ is the set of real numbers, $+, \cdot$ are binary function symbols, $-$ is a monary function symbol, $<$ is a binary relation symbol and $0, 1$ are constants with their usual meaning.
Note that given an arbitrary L-structure $\mathcal{M}$ we usually write each symbol with the superscript $\mathcal{M}$ indicating the interpretation of the corresponding symbol in $\mathcal{M}$. However, when talking about the structure $\mathbb{R}$ all such superscripts will be dropped in order to simplify notation.

As it is known, A. Tarski proved that the theory of the structure $\mathbb{R}$, $\mathrm{Th}(\mathbb{R})$, can be axiomatized in first-order logic by the axioms for *real closed fields* (see, for example, [6] or [11]):

1. Field axioms

   $\forall x \forall y \forall z((x + y) + z \doteq x + (y + z))$
   $\forall x \forall y(x + y \doteq y + x)$
   $\forall x(x + 0 \doteq x)$
   $\forall x(x + (-x) \doteq 0)$
   $\forall x \forall y \forall z((x \cdot y) \cdot z \doteq x \cdot (y \cdot z))$
   $\forall x \forall y(x \cdot y \doteq y \cdot x)$
   $\forall x \forall y \forall z((x + y) \cdot z \doteq x \cdot z + y \cdot z)$
   $\forall x(x \cdot 1 \doteq x)$
   $\forall x(\neg x \doteq 0 \to \exists y(x \cdot y \doteq 1))$

2. Order axioms

   $\neg x < x$
   $\forall x \forall y (x < y \ \vee \ x \doteq y \ \vee \ y < x)$
   $\forall x \forall y \forall z (x < y \ \wedge \ y < z \to x < z)$
   $\forall x \forall y \forall z (x < y \to x + z < y + z)$
   $\forall x \forall y (0 < x \ \wedge \ 0 < y \to 0 < x \cdot y)$

3. Positive elements are squares

   $\forall x (0 < x \to \exists y (x \doteq y^2))$

4. Polynomials of degree $2n + 1$ have zeros

   $\forall x_0 x_1 \ldots x_{2n} \exists z (z^{2n+1} + x_{2n} z^{2n} + \ldots + x_1 z + x_0 \doteq 0)$.

The field and order axioms determine an ordered field. Since the characteristic of any ordered field F is 0, there is an embedding of the set of natural numbers $\mathbb{N}$ and the set of integers $\mathbb{Z}$ into F, whence we may assume that $\mathbb{N}$ and $\mathbb{Z}$ are subsets of F. In particular, $\mathrm{char}(\mathbb{R}) = 0$ and $\mathbb{N}$ and $\mathbb{Z}$ are subrings of $\mathbb{R}$.

Definition 3.1. *Let F be an ordered field. F is said to be complete if every nonempty subset of F that is bounded above has a supremum in F.*

Completeness cannot be expressed by formulas of first-order logic. Indeed, as it is known, the above axioms do not characterize $\mathbb{R}$ up to isomorphism. Given an ordered field F, it is well-known that

$$F \cong \mathbb{R} \quad \text{if and only if} \quad F \text{ is complete,}$$

(for reference see [10]).

Definition 3.2. *We say that an ordered field F has the* Archimedean property *if for every positive elements x, y of F there is a natural number n such that $x < ny$. In other words, the set $\mathbb{N}$ of natural numbers has no upper bound in F.*

Lemma 3.3. *In any ordered field F, the set $\mathbb{N}$ does not have a least upper bound.*

Proof. Suppose $x \in F$ is an upper bound of $\mathbb{N}$. For any $y \in \mathbb{N}$ we have that $y+1 \in \mathbb{N}$, so $y + 1 \leq x$ whence $y \leq x - 1$. Thus $x - 1$ is also an upper bound of $\mathbb{N}$. Therefore x cannot be a least upper bound of $\mathbb{N}$, since $x - 1 < x$. □

Corollary 3.4. *Every complete ordered field has the Archimedean Property.*

Proof. It follows immediately by the preceding lemma. □

3.2 Hyperreal numbers

In the preceding section we have seen the axiomatization of the theory of the structure $\mathbb{R}$. Now we are ready to introduce the hyperreal numbers from $\mathbb{R}$. Let D be a non-principal ultrafilter over the set of natural numbers $\mathbb{N}$ and let $\mathbb{R}^*$ be the ultrapower $\mathbb{R}^D$. Therefore,

$$\mathbb{R}^* = (\mathbb{R}^*, +^*, -^*, \cdot^*, <^*, 0^*, 1^*)$$

and we call it the field of *hyperreal numbers.* By the last paragraph of the previous chapter, we may assume that $\mathbb{R}^*$ is an elementary extension of $\mathbb{R}$, that is

$$\mathbb{R} \preceq \mathbb{R}^*.$$

Therefore, each element of $\mathbb{R}$ can be identified with its image under the canonical embedding in $\mathbb{R}^D$.

Notation. In many cases, we will work with expansions of $\mathbb{R}$ and their elementary extensions, which by the Expansion Theorem 2.14 are expansions of $\mathbb{R}^*$. Usually we will denote by $\mathbb{R}$ the expansions of $\mathbb{R}$ and by $\mathbb{R}^*$ the expansions of $\mathbb{R}^*$. For example, given a subset $X \subseteq \mathbb{R}$, we may consider the expansion

$$\mathbb{R} = (\mathbb{R}, +, -, \cdot, <, 0, 1, X)$$

and its elementary extension

$$\mathbb{R}^* = (\mathbb{R}, +, -, \cdot, <, 0, 1, X)^D = (\mathbb{R}^*, +^*, -^*, \cdot^*, <^*, 0^*, 1^*, X^*),$$

which by the Expansion Theorem 2.14 is an expansion of $\mathbb{R}^*$. This also holds for relations and functions of $\mathbb{R}$.

When no ambiguity can arise, we will omit the stars $(^*)$ on the hyperreal symbols. Moreover, we will usually write known operations such as the absolute value $|\cdot|$ or dom(), instead of their expressions as formulas of first-order logic. For instance, the absolute value can be defined by the following formula

$$\varphi(x, y) = (x \geq 0 \wedge y = x) \vee (x < 0 \wedge y = -x).$$

We will also write $x \in X$ instead of $R_X x$, where R_X is a relation symbol introduced to denote the set X.

The following corollary will be proved later, since the proof involves definitions that we do not have yet.

Corollary 3.5. *The ordered field $\mathbb{R}^*$ of hyperreal numbers does not have the Archimedean Property.*

Now we define the elements of $\mathbb{R}^*$, and we prove that they exist.

Definition 3.6. *Given an element $x \in \mathbb{R}^*$. We say that x is* finite *if $|x| < r$ for some real r. The element x is called* infinitesimal *if $|x| < r$ for all positive real r. When $|x| > r$ for all real r, then x is* infinite.

Proposition 3.7. *There exists $x \in \mathbb{R}^*$ such that x is infinite.*

Proof. Let $r \in \mathbb{R}$ and let $f : \mathbb{N} \to \mathbb{R}$ be a function such that $f(n) = n$.

Claim. $f_D > r$.

Take $m > r$ and let $Y = \{n \in \mathbb{N} : n \leq m\}$. Since D is a non-principal ultrafilter, $Y \notin D$ and consequently the set $\mathbb{N} \setminus Y = \{n \in \mathbb{N} : n > m\} \in D$. Therefore $\{n \in \mathbb{N} : \mathbb{R} \models f(n) > r\} \in D$ and by the Fundamental Theorem of Ultraproducts we have that $\mathbb{R}^* \models f_D > r$. $\square$

Now that the existence of infinite numbers in $\mathbb{R}^*$ has been proved, we can prove the corollary stated before.

Proof. (Corollary 3.5) Let x be an infinite element of $\mathbb{R}^*$. By definition, x is an upper bound of $\mathbb{R}$ and since $\mathbb{N} \subseteq \mathbb{R}$, x is an upper bound of $\mathbb{N}$. $\square$

Proposition 3.8. *There exists $x \in \mathbb{R}^*$ such that x is infinitesimal.*

Proof. By the above proposition there exist hyperreal numbers which are infinite. Let $y \in \mathbb{R}^*$ be infinite. Then, in particular, $y > r$ for each positive real r, whence $1/r > 1/y$. Therefore there exists a hyperreal number $x = 1/y$ such that $s > x$ for each positive real s. $\square$

We say that two elements $x, y \in \mathbb{R}^*$ are *infinitely close*, and we write $x \approx y$, if $x - y$ is infinitesimal. Observe that x is infinitesimal if and only if $x \approx 0$.

Definition 3.9. *Let x be a hyperreal number. The* monad *of x is defined as the set*

$$\text{monad}(x) = \{y \in \mathbb{R}^* : x \approx y\}.$$

The galaxy *of x is the set*

$$\text{galaxy}(x) = \{y \in \mathbb{R}^* : x - y \textit{ is finite}\}.$$

Therefore, monad(0) is the set of infinitesimal numbers and galaxy(0) is the set of finite hyperreal numbers.

Theorem 3.10. *The set* galaxy(0) *is a subring of* $\mathbb{R}^*$.

Proof. Let x and y be finite, i.e.,

$$|x| < r \text{ and } |y| < s,$$

where r and s are real. Then, since

$$|x + y| < r + s, \quad |x - y| < r + s, \quad |xy| < rs$$

we have that $x + y, x - y$ and xy are finite. □

Corollary 3.11. *Given x, y hyperreal numbers, "$x - y$ is finite" defines an equivalence relation and* galaxy(x) *is the equivalence class of x.*

Proof. Reflexivity and symmetry are obvious. Transitivity follows from Theorem 3.10. □

Corollary 3.12. *The relation $\approx$ is an equivalence relation and* monad(x) *is the equivalence class of x, where x is a hyperreal number.*

Proof. Reflexivity and symmetry are easy to prove. Let $x, y, z \in \mathbb{R}^*$ be such that $x \approx y$ and $y \approx z$. Then by definition

$$|x - y| < r \quad \text{and} \quad |y - z| < s,$$

for all positive real numbers r, s. Therefore,

$$|x - z| = |x - y + y - z| < r + s.$$

Thus we have that $x \approx z$. □

By the above corollaries, any two galaxies are either equal or disjoint and also any two monads are equal or disjoint.

Theorem 3.13. *The set* monad(0) *of infinitesimal elements is a subring of* $\mathbb{R}^*$ *and an ideal in* galaxy(0)*, that is*

1. *Sums, differences and products of infinitesimals are infinitesimal.*

2. *The product of an infinitesimal and a finite element is infinitesimal.*

Proof. Let $\varepsilon, \delta \approx 0$. Then $|\varepsilon| < r/2, |\delta| < r/2$, for any positive real r. Since $|\varepsilon + \delta| < r, |\varepsilon - \delta| < r$, $\varepsilon + \delta$ and $\varepsilon - \delta$ are infinitesimal.
Let a be finite, say $|a| < t, 1 \leq t \in \mathbb{R}$. Hence, for any positive real number r we have $|\varepsilon| < r/t, |\varepsilon a| < r$. Thus εa is infinitesimal. □

Theorem 3.14. *The following statements are true:*

1. *x is infinite if and only if x^{-1} is infinitesimal.*

2. $\operatorname{monad}(0)$ *is a maximal ideal in* $\operatorname{galaxy}(0)$*.*

Proof. 1. We have the following

$$\begin{aligned} |x| \geq r, \text{ for each positive real } r &\iff |x^{-1}| \geq r^{-1}, \text{ for each positive real } r \\ &\iff x^{-1} \text{ is infinitesimal.} \end{aligned}$$

2. Let I be an ideal that contains $\operatorname{monad}(0)$ and let $x \in I \setminus \operatorname{monad}(0)$. Since $x = (x^{-1})^{-1}$ is not infinitesimal, we have that x^{-1} is finite by (1). Therefore $x^{-1} \in \operatorname{galaxy}(0)$, so $1 = x \cdot (x^{-1}) \in I$. Then for any $y \in \operatorname{galaxy}(0)$, $1 \cdot y = y \in I$, so $I = \operatorname{galaxy}(0)$.

□

We will now define the standard part of a hyperreal number, but before that, we prove the Standard Part Principle.

Theorem 3.15 (Standard Part Principle)**.** *Every finite $x \in \mathbb{R}^*$ is infinitely close to a unique real number r. That is, every finite monad contains a unique real number.*

Proof. Let $x \in \mathbb{R}^*$ be finite. Firstly, we will prove uniqueness and then existence. Suppose $x \approx r$ and $x \approx s$, where $r, s \in \mathbb{R}$. Then, $r \approx s$, because $\approx$ is an equivalence relation, and furthermore $r - s \approx 0$. Since $r - s$ is real, $r - s = 0$, so $r = s$.

Let $X = \{s \in \mathbb{R} : s < x\}$. X is nonempty and since there exists a real number $r > 0$ such that $|x| < r$, the set X has an upper bound. Whence $-r < x < r$, so $-r \in X$ and r is an upper bound of X. In fact, X has a least upper bound $t \in \mathbb{R}$, as $\mathbb{R}$ is a complete ordered field. For every real number $\varepsilon > 0$ we have

$$x \leq t + \varepsilon, \quad x - t \leq \varepsilon \quad \text{and} \quad t - \varepsilon \leq x, \quad -(x - t) \leq \varepsilon.$$

Consequently, $x - t \approx 0$, so $x \approx t$. □

Definition 3.16. *Let $x \in \mathbb{R}^*$ be finite. We say that r is the* standard part *of x, and we denote it by* $\mathrm{st}(x) = r$*, if it is the unique real number such that $r \approx x$.*

If x is infinite, $\mathrm{st}(x)$ is undefined.

Here are some properties of the standard part.

Corollary 3.17. *Let x and y be finite, then*

1. *$x \approx y$ if and only if* $\mathrm{st}(x) = \mathrm{st}(y)$.
2. $x \approx \mathrm{st}(x)$.
3. *If $r \in \mathbb{R}$, then* $\mathrm{st}(r) = r$.
4. *If $x \leq y$, then* $\mathrm{st}(x) \leq \mathrm{st}(y)$.

Proof. 1, 2 and 3 follow from the Theorem 3.15.

4 Let $x = \mathrm{st}(x) + \varepsilon$, $\quad y = \mathrm{st}(y) + \delta$, for some infinitesimal ε and δ. Suppose $x \leq y$. Then

$$\mathrm{st}(x) + \varepsilon \leq \mathrm{st}(y) + \delta,$$

$$\mathrm{st}(x) \leq \mathrm{st}(y) + (\delta - \varepsilon).$$

Since ε, δ are infinitesimal, for any real $r > 0$,

$$\delta - \varepsilon < r,$$

$$\mathrm{st}(x) < \mathrm{st}(y) + r.$$

Therefore $\mathrm{st}(x) \leq \mathrm{st}(y)$.

□

Theorem 3.18. *The standard part function is a homomorphism of* $\mathrm{galaxy}(0)$ *onto the field of real numbers.*

Proof. Let $x = r + \varepsilon, y = s + \delta$ where $r = \mathrm{st}(x)$ and $s = \mathrm{st}(y)$. Then ε and δ are infinitesimal. Therefore,

$$\begin{aligned} \mathrm{st}(x+y) &= \mathrm{st}((r+\varepsilon) + (s+\delta)) \\ &= \mathrm{st}((r+s) + (\varepsilon+\delta)) \\ &= r + s. \end{aligned}$$

In a similar way we can prove that $\mathrm{st}(x-y) = \mathrm{st}(x) - \mathrm{st}(y)$. Moreover, we have that

$$\begin{aligned} \mathrm{st}(xy) &= \mathrm{st}((r+\varepsilon)\cdot(s+\delta)) \\ &= \mathrm{st}(rs + r\delta + s\varepsilon + \varepsilon\delta)) \\ &= rs, \end{aligned}$$

because $r\delta + s\varepsilon + \varepsilon\delta$ is infinitesimal, by Theorem 3.13. □

Corollary 3.19. *Given x,y finite. The following statements hold.*

1. *If* $\mathrm{st}(y) \neq 0$, *then* $\mathrm{st}(x/y) = \mathrm{st}(x)/\,\mathrm{st}(y)$.
2. *If* $x \geq 0$ *and* $y = \sqrt[n]{x}$ *then* $\mathrm{st}(y) = \sqrt[n]{\mathrm{st}(x)}$.

Proof. 1. Suppose $\mathrm{st}(y) \neq 0$. Then

$$\mathrm{st}(x) = \mathrm{st}((x/y)\cdot y) = \mathrm{st}(x/y)\cdot \mathrm{st}(y).$$

2. If $x \geq 0$ and $y = \sqrt[n]{x}$, then $y^n = x$ and $y \geq 0$. Taking standard parts,

$$\mathrm{st}(x) = \mathrm{st}(y^n) = \mathrm{st}(y)^n,$$

where $\mathrm{st}(x) \geq 0$, $\mathrm{st}(y) \geq 0$, so $\mathrm{st}(y) = \sqrt[n]{\mathrm{st}(x)}$.

□

3.3 Basic topological notions

In this section we will give some basic topological concepts and we will see that they also admit definitions in terms of hyperreal numbers.

Definition 3.20. *Given a point* $x \in \mathbb{R}$, *a* neighborhood *of* x *is a set* X *such that*

$$(x-r, x+r) \subseteq X,$$

where $(x-r, x+r)$ *is an open interval and* r *is a positive real number.*

A set X *is called* open set *if for every* $x \in X$, X *is a neighborhood of* x. *The* interior *of a set* X *is defined as the set*

$$\mathrm{int}(X) = \{x \in X \,:\, \textit{there is a neighborhood}\; Y \textit{of}\; x \;\textit{such that}\; Y \subseteq X\}.$$

Note that X *is an open set if and only if* $X = \mathrm{int}(X)$.

We say that X *is a* closed set *if it is the complement of an open set. The* closure $\bar{X}$ *of a set* X *is defined as follows*

$$\bar{X} = \{x \in X \,:\, Y \cap X \neq \emptyset \;\textit{for every neighborhood}\; Y \textit{of}\; x\}.$$

Notice that X *is a closed set if and only if* $X = \bar{X}$.

Theorem 3.21. *Let $a \in \mathbb{R}$ and $X \subseteq \mathbb{R}$. X is a neighborhood of a if and only if X^* includes the monad of a.*

Proof. Suppose X is a neighborhood of a, that is for some real number $r > 0$, $(a - r, a + r) \subseteq X$. Then,

$$\mathbb{R} \models \forall x(|a - x| < r \rightarrow x \in X).$$

Since $\mathbb{R}^*$ is an elementary extension of $\mathbb{R}$, we have

$$\mathbb{R}^* \models \forall x(|a - x| < r \rightarrow x \in X^*),$$

whence $\mathrm{monad}(a) \subseteq X^*$.

Now suppose X is not a neighborhood of a, that is $(a - r, a + r) \nsubseteq X$ for each positive real r. Then,

$$\mathbb{R} \models \forall y(y > 0 \rightarrow \neg\forall x(|a - x| < y \rightarrow x \in X)),$$

and, consequently,

$$\mathbb{R}^* \models \forall y(y > 0 \rightarrow \neg\forall x(|a - x| < y \rightarrow x \in X^*)).$$

Now let ε be infinitesimal, thus $\varepsilon > 0$. Therefore,

$$\mathbb{R}^* \models \exists x(|a - x| < \varepsilon \wedge x \notin X^*),$$

whence for some x, $x \in \mathrm{monad}(a)$ and $x \notin X^*$. So the theorem is proved. □

Corollary 3.22. *The closure of a set $X \subseteq \mathbb{R}$ is equal to the set*

$$\{\mathrm{st}(x) \,:\, x \textit{ is finite and } x \in X^*\}.$$

Proof. The following statements are equivalent:

$$\begin{aligned} a \text{ belongs to the closure of X} &\iff a \text{ does not belong to the interior of } \mathbb{R} \smallsetminus X \\ &\iff \mathrm{monad}(a) \nsubseteq \mathbb{R}^* \smallsetminus X^* \\ &\iff \text{there is an } x \in X^* \text{ such that } x \approx a \\ &\iff a = \mathrm{st}(x) \text{ for some finite } x \in X^*. \end{aligned}$$

□

By the above corollary we have that X is closed if and only if $\mathrm{st}(x) \in X$ for each finite element x of X^*.

Corollary 3.23. *Given a positive real number r. A real function f is defined at every point of $(a-r, a+r)$ if and only if f^* is defined at every point of the monad of a.*

Proof. It follows from Theorem 3.21, by taking $X = \mathrm{dom}(f)$ and $X^* = \mathrm{dom}(f^*)$. $\square$

Definition 3.24. *A set $X \subseteq \mathbb{R}$ is said to be* bounded *if it is included in some closed real interval $[a, b]$.*

Theorem 3.25. *Let $X \subseteq \mathbb{R}$. Then X is bounded if and only if every element of X^* is finite.*

Proof. Suppose X is bounded, i.e., $X \subset [a, b]$. Then

$$\mathbb{R} \models \forall x(x \in X \to a \leq x \wedge x \leq b)$$

and since $\mathbb{R} \preceq \mathbb{R}^*$,

$$\mathbb{R}^* \models \forall x(x \in X^* \to a \leq x \wedge x \leq b).$$

Therefore, every element of X^* is finite.

Now suppose that X is not bounded. Then either X has no upper bound or no lower bound. Consider X has no upper bound. Hence

$$\mathbb{R} \models \forall y \exists x(x \in X \wedge y < x).$$

Again, by Expansion, we have that

$$\mathbb{R}^* \models \forall y \exists x(x \in X^* \wedge y < x).$$

Hence there exists an infinite element of X^*. $\square$

Definition 3.26. *A set of real numbers is called* compact *if it is closed and bounded.*

Corollary 3.27. *Let X be a set of real numbers. X is compact if and only if for every $x \in X^*$, x is finite and $\mathrm{st}(x) \in X$.*

Proof. By Corollary 3.22 and Theorem 3.25. $\square$

3.4 Differentiation

We will define the notion of derivative by using infinitesimals and we will give some known properties.

Definition 3.28. *A real number s is called the* slope *of a real function f at a real point a if*

$$s = \mathrm{st}\left(\frac{f^*(a+\Delta x) - f(a)}{\Delta x}\right)$$

for every nonzero infinitesimal Δx.

The derivative *of a real function f is the real function f' such that $f'(x)$ is the slope of f at x if it exists, otherwise, $f'(x)$ is undefined.*

The function f is said to be differentiable *at a if the slope of f at a exists.*

In the next section of limits and continuity we will see that the above definition of derivative coincides with the standard one.

The following results are left without proof, since they are easy consequences of the definition.

Corollary 3.29. *f is differentiable at $a \in \mathbb{R}$ if and only if*

1. *$f^*(x)$ is defined for all $x \approx a$ and*
2. *the quotient $(f^*(a+\Delta x) - f(a))/\Delta x$ is finite and has the same standard part for all nonzero $\Delta x \approx 0$.*

Corollary 3.30. *If f is differentiable at $a \in \mathbb{R}$, then $f(x)$ is defined for all real x in some neighborhood of a.*

Given an infinitesimal Δx, the dependence equation of the increment of y, denoted by Δy, on x and Δx is the following

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

Therefore, if $f'(x)$ exists, its value is

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

Now we give some known properties of derivatives and we prove them by using infinitesimals. First, we need to show the following result.

Proposition 3.31. *Let x be a real number and Δx a nonzero infinitesimal. If $f'(x)$ exists, then there is an infinitesimal ε such that*

$$\Delta y = f'(x)\Delta x + \varepsilon\Delta x.$$

Proof. Take

$$\varepsilon = \frac{\Delta y}{\Delta x} - f'(x).$$

Therefore $\varepsilon \approx 0$. Now, multiplying by Δx, we have $\varepsilon\Delta x = \Delta y - f'(x)\Delta x$, and so the proof is completed. □

Theorem 3.32. *Consider the real functions f and g. Then, for any real value of x such that $f'(x)$ and $g'(x)$ exist, we have*

1. *(Sum Rule)*
$$(f(x) + g(x))' = f'(x) + g'(x).$$

2. *(Constant Rule) For any real number c,*
$$(cf(x))' = c \cdot f'(x).$$

3. *(Product Rule)*
$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

4. *(Quotient Rule) If $g(x) \neq 0$,*
$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}.$$

Proof. Let Δx be a nonzero infinitesimal.

1. Sum Rule: Let $h = f + g$. Then

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

   Therefore, taking standard parts, the proof is completed.

2. Constant Rule: Let $f = cg$. Similarly as above, we have

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

Again, taking standard parts, the rule is proved.

3. Product Rule: Let $h = f \cdot g$. Thus

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

Taking standard parts, the proof is completed.

4. Quotient Rule: Let $h = f/g$, where $g \neq 0$. Then

$$\begin{aligned}\frac{h(x+\Delta x)-h(x)}{\Delta x} &= \frac{f(x+\Delta x)}{g(x+\Delta x)}\cdot\frac{1}{\Delta x}-\frac{f(x)}{g(x)}\cdot\frac{1}{\Delta x}\\ &= \frac{f(x+\Delta x)\cdot g(x)-f(x)\cdot g(x+\Delta x)}{g(x+\Delta x)\cdot g(x)\cdot\Delta x}\\ &= \frac{(f(x+\Delta x)/\Delta x)\cdot g(x)-f(x)\cdot(g(x+\Delta x)/\Delta x)}{g(x+\Delta x)\cdot g(x)}.\end{aligned}$$

Taking standard parts,

$$\begin{aligned}\mathrm{st}\left(\frac{h(x+\Delta x)-h(x)}{\Delta x}\right) &= \mathrm{st}\left(\frac{(f(x+\Delta x)/\Delta x)\cdot g(x)-f(x)\cdot(g(x+\Delta x)/\Delta x)}{g(x+\Delta x)\cdot g(x)}\right)\\ &= \frac{f'(x)\cdot g(x)-f(x)\cdot g'(x)}{\mathrm{st}(g(x+\Delta x))\cdot g(x)} = \frac{f'(x)\cdot g(x)-f(x)\cdot g'(x)}{g^2(x)}.\end{aligned}$$

Let us show the last equivalence. Observe that

$$g(x+\Delta x) = g(x) + (g(x+\Delta x)-g(x)).$$

Thus by Proposition 3.31,

$$g(x + \Delta x) = g(x) + f'(x)\Delta x + \varepsilon \Delta x,$$

for some infinitesimal ε. Therefore, $\mathrm{st}(g(x + \Delta x)) = \mathrm{st}(g(x)) + 0 = g(x)$.

□

Theorem 3.33 (Chain Rule). *Let f and g be real functions en let h be the composition*

$$h(t) = (g \circ f)(t).$$

For any real value of t where $f'(t)$ and $g'(f(t))$ exist, $h'(t)$ also exists and it is

$$h'(t) = g'(f(t))f'(t).$$

Proof. Let $x = f(t), y = h(t) = g(x)$ and let Δt be a nonzero infinitesimal. Then $\Delta y = h(t + \Delta t) - h(t)$ and $\Delta x = f(t + \Delta t) - f(t)$. By Proposition 3.31,

$$\Delta y = g'(x)\Delta x + \varepsilon \Delta x,$$

for some infinitesimal ε. Thus

$$\frac{h(t + \Delta t) - h(t)}{\Delta t} = \frac{g'(x)\Delta x + \varepsilon \Delta x}{\Delta t},$$

and taking standard parts we obtain what we wanted. □

Theorem 3.34 (Power Rule). *If x is a positive real number, r is any rational number and $f(x) = x^r$, then*

$$f'(x) = rx^{r-1}.$$

Proof. There are four different cases.

Case 1: If r is a positive integer, the Power Rule can be easily proved by induction using the Product Rule.

Case 2: $r = 1/n$ for some positive integer n. Consider the function $g(x) = x^n$. It follows from Case 1 that $g'(x) = nx^{n-1}$. By the Chain Rule 3.33,

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

Therefore, since $(f \circ g)(x) = x$ and $(f \circ g)'(x) = 1$, we have that

$$f'(g(x)) = \frac{1}{g'(x)} = \frac{1}{nx^{n-1}}.$$

Now we can consider $y = (y^{1/n})^n$. Then

$$f'(y) = \frac{1}{n(y^{1/n})^{n-1}} = \frac{1}{n} y^{\frac{1}{n}-1},$$

and the proof of this case is completed.

Case 3: r is a positive rational. This is a result of Cases 1 and 2 using the fact that

$$x^{m/n} = x^{m(1/n)}.$$

Case 4: r is a negative rational. This is a consequence of Case 3 using the Quotient Rule. □

Consider $r = m/n$. The Power Rule holds for negative values of x when n is odd. If n is even, then x^r is undefined in the real number system when x is negative.

3.5 Limits and continuous functions

As in the preceding sections, in this one we give some notions of limits and continuity by using infinitesimals.

Definition 3.35. *The* limit *of $f(x)$ as x approaches a real number a, is the real number L such that whenever $x \approx a$ but $x \neq a$, we have $f(x) \approx L$. We denote it by*

$$L = \lim_{x \to a} f(x).$$

If there is no such L, we say that the limit does not exist.

We can also define infinite limits.

Definition 3.36. *Let a and L be real numbers.*
$\lim_{x\to\infty} f(x) = L$ *if $f(y) \approx L$ for every positive infinite y.*
$\lim_{x\to a} f(x) = \infty$ *if $f(x)$ is positive infinite whenever $x \approx a$ but $x \neq a$.*
Again, if there is no such L, we say that the limit does not exist.

The next theorem shows that the above infinitesimal definition of limit when x approaches a is equivalent to the standard ε, δ definition.

Theorem 3.37. *Let f be a real function and let c and L be real numbers. The following are equivalent:*

1. $\lim_{x\to a} f(x) = L$. *That is, whenever $x \approx a$, then $f(x) \approx L$.*
2. *There exists a hyperreal $\delta > 0$ such that whenever $|x - a| < \delta$, then $f(x) \approx L$.*

3. *For every real* $\varepsilon > 0$ *there exists a real* $\delta > 0$ *such that whenever* x *is real and* $|x - a| < \delta$*, then* $|f(x) - L| < \varepsilon$.

Proof. (1) implies (2), with δ being any positive infinitesimal. We want to see that (2) implies (3). Suppose (3) fails for some real $\varepsilon > 0$. Then we have the following

$$\mathbb{R} \models \forall x(x > 0 \to \exists y(|y - a| < x \wedge |f(y) - L| \geq \varepsilon)).$$

Therefore,

$$\mathbb{R}^* \models \forall x(x > 0 \to \exists y(|y - a| < x \wedge |f(y) - L| \geq \varepsilon)).$$

Let $\delta > 0$ be hyperreal. Then there is a hyperreal x such that $|x - a| < \delta$ and $f(x) \not\approx L$. Thus (2) doesn't hold.

Now assume (3). Let $\varepsilon > 0$ be any real number and let $\delta > 0$ be the corresponding real number in the ε, δ condition. Then

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

and so

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

Hence taking $x_1 \approx a$ we have $|x_1 - a| < \delta$, and consequently $|f(x_1) - L| < \varepsilon$. Since this holds for each real $\varepsilon > 0$, $f(x_1) \approx L$. $\square$

Similarly, it can be proved that definition 3.36 is equivalent to the ε, δ one.

From the definitions of limit and standard part, we see that if $\mathrm{st}(f(x)) = L$ for all x infinitely close but not equal to a, then

$$\lim_{x \to a} f(x) = L.$$

Corollary 3.38. *If* $\lim_{x \to a} f(x)$ *exists, then* $f(x)$ *is defined for all real* $x \neq a$ *in some neighborhood of* a.

Proof. Let $X = dom(f) \cup \{a\}$. Then by Expansion $X^* = dom(f^*) \cup \{a\}$ and by the definition of limit, $f(x)$ must be defined for all $x \neq a$ in monad(a), so monad$(a) \subseteq X^*$. Therefore, by Theorem 3.21, X is a neigborhood of a, thus $f(x)$ is defined for all real $x \neq a$ in that neighborhood. $\square$

The following corollary states that the definition of derivative given in the previous section and the standard one coincide.

Corollary 3.39. *The slope of f at a is given by the limit*

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

Proof. It follows easily from the definitions. $\square$

The limit with respect to a subset X of $\mathbb{R}$ is defined as follows.

Definition 3.40. *Let L, a be real numbers. We say that L is the* limit of $f(x)$ as x approaches a in X, *and we write*

$$L = \lim_{x \to a, x \in X} f(x)$$

if whenever $x \in X^*$ and $x \approx a$ but $x \neq a$, we have $f(x) \approx L$.

The one-sided limits are defined by

$$\lim_{x \to a^-} f(x) = \lim_{x \to a, x < a} f(x)$$
$$\lim_{x \to a^+} f(x) = \lim_{x \to a, x > a} f(x).$$

Proposition 3.41. $\lim_{x \to a} f(x)$ *exists if and only if both one-sided limits exist and are equal.*

Proof. It follows from the definitions. $\square$

Let us see some properties of limits, which follow easily from the ones of standard parts.

Theorem 3.42. *Suppose the limits*

$$\lim_{x \to a} f(x), \quad \lim_{x \to a} g(x)$$

both exist. Then

1. *For any constant c,* $\lim_{x \to a}(cf(x)) = c \lim_{x \to a} f(x)$
2. $\lim_{x \to a}(f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$
3. $\lim_{x \to a} f(x) \cdot g(x) = (\lim_{x \to a} f(x)) \cdot \lim_{x \to a} g(x))$
4. *If* $\lim_{x \to a} g(x) \neq 0, \lim_{x \to a}(f(x)/g(x)) = (\lim_{x \to a} f(x))/(\lim_{x \to a} g(x))$.

Proof. It can be easily proved by using the Theorem 3.18 and the fact that $\lim_{x \to a} f(x) = \mathrm{st}(f(x))$. $\square$

We now give the definition of continuity in terms of infinitesimals, and the corollary that states this definition and the standard ε, δ one are equivalent.

Definition 3.43. *We say that f is* continuous *at $a \in \mathbb{R}$ if $f(a)$ is defined and $f(x) \approx f(a)$ whenever $x \approx a$.*

Corollary 3.44. *Let f be a real function and let c be a real number. The following are equivalent:*

1. *f is continuous at a. That is, whenever $x \approx a$, then $f(x) \approx f(a)$.*
2. *There exists a hyperreal $\delta > 0$ such that whenever $|x-a| < \delta$, then $f(x) \approx f(a)$.*
3. *For every real $\varepsilon > 0$ there exists a real $\delta > 0$ such that for all real $x \in (a-\delta, a+\delta)$, we have $|f(x) - f(a)| < \varepsilon$.*

Proof. It follows from Theorem 3.2. □

The following results are immediate consequences of the definitions. The first one is the usual condition for continuity in terms of limits.

Corollary 3.45. *f is continuous at a real point a if and only if $f(a)$ is defined and $\lim_{x\to a} f(x) = f(a)$.*

Corollary 3.46. *If f is continuous at a, then $f(x)$ is defined for all real x in some neighborhood of a.*

It follows from Theorem 3.42 that sums, products, and quotients of continuous functions are continuous, provided that the denominator is not 0.

Theorem 3.47. *If f is differentiable at a, then f is continuous at a.*

Proof. Let f be differentiable at a and let $x \approx a$ but $x \neq a$. Then $f(a)$ is defined and

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

is finite and $x - a$ is infinitesimal. It follows from Theorem 3.13 that $f(x) - f(a)$ is infinitesimal. Hence f is continuous at a. □

Proposition 3.48. *Compositions of continuous functions are continuous, that is if f is continuous at a and g is continuous at $f(a)$, then $h(x) = g(f(x))$ is continuous at a.*

Proof. Let $x \approx a$. Then $f(x) \approx f(a)$ and $g(f(x)) \approx g(f(a))$, whence $h(x) \approx h(a)$ obviously. □

We now give the definition of continuity and uniform continuity on a subset X of the real numbers $\mathbb{R}$.

Definition 3.49. *Let X be a subset of the domain of a real function f. We say that f is* continuous *on X if whenever $a \in X$, $x \approx a$, and $x \in X^*$, we have $f^*(x) \approx f^*(a)$.*

f is uniformly continuous *on X if whenever $x, y \in X^*$ and $x \approx y$, we have $f^*(x) \approx f^*(y)$.*

It follows from the above definitions that if f is uniformly continuous on X, then f is continuous on X.

Theorem 3.50. *Let $X \subseteq \mathbb{R}$ be compact. If f is continuous on X, then f is uniformly continuous on X.*

Proof. Suppose f is continuous on X. Let $x, y \in X^*$ and $x \approx y$. By Corollary 3.57, x is finite and $a := \mathrm{st}(x) \in X$. Then $a = \mathrm{st}(y)$, as $x \approx y$. Since f is continuous on X,

$$f(x) \approx f(a) \quad \text{and} \quad f(y) \approx f(a).$$

Therefore $f(x) \approx f(y)$ and f is uniformly continuous on X. □

The following theorem states that one can extend the domain of a uniformly continuous function from an interval to the whole real line.

Theorem 3.51.

1. *Let f be uniformly continuous on an interval I. Then there exists a function g such that $g \restriction_I = f$ and it is uniformly continuous on the whole real line.*

2. *Suppose the derivative f' of f is uniformly continuous on an interval I. Then there exists a function g such that $g \restriction_I = f$ and g' is uniformly continuous on the whole real line.*

Proof. We give the proof for the case where I is a half-open interval of the form $[a, b)$. The other cases can be proved similarly.

1. Firstly, we show that $\lim_{x\to b^-} f(x)$ exists. Let $x < b$ such that $x \approx b$. Suppose that $f(x)$ is infinite. Then for each real number $r > 0$,

$$\mathbb{R}^* \models r < b \to \exists y(y \in (r,b) \wedge |f(r) - f(y)| \geq 1)$$

Thus,

$$\mathbb{R} \models \forall u(0 < u < b \to \exists y(y \in (u,b) \wedge |f(u) - f(y)| \geq 1),$$

and so

$$\mathbb{R}^* \models \forall u(0 < u < b \to \exists y(y \in (u,b) \wedge |f(u) - f(y)| \geq 1).$$

Therefore, for $u_1 < b$ such that $u_1 \approx b$ there exists $y_1 \in (u_1, b)$ such that $|f(u) - f(y_1)| \geq 1$. Whence $u_1 \approx y_1$, but $f(u_1) \not\approx f(y_1)$. Consequently $f(x)$ is finite and has a standard part B. Then for all $y < b$ such that $y \approx x$, we have $f(y) \approx f(x)$ and thus $f(y) \approx B$. Hence $B = \lim_{x\to b^-} f(x)$. Now consider the function g such that

$$g(x) = \begin{cases} f(a), \text{ if } x < a \\ f(x), \text{ if } x \in [a,b) \\ B, \text{ if } x \geq b. \end{cases}$$

Then g restricted to $[a,b)$ is f and is uniformly continuous on the whole real line.

2. From the proof of (1), we have that the limits

$$B = \lim_{x\to b^-} f(x), \quad C = \lim_{x\to b^-} f'(x)$$

both exist. Let g be the function

$$g(x) = \begin{cases} f(a) + f'(a)(x-a), \text{ if } x < a \\ f(x), \text{ if } x \in [a,b) \\ B + C(x-b), \text{ if } x \geq b. \end{cases}$$

Then g restricted to $[a,b)$ is f and has a uniformly continuous derivative on the whole real line.

$\square$

Now we introduce the notion of a partition of a closed interval $[a,b]$ into infinitely many subintervals of equal infinitesimal length. Consider the hyperreal numbers x, y. The set

$$[x,y]^* = \{z \in \mathbb{R}^* : x \leq z \leq y\}$$

is called a *hyperreal closed interval.* When $a \le x \le y \le b$, we call $[x, y]^*$ a *hyperreal subinterval* of $[a, b]^*$. An *infinitesimal interval* is an interval of the form $[x, x+\Delta x]^*$, where $0 < \Delta x$ and $\Delta x \approx 0$. Since there is no confusion, we usually write $[x, x+\Delta x]$ instead of $[x, x+\Delta x]^*$. Given $n \in \mathbb{N}^*$, the closed hyperreal interval $[a, b]^*$ may be partitioned into subintervals of length $\delta = (b-a)/n$. The partition points are

$$a, a+\delta, a+2\delta, \ldots, a+k\delta, \ldots, a+n\delta = b,$$

where $k \in \mathbb{N}^*$ runs from 0 to n. When n is infinite, the length of each subinterval will be infinitesimal, and the partition is called an *infinite partition of* $[a, b]^*$.

The following corollary will be needed later.

Corollary 3.52. *Let $[a, b]$ be a real interval, n a positive hypernatural number and let $\delta = (b-a)/n$. Then $[a, b]^*$ is the union of the subintervals*

$$[a+k\delta, a+(k+1)\delta)^*,$$

where $k \in \mathbb{N}^*$ and $k < n$.

Proof. Consider $x \in [a, b]^*$ and let k be the greatest hypernatural number such that $k \le (x-a)/\delta$. Then

$$k \le \frac{x-a}{\delta} < k+1.$$

Therefore

$$0 \le k < \frac{b-a}{\delta} = n \quad \text{and} \quad a+k\delta \le x < a+(k+1)\delta.$$

$\square$

We will now give some well-known results of continuous functions and use hyperreals to prove them.

Theorem 3.53 (Intermediate Value Theorem)**.** *Suppose f is continuous on the closed interval $[a, b]$. Then for every real number $d \in [f(a), f(b)]$ there is a point $c \in [a, b]$ such that $f(c) = d$.*

Proof. We may assume that $f(a) \le f(b)$. Since the result is obvious if $d = f(a)$ or $d = f(b)$, we assume that $a < b$ and $f(a) < d < f(b)$. Let $n \in \mathbb{N}^+$ and let

$$a, a+\delta, a+2\delta, \ldots, a+n\delta = b$$

be a finite partition, where $\delta = (b-a)/n$. As the value of f must cross d in one of the subintervals. we have

$$\mathbb{R} \models \forall x \forall y (x \in \mathbb{N}^+ \wedge xy = (b-a) \rightarrow \exists z (z \in \mathbb{N} \wedge z < x \wedge f(a+zy) \le d \le f(a+(z+1)y))).$$

Hence

$$\mathbb{R}^* \models \forall x \forall y (x \in \mathbb{N}^{+*} \wedge xy = (b-a) \to \exists z (z \in \mathbb{N}^* \wedge z < x \wedge f(a+zy) \leq d \leq f(a+(z+1)y))).$$

Now we can take an infinite $n_1 \in \mathbb{N}^*$ and the infinitesimal $\delta_1 = (b-a)/n$. Therefore, there exists $m \in \mathbb{N}^*$ such that $m < n_1$ and $f(a + m\delta_1) \leq d \leq f(a + (m+1)\delta_1)$. Let $c = \mathrm{st}(a + m\delta_1)$. We want to see that $a \leq c \leq b$ and $f(c) = d$. We have

$$a \leq a + m\delta_1 \leq a + (m+1)\delta_1 \leq a + n_1\delta_1 = b.$$

Now taking standard parts, $a \leq c \leq b$. Since f is continuous on $[a, b]$,

$$f(c) = \mathrm{st}(f(a + m\delta_1)) \leq d$$

$$f(c) = \mathrm{st}(f(a + (m+1)\delta_1)) \geq d.$$

Thus $f(c) = d$. □

Definition 3.54. *A function f is called* increasing *if $f(x) < f(y)$ whenever $x < y$ and $x, y \in \mathrm{dom}(f)$. f is called* decreasing *$f(x) > f(y)$ whenever $x < y$ and $x, y \in \mathrm{dom}(f)$. Otherwise, f is said to be* constant *when $f(x) = f(y)$ for all $x, y \in \mathrm{dom}(f)$.*

Theorem 3.55. *If f is a continuous one to one function such that $\mathrm{dom}(f) = I$, where I is an interval, then f is either strictly increasing or decreasing.*

Proof. The proof of this theorem uses standard methods only, so we will not give it here. □

Definition 3.56. *We say that f has a* maximum *at a if $f(a) \geq f(x)$ for all element x of the domain of f. f has a* minimum *at a if $f(a) \leq f(x)$ for all x of the domain of f.*

Proposition 3.57. *Suppose f has a maximum at a. Then f^* also has a maximum at a, i.e., $f^*(a) \geq f^*(x)$ for all hyperreal $x \in \mathrm{dom}(f^*)$.*

Proof. Let f be a function which has a maximum at a. Then

$$\mathbb{R} \models \forall x (f(x) \text{ is defined} \to f(a) \geq f(x)).$$

Therefore,

$$\mathbb{R}^* \models \forall x (f^*(x) \text{ is defined} \to f^*(a) \geq f^*(x)).$$

□

Definition 3.58. *We say that f has a* local maximum *at a if there exists an open real interval $(a-r, a+r)$ such that $f(x)$ is defined and $f(a) \geq f(x)$, for all $x \in (a-r, a+r)$. A* local minimum *is defined analogously.*

The above definition of local maximum can be characterized in terms of hyperreals as follows.

Theorem 3.59. *f has a local maximum at a if and only if $f(x)$ is defined and $f^*(a) \geq f^*(x)$ for all hyperreal $x \approx a$.*

Proof. Suppose f has a local maximum at a. Then $f(a) \geq f(x)$ for all x in some open real interval $(a-r, a+r)$. Similarly to Proposition 3.57, we can see that $f(a) \geq f(x)$ for each $x \in \mathbb{R}^*$ such that $a - r < x$ and $x < a + r$, whence $f(a) \geq f(x)$ for each hyperreal $x \approx a$.

Now suppose f does not have a local maximum at a. Assume first there is no real open interval $(a - r, a + r)$ on which f is defined and therefore, by Corollary 3.23, there is a hyperreal number $x \approx a$ at which $f(x)$ is not defined. Now suppose that f is defined on some real open interval $(a - r, a + r)$. Let $s < r$ be a positive real number. Since f does not have a local maximum at a,

$$\mathbb{R} \models \forall z(0 < z < r \to \exists x(a - z < x \ \wedge \ x < a + z \ \wedge \ f(a) < f(x))).$$

Then,

$$\mathbb{R}^* \models \forall z(0 < z < r \to \exists x(a - z < x \ \wedge \ x < a + z \ \wedge \ f(a) < f(x))).$$

Therefore taking an infinitesimal number z, there is a hyperreal number $x \approx a$ such that $f(a) < f(x)$. $\square$

Theorem 3.60 (Extreme Value Theorem)**.** *If the domain of f is a closed interval $[a, b]$ and f is continuous on $[a, b]$, then f has a maximum and a minimum.*

Proof. We may assume $a < b$, since the result is trivial if $a = b$. Let $n \in \mathbb{N}^+$ and consider the finite partition

$$a, a + \delta, a + 2\delta, \ldots, a + n\delta = b,$$

where $\delta = (b - a)/n$. Let $f(a + m\delta)$ be the greatest of the values

$$f(a), f(a + \delta), \ldots, f(a + n\delta),$$

and let g be the function on $\mathbb{N}^+$ such that $g(n) = m$. Then

$$\mathbb{R} \models \forall x \forall y \forall z(x \in \mathbb{N}^+ \wedge xy = (b - a) \wedge z = g(x) \to a \leq a + zy \leq b) \tag{3.1}$$

and consequently,

$$\mathbb{R}^* \models \forall x \forall y \forall z (x \in \mathbb{N}^{+*} \wedge xy = (b-a) \wedge z = g(x) \rightarrow a \leq a + zy \leq b).$$

Furthermore, since the following formula holds in $\mathbb{R}$:

$$\forall x \forall y \forall z \forall u (x \in \mathbb{N}^{+} \wedge xy = (b-a) \wedge z = g(x) \wedge u \in \mathbb{N}^{+} \wedge u \leq x \rightarrow f(a+zy) \geq f(a+uy)),$$

the following one holds in $\mathbb{R}^*$:

$$\forall x \forall y \forall z \forall u (x \in \mathbb{N}^{+*} \wedge xy = (b-a) \wedge z = g(x) \wedge u \in \mathbb{N}^{+*} \wedge u \leq x \rightarrow f(a+zy) \geq f(a+uy)),$$

Now we can take a positive infinite hypernatural n_1, $\delta_1 = (b-a)/n_1$ and $m = g(n_1)$. We show that f has a maximum at $c = \mathrm{st}(a + m\delta_1)$. By (3.1) we have that

$$a \leq a + m\delta_1 \leq b.$$

Taking standard parts,

$$a \leq c \leq b.$$

Let x be any real number such that $x \in [a, b]$. By Corollary 3.52,

$$x \in [a + k\delta_1, a + (k+1)\delta_1]^*,$$

where k is a hypernatural number between 0 and n_1. Then $x = \mathrm{st}(a + k\delta_1)$. Now considering k in addition to n_1, δ_1 and m, we have that

$$f(a + m\delta_1) \geq f(a + k\delta_1).$$

Since f is continuous on $[a, b]$,

$$f(c) = \mathrm{st}(f(a + m\delta_1)) \geq \mathrm{st}(f(a + k\delta_1)) = f(x).$$

Hence f has a maximum at c.

$\square$

Theorem 3.61 (Critical point Theorem)**.** *Suppose the domain of f is an interval I, f is continuous on I and f has a maximum or a minimum at a point $a \in I$. Then one of the following occurs:*

1. *a is an endpoint of I*
2. *$f'(a)$ is undefined*
3. $f'(a) = 0$

Proof. Suppose that neither (1) nor (2) holds. We must show that (3) is true. Suppose f has a maximum at a and let $\Delta x > 0$ be infinitesimal. Then

$$f(a+\Delta x) \leq f(a) \quad \text{and} \quad f(a-\Delta x) \leq f(a).$$

Therefore

$$\frac{f(a+\Delta x) - f(a)}{\Delta x} \leq 0 \leq \frac{f(a-\Delta x) - f(a)}{-\Delta x}$$

and taking standard parts,

$$f'(a) \leq 0 \leq f'(a).$$

Thus $f'(a) = 0$. □

A point a where $(1), (2)$ or (3) happens is called a *critical point* of f. A critical point which is not an endpoint of I is called an *interior critical point* of f.

Theorem 3.62 (Mean Value Theorem)**.** *If $a < b$ and f is continuous on the closed interval $[a, b]$ and differentiable on (a, b). Then there exists a point $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

We do not give the proof of this Theorem, since it follows from the Extreme Value and Critical Point Theorems and uses standard procedures only.

The Intermediate, Extreme, and Mean Value Theorems have the following useful consequences which involve hyperreal numbers. They can be easily proved, so we omit the proofs.

In each of the following theorems we suppose that f is a real function which is continuous on a closed interval I.

Theorem 3.63 (Hyperreal Intermediate Value Theorem)**.** *For each $a, b \in I^*$ such that $a < b$, if $y \in [f^*(a), f^*(b)]$ is a hyperreal number, then there is a hyperreal $x \in [a, b]$ such that $f(x) = y$.*

Theorem 3.64 (Hyperreal Extreme Value Theorem)**.** *For each $a, b \in I^*$ such that $a < b$, f^* has a maximum and a minimum on the hyperreal closed interval $[a, b]^*$.*

Theorem 3.65 (Hyperreal Mean Value Theorem)**.** *If $a < b$ and f is continuous on the closed interval $[a, b]$ and differentiable on (a, b), and let $g = f'$. Then there exists a hyperreal number $x \in (a, b)^*$ such that*

$$g^*(x) = \frac{f^*(b) - f^*(a)}{b - a}.$$

3.6 Integration

Throughout this section we assume that f and g are real functions which are continuous on an interval I.

Definition 3.66. *Let $[a, b]$ be a subinterval of I and let Δx be a positive real number. We define the* Riemann sum $\sum_a^b f(x)\Delta x$ *as the sum*

$$\sum_a^b f(x)\Delta x = f(x_0)\Delta x + f(x_1)\Delta x + \ldots + f(x_{n-1})\Delta x + f(x_n)(b - x_n),$$

where n is the largest integer such that $a + n\Delta x < b$ and

$$x_0 = a, x_1 = a + \Delta x, \ldots, x_n = a + n\Delta x.$$

We observe that the Riemann sum is a real function of the three variables $a, b, \Delta x$. If we fix a and b and we replace the positive real Δx in this function by a positive infinitesimal dx, then the elementary extension gives us the infinite Riemann sum.

Definition 3.67. *Let f be a continuous real function I, $[a, b]$ a subinterval of I and let*

$$S(\Delta x) = \sum_a^b f(x)\Delta x$$

be the finite Riemann sum. Then

$$S^*(dx) = \sum_a^b f(x)dx$$

is called the infinite Riemann sum.

The infinite Riemann sum is defined for every hyperreal $dx > 0$, since the finite Riemann sum is defined for every real $\Delta x > 0$. We will define the integral as the standard part of the infinite Riemann sum, hence we must prove that this sum is finite, so its standard part exists.

Lemma 3.68. *Let $a < b$ in I and let dx be a positive infinitesimal. Then the infinite Riemann sum $\sum_a^b f(x)dx$ is a finite hyperreal number.*

Proof. By the Extreme Value Theorem 3.60, f has a minimum m and a maximum M on $[a, b]$. Then for each positive real Δx we have

$$\sum_a^b m\Delta x \leq \sum_a^b f(x)\Delta x \leq \sum_a^b M\Delta x,$$

and

$$\sum_a^b m\Delta x = m(b-a), \quad \sum_a^b M\Delta x = M(b-a).$$

That is

$$\mathbb{R} \models \forall y(0 < y \to m(b-a) \leq \sum_a^b f(x)y \leq M(b-a)).$$

Hence

$$\mathbb{R}^* \models \forall y(0 < y \to m(b-a) \leq \sum_a^b f(x)y \leq M(b-a)).$$

We can take a positive infinitesimal dx, so $\sum_a^b f(x)dx$ is finite. $\square$

Definition 3.69. *Let $a < b$ in I and let dx be a positive infinitesimal. We define the* definite integral *of f from a to b with respect to dx as the standard part of the infinite Riemann sum, that is*

$$\int_a^b f(x)dx = \mathrm{st}\left(\sum_a^b f(x)dx\right).$$

Furthermore,

$$\int_a^a f(x) = 0, \quad \int_b^a f(x)dx = -\int_a^b f(x)dx.$$

We now give some properties of the definite integrals.

Theorem 3.70. *Let $a, b \in I$ such that $a < b$, let $c \in \mathbb{R}$ be a constant and let dx be a positive infinitesimal. Then*

1. $\int_a^b c\,dx = c(b-a)$
2. $\int_a^b c\,f(x)dx = c\int_a^b f(x)dx$
3. $\int_a^b (f(x)+g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
4. *If $f(x) \leq g(x)$ for all $x \in [a,b]$, then $\int_a^b f(x)dx \leq \int_a^b g(x)dx$*

Proof. We give the proof of 2. The other cases can be proved analogously. We have

$$\mathbb{R} \models \forall y(0 < y \to \sum_a^b cf(x)y = c\sum_a^b f(x)y).$$

Therefore

$$\mathbb{R}^* \models \forall y(0 < y \to \sum_a^b cf(x)y = c\sum_a^b f(x)y).$$

Then we can take a positive infinitesimal dx, so $\sum_a^b cf(x)dx = c\sum_a^b f(x)dx$. Now taking standard parts, we obtain what we wanted. $\square$

The definite integral $\int_a^b f(x)dx$ does not depend on the infinitesimal dx. The following theorem shows it.

Theorem 3.71. *Let $a, b \in I$ such that $a < b$ and let dx and du be positive infinitesimals. Then*

$$\int_a^b f(x)dx = \int_a^b f(u)du.$$

Proof. It is sufficient to prove that for every positive real number r,

$$\int_a^b f(x)dx \leq \int_a^b f(u)du + r.$$

Consider $c = r/(b-a)$. If we show that

$$\sum_a^b f(x)dx \leq \sum_a^b (f(u)+c)du, \tag{3.2}$$

then, by Theorem 3.70,

$$\int_a^b f(x)dx \leq \int_a^b (f(u)+c)du = \int_a^b f(u)du + r.$$

Let Δx and Δu be positive real numbers. If

$$\sum_a^b f(x)\Delta x > \sum_a^b (f(u)+c)\Delta u,$$

then there exist x, u in $[a, b]$ such that $u \in [x - \Delta u, x + \Delta x]$ and $f(x) > f(u) + c$. Therefore the following formula holds in $\mathbb{R}$

$$\begin{aligned}&\forall y \forall z (y > 0 \wedge z > 0 \wedge \sum_a^b f(x)y > \sum_a^b (f(u)+c)z \rightarrow \\ &\qquad \exists u \exists v (u \in [a,b] \wedge v \in [a,b] \wedge v \in [u-z, u+y] \wedge f(u) > f(v)+c)).\end{aligned}$$

Thus the following is true in $\mathbb{R}^*$

$$\begin{aligned}&\forall y \forall z (y > 0 \wedge z > 0 \wedge \sum_a^b f(x)y > \sum_a^b (f(u)+c)z \rightarrow \\ &\qquad \exists u \exists v (u \in [a,b] \wedge v \in [a,b] \wedge v \in [u-z, u+y] \wedge f(u) > f(v)+c)).\end{aligned}$$

Now suppose dx and du are infinitesimals such that

$$\sum_a^b f(x)dx > \sum_a^b (f(u)+c)du.$$

Then there exist hyperreals x_1, u_1 that satisfy

$$x_1 \in [a,b], \quad u_1 \in [a,b], \quad u_1 \in [x_1 - du, x_1 + dx], \quad f(x_1) > f(u_1) + c.$$

Since dx, du are infinitesimals, we have that $x_1 \approx u_1$ and $f(x_1) \not\approx f(u_1)$. This contradicts the continuity of f. Therefore (3.2) is true and the proof is completed. $\square$

Conclusions

Throughout this work we have reviewed some basic notions of first-order logic and we have learnt new concepts, such as filters or ultraproducts, as well as we have seen important results related to them. A couple of examples could be the Łoś theorem or the ultraproduct version of the compactness theorem of first-order logic. Moreover, we have defined the hyperreal numbers as an elementary extension of the structure of the real numbers. Finally, we have been able to develop rigorously notions of calculus as derivatives, limits, or continuity, based on an infinitesimal approach. Therefore, the initial objectives have been met.

This work could be continued in a natural way by showing more properties and results of integration. Besides trigonometric and exponential functions, infinite series or differential equations can be defined by using infinitesimals.

In this project we have shown that many branches of mathematics can be related to others, although it may not seem so at first glance. In this case, model theory and analysis together develop non-standard analysis.

Bibliography

[1] J.L. Bell and A.B. Slomson, *Models and Ultraproducts: An Introduction*, North Holland Publishing Company, 1974.

[2] C.C. Chang and H.J. Keisler, *Model Theory*, North Holland, Amsterdam, 1973.

[3] J.W. Dauben, *Abraham Robinson: The Creation of Nonstandard Analysis, A Personal and Mathematical Odyssey*, Princeton University Press, 1998.

[4] H.J. Keisler, *Foundations of Infinitesimal Calculus*, `https://people.math.wisc.edu/~keisler/foundations.pdf`, 2022.

[5] P.A. Loeb, Nonstandard Analysis for the Working Mathematician, In P.A. Loeb and M. Wolff, editors, *A simple introduction to nonstandard analysis*, 3-33, Kluwer Academic Publishers, 2000.

[6] A. Prestel, *Model Theory for the real algebraic geometer*, Istituti Editoriali e Poligrafici Internazionali, 1998.

[7] A. Robinson, *Non-standard Analysis*, North Holland Publishing Company, Amsterdam, 1966.

[8] A. Robinson, Non-standard Analysis, In H.J. Keisler, S. Körner, W.A.J. Luxemburg, and A.D. Young, editors, *Selected Papers of A. Robinson*, Vol. 2, 3-11, North Holland Publishing Company, Amsterdam, 1979.

[9] K. Tent and M. Ziegler, *A Course in Model Theory*, Cambridge University Press, 2012.

[10] J.K. Truss *Foundations of Mathematical Analysis*, Clarendon Press, Oxford, 1997.

[11] L. Van Den Dries, *Alfred Tarski's elimination theory for real closed fields*, The Journal of Symbolic Logic, Vol. 53, No. 1, 7-19, 1988.

Index