SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Quantifier elimination and decidability of infinitary theories of the real line

av

Daniel Zavala-Svensson

2016 - No 20

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Quantifier elimination and decidability of infinitary theories of the real line

Daniel Zavala-Svensson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Erik Palmgren

2016

Abstract

In this thesis, we extend first extend logic language to infinitary languages, where we allow for con- and disjunctions of infinite sets of formulas, and quantifiers can bind infinite sets of variables. The cardinalities of those sets are bounded however, and based on those bounds we investigate the existence of quantifier elimination and decision methods for infinitary theories on the ordered field of reals. With analytic sets from descriptive set theory as a counterexample we prove the main result: The countably infinite theory of the ordered field of reals does not have quantifier elimination.

Sammanfattning

I denna uppsats börjar vi med att utöka ändlig logik till oändlig logik, där vi tillåter kon- och disjunktioner av oändliga mängder av formler, och där kvantorer kan binda oändliga mängder av variabler. Dessa mängders kardinalitet är begränsad, och beroende på de begränsningarna undersöker vi huruvida existensen av kvantorelimination och avgörbarhetsmetoder hos de reella talen som ordnad kropp. Vi använder analytiska mängder från den deskriptiva mängdläran som motexempel för att bevisa uppsatsens huvudresultat: Den uppräkneligt oändliga teorin om de reella talen som ordnad kropp har inte kvantorelimination

Acknowledgements

I want to thank my supervisor, professor Erik Palmgren, for obvious reasons such as his patience and support and technical advise. But also for the very kernel of this thesis: The hunch that analytic sets were the key.

I would also like to thank Lina Johansson, without whom this thesis would not exist.

Contents

1. Introduction

In 1931, Gödel [1] with his incompleteness theorems put an end to the great formalist project, championed by Hilbert [2, problem 2], to show that all mathematics could be algorithmically proved using only syntactic manipulation of logic notation. The theorems were not, however, the death of the concept of pure syntactic manipulation, or even proofs by algorithms. The scope was merely reduced, from *everything* to classes of mathematical structures where such algorithms are possible, and to the specific kinds of proofs that are amenable to such syntactic games.

Tarski [8] showed in 1951 that one such structure is the ordered field of reals, i.e. $\mathbb{R}$ described by the symbols $(+, -, \cdot, <, 0, 1,)$ and variables. He showed that every (finite) logical statement, without free variables, about equalities and inequalities of polynomials with integer coefficients, can be algorithmically decided upon (see section 4.1). Further, he also proved that all such statements, even those with free variables, can be reduced to statements without quantifiers ($\forall$ and $\exists$), in a sense simplifying them.

This thesis expands slightly on this matter. In chapter 2 it introduces infinitary languages, which allow for logical statements of "infinite length", and proves some propositions about these languages analogous to well-known theorems of finitary logic. Chapter 3 introduces some notions of topology, paving the way for the next chapter. In chapter 4 we give some examples of theorems which can be generalised from finitary logic to infinitary logic. There we also reach the main points: That contrary to in finitary languages, in infinitary languages the statements about $\mathbb{R}$ can *not* in general be expressed without quantifiers, and there is *not* an algorithm for deciding infinitary sentences about $\mathbb{R}$.

1.1 Model Theory

The subject area is model theory. As always, whenever one tries to provide a compact definition of any given branch of mathematics, there will be instances when something which undoubtedly belongs to that branch still falls outside of that definition. One can try to remedy this by giving vaguer and vaguer definitions, until the definition is so ambiguous that nothing can be gained from reading it. Therefore, we shall not try to give an all-encompassing definition of model theory, but rather a concise and eloquent definition, which covers at least the core of the subject:

Model theory is the classification and study of algebraic universal structures, by defining the structures in terms of logical formulas which are true on the underlying domain, and the subsequent study of such formulas. Or, in the words of Keisler and Chang [5]:

$$\text{universal algebra} + \text{logic} = \text{model theory}.$$

The subject is characterised by the dichotomy between *syntax* and *semantics*, or as we shall call them, *language* and *model*. Facts and properties which can be regarded as true or false belong to the model. The language in turn is the precise systematisation of how we make statements about those facts and properties, in the form of strings of logical symbols. We can talk and prove things about the model, and the language is the grammar, sterile but giving a strict structure to our thoughts.

Separating the language of statements from the meaning of those statements is a powerful thing. It is a form of abstraction, and like every other form of abstraction, it shows us how objects we previously held as separate can be unified. When we consider the structure of the statements separate from their meaning we can identify models which share languages, and consider them as equivalent (technically we say that they are logically equivalent.) Thus model theory provides us with a way of classifying models based on similarities between the syntactic structure of statements about them. It follows that we can share proofs and theorems freely between equivalent models, so long as those proofs and theorems have the correct structure.

This, in the author's opinion, is at the heart of model theory.

2. Logic or Talking about Structures

2.1 Universal algebra

Model theory uses logic as a tool and is therefore regarded as a branch of logic, but the objects of study are those of universal algebra: Structures.

Structures in universal algebra are generalisations of algebraic structures such as groups, fields, graphs, ordered sets, etc. Recognising that there are similarities in how such objects are constructed and studied, a general framework is given in which all such structures fit:

Definition. Following Hodges [3], a **structure** A is an object consisting of four parts:

- An underlying set called the **domain** of A, written $\operatorname{dom} A$. By the **elements** of A we mean the elements of $\operatorname{dom} A$. A and $\operatorname{dom} A$ are oftentimes used interchangeably.
- A subset of $\operatorname{dom} A$ whose elements are called the **constants** of A.
- For each positive integer n, a set of subsets of $(\operatorname{dom} A)^n$, called n-ary **relations** of A.
- For each positive integer n, a set of n-ary functions on $\operatorname{dom} A$.

Sometimes the constants are omitted and replaced by 0-ary functions. Structures are sometimes called **models**.

2.2 The words

The promise of model theory is to translate algebra into logic, and the goal of this section is to develop a logical language $L_{\alpha\beta}$, where α and β are ordinals, for this purpose. Every structure will have its own language, and the first step is to abstract the symbols of the structure, so that we can use the logician's symbolic manipulation powers without directly assigning any meaning to the symbols.

Definition. The **signature** L of a structure A is the collection of

- The constant symbols of A.
- For each $n \geq 1$, the symbols of n-ary relations of A.
- For each $n \geq 1$, the symbols of n-ary functions of A.

In this definition, we regard signatures as being generated from structures. But we can also regard signatures as the basic objects, and let these generate structures. A structure would then be a triple containing a domain A, a signature σ, and an interpretation function ι. ι tells us how to interpret σ in terms of A, so takes the constants of σ to elements of A, functions of σ to functions on A, and n-ary relations of σ to subsets of A^n.

Note that the signature does not contain any symbols representing the domain of the structure. To be able to talk about elements of a structure, we introduce variables, to be used with a given signature. A **variable** can be any symbol (often x, y or z, or $v_0, v_1, v_2, \ldots$) which is not already in use in the signature, and they pertain to the language. One can think of them as place holders in the language, for elements of the structure we wish to talk about.

But talking generally about elements of A is not enough. We want to be able to talk about the functional and relational structure of the elements in it, to talk about elements that as function values, and relations between them. To this end we introduce symbols representing elements not only in the form of variables and constants, but also in the form of function values of A.

Definition. The **terms** of a signature L are the symbols generated by variables, the constants of L and the functions of L:

- Every constant of L is a term.

- Every variable is a term.
- For every $n \geq 1$, if $\vec{t}$ is an n-tuple of terms and F is an n-ary function symbol of L, then $F(\vec{t})$ is a term.

We will use the notation $t(\vec{x})$ to denote a term t which contains no other variables than those in $\vec{x}$. $t(\vec{s})$ then, where t is of the form $t(\vec{x})$ for some n-tuple of variables $\vec{x}$ and $\vec{s}$ is a tuple of at least length n, means the string t, with all instances of $x_i \in \vec{x}$ replaced by $s_i \in \vec{s}$. If X is a set of variables, $t(X)$ means that t contains no other variables than those in X.

Next, we wish to be able to make to make statements about the terms of A. The basic building blocks for these statements will be identities and the relation symbols inherited from A.

Definition. An **atomic formula** of a signature L is a string of symbols of one of the forms

- $s = t$, where s and t are terms of L
- R $\left(\vec{t}\right)$, where R is an n-ary relation symbol of L, and $\vec{t}$ is an n-tuple of terms of L.

Note that we assume that the symbol $=$ is not already in use in L.

By a **negated** atomic formula, we mean a string of the form $\neg\phi$, were ϕ is an atomic formula. A **literal** is an atomic formula or a negated atomic formula.

2.3 The sentences

From the atoms we build formulas. They are the last step in creating the language and they represent declarative sentences, or statements, about the structure.

Definition 1. A **formula of the language** $L_{\alpha\beta}$, where L is a signature and α and β are ordinals, is a string of symbols generated by atomic formulas and the characters $\bigwedge$, $\bigvee$, $\neg$, $\top$, $\bot$, $\forall$ and $\exists$ in the following way:

- Every atomic formula is a formula.
- $\top$ and $\bot$ are formulas.

- If ϕ is a formula then $\neg\phi$ is a formula.
- If Φ is a set of formulas, of cardinality $< \alpha$, then $\bigwedge \Phi$ and $\bigvee \Phi$ are formulas.
- If X is a set of variables of cardinality $< \beta$, and ϕ is a formula, then $\forall X \phi$ and $\exists X \phi$ are formulas.
- Nothing else is a formula.

Note that the formulas of $L_{\alpha\beta}$ is the smallest set which contains the atomic formulas and $\top$ and $\bot$, and is closed under the concatenations of symbols described above.

A word on notation: Throughout this thesis we shall need to refer to long formulas by short abbreviations. In these cases we will use the symbol $\equiv$. For instance, $\phi \equiv \bigwedge\{x = y, y = z\}$ means that ϕ is shorthand for what we actually mean, which is $\bigwedge\{x = y, y = z\}$. Moreover we will use the notations X_I or $\{x_i\}_{i \in I}$ to denote **indexed sets** X with index set I. If the symbol for the index is clear from context we shall often omit it like so: $\{x_i\}_I$ instead of $\{x_i\}_{i \in I}$.

If $\Phi = \{\phi_i\}_{i \in I}$ is a set of formulas, then $\neg\Phi = \{\neg\phi_i\}_{i \in I}$. Similarly we will take other operations on elements, applied to a set of those elements, to mean that set with the operation applied to every element in it. It should be clear from context when this is the case. If Φ is the finite set $\Phi = \{\phi_0, \ldots, \phi_n\}$, then $\bigwedge \Phi$ and $\bigvee \Phi$ are usually written with infix notation $\phi_0 \wedge \phi_1 \wedge \ldots \wedge \phi_n$ and $\phi_0 \vee \phi_1 \vee \ldots \vee \phi_n$ respectively. $\bigwedge_{i \in I} \phi_i$ means $\bigwedge\{\phi_i\}_{i \in I}$, and similarly for $\bigvee_{i \in I} \phi_i$, and also for $\bigwedge_{i=m}^{n} \phi_i$, and $\bigvee_{i=m}^{n} \phi_i$ and other similar notations. If y is a single variable, and $\vec{y} = (y_1, \ldots, y_n)$ is an n-tuple of variables, then $\exists y$ means $\exists\{y\}$ and $\exists\vec{y}$ means $\exists\{y_1, \ldots, y_n\}$, and similarly for $\forall y$ and $\forall\vec{y}$. These variations in notation are used to make the text flow easier, and their meaning should be obvious when encountered.

We have not yet given these symbols semantic meaning, but the intended interpretation is clear: Atomic formulas are first-order finitary statements, and $\bigwedge$, $\bigvee$, $\neg$, $\top$, $\bot$, $\forall$ and $\exists$ represent the usual connectives, negation and quantifiers. Some of these operators are redundant, since they can be constructed from each other. Therefore we could have made another choice of symbols to generate the formulas, without inducing anything other than cosmetic changes to the theory and the expressive power the formulas will have once we give them meaning. For example, in propositional logic

$$\begin{aligned} \bot &\longleftrightarrow \bigvee \varnothing \\ \top &\longleftrightarrow \neg\bot \\ \neg\psi \vee \phi &\longleftrightarrow \psi \to \phi. \end{aligned}$$

This means we could restrict the language to the symbols $\bigwedge$, $\bigvee$, $\neg$, $\forall$ and $\exists$. In fact, this is how it is done in Hodges [3]. We could also have added the implication symbol '$\rightarrow$'. If we wish to restrict the set of symbols even further, an even smaller set of symbols is $\bar{\bigwedge}$ ("nand") and $\exists$, where $\bar{\bigwedge}\Phi \leftrightarrow \neg \bigwedge \Phi$, since

$$\begin{aligned} \neg\phi &\longleftrightarrow \bar{\bigwedge}\{\phi\} \\ \bigwedge \Phi &\longleftrightarrow \neg\bar{\bigwedge}\Phi \\ \bigvee \Phi &\longleftrightarrow \bar{\bigwedge}\neg\Phi \\ \forall x\phi &\longleftrightarrow \neg\exists x\neg\phi \end{aligned}$$

in propositional logic.

For a set of formulas Φ, we say that ϕ is a **boolean combination** of the formulas in Φ, if it is generated (in a finite number of steps) by $\neg$, $\bigwedge$ and $\bigvee$ acting on the elements and subsets of $\Phi \cup \{\top, \bot\}$, as per the rules in Definition 1 above. $\langle\Phi\rangle$ means the set of all boolean combinations of the formulas in Φ. If we need to distinguish between boolean combinations in $L_{\alpha\beta}$ where $\alpha > \omega$, and $L_{\omega\beta}$, we may call the former **infinite boolean combinations** and the latter **finite boolean combinations**.

As is evident from the definition, formulas of $L_{\omega\omega}$ are just strings of symbols, built from atomic formulas and $\top$ and $\bot$, and concatenated with the connectives and quantifiers of Definition 1 into ever more complex strings. Many theorems and arguments use recursive manipulation of these strings, and to facilitate this we shall define the complexity of a formula, which may intuitively be taken as a measure of the number of "steps" taken when creating a formula from atomic formulas.

Sometimes it will suffice to consider the number of steps taken not from the atomic formulas, but from some other given set of formulas which acts as the base case. This happens for example in Theorem 2, and for this reason we will define not complexity, but complexity *above* (relative to) a set of formulas.

Also, since we will be dealing with infinitary languages, the complexity will be allowed to be be any ordinal. This will allow us to do transfinite induction (see Jech [4, Chapter 2]).

Definition 2. The following formulas have **complexity** α **above** Φ, where α is an ordinal:

- Every atomic formula, $\top$, $\perp$ and every formula in Φ, has complexity 0 above Φ.
- If ψ is a formula with complexity α above Φ, then $\neg\psi$ and $\exists Y\psi$ and $\forall Y\psi$ have complexities $\alpha + 1$ above Φ.
- If Ψ is a set of formulas, let A be the set of complexities of the elements in Ψ. Then $\bigwedge \Psi$ and $\bigvee \Psi$ have complexities $\sup(A + 1)$ above Φ. Note that this new complexity exists, and is strictly greater than any in A (Jech [4]).

If ψ has complexity α above the empty set, we simply say that ψ has **complexity** α. The set of formulas that have complexity contains the atomic formulas, $\top$ and $\perp$ and is closed under under the logical operators of Definition 1, and since the set of formulas is the smallest such set, every formula has complexity. Complexity implies complexity above any set of formulas, and therefore every formula has complexity above every set of formulas.

We can now define the language:

Definition 3. The **language** $L_{\alpha\beta}$, where L is a signature and α and β are ordinals, is the set of all formulas of $L_{\alpha\beta}$ (see Definition 1). If α and β are both at most ω, so that any formula contains only a finite number of atomic formulas and variables, then $L_{\alpha\beta}$ is a **finitary language**. Otherwise it is an **infinitary language**.

Finally we shall need the notion of *free* and *bound* variables. For this we need to define *occurrences* of variables:

Definition. A variable x **occurs** in a formula if it is used as a symbol in the formula. To be more specific, we define

- If t is the term consisting only of x, then x occurs in t.
- If F is an n-ary function symbol of L and $\vec{t}$ is an n-tuple of terms, then x occurs in $F(\vec{t})$ if x occurs in any one of the terms in $\vec{t}$.
- If s and t are terms, and x occurs in s or t, then x occurs in the atomic formula $s = t$.
- If $\vec{t}$ is an n-tuple of terms, at least one in which x occurs, and R is an n-ary relation, then x occurs in the atomic formula $\mathrm{R}(\vec{t})$.
- If Φ is a formula in which x occurs, $\phi \in \Phi$ and ψ is a formula, then x occurs in $\neg\phi$, $\bigwedge \Phi$ and $\bigvee \Phi$.
- If x occurs in Φ, then x occurs in $\forall X \Phi$ and $\exists X \Phi$. Furthermore, if $x \in X$, then all occurrences of x in Φ are said to be **bound** occurrences.

A variable x is **free** in a formula Φ if it occurs not bound somewhere in Φ. Note that x can both occur bounded in Φ and be free in Φ, as in for example $\psi \equiv y = x \wedge \forall x\ y = x$. x occurs bound in ψ, but also unbound, and is therefore free in ψ. In a sense, the bound x and the free x are different variables.

We shall use the notation $\phi(X)$, where ϕ is a formula and X is a set of variables, to specify that ϕ is a formula where no other free variables than those in X occur. For formulas with a finite number of free variables, denoted here as an n-tuple $\vec{x}$, a common notation is also $\phi(\vec{x})$. With our notation using sets, this means $\phi(X_I)$, where X_I is any indexed set of variables such that every coordinate x_i of $\vec{x}$ is an element of X_I and $I = \{1, \ldots, n\}$.

We shall frequently abuse this notation to in the following way: If $\phi(X_I)$ is a formula with precisely X_I as free variables, then $\phi(T_I)$, where T_I is an I-indexed set of terms, means the formula constructed in the same way as ϕ, but with every occurrence of x_i recursively replaced by the term t_i. The reader will probably not even notice this, since it is just the same convention we use to denote function composition, for example defining $f(x) = x^2$ and then letting $f(-x)$ mean $(-x)^2$. We call this **substitution**, and we say that we **substitute** X_i for T_i.

In the language $L_{\alpha\beta}$, the index set I must always be of cardinality $< \beta$ (and therefore so must X_I be.)

2.4 What the words mean

We have talked figuratively about formulas as "sentences," but formally, a **sentence** is a formula with no free variables, i.e. a formula with what can be considered a well defined semantic meaning with no "meaningless" words (i.e. free variables). A **theory** is a set of sentences of a language (we shall allow only sets, not proper classes.) The idea behind this name is of course that a theory contains a set of statements which accurately describes some structure.

Having constructed a language using the symbols of a structure but independently from it, we will now turn back to our its intended purpose and give meaning to the language. All structures that have the same number of constants, and n-ary functions and relations for every n (i.e. have the same signature) share the same languages, but the languages have different *meanings* for different structures. We introduce anew the interpretation function $\iota_A : \sigma \setminus X \mapsto A$ (where $\sigma \setminus X$ is every symbol of the signature except the variables). Let L be a language of a signature σ. For every structure A with signature σ we let ι_A be a the function which takes every symbol s of σ, except variables, to the respective constant, function or relation in A named by s in σ. Sometimes we denote $\iota_A(s)$ by s^A, as in Hodges [3]. There may of course be more than one interpretation function $\sigma \setminus X \mapsto A$, if there is more than one constant, one n-ary function for some n, or one n-ary relation for some n. In this case, the particular interpretation function s^A uses is considered canonical. It will usually be clear from the context which particular interpretation function is in use.

We cannot extend ι_A in any natural way to all terms of σ, since a variable does not represent a unique element in $\operatorname{dom} A$, but rather any of them. If we specify which element a variable represents however, we can. Let X_I be a set variables, and $S_I \subset \operatorname{dom} A$. We can then let ι_{A,S_I} extend the domain of ι_A to include X_I, by letting ι_{A,S_I} take $x_i \in X_I$ to $s_i \in S_I$.

We can now extend ι_A to a function on all closed terms of σ recursively:

Definition 4. The **interpretation functions** ι_A and ι_{A,S_I} are defined recursively as follows:

- For every constant and function s of σ, $\iota_A(s) = s^A$ as above.
- $\iota_A(F(t_1, t_2, \ldots, t_n)) = F^A(\iota_A t_1, \iota_A t_2, \ldots, \iota_A t_n)$.

This has only extended ι_A to closed terms. If we choose an indexed subset $S_I \subseteq \operatorname{dom} A$, then we extend ι_A to every term of the form $t(X_J)$ where $J \subseteq I$:

- $\iota_{A,S_I}(x_j) = s_j$.

$t[S_I]$ means $\iota_{A,S_I}(t(X_I))$. If $\vec{a}$ is an n-tuple and t is of the form $t(X_I)$, where $\{1, \ldots, n\} \subseteq I$, then $t[\vec{a}]$ means $t[S_{\{1,\ldots,n\}}]$ where $S_{\{1,\ldots,n\}}$ is the indexed set $\{s_j \in \operatorname{dom} A | s_j$ is the j:th coordinate of $\vec{a}\}$.

In one fell swoop, we can now give meaning to all sentences of L in terms of A:

Definition. In all of the following, ϕ must be of the form $\phi(X_J)$, where X_J is a set of variables with index set J, and $J \subseteq I$ which is used as index set for S_I, a subset of $\operatorname{dom} A$ with index set I. This is necessary so that we do not mistakenly try to give meaning in a structure A to formulas with free variables (for example, in the field $\mathbb{R}$, $4 = 5$ and $5 = 5$ means something in the sense we use here, but $x = 5$ does not.)

Given a language L, for a sentence ϕ, $A \vDash \phi$ is read 'A is a **model of** ϕ' or 'ϕ is **true in** A', and:

- If $\phi(X_J)$ is the atomic formula $t(X_J) = s(X_J)$ then $A \vDash \phi[S_I]$ iff $t[S_I] = s[S_I]$.
- If $\phi(X_J)$ is the atomic formula $R(X_J)$, then $A \vDash \phi[S_I]$ iff $R[S_I] \in A$.
- $A \vDash \top$ is always true, and $A \vDash \bot$ is never true.
- $A \vDash \neg\phi[S_I]$ is true iff $A \vDash \phi[S_I]$ is not true.
- If Φ is a set of formulas, then $A \vDash \bigwedge \Phi[S_I]$ iff $A \vDash \phi[S_I]$ for every $\phi \in \Phi$, and $A \vDash \bigvee \Phi[S_I]$ iff there is a $\phi \in \Phi$ such that $A \vDash \phi[S_I]$.

Furthermore, let X_J be a set of variables with index set J and S_I a subset of $\operatorname{dom} A$ with index set I. Then

- $A \vDash (\forall X_J \phi)[S_I]$ iff for every subset $S'_J \subseteq \operatorname{dom} A$ with index set J, $A \vDash \phi[\hat{S}_{I\cup J}]$, where $\hat{S}_{I\cup J} = \{\hat{s}_k | \hat{s}_k = s'_k \text{ if } k \in J, \text{ and } \hat{s}_k = s_k \text{ otherwise.}\}$.

- $A \vDash (\exists X_J \phi)[S_I]$ iff for some subset $S'_J \subseteq \operatorname{dom} A$ with index set J, $A \vDash \phi[\hat{S}_{I \cup J}]$, where $\hat{S}_{I \cup J} = \{\hat{s}_k | \hat{s}_k = s'_k \text{ if } k \in J, \text{ and } \hat{s}_k = s_k \text{ otherwise.}\}$.

For a theory Φ, $A \vDash \Phi$ (read 'A is a **model of** Φ') if A is a model of every sentence in Φ.

The **theory of a structure** A, $\operatorname{Th}_L A$, is the set (or family) of all sentences ϕ of the language L that are true in A. $\operatorname{Th} A$, without specifying the language, means $\operatorname{Th}_L A$ where L is the first-order language of A.

The **models of a theory** Φ, $\operatorname{Mod} \Phi$, is the set (or collection) of all models of Φ. If T is a theory in $L_{\alpha\beta}$ and K is a class of L-structures, we say that T **axiomatises** K if $\operatorname{Mod} T = K$. If A is a structure, we say that T axiomatises A if $\operatorname{Mod} T = \operatorname{Mod} \operatorname{Th} A$. The formulas of T are called **axioms** of K and A, respectively.

Two formulas ϕ and ψ are **equivalent modulo a theory** T if for every structure $A \in \operatorname{Mod} T$, and admissible subset $S \subseteq \operatorname{dom} A$, $A \vDash \phi[S] \iff A \vDash \psi[S]$. We write $\phi \leftrightarrow \psi \pmod{T}$.

Two formulas $\phi(X)$ and $\psi(X)$ of a language $L_{\alpha\beta}$, where X is finite, are **logically equivalent**, or simply **equivalent**, if they are equivalent modulo the empty theory, and we write $\phi \leftrightarrow \psi$. If the language is $L_{\omega\omega}$ they are **elementarily equivalent**. Well-known examples of logically equivalent sentences are $\forall Y \phi \leftrightarrow \neg\exists\neg\phi$, $\neg\neg\phi \leftrightarrow \phi$ and $\bigwedge \Phi \leftrightarrow \neg \bigvee \neg\Phi$. It is easy to see that these hold in infinitary languages.

A formula $\phi(x)$ **defines** the subset S of the structure A if $A \vDash \phi[s] \Leftrightarrow s \in S$. A subset $S \subseteq A$ is **definable** in the language $L_{\alpha\beta}$ if there is a formula in $L_{\alpha\beta}$ which defines S.

It may not be as easy to see that the distributive laws hold in infinitary languages however. These will be needed for example in Theorem 2, so we shall prove them.

Theorem 1 (Distributivity)**.** *In a language $L_{\alpha\beta}$, if $\left|J^I\right| \leq |\alpha|$,*

$$\bigwedge_I \bigvee_J \phi_{ij} \leftrightarrow \bigvee_{f \in J^I} \bigwedge_I \phi_{if(i)}$$
$$\text{and} \tag{2.1}$$
$$\bigvee_I \bigwedge_J \phi_{ij} \leftrightarrow \bigwedge_{f \in J^I} \bigvee_I \phi_{if(i)}.$$

Note that this holds for all formulas if the language is $L_{\omega\beta}$ or $L_{\infty\beta}$ where β is any ordinal.

Proof. We prove the first equivalence: The LHS is true when for every $i \in I$ there is a $j \in J$ such that ϕ_{ij} is true, i.e. there is a function $h : I \mapsto J$ such that $\phi_{if(i)}$ is true for every i. But then for this specific h, $\bigwedge_I \phi_{ih(i)}$ is true, so the RHS is true. The LHS is false when there is an i such that ϕ_{ij} is false for every j. But then every conjunction $\bigwedge_I \phi_{if(i)}$ is false since $\phi_{if(i)}$ will fail for the one i. So the RHS is also false. So the RHS and LHS are both true or both false. The second equivalence is proved in a similar manner. □

Two logical equivalences that hold for finite languages but not necessarily for infinitary languages are the disjunctive and conjunctive normal forms. A formula ϕ is in **conjunctive normal form over** a set of formulas Φ if $\phi \equiv \bigwedge \{\bigvee \Psi_i\}_I$, where $\Psi \subseteq \Phi \cup \neg\Phi$. ϕ is in **disjunctive normal form over** Φ if $\phi \equiv \bigvee \{\bigwedge \Psi_i\}_I$. If we need to distinguish between normal forms in $L_{\alpha\beta}$ where $\alpha > \omega$ and $L_{\omega\beta}$, we may call the former **finitary conjunctive normal forms** and the latter **infinitary conjunctive normal forms**, and similarly for disjunctive normal forms. Any finite boolean combination of Φ is equivalent to a formula on finitary disjunctive normal form, and a formula on finitary conjunctive normal form, over Φ. If we relax our limits on the cardinalities of connectives, we make corresponding claims for infinitary languages:

Theorem 2. *If $\phi(X)$ is an infinite boolean combination of Φ of a language $L_{\omega\beta}$ or $L_{\infty\beta}$, then there is a formula $\mu(X)$ on infinitary conjunctive normal form over Φ, and a formula $\pi(X)$ on infinitary disjunctive normal form over Φ, such that ϕ, μ and π are logically equivalent.*

Proof. By induction on complexity. Every formula with complexity 0 above

Φ is trivially equivalent to both a disjunctive and a conjunctive normal form over Φ. Let $\phi(X)$ be an infinite boolean combination of Φ, with complexity $\alpha > 0$ above Φ, and assume that the theorem holds for all boolean combinations of Φ with complexity $< \alpha$ above Φ. ϕ is on one of the forms $\neg\theta$, $\bigvee\Theta$ or $\bigwedge\Theta$, where θ and every formula in Θ have complexities $< \alpha$ above Φ. By the induction hypothesis, and for the last equivalence Theorem 1,

$$\begin{aligned}
\neg\theta &\longleftrightarrow \neg\bigwedge_I\bigvee\Psi_i \longleftrightarrow \bigvee_I\bigwedge\neg\Psi_i \\
\bigvee\Theta &\longleftrightarrow \bigvee_J\bigvee_I\bigwedge\Psi_{ij} \longleftrightarrow \bigvee_{I\times J}\bigwedge\Psi_{ij} \\
\bigwedge\Theta &\longleftrightarrow \bigwedge_J\bigwedge_I\bigvee\Psi_{ij} \longleftrightarrow \bigwedge_{I\times J}\bigvee_K\psi_{ijk} \longleftrightarrow \bigvee_{f\in K^{I\times J}}\bigwedge_{I\times J}\psi_{ijf(ij)}
\end{aligned}$$

where every $\Psi_{ij} \subseteq \langle\Phi\rangle$. Similarly we can reduce ϕ to conjunctive normal form over Φ, so by induction the theorem holds. $\square$

3. Topology

A **topology** on a set Ω is a collection of subsets of Ω that is closed under finite intersections and arbitrary unions, and includes the empty set and Ω itself. These subsets are called **open sets**, and a set together with a topology on it is called a **topological space**. The complement of an open set is called a **closed** set.

A collection B of open sets in a topology T such that every open set in T is a (possibly empty) union of elements in B, is called a **base** for T, and B is said to **generate** T. The elements of a base are called **basic open sets**. A collection S of open sets of T such that T is the smallest topology containing S, is called a **subbase** of T.

If $X_1, \ldots, X_n$ are topological spaces, the **projection** $\pi_i : \prod_{i=1}^n X_i \to X_i$ is the function $(x_1, \ldots, x_i, \ldots, x_n) \mapsto x_i$. The **product topology** on the Cartesian product $\prod_{i=1}^n X_i$ is the topology where the preimages of open sets in $X_1, \ldots, X_n$ under the projections $\pi_1, \ldots, \pi_n$ are the subbase. This is equivalent to the smallest topology such that the projections are continuous.

The **subspace topology** of a subset Ω' of a topological space Ω is the topology of all open subsets of Ω intersected with Ω'.

A **continuous** function is a function between two topological spaces such that the preimage of every open set is an open set. A **homeomorphism** is a continuous function with a continuous inverse. Two topological spaces are said to be **homeomorphic** if there is a homeomorphism between them.

Proposition 3. *A function $f : Y \to \prod_{i=0}^n X_i$, where the codomain has the product topology, is continuous iff every composition $\pi_i f$ is continuous.*

Proof. If f is continuous then $\pi_i f$ is continuous since compositions of continuous functions are continuous. For the other opposite implication, note that the collection of all sets of the form $\prod_{i=0}^m A_i$, where each A_i is open

X_i, forms a subbase in $\prod_{i=0}^{n} X_i$. For any such set,

$$f^{-1}\left(\prod_{i=0}^{m} A_i\right) = f^{-1}\left(\bigcap_{i=0}^{m} \pi_i^{-1} A_i\right) = \bigcap_{i=0}^{m} (\pi_i f)^{-1} A_i,$$

which is open since $\pi_i f$ is continuous. So the preimage of every open set in the subbase is open, and thus the preimage of every open set is open. Therefore f is continuous. $\square$

A sequence $\{a_i\}_{\mathbb{N}}$ in a topological space Ω **converges** to $a \in \Omega$ if for every open set A which contains a there is an $N \in \mathbb{N}$ such that $N < i \Rightarrow a_i \in A$. We write $\lim_{i\to\infty} a_i = a$. A sequence that converges to a point is said to be **convergent**.

Proposition 4. *Continuous functions between topological spaces preserve limits. I.e. If $f : X \to Y$ is a continuous function and $\{s_i\}_{\mathbb{N}}$ is convergent, then*

$$\lim_{i\to\infty} f(s_i) = f\left(\lim_{i\to\infty}(s_i)\right).$$

Proof. Let $\lim_{i\to\infty}(s_i) = s$. Let A be any open set containing $f(s)$. Let $B = f^{-1}(A)$. B is open since f is continuous, and $s \in B$. Since $\{s_i\}_{\mathbb{N}}$ converges to s there is an N such that $N < i \Rightarrow s_i \in B$. But this means that $N < i \Rightarrow f(s_i) \in f(B) = A$. So $\{f(s_i)\}_{\mathbb{N}}$ converges to $f(s)$. $\square$

A **metric**, or **distance function**, d on a set Ω is a function $\Omega^2 \to \mathbb{R}$ such that for all $x, y, z \in \Omega$

(i) $\mathrm{d}(x, y) \geq 0$,
(ii) $\mathrm{d}(x, y) = \mathrm{d}(y, x)$,
(iii) $\mathrm{d}(x, y) = 0 \Leftrightarrow x = y$,
(iv) $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.

A set together with a metric on that space is called a **metric space**. An **open ball** of **radius** $r > 0$ **around** a point $a \in \Omega$, denoted $B_r(a)$, is the set of all points $x \in \Omega$ such that $\mathrm{d}(a, x) < r$. The open balls are the basis of a topology (see Waldmann [9]). A topological space with topology T such that there is a metric which generates T in the above sense, is called **metrisable**.

A **Cauchy sequence** is a sequence $\{a_i\}_{\mathbb{N}}$ in a metric space with metric d, such that for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $i, j > N \Rightarrow \mathrm{d}(a_i, a_j) < \varepsilon$. A metric space is **complete** if every Cauchy sequence is convergent.

3.1 Descriptive set theory

The **Baire space** $\mathcal{N} = \mathbb{N}^{\mathbb{N}}$ is the space of all infinite sequences of natural numbers, with the following topology: Define Seq as the set of all *finite* sequences of natural numbers (note that Seq is countable since countable unions of countable sets are countable). For every sequence $s \in \text{Seq}$, let $\mathcal{O}(s)$ be the subset of $\mathcal{N}$ consisting of all sequences starting with s. I.e.

$$\mathcal{O}(s) = \Big\{ r \in \mathcal{N} \mid r = s \restriction n \Big\}. \tag{3.1}$$

Where $s \restriction n$ means the subsequence of s consisting of its n first elements. Let $\widehat{\mathcal{O}}$ be the set of all such sets,

$$\widehat{\mathcal{O}} = \{ \mathcal{O}(s) \mid s \in \text{Seq} \}. \tag{3.2}$$

We let $\widehat{\mathcal{O}}$ be the subbase for $\mathcal{N}$. We give $\mathcal{N}^2$ the product topology.

It turns out that $\widehat{\mathcal{O}}$ is in fact a base for $\mathcal{N}$. This can be proved with induction: Consider an intersection between two elements $\mathcal{O}(s)$ and $\mathcal{O}(t)$ of $\widehat{\mathcal{O}}$. If $s = t \restriction n$ or $t = s \restriction n$ then $\mathcal{O}(s) \cap \mathcal{O}(t) = \mathcal{O}(s)$ or $= \mathcal{O}(t)$ respectively. If not, $\mathcal{O}(s) \cap \mathcal{O}(t) = \emptyset$. So any finite intersection of unions of elements in $\widehat{\mathcal{O}}$ is still a union of elements in $\widehat{\mathcal{O}}$. And obviously the same holds for unions, so by induction $\widehat{\mathcal{O}}$ generates the topology. And since Seq is countable, every such union is equal to an at most countable union. This gives us:

Proposition 5. *Every open subset of $\mathcal{N}$ is a countable union of elements in $\widehat{\mathcal{O}}$.*

A subset A of a topological space Ω is **dense** in Ω if every non-empty open set intersects A. A space is **separable** if it has a countable dense subset. A **Polish space** is a topological space which is homeomorphic to a complete separable metric space.

A **σ-algebra** over a set X is a non-empty family of subsets of X which is closed under complementation and countable unions. Note that a σ-algebra will also be closed under countable intersections. In a Polish space P, the **Borel sets**, denoted $\mathscr{B}$, are the σ-algebra generated by the open sets of P.

Proposition 6. *There exists an open set $U \subset \mathcal{N}^2$ such for every open set $\mathcal{O} \subset \mathcal{N}$ there is some sequence $s \in \mathcal{N}$ such that*

$$\mathcal{O} = \{ x \mid (s, x) \in U \}.$$

*We call U a **universal** open $\mathcal{N}$-set.*

Proof. Let $\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \ldots$ be an enumeration of the elements in $\widehat{O}$, and $\mathcal{O}_0 = \emptyset$. We let $U \subset \mathcal{N}^2$ be defined by

$$(s, x) \in U \text{ iff } x \in \mathcal{O}_n \text{ for some } n \in s.$$

U is universal since by Proposition 5, for any open set A there is a sequence $N = \{n_i\}_{\mathbb{N}}$ of natural numbers such that $A = \bigcup_N \mathcal{O}_{n_i}$. Then $A = \{x \mid (N, x) \in U\}$.
To see that U is open, we note that we can divide U into a union of subsets of the form $H_n = \{(s, x) \mid x \in \mathcal{O}_{s_n}\}$. Every H_n can be further divided into a union of sets of the form $G_{n,i} = \{(s, x) \mid s_n = i \wedge x \in \mathcal{O}_i\}$. The subset $A_{n,i} = \{s \in \mathcal{N} \mid s_n = i\}$ of $\mathcal{N}$ is a union of basic open sets and therefore open. But $G_{n,i} = \pi_1^{-1} A_{n,i} \cap \pi_2^{-1} \mathcal{O}_i$ and thus is also open. Therefore U, which is a union of sets $G_{n,i}$, is open. $\square$

An **analytic set** is a subset A of a Polish space P such that A is the image of a continuous function $f : \mathcal{N} \to P$.

Lemma 7. *For every $n \geq 1$, $\mathcal{N}^n$ is homeomorphic to $\mathcal{N}$.*

Proof. Let $h : \mathcal{N}^2 \to \mathcal{N}$ riffle the sequences:

$$h : (\{s_i\}_{\mathbb{N}}, \{t_i\}_{\mathbb{N}}) \mapsto \{s_0, t_0, s_1, t_1, s_2, t_2, \ldots\}$$

Clearly h is a bijection. Now let $A \subset \mathcal{N}$ be open so that for some subset $S \subset \mathrm{Seq}$,

$$h^{-1}A = h^{-1}\left(\bigcup_{s \in S} \mathcal{O}(s)\right) = \bigcup_S h^{-1}\mathcal{O}(s).$$

If s has an odd number of entries, $s = \{n_0, n_1, \ldots, n_{2j}\}$, then $\mathcal{O}(s)$ is the set of all $\bar{s} \in \mathcal{N}$ starting with s. By "de-riffling",

$$\begin{aligned} h^{-1}(\mathcal{O}(s)) &= \{(\bar{s}, \bar{t}) \mid \bar{s} \in \mathcal{O}\{n_0, n_2, \ldots, n_{2j}\} \wedge \bar{t} \in \mathcal{O}\{n_1, n_3, \ldots, n_{2j-1}\}\} \\ &= \pi_1^{-1}(\mathcal{O}\{n_0, n_2, \ldots, n_{2j}\}) \cap \pi_2^{-1}(\mathcal{O}\{n_1, n_3, \ldots, n_{2j-1}\}) \end{aligned}$$

which is a finite intersection of open sets and thus open. Similarly we can prove that $h^{-1}(\mathcal{O}(s))$ is open if s has length one, or another odd number of entries. Thus $h^{-1}A$ is open, since it is a union of open sets, and h is continuous.

By Proposition 3, to show that h^{-1} is continuous it is enough to show that $\pi_1 h^{-1}$ and $\pi_2 h^{-1}$ are continuous. Once again let $A \in \mathcal{N}$ be open. Then

$$\left(\pi_1 h^{-1}\right)^{-1} A = \bigcup_{s \in S} h\, \pi_1^{-1}\, \mathcal{O}(s)$$

for some subset $S \subset \mathrm{Seq}$. Let $s = \{s_i\}_{i=0}^{n}$ and denote by $\mathcal{O}(k,i)$ the subset of $\mathcal{N}$ of all sequences such that their k:th entry is i. $\mathcal{O}(k,i)$ is open since it is a union of basic open sets. We note that

$$\begin{aligned} h\, \pi_1^{-1}\, \mathcal{O}(s) &= \{\{s_0, x_0, s_1, x_1, \ldots, s_n, x_n, x_{n+1}, x_{n+2}, \ldots\} \mid x_{i\mathbb{N}} \in \mathcal{N}\} \\ &= \mathcal{O}(0, s_0) \cap \mathcal{O}(2, s_1) \cap \mathcal{O}(4, s_2) \cap \ldots \cap \mathcal{O}(2n, s_n) \end{aligned}$$

which is a finite intersection of open sets and therefore open. So $\pi_1 h^{-1}$ is continuous, and the continuity of $\pi_2 h^{-1}$ is proved in the same way. Thus h is a homeomorphism $\mathcal{N} \to \mathcal{N}^2$.

Assume that there is a homeomorphism $h_n : \mathcal{N}^n \to \mathcal{N}^{n-1}$, for some $n \geq 2$. Then the function

$$\begin{aligned} h_{n+1} :& \mathcal{N}^{n+1} \to \mathcal{N}^n \\ & (y, \vec{x}) \mapsto (y, h_n(\vec{x})) \end{aligned}$$

is bijective, and a homeomorphism by Proposition 3. Thus by induction there is a homeomorphism from every $\mathcal{N}^n$ to $\mathcal{N}$. □

We shall subsequently need the following lemma, due to Jech [4, Lemma 11.6], which states

Lemma 8 (Jech)**.** *The following are equivalent, for any set A in a Polish space X:*

(i) *A is the continuous image of $\mathcal{N}$.*
(ii) *A is the continuous image of a Borel set B (in some Polish space Y).*
(iii) *A is the projection of a Borel set in $X \times Y$, for some Polish space Y.*
(iv) *A is the projection of a closed set in $X \times \mathcal{N}$.*

Theorem 9. *There exists a universal analytic $\mathcal{N}$-set. I.e. an analytic set $U \in \mathcal{N}^2$ such that for every analytic space $A \subseteq \mathcal{N}$ there is an $s \in \mathcal{N}$ such*

that
$$A = \{x \mid (s,x) \in U\}.$$

Proof. By Lemma 7 there is a homeomorphism $h : \mathcal{N}^2 \to \mathcal{N}$. Let V be a universal open set in $\mathcal{N}$. We construct our universal analytic set U by defining

$$(s,x) \in U \quad \text{iff} \quad \exists a \in \mathcal{N} \quad \text{such that} \quad (s, h(a,x)) \in V^{\complement}.$$

First we need to show that U is analytic. Consider the function

$$\begin{aligned} f: \quad \mathcal{N}^3 &\to \mathcal{N}^2 \\ (s,a,x) &\mapsto (s, h(a,x)). \end{aligned}$$

$\pi_1 f = \pi_1$ and $\pi_2 f = h\,\pi_{2,3}$, both of which are continuous. Hence, by Proposition 3 f is continuous.
Since $V^{\complement}$ is closed and f is continuous the preimage

$$f^{-1}(V^{\complement}) = \{(s,a,x) \mid (s, h(a,x)) \in V^{\complement}\}$$

is closed, and by Lemma 8 the projection $\pi_{1,3}\, f^{-1}(V^{\complement}) = U$ is analytic.
Secondly we need to show that U is universal. Let A be an analytic $\mathcal{N}$-set. Once again, by Lemma 8, A is the projection of a closed set B in $\mathcal{N}^2$, so that
$$x \in A \quad \Longleftrightarrow \quad (s,x) \in B \text{ for some } s \in \mathcal{N}.$$
Let $C = h(B)^{\complement}$. C is open since $h(B)$ is closed. Therefore we can use the universal open set V and say that there is an element $u \in \mathcal{N}$ such that $C = \{v \mid (u,v) \in V\}$. Then, with this u:

$$\begin{aligned} x \in A &\Leftrightarrow \exists s \in \mathcal{N}\ (s,x) \in B \Leftrightarrow \exists s \in \mathcal{N}\ h(s,x) \in h(B) \\ &\Leftrightarrow \exists s \in \mathcal{N}\ h(s,x) \notin C \Leftrightarrow \exists s \in \mathcal{N}\ (u, h(s,x)) \notin V \Leftrightarrow (u,x) \in U. \end{aligned}$$

I.e. U is a universal analytic $\mathcal{N}$-set in $\mathcal{N}^2$. □

Lemma 10. *For every Cartesian product $\mathcal{N}^n$ the* ***diagonal*** $\operatorname{diag}(\mathcal{N}^n) = \{(x,\ldots,x) \in \mathcal{N}^n \mid x \in \mathcal{N}\}$ *is closed.*

Proof. If $n = 1$ the diagonal is the entire space and therefore closed. If

$n > 2$ then for every point $p = (x_1, \ldots, x_n) \in \mathcal{N}^n$ not on the diagonal, let k be the first index at which the coordinates of p do not all have the same integer. For $1 \le i \le n$ let s_i be the partial sequence of x_i consisting of its first k entries.
$\mathcal{O}(s_i)$ is an open set in $\mathcal{N}$, and therefore in the product topology the cylinder $\pi_i^{-1}\,\mathcal{O}(s_i) \in \mathcal{N}^n$ is open. Hence the "k-cell"

$$C_p = \bigcap_{i=1}^{n} \pi_i^{-1}\,\mathcal{O}(s_i)$$

is open. Note that every point in C_p has a pair of coordinates that differ on their k:th entry; therefore C_p does not intersect the diagonal. Also note that $p \in C_p$.
Now consider the union

$$\bigcup_{p \in (\operatorname{diag} \mathcal{N}^n)^{\complement}} C_p.$$

It is a union of open sets and therefore open, and it contains every point of $\mathcal{N}^n$ except the diagonal. Thus its complement, the diagonal, is closed. □

Lemma 11. *There is a subset $A \subset \mathcal{N}$ which is analytic, but not the complement of an analytic set.*

Proof. Let $U \subset \mathcal{N}^2$ be a universal analytic $\mathcal{N}$-set, and let

$$A = \{x \mid (x, x) \in U\}.$$

Since U is analytic, by Lemma 8 it is the projection of a closed set $X \subset \mathcal{N}^3$. $\operatorname{diag}(\mathcal{N}^2)$ is closed by Lemma 10 and therefore the cylinder $Y = \operatorname{diag}(\mathcal{N}^2) \times \mathcal{N}$ is closed. Thus $X \cap Y$ is closed, which means that the projection

$$\pi_{1,2}(X \cap Y) = U \cap \operatorname{diag}(\mathcal{N}^2) = A$$

is analytic.
To see that A is not the complement of an analytic set, suppose it was; $A = B^{\complement}$ where B is analytic. Then there is an $s \in \mathcal{N}$ such that $B = \{x \mid (s, x) \in U\}$. If $s \in B$ then $(s, s) \in U$ meaning that $s \in A = B^{\complement}$. If $s \notin B$ then $s \in A$ which means that $(s, s) \in U$ and thus $u \in B$. In both cases we get a contradiction. Therefore A cannot be the complement of an analytic set. □

Proposition 12. *There is a subset $A \subset \mathcal{N}$ which is analytic but not Borel.*

Proof. By Lemma 8 every Borel subset of $\mathcal{N}$ is analytic, since the identity function is continuous. Therefore, if every analytic subset of $\mathcal{N}$ was Borel, the Borel and the analytic subsets of $\mathcal{N}$ would be precisely the same sets. But the Borel sets are closed under complementation, so this would contradict Lemma 11. □

Proposition 13. *If X and Y are topological spaces and $f : X \to Y$ is continuous, then the preimage of every Borel set in Y is Borel.*

Proof. Let $\mathcal{S} = \{S \in Y \mid f^{-1}(S) \in \mathscr{B}\}$. Since f is continuous $\mathcal{S}$ contains every open set. If $S \in \mathcal{S}$ then $f^{-1}(S^{\complement}) = f^{-1}(S)^{\complement} \in \mathscr{B}$, and if $\{S_i\}_I$ is a countable sequence of elements in $\mathcal{S}$ then $f^{-1} \bigcup_I S_i = \bigcup_I f^{-1}(S_i) \in \mathscr{B}$. So $\mathcal{S}$ contains the open sets of Y, and is closed under complementation and countable unions, which means that it contains the Borel sets of Y. □

$\mathcal{N}$ is homeomorphic to the irrationals $\mathbb{P}$, and a well-known example of a homeomorphism is the function mapping the sequence $\{s_i\}_{\mathbb{N}}$ of natural numbers to continued fractions:

$$\begin{aligned} h : \quad \mathcal{N} &\to (0,1) \\ \{s_i\}_{\mathbb{N}} &\mapsto \frac{1}{s_0 + \frac{1}{s_1 + \ddots}} \end{aligned} \tag{3.3}$$

This can then be composed with a homeomorphism to all of $\mathbb{R}$, and homeomorphy of $\mathcal{N}$ and $\mathbb{P}$ follows. However, to avoid having to dig into the theory of continued fractions we shall prove the existence of another homeomorphism, due to Miller [7, Theorem 1.1]. The construction is repeated here, only in more detail.

Lemma 14. *If $\{I_n\}_{\mathbb{N}}$ is a sequence of non-empty intervals in $\mathbb{R}$ such that their lengths converge to 0 and for every closure $\bar{I}_{n+1} \subset I_n$, then $\bigcap_{\mathbb{N}} I_n$ is a singleton.*

Proof. Let each $I_n = (a_n, b_n)$. Let A be the set of all a_n. A is non-empty and bounded above by every b_n. Therefore, if we let $x = \sup A$ then $x \leq b_n$ for every n. So $a_n \leq x \leq b_n$ for every n. So x is in every closure $\bar{I}_n$. But $\bar{I}_n \subset I_{n+1}$ so $x \in I_n$ for every n. Hence their intersection is non-empty.

There cannot be two points in the intersection, since for every pair of distinct points there is an I_n shorter than the distance between them. □

Theorem 15 (Miller). *$\mathcal{N}$ is homeomorphic to the irrationals $\mathbb{P}$ (under the subspace topology).*

Proof. For the purpose of this proof, let a **semi-partitioning** of an open interval (a, b) be a sequence of intervals $\{(a_i, b_i)\}_{\mathbb{Z}}$ such that for every i, $b_i = a_{i+1}$ and the closure of the union the sequence is the closed interval $[a, b]$. I.e. a semi-partitioning of (a, b) is a division of (a, b) into countably infinitely many disjoint open subintervals that lie shoulder to shoulder, and except for their endpoints cover all of (a, b). A semi-partitioning of a union of open intervals is a union of semi-partitionings of every interval.
If s is a finite sequence, let $s\hat{\ }n$ denote the sequence of s with n appended to it. The set of all finite sequences can be considered an infinite tree (with $\emptyset$ as its root), where each $s\hat{\ }n$ branches off from s. Each infinite branch then corresponds to an element of $\mathcal{N}$.
The idea is to construct a sequence of successive semi-partitionings of the real line such that the lengths of the intervals tend to zero, and every rational is the end point of some interval. Each infinite sequence of subintervals will correspond to an element of $\bar{s} \in \mathcal{N}$, and the intersections of each such sequence will be a singleton whose element will be the homeomorphisms value at $\bar{s}$.

Now for the details. Let $\{z_i\}_{\mathbb{N}}$ be an enumeration of the integers, $\{q_i\}_{\mathbb{N}}$ an enumeration of the rationals, and let $\mathrm{Seq}_n \subset \mathrm{Seq}$ be the sequences of length n. Let $I_\emptyset = \mathbb{R}$. For each $s \in \mathrm{Seq}_1$ let $I_s = (z_{s_0}, z_{s_0} + 1)$. This semi-partitions $\mathbb{R}$ into unit intervals with integer endpoints. From here, recursively define intervals I_s for each Seq_n as follows:

1. Semi-partition every I_s, $s \in \mathrm{Seq}_n$, in the following way: Let $I' \subset I_s$ be the open interval with centre in the middle of I_s and half the radius. To the left of I', define a new open interval stretching from I' to half to the remaining length of I_s. To the left of that new interval, define a new open interval of half the remaining length. Etc. ad infinitum. Do the same thing to the right of I'. All these open subintervals form a semi-partitioning of I_s. Note that all end points are rational.

2. q_n is either an endpoint of a previously constructed interval, or it sits in the interior of exactly one of the semi-partitions constructed above.

In the latter case, let q_n divide the corresponding semi-partition into two new open subintervals.

3. Now every I_s, $s \in \mathrm{Seq}_n$, has been semi-partitioned, and every rational up to and including q_n is the end point of one these semi-partitions, or a semi-partition earlier in the process. For each $s \in \mathrm{Seq}_n$ let $\{I'_i\}_{\mathbb{N}}$ be an enumeration of the semi-partitioning of I_s, and let $I_{s\hat{\ }i} = I'_i$. (Note that $s\hat{\ }i \in \mathrm{Seq}_{n+1}$. This is how the recursion continues.)

By construction, at every step the diameters of the intervals are at least halved. Also, since every created interval has non-zero distances to the endpoints of its superset, the closure $\bar{I}_{s\hat{\ }i} \subset I_s$ for every i and s. Thus by Lemma 14, for every $\bar{s} \in \mathcal{N}$

$$\bigcap_{n\in\mathbb{N}} I_{\bar{s}\restriction n} \tag{3.4}$$

is a singleton $\{x\}$. Let h be the function $\bar{s} \mapsto x$. x must be irrational since every rational is the end point of some open interval $I_{\bar{s}\restriction n}$, and so cannot be in any of the unions (3.4).

h is a bijection onto the irrationals since every irrational p is in $I_\emptyset$, and if $p \in I_s$ then p is in exactly one of the subsets $I_{s\hat{\ }i}$. h is continuous since if $A \subset \mathbb{P}$ is open, $A = A' \cap \mathbb{P}$, where A' is open in $\mathbb{R}$. This means that A' is a union of open intervals with rational endpoints. A' is semi-partitioned by a set of intervals $\{I_s\}_{s\in\mathbf{S}}$, $\mathbf{S} \subset \mathrm{Seq}$, so that

$$h^{-1}(A) = h^{-1}\left(\mathbb{P} \cap \bigcup_{s\in\mathbf{S}} I_s\right) = \bigcup_{s\in\mathbf{S}} h^{-1} I_s = \bigcup_{s\in\mathbf{S}} \mathcal{O}(s)$$

which is open. Similarly we can show that h is open: If $B \subset \mathcal{N}$ is open then by Proposition 5, $B = \bigcup_{\mathbf{S}} \mathcal{O}_s$ for some subset $\mathbf{S} \subset \mathrm{Seq}$. And then

$$h(B) = h\left(\bigcup_{\mathbf{S}} \mathcal{O}(s)\right) = \bigcup_{\mathbf{S}} h\left(\mathcal{O}(s)\right) = \bigcup_{\mathbf{S}} \mathbb{P} \cap I_s = \mathbb{P} \cap \bigcup_{\mathbf{S}} I_s$$

which is open in the subspace topology of $\mathbb{P}$. □

Proposition 16. *There is a subset of* $\mathbb{R}$ *which is analytic but not Borel.*

Proof. By Theorem 15 there is a homeomorphism $h : \mathcal{N} \to \mathbb{P}$, and by Proposition 12 there is an analytic non-Borel subset $A \subset \mathcal{N}$. Note that a homeomorphism $X \to Y \subset Z$ is a continuous function $X \to Z$. Therefore h is a continuous function $\mathcal{N} \to \mathbb{R}$ and thus, by Proposition 13 the image $C = h(A)$ is not Borel in $\mathbb{R}$.
To see that C is analytic, we note that since A is analytic there is a continuous function $f : \mathcal{N} \to \mathcal{N}$ whose image is A, and that thus the composition $h \circ f : \mathcal{N} \to \mathbb{R}$ is a continuous function with image C $\square$

4. Quantifier elimination and decision methods

Having introduced the necessary topology, we have now reached the point where we are ready to investigate decidability and quantifier elimination of $\mathbb{R}$. We start with two examples of how theorems about quantifier elimination in Hodges [3] may be generalised to infinitary languages.

Lemma 17. *Let T be a theory of $L_{\alpha\beta}$, and let Φ be a set formulas of $L_{\alpha\beta}$ such that:*

(a) Every atomic formula of L is in Φ.

(b) Φ is closed under boolean combinations.

(c) For every formula $\phi(X \cup Y) \in \Phi$, X and Y disjoint, $\exists Y \phi(X \cup Y)$ is equivalent modulo T to a formula $\psi(X) \in \Phi$.

Then every *formula is equivalent modulo T to a formula in Φ.*

Proof. By induction on complexity. If ϕ is a formula with complexity 0 then either ϕ is atomic, or it is $\top$ or $\bot$. In either case, by (a) and (b) it is in Φ.
Now let $\alpha > 0$ and assume that all formulas of complexity $< \alpha$ are $\leftrightarrow \pmod{T}$ to formulas in Φ. A formula ϕ of complexity α is either a boolean combination of formulas of lower complexity and thus $\leftrightarrow \pmod{T}$ to a formula in ϕ by (b), or of the form $\exists Y \psi$ or $\forall Y \psi$, where ψ has $< \alpha$. ψ is $\leftrightarrow \pmod{T}$ some formula $\pi \in \Phi$, so $\exists Y \phi \leftrightarrow \exists Y \pi \leftrightarrow \pmod{T}$ some formula in Φ by (c), and $\forall Y \phi \leftrightarrow \forall Y \pi \leftrightarrow \neg\exists\neg\pi \leftrightarrow \pmod{T}$ some formula in Φ by (b) and (c). $\square$

Given a language $L_{\alpha\beta}$ and its signature L, an **elimination set** for a class K of L-structures is a set Φ of $L_{\alpha\beta}$-formulas such that every formula $\phi(X)$ is equivalent in every structure in K to a boolean combination $\phi^*(X)$ of formulas in Φ. Since a boolean combination of some set of formulas Φ cannot have any other quantifiers than those already in the formulas of Φ, the process of finding the equivalent formula ϕ^* in the elimination set is called **quantifier elimination**. We say that a theory T **has quantifier elimination** if $\operatorname{Mod} T$ has a quantifier free elimination set. See Hodges [3, section 2.7].

Usually K will be some well-known class of structures such as the class of linear orderings, or the class of real-closed fields. Of course every class K always has the trivial elimination set consisting of every formula of $L_{\alpha\beta}$, but often the aim is to find elimination sets consisting of only very few formulas, or elimination sets consisting of particularly simple formulas. Such a "simple" elimination set tells us that when we work with some specific structure A in K we can, at least in principle, restrict ourselves the formulas of Φ. The truth of every $L_{\alpha\beta}$-statement we make about A depends only on the truth of the formulas in Φ.

Of particular interest are elimination sets without quantifiers (i.e. the formulas in the elimination sets are themselves boolean combinations of atomic formulas.) They are important because they tell us that for every $L_{\alpha\beta}$-formula $\phi(X_I)$, L-structure A and $S_I \subseteq \operatorname{dom} A$, our ability to determine the veracity of the statement $\phi[S_I]$ depends solely on our ability to determine the veracity of atomic formulas.

Since $\forall X\phi \leftrightarrow \neg\exists X\neg\phi$, Lemma 17 already hints at elimination sets for K, namely any set of generators (under boolean combinations) for Φ, with $T = \operatorname{Th}\mathbf{K}$. We shall now refine this idea lemma to arrive at very specific conditions that can tell us what an elimination set will look like:

Theorem 18. *Let K be a class of L-structures, $\alpha = \omega$ or $\alpha = \infty$, and Φ a set of $L_{\alpha\beta}$-formulas. If:*

(i) Every atomic formula of L is in Φ.

(ii) for every formula $\phi(X)$ of the form $\exists Y \bigwedge \Psi(X \cup Y)$, where $\Psi \subseteq \Phi \cup \neg\Phi$, there is a formula $\phi^*(X) \in \langle\Phi\rangle$ of $L_{\alpha\beta}$ such that $\phi \leftrightarrow \phi^*$ $(\mathrm{mod}\ \operatorname{Th}\mathbf{K})$.

Then Φ is an elimination set for K.

Proof. Let $\phi(X)$ be a formula of complexity $\alpha > 0$ above Φ, and assume every formula $\psi(X)$ of complexity $< \alpha$ above Φ is equivalent to some formula $\psi'(X) \in \langle\Phi\rangle$ in every structure in K (remember that $\langle\Phi\rangle$ is the set of Boolean combinations of elements in Φ). $\phi(X)$ is on one of the following forms:

(a) $\neg\psi$, $\bigvee \Psi$ or $\bigwedge \Psi$.

(b) $\exists Y \psi(X \cup Y)$.

(c) $\forall Y \psi(X \cup Y)$.

Here ψ, π and every formula in Ψ have complexities $< \alpha$, and X and Y are disjoint. If it is (a), then by the induction hypothesis $\psi \in \langle\Phi\rangle$.
If it is (b), then by the induction hypothesis $\psi(X \cup Y)$ is equivalent to some boolean combination of formulas in Φ, which by Theorem 2 is equivalent to a formula on disjunctive normal form over Φ, so that for some $\Psi(X \cup Y) \subseteq \Phi$

$$\exists Y \psi(X \cup Y) \longleftrightarrow \exists Y \bigvee \bigwedge \Psi \longleftrightarrow \bigvee \exists Y \bigwedge \Psi.$$

by (ii) the last formula is equivalent to $\bigvee \theta^*(X \cup Y)$ where $\theta^* \in \langle\Phi\rangle$. But $\bigvee \theta^* \leftrightarrow \theta^*$, which is the formula we are looking for. If it is (c), then we use $\forall Y \psi \leftrightarrow \neg\exists Y \neg\psi$ and then proceed in the same way as for (b).
Every formula of complexity 0 is already in Φ, so by transfinite induction every formula is equivalent to a formula in $\langle\Phi\rangle$ modulo $\mathrm{Th}\,\mathbf{K}$. $\square$

A notion related to quantifier elimination is decidability. A sentence ϕ is a **consequence** of a theory T if $\phi \in \mathrm{Th}\,\mathrm{Mod}\,T$, i.e. if ϕ is true in every model of T (this definition of consequence is equivalent to that of Hodges [3].) A theory T of a language $L_{\alpha\beta}$ is **decidable** if there is a terminating algorithm which determines whether any given sentence of $L_{\alpha\beta}$ is a consequence of T. Note that decidability implies completeness.

Decidability is related to quantifier elimination both spiritually and practically: Spiritually, decidability of a theory T means that every sentence of $\mathrm{Mod}\,T$ can be reduced algorithmically to the "elimination set" $\{\top, \bot\}$. So decidability in a sense is a stronger form of quantifier elimination, but only on sentences. Practically, an algorithm for decidability may often begin with an algorithm for quantifier elimination. Se Tarski [8] for an example of this.

4.1 The real case

Let $\mathcal{R}$ be the signature $(0, 1, -, +, \cdot, <)$, and let $\mathbb{R}$ be the structure of the reals as an ordered field with signature $\mathcal{R}$ (used of course interchangeably with $\mathbb{R}$ as simply the set of reals). The **real-closed fields**, abbreviated **RCF**, are the structures that are elementarily equivalent to $\mathbb{R}$, or in other words

$$A \in \mathrm{RCF} \iff \mathrm{Th}_{\mathcal{R}_{\omega\omega}} A = \mathrm{Th}_{\mathcal{R}_{\omega\omega}} \mathbb{R}.$$

If we wish to talk about the theory of $\mathbb{R}$ in other languages, we shall use the notation $\mathbb{R}_{\alpha\beta} = \mathrm{Th}_{\mathcal{R}_{\alpha\beta}} \mathbb{R}$. It turns out (see Hodges [3]) that RCF ($= \mathbb{R}_{\omega\omega}$) can be axiomatised by a countably infinite of formulas.

As famously shown by Tarski [8], the atomic formulas are an elimination set for the class of real-closed fields where $\mathcal{R}_{\omega\omega}$ has been exchanged for in the language $(0, 1, -1, +, \cdot, <)_{\omega\omega}$. This is only a cosmetic change since in every use of $-$ and -1 from the two signatures, we can simply switch between $(-1) \cdot a$ and $-a$ (or $0 - a$ if we consider $-$ a binary function.) So the theory of RCF has quantifier elimination. Furthermore, Tarski shows that the theory of RCF is decidable (this statement would of course suffice, since quantifier elimination follows from decidability). Since $\mathbb{R}$ is an RCF, we arrive at

Theorem 19 (Tarski)**.** *Let $\mathcal{R}$ be the signature of $\mathbb{R}$ as an ordered field, and $\mathbb{R}_{\omega\omega} = \mathrm{Th}_{\mathcal{R}_{\omega\omega}} \mathbb{R}$. Then $\mathbb{R}_{\omega\omega}$ has quantifier elimination, and is decidable.*

Remains the questions of whether $\mathbb{R}_{\alpha\beta}$ is decidable or has quantifier elimination when $\alpha, \beta \geq \omega_1$. Let us call $\mathbb{R}_{\alpha\beta}$, where $\alpha, \beta \geq \omega_1$, **infinitary RCF**. It is intuitively clear that the answer to the question of decidability is no. For in infinitary RCF with domain $\mathbb{R}$ we can express the set of integers, and thus decidability would imply an oracle for first-order statements about integers.

To be more precise about this idea, we will first construct a framework of formulas in $\mathcal{R}_{\omega\omega}$ and $\mathcal{R}_{\omega_1\omega_1}$ that let us talk about specific numbers and sets of numbers, and then use these to show that decidability of infinitary RCF implies a solution to Hilbert's tenth problem. We will demarcate the formulas with $(\alpha\beta)$ to show which language $\mathcal{R}_{\alpha\beta}$ they are part of (note that (00) means an atomic formula. First of all let us define the non-negative integers:

$$\begin{aligned} &\text{(00):} \quad \zeta_0(x) \equiv x = 0, \\ &\text{(00):} \quad \zeta_i(x) \equiv x = \underbrace{1 + \cdots + 1}_{i \text{ times}} \quad \text{for } i \geq 1. \end{aligned} \tag{4.1}$$

In $\mathbb{R}$, the interpretation is that ζ_n defines the natural number n, and in an infinitary language we can actually use *all* of these atomic formulas and thus construct formulas that talk about *all* natural numbers instead of only a finite subset of them. For example we can define the formulas

$$\begin{aligned}
&(\omega_1 0): && \nu(x) \equiv \bigvee_{\mathbb{N}} \zeta_i(x), \\
&(\omega_1 0): && \zeta(x) \equiv \nu(x) \vee \nu(-x), \\
&(\omega_1 \omega): && \koppa(x) \equiv \exists\{y,z\}\, \zeta(y) \wedge \zeta(z) \wedge y \neq 0 \wedge (x \cdot y = z).
\end{aligned} \tag{4.2}$$

$\nu[s]$ is true in $\mathbb{R}$ iff $s \in \mathbb{N}$, $\mathbb{R} \vDash \zeta[s]$ iff $s \in \mathbb{Z}$ and $\mathbb{R} \vDash \koppa[s]$ iff $s \in \mathbb{Q}$. So using ν, ζ and $\koppa$ ("qoppa") we can construct formulas that deal explicitly with the natural numbers, the integers and the rationals respectively. It will also be of use to be able to talk about specific integers and rationals, for example to construct specific Diophantine equations. Therefore let

$$\begin{aligned}
&(00): && \zeta_n(x) \equiv \zeta_{-n}(-x) \quad \text{for } n \leq -1, \\
&(\omega\omega): && \koppa_{a,b}(x) \equiv \exists\{y,z\} \zeta_a(y) \wedge \zeta_b(z) \wedge z \neq 0 \wedge (x \cdot z = y).
\end{aligned} \tag{4.3}$$

Note that because of Theorem 19, formulas (4.3) above are in fact equivalent to $\mathcal{R}_{\omega 0}$-formulas. $\zeta_n[s]$ is true in $\mathbb{R}$ iff s is the integer n, and $\mathbb{R} \vDash \koppa_{a,b}[s]$ iff s is the rational a/b where a and b are integers. If q is a rational we shall often be lazy and use the notation $\koppa_q$, taken to mean $\koppa_{a,b}$ where a, b is some pair of integers such that $q = a/b$.

It is interesting to note that to define a given rational, as in equations (4.3) above, $\mathcal{R}_{\omega\omega}$ suffices. But to define the corresponding sets, as in equations (4.2), we have used $\mathcal{R}_{\omega_1\omega}$. To define a specific real however, we immediately make use of $\mathcal{R}_{\omega_1\omega}$: For every real r choose a sequence $\{q_i^-\}_{i\in\mathbb{N}}$ and a sequence $\{q_i^+\}_{i\in\mathbb{N}}$ where q_i^- and q_i^+ are rationals such that all $q_i^- < r$, all $q_i^+ > r$ and $\lim q_i^- = \lim q_i^+ = r$. Then the formula

$$(\omega_1\omega): \quad \hat{\rho}_r(x) \equiv \bigwedge_{i\in\mathbb{N}} \Big(\exists\{y,z\}\ \koppa_{q_i^-}(y) \wedge \koppa_{q_i^+}(z) \wedge y < x \wedge x < z \Big) \tag{4.4}$$

defines r. That is, $\mathbb{R} \vDash \hat{\rho}_r[s]$ iff $s = r$. Note that the formula inside the conjunction above is an $\mathcal{R}_{\omega\omega}$-formula, and thus by Theorem 19 equivalent to a quantifier-free formula. Thus $\hat{\rho}_r$ is, in fact, equivalent to a formula of $\mathcal{R}_{\omega_1 0}$, and we can assert the following:

Proposition 20. *For every real r there is a formula*

$$\rho_r(x) \in \mathcal{R}_{\omega_1 0} \tag{4.5}$$

which defines r.

Using ν_n we can express solutions to Diophantine equations with rational coefficients. Every Diophantine equation with rational coefficients is equivalent to a Diophantine equation with integer coefficients, so we need only consider those. To express solutions to such an equation D, replace every integer coefficient n with the variable v_n and take integer powers to mean repeated multiplication. Then we get a corresponding atomic formula α_D. With N as the set of coefficients in D and X the set of variables, let

$$\begin{aligned} &{\scriptstyle (\omega\omega):} && \delta_D(X) \equiv \exists \{v_i\}_N \bigwedge_N \zeta_i(v_i) \wedge \alpha_D, \\ &{\scriptstyle (\omega_1\omega):} && \varsigma_D \equiv \exists X\, \delta_D(X) \wedge \bigwedge_{x \in X} \text{♀}(x). \end{aligned} \tag{4.6}$$

and note that $\mathbb{R} \vDash \delta_D[S]$ iff S is a solution to D, and $\mathbb{R} \vDash \varsigma_D$ iff D has a rational solution. Hilbert in his tenth problem proposed to seek an algorithm by which the solvability in rationals of any rational Diophantine equation could be decided (see [2]). Verbatim:

> 10. DETERMINATION OF THE SOLVABILITY OF A DIOPHANTINE EQUATION.
>
> Given a diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers.

This was proved impossible by Matiyasevich [6]. Therefore, since decidability of infinitary RCF implies an algorithm for deciding, for every Diophantine equation D, whether $\mathbb{R} \vDash \varsigma_D$ or not, infinitary RCF is not decidable.

Theorem 21 (Matiyasevich)**.** *$\mathbb{R}_{\alpha\beta}$ is not decidable for $\alpha \geq \omega_1$ and $\beta \geq \omega$.*

4.2 A Counterexample to quantifier elimination in $\mathbb{R}_{\omega_1\omega_1}$

As discussed in Section 4.1, Tarski [8] showed that $\mathbb{R}_{\omega\omega}$ has quantifier elimination, and Matiyasevich [6] implies that $\mathbb{R}_{\omega_1\omega_1}$ is not decidable (and therefore neither does $\mathbb{R}_{\alpha\beta}$ for higher ordinals α, β.) A natural question arises: For which $\alpha, \beta > \omega$, if any, does $\mathbb{R}_{\alpha\beta}$ also have quantifier elimination? Turns out, none. Our strategy for proving this will be to show that sets definable without quantifiers are Borel sets, and that there are sets definable with quantifiers in infinitary $\mathbb{R}$ that are not Borel sets. We will start by defining the Borel sets, following Jech [4]. Jech defines Borel sets in the more general topological setting of Polish spaces, but we will do it only for the reals (which is indeed a Polish space.) For this we shall need some topological concepts, but for simplicity we shall define these too only for $\mathbb{R}$. They will however agree with the ordinary more general definitions.

The topology of $\mathbb{R}$ is generated by its open balls, the open intervals.

Every open set in $\mathbb{R}$ is a countable union of open intervals in $\mathbb{R}$ (note that the empty set is an open set).

Lemma 22. *The Borel subsets of $\mathbb{R}$ are definable without quantifiers in $\mathbb{R}_{\omega_1\omega_1}$.*

Proof. First we show that the open sets of $\mathbb{R}$ are definable without quantifiers in $\mathbb{R}_{\omega_1\omega_1}$: The standard topology of $\mathbb{R}$ is generated by the open intervals of the forms (a, b), $(-\infty, b)$ and (a, ∞), where a and b are reals. Let Q be the set of all open intervals with rational end-points, and Q^* the set of all unions of intervals in Q.
Every open interval is the union of its subintervals of finite length with rational end-points, and is thus in Q^*. Therefore, the topology on $\mathbb{R}$ is also generated by Q^*. Since Q is closed under finite intersections, so is Q^*, and quite clearly Q^* is closed under unions. Hence, Q^* is the standard topology on $\mathbb{R}$.
Since $\mathbb{Q}$ is countable Q must be countable, and therefore every element in Q^* is equal to an (at most) countable union of elements in Q. So every open subset $S \subseteq R$ is either the empty set, which is defined by the formula $\alpha_\emptyset(x) \equiv \bot$, or of the form

$$\bigcup_{i \in I} (q_i, p_i)$$

where q_i and p_i are rationals and I is non-empty and at most countable. This set is defined by the quantifier-free $\mathbb{R}_{\omega_1\omega_1}$-formula

$$\alpha_S(x) \equiv \bigvee_{i\in I} q_i < x < p_i.$$

If $\{B_i\}_{i\in I}$ is an at most countable set of subsets of $\mathbb{R}$, and each B_i is defined by a quantifier-free formula $\psi_{B_i}(x)$ of $\mathbb{R}_{\omega_1\omega_1}$, then the union $\bigcup_{i\in I} B_i$ is defined by the quantifier-free formula $\bigvee_{i\in I}\psi_{B_i}(x)$ of $\mathbb{R}_{\omega_1\omega_1}$, and the complement of any B_i is defined by $\neg\psi_{B_i}(x)$ of $\mathbb{R}_{\omega_1\omega_1}$ which is also quantifier-free. So by induction every set in the σ-algebra generated by $\{B_i\}_I$ is defined by a quantifier-free formula, and by Definition 1 these formulas are of $\mathbb{R}_{\omega_1\omega_1}$. Thus, the Borel subsets of R are defined by quantifier-free formulas. □

Lemma 23. *Let A be a structure with signature L and a topology. Suppose every atomic formula with at most one free variable defines a Borel set. Then every quantifier-free formula in $L_{\omega_1\omega_1}$ with at most one free variable defines a Borel set.*

Proof. Let Φ_x be the atomic formulas in $L_{\omega_1\omega_1}$ with at most x as a free variable. The quantifier-free formulas with at most x are precisely $\langle\Phi_x\rangle$. Now

- Every atomic formula in Φ_x is in $\mathscr{B}$ per assumption.
- $\top$ defines A and $\bot$ defines $\emptyset$, both of which are in $\mathscr{B}$.
- If $\Psi = \{\psi_i(x)\}_I$ is a countable set of formulas where each ψ_i defines a set $B_i \in \mathscr{B}$ then
$$\bigwedge \Psi \text{ defines } \bigcap_I B_i$$
and
$$\bigvee \Psi \text{ defines } \bigcup_I B_i,$$
both of which are in $\mathscr{B}$.

Therefore, following Defintion 1, $\langle\Phi_x\rangle$ all define Borel sets. The same proof of course holds for every single variable in place of x. □

Theorem 24. *The Borel sets of $\mathbb{R}$ are precisely the sets defined by the quantifier-free formulas in $\mathbb{R}_{\omega_1\omega_1}$.*

Proof. By Lemma 22 every Borel set is defined by a quantifier-free formula. Every atomic formula is a polynomial equation or strict inequality in one variable. Polynomial equations in $\mathbb{R}$ have at a finite number of roots, and so define closed sets, which are Borel. Strict polynomial inequalities define unions of open intervals, which are Borel. Therefore Lemma 23 implies that quantifier-free formulas which define sets, define Borel sets. $\square$

Theorem 25. *Every analytic subset of $\mathbb{R}$ is definable in $\mathbb{R}_{\omega_1\omega_1}$.*

Proof. Let A be an analytic subset of $\mathbb{R}$. Then A is the image of a continuous function $f : \mathcal{N} \to \mathbb{R}$. The idea is that for a continuous function it is enough to know the values of a dense subset of the domain to be able to calculate the values of the entire domain. We will construct a formula ϕ of $\mathbb{R}_{\omega_1\omega_1}$ that defines A in terms of such a set.
For every $s \in \mathrm{Seq}$ choose a point $s^* \in \mathcal{O}(s)$. Let $\{p_{1,i}\}_{\mathbb{N}}$ be the sequence of every second prime, $\{p_{2,i}\}_{\mathbb{N}}$ every second of those which are left, $\{p_{3,i}\}_{\mathbb{N}}$ those which are left after $\{p_{1,i}\}_{\mathbb{N}}$ and $\{p_{2,i}\}_{\mathbb{N}}$, etc. Let $\mathcal{P}$ be the set of all products of at least one distinct prime. Define the encoding function

$$\begin{aligned} \mathrm{enc}: \quad \mathrm{Seq} &\to \mathcal{P} \\ \{s_i\}_{i=0}^n &\mapsto \Pi_{i=0}^n p_{i,s_i}. \end{aligned}$$

Note that enc is a bijection, and that $\mathrm{enc}(s)$ divides $\mathrm{enc}(t)$ iff t begins with the sequence s. Re-index the values $\{f(s^*)\}_{s\in\mathrm{Seq}}$ by letting $r_{\mathrm{enc}\,s} = f(s^*)$. We will need a way to access elements of indexed sets with formulas. Let

$$\iota(x, y) \equiv \bigvee_{i\in\mathcal{P}} \zeta_i(x) \wedge \rho_{r_i}(y),$$

where ζ_i and ρ_{r_i} are the formulas in (4.3). $\mathbb{R} \vDash \iota[i, r]$ iff $i \in \mathcal{P}$ and $r = r_i$. Let $\delta(x, y) \equiv x \neq y \wedge \exists z\, \zeta(z) \wedge x \cdot z = y$ and $\pi(x) \equiv \bigvee_{i\in\mathcal{P}} \zeta_i(x)$. $\mathbb{R} \vDash \delta(a, b)$ iff b is an integer multiple of a (i.e. $a|b$ if a and b are integers) and $a \neq b$.

$\mathbb{R} \vDash \pi[a]$ iff $a \in \mathcal{P}$. Finally, let

$$\alpha(x) \equiv \exists\{x_i\}_{\mathbb{N}} \left[\bigwedge_{\mathbb{N}} \pi(x_i)\right]_{(1)} \wedge \left[\bigwedge_{\mathbb{N}} \delta(x_i, x_{i+1})\right]_{(2)} \wedge$$
$$\neg\left[\exists z\ \pi(z) \wedge \left(\bigwedge_{\mathbb{N}} z \neq x_i\right) \wedge \bigvee_{\mathbb{N}} \big(\delta(x_i, z) \wedge \delta(z, x_{i+1})\big)\right]_{(3)} \wedge$$
$$\left[\forall\varepsilon\ \varepsilon > 0 \to \exists N \bigwedge_{\mathbb{N}} x_i > N \to \exists y\ \iota(x_i, y) \wedge \big((x-y)^2 < \varepsilon^2\big)\right]_{(4)}.$$

If t is a sequence, let $t \restriction n$ mean the subsequence of t consisting of the first n elements. (1) indicates that $\{x_i\}_{\mathbb{N}}$ is a subset of $\mathcal{P}$. (2) indicates that $\{x_i\}_{\mathbb{N}}$ that every x_i divides x_{i+1}. Paired with (1) this means that $\{x_i\}_{\mathbb{N}}$ is a subsequence of $\{\operatorname{enc}\bar{s} \restriction n\}_{n\in\mathbb{N}}$ for some $\bar{s} \in \mathcal{N}$. (3) indicates that there are no other elements of $\mathcal{P}$ in-between the elements of $\{x_i\}_{\mathbb{N}}$, so that together with (1) and (2) $\{x_i\}_{\mathbb{N}} = \{\operatorname{enc}\bar{s} \restriction n\}_{n\in\mathbb{N}}$. (4) Uses the epsilon-delta definition of limits to say that x is the limit of the sequence indexed by $\{x_i\}_{\mathbb{N}}$. So $\alpha(x)$ says that x is the limit of the subsequence $\{r_i\}_{\{x_i\}_{\mathbb{N}}}$ of $\{r_i\}_{\mathcal{P}}$, i.e. that for some $\bar{s} \in \mathcal{N}$

$$x = \lim_{n\to\infty} r_{\operatorname{enc}(\bar{s}\restriction n)} = \lim_{n\to\infty} f\left((\bar{s} \restriction n)^*\right). \tag{4.7}$$

By Proposition 5 every open set A containing $\bar{s}$ is a union of elements in $\widehat{\mathcal{O}}$. So there is an $\mathcal{O}(t) \subseteq A$, $t \in \mathrm{Seq}$, such that $\bar{s} \in \mathcal{O}(t)$. But this means that $t = \bar{s} \restriction N$ for some N, which in turn means that if $N < n$ then $(\bar{s} \restriction n)^* \in \mathcal{O}(t) \subseteq A$. Thus

$$\lim_{n\to\infty} (\bar{s} \restriction n)^* = \bar{s}. \tag{4.8}$$

Equations (4.7) and (4.8) together with Proposition 4 which says that f preserves limits, then tell us that $f(\bar{s}) = x$. So

$$\mathbb{R} \vDash \alpha[r] \text{ iff } r \in \operatorname{im} f = A. \qquad \square$$

Using Proposition 16 and Theorems 24 and 25, we draw the conclusion that there is a formula $\alpha(x)$ of $\mathbb{R}_{\omega_1\omega_1}$, which defines an analytic set but not a Borel set, and therefore is not logically equivalent to a quantifier-free formula. Which in turn allows us to draw this thesis final conclusion:

Theorem 26. *Let $\mathcal{R}$ be the signature of $\mathbb{R}$ as an ordered field, and $\mathbb{R}_{\omega_1\omega_1} = \mathrm{Th}_{\mathcal{R}_{\omega_1\omega_1}} \mathbb{R}$. Then $\mathbb{R}_{\omega_1\omega_1}$ does not have quantifier elimination.*

4.3 Beyond countability

So what about languages of other cardinalities? Let us go about it systematically and fill out table (4.1). We assume the continuum hypothesis so that we can limit the cardinalities we examine. The subject matter is the existence of quantifier elimination and decidability, and the variables are α and β in $\mathbb{R}_{\alpha\beta}$. Remember that $|\alpha|$ is the maximum cardinality of the set of formulas that a conjunction or a disjunction can act upon, and $|\beta|$ is the maximum cardinality of the set of variables that a quantifier binds.

$\alpha\backslash\beta$	0		ω		ω_1		ω_2	
0	Q.E	dec.	?	dec.				
ω	Q.E	dec.	Q.E	dec.				
ω_1	Q.E	?	?	~~dec.~~	~~Q.E~~	~~dec.~~		
ω_2	Q.E	?	?	~~dec.~~	?	~~dec~~	Sub.	~~dec.~~

Table 4.1: Q.E. and decidability of $\mathbb{R}_{\alpha\beta}$.

If $0 < \beta < \omega$ then even though each quantifier is limited to some finite number of variables, we can circumvent this limit by stringing multiple quantifiers together. Thus every formula of $\mathbb{R}_{\alpha\omega}$ is equivalent to a formula of $\mathbb{R}_{\alpha\beta}$, and the reverse is clear. So we need only consider $\mathbb{R}_{\alpha\omega}$. Similarly we can omit the cases $0 < \alpha < \omega$.

If $|\omega| < |\alpha|$ and $\alpha < \beta$, then even though every quantifier can bind $|\beta|$ variables, α still limits how many variables can occur in a formula. So the formulas of $\mathbb{R}_{\alpha\beta}$ are equivalent to formulas of $\mathbb{R}_{\alpha\alpha}$. And since $\mathbb{R}_{\alpha\alpha} \subset \mathbb{R}_{\alpha\beta}$ we need not consider the theories where $\alpha < \beta$. Having decided on which theories to consider, we can start examining them:

- If $\beta = 0$ then every formula is already quantifier free.
- If $\mathbb{R}_{\alpha\beta}$ is decidable, then so is $\mathbb{R}_{\gamma\delta}$ for every $\gamma \le \alpha$ and $\delta \le \beta$. By Theorem 19, $\mathbb{R}_{\omega\omega}$ is decidable.

- The contrapositive of this is of course that if $\mathbb{R}_{\alpha\beta}$ is not decidable, then neither is $\mathbb{R}_{\gamma\delta}$ where $\gamma \geq \alpha$ and $\delta \geq \beta$. By Theorem 21, $\mathbb{R}_{\omega_1\omega}$ is not decidable.

- By Theorem 26, $\mathbb{R}_{\omega_1\omega_1}$ does not have quantifier elimination.

What about quantifier elimination for $\mathbb{R}_{\omega_2\omega_2}$? In $\mathbb{R}_{\omega_2\omega_2}$ every set is definable without quantifiers (using ρ_r of formula (4.5)). This lets us construct the following formulas: Let $\phi(X_I)$ be a formula of $\mathbb{R}_{\alpha\beta}$ where X_I is precisely the set of free variables of ϕ. Let $\mathbf{S}_J$ be the set of all I-indexed subsets S_{jI} of $\mathbb{R}$ such that $\mathbb{R} \vDash \phi[S_{jI}]$. Then

$$\phi(X_I) \leftrightarrow \bigvee_J \bigwedge_I \rho_{s_{ji}}(x_i) \pmod{\mathbb{R}}. \tag{4.9}$$

If we want every formula of the form of the LHS above to be a formula of $\mathbb{R}_{\gamma\delta}$, what do the cardinalities of γ and δ need to be? Well, $\rho_{s_{ji}}$ is quantifier free, so $\delta = 0$. γ needs to allow for connectives of cardinalities $|J|$, $|I|$ and $|\omega|$. $|\omega|$ immediately implies that $\gamma \geq \omega_1$. Depending on ϕ, J can be as big as $\mathbb{R}^I$ since $\mathbf{S}_J$ is a set of indexed subsets of $\mathbb{R}$ which have cardinalities $|I|$. But $|\mathbb{R}|^{|I|} \geq 2^{|I|}$ which is the cardinality of the power set of I and strictly greater than the cardinality if I itself. Therefore $|J|$ could be strictly greater than $|I|$, which is limited by α. Thus $|\gamma| > |\alpha|$.

So we cannot use equation (4.9) to eliminate quantifiers in arbitrary formulas of a language. But we limit the number of free variables in a formula, we *can* use it to perform quantifier elimination on a subset of the language. If we allow no more than countably many free variables then $|I| = |\mathbb{N}|$, and thus $|J| = |\mathbb{R}|^{|\mathbb{N}|} = |\mathbb{R}|$ at most. Then it suffices with $\gamma = \omega_2$. Hence:

- The subset of $\mathbb{R}_{\omega_2\beta}$ where the sets of free variables are countable, has quantifier elimination.

Bibliography

[1] Kurt Gödel. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatshefte für Mathematik und Physik* 38.1 (1931), pp. 173–198.

[2] David Hilbert. "Mathematical problems". In: *Bulletin of the American Mathematical Society* 8.10 (1902), pp. 437–479.

[3] Wilfrid Hodges. *Model theory.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, Mar. 1993.

[4] Thomas J. Jech. *Set theory: Millennium edition.* Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Oct. 2002.

[5] H. J. Keisler and C. C. Chang. *Model Theory.* Amsterdam and New York: North Holland and American Elsevier, 1973.

[6] Yuri V. Matiyasevich. *Hilbert's Tenth Problem.* Foundations of Computing. Cambridge, Massachusetts: MIT Press, 1993.

[7] Arnold W. Miller. *Descriptive set theory and forcing: How to prove Theorems about Borel sets the hard way.* Vol. 4. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Sept. 1995.

[8] Alfred Tarski. "A Decision Method for Elementary Algebra and Geometry". In: *Quantifier Elimination and Cylindrical Algebraic Decomposition.* Ed. by Bob F. Caviness and Jeremy R. Johnson. Wien: Springer-Verlag, Jan. 1998, pp. 24–84.

[9] Stefan Waldmann. *Topology: An Introduction.* Springer International Publishing, 2014.

Nomenclature

dom A	Domain of A . . . 3
L	signature . . . 4
$\bigwedge$	setwise conjunction . . . 5
$\bigvee$	setwise disjunction . . . 5
R	relation . . . 5
$\equiv$	Syntactic abbreviation . . . 6
$\langle\Phi\rangle$	The set of boolean combinations of the formulas in Φ . . . 7
$L_{\alpha\beta}$	Language. α is a limit on the cardinalities of con- and disjunctions. β is a limit on the number of variables in quantifiers. . . . 8
ι_A	Interpretation function to a structure A . . . 10
$t[S_I]$	Interpretation of term t of variables S_I . . . 11
Mod	Models of a theory . . . 12
Th	Theory of a structure . . . 12
$\leftrightarrow$	Logically equivalent . . . 12
mod	Modulo a theory . . . 12
π_i	Projection on i:th coordinate. 15
$B_r(a)$	Open ball of radius r around a 16
$\mathcal{O}(s)$	Subset of $\mathcal{N}$ beginning with s 17
$\mathcal{N}$	Baire space, $\mathbb{N}^{\mathbb{N}}$ . . . 17
σ-algebra	Sets closed under complementation and ctble unions. . . . 17
$\widehat{\mathcal{O}}$	Set of all $\mathcal{O}(s)$-sets . . . 17
$s \restriction n$	First n entries of s . . . 17
$\mathscr{B}$	Borel sets . . . 17
diag	Diagonal . . . 20
$\mathcal{R}$	Signature of the real-closed fields 29
$\mathbb{R}_{\alpha\beta}$	$\mathrm{Th}_{\mathcal{R}_{\alpha\beta}}\,\mathbb{R}$ . . . 29

Index