# FIRST ORDER LOGIC AND GÖDEL INCOMPLETENESS

ANUSH TSERUNYAN

## Contents

These notes owe a great deal to [Mos08] and [vdD10]; in fact, some parts are almost literally copied from them. I also used my handwritten lecture notes from Matthias Aschenbrenner's model theory course taught at UCLA, as well as [Mar02] and [End01].

## 1. Introduction

At the beginning of the 20th century mathematics experienced a crisis due to the discovery of certain paradoxes (e.g. Russell's paradox) in previous attempts to formalize abstract notions of sets and functions. To put analysis on a firm foundation, similar to the axiomatic foundation for geometry, Hilbert proposed a program aimed at a direct consistency proof of analysis. This would involve a system of axioms that is consistent, meaning free of internal contradictions, and complete, meaning rich enough to prove all true statements. But the search for such a system was doomed to fail: Gödel proved in the early 1930s that any system of axioms that can be listed by some computable process, and subsumes Peano arithmetic, is either incomplete or inconsistent. This is the Gödel Incompleteness Theorem, and we will prove it in the second half of this course. In the first half, we will develop the framework of First Order Logic (FOL), culminating in a proof of the Completeness theorem, yet another foundational theorem by Gödel. From this we will derive the Compactness theorem, which is one of the most useful tools of logic. In addition, we will discuss applications in various fields of mathematics such as combinatorics and algebra.

## 2. First order logic

Like any other field of mathematics, mathematical logic starts with a pile of definitions, the importance and use of which will become apparent as we go. Right now, our position is analogous to that of an instructor of geometry who has to define the concept of a differential manifold from scratch without assuming knowledge of point set topology and differentiability. So one has to patiently make his way through the definitions keeping in mind that the end goal is worth it. Let the story begin...

### 2.1. **Structures**

Every mathematician recognizes a mathematical structure as such when he sees it. Here are some

**Examples 2.1.**

(a) A *graph* is a pair $\mathbf{\Gamma} = (\Gamma, E)$, where $\Gamma \neq \varnothing$ is the set of nodes and $E$ is a binary relation on $\Gamma$, i.e. $E \subseteq \Gamma^2$.

(b) A *partial ordering* is a pair $\mathbf{P} = (P, \leq)$, where $P$ is a set and $\leq$ is a binary relation on it satisfying the following conditions:
   (i) (Reflexivity) $\forall x \in P$, $x \leq x$,
   (ii) (Antisymmetry) $\forall x, y \in P$, if $x \leq y$ and $y \leq x$, then $x = y$,
   (iii) (Transitivity) $\forall x, y, z \in P$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

(c) A *group* is a triple $\mathbf{G} = (G, 1, \cdot)$, where $G$ is a set, 1 is a fixed element of $G$ (a constant) and $\cdot$ is a binary operation on $G$ such that the following conditions hold:
   (i) (Associativity) $\forall x, y, z \in G$, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$,
   (ii) (Identity) $\forall x \in G$, $1 \cdot x = x \cdot 1 = 1$,
   (iii) (Inverse) $\forall x \in G \; \exists y \in G$, $xy = 1$.

(d) An *ordered field* is a 6-tuple $\mathbf{F} = (F, 0, 1, +, \cdot, <)$, where $F$ is a set, $0, 1$ are some fixed elements of $F$, $+$ and $\cdot$ are binary operations, and $<$ is a binary relation on $F$ such that certain conditions are satisfied (too many to list here).

What is common between these examples? Well, they all have an underlying set together with either relations, operations or constant elements (or all of the above as in example (d)) defined on it. Let's formalize this and give an abstract definition of a *mathematical structure.*

**Definition 2.2** (Structure)**.** *A structure is a quadruple* $\mathbf{S} = (S, \mathcal{C}, \mathcal{F}, \mathcal{R})$*, where* $S$ *is a set,* $\mathcal{C}$ *is a set of elements from* $S$ *(constants),* $\mathcal{F}$ *is a set of operations on* $S$ *(i.e. each element of* $F$ *is a function from* $S^n$ *to* $S$ *for some* $n \geq 1$*) and* $\mathcal{R}$ *is a set of relations on* $S$ *(not necessarily binary).*

Although this definition covers all of the examples above, it is awkward to use when defining a subclass of structures, say groups. In order to define the class of groups, we have to not only demand that in those structures $\mathcal{C}$ and $\mathcal{F}$ are singletons, $\mathcal{R} = \varnothing$ and the operation in $\mathcal{F}$ is binary, but we also have to make sure that the conditions (i)-(iii) of example (c) are satisfied. For these conditions to even make sense, we have to specify that 1 refers to the unique element in $\mathcal{C}$ and $\cdot$ refers to the unique element in $\mathcal{F}$. So why don't we first fix a set of names (like $\{1, \cdot\}$) and then include their correspondence with the actual constants, functions and relations in the definition of a structure? In fact, that is exactly what we will do.

**Definition 2.3** (Signature)**.** *A signature is a quadruple*

$$\tau = (\mathcal{C}, \mathcal{F}, \mathcal{R}, a),$$

*where* $\mathcal{C}, \mathcal{F}, \mathcal{R}$ *are pairwise disjoint sets (of symbols), which we refer to as the sets of constant, relation and function symbols, respectively, and*

$$a : \mathcal{F} \cup \mathcal{R} \to \mathbb{N}^{>0}.$$

*(Here* $\mathbb{N}^{>0}$ *denotes the set of positive natural numbers because in logic* $\mathbb{N}$ *includes* $0$*.)*

A relation or function symbol $P$ (i.e. an element of $\mathcal{F} \cup \mathcal{R}$) is said to be $n$-ary if $a(P) = n$. The sets $\mathcal{C}, \mathcal{F}, \mathcal{R}$ should be thought of as *names* for constant elements, relations and functions (operations), and not the actual constant elements, relations and functions themselves! It is also good to keep in mind that any of the sets $\mathcal{C}, \mathcal{F}, \mathcal{R}$ can be empty.

**Examples 2.4.**

(a) The signature for graphs is

$$\tau_{\text{graphs}} = (\varnothing, \varnothing, \{E\}, (E \mapsto 2)),$$

However, this is too formal and hard to read, so in order to avoid headache (think of a signature for ordered fields!) we simply write

$$\tau_{\text{graph}} = (E),$$

and then specify that $E$ is a binary relation symbol.

(b) The signature for groups (or monoids) is

$$\tau_{\text{group}} = (1, \cdot),$$

where $\cdot$ is a binary function symbol and 1 is a constant symbol.

(c) The signature for rings is

$$\tau_{\text{ring}} = (0, 1, +, -, \cdot),$$

where $+, -, \cdot$ are binary function symbols and $0, 1$ are constant symbols.

(d) The signature for arithmetic is

$$\tau_{\mathrm{a}} = (0, S, +, \cdot),$$

where 0 is a constant symbol, $S$ is a unary function symbol ($S$ stands for "successor"), and $+, \cdot$ are binary function symbols.

(e) The signature for sets is

$$\tau_{\mathrm{set}} = (\in),$$

where $\in$ is a binary relation symbol.

Although in this examples the signatures are finite, it is not required by the definition. Now we are ready to define a structure in a given signature $\tau = (\mathcal{C}, \mathcal{R}, \mathcal{F})$.

**Definition 2.5** ($\tau$-structure)**.** *A $\tau$-structure is a pair $\mathbf{S} = (S, i)$, where $S$ is a set and $i$ is a map (correspondence) that assigns*

- *to each constant symbol $c$ in $\tau$ a member $i(c)$ of $S$;*
- *to each $n$-ary relation symbol $R$ in $\tau$ an $n$-ary function $i(R) \subseteq S^n$;*
- *to each $n$-ary function symbol $f$ in $\tau$ an $n$-ary function $i(f) : S^n \to S$.*

We call $S$ the universe of the structure $\mathbf{S}$. The choice of the letter $i$ is because we think of $i$ as the *interpretation* of the symbols of $\tau$ in the structure $\mathbf{S}$. To simplify the notation, we write $q^{\mathbf{S}}$ instead of $i(q)$, for all symbols $q$ in $\tau$, and so instead of $(S, i)$, we write

$$\mathbf{S} = (S, \{c^{\mathbf{S}}\}_{c \in \mathcal{C}}, \{R^{\mathbf{S}}\}_{R \in \mathcal{R}}, \{f^{\mathbf{S}}\}_{f \in \mathcal{F}}).$$

For finite signatures, we use an even simpler notation as in the following examples.

**Examples 2.6.**

(a) A complete graph on $n$ vertices is a $\tau_{\mathrm{graphs}}$-structure

$$\mathbf{K_n} = (\Gamma, E^{\mathbf{K_n}}),$$

where $\Gamma$ is a set of $n$ elements and $E^{\mathbf{K_n}} = \Gamma^2$.

(b) $\mathbb{Z}$ as a group is a $\tau_{\mathrm{group}}$-structure

$$\mathbf{Z} = (\mathbb{Z}, 1^{\mathbf{Z}}, \cdot^{\mathbf{Z}}),$$

where $1^{\mathbf{Z}}$ is $0 \in \mathbb{Z}$ and $\cdot^{\mathbf{Z}}$ is the usual addition operation.

(c) Here is a useless $\tau_{\mathrm{ring}}$-structure:

$$\mathbf{R}_{\mathrm{crazy}} = (\mathbb{R}, 0^{\mathbf{R}_{\mathrm{crazy}}}, 1^{\mathbf{R}_{\mathrm{crazy}}}, +^{\mathbf{R}_{\mathrm{crazy}}}, -^{\mathbf{R}_{\mathrm{crazy}}}, \cdot^{\mathbf{R}_{\mathrm{crazy}}}),$$

where $0^{\mathbf{R}_{\mathrm{crazy}}}, 1^{\mathbf{R}_{\mathrm{crazy}}}$ are equal to $\pi$, $+^{\mathbf{R}_{\mathrm{crazy}}}$ is the $\sin(x+y)$ function, $-^{\mathbf{R}_{\mathrm{crazy}}}$ is the $x+y$ function and $\cdot^{\mathbf{R}_{\mathrm{crazy}}}$ is the $x+4y$ function. Clearly $\mathbf{R}_{\mathrm{crazy}}$ is far from being a ring although it is a structure in the signature of rings.

(d) $\mathbb{R}$ as a field is a $\tau_{\mathrm{ring}}$-structure:

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

where all of the symbols are interpreted in the usual way.

(e) The structure of natural numbers as a $\tau_{\mathrm{a}}$-structure will be the central object of this course:

$$\mathbf{N} = (\mathbb{N}, 0^{\mathbf{N}}, S^{\mathbf{N}}, +^{\mathbf{N}}, \cdot^{\mathbf{N}}),$$

where $0^{\mathbf{N}}, +^{\mathbf{N}}, \cdot^{\mathbf{N}}$ are defined in the usual way, and $S^{\mathbf{N}}$ is the successor operation (i.e. the unary function of adding 1).

Since it is annoying to keep writing $\mathbf{S}$ in the superscript to denote the interpretation of symbols of $\tau$ in a $\tau$-structure $\mathbf{S}$, we omit it as long as it is clear from the context that we mean the interpretations rather than the symbols. For example, we will write $\mathbf{N} = (\mathbb{N}, 0, S, +, \cdot)$ instead of $\mathbf{N} = (\mathbb{N}, 0^{\mathbf{N}}, S^{\mathbf{N}}, +^{\mathbf{N}}, \cdot^{\mathbf{N}})$.

In algebra, one of the first things you learn after the definition of a group is the definition of a subgroup, homomorphism and isomorphism. We do the same with arbitrary structures.

**Definition 2.7** (Substructure)**.** *For $\tau$-structures $\mathbf{A}, \mathbf{B}$, we say that $\mathbf{A}$ is a substructure of $\mathbf{B}$ and write $\mathbf{A} \subseteq \mathbf{B}$ if $A \subseteq B$ and the interpretations of $\tau$ by $\mathbf{A}$ and $\mathbf{B}$ coincide on $A$, more precisely:*

- $c^{\mathbf{A}} = c^{\mathbf{B}}$, *for any constant symbol $c$ in $\tau$,*
- $f^{\mathbf{A}}(\vec{a}) = f^{\mathbf{B}}(\vec{a})$, *for any $n$-ary function symbol $f$ in $\tau$ and for all $\vec{a} \in A^n$,*
- $R^{\mathbf{A}}(\vec{a}) \iff R^{\mathbf{B}}(\vec{a})$, *for any $n$-ary relation symbol $R$ in $\tau$ and for all $\vec{a} \in A^n$.*

For example, $(\mathbb{N}, 0, +)$ is a substructure of $(\mathbb{Z}, 0, +)$, $\mathbf{Z} = (\mathbb{Z}, 0, 1, +, \cdot)$ is a substructure of $\mathbf{R} = (\mathbb{R}, 0, 1, +, \cdot)$. If $\tau$ only contains relation symbols, then any subset is (the universe of) a substructure.

Note that the intersection of substructures of the same structure is again a substructure. Let $\mathbf{B}$ be a $\tau$-structure and $S \subseteq B$. The *substructure generated by $S$* is the smallest substructure containing $S$, i.e. it is the intersection of all substructures of $\mathbf{B}$ that contain $S$. We denote this fact by $\mathbf{A} =<S>_{\mathbf{B}}$. For example, the substructure of $\mathbf{R} = (\mathbb{R}, 0, 1, +, \cdot)$ generated by $\varnothing$ is $(\mathbb{N}, 0, 1, +, \cdot)$ (why?).

**Definition 2.8** (Homomorphism)**.** *Let $\mathbf{A}, \mathbf{B}$ be $\tau$-structures. A function $h : A \to B$ is called a $\tau$-homomorphism (or just homomorphism) if $h$ respects the interpretation of $\tau$, more precisely:*

- $h(c^{\mathbf{A}}) = c^{\mathbf{B}}$, *for any constant symbol $c$ in $\tau$,*
- $h(f^{\mathbf{A}}(\vec{a})) = f^{\mathbf{B}}(h(\vec{a}))$, *for any $n$-ary function symbol $f$ in $\tau$ and for all $\vec{a} \in A^n$,*
- $R^{\mathbf{A}}(\vec{a}) \implies R^{\mathbf{B}}(h(\vec{a}))$, *for any $n$-ary relation symbol $R$ in $\tau$ and for all $\vec{a} \in A^n$,*

*where for $\vec{a} = (a_1, ..., a_n)$, $h(\vec{a}) := (h(a_1), ..., h(a_n))$.*

We write $h : \mathbf{A} \to \mathbf{B}$ to mean that $h$ is a homomorphism between the structures $\mathbf{A}, \mathbf{B}$ (note that it is different from $h : A \to B$).

**Definition 2.9** (Isomorphism)**.** *Let $\mathbf{A}, \mathbf{B}$ be $\tau$-structures. A function $h : A \to B$ is called a $\tau$-isomorphism (or just isomorphism) if $h$ is bijective and both $h$ and $h^{-1}$ are $\tau$-homomorphisms. $\mathbf{A}, \mathbf{B}$ are called isomorphic if there is an isomorphism between them.*

Sometimes in algebra we consider the universe of a ring as an abelian group under addition, in other words, we "forget" the multiplication operation. We make this precise here.

**Definition 2.10.** *Let $\tau, \tau'$ be signatures with $\tau \subseteq \tau'$, let $\mathbf{A}$ be a $\tau$-structure and $\mathbf{B}$ be a $\tau'$-structure. We say that $\mathbf{A}$ is a reduct of $\mathbf{B}$ (or $\mathbf{B}$ an expansion of $\mathbf{A}$) and write $\mathbf{A} = \mathbf{B}\restriction_{\tau}$ if $\mathbf{A}$ and $\mathbf{B}$ have the same underlying set and the same interpretations of the symbols of $\tau$.*

For example, $(\mathbb{R}, 0, +)$ is a reduct of $(\mathbb{R}, 0, 1, +, \cdot)$, which in its turn is a reduct of $(\mathbb{R}, 0, 1, +, \cdot, <)$.

### 2.2. **Language and interpretation**

Now we have to define the language of the First Order Logic (FOL) that will allow us to express statements about $\tau$-structures, like axioms (i)-(iii) in Example 2.1(c). Although the definitions below are very natural, they are somewhat annoying to write and even to read. The readers are advised to try to come up with the definitions themselves before (instead of?) reading.

Let $\tau$ denote a signature for the rest of the section.

**Definition 2.11** (Alphabet)**.** *The alphabet $\mathbb{FOL}(\tau)$ of the first order language in the signature $\tau$ comprises of the symbols in $\tau$ and the following additional symbols:*

- *logical symbols* $= \neg\ \wedge\ \vee\ \rightarrow\ \ \forall\ \exists$
- *punctuation symbols* , ( )
- *variables* $v_0, v_1, v_2, ...$

Words in $\mathbb{FOL}(\tau)$ are finite sequences of symbols from $\mathbb{FOL}(\tau)$.

**Definition 2.12** (Terms)**.** *A term in $\mathbb{FOL}(\tau)$ (or a $\tau$-term) is a word formed by the following recursive rules:*

*(i) each constant symbol is a term;*
*(ii) each variable is a term;*
*(iii) if $t_1, ..., t_n$ are terms and $f \in \tau$ is an $n$-ary function symbol, then $f(t_1, ..., t_n)$ is a term.*

**Examples 2.13.**

(a) $(v_0 \cdot 1) \cdot v_3$ is a term in $\mathbb{FOL}(\tau_{\text{group}})$. Note that the way this term is written is technically incorrect, we should have written $\cdot(\cdot(v_0, 1), v_3)$, but the latter is almost impossible to read, so we will keep abusing notation and write the former way.
(b) $S(0 + v_2) + S(S(S(v_2)))$ is a term in $\mathbb{FOL}(\tau_{\text{a}})$ (the language of arithmetic).
(c) Variables $v_0, v_1, ...$ are the only terms in $\mathbb{FOL}(\tau_{\text{graph}})$.

We also often use letters different than $v_0, v_1, ...$ to denote variables, e.g. $v, u, x, y, z$.

**Definition 2.14** (Interpretation of terms)**.** *Let $\mathbf{M}$ be a $\tau$-structure and $t$ be a $\tau$-term build using variables from $\vec{v} = (v_1, ..., v_n)$. We define the interpretation of $t(\vec{v})$ in $\mathbf{M}$ as a function $t^{\mathbf{M}} : M^n \to M$ by induction on the construction of $t$ as follows: for $\vec{a} = (a_1, ..., a_n) \in M^n$*

*(i) if $t = c$, where $c$ is a constant symbol in $\tau$, then $t^{\mathbf{M}}(\vec{a}) = c^{\mathbf{M}}$;*
*(ii) if $t = v_i$, then $t^{\mathbf{M}}(\vec{a}) = a_i$;*
*(iii) if $t = f(t_1, ..., t_n)$, where $t_1, ..., t_k$ are terms and $f$ is an $k$-ary function symbol in $\tau$, then $t^{\mathbf{M}}(\vec{a}) = f^{\mathbf{M}}(t_1^{\mathbf{M}}(\vec{a}), ..., t_n^{\mathbf{M}}(\vec{a}))$.*

So one should think of the term $t(\vec{v})$ as a name of the function $t^{\mathbf{M}}$. Note that if $t = v_1$, then $t(v_1)$ is interpreted as a unary function, while $t(v_1, v_2)$ as a binary function (although it does not depend on $v_2$). This is exactly what we do with polynomials for example: we write $p(x, y) = x^2 + 1$ to mean that this is a polynomial in two variables $x$ and $y$ although it doesn't depend on $y$.

**Definition 2.15** (Formulas)**.** *A formula in $\mathbb{FOL}(\tau)$ is a word formed by the following recursive rules:*

*(i) if $s, t$ are terms then $s = t$ is a formula;*

*(ii) if $t_1, ..., t_n$ are terms and $R \in \tau$ is an $n$-ary relation symbol, then $R(t_1, ..., t_n)$ is a formula;*

*(iii) if $\phi$ and $\psi$ are formulas then $\neg(\phi)$, $(\phi) \wedge (\psi)$, $(\phi) \vee (\psi)$, $(\phi) \to (\psi)$, $\forall v\phi$, $\exists v\phi$ are formulas.*

According to this definition, $(\forall x(x = y)) \ \wedge \ (x \neq z)$ is a valid formula (in any signature), although the third occurrence of $x$ has nothing to do with its first two occurrences, where $x$ is used as the variable of the quantifier. The use of $x$ as the variable for the quantifier is a bad idea because it makes reading of the formula hard and confusing. (Imagine writing $x \int_0^1 x dx$ instead of $x \int_0^1 t dt$ in a calculus course!) Thus we make a convention to not use such bad notation.

**Convention.** We say that the variable $v$ is *quantified* in the formula $\phi$ if $\forall v\psi$ or $\exists v\psi$, for some formula $\psi$, occurs in some stage of the recursive construction of $\phi$. We make the convention that each variable $v$ can be used with a quantifier only once ($\forall v\psi$ or $\exists v\psi$ occurs at most once) and in this case $v$ is not allowed to be used elsewhere other than in $\psi$.

This convention makes things like $(\forall x(x = y)) \ \wedge \ (x \neq z)$ invalid, and one should write $(\forall t(t = y)) \ \wedge \ (x \neq z)$ instead.

A variable $v$ is *free* in a formula $\phi$ if it occurs in $\phi$ and is not quantified. A formula without free variables is called a *sentence*. Note that all statements (theorems, conjectures, etc.) in mathematics are sentences (in the language of set theory).

We interpret formulas in a given structure $\mathbf{M}$ as $n$-ary relations on $M$, for some $M$, or, equivalently, as functions from $M^n$ to $\{$`true, false`$\}$. Just like we did with terms, we define interpretation for $\phi(\vec{v})$ (as opposed to just $\phi$), for a vector of variables $\vec{v} = (v_1, ..., v_n)$, as long as the free variables of $\phi$ are among $v_1, ..., v_n$ and none of $v_1, ..., v_n$ is quantified in $\phi$.

**Definition 2.16** (Interpretation of formulas)**.** *Let $\mathbf{M}$ be a $\tau$-structure, $\phi$ a $\tau$-formula and let $\vec{v}$ be as above. For $\vec{a} = (a_1, ... a_n) \in M^n$, we define the relation $\mathbf{M} \vDash \phi(\vec{a})$ by induction on the construction of $\phi$ as follows:*

*(i) if $\phi$ is $t_1 = t_2$, then $\mathbf{M} \vDash \phi(\vec{a})$ if $t_1^{\mathbf{M}}(\vec{a}) = t_2^{\mathbf{M}}(\vec{a})$;*

*(ii) if $\phi$ is $R(t_1, ..., t_n)$, then $\mathbf{M} \vDash \phi(\vec{a})$ if $R^{\mathbf{M}}(t_1^{\mathbf{M}}(\vec{a}), ..., t_n^{\mathbf{M}}(\vec{a}))$, i.e. $(t_1^{\mathbf{M}}(\vec{a}), ..., t_n^{\mathbf{M}}(\vec{a})) \in R^{\mathbf{M}}$;*

*(iii) if $\phi$ is $\neg\psi$, then $\mathbf{M} \vDash \phi(\vec{a})$ if $\mathbf{M} \nvDash \phi(\vec{a})$;*

*(iv) if $\phi$ is $\psi \wedge \theta$, then $\mathbf{M} \vDash \phi(\vec{a})$ if $\mathbf{M} \vDash \psi(\vec{a})$ and $\mathbf{M} \vDash \theta(\vec{a})$;*

*(v) if $\phi$ is $\psi \vee \theta$, then $\mathbf{M} \vDash \phi(\vec{a})$ if $\mathbf{M} \vDash \psi(\vec{a})$ or $\mathbf{M} \vDash \theta(\vec{a})$;*

*(vi) if $\phi$ is $\forall u\psi(\vec{v}, u)$ (hence $u$ is not in $\vec{v}$ by our assumption), then $\mathbf{M} \vDash \phi(\vec{a})$ if for all $b \in M$, $\mathbf{M} \vDash \psi(\vec{a}, b)$;*

*(vii) if $\phi$ is $\exists u\psi(\vec{v}, u)$, then $\mathbf{M} \vDash \phi(\vec{a})$ if there exists $b \in M$, $\mathbf{M} \vDash \psi(\vec{a}, b)$.*

We read $\mathbf{M} \vDash \phi(\vec{a})$ as $\phi$ is true (holds) about $\vec{a}$ in $\mathbf{M}$. Note that the above definition applies when $\phi$ is a sentence and $n = 0$. In this case, we read $\mathbf{M} \vDash \phi$ as $\phi$ is true/valid (holds) in $\mathbf{M}$. For a vector of variables $\vec{v} = (v_1, ..., v_n)$, we say that a formula $\phi(\vec{v})$ is valid in $\mathbf{M}$ and write $\mathbf{M} \vDash \phi(\vec{v})$ if $\mathbf{M} \vDash \forall\vec{v}\phi(\vec{v}))$, where $\forall\vec{v}$ abbreviates $\forall v_1 \forall v_2 ... \forall v_n$.

**Examples 2.17.**

(a) $\mathbf{N} \vDash S(S(0)) = 2$

(b) Thanks to A. Wiles, we now know that $\mathbf{N} \vDash \forall n \forall x \forall y \forall z[(n \geq 3 \wedge x^n + y^n = z^n) \to (x = 0 \vee y = 0)]$.

(c) $\mathbf{R} \vDash \exists y(a = y \cdot y)$ holds for all non-negative $a \in \mathbb{R}$.

**Lemma 2.18.** *Let $\mathbf{A}, \mathbf{B}$ be two $\tau$-structures. If $h : \mathbf{A} \to \mathbf{B}$ is a homomorphism, then for any term $t(\vec{v})$ and $\vec{a} \in A^n$,*

$$h(t^{\mathbf{A}}(\vec{a})) = t^{\mathbf{B}}(h(\vec{a})),$$

*where $h(\vec{a}) = (h(a_1), ..., h(a_n))$.*

*Proof.* We prove by induction on the construction (length) of $t$.

- If $t = c$, for a constant symbol $c$ in $\tau$, then $t^{\mathbf{A}}(\vec{a}) = c^{\mathbf{A}}$ and hence we have
$$h(t^{\mathbf{A}}(\vec{a})) = h(c^{\mathbf{A}}) = c^{\mathbf{B}} = t^{\mathbf{B}}(h(\vec{a}))$$
because $h$ is a homomorphism.
- If $t = v_i$, for a variable $v_i$, then $t^{\mathbf{A}}(\vec{a}) = a_i$ and hence we have
$$h(t^{\mathbf{A}}(\vec{a})) = h(a_i) = t^{\mathbf{B}}(h(\vec{a})).$$
- If $t = f(t_1, ..., t_k)$, for a function symbol $f$ in $\tau$, then
$$\begin{aligned} h(t^{\mathbf{A}}(\vec{a})) &= h(f^{\mathbf{A}}(t_1^{\mathbf{A}}(\vec{a}), ..., t_k^{\mathbf{A}}(\vec{a}))) \\ &= f^{\mathbf{B}}(h(t_1^{\mathbf{A}}(\vec{a})), ..., h(t_k^{\mathbf{A}}(\vec{a}))) && (h \text{ is a homomorphism}) \\ &= f^{\mathbf{B}}(t_1^{\mathbf{B}}(h(\vec{a})), ..., t_k^{\mathbf{A}}(h(\vec{a}))) && (\text{by the induction hypothesis}) \\ &= t^{\mathbf{B}}(h(\vec{a})). \end{aligned}$$

□

**Proposition 2.19.** *Let $\mathbf{A}, \mathbf{B}$ be two $\tau$-structures. If $h : \mathbf{A} \to \mathbf{B}$ is an isomorphism, then for any formula $\phi(v_1, ..., v_n)$ and $(a_1, ..., a_n) \in A^n$,*

$$\mathbf{A} \vDash \phi(a_1, ...a_n) \iff \mathbf{B} \vDash \phi(h(a_1), ..., h(a_n)).$$

*Proof.* We prove by induction on the construction (length) of $\phi$. For the step of induction, it is enough to consider only the following cases: $\phi \equiv \neg\psi$, $\phi \equiv \neg\psi_1 \wedge \psi_2$ and $\phi \equiv \exists v \psi$.

- If $\phi \equiv t_1 = t_2$, then
$$\begin{aligned} \mathbf{A} \vDash \phi(\vec{a}) &\iff t_1^{\mathbf{A}}(\vec{a}) = t_2^{\mathbf{A}}(\vec{a}) \\ &\iff h(t_1^{\mathbf{A}}(\vec{a})) = h(t_2^{\mathbf{A}}(\vec{a})) && (h \text{ is injective}) \\ &\iff t_1^{\mathbf{B}}(h(\vec{a})) = t_2^{\mathbf{B}}(h(\vec{a})) && (\text{by Lemma 2.18}) \\ &\iff \mathbf{B} \vDash \phi(h(\vec{a})). \end{aligned}$$
- If $\phi \equiv R(t_1, ..., t_k)$, then the calculation is similar to the previous case (also uses Lemma 2.18).
- If $\phi \equiv \neg\psi$, then
$$\begin{aligned} \mathbf{A} \vDash \phi(\vec{a}) &\iff \mathbf{A} \nvDash \psi(\vec{a}) \\ &\iff \mathbf{B} \nvDash \psi(\vec{a}) && (\text{by the induction hypothesis}) \\ &\iff \mathbf{B} \vDash \phi(h(\vec{a})). \end{aligned}$$
- If $\phi \equiv \psi_1 \wedge \psi_2$, then the calculation is similar to the previous case.

- If $\phi \equiv \exists v \psi$, then

$$\begin{aligned}
\mathbf{A} \models \phi(\vec{a}) &\iff \exists a' \in A, \mathbf{A} \models \psi(\vec{a}, a') \\
&\iff \exists a' \in A, \mathbf{B} \models \psi(h(\vec{a}), h(a')) && \text{(by the induction hypothesis)} \\
&\iff \exists b \in B, \mathbf{B} \models \psi(h(\vec{a}), b) && \text{(use surjectivity of } h \text{ for } \Longleftarrow) \\
&\iff \mathbf{B} \models \phi(h(\vec{a})).
\end{aligned}$$

□

**Proposition 2.20.** *If a $\tau$-structure $\mathbf{A}$ is a reduct of a $\tau'$-structure $\mathbf{B}$, then for every $\tau$-formula $\phi(\vec{v})$ and $\vec{a} \in A^n$ ($= B^n$),*

$$\mathbf{A} \models \phi(\vec{a}) \iff \mathbf{B} \models \phi(\vec{a}).$$

*Proof.* Trivial induction on formulas and possibly also terms. □

## 2.3. **Definability**

**Definition 2.21** (Definability)**.** *Let $\mathbf{M}$ be a $\tau$-structure and $A \subseteq M$. $D \subseteq M^n$ is called $A$-definable (or definable from $A$) in $\mathbf{M}$ if there is a formula $\phi(\vec{x}, \vec{y})$, where $\vec{x} = (x_1, ..., x_n)$ and $\vec{y} = (y_1, ..., y_m)$ (for some $m \geq 0$), and $\vec{a} \in M^m$ such that $\forall \vec{b} \in M^n$*

$$\vec{b} \in D \Leftrightarrow \mathbf{M} \models \phi(\vec{b}, \vec{a}).$$

If $A = \varnothing$, we say that $D$ is 0-definable, and if $A = M$, we say that $D$ is definable. We say that an element $\vec{b} \in M^n$ is definable if so is the singleton $\{\vec{b}\}$. An $n$-ary function $f : A^n \to A$ is called *definable* in $\mathbf{A}$ if so is its graph $\{(\vec{a}, b) \in A^n \times A : f(\vec{a}) = b\}$.

Note that the set $\mathcal{D}^n_A$ of $A$-definable subsets of $M^n$ is an algebra, i.e. it is closed under finite unions and complements and contains $\varnothing$ and $M^n$. It is very useful to consider the topology on $M^n$ generated by $\mathcal{D}^n_A$. It is clear that $\mathcal{D}^n_A$ is actually a base for that topology. Note that the topology might not be Hausdorff and whether it is compact or not is tightly related to a property called saturation, which however is outside the scope of the course.

**Examples 2.22.**

(a) In $\mathbf{R} = (\mathbb{R}, 0, 1, +, \cdot)$, the set of positive numbers is 0-definable by the formula $\phi_{>0}(x) \equiv x \neq 0 \wedge \exists y(x = y^2)$, where $y^2$ is the abbreviation for $y \cdot y$. Using this, one can define the binary relation $< \subseteq \mathbb{R}^2$ by the formula $\phi_<(x, y) \equiv \phi_{>0}(y - x)$ (0-definable). Thus $\mathbf{R}$ and $\mathbf{R}_< = (\mathbb{R}, 0^{,}1, +, \cdot, <)$ have the same definable sets.

(b) In $\mathbf{R}_<$ the set $\{r \in \mathbb{R} : r < \pi\}$ is definable by the formula $x < \pi$. It turns out that this set is not 0-definable. This follows from the fact that $\pi$ is transcendental and a famous theorem of Tarski that $\mathbf{R}_<$ admits "quantifier elimination", which implies that all 0-definable sets are just finite unions of intervals with algebraic (or infinite) endpoints.

(c) In any graph $\mathbf{\Gamma} = (\Gamma, E)$, the set

$$\{(u, v) \in G^2 : \text{the edge-distance between } u \text{ and } v \text{ is } \leq 2\}$$

is 0-definable by the formula

$$\phi(x, y) \equiv xEy \vee \exists z(xEz \wedge zEy).$$

Similarly, one can show that for any $n \geq 1$, the set

$$\{(u, v) \in G^2 : \text{the edge-distance between } u \text{ and } v \text{ is } \leq n\}$$

is 0-definable. However it turns out that the set

$$\{(u,v)\in G^2 : u \text{ and } v \text{ are connected}\}$$

is not even definable in some (actually most) graphs. We will prove this later on in the course after proving the Compactness theorem.

The definable subsets of $\mathbf{N} = (\mathbb{N}, 0, S, +, \cdot)$ are called *arithmetical*. It is easy to see that a set is definable in $\mathbb{N}$ if and only if it is 0-definable.

### 2.4. **Theories and models**

Given a signature $\tau$, a set of $\tau$-sentences is called a *$\tau$-theory*. The sentences in a theory $T$ are often referred to as *axioms*.

We say that a nonempty $\tau$-structure $\mathbf{M}$ *satisfies* (or models) a $\tau$-theory $T$ and write $\mathbf{M} \vDash T$ if $\mathbf{M} \vDash \phi$, for every $\phi \in T$. Equivalently, we also say that $\mathbf{M}$ is a *model* of $T$.

A theory $T$ is called *satisfiable* (or *semantically consistent*) if it has a model.

Given a $\tau$-structure $\mathbf{M}$, we put $\mathsf{Th}(\mathbf{M}) = \{\phi : \phi \text{ is a } \tau\text{-sentence and } \mathbf{M} \vDash \phi\}$. Note that $\mathsf{Th}(\mathbf{M})$ is satisfiable and for every $\tau$-sentence $\phi$, $\mathsf{Th}(\mathbf{M})$ contains exactly one of $\phi$ and $\neg\phi$.

We say that a $\tau$-theory $T$ satisfies a $\tau$-sentence $\phi$ and write $T \vDash \phi$, if every model of $T$ satisfies $\phi$, i.e. $\forall \mathbf{M} \vDash T(\mathbf{M} \vDash \phi)$. Equivalently, we say that $T$ semantically implies $\phi$.

**Examples 2.23.**

(a) The theory $\mathsf{GRAPHS}$ of undirected graphs with no loops in the signature $\tau_{\text{graph}} = (E)$ consists of the following axioms:
  (i) (Undirected) $\forall x \forall y(xEy \to yEx)$,
  (ii) (No loops) $\forall x(\neg xEx)$.

(b) The theory of undirected infinite graphs with no loops in the signature $\tau_{\text{graph}} = (E)$:

$$\mathsf{GRAPHS}_\infty = \mathsf{GRAPHS} \cup \{\exists v_1 \exists v_2 \dots \exists v_n \bigwedge_{i<j} v_i \neq v_j : n \geq 2\}.$$

(c) The theory $\mathsf{PO}$ of partial orderings in the signature $\tau_{\text{PO}} = (\leq)$ consists of the following axioms:
  (PO1) (Reflexivity) $\forall x(x \leq x)$.
  (PO2) (Antisymmetry) $\forall x \forall y(x \leq y \wedge y \leq x \to x = y)$,
  (PO3) (Transitivity) $\forall x \forall y \forall z(x \leq y \wedge y \leq z \to x \leq z)$.

(d) The theory $\mathsf{GROUPS}$ of partial orderings in the signature $\tau_{\text{group}} = (1, \cdot)$ consists of the following axioms:
  (G1) (Associativity) $\forall x \forall y \forall z[x \cdot (y \cdot z) = (x \cdot y) \cdot z]$,
  (G2) (Identity) $\forall x[1 \cdot x = x \cdot 1 = 1]$,
  (G3) (Inverse) $\forall x \exists y[xy = 1]$.

  We know from group theory that GROUPS $\vDash \forall x \forall y \forall y'(yx = 1 = xy' \to y = y')$.

(e) Similarly, one defines the theory $\mathsf{RINGS}$ of rings in the signature $\tau_{\text{ring}} = (0, 1, +, -, \cdot)$ (too many axioms to write, but still finitely many), and then one the theory $\mathsf{FIELDS}$ of fields is defined as $\mathsf{RINGS}$ together with the following two axioms:
  (F1) (Nonzero) $0 \neq 1$,
  (F2) (Commutativity) $\forall x \forall y[x \cdot y = y \cdot x]$,
  (F3) (Multiplicative inverse) $\forall x \exists y[xy = 1]$,

(f) So far all the theories were finite. Here is an example of an infinite theory. The theory of algebraically closed fields in the signature $\tau_{\text{ring}}$:

$$\mathsf{ACF} = \mathsf{FIELDS} \cup \{\forall a_0 \forall a_1 ... \forall a_n \exists r[a_n r^n + a_{n-1} r^{n-1} + ... + a_1 r + a_0 = 0] : n \in \mathbb{N}\}.$$

(g) The theory of fields of characteristic $p$, for a prime number $p$:

$$\mathsf{FIELDS}_p = \mathsf{FIELDS} \cup \{\underbrace{1 + 1 + ... + 1}_{p} = 0\}.$$

One can easily show that for any $n \geq 0$,

$$\mathsf{FIELDS}_p \vDash \underbrace{1 + 1 + ... + 1}_{n} = 0 \iff p \text{ divides } n.$$

(h) The theory of fields of characteristic 0:

$$\mathsf{FIELDS}_0 = \mathsf{FIELDS} \cup \{\underbrace{1 + 1 + ... + 1}_{p} \neq 0 : p \text{ prime}\}.$$

It is easy to see that for all $n \geq 1$, $\mathsf{FIELDS}_0 \vDash \underbrace{1 + 1 + ... + 1}_{n} \neq 0$.

(i) The theory of algebraically closed fields of fixed characteristic $n$, where $n$ is either 0 or prime:

$$\mathsf{ACF}_n = \mathsf{ACF} \cup \mathsf{FIELDS}_n.$$

(j) The theory $\mathsf{PA}$ of arithmetic, called Peano Arithmetic (defined by Peano), in the signature $\tau_{\text{a}} = (0, S, +, \cdot)$ consists of the following (infinitely many) axioms:

(PA1) $\forall x[\neg S(x) = 0]$,

(PA2) $\forall x \forall y[S(x) = S(y) \rightarrow x = y]$,

(PA3) $\forall x[x + 0 = x]$,

(PA4) $\forall x \forall y[S(x + y) = x + S(y)]$,

(PA5) $\forall x[x \cdot 0 = 0]$,

(PA6) $\forall x \forall y[x \cdot S(y) = x \cdot y + x]$,

(PA7) (Axiom schema of induction) for all $\tau_{\text{a}}$-formulas $\phi(x, \vec{y})$, where $x$ is a variable and $\vec{y}$ is a vector of variables, the following is an axiom:

$$[\phi(0, \vec{y}) \wedge \forall x(\phi(x, \vec{y}) \rightarrow \phi(x + 1, \vec{y}))] \rightarrow \forall x \phi(x, \vec{y}).$$

Clearly, $\mathbf{N} \vDash \mathsf{PA}$, where $\mathbf{N} = (\mathbb{N}, 0, S, +, \cdot)$.

(k) The Zermelo-Fraenkel set theory, ZFC, is a theory in the signature $\tau_{\text{set}} = (\in)$, in which all of the mathematics is derived. Its list of axiom schemas is a little too long to be listed here, so it is enough to mention that they express some basic facts about sets such as existence of unions, definable subsets, an infinite set, etc.

**Definition 2.24.** *A property $\Phi$ of $\tau$-structures is called axiomatizable if there is a $\tau$-theory $T$ such that for each $\tau$ structure $\mathbf{M}$*

$$M \textit{ has property } \Phi \iff \mathbf{M} \vDash T.$$

One can think of the property $\Phi$ as the class of $\tau$-structures satisfying that property. We showed above that for example the classes of infinite graphs, groups, algebraically closed fields, etc., are axiomatizable. However, we will show later on in the course that the class of connected graphs (as well as of disconnected graphs) is not axiomatizable (try proving that it is axiomatizable to see where the problem is).

## 2.5. **Elementarity**

In the signature $\tau_{\text{group}}$, a substructure of a group may not be a subgroup because not all elements might have inverses in the substructure. Even if it was a subgroup, it might disagree with the ambient group about the truth of statements like "being abelian" or "a particular element commutes with everybody" (they may be true in the subgroup, but false in the ambient group). The following definitions isolate those substructures which agree with the ambient structure on the statements about the elements of the substructure.

**Definition 2.25** (Elementary embedding)**.** *Let* $\mathbf{A}, \mathbf{B}$ *be* $\tau$*-structures. A map* $f : A \to B$ *is called an elementary (or an elementary embedding) if for all formulas* $\phi(\vec{x})$ *and tuples* $\vec{a} \in A^n$*,*

$$\mathbf{A} \vDash \phi(\vec{a}) \iff \mathbf{B} \vDash \phi(f(\vec{a})).$$

*If such* $f$ *exists, we say that* $\mathbf{A}$ *elementarily embeds into* $\mathbf{B}$ *and write* $\mathbf{A} \prec_e \mathbf{B}$*.*

Note that such $f$ is automatically an injective homomorphism.

**Definition 2.26** (Elementary substructure)**.** *A substructure* $\mathbf{A}$ *of a* $\tau$*-structure* $\mathbf{B}$ *is called elementary if the inclusion map is elementary. We denote this by* $\mathbf{A} \prec \mathbf{B}$*.*

**Proposition 2.27** (Tarski-Vaught test)**.** *Let* $\mathbf{A}$ *be a substructure of* $\mathbf{B}$*.* $\mathbf{A}$ *is an elementary substructure of* $\mathbf{B}$ *if and only if for every formula* $\phi(x, \vec{y})$ *and* $\vec{a} \in A^n$*,*

$$\mathbf{B} \vDash \exists x \phi(x, \vec{a}) \iff \exists a' \in A \text{ such that } \mathbf{B} \vDash \phi(a', \vec{a}).$$

*Proof.* Left as homework. □

Given a $\tau$-structure $(B)$ and $S \subseteq B$, we could define a substructure generated by $S$ as the smallest substructure containing $S$ mainly because intersection of substructures is still a substructure. However, intersection of elementary substructures may not be elementary. So we cannot define "the elementary substructure generated by $S$", however the following theorem brings us as close as possible.

**Theorem 2.28** (Löwenheim-Skolem)**.** *Let* $\mathbf{B}$ *be a* $\tau$*-structure and* $S \subseteq B$*. There exists* $\mathbf{A} \prec \mathbf{B}$ *with* $A \supseteq S$ *such that* $|A| \leq \max(|S|, |\tau|, \aleph_0{}^1)$*.*

*Proof.* Left as homework. □

**Definition 2.29** (Elementary equivalence)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $\tau$*-structures. We say that* $\mathbf{A}$ *and* $\mathbf{B}$ *are called elementarily equivalent and write* $\mathbf{A} \equiv \mathbf{B}$ *if* $\mathsf{Th}(\mathbf{A}) = \mathsf{Th}(\mathbf{B})$*.*

Note that isomorphic structures are elementarily equivalent (homework). However, the converse is false! For example, it is a homework problem to show that $(\mathbb{Q}, <)$ and $(\mathbb{R}, <)$ are elementarily equivalent, but they clearly cannot be isomorphic (simply because of cardinality considerations).

## 2.6. **Formal proofs**

So far, we have been dealing with the semantic (model-theoretic) aspect of FOL, i.e. structures/models, satisfiability, definability, etc. In this section we turn to the syntactic aspect, namely proof systems and formal proofs.

[1] $\aleph_0$ denotes the cardinality of $\mathbb{N}$.

We fix a signature $\tau$ for this subsection and everything below is assumed to be in this signature.

We need the following technical definition in order to state some of the axioms:

**Definition 2.30.** *Let $\phi$ be a formula and $t$ be a term. We say that $t$ is free for $v$ in $\phi$ if no variable in $t$ is quantified in $\phi$ and $v$ is not quantified in $\phi$. If $t$ is free for $v$ in $\phi$, we define $\phi(t/v)$ to be the formula obtained from $\phi$ by replacing all occurrences of $v$ by $t$.*

Thus whenever we write $\phi(t/v)$, it is assumed that $t$ is free for $v$ in $\phi$.

The following are the *axioms* (or *axiom schemes*) and *rules of inference* of $\mathbb{FOL}(\tau)$.

**Logical axioms**. For each $\tau$-formula $\phi, \psi, \chi$, and $\tau$-term $t$, we have:

Axioms for →:

(1) $\phi \to (\psi \to \phi)$
(2) $(\phi \to \psi) \to [(\phi \to (\psi \to \chi)) \to (\phi \to \chi)]$

Axiom for ¬:

(3) $(\phi \to \psi) \to [(\phi \to \neg\psi) \to \neg\phi]$
(4) $\neg\neg\phi \to \phi$

Axioms for ∧:

(5) $\phi \to [\psi \to (\phi \wedge \psi)]$
(6a) $(\phi \wedge \psi) \to \phi$; (6b) $(\phi \wedge \psi) \to \psi$

Axioms for ∨:

(7) $(\phi \to \chi) \to [(\psi \to \chi) \to ((\phi \vee \psi) \to \chi)]$
(8a) $\phi \to (\phi \vee \psi)$; (8b) $\psi \to (\phi \vee \psi)$

Quantifier axioms:

(9) $\forall v(\phi \to \psi) \to (\phi \to \forall v\psi)$ ($v$ does not occur in $\phi$)
(10) $\forall v\phi \to \phi(t/v)$ ($t$ is free for $v$ in $\phi$)
(11) $\phi(t/v) \to \exists v\phi$ ($t$ is free for $v$ in $\phi$)

**Axioms for equality**. For each $n$-ary relation symbol $R$ and $n$-ary function symbol $f$ in $\tau$, we have:

(12) $v = v$; $v = v' \to v' = v$; $(v = v' \wedge v' = v'') \to v = v''$
(13) $(\bigwedge_{i=1}^{n} v_i = w_i) \to (R(v_1, ..., v_n) \to R(w_1, ..., w_n))$
(14) $(\bigwedge_{i=1}^{n} v_i = w_i) \to (f(v_1, ..., v_n) = f(w_1, ..., w_n))$

**Rules of inference**. For each $\tau$-formula $\phi, \psi$ and $\tau$-term $t$, we have:

(15) Modus Ponens: $\phi, \phi \to \psi \implies \psi$
(16) Generalization: $\phi \implies \forall v\phi$ ($v$ is not quantified in $\phi$)
(17) $\exists$-elimination: $\phi \to \psi \implies \exists v\phi \to \psi$ ($v$ is not quantified in $\phi$ and does not occur in $\psi$)

The proof of the following lemma is an easy but tedious verification:

**Lemma 2.31.** *All of the axioms above are valid in every $\tau$-structure and the rules of inference preserve validity.*

**Definition 2.32** (Formal proof)**.** *Let $T$ be a theory and $\phi$ be a formula. A proof of $\phi$ from $T$ is a finite sequence $\phi_1, \phi_2, ...\phi_n$ of formulas such that $\phi_n = \phi$ and for each $i$*

***either*** *$\phi_i$ is an axiom of $\mathbb{FOL}(\tau)$,*
***or*** *$\phi_i \in T$,*
***or*** *$\phi_i$ follows from the previous $\phi_j$-s by one of the rules of inference, more precisely:*
*<u>either</u> for some $j, k < i$, $\phi_i$ is obtained from $\phi_j$ and $\phi_k$ by Modus Ponens,*
*<u>or</u> for some $j < i$, $\phi_i$ is obtained from $\phi_j$ by Generalization,*
*<u>or</u> for some $j < i$, $\phi_i$ is obtained from $\phi_j$ by $\exists$-elimination.*

We say that $T$ *proves* $\phi$ (or $\phi$ is proved in $T$) and write $T \vdash phi$ if there exists a proof of $\phi$ from $T$. When $T = \varnothing$, we just write $\vdash \phi$.

The following example illustrates formal proofs and how tedious (even hard) it can be to find formal proofs of statements that are "obviously" true.

**Example 2.33.** Here is a formal proof of $\theta \to \theta$ from the empty theory, for all formulas $\theta$:

(i) $(\theta \to (\theta \to \theta)) \to [(\theta \to ((\theta \to \theta) \to \theta)) \to (\theta \to \theta)]$ (axiom schema (2) for $\phi \equiv \chi \equiv \theta$ and $\psi \equiv (\theta \to \theta)$),
(ii) $\theta \to (\theta \to \theta)$ (axiom schema (1) for $\phi \equiv \psi \equiv \theta$),
(iii) $(\theta \to ((\theta \to \theta) \to \theta)) \to (\theta \to \theta)$ (Modus Ponens (i), (ii)),
(iv) $\theta \to ((\theta \to \theta) \to \theta)$ (axiom schema (1) for $\phi \equiv \theta$ and $\psi \equiv (\theta \to \theta)$),
(v) $\theta \to \theta$ (Modus Ponens (iii), (iv)).

The following proposition justifies why we introduced a proof system and formal proofs:

**Proposition 2.34** (Soundness)**.** *If $T \vdash \phi$ then $T \vDash \phi$.*

*Proof.* This follows by induction on the length of the formal proof of $\phi$ and Lemma 2.31. □

The next two propositions are again proved by induction on the length of the formal proof and we leave the (somewhat nontrivial) details as homework.

**Proposition 2.35** (Deduction theorem)**.** *For a theory $T$, a sentence $\chi$ and a formula $\phi$,*

$$T, \chi \vdash \phi \iff T \vdash \chi \to \phi.$$

Let $\mathcal{S}$ be a set of symbols neither of which is in $\tau$. Then we denote by $\tau(S)$ the extension of $\tau$ obtained by adding to it the symbols in $S$ as constant symbols. If $S = \{s_1, ..., s_n\}$ is finite, we just write $\tau(s_1, ..., s_n)$.

**Proposition 2.36** (Constant Substitution)**.** *Let $c$ be a symbol that is not in $\tau$ and let $v$ be free in a $\tau$-formula $\phi$. For a $\tau$-theory $T$,*

$$T \vdash \phi(c/v) \iff T \vdash \phi,$$

*where in the first statement $T$ is viewed as a $\tau(c)$-theory.*

## 3. Completeness of FOL and its consequences

Proposition 2.34 (the soundness of the proof system) says that if we have a "first order (finite) certificate" that something is true (is a syntactic consequence of $T$), then it is indeed true (in every model of $T$). What about the converse: is the validity of $\phi$ in every model of $T$ witnessed by an actual formal proof from $T$? If the answer to this question was no, mathematicians would appear in a pretty rough shape since it would be possible that some (first order) statement was true in every model of $T$ (e.g. Hilbert's Nullstellensatz for algebraically closed fields), but we would have no (first order) way of proving that. Fortunately, the answer is YES and that is the content of the Completeness Theorem to which this section is devoted.

### 3.1. Syntactic-semantic duality, completeness and compactness

We have already defined some syntactic and semantic notions for a theory $T$ such as $T \vdash \phi$, $T \vDash \phi$, $T$ is satisfiable. In this subsection, we define some more notions and draw analogies between semantic and syntactic ones. Finally, we state the Completeness theorem, which in my opinion should have been called the Syntactic-Semantic Duality theorem (I still don't understand why it is called Completeness).

Let $\top$ denote the sentence $\forall x(x = x)$ and set $\bot {=}{\equiv} \neg\top$.

**Definition 3.1.** *A $\tau$-theory $T$ is said to be*

- *consistent if there is no $\tau$-sentence $\phi$ such that $T \vdash \phi \wedge \neg\phi$;*
- *complete if for any $\tau$-sentence $\phi$, $T \vdash \phi$ or $T \vdash \neg\phi$;*
- *semantically complete if for any $\tau$-sentence $\phi$, $T \vDash \phi$ or $T \vDash \neg\phi$.*

Note that a satisfiable theory is consistent by the Soundness of the proof system. Also, any inconsistent theory is automatically complete because $\vdash \bot \to \phi$ for any $\tau$-formula $\phi$. Clearly, $\mathsf{Th}(\mathbf{M})$ is complete for any $\tau$-structure $\mathbf{M}$.

The following is a more convenient characterization of semantic completeness.

**Proposition 3.2** (Semantic completeness, rephrased)**.** *A $\tau$-theory $T$ is semantically complete if and only if for any $\mathbf{A}, \mathbf{B} \vDash T$, $\mathbf{A} \equiv B$.*

*Proof.* Left as homework. □

**Lemma 3.3** (About consistency)**.** *Let $T$ be a $\tau$-theory.*

*(a) $T$ is consistent if and only if there is a sentence $\chi$ such that $T \nvdash \chi$.*
*(b) $T$ is consistent if and only if every finite subset of $T$ is consistent.*
*(c) For any sentence $\chi$, $T \cup \{\chi\}$ is inconsistent if and only if $T \vdash \neg\chi$.*
*(d) If $T$ is consistent, then for any sentence $\chi$, at least one of $T \cup \{\chi\}$ and $T \cup \{\neg\chi\}$ is consistent.*
*(e) If $\exists v\phi(v)$ is a sentence, $T \cup \{\exists v\phi(v)\}$ is consistent, and $c$ is a constant symbol that does not occur in $T \cup \{\exists v\phi(v)\}$, then $T \cup \{\phi(c)\}$ is consistent.*

*Proof.* Part (a) just expresses the fact that once a theory proves a contradiction, then it proves every sentence. (b) follows from the fact that proofs are finite. We prove the rest in detail.

The right-to-left direction of (c) is immediate, and we show the other direction. Assume $T \cup \{\chi\}$ is inconsistent and hence $T, \chi \vdash \bot$. By the Deduction theorem (this is where we really need this theorem), $T \vdash \chi \to \bot$, and since $\vdash (\chi \to \bot) \to (\top \to \neg\chi)$ (it's a tautology, but one has to actually verify this), by Modus Ponens we get $T \vdash \top \to \neg\chi$. But $\top$ is an axiom, so $T \vdash \top$, and hence by applying Modus Ponens again, we get $T \vdash \neg\chi$.

For (d), we prove the contrapositive. Assume both $T \cup \{\chi\}$ and $T \cup \{\neg\chi\}$ are inconsistent. Then by (c), $T \vdash \neg\chi$ and $T \vdash \neg\neg\chi$. Thus $T \vdash \chi \wedge \neg\chi$ and hence is inconsistent.

For (e), we also prove the contrapositive. Assume $T \cup \{\phi(c)\}$ is inconsistent. Then by (c), $T \vdash \neg\chi(c)$. By the constant substitution lemma (2.36), $T \vdash \neg\phi(v)$, and by the Generalization rule, $T \vdash \forall v\neg\phi(v)$ and hence $T \vdash \neg\exists v\phi(v)$. Thus, by (c), $T \cup \{\exists v\phi(v)\}$ is inconsistent. □

Note the following "compactness" phenomenon: if $T \vdash \phi$, then there is a finite $T_0 \subseteq T$ with $T_0 \vdash \phi$. This is an immediate consequence of the fact that formal proofs are finite and hence they only use finitely many axioms from $T$. This "compactness" statement is actually equivalent to the fact that the following topological space is compact: let $\mathcal{T}$ be the set of all consistent complete theories and take the topology generated by the sets $<\phi>:=\{T\in\mathcal{T}:T\vdash\phi\}$, for all $\tau$-sentences $\phi$.

The following table compares the notions we have defined.

| Notions | **Syntactic** (Proof-theoretic) | **Semantic** (Model-theoretic) |
|---|---|---|
| **Consistency** | $T\nvdash\perp$ | $\exists$ (nonempty) $\mathbf{M}\vDash T$ |
| **Implication** | $T\vdash\phi$ | $T\vDash\phi$ |
| **Completeness** | $\forall\phi$, $T\vdash\phi$ or $T\vdash\neg\phi$ | $\forall\mathbf{A},\mathbf{B}\vDash T$, $\mathbf{A}\equiv\mathbf{B}$ |
| **Compactness** | $T\vdash\phi\implies\exists$ finite $T_0\subseteq T$, $T_0\vdash\phi$ | $T\vDash\phi\implies\exists$ finite $T_0\subseteq T$, $T_0\vDash\phi$ |

Although the statements in each row are clearly analogous, there is no immediate reason to think that they may be equivalent. For example, it is not clear at all whether the semantic version of the compactness statement is true. This is why one should appreciate the following

**Theorem 3.4** (Completeness of FOL; Gödel, 1929). *Any consistent $\tau$-theory $T$ is satisfiable. In fact, it has a model of cardinality at most* $\max\{|\tau|,\aleph_0\}$.

**A silly remark.** Completeness of FOL should NOT be confused with completeness of a theory; these are two completely different notions, they just use the same adjective (unfortunate terminology). I put this remark here because I have had students ask me whether Gödel's Completeness theorem contradicts his Incompleteness theorem. The first one means Completeness of FOL, the second means Incompleteness of PA (as a theory).

Before proceeding to the proof of this theorem, let us mention a couple of very important immediate corollaries.

**Corollary 3.5** (Syntactic-semantic duality). *The statements in each row of the above table are equivalent. In particular, for any $\tau$-sentence $\phi$,*

$$T\vdash\phi\iff T\vDash\phi.$$

*Proof.* We only prove that $T\vDash\phi$ implies $T\vdash\phi$ since the rest easily follows from it. We show the contrapositive. Suppose $T\nvdash\phi$, in particular $T$ is consistent (inconsistent theories prove everything). Moreover, $T\cup\{\neg\phi\}$ is consistent because otherwise, by Deduction theorem we would have $T\vdash\neg\phi\rightarrow\perp$, hence $T\vdash\top\rightarrow\phi$ (taking the contrapositive) and thus $T\vdash\phi$, a contradiction. Now, by the Completeness theorem, $T\cup\{\neg\phi\}$ has a model $\mathbf{M}$. Thus $\mathbf{M}\nvDash\phi$, and hence $T\nvDash\phi$. □

From now on, we will not differentiate between the notions of consistent theory and satisfiable theory, completeness and semantic completeness, etc.

**Remark.** If one somehow manages to prove a first-order statement $\phi$ about all models of $T$ using methods from outside of FOL, the syntactic-semantic duality implies that there is a first-order proof of $\phi$ from $T$ and using external methods was an overkill.

A theory is called *finitely satisfiable* if every finite subset of it is satisfiable. Rephrasing the semantic version of the compactness statement above, we get (probably) the most useful theorem of logic:

**Theorem 3.6** (Compactness). *If a $\tau$-theory $T$ is finitely satisfiable, then it is satisfiable. In fact, it has a model of cardinality at most* $\max\{|\tau|,\aleph_0\}$.

*Proof.* Because $T$ is finitely satisfiable, every finite subset of it is consistent. Hence $T$ is consistent and the Completeness theorem applies. □

The Compactness theorem has a wide range of applications and we will mention some of them in the upcoming lectures.

### 3.2. **Henkin's proof of Gödel's Completeness Theorem**

In this subsection we give a proof of Gödel's Completeness theorem that is due to Henkin.

**Definition 3.7.** *A $\tau$-theory $H$ is called a ($\tau$-)Henkin set if*
*(H1) $H$ is consistent,*
*(H2) for each $\tau$-sentence $\chi$, $\chi \in H$ or $\neg\chi \in H$,*
*(H3) if $\exists v \phi(v) \in H$, then there is a constant symbol $c$ in $\tau$ such that $\phi(c) \in H$.*

Note that the existence of a Henkin set implies that $\tau$ has at least one constant symbol. The constant $c$ in (H3) is called a *Henkin witness* for $\exists v \phi(v)$, so basically a Henkin set is a consistent (strongly) complete theory with Henkin witnesses.

Given a consistent theory $T$, it is easy to construct a consistent completion of it. A slight modification of this argument gives a construction of a Henkin set containing $T$.

Set $\kappa = \max\{|\tau|, \aleph_0\}$, where $\aleph_0 = |\mathbb{N}|$. Take a sequence $\mathcal{D} = \{d_i\}_{i<\kappa}$ of distinct constants that are not in $\tau$, and denote by $\bar{\tau}$ the extension of $\tau$ obtained by adding these new constant symbols.

**Lemma 3.8** (Constructing a Henkin set, uses the Axiom of Choice)**.** *If $\tau$ is consistent, then there exists a $\bar{\tau}$-Henkin set $H \supseteq T$.*

*Proof.* We will prove this assuming $\tau$ is countable (and hence $\kappa = |\mathbb{N}|$) to make the exposition easier to understand for those readers who are not familiar with the ordinals. However, the readers who are familiar are invited to prove this for general $\tau$. Since $\kappa$ is countable, we can write $\mathcal{D} = \{d_n\}_{n\in\mathbb{N}}$ for the sequence of new constant symbols.

**Claim 1.** *There is an enumeration*

$$\chi_0, \chi_1, \dots, \chi_n, \cdots$$

*of $\bar{\tau}$-sentences such that $d_n$ does not occur in $\chi_0, \dots, \chi_n$.*

*Proof of Claim.* Let $\mathcal{S}$ be the set containing all the symbols in $\tau$ and variables, i.e.

$$\mathcal{S} = \tau \cup \{v_n\}_{n\in\mathbb{N}}.$$

$\mathcal{S}$ is still countable, so fix an enumeration $\mathcal{S} = \{s_n\}_{n\in\mathbb{N}}$ (here we use the Axiom of Choice to well-order $S$ and it is the only place where we use it).

For $n \in \mathbb{N}$, let $L_n$ be the set of all $\bar{\tau}$-sentences of length $\leq 7 + n$ that use only symbols $s_0, \dots, s_n$ and $d_0, \dots, d_{n-1}$. Clearly we have

(i) $L_n \supset L_{n-1}$,
(ii) $L_n \setminus L_{n-1} \neq \varnothing$ because it contains $\forall v_n(v_n = v_n)$ (this sentence consists of 7 symbols, hence the choice of 7 above),
(iii) $L_n$ is finite,
(iv) $d_n$ does not occur in any sentence in $L_n$.

Since each $L_n$ is finite, we fix an order on $L_n \setminus L_{n-1}$ (one can use the lexicographical ordering on sentences induced by the orderings of $\mathcal{S}$ and $\mathcal{D}$, giving $\mathcal{S}$ the priority) and construct an enumeration as follows:

$$L_0, L_1 \setminus L_0, L_2 \setminus L_1, ...$$

Because each $L_n \setminus L_{n-1}$ is nonempty, it is clear that this enumeration of $\bar{\tau}$-sentences satisfies the claim. $\dashv$

Now we construct a sequence $\phi_0, \phi_1, ...$ of $\bar{\tau}$-sentences such that for each $n$,

(i) $\phi_{2n} \equiv \chi_n$ or $\phi_{2n} \equiv \neg\chi_n$,
(ii) if $\phi_{2n} \equiv \exists v \psi(v)$, then $\phi_{2n+1} \equiv \psi(d_n)$, otherwise $\phi_{2n+1} \equiv \phi_{2n}$,
(iii) $T \cup \{\phi_0, ..., \phi_{2n}\}$ is consistent.

The sequence $\phi_{2n}$ is defined by recursion on $n$ using (d) and (e) of the lemma about consistency (3.3) and $\phi_{2n+1}$ is uniquely determined by (ii).

It follows from (i)-(iii) that $H = \{\phi_n\}_{n\in\mathbb{N}}$ is a Henkin set. It remains to show that $T \subseteq H$. Indeed, let $\chi \in T$. Then $\chi \equiv \chi_n$, for some $n$, and hence $\phi_{2n} \equiv \chi_n$ or $\phi_{2n} \equiv \neg\chi_n$. But if $\phi_{2n} \equiv \neg\chi_n$, then $T \cup \{\phi_0, ..., \phi_{2n}\}$ is inconsistent, contradicting (iii). □

Thus it is enough to construct a model for the Henkin set $H$ provided by the last lemma and then take its reduct to the signature $\tau$.

**Lemma 3.9** (Constructing a model for a Henkin set). *If $H$ is a Henkin set in a signature $\sigma$, then it has a model. In fact, it has a model whose cardinality is at most the cardinality of the set of constants in $\sigma$.*

*Proof.* As our first attempt, we take the set of constant symbols $C$ of $\sigma$ as the universe of our future model $\mathbf{C}$ with the following interpretations: for all $e_1, ..., e_n, e \in C$,

$$\begin{array}{rcll} c^{\mathbf{C}} & = & c, & \text{for every constant symbol } c \text{ in } \sigma \\ R^{\mathbf{C}}(e_1, ..., e_n) & \iff & R(e_1, ..., e_n) \in H, & \text{for every } n\text{-ary relation symbol } R \text{ in } \sigma \\ f^{\mathbf{C}}(e_1, ..., e_n) = e & \iff & f(e_1, ..., e_n) = e \in H, & \text{for every } n\text{-ary function symbol } f \text{ in } \sigma. \end{array}$$

This construction almost works except that it may well be that $c = c' \in H$, for distinct constant symbols $c$ and $c'$ in $\bar{\mathbb{C}}$. Because of this, $\mathbf{C}$ is not even a $\sigma$-structure since the last clause defines a multi-valued function. Even if we managed to choose a single valued branch for $f^{\mathbf{C}}$, $\mathbf{C}$ would still not be a model of $H$ because it would not satisfy $c = c'$. So what we do is we mod out $C$ by the equivalence relation $c = c' \in H$. More precisely, for all $c, c' \in C$, define

$$c \sim c' \iff c = c' \in H.$$

It follows from Axioms (12) for equality that $\sim$ is an equivalence relation on $C$.

Put $M = C/\sim$, so $M = \{[c] : c \in C\}$, where $[c]$ denotes the equivalence class of $c$. We define a $\sigma$-structure $\mathbf{M}$ with universe $M$ and the following interpretations: for all $e_1, ..., e_n, e \in C$,

$$\begin{array}{rcll} c^{\mathbf{M}} & = & [c], & \text{for every constant symbol } c \text{ in } \sigma \\ R^{\mathbf{M}}([e_1], ..., [e_n]) & \iff & R(e_1, ..., e_n) \in H, & \text{for every } n\text{-ary relation symbol } R \text{ in } \sigma \\ f^{\mathbf{M}}([e_1], ..., [e_n]) = e & \iff & f(e_1, ..., e_n) = e \in H, & \text{for every } n\text{-ary function symbol } f \text{ in } \sigma. \end{array}$$

**Claim 1.** $\mathbf{M}$ *is well-defined.*

*Proof of Claim.* One has to prove that the definitions of $R^{\mathbf{M}}$ and $f^{\mathbf{M}}$ do not depend on the choice of the representatives of the equivalence classes. Moreover, for $f$, we need to show that for all $e_1, ..., e_n \in C$, there exists a unique up to $\sim$ $e \in C$ such that $f(e_1, ..., e_n) = e \in H$.

The verification of these statements is left to the reader. To show that there exists such $e$ argue as follows: $H \vdash f(e_1,...,e_n) = f(e_1,...,e_n)$. By the $\exists$-elimination rule applied to $\phi(f(e_1,...,e_n)/v)$, where $\phi(v) \equiv f(e_1,...,e_n) = v$, we get $H \vdash \exists v(f(e_1,...,e_n) = v)$. Since $H$ is a Henkin set, there is a Henkin witness $e$ for $\exists v(f(e_1,...,e_n) = v) \in H$, so $f(e_1,...,e_n) = e \in H$. $\dashv$

**Claim 2.** *For every $\sigma$-term $t$ with no variables, there exists $c \in C$ such that $t = c \in H$ and $\mathbf{M} \vDash t = [c]$.*

*Proof of Claim.* We do induction on the construction (length) of $t$. The case of $t$ being a variable is excluded, so the only base case is $t \equiv e$, where $e$ is a constant symbol of $\sigma$. Then take $c \equiv e$, and clearly $e = e \in H$ since $\vdash e = e$. Also, by the way we defined the interpretation for constants, $\mathbf{M} \vDash e = [e]$.

Now assume that $t = f(t_1,...,t_n)$. By the induction hypothesis, we have $c_1,...,c_n \in C$ such that $t_i = c_i \in H$ and $\mathbf{M} \vDash t_i = [c_i]$, for $i = 1,...,n$. Thus by Axiom (14), $H \vdash f(t_1,...,t_n) = f(c_1,...,c_n)$. Also, because $f^{\mathbf{M}}$ is a function, $\mathbf{M} \vDash f(t_1,...,t_n) = f(c_1,...,c_n)$. Now, by the definition of the interpretation of function symbols, there is $e \in C$ such that $f(c_1,...,c_n) = e \in H$ and $\mathbf{M} \vDash f([c_1],...,[c_n]) = [e]$. Thus, by Axiom (12), $f(t_1,...,t_n) = e \in H$, and by transitivity of equality, $\mathbf{M} \vDash f(t_1,...,t_n) = [e]$. $\dashv$

**Claim 3.** $\mathbf{M} \vDash H$.

*Proof of Claim.* We show that for every $\sigma$-formula $\phi$ and $c_1,...,c_n \in C$,

$$\mathbf{M} \vDash \phi([c_1],...,[c_n]) \iff \phi(c_1,...,c_n) \in H$$

by structural induction on the construction of $\phi$. The cases of equality and a relation symbol are handled by applying Claim 2 and Axioms (12)-(13). The cases of $\neg$ and $\wedge$ follow trivially from the induction hypothesis, and we handle the case of $\phi([c_1],...,[c_n]) \equiv \exists v \psi([c_1],...,[c_n],v)$ as follows:

$$\begin{aligned} \mathbf{M} \vDash \phi([c_1],...,[c_n]) &\iff \text{there is } [b] \in M \text{ such that } \mathbf{M} \vDash \psi([c_1],...,[c_n],[b]) \\ &\iff \text{there is } b \in C \text{ such that } \psi(c_1,...,c_n,b) \in H \text{ (by induction)} \\ &\iff \exists v \phi(c_1,...,c_n) \in H, \end{aligned}$$

where in the last equivalence, $\Longrightarrow$ is by $\exists$-elimination rule, and $\Longleftarrow$ is because there are Henkin witnesses. $\dashv$

The last claim finishes the proof of the lemma. □

*Proof of the Completeness Theorem 3.4* (Henkin, 1949). By Lemma 3.8, there is a $\bar{\tau}$-Henkin set $H \supseteq T$. Now applying Lemma 3.9 to $\sigma = \bar{\tau}$ and $H$, we get a model $\mathbf{M}$ of $H$ of cardinality at most $|\sigma|$ and hence at most $\kappa = \max\{|\tau|, \aleph_0\}$. Finally, take the reduct of $\mathbf{M}$ to the signature $\tau$. □

### 3.3. The Skolem "paradox" and weak Löwenheim-Skolem theorem

The Completeness theorem has the following striking consequence: if ZFC is consistent (which we really hope it is), then it has a countable model. This is maybe strange because that countable model $\mathbf{M}$ believes that there is an uncountable set since Cantor's theorem that $\mathbb{R}$ is uncountable is true in $\mathbf{M}$. Does this imply that ZFC is inconsistent?

The answer is of course NO and here are the two reasons why (the main reason is (2)):

(1) It may well be that $M = \mathbb{N}$ with a binary relation $\in^{\mathbf{M}}$ defined on it. So what if somehow $\mathbf{M}$ satisfies the statement that reads "there is an uncountable set"? It is just some statement about this binary relation $\in^{\mathbf{M}}$ and it does not imply anything about the actual sets and the cardinality of $M$.

(2) Even if $M$ was a set of sets and $\in^{\mathbf{M}}$ was the true $\in$, then the countability of $M$ would simply imply that $\mathbf{M}$'s version of the real numbers, $\mathbb{R}^{\mathbf{M}}$, is indeed countable (for us), i.e. there is a bijection of $\mathbb{R}^{\mathbf{M}}$ with $\mathbb{N}$. This bijection is a set (any function is a set of pairs), but it may not be an element of $M$. In fact, since $\mathbf{M}$ satisfies the statement "$\mathbb{R}^{\mathbf{M}}$ is uncountable", we conclude that NO bijection of $\mathbb{R}^{\mathbf{M}}$ with $\mathbb{N}$ is an element of $M$. In other words, $M$ does not "see" the countability of $\mathbb{R}^{\mathbf{M}}$ and thus thinks that $\mathbb{R}^{\mathbf{M}}$ is uncountable. It's like how people thought the world was endless before they discovered it was round since all they could see was the ocean up to the line of the horizon and for all they knew it continued forever. It was also not too long ago that we still thought the universe was infinite until we discovered the big bang theory. The difference is that we eventually obtained this (perhaps still questionable) knowledge, while $\mathbf{M}$ never will.

The following is a general statement about cardinalities of models.

**Theorem 3.10** (Löwenheim-Skolem, weak version)**.** *If a $\tau$-theory $T$ has an infinite model, then it has a model of any cardinality $\kappa \geq \max\{|\tau|, \aleph_0\}$.*

*Proof.* Put $\bar{\tau} = \tau \cup \{c_\alpha\}_{\alpha<\kappa}$, where $c_\alpha$ are constant symbols that are not in $\tau$. Define

$$T' = T \cup \{c_\alpha \neq c_\beta : \alpha \neq \beta, \alpha, \beta < \kappa\}.$$

$T'$ is finitely satisfiable since it has an infinite model. Thus, by the Compactness theorem, $T'$ has a model $\mathbf{M}$ of cardinality at most $\kappa$ since $|\bar{\tau}| = \kappa \geq \aleph_0$. On the other hand, $|M| \geq \kappa$ since $c_\alpha^{\mathbf{M}} \neq c_\beta^{\mathbf{M}}$ for distinct $\alpha, \beta < \kappa$. Thus $|M| = \kappa$. □

This theorem implies for example that PA has uncountable models!

## 3.4. Nonstandard models of arithmetic

A *nonstandard model of Peano arithmetic* is any model of PA that is not isomorphic to $\mathbf{N} = (\mathbb{N}, 0, S, +, \cdot)$. As mentioned above, PA has uncountable models and hence they are nonstandard. In this subsection we construct a countable nonstandard model of PA.

For the rest of the subsection we work in the signature $\tau_{\mathrm{a}} = (0, S, +, \cdot)$ of arithmetic.

For each $n \in \mathbb{N}$, recursively define a $\tau_{\mathrm{a}}$-term $\Delta(n)$ as follows:

$$\begin{cases} \Delta(0) \equiv 0 \\ \Delta(n+1) =\equiv S(\Delta(n)) \end{cases}.$$

Note that for every $n \in \mathbb{N}$, $\mathbf{N} \models \Delta(n) = n$ and hence $\mathbb{N} = \{\Delta(n)^{\mathbf{N}} : n \in \mathbb{N}\}$.

**Proposition 3.11.** *There is a countable nonstandard model of* PA.

*Proof.* Let $w$ be a new constant symbol not in $\tau_{\mathrm{a}}$ and consider the extension $\sigma = \tau_{\mathrm{a}} \cup \{w\}$. Put

$$T = \mathsf{PA} \cup \{w \neq \Delta(n) : n \in \mathbb{N}\}.$$

$T$ is finitely satisfiable because for any finite $T_0 \subseteq T$, letting $n$ be the maximum number with $w \neq \Delta(n) \in T_0$, the expansion of $\mathbb{N}$ to a $\sigma$-structure with $w$ being interpreted as $n+1$ satisfies $T_0$. Thus, by the Compactness theorem, $T$ has a countable model $\mathbf{M}$.

To see that this $\mathbf{M}$ is nonstandard, assume for contradiction that there is an isomorphism $f : \mathbf{N} \to \mathbf{M}$. Since $f(\Delta(n)^{\mathbf{N}}) = \Delta(n)^{\mathbf{M}}$, $f[\mathbb{N}] = \{\Delta(n)^{\mathbf{M}} : n \in \mathbb{N}\}$. But then $w^{\mathbf{M}} \notin f[\mathbb{N}]$ and thus $f$ is not surjective, a contradiction. □

### 3.5. Applications to combinatorics

In this section we give one application of the Compactness theorem to combinatorics, but many more are given as homework problems.

The following is the most basic and well known theorem of infinite combinatorics. For a set $S$, let $[S]^2$ denote the set of two element subsets of $S$ (think of it as the set of edges of the undirected complete graph on $S$). Given a 2-coloring of $[\mathbb{N}]^2$, i.e. a function $c : [\mathbb{N}]^2 \to \{0,1\}$, a set $E \subseteq [\mathbb{N}]^2$ is said to be monochromatic if all elements of $E$ have the same color, i.e. $c\downharpoonright_E$ is constant. A set $A \subseteq \mathbb{N}$ is called monochromatic if $[A]^2$ is monochromatic.

**Theorem 3.12** (Infinite Ramsey)**.** *For any 2-coloring of $[\mathbb{N}]^2$, there exists an infinite monochromatic subset of $\mathbb{N}$.*

*Proof.* For $a \in \mathbb{N}$ and $A \subseteq \mathbb{N}$, put $(a, A) = \{\{a, a'\} : a' \in A \setminus \{a\}\}$. Set $A_0 = \mathbb{N}$ and take sequences $a_n \in \mathbb{N}$ and $A_n \subseteq \mathbb{N}$ satisfying:

(i) $a_n \in A_n$,
(ii) $A_{n+1} \subseteq A_n$ is infinite and $(a_n, A_{n+1})$ is monochromatic.

It is easy to see that such sequences $(a_n)_{n\in\mathbb{N}}$ and $(A_n)_{n\in\mathbb{N}}$ exist (define them recursively). Call $a_n$ red if all elements of $(a_n, A_{n+1})$ have color 0, otherwise call it blue. Clearly, there is a subsequence $(a_{n_k})_{k\in\mathbb{N}}$ with all $a_{n_k}$ having the same color (red or blue). Now it is straightforward to check that $A = \{a_{n_k}\}_{k\in\mathbb{N}}$ is monochromatic. □

We now derive the Finite Ramsey theorem from this using the Compactness theorem. The original combinatorial proof is much messier (look it up).

Let $\bar{n} = \{0, 1, ..., n-1\}$.

**Theorem 3.13** (Finite Ramsey)**.** *For every $m \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that for any 2-coloring of $[\bar{n}]^2$, there exists a monochromatic subset $A \subseteq \bar{n}$ of cardinality $m$.*

*Proof.* Let $\tau$ be the signature containing constant symbols $c_n$, for every $n \in \mathbb{N}$, and a binary relation symbol $R$ (think of $R$ as a symbol for coloring: the color of $\{x, y\}$ is 1 if $R(x, y)$ and 0 otherwise). Fix $m \in \mathbb{N}$, and for each $n \in \mathbb{N}$, let $\phi_n$ be a $\tau$-sentence expressing that $\{c_0, c_1, ..., c_{n-1}\}$ does not have a monochromatic subset of cardinality $m$ (there are only finitely many such subsets, so we can express it).

Now assume for contradiction that for any $n$, there is a 2-coloring of $[\bar{n}]^2$ such that $\bar{n}$ has no monochromatic subsets of cardinality $m$. Thus the theory $T = \{\phi_n : n \in \mathbb{N}\}$ is finitely satisfiable and hence has a model $\mathbf{M}$. Let $C = \{c_n^{\mathbf{M}} : n \in \mathbb{N}\}$. By Infinite Ramsey theorem, $C$ has an infinite monochromatic subset $A$, i.e. either for all distinct $a, a' \in A$, $R^{\mathbf{M}}(a, a')$ or for all distinct $a, a' \in A$, $\neg R^{\mathbf{M}}(a, a')$. Let $n$ be large enough so that $A \cap \{c_i : i < n\}$ has at least $m$ elements. Then it is clear that $\mathbf{M} \nvDash \phi_n$, a contradiction. □

**A silly remark.** Mathematicians sometimes refer to this kind of arguments as "compactness and contradiction arguments". Now that we have learnt the Compactness theorem, we can just say "by the Compactness theorem" as opposed to "by a compactness and contradiction argument".

## 4. Complete theories

As mentioned above, it is easy to see that every consistent theory has a (consistent) completion. So why don't we only consider complete theories and not have to deal with the issues that come with incomplete theories? For example, why don't we just work with $\mathsf{Th}(\mathbf{N})$ instead of $\mathsf{PA}$? The problem is that it is hard (in a very precise sense) to check whether a given statement is an axiom of $\mathsf{Th}(\mathbf{N})$ or not. For example, is the Twin Prime Conjecture in $\mathsf{Th}(\mathbf{N})$? We wish we knew. The whole point of mathematics is to derive complicated statements from "easy-to-verify" axioms. We will see in the next section that "easy-to-verify" means that we can write a computer program that checks whether a given sentence is an axiom or not. For example, all of the theories in Examples 2.23 satisfy this criterion.

Now the question is: having defined some reasonable theory, like $\mathsf{ACF}_p$, is it complete? In other words, are these axioms enough to capture the first-order essence of say algebraically closed fields of characteristic $p$? In this section we develop a sufficient condition for verifying completeness, using which we show that $\mathsf{ACF}_p$ is complete.

### 4.1. **The Łoś-Vaught test**

**Definition 4.1.** *Let $\kappa$ be a cardinal. A $\tau$-theory $T$ is called $\kappa$-categorical if any two models of $T$ of cardinality $\kappa$ are isomorphic. We say that $T$ is uncountably categorical if it is $\kappa$-categorical for some uncountable cardinal $\kappa$.*

For example, the theory of vector spaces over $\mathbb{Q}$ is uncountably categorical; in fact, it is $\kappa$-categorical, for every uncountable cardinal $\kappa$. This is by virtue of the fact that every vector space has a basis and to construct an isomorphism between vector spaces it is enough to find a bijection between their bases. We will see shortly that a similar argument shows that $\mathsf{ACF}_p$ is $\kappa$-categorical as well (for every uncountable cardinal $\kappa$).

**Proposition 4.2** (Łoś-Vaught test)**.** *Let $T$ be a $\tau$-theory that has an infinite model. If $T$ is $\kappa$-categorical for some $\kappa \geq \max\{|\tau|, \aleph_0\}$, then $T$ is complete.*

*Proof.* Let $\mathbf{A}, \mathbf{B} \vDash T$ and we need to show that $\mathbf{A} \equiv \mathbf{B}$, by Proposition 3.2. By the weak version of Löwenheim-Skolem (3.10), there are $\mathbf{A}' \vDash \mathsf{Th}(\mathbf{A})$ and $\mathbf{B}' \vDash \mathsf{Th}(\mathbf{B})$ such that $|A'| = \kappa = |B'|$. Since $T$ is $\kappa$-categorical, $\mathbf{A}' \cong \mathbf{B}'$ and hence $\mathbf{A}' \equiv \mathbf{B}'$. Thus $\mathbf{A} \equiv \mathbf{A}' \equiv \mathbf{B}' \equiv \mathbf{B}$. □

This immediately gives that the theory of vector spaces over $\mathbb{Q}$ is complete.

One cannot help mentioning the following very important theorem that started the modern model theory:

**Theorem Morley, 1965.** *Let $T$ be a theory in a countable signature $\tau$. If $T$ is uncountably categorical, then it is $\kappa$-categorical for every uncountable cardinal $\kappa$.*

Thus it is not a coincidence that the theory of vector spaces is $\kappa$-categorical for all uncountable cardinals $\kappa$. The proof of this theorem is far outside the realm of this course, but it is worth mentioning that the most important ingredient of it is showing that if a structure is such that all of its definable sets are either finite or cofinite (complement is finite), then it admits a "basis" similar to the vector space basis, and so one can use the same argument as for vector spaces to construct isomorphisms.

Lastly, we would like to mention the following long standing open problem that, although being model-theoretic in nature, has been best understood (but not completely solved) in the context of descriptive set theory:

**Vaught's conjecture.** Let $\tau$ be a countable signature and $T$ be a complete $\tau$-theory having infinite models. If $T$ has uncountably many nonisomorphic countable models, does it have continuum many nonisomorphic countable models?

## 4.2. **Algebraically closed fields and the Lefschetz Principle**

We now aim at satisfying the conditions of the Łoś-Vaught test for $\mathsf{ACF}_p$.

**Lemma 4.3.** *Every algebraically closed field is infinite.*

*Proof.* For any finite field $F = \{a_1, ..., a_n\}$, the polynomial $(x - a_1)(x - a_2)...(x - a_n) + 1$ does not have a root in $F$. Thus $F$ is not algebraically closed. □

The proof of the following is similar to that of the theory of vector spaces being uncountably categorical, and can be safely omitted by the reader if (s)he does not feel like remembering field theory.

**Proposition 4.4.** *For $p$ prime or $0$, $\mathsf{ACF}_p$ is $\kappa$-categorical for any uncountable cardinal $\kappa$.*

*Proof.* Let $\mathbf{K}_1, \mathbf{K}_2 \vDash \mathsf{ACF}_p$ with $|K_1| = |K_2| = \kappa$. For $i = 1, 2$, let $F_i$ be the base field of $\mathbf{K}_i$, i.e. the substructures of $\mathbf{K}_i$ generated by $\varnothing$. (If $p = 0$, then $F_i$ is a copy of $\mathbb{Q}$; otherwise it is a copy of $\mathbb{Z}/p\mathbb{Z}$.) Since $F_1$ and $F_2$ are clearly isomorphic (as rings), we can assume without loss of generality that $F_1 = F_2 =: F$. Let $B_i$ be transcendence base over $F$ in $\mathbf{K}_i$. (Transcendence base is a maximal collection of algebraically independent elements over $F$.) Now it is not hard to see that $K_i = \overline{F(B_i)}$, where $F(B_i)$ denotes the field generated by $B_i$ over $F$ and $\overline{F(B_i)}$ denotes its algebraic closure in $K_i$.

Thus, because $F$ is countable, $|K_i| = |B_i| \cdot \aleph_0 + |F|$. If $B_i$ is countable then so is $|B_i| \cdot \aleph_0$, but $K_i$ is uncountable, and hence $B_i$ is uncountable. Then, by basic cardinal arithmetic, $|B_i| \cdot \aleph_0 + |F| = |B_i|$ and so $\kappa = |K_i| = |B_i|$. Hence, there is a bijection $f : B_1 \to B_2$. This $f$ uniquely extends to an isomorphism of $F(B_1)$ onto $F(B_2)$, which in its turn extends (not necessarily uniquely) to an isomorphism of $K_1 = \overline{F(B_1)}$ onto $K_2 = \overline{F(B_2)}$. □

**Corollary 4.5.** $\mathsf{ACF}_p$ *is complete, for any $p$ prime or $0$.*

*Proof.* Follows from 4.3, 4.4 and the Łoś-Vaught test (4.2). □

The following was once just a principle (a belief) in algebraic geometry, but it was later on formalized and turned into a theorem by A. Robinson (who was by the way a professor at UCLA):

**Theorem 4.6** (Lefschetz Principle)**.** *Let $\mathbf{C} = (\mathbb{C}, 0, 1, +, -, \cdot)$. For a $\tau_{ring}$-sentence $\phi$ the following are equivalent:*

*(1)* $\mathbf{C} \vDash \phi$.
*(2)* $\mathbf{K} \vDash \phi$, *for some* $\mathbf{K} \vDash \mathsf{ACF}_0$.
*(3)* $\mathsf{ACF}_0 \vDash \phi$.
*(4) For sufficiently large primes $p$,* $\mathsf{ACF}_p \vDash \phi$.
*(5) For infinitely many primes $p$, there is* $\mathbf{K} \vDash \mathsf{ACF}_p$ *such that* $\mathbf{K} \vDash \phi$.

*Proof.* (1) $\iff$ (2) $\iff$ (3): Follows from the completeness of $\mathsf{ACF}_0$.
(3) $\implies$ (4): $\mathsf{ACF}_0 \vDash \phi$ implies $\mathsf{ACF}_0 \vdash \phi$ by the Completeness theorem. Hence, because proofs are finite, there is a finite $T \subseteq \mathsf{ACF}_0$ such that $T \vdash \phi$. But then, by the definitions of $\mathsf{ACF}_0$ and $\mathsf{ACF}_p$, for sufficiently large prime $p$, $T \subseteq \mathsf{ACF}_p$. Thus $\mathsf{ACF}_p \vdash \phi$ and hence $\mathsf{ACF}_p \vDash \phi$.

(4) $\implies$ (5): Trivial.

(5) $\implies$ (3): We prove the contrapositive: assume (3) fails. But then $\mathsf{ACF}_0 \vDash \neg\phi$ and hence, by (3) $\implies$ (4), for sufficiently large primes $p$, $\mathsf{ACF}_p \vDash \neg\phi$. Therefore (5) is false. □

### 4.3. **Reducts of arithmetic**

**Definition 4.7.** *Let $T$ be a $\tau$-theory. A $\tau$-theory $T'$ is called an axiomatization for $T$ if for all $\tau$-sentences,*

$$T \vdash \tau \iff T' \vdash \phi.$$

$\mathsf{PA}$ was constructed as an attempt to "conveniently" axiomatize $\mathsf{Th}(\mathbf{N})$, where "convenient" means that there is a computer program recognizing the axioms (we will make this more in the next section). However, as we will see, Gödel's Incompleteness theorem states that $\mathsf{PA}$ is incomplete. In fact, there is no convenient axiomatization for $\mathsf{Th}(\mathbf{N})$, i.e. any subtheory $T \subseteq \mathsf{Th}(\mathbf{N})$ is either incomplete or inconvenient.

What about reducts of $\mathbf{N}$? Does the theory of $(\mathbb{N}, 0, S)$ or even of $(\mathbb{N}, 0, S, +)$ admit a convenient axiomatization? In other words, where is the boundary of incompleteness? It turns out that unlike $\mathbf{N}$, the theories of $(\mathbb{N}, 0, S)$ and $(\mathbb{N}, 0, S, +)$ admit convenient axiomatizations, and this is what we will focus on in this subsection.

We start with $\mathbf{N}_S := (\mathbb{N}, 0, S)$. Let $\tau_S = (0, S)$. Here is our first (and last) attempt of axiomatizing $\mathsf{Th}(\mathbf{N}_S)$. Let theory $T_S$ consist of the following axioms:

(S1) Zero has no predecessor: $\forall x(S(x) \neq 0)$.
(S2) The successor function is one-to-one: $\forall x \forall y(S(x) = S(y) \to x = y)$.
(S3) Any nonzero number is a successor of something: $\forall x(x \neq 0 \to \exists y(x = S(y)))$.
(S4) For all $n \in \mathbb{N}$, there are no $n$-loops: $\forall x(S^n(x) \neq x)$, where $S^n$ stands for the $n$-fold composition of $S$.

Note that (S4) is an axiom schema, i.e. it contains an axiom for every $n \in \mathbb{N}$; in particular, $T_S$ is infinite.

It is clear that any model $\mathbf{M}$ of $T_S$ has a standard part $\bar{\mathbb{N}} = \{\Delta(n)^{\mathbf{M}} : n \in \mathbb{N}\}$, where $\Delta(n) := S^n(0)$. Define a binary relation $\sim$ on $M$ as follows: for all $a, b \in M$,

$$a \sim b \iff \text{if for some } n \in \mathbb{N},\ \mathbf{M} \vDash S^n(a) = b \text{ or } \mathbf{M} \vDash S^n(b) = a.$$

If $a$ is standard, i.e. $a \in \bar{\mathbb{N}}$, then the equivalence class $[a]$ of $a$ is exactly $\bar{\mathbb{N}}$. If $a \in M$ is nonstandard, then $[a]$ does not have a least element (why?) and hence looks like a $\mathbb{Z}$-chain:

$$\ldots \to * \to a \to S^{\mathbf{M}}(a) \to S^{\mathbf{M}}(S^{\mathbf{M}}(a)) \to \ldots$$

Thus $\mathbf{M}$ is a union of $\bar{\mathbb{N}}$ and a bunch of $\mathbb{Z}$-chains. Let $\Lambda_{\mathbf{M}}$ denote the set of $\mathbb{Z}$-chains in $\mathbf{M}$ and put $\lambda_{\mathbf{M}} = |\Lambda_{\mathbf{M}}|$. Then $|M| = |\mathbb{N}| + \lambda_{\mathbf{M}} \cdot |\mathbb{Z}|$ and hence, by basic cardinal arithmetic, $M$ has cardinality $\lambda_{\mathbf{M}}$ unless $\lambda_{\mathbf{M}}$ is finite, i.e. $|M| = \max\{\lambda_{\mathbf{M}}, \aleph_0\}$. In particular, if $M$ is uncountable, then $|M| = \lambda_{\mathbf{M}}$.

**Proposition 4.8.** *$T_S$ is $\kappa$-categorical, for any uncountable cardinal $\kappa$.*

*Proof.* Let $\mathbf{A}, \mathbf{B} \models T_S$ with $|A| = |B| = \kappa$. By above, $\lambda_{\mathbf{A}} = |A| = \kappa = |B| = \lambda_{\mathbf{B}}$. Thus, there is a bijection $f : \Lambda_{\mathbf{A}} \to \Lambda_{\mathbf{B}}$. Now the standard parts of $\mathbf{A}$ and $\mathbf{B}$ are clearly isomorphic. Moreover, any $\mathbb{Z}$-chain $C \in \Lambda_{\mathbf{A}}$ is isomorphic to $f(C)$ because any two $\mathbb{Z}$-chains are clearly isomorphic. Thus, combining all these individual isomorphisms together, we get an isomorphism of $\mathbf{A}$ onto $\mathbf{B}$. □

From this and the Łoś-Vaught test, we get

**Corollary 4.9.** *$T_S$ is complete.*

Now we turn to $\mathbf{N}_+ := (\mathbb{N}, 0, S, +)$. Let $\tau_+ = (0, S, +)$ and let $T_+$ be the theory consisting of all of the axioms of $\mathsf{PA}$ except for the ones involving multiplication (hence it is a convenient theory). The proof of the following theorem will be omitted since it uses the technique of quantifier elimination, which is not covered in these notes.

**Theorem 4.10** (Presburger, 1929). *$T_+$ is complete.*

Thus, as we will see, the incompleteness phenomenon starts with $\mathbf{N} = (\mathbb{N}, 0, S, +, \cdot)$.

## 5. Incomplete theories

We start with an informal definition, which we will formalize later on.

**Definition 5.1** (Informal). *A $\tau$-theory $T$ is called recursive if there is a computer program such that given a $\tau$-sentence $\phi$, it returns YES if $\phi \in T$, and NO otherwise.*

We saw in the previous section that the theories of $(\mathbb{N}, 0, S)$ and $(\mathbb{N}, 0, S, +)$ admit primitive recursive axiomatizations. However, the situation changes once we add multiplication because it enables prime numbers and makes it possible to code tuples of natural numbers into a single number, and we have the following ground-breaking theorem:

**Theorem 5.2** (Incompleteness; Gödel, 1931). *Any recursive theory $T \subseteq \mathsf{Th}(\mathbf{N})$ is incomplete. In particular, $\mathsf{PA}$ is incomplete.*

This section is devoted to the proof of several versions of this theorem and some of its consequences, as well as making the definition of primitive recursive precise.

### 5.1. Sketch of proof of the Incompleteness theorem

Below, we sketch the proof of the Incompleteness theorem stated above to make the idea of the proof apparent and not get lost in the technical details that one has to go through in order to rigorously prove the theorem.

**Definition 5.3** (Informal). *A function $f : \mathbb{N}^k \to \mathbb{N}$ is called recursive if there is a computer program such that given $\vec{a} \in \mathbb{N}^k$ as input, it outputs $f(\vec{a})$. A set/relation $A \subseteq \mathbb{N}^k$ is called recursive if so is its indicator function.*

First thing one shows is that recursive functions are arithmetical. Thus any function we can write a computer program for is expressible in the language of arithmetic.

For a finite signature $\tau$, whose symbols are $s_0, \ldots s_n$ we enumerate the symbols of $\mathbb{FOL}(\tau)$ as follows:

$$s_0 \; s_1 \; \ldots \; s_n \; = \neg \; \wedge \; \vee \; \to \; \forall \; \exists \; , \; ( \; ) \; v_0, \; v_1, \; v_2, \; \ldots$$

and call the index of a symbol its *code*. For example, the code of $s_0$ is 0, the code of $=$ is $n+1$ and the code of $v_i$ is $n+11+i$. Using prime numbers and the fact that prime number factorization is unique, we can code a tuple of natural numbers into a single natural number ($<n_1, ..., n_k>= p_1^{n_1+1} \cdot ... \cdot p_k^{n_k+1}$), and so we can code formulas since they are just tuples of symbols of $\mathbb{FOL}(\tau)$. In fact, we can make sure that the coding and decoding operations are recursive (think of computer programs that would do this).

Thus let $\ulcorner t \urcorner$ and $\ulcorner \phi \urcorner$ denote the codes of a $\tau$-term and a $\tau$-formula $\phi$, respectively. It is now not hard to see that a $\tau$-theory $T$ is recursive if and only if the set of codes of its axioms is recursive (as a subset of $\mathbb{N}$).

Now let $\tau$ be the signature of arithmetic, i.e. $\tau = \tau_{\mathrm{a}}$, and thus we have the above coding since $\tau_{\mathrm{a}}$ is finite. For every $n \in \mathbb{N}$, set $\Delta(n) = S^n(0)$. It is tedious but straightforward to show that there is a recursive function $\mathsf{Sub}_0 : \mathbb{N}^2 \to \mathbb{N}$ such that for any $\tau_{\mathrm{a}}$-formula $\phi$ in which $v_0$ is not quantified, and for any $m \in \mathbb{N}$,

$$\mathsf{Sub}_0(\ulcorner \phi \urcorner, m) = \ulcorner \phi(\Delta m / v_0) \urcorner.$$

In words, this function takes $m$ and the code of $\phi$, and returns the code of the formula obtained from $\phi$ by replacing all occurrences of $v_0$ by the term $\Delta m$.

As mentioned above, all recursive functions are arithmetical. Hence, there is a $\tau_{\mathrm{a}}$-formula $\mathbf{Sub}_0(x, y, z)$ such that for all $a, b, c \in \mathbb{N}$,

$$\mathsf{Sub}_0(a, b) = c \iff \mathbf{N} \models \mathbf{Sub}_0(a, b, c).$$

Without loss of generality, we can assume $v_0$ is not quantified in $\mathbf{Sub}_0(x, y, z)$.

**Lemma 5.4** (Fixed point for **N**). *For each $\tau_a$ formula $\phi(v)$ there is a $\tau_a$-sentence $\theta$ such that*

$$\mathbf{N} \models \theta \leftrightarrow \phi(\ulcorner \theta \urcorner).$$

*Proof.* Put $\psi(v_0) \equiv \exists z(\mathbf{Sub}_0(v_0, v_0, z) \wedge \phi(z))$ and $e = \ulcorner \psi(v_0) \urcorner$. Now we feed $\psi(v_0)$ its own code by letting $\theta \equiv \psi(\Delta e)$, and thus $\mathsf{Sub}_0(e, e) = \ulcorner \psi(\Delta(e)) \urcorner = \ulcorner \theta \urcorner$. Now magic happens:

$$\begin{aligned} \mathbf{N} \models \theta &\iff \mathbf{N} \models \psi(e) \\ &\iff \mathbf{N} \models \exists z(\mathbf{Sub}_0(e, e, z) \wedge \phi(z)) \\ &\iff \text{there exists } b \in \mathbb{N} \text{ such that } b = \mathrm{sub}(e, e) \text{ and } \mathbf{N} \models \phi(b) \\ &\iff \mathbf{N} \models \phi(\ulcorner \theta \urcorner). \end{aligned}$$

If you feel cheated, join the club. □

This lemma says that every unary arithmetical relation $\phi(v)$ asserts of (the code of) some sentence $\theta$ exactly what $\theta$ asserts about **N**. It enables self-reference in the language of arithmetic, using which we can express the Liar Paradox (i.e. Cantor's diagonalization method), which is what lies at the heart of the proof of the Incompleteness theorem.

As an immediate corollary we get the following result that is actually stronger than the Gödel's Incompleteness theorem:

**Theorem 5.5** (Tarski, 1939). *$Th(\mathbf{N})$ is not arithmetical, i.e. the set $\ulcorner \mathsf{Th}(\mathbf{N}) \urcorner := \{\ulcorner \phi \urcorner : \phi \in \mathsf{Th}(\mathbf{N})\}$ is not definable in* **N**.

*Proof.* Left as a homework problem. □

Because formal proofs are just finite sequences of formulas, we can code them using the operation of coding $n$-tuples. Given a *recursive* $\tau_{\mathrm{a}}$-theory $T$, it is straightforward to check that the following relation is recursive if such is $T$: for $a, e \in \mathbb{N}$,

$$\mathsf{Proof}_T(a, e) \iff a \text{ is a code of a } \tau_{\mathrm{a}}\text{-formula } \phi \text{ and } e \text{ is a code of a proof of } \phi \text{ from } T.$$

To write a program for this, one has to check the definition of the formal proof, i.e. that every formula in the finite sequence coded by $e$ is either an axiom of $\mathbb{FOL}(\tau_{\mathrm{a}})$, or belongs to $T$ (this is where we need $T$ to be recursive), or can be obtained from the previous formulas in the sequence by applying one of the three operations: Modus Ponens, Generalization or $\exists$-elimination.

As before, since all recursive functions are arithmetical, there is a $\tau_{\mathrm{a}}$-formula $\mathbf{Proof}_T(x, y)$ such that for all $a, b \in \mathbb{N}$,

$$\mathsf{Proof}_T(a, b) \iff \mathbf{N} \models \mathbf{Proof}_T(a, b).$$

Given this, we have a $\tau_{\mathrm{a}}$-formula defining the relation of provability in $\mathbf{N}$:

$$\mathbf{Provable}_T(x) \equiv \exists y \mathbf{Proof}_T(x, y),$$

and hence, for any $\tau$-formula $\phi$,

$$\phi \text{ is provable in } T \iff \mathbf{N} \models \mathbf{Provable}_T(\ulcorner \phi \urcorner).$$

*Proof of the Incompleteness theorem 5.2.* Applying the Fixed Point lemma to

$$\phi(v) \equiv \neg \mathbf{Provable}_T(v),$$

we get a $\tau_{\mathrm{a}}$-sentence $\gamma_T$ such that

$$\mathbf{N} \models \gamma_T \leftrightarrow \neg \mathbf{Provable}_T(\ulcorner \gamma_T \urcorner).$$

The *Gödel sentence* $\gamma_T$ says about itself that it is not provable in $T$ (just like in the Liar Paradox, the liar says "I am a liar"). Hence, we have

$$\begin{aligned} \mathbf{N} \models \gamma_T &\iff \mathbf{N} \models \neg \mathbf{Provable}_T(\ulcorner \gamma_T \urcorner) \\ &\iff \text{for all } e \in \mathbb{N}, \mathbf{N} \models \neg \mathbf{Proof}_T(\ulcorner \gamma_T \urcorner, e) \\ &\iff \text{for all } e \in \mathbb{N}, e \text{ is not a code of a proof of } \gamma_T \\ &\iff T \nvdash \gamma_T. \end{aligned}$$

This equivalence implies that $\mathbf{N} \models \gamma_T$ since otherwise, $T \vdash \gamma_T$ and $\gamma_T \notin \mathsf{Th}(\mathbb{N})$, which contradicts $T \subseteq \mathsf{Th}(\mathbf{N})$. Thus, again by the equivalence above, $T \nvdash \gamma_T$. It also cannot be that $T \vdash \neg \gamma_T$ since $\mathbf{N} \models T$ and $\mathbf{N} \models \gamma_T$. □

Here is another proof of the Incompleteness theorem that is shorter but nonconstructive:

*Another proof of the Incompleteness theorem 5.2.* If $T$ was recursive and complete, then the formula $\mathbf{Provable}_T(x)$ would define the set $\ulcorner \mathsf{Th}(\mathbf{N}) \urcorner$ in $\mathbf{N}$ because, by the completeness of $T$, for every sentence $\phi$, $\phi$ is provable from $T$ if and only if $\ulcorner \phi \urcorner \in \ulcorner \mathsf{Th}(\mathbf{N}) \urcorner$. Thus $\ulcorner \mathsf{Th}(\mathbf{N}) \urcorner$ would be arithmetical, contradicting Tarski's theorem (5.5). □

For the rest of the section, we will be occupied with making the notion of recursive precise and developing tools for proving a stronger version of Gödel's Incompleteness theorem that applies not only to subtheories of $\mathsf{Th}(\mathbf{N})$, but also to theories (in an arbitrary finite signature $\tau$), which have $\mathsf{PA}$ "encoded" in them; for example, $\mathsf{PA} \cup \{\neg \gamma_{\mathsf{PA}}\}$ and ZFC.

### 5.2. **A quick introduction to recursion theory**

In this subsection we give a model (of computation) to capture intuitive notions such as algorithm, computable functions, etc. It is a general belief, known as the Church-Turing thesis, that this model captures the mentioned notions pretty well. One evidence of it is that it is very robust in the sense that all other seemingly different models of computation that people had defined turned out to be equivalent.

**Definition 5.6** (Search operation)**.** *For a unary relation $R \subseteq \mathbb{N}$, define $\mu x(R(x))$ as the smallest $x \in \mathbb{N}$ for which $R(x)$ holds, if such $x$ exists, and it is undefined, otherwise. In the latter case, we write $\mu x(R(x)) =\perp$.*

For example, $\mu x(x^2 > 7) = 3$. This operation is also called *minimalization*.

**Definition 5.7** (Recursive functions)**.** *A function $f : \mathbb{N}^k \to \mathbb{N}$ is called recursive (or computable) if it is obtained by inductively applying the following rules:*

*(R1)*
- $+ : \mathbb{N}^2 \to \mathbb{N}$ *and* $\cdot : \mathbb{N}^2 \to \mathbb{N}$ *are recursive;*
- $\chi_\leq : \mathbb{N}^2 \to \mathbb{N}$ *is recursive, where $\chi_\leq$ is the characteristic function of $\leq$, i.e. $\chi_\leq(x, y) = 1$ if $x \leq y$, and $0$, otherwise;*
- *The projection functions $P_i^n(x_1, ..., x_n) = x_i$ are recursive, for all $i = 1, ..., n$ and $n \in \mathbb{N}$;*

*(R2) Composition: if $g : \mathbb{N}^m \to \mathbb{N}$ and $h_1, ..., h_m : \mathbb{N}^k \to \mathbb{N}$ are recursive, then so is the composition function $f = g(h_1, ..., h_2) : \mathbb{N}^k \to \mathbb{N}$ defined by*

$$f(\vec{a}) = g(h_1(\vec{a}), ..., h_m(\vec{a}));$$

*(R3) Well-defined search: if $g : \mathbb{N}^{n+1} \to \mathbb{N}$ is recursive and for all $\vec{a} \in \mathbb{N}^n$ there is $x \in \mathbb{N}$ with $g(\vec{a}, x) = 0$, then the function $f : \mathbb{N}^n \to \mathbb{N}$ defined by*

$$f(\vec{a}) = \mu x(g(\vec{a}, x) = 0)$$

*is recursive.*

*A relation $R \subseteq \mathbb{N}^n$ is called recursive if so is its characteristic function $\chi_R : \mathbb{N}^n \to \mathbb{N}$.*

Although the class of recursive functions is obtained by closing the set of functions in (R1) under operations (R2) and (R3), it is closed under many other operations. The most important among those is the operation of *primitive recursion*, which is often included in the definition of recursive functions. However, we prefer showing that it is a consequence of the definition rather than including it in the latter since keeping the definition minimalistic makes it easier to prove that the class of recursive functions is contained in other classes of functions (less cases to consider).

The following proposition provides some closure properties of the class of recursive functions together with some examples.

**Lemma 5.8.**

*(a) The relations $\geq, =$ are recursive.*
*(b) Constant functions $C_k^n : \mathbb{N}^n \to \mathbb{N}$ are recursive, where $C_k^n(\vec{a}) = k$, for all $\vec{a} \in \mathbb{N}^n$.*
*(c) The successor function $S : \mathbb{N} \to \mathbb{N}$ is recursive.*
*(d) If $n$-ary relations $P, Q$ on $\mathbb{N}^n$ are recursive, then so are the following*

$$\neg P := \mathbb{N}^n \smallsetminus P, P \wedge Q := P \cap Q, P \vee Q := P \cup Q.$$

(e) *(Definition by Cases) Let $R_1$, ..., $R_k \subseteq \mathbb{N}^n$ be recursive such that for each $\vec{a} \in \mathbb{N}^n$ exactly one of $R_1(\vec{a}), ..., R_k(\vec{a})$ holds, and suppose that $g_1, ..., g_k : \mathbb{N}^n \to \mathbb{N}$ are recursive. Then $g : \mathbb{N}^n \to \mathbb{N}$ given by*

$$g(\vec{a}) = \begin{cases} g_1(\vec{a}) & \text{if } R_1(\vec{a}) \\ \vdots & \vdots \\ g_k(\vec{a}) & \text{if } R_k(\vec{a}) \end{cases}.$$

*is recursive.*

*Proof.* For (a), let note that $\chi_{\geq}(x,y) = \chi_{\leq}(P^2_2(x,y), P^2_1(x,y))$ and $\chi_{=}(x,y) = \chi_{\leq}(x,y) \cdot \chi_{\geq}(x,y)$.

We prove (b) by induction on $k$. For $k = 0$, observe that $c^n_0(\vec{a}) = \mu x(P^{n+1}_{n+1}(\vec{a}, x) = 0)$. Assume $c^n_k$ is recursive and note that

$$c^n_{k+1}(\vec{a}) = \mu x(c^n_k(\vec{a}) < x) = \mu x(\chi_{\geq}(c^{n+1}_k(\vec{a}, x), P^{n+1}_{n+1}(\vec{a}, x)) = 0).$$

For (c), just note that $S(a) = a + c^1_1(a)$.

For (d), observe that $\neg P(\vec{a}) \iff \chi_P(\vec{a}) = c^n_0(\vec{a})$ and $\chi_{P \wedge Q}(\vec{a}) = \chi_P(\vec{a}) \cdot \chi_Q(\vec{a})$. Thus $\neg P$ and $P \wedge Q$ are recursive if so are $P$ and $Q$. Recursiveness of the rest of the Boolean combinations follows from this because they are expressible in terms of $\wedge$ and $\neg$.

Part (e) is left to the reader. □

**Lemma 5.9.** *Let $R \subseteq \mathbb{N}^{n+1}$ be recursive such that for all $\vec{a} \in \mathbb{N}^n$ there exists $x \in \mathbb{N}$ with $(\vec{a}, x) \in R$. Then the function $f : \mathbb{N}^n \to \mathbb{N}$ given by*

$$f(\vec{a}) = \mu x R(\vec{a}, x)$$

*is recursive.*

*Proof.* Note that $f(\vec{a}) = \mu x(\chi_{\neg R}(\vec{a}, x) = 0)$. □

Using this we get the following convenient property for verifying recursiveness of functions:

**Proposition 5.10** (Graph property)**.** *Let $f : \mathbb{N}^n \to \mathbb{N}$. Then $f$ is recursive if and only if so is its graph (as a subset of $\mathbb{N}^{n+1}$).*

*Proof.* let $R \subseteq \mathbb{N}^{n+1}$ be the graph of $f$. Then for all $\vec{a} \in N^n$ and $b \in \mathbb{N}$,

$$R(\vec{a}, b) \iff f(\vec{a}) = b,$$

and hence

$$f(\vec{a}) = \mu x R(\vec{a}, x),$$

from which the proposition follows immediately. □

**Definition 5.11** (Primitive recursion)**.** *Let $g : \mathbb{N}^k \to \mathbb{N}$ and $h : \mathbb{N}^{k+2} \to \mathbb{N}$. We say that $f : \mathbb{N}^{k+1} \to \mathbb{N}$ is defined by primitive recursion from $g, h$ if for all $\vec{a} \in \mathbb{N}^n$ and $n \in \mathbb{N}$,*

$$f(\vec{a}, 0) = g(\vec{a})$$
$$f(\vec{a}, n+1) = h(\vec{a}, n, f(\vec{a}, n))$$

We aim at showing that the class of recursive functions is closed under this operation. For that, we first convert the recursive definition into an explicit (iterative) one as follows.

**Proposition 5.12** (Dedekind's analysis of recursion). *If $f : \mathbb{N}^{k+1} \to \mathbb{N}$ is defined by primitive recursion from $g, h$ as in 5.11, then for all $\vec{a} \in \mathbb{N}^k$, $n \in \mathbb{N}$ and $w \in \mathbb{N}$,*

$$\begin{aligned} f(\vec{a}, n) = w \iff & \textit{there exists a sequence } (w_0, ..., w_n) \textit{ such that} \\ & w_0 = g(\vec{a}) \wedge (\forall i < n)[w_{i+1} = h(\vec{a}, i, w_i)] \wedge w_n = w. \end{aligned}$$

*Proof.* Obvious. □

To be able to express the right hand side of Dedekind's analysis of recursion, we need to be able to recursively code and decode tuples of natural numbers of arbitrary length into a single natural number. We do it using the

**Chinese Remainder Theorem 5.13.** *Let $d_0, ..., d_n$ be pairwise coprime and put $d = d_0 d_1 ... d_n$. Then the natural projection map*

$$h : \mathbb{Z}/d\mathbb{Z} \to \mathbb{Z}/d_0\mathbb{Z} \times ... \times \mathbb{Z}/d_n\mathbb{Z}$$

*defined by*

$$[a]_d \mapsto ([a]_{d_0}, ..., [a]_{d_n})$$

*is a well-defined group isomorphism.*

*Proof.* That $h$ is well-defined follows from the fact that every $d_i$ divides $d$, and that $h$ is a homomorphism follows from the fact that the remainder function respects addition. Since the groups on the left and right of the homomorphism have the same number of elements, by Pigeon Hole Principle, we only have to show that $h$ is injective. To this end, assume that $h([a]_d) = 0$. Thus every $d_i$ divides $a$ and hence $d$ divides $a$ because $d_i$ are pairwise coprime. Therefore, $[a]_d = 0$ and hence $\ker(h)$ is trivial. □

**Lemma 5.14.**

(a) *If relation $R \subseteq \mathbb{N}^{k+1}$ is recursive, then so are the relations*

$$P(\vec{a}, y) \iff \exists x_{<y} R(\vec{a}, x), Q(\vec{a}, y) \iff \forall x_{<y} R(\vec{a}, x),$$

*for all $\vec{a} \in \mathbb{N}^k$, $y \in \mathbb{N}$.*

(b) *The function $\dot{-} : \mathbb{N}^2 \to \mathbb{N}$ defined by $n \dot{-} m = \max\{n - m, 0\}$ is recursive.*

(c) *The remainder function* $\mathsf{Rem} : \mathbb{N}^2 \to \mathbb{N}$, *defined by $(a, b) \mapsto$ the remainder of $a$ when divided by $b$, is recursive.*

(d) *The function* $\mathsf{Pair} : \mathbb{N}^2 \to \mathbb{N}$ *defined by*

$$(x, y) \to \frac{(x + y)(x + y + 1)}{2} + x$$

*is a recursive bijection.*

(e) *The functions* $\mathsf{Left}$, $\mathsf{Light} : \mathbb{N} \to \mathbb{N}$ *defined by*

$$\mathsf{Pair}(x, y) = z \iff \mathsf{Left}(z) = x \wedge \mathsf{Right}(z) = y$$

*are recursive.*

*Proof.* We leave parts (a),(b) and (c) to the reader. For (d), $\mathsf{Pair}(x, y) = \mu z(2z \dot{-} (x + y)(x + y + 1) = 0) + x$ and hence is recursive. It is a bijection because it enumerates pairs $(x, y)$ as follows:

$$\underbrace{(0,0)}_{x+y=0}\,\underbrace{(0,1)(1,0)}_{x+y=1}\,\underbrace{(0,2)(1,1)(2,0)}_{x+y=2}\ldots$$

For (e), observe that $\mathsf{Left}(z) = \mu x(\exists y_{<z+1} \mathsf{Pair}(x, y) = z)$ and similarly for $\mathsf{Right}$. □

**Lemma 5.15** (Gödel's $\beta$-function)**.** *The function $\beta : \mathbb{N}^2 \to \mathbb{N}$ defined by*

$$\beta(w, i) = \mathit{Rem}(\mathit{Left}(w), 1 + (i + 1)\mathit{Right}(w))$$

*is recursive and has the property that for every sequence $(w_0, ..., w_n)$, there exists $w \in \mathbb{N}$ such that for all $i \leq n$,*

$$\beta(w, i) = w_i.$$

*Proof.* The fact that $\beta$ is recursive follows from 5.14, so we prove the second statement. Let $s = \max\{n, w_0, w_1, ..., w_n\}$, set $b = s!$ and verify that

$$d_0 = 1 + (0 + 1)b, d_1 = 1 + (1 + 1)b, ..., d_n = 1 + (n + 1)b$$

are pairwise coprime as follows: if a prime $p$ divides $1 + (i + 1)b$ and $1 + (j + 1)b$, for $i < j$, then it divides their difference $(j - i)b = (j - i)s!$. Since $j - i \leq n \leq s$, $p$ must divide $s! = b$, contradicting $p$ dividing $1 + (i + 1)b$.

By the Chinese Remainder Theorem, there is $a < d_0 \cdot ... \cdot d_n$ such that $\mathsf{Rem}(a, d_i) = w_i$. Thus setting $w = \mathsf{Pair}(a, b)$, we get

$$w_i = \mathsf{Rem}(a, d_i) = \mathsf{Rem}(\mathsf{Left}(w), 1 + (i + 1)\mathsf{Right}(w)) = \beta(w, i).$$

□

Using Gödel's $\beta$-function, we define the following coding/decoding tuples functions, which are clearly recursive:

- $<a_0, ..., a_{n-1}>:= \mu x(\beta(x, 0) = n \wedge \bigwedge_{i=1}^{n} \beta(x, i) = a_{i-1})$. Note that $<>= 0$ (as a nullary function).
- $\mathsf{lh} : \mathbb{N} \to \mathbb{N}$ by $\mathsf{lh}(a) = \beta(a, 0)$.
- $(a)_i := \beta(a, i + 1)$. Note that $(<a_0, ..., a_{n-1}>)_i = a_i$.
- $\mathsf{InitSeg}(a, i) = \mu x(\mathsf{lh}(x) = i \wedge \forall j_{<i}(x)_j = (a)_j)$. Thus $\mathsf{InitSeg}(< a_0, ..., a_n >, i) =< a_0, ...a_{i-1}>$.
- $a * b = \mu x(\mathsf{lh}(x) = \mathsf{lh}(a) + \mathsf{lh}(b) \wedge \forall i_{<\mathsf{lh}(a)})(x)_i = (a)_i \wedge \forall i_{<\mathsf{lh}(b)})(x)_{\mathsf{lh}(a)+i} = (b)_i$. Thus $<a_0, ...a_{n-1}> * <b_0, ...b_{m-1}>=<a_0, ...a_{n-1}, b_0, ..., b_{m-1}>$.

**Proposition 5.16.** *Recursive functions are closed under the operation of primitive recursion, i.e. if $g, h, f$ are as in Definition 5.11 and $g, h$ are recursive, then so is $f$.*

*Proof.* We implement Dedekind's analysis of recursion as follows. Define an auxiliary function $\tilde{f} : \mathbb{N}^{k+1} \to \mathbb{N}$ by

$$\tilde{f}(\vec{a}, n) = \mu x(\mathsf{lh}(x) = n + 1 \wedge (x)_0 = g(\vec{a}) \wedge \forall i_{<n}(x)_{i+1} = h(\vec{a}, i, (x)_i)),$$

and note that $f(\vec{a}, n) = (\tilde{f}(\vec{a}, n))_n$. Since $\tilde{f}$ is clearly recursive, so is $f$. □

Primitive recursion enables us to show that any function that admits a recursive definition is recursive. E.g. $n \to 2^n$ is recursive because

$$\begin{cases} 2^0 & = & 1 \\ 2^{n+1} & = & 2 \cdot 2^n \end{cases}.$$

We now define a nice subclass of recursive functions, namely that of *primitive recursive* functions, which is still rich enough to contain most of the functions that can be implemented as computer programs. In fact, most of the recursive functions mentioned so far are actually primitive recursive.

**Definition 5.17** (Primitive recursive functions)**.** *The class of primitive recursive functions is the smallest class containing the successor function $S : \mathbb{N} \to \mathbb{N}$, the constant functions $C^n_k : \mathbb{N}^n \to \mathbb{N}$, $k, n \in \mathbb{N}$ and the projection functions $P^n_i(x_1, ..., x_n) = x_i$, $i \leq n, n \in \mathbb{N}$, and is closed under composition and primitive recursion. A relation $R \subseteq \mathbb{N}^n$ is called primitive recursive if so is its characteristic function $\chi_R : \mathbb{N}^n \to \mathbb{N}$.*

The reader can verify that the functions in (R1) of the definition of recursive functions are primitive recursive. It is also easy to check that Lemma 5.8 holds with *recursive* replaced by *primitive recursive.*

The following makes it easy to verify that Lemmas 5.14 and 5.15 also hold with *recursive* replaced by *primitive recursive.*

**Lemma 5.18** (Bounded search)**.** *Let $R \subseteq \mathbb{N}^{n+1}$ be a recursive relation. Then the function $f : \mathbb{N}^{n+1} \to \mathbb{N}$ defined by $f(\vec{a}, y) = \mu x_{<y} R(\vec{a}, x)$ is primitive recursive, where*

$$\mu x_{<y} R(\vec{a}, x) = \begin{cases} \mu x R(\vec{a}, x) & \textit{if } \exists x_{<y} R(\vec{a}, x) \\ y & \textit{otherwise} \end{cases} .$$

*Proof.* We define $f(\vec{a}, y)$ by primitive recursion as follows: let $f(\vec{a}, 0) = 0$ and

$$f(\vec{a}, y+1) = \begin{cases} f(\vec{a}, y) & \text{if } f(\vec{a}, y) < y \\ y & \text{if } f(\vec{a}, y) = y \wedge R(\vec{a}, y) \\ y+1 & \text{otherwise} \end{cases} .$$

□

The proof of 5.15 yields a primitive recursive function $B : \mathbb{N} \to \mathbb{N}$, defined by $B(N) = \prod_{i<n}(1 + (1+i)N!)$, such that for every $n \in \mathbb{N}$ and $\vec{a} \in \mathbb{N}^n$,

*whenever $N \geq \max\{N, a_0, ..., a_{n-1}\}$, there is $a < B(N)$ such that $\beta(a, i) = a_i$, $\forall i < n$.*

Using this together with 5.18 one can easily show that the coding/decoding functions $<a_0, ..., a_{n-1}>$, $\mathsf{lh}(a)$, $(a)_i$, $\mathsf{InitSeg}(a, i)$, $a * b$ are primitive recursive.

The following lemma allows recursive definitions using all previously computed values of a function as opposed to only the last computed value.

**Lemma 5.19** (Complete primitive recursion)**.** *For $f : \mathbb{N}^{n+1} \to \mathbb{N}$, let*

$$\bar{f}(\vec{a}, n) =<f(\vec{a}, 0), ..., f(\vec{a}, n-1)> .$$

*Then:*

*(a) $f$ is primitive recursive if and only if $\bar{f}$ is primitive recursive.*
*(b) If $g : \mathbb{N}^{k+1} \to \mathbb{N}$ is primitive recursive, then so is $f : \mathbb{N}^{k+1} \to \mathbb{N}$ defined by $f(\vec{a}, n) = g(\vec{a}, \bar{f}(\vec{a}, n))$.*

*Proof.* We prove part (a) and leave (b) to the reader.
$\Leftarrow$: Put $f(\vec{a}, n) = (\bar{f}(\vec{a}, n+1))_n$.
$\Rightarrow$: We define $\bar{f}(\vec{a}, n)$ by primitive recursion as follows:

$$\begin{cases} \bar{f}(\vec{a}, 0) &= <> \\ \bar{f}(\vec{a}, n+1) &= \bar{f}(\vec{a}, n)* <f(\vec{a}, n)> \end{cases} .$$

□

One may ask if there are any recursive functions that are not primitive recursive. The answer is YES (of course) and here is why:

**Proposition 5.20.** *There exists a recursive function $\phi : \mathbb{N}^2 \to \mathbb{N}$ such that $\phi_n := \phi(n, \cdot)$ enumerates all the primitive recursive functions (possibly with repetitions), i.e. for every $n$, $\phi_n$ is primitive recursive and for every primitive recursive function $f$, there is $n$ such that $f = \phi_n$. Moreover, any such function $\phi$ is not primitive recursive.*

*Proof.* A proof of the existence of such $\phi$ is outlined in one of the homework problems and here we show that such $\phi$ is not primitive recursive by applying Cantor's diagonalization[2] method. Assume for contradiction that $\phi$ is primitive recursive. Then so is the function $\psi(n) = \phi(n,n) + 1$, for all $n$, and thus there is $n_0 \in \mathbb{N}$ such that $\phi_{n_0} = \psi$. But then we have

$$\psi(n_0) = \phi_{n_0}(n_0) = \phi(n_0, n_0)$$

on one hand, and

$$\psi(n_0) = \phi(n_0, n_0) + 1$$

on the other, which is a contradiction. □

Note that the same proof shows that there is no recursive enumeration of recursive functions. Similarly, the set of codes of recursive functions is not recursive, i.e. there is no recursive binary relation $R$ such that for any unary recursive relation $Q$ there is $n$ such that for all $x$,

$$Q(x) \iff R(n, x).$$

This is known as the undecidability of the *halting problem.*

Here is a more concrete and important example of a recursive function that is not primitive recursive:

**Definition 5.21** (Ackermann function)**.** *Ackermann function is the function $A : \mathbb{N}^2 \to \mathbb{N}$ inductively defined as follows:*

$$\begin{cases} A(0,x) & = & x+1 \\ A(n+1,0) & = & A(n,1) \\ A(n+1,x+1) & = & A(n, A(n+1,x)) \end{cases}.$$

The proof that this function is recursive but not primitive recursive is left as a homework problem together with the proof that the graph of this function is primitive recursive. The last fact shows that the graph property (Proposition 5.10) does not hold for primitive recursive functions.

## 5.3. **Representability in Robinson's system Q**

In the sketch of the proof of the Incompleteness theorem above, we used the fact that recursive functions are arithmetical, i.e. definable in $\mathbf{N}$. Thus the proof only applied to theories that $\mathbb{N}$ satisfies. If we want to prove incompleteness for other theories, like $\mathsf{PA} \cup \{\neg\gamma_{\mathsf{PA}}\}$, we have to develop a notion of definability inside a theory rather than a structure. This is what the following definition is supposed to capture.

**Definition 5.22** (Representability)**.** *Let $T$ be a theory in the signature $\tau_a$ of arithmetic.*

[2] As van den Dries suggests, perhaps *antidiagonalization* would be a better name.

- *We say that a relation $R \subseteq \mathbb{N}^n$ is representable in $T$ if there is a formula $\phi(\vec{x})$ such that for all $\vec{a} \in \mathbb{N}^n$,*
$$R(\vec{a}) \implies T \vdash \phi(\Delta(\vec{a})) \text{ and } \neg R(\vec{a}) \implies T \vdash \neg\phi(\Delta(\vec{a})),$$
*where $\Delta(\vec{a}) = (\Delta(a_1), ..., \Delta(a_n))$. Such $\phi$ is said to represent the relation $R$ in $T$.*
- *We say that a function $f : \mathbb{N}^n \to \mathbb{N}$ is representable in $T$ if there is a formula $\phi(\vec{x}, y)$ such that for all $\vec{a} \in \mathbb{N}^n$,*
$$T \vdash \phi(\Delta(\vec{a}), y) \leftrightarrow y = \Delta(f(\vec{a})).$$
*Such $\phi$ is said to represent the function $f$ in $T$.*
- *A $\tau_a$-term $t(\vec{x})$ is said to represent the function $f : \mathbb{N}^n \to \mathbb{N}$ in $T$ if for all $\vec{a} \in \mathbb{N}^n$,*
$$T \vdash t(\Delta(\vec{a})) = \Delta(f(\vec{a})).$$

The following shows that we could have defined representability of relations using that of functions (not the other way around). Below we use the expression "arguing in models" to mean that we prove something for every model of a theory and then conclude that the theory proves it by the Completeness theorem.

**Lemma 5.23.** *If $T$ is a $\tau_a$-theory such that $T \vdash 0 \neq S(0)$ and $R \subseteq \mathbb{N}^n$, then*
$$R \text{ is representable in } T \text{ if and only if } \chi_R \text{ is representable in } T.$$

*Proof.* $\Rightarrow$: Let $\phi(\vec{x})$ represent $R$ in $T$ and put
$$\psi(\vec{x}, y) \equiv (\phi(\vec{x}) \wedge y = S(0)) \vee (\neg\phi(\vec{x}) \wedge y = 0).$$
We show that $\psi(\vec{x}, y)$ represents $\chi_R$ in $T$, that is: for all $\vec{a} \in \mathbb{N}^n$,
$$T \vdash \psi(\Delta(\vec{a}), y) \leftrightarrow y = \chi_R(\Delta(\vec{a})).$$
Assume $\vec{a} \in R$. Then $T \vdash \phi(\Delta(\vec{a}))$ and thus, arguing in models of $T$,
$$(\phi(\Delta(\vec{a})) \wedge y = S(0)) \vee (\neg\phi(\Delta(\vec{a})) \wedge y = 0)$$
holds if and only if $y = S(0)$ holds (here is where we use that $T \vdash 0 \neq S(0)$). Thus, by the Completeness theorem, $T \vdash \psi(\Delta(\vec{a}), y) \leftrightarrow y = S(0)$. Similarly, one shows that for $\vec{a} \notin R$, $T \vdash \psi(\Delta(\vec{a}), y) \leftrightarrow y = 0$.
$\Leftarrow$: Let $\phi(\vec{x}, y)$ represent $\chi_R$ and put
$$\psi(\vec{x}) \equiv \phi(\vec{x}, S(0)).$$
We show that $\psi(\vec{x})$ represents $R$ in $T$. For every $\vec{a} \in \mathbb{N}^n$,
$$\begin{aligned} R(\vec{a}) &\implies \chi_R(\vec{a}) = 1 \\ &\implies T \vdash \phi(\Delta(\vec{a}), y) \leftrightarrow y = S(0) \\ &\implies T \vdash \phi(\Delta(\vec{a}), S(0)) \qquad \text{(substitute } y = S(0)) \\ &\implies T \vdash \psi(\Delta(\vec{a})). \end{aligned}$$
Similarly,
$$\begin{aligned} \neg R(\vec{a}) &\implies \chi_R(\vec{a}) = 0 \\ &\implies T \vdash \phi(\Delta(\vec{a}), y) \leftrightarrow y = 0 \\ &\implies T \vdash \neg\phi(\Delta(\vec{a}), S(0)) \qquad \text{(substitute } y = S(0)) \\ &\implies T \vdash \neg\psi(\Delta(\vec{a})). \end{aligned}$$
□

Now we describe a finite subtheory of $Th(\mathbb{N})$, namely Robinson's[3] system $\mathsf{Q}$, which is much weaker than $\mathsf{PA}$, but still rich enough to represent recursive functions. The advantage of it over $\mathsf{PA}$ is that it is finite, and we will use this later in proving that the empty $\tau_{\mathrm{a}}$-theory is undecidable.

**Definition 5.24** (Robinson's system $\mathsf{Q}$)**.** *The following are the axioms of* $\mathsf{Q}$*:*
*(*$\mathsf{Q}$*1)* $\forall x[\neg S(x) = 0]$,
*(*$\mathsf{Q}$*2)* $\forall x \forall y[S(x) = S(y) \to x = y]$,
*(*$\mathsf{Q}$*3)* $\forall x[x + 0 = x]$,
*(*$\mathsf{Q}$*4)* $\forall x \forall y[S(x + y) = x + S(y)]$,
*(*$\mathsf{Q}$*5)* $\forall x[x \cdot 0 = 0]$,
*(*$\mathsf{Q}$*6)* $\forall x \forall y[x \cdot S(y) = x \cdot y + x]$,
*(*$\mathsf{Q}$*7)* $\forall x(x \neq 0 \to \exists y(x = S(y)))$.

So the difference between $\mathsf{PA}$ and $\mathsf{Q}$ is that the induction schema of $\mathsf{PA}$ is replaced by a single axiom stating that every nonzero element has a predecessor (which is clearly provable in $\mathsf{PA}$). This theory is pretty weak: for example, it does not prove the associativity/commutativity of the addition/multiplication. However, every model of $\mathsf{Q}$ has a standard part:

**Proposition 5.25.**
*(a) For any model* $\mathbf{M}$ *of* $\mathsf{Q}$*, there is a unique homomorphism* $f : \mathbf{N} \to \mathbf{M}$*. In fact, this* $f$ *is a* $\tau_a$*-embedding and hence we can view* $\mathbf{N}$ *as a substructure of* $\mathbf{M}$*.*
*(b) For any quantifier free formula* $\phi(\vec{x})$ *and* $\vec{a} \in \mathbb{N}^k$*,*

$$\mathbf{N} \vDash \phi(\vec{a}) \iff \mathsf{Q} \vdash \phi(\Delta(\vec{a})),$$

*where* $\Delta(\vec{a}) = (\Delta(a_1), ..., \Delta(a_k))$*.*

*Proof.* Part (b) follows from (a) since for $\mathbf{M} \vDash \mathsf{Q}$, $\mathbf{N} \subseteq \mathbf{M}$ and hence

$$\mathbf{N} \vDash \phi(\vec{a}) \iff \mathbf{M} \vDash \phi(\Delta(\vec{a})),$$

because $\phi$ is quantifier free. Because $\mathbf{M}$ was an arbitrary model of $\mathsf{Q}$, we are done by the Completeness theorem.

As for part (a), the proof is exactly the same as for models of $\mathsf{PA}$. The uniqueness is clear because we $f$ has to preserve 0 and $S$ and thus $f(\Delta(n)^{\mathbf{N}}) = \Delta(n)^{\mathbf{M}}$. This function is injective because $S^{\mathbf{M}}$ is injective and $0^{\mathbf{M}}$ is does not have a predecessor. It remains to show that $f$ preserves $+$ and $\cdot$. We show that $f(n + m) = f(n) + f(m)$ by induction on $m$, and we leave the case of $\cdot$ to the reader. For $m = 0$, this follows from axiom ($\mathsf{Q}$3). Now assume $f(n+m) = f(n)+f(m)$. Then $f(n+S(m)) = f(S(n+m)) = S(f(n+m)) = S(f(n)+f(m)) = f(n) + S(f(m)) = f(n) + f(S(m))$, where we used the facts that $f$ respects $S$ and that $\mathbf{M}$ satisfies axiom ($\mathsf{Q}$4). □

Let $x \leq y$ and $x < y$ abbreviate the formulas $\exists z(z + x = y)$ and $x \neq y \wedge \exists z(z + x = y)$, respectively. Keep in mind that $z + x$ may not be equal to $x + z$ in a model of $\mathsf{Q}$. Since the statement $x \leq y$ is not quantifier free, it does not follow from the previous lemma that a model of $\mathsf{Q}$ and $\mathbb{N}$ have to agree on the ordering of natural numbers (the standard part of $\mathbf{M}$). However, it turns out to still be true:

[3] This is due to Raphael Robinson and not Abraham or Julia Robinsons as I falsely thought.

**Lemma 5.26** (Q preserves the ordering on $\mathbb{N}$). *For all $n, m \in \mathbb{N}$,*

*(a)* $\mathsf{Q} \vdash x \leq \Delta(n) \to \bigvee_{i=0}^{n} x = \Delta(i)$*;*
*(b)* $n \leq m \iff \mathsf{Q} \vdash \Delta(n) \leq \Delta(m)$*;*
*(c)* $\neg n \leq m \iff \mathsf{Q} \vdash \neg\Delta(n) \leq \Delta(m)$*;*
*(d)* $\mathsf{Q} \vdash x \leq \Delta(n) \vee \Delta(n+1) \leq x$*;*
*(e)* $\mathsf{Q} \vdash x \leq \Delta(n) \vee \Delta(n) < x$*.*

*Proof.* For part (b), the right-to-left direction follows immediately from (a). As for the other direction, if $n \leq m$, then let $k = m - n$ and thus $\mathbb{N} \vDash \Delta(k) + \Delta(n) = \Delta(m)$. By (b) of 5.25, $\mathsf{Q} \vdash \Delta(k) + \Delta(n) = \Delta(m)$ and thus $\mathsf{Q} \vdash \Delta(n) \leq \Delta(m)$.

For (e), first consider $n = 0$. Then by (Q3), $\mathsf{Q} \vdash 0 \leq x$, so the desired statement follows from the definition of the formula $y < z$. Now let $n \neq 0$ and hence $n = m + 1$. By (d), $\mathsf{Q} \vdash x \leq \Delta(m) \vee \Delta(n) \leq x$. Thus, arguing in $\mathsf{Q}$ and using (a), either $x = \Delta(k)$ for some $k < n$, or $x = \Delta(n)$, or $\Delta(n) \geq x$. Hence, again using (a) and the definition of the formula $y < z$, we get that either $x \leq \Delta(n)$ or $\Delta(n) < x$.

We leave the proofs of (c) and (d) to the reader, and we prove (a) by induction on $n$. Let $\mathbf{M} \vDash \mathsf{Q}$. For $n = 0$, assume $a \in M$ and $\mathbf{M} \vDash a \leq 0$. Thus, there is $b \in M$ such that $\mathbf{M} \vDash b + a = 0$. Now if $a \neq 0^{\mathbf{M}}$, then $a$ has a predecessor, i.e. for some $c \in \mathbf{M}$, $\mathbf{M} \vDash a = S(c)$ and thus $\mathbf{M} \vDash b + S(c) = 0$. Arguing inside $\mathbf{M}$, $0 = b + S(c) = S(b + c)$, which contradicts the fact that 0 is not a successor. Thus $a = 0$.

Now assume the statement is true for $n$ and assume $\mathbf{M} \vDash a \leq \Delta(n+1)$. Hence there is $b \in M$ such that $b + a = \Delta(n+1)$ (arguing inside $\mathbf{M}$). Now if $a = 0$, we are done. Otherwise, it has a predecessor $c \in M$ and thus $S(b + c) = b + S(c) = \Delta(n+1)$. By injectivity of $S$, we get $b + c = \Delta(n)$ and hence $c \leq \Delta(n)$. By the induction hypothesis, $c$ is equal to one of $\Delta(i)$ for $i = 0, ..., n$ and thus $a$ is equal to one of $\Delta(j)$ for $j = 1, ..., n + 1$. □

**Proposition 5.27.** *All recursive functions and relations are representable in* $\mathsf{Q}$.

*Proof.* By Lemma 5.23, it is enough to show for functions.

It follows from (b) of 5.25 that the terms $x+y$, $x \cdot y$ represent the addition and multiplication functions. It is clear that the term $t(x_1, ..., x_n) = x_i$ represents the projection function $P_i^n(x_1, ..., x_n) = x_i$, and it follows from (c) and (d) of 5.26 that the formula $x \leq y$ represents the relation $\leq$ and hence $\chi_{\leq}$ is representable by 5.23. It remains to show that representability is closed under (R2) and (R3).

For (R2), assume that $\phi(\vec{x}, y)$ represents the function $g : \mathbb{N}^k \to \mathbb{N}$ and $\psi_i(\vec{v}, u)$ represent the functions $h_i : \mathbb{N}^n \to \mathbb{N}$, where $\vec{x}$ is an $k$-vector and $\vec{v}$ is a $n$-vector. We show that

$$\theta(\vec{v}, y) \doteq \exists \vec{x} \bigwedge_{i=1}^{k} \psi_i(\vec{v}, x_i) \wedge \phi(\vec{x}, y)$$

represents $f = g(h_1, ..., h_k)$. Fix $\vec{a} \in \mathbb{N}^n$ and let $c = f(\vec{a})$. We have to show that

$$\mathsf{Q} \vdash \theta(\Delta(\vec{a}), y) \leftrightarrow y = \Delta(c).$$

Let $b_i = h_i(\vec{a})$ and put $\vec{b} = (b_1, ..., b_k)$. Then $f(\vec{a}) = g(\vec{b}) = c$. Therefore,

$$\mathsf{Q} \vdash \phi(\vec{b}, y) \leftrightarrow y = c \text{ and } \mathsf{Q} \vdash \psi_i(\Delta(\vec{a}), z) \leftrightarrow z = \Delta(b_i), \text{ for } i = 1, ..., k.$$

Thus, arguing in models, we conclude that $\mathsf{Q} \vdash \theta(\Delta(\vec{a}), y) \leftrightarrow y = \Delta(c)$.

For (R3), let $\phi(\vec{x}, y, z)$ represent the function $g : \mathbb{N}^{n+1} \to \mathbb{N}$, where $\vec{x}$ is an $n$-vector and $g$ is such that for all $\vec{a} \in \mathbb{N}^n$ there is $b \in \mathbb{N}$ such that $g(\vec{a}, b) = 0$. We show that

$$\psi(\vec{x}, z) \doteq \phi(\vec{x}, y, 0) \wedge \forall u(u < z \to \neg\phi(\vec{x}, u, 0))$$

represents $f(\vec{a}) = \mu x(g(\vec{a}, x) = 0)$. Fix $\vec{a} \in \mathbb{N}^n$ and let $b = f(\vec{a})$. We have to show that

$$\mathsf{Q} \vdash \psi(\Delta(\vec{a}), z) \leftrightarrow z = \Delta(b).$$

By definition, $g(\vec{a}, i) = c_i \neq 0$ for all $i < b$, and $g(\vec{a}, b) = 0$. Thus

$$\mathsf{Q} \vdash \phi(\vec{a}, \Delta(b), v) \leftrightarrow v = 0 \text{ and } \mathsf{Q} \vdash \phi(\vec{a}, \Delta(i), v) \leftrightarrow v = \Delta(c_i) \text{ for all } i < b.$$

Arguing in models and using part (a) of 5.26, we conclude that

$$\mathsf{Q} \vdash \psi(\Delta(\vec{a}), z) \leftrightarrow z = \Delta(b).$$

□

The converse of this proposition is also true and we will prove it in a later subsection. Thus representability in $\mathsf{Q}$ characterizes recursive functions.

## 5.4. **Gödel coding**

Here we describe a coding of formulas and proofs, and all functions necessary to prove the fixed point lemma and the Incompleteness theorem.

For the rest of the section, let $\tau$ be a finite signature.

- We code the symbols of $\mathbb{FOL}(\tau)$ as follows: for $s \in \tau \cup \{\text{logical symbols}\} \cup \{v_0, v_1, ...\}$, assign a number $\mathsf{SN}(s)$ as follows: put $\mathsf{SN}(s) = 2i$ if $s = v_i$ and assign an odd number to each of the remaining symbols (finitely many) such that different symbols get different numbers.
- For a $\tau$-term $t$, define its Gödel code $\ulcorner t \urcorner$ as follows

$$\ulcorner t \urcorner = \begin{cases} <\mathsf{SN}(s)> & \text{if } t = s \text{ is a variable or a constant symbol} \\ <\mathsf{SN}(f), \ulcorner t_1 \urcorner, ..., \ulcorner t_n \urcorner> & \text{if } f \text{ is an } n\text{-ary function symbol and } t = f(t_1, ..., t_n) \end{cases}.$$

Note that for a variable or a constant symbol $s$, $\ulcorner s \urcorner$ may not be equal to $\mathsf{SN}(s)$.

- For a $\tau$-formula $\phi$, define its Gödel code $\ulcorner \phi \urcorner$ as follows

$$\ulcorner \phi \urcorner = \begin{cases} <\mathsf{SN}(=), \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner> & \text{if } \phi \equiv (t_1 = t_2) \\ <\mathsf{SN}(R), \ulcorner t_1 \urcorner, ..., \ulcorner t_n \urcorner> & \text{if } R \text{ is an } n\text{-ary relation symbol and } \phi \equiv R(t_1, ..., t_n) \\ <\mathsf{SN}(\neg), \ulcorner \psi \urcorner> & \text{if } \phi \equiv \neg\psi \\ <\mathsf{SN}(\wedge), \ulcorner \psi_1 \urcorner, \ulcorner \psi_2 \urcorner> & \text{if } \phi \equiv \psi_1 \wedge \psi_2 \\ <\mathsf{SN}(\vee), \ulcorner \psi_1 \urcorner, \ulcorner \psi_2 \urcorner> & \text{if } \phi \equiv \psi_1 \vee \psi_2 \\ <\mathsf{SN}(\to), \ulcorner \psi_1 \urcorner, \ulcorner \psi_2 \urcorner> & \text{if } \phi \equiv \psi_1 \to \psi_2 \\ <\mathsf{SN}(\exists), \ulcorner v \urcorner, \ulcorner \psi \urcorner> & \text{if } \phi \equiv \exists v \psi \\ <\mathsf{SN}(\forall), \ulcorner v \urcorner, \ulcorner \psi \urcorner> & \text{if } \phi \equiv \forall v \psi \end{cases}.$$

**Lemma 5.28.** *The following subsets of* $\mathbb{N}$ *are primitive recursive:*

*(i)* $\mathsf{Variable} := \{\ulcorner x \urcorner : x \textit{ is a variable}\}$
*(ii)* $\mathsf{Term} := \{\ulcorner t \urcorner : t \textit{ is a } \tau\textit{-term}\}$
*(iii)* $\mathsf{Formula} := \{\ulcorner \phi \urcorner : \phi \textit{ is a } \tau\textit{-formula}\}$

*Proof.* In all proofs we use complete primitive recursion (Lemma 5.19).
(i) $a \in \mathsf{Variable}$ if and only if $\mathsf{lh}(a) = 1$ and $(a)_0$ is even.
(ii) $\mathsf{Term}(a)$ if and only if $\mathsf{Variable}(a)$ or $a$ is a code for a constant symbol or $(a)_0$ is a code for an $n$-ary functions symbol with $n = \mathsf{lh}(a) - 1$ and $\forall i < n, \mathsf{Term}((a)_{i+1})$.
(iii) is left to the reader. It gets messy if one wants to also check our convention about quantified variables. □

**Lemma 5.29.** *There is a primitive recursive function* $\mathsf{Sub} : \mathbb{N}^3 \to \mathbb{N}$ *such that for any* $\tau$*-formula* $\phi$*, variable* $v$ *and* $\tau$*-term* $t$ *that is free for* $\phi$*,*

$$\mathsf{Sub}(\ulcorner\phi\urcorner, \mathsf{SN}(v), \ulcorner t\urcorner) = \ulcorner\phi(t/v)\urcorner.$$

*Proof.* Define $\mathsf{Sub}(a, m, k) =$

$$\begin{cases} k & \text{if } \mathsf{Variable}(a) \text{ and } (a)_0 = m \\ <(a)_0, \mathsf{Sub}((a)_1, m, k), ..., \mathsf{Sub}((a)_{\mathsf{lh}(a)-1}, m, k)> & \text{if } \mathsf{lh}(a) > 0 \text{ and } (a)_0 \neq \mathsf{SN}(\exists) \\ <(a)_0, (a)_1, \mathsf{Sub}((a)_2, m, k)> & \text{if } \mathsf{lh}(a) > 0 \text{ and } (a)_0 = \mathsf{SN}(\exists) \text{ and } (a)_1 \neq m \\ a & \text{otherwise} \end{cases}.$$

This is clearly primitive recursive (using complete recursion). □

**Lemma 5.30.** *The following relations are primitive recursive:*
*(1)* $\mathsf{FreeVar} := \{(\ulcorner\phi\urcorner, \mathsf{SN}(v)) : v$ *occurs free in* $\phi\} \subseteq \mathbb{N}^2$
*(2)* $\mathsf{FreeSub} := \{(\ulcorner\phi\urcorner, \ulcorner t\urcorner) : t$ *is free for* $\phi\} \subseteq \mathbb{N}^3$
*(3)* $\mathsf{Sentence} := \{\ulcorner\phi\urcorner : \phi$ *is a sentence*$\} \subseteq \mathbb{N}$
*(4)* $\mathsf{Axiom} := \{\ulcorner\phi\urcorner : \phi$ *is an axiom of* $\mathbb{FOL}(\tau)\} \subseteq \mathbb{N}$
*(5)* $\mathsf{MP} := \{(\ulcorner\phi\urcorner, \ulcorner\phi \to \psi\urcorner, \psi) : \phi, \psi$*are* $\tau$*-formulas*$\} \subseteq \mathbb{N}^3$
*(6)* $\mathsf{Gen} := \{(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) : \psi$ *is obtained from* $\phi$ *by Generalization*$\} \subseteq \mathbb{N}^2$
*(7)* $\mathsf{ExistsElim} := \{(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) : \psi$ *is obtained from* $\phi$ *by* $\exists$*-elimination*$\} \subseteq \mathbb{N}^2$
*where* $\phi$*,* $t$ $v$ *range over formulas, terms and variables of* $\mathbb{FOL}(\tau)$*.*

*Proof.* This is an easy but tedious programming exercise. For example: for all $a \in \mathbb{N}$,

$$\mathsf{Sentence}(a) \iff \mathsf{Formula}(a) \text{ and } \forall i_{<a} \neg \mathsf{FreeVar}(a, i).$$

The readers are invited to check the rest of the relations themselves if they feel like programming. □

**Definition 5.31.** *For a* $\tau$*-theory* $T$*, define*

$$\mathsf{Proof}_T := \{(<\ulcorner\phi_1\urcorner, ..., \ulcorner\phi_n\urcorner>, \ulcorner\phi\urcorner) : (\phi_1, ..., \phi_n) \textit{ is a proof of } \phi \textit{ from } T\} \subseteq \mathbb{N}^2,$$

*where* $\phi_i$ *and* $\phi$ *vary over* $\tau$*-formulas.*

For a $\tau$-theory $T$, put $\ulcorner T\urcorner := \{\ulcorner\phi\urcorner : \phi \in T\}$. We say that $T$ is recursive (primitive recursive, arithmetical) if such is $\ulcorner T\urcorner$.

**Lemma 5.32.** *If a* $\tau$*-theory* $T$ *is recursive (primitive recursive, arithmetical), then so is* $\mathsf{Proof}_T$*.*

*Proof.* This is because for all $a \in \mathbb{N}$, $\mathsf{Proof}_T(a, b)$ if and only if $\mathsf{lh}(a) > 0$ and $(a)_{\mathsf{lh}(a)-1} = b$ and for every $k < \mathsf{lh}(a)$ either $(a)_k \in \mathsf{Axiom}$ or $(a)_k \in \ulcorner T\urcorner$ or $\exists i_{<k} \exists j_{<k} \mathsf{MP}((a)_i, (a)_j, (a)_k)$ or $\exists j_{<k} \mathsf{Gen}((a)_j, (a)_k)$ or $\exists j_{<k} \mathsf{ExistsElim}((a)_j, (a)_k)$. □

## 5.5. **The First Incompleteness Theorem (Rosser's form)**

Define a function $\mathsf{Sub}_0 : \mathbb{N}^2 \to \mathbb{N}$ by $\mathsf{Sub}_0(a,n) = \mathsf{Sub}(a, \mathsf{SN}(v_0), \Delta(n))$. It is clear that $\mathsf{Sub}_0$ is primitive recursive since such is $\mathsf{Sub}$.

For a $\tau_{\mathrm{a}}$-formula $\theta$, put $[\theta] \coloneqq \Delta(\ulcorner\theta\urcorner)$.

**Lemma 5.33** (Fixed point for $\mathsf{Q}$). *For every $\tau_a$-formula $\phi(v)$, there is a $\tau_a$-sentence $\theta$ such that*

$$\mathsf{Q} \vdash \theta \leftrightarrow \phi([\theta]).$$

*Proof.* Let $\mathbf{Sub}_0(x,y,z)$ be a $\tau_{\mathrm{a}}$-formula representing $\mathsf{Sub}_0$ in $\mathsf{Q}$. We can assume without loss of generality that the variable $v_0$ does not appear in $\mathbf{Sub}_0$ and $\phi$. Put

$$\psi(v_0) \equiv \exists z(\mathbf{Sub}_0(v_0, v_0, z) \wedge \phi(z)),$$

and let $e = \ulcorner\psi\urcorner$. Put $\theta \equiv \psi(\Delta(e))$. Then $\mathsf{Sub}_0(e,e) = \ulcorner\psi(\Delta(e))\urcorner = \ulcorner\theta\urcorner$ and hence, by the definition of representability,

$$\mathsf{Q} \vdash \mathbf{Sub}_0(\Delta(e), \Delta(e), z) \leftrightarrow z = [\theta]. \tag{i}$$

In particular,

$$\mathsf{Q} \vdash \mathbf{Sub}_0(\Delta(e), \Delta(e), [\theta]). \tag{ii}$$

Therefore, we have

$$\begin{aligned} \mathsf{Q} \vdash \theta &\iff \mathsf{Q} \vdash \psi(\Delta(e)) \\ &\iff \mathsf{Q} \vdash \exists z(\mathbf{Sub}_0(\Delta(e), \Delta(e), z) \wedge \phi(z)) \\ &\iff \mathsf{Q} \vdash \mathbf{Sub}_0(\Delta(e), \Delta(e), [\theta]) \wedge \phi([\theta]) && (\Longrightarrow \text{ is because of (i)}) \\ &\iff \mathsf{Q} \vdash \phi([\theta]). && (\Longleftarrow \text{ is because of (ii)}) \end{aligned}$$

□

Now we are ready to prove the Incompleteness theorem for all $\tau_{\mathrm{a}}$-theories $T \supseteq \mathsf{Q}$. However, we would like to prove a slightly stronger version that applies to theories in signatures other than $\tau_{\mathrm{a}}$ that are rich enough to encode $\mathsf{Q}$ in them. We make this precise in the following

**Definition 5.34.** *Let $T_1, T_2$ be theories in finite signatures $\tau_1, \tau_2$, respectively. An interpretation of $T_1$ in $T_2$ is a map $\pi$ from the set of $\tau_1$-sentences to the set of $\tau_2$-sentences such that*

*(i)* $T_1 \vdash \theta \implies T_2 \vdash \pi(\theta)$,
*(ii)* $T_2 \vdash \pi(\neg\theta) \leftrightarrow \neg\pi(\theta)$,
*(iii)* $T_2 \vdash \pi(\phi \wedge \psi) \leftrightarrow \pi(\phi) \wedge \pi(\psi)$,
*(iv) there is a primitive recursive function* $\pi^* : \mathbb{N} \to \mathbb{N}$ *such that* $\pi^*(\ulcorner\theta\urcorner) = \ulcorner\pi(\theta)\urcorner$,

*where $\theta, \phi, \psi$ range over $\tau_1$-sentences, and in the last equality, $\ulcorner\ \urcorner$ denotes the coding function of $\mathbb{FOL}(\tau_1)$ on the left and of $\mathbb{FOL}(\tau_2)$ on the right.*

If there is an interpretation of $T_1$ in $T_2$, we say that $T_2$ interprets $T_1$. For example, $\mathsf{ZFC}$ interprets $\mathsf{Q}$. Also, if $T_1 \subseteq T_2$, then by taking the identity function as $\pi^*$, we see that $T_2$ interprets $T_1$.

Below let $\tau$ be a finite signature.

**Lemma 5.35.** *Let $T$ be a (primitive) recursive $\tau$-theory that interprets* $\mathsf{Q}$ *and let $\pi$ be an interpretation of* $\mathsf{Q}$ *in $T$. Then the following relations are (primitive) recursive:*

$$\begin{aligned} \mathsf{Proof}_{\pi,T}(a,b) &\iff \text{$b$ is an $\mathbb{FOL}(\tau_a)$-code of a $\tau_a$-sentence $\phi$ and} \\ &\qquad \text{$a$ is an $\mathbb{FOL}(\tau)$-code of a proof of $\pi(\phi)$ from $T$,} \\ \mathsf{Refute}_{\pi,T}(a,b) &\iff \text{$b$ is an $\mathbb{FOL}(\tau_a)$-code of a $\tau_a$-sentence $\phi$ and} \\ &\qquad \text{$a$ is an $\mathbb{FOL}(\tau)$-code of a proof of $\pi(\neg\phi)$ from $T$.} \end{aligned}$$

*Proof.* Observe that

$$\begin{aligned} \mathsf{Proof}_{\pi,T}(a,b) &\iff \mathsf{Sentence}_{\tau_a}(b) \text{ and } \mathsf{Proof}_T(a,\pi^*(b)), \\ \mathsf{Refute}_{\pi,T}(a,b) &\iff \mathsf{Sentence}_{\tau_a}(b) \text{ and } \mathsf{Proof}_T(a,\pi^*(\langle \mathsf{SN}(\neg), b\rangle)). \end{aligned}$$

□

**First Incompleteness Theorem 5.36** (Rosser's form)**.** *Any consistent recursive $\tau$-theory that interprets* $\mathsf{Q}$ *is incomplete.*

Let us contemplate about the proof a bit before we present it. In the proof of the Incompleteness theorem for $T \subseteq \mathsf{Th}(\mathbf{N})$, we constructed a sentence $\gamma$ that basically expressed the Liar Paradox: it said about itself that it is not provable. Let us try to use the same idea here: let $\pi$ be an interpretation of $\mathsf{Q}$ in $T$ and let $\mathbf{Proof}_{\pi,T}(x,y)$ be a $\tau_\mathrm{a}$-formula representing $\mathsf{Proof}_{\pi,T}$ in $\mathsf{Q}$. Then by the fixed point lemma for $\mathsf{Q}$, we get a $\tau_\mathrm{a}$-sentence $\gamma$ such that

$$\mathsf{Q} \vdash \gamma \leftrightarrow \forall x \neg \mathbf{Proof}_{\pi,T}(x,[\gamma]). \tag{$*$}$$

It is true that $T \nvdash \pi(\gamma)$ since otherwise there will be a code $a \in \mathbb{N}$ of a proof of $\pi(\gamma)$ from $T$ and hence $\mathsf{Q} \vdash \mathbf{Proof}_{\pi,T}(\Delta(a),[\gamma])$. But then by $(*)$, $\mathsf{Q} \vdash \neg\gamma$ and thus $T \vdash \pi(\neg\gamma)$, so $T \vdash \neg\pi(\gamma)$, contradicting the consistency of $T$.

However, we don't get any contradiction if we assume $T \vdash \neg\pi(\gamma)$. Indeed, assuming the latter, the consistency of $T$ implies that $T \nvdash \pi(\gamma)$ and hence there is no natural number that is a code of a proof of $\pi(\gamma)$ from $T$, i.e. $\neg\mathsf{Proof}_{\pi,T}(a,\ulcorner\gamma\urcorner)$, for all $a \in \mathbb{N}$. Then, for every $a \in \mathbb{N}$, $\mathsf{Q} \vdash \neg\mathbf{Proof}_{\pi,T}(\Delta(a),[\gamma])$. Unfortunately, this does NOT imply that $\mathsf{Q} \vdash \forall x \neg\mathbf{Proof}_{\pi,T}(x,[\gamma])$ because there may well be a model $\mathbf{M}$ of $\mathsf{Q}$ with a nonstandard element $w \in M \setminus \mathbb{N}$ such that $\mathbf{M} \vDash \mathbf{Proof}_{\pi,T}(w,[\gamma])$ and there is no contradiction here.

So, the Liar Paradox doesn't work here and Rosser's trick is to use an idea somewhat similar to Berry's Paradox, which is the following:

*The smallest natural number not definable in less than 100 characters.*

This is a "paradox" because we can only define finitely many different numbers using less than 100 characters (English letters and numbers) and hence there surely are numbers which cannot be defined in less than 100 characters. However, we just described the smallest of them using (I believe) 70 characters.

*Rosser's proof of the Incompleteness Theorem 5.36.* Let $\pi$ be an interpretation of $\mathsf{Q}$ in $T$, and let $\mathbf{Proof}_{\pi,T}(x,y)$ and $\mathbf{Refute}_{\pi,T}(x,y)$ be $\tau_\mathrm{a}$-formulas representing $\mathsf{Proof}_{\pi,T}$ and $\mathsf{Refute}_{\pi,T}$ in $\mathsf{Q}$. Then by the fixed point lemma for $\mathsf{Q}$, we get a $\tau_\mathrm{a}$-sentence $\rho$ such that

$$\mathsf{Q} \vdash \rho \leftrightarrow \forall x(\mathbf{Proof}_{\pi,T}(x,[\rho]) \to (\exists u < x)\mathbf{Refute}_{\pi,T}(u,x)). \tag{1}$$

The *Rosser sentence* $\rho$ expresses the unprovability of its translation in $T$ in a round-about way: it asserts

*For every proof of myself, there is a shorter proof of my negation.*

We show that neither $T \vdash \pi(\rho)$ nor $T \vdash \neg\pi(\rho)$.
**Case 1**: suppose $T \vdash \pi(\rho)$. Then there is a code $m \in \mathbb{N}$ of a proof of $\pi(\rho)$ from $T$ and hence

$$\mathsf{Q} \vdash \mathbf{Proof}_{\pi,T}(\Delta(m), [\rho]). \tag{2}$$

Because $T$ is consistent, $T \nvdash \pi(\neg\rho)$ since $\pi(\neg\rho) \equiv \neg\pi(\rho)$. Thus $\forall k \in \mathbb{N}$, $\neg\mathsf{Refute}_{\pi,T}(k, \ulcorner\rho\urcorner)$ and hence $\mathsf{Q} \vdash \neg\mathbf{Refute}_{\pi,T}(\Delta(k), [\rho])$; in particular, this is true for all $k < m$. Therefore, by (a) of Lemma 5.26,

$$\mathsf{Q} \vdash (\forall u < \Delta m)\neg\mathbf{Refute}_{\pi,T}(u, [\rho]). \tag{3}$$

From (2) and (3), we get

$$\mathsf{Q} \vdash \exists x(\mathbf{Proof}_{\pi,T}(x, [\rho]) \wedge (\forall u < x)\neg\mathbf{Refute}_{\pi,T}(u, x)),$$

which implies $\mathsf{Q} \vdash \neg\rho$ by (1). Therefore, $T \vdash \pi(\neg\rho)$ and hence $T \vdash \neg\pi\rho$, contradicting the consistency of $T$.

**Case 2**: suppose $T \vdash \neg\pi(\rho)$. Thus $T \vdash \pi(\neg\rho)$, so there is a code $k \in \mathbb{N}$ of a proof of $\pi(\neg\rho)$ from $T$. Hence $\mathsf{Refute}_{\pi,T}(k, \ulcorner\rho\urcorner)$ holds and by representability in $\mathsf{Q}$,

$$\mathsf{Q} \vdash \mathbf{Refute}_{\pi,T}(\Delta(k), [\rho]). \tag{4}$$

Also, for any $n \in \mathbb{N}$, $\neg\mathsf{Proof}_{\pi,T}(n, \ulcorner\rho\urcorner)$ holds by the consistency of $T$, and thus

$$\mathsf{Q} \vdash \neg\mathbf{Proof}_{\pi,T}(\Delta(n), [\rho]). \tag{5}$$

We argue in models, so fix $\mathbf{M} \vDash \mathsf{Q}$. By (e) of Lemma 5.26, for every $a \in M$, $a \le \Delta(k)$ or $\Delta(k) < a$. In the first case, by (a) of Lemma 5.26, we get that $a = \Delta(n)$ for some $n \le k$, and thus $\mathbf{M} \vDash \neg\mathbf{Proof}_{\pi,T}(a, [\rho])$, by (5). In the second case, i.e. if $\Delta(k) < a$,

$$\mathbf{M} \vDash (\exists u < a)\mathbf{Refute}_{\pi,T}(u, [\rho]),$$

by (4). Therefore, for all $a \in M$,

$$\mathbf{M} \vDash \mathbf{Proof}_{\pi,T}(a, [\rho]) \to (\exists u < a)\mathbf{Refute}_{\pi,T}(u, [\rho]).$$

Thus

$$\mathsf{Q} \vdash \forall x(\mathbf{Proof}_{\pi,T}(x, [\rho]) \to (\exists u < x)\mathbf{Refute}_{\pi,T}(u, x)),$$

and hence $\mathsf{Q} \vdash \rho$, by (1). But then $T \vdash \pi(\rho)$, contradicting the consistency of $T$. □

## 5.6. **The Second Incompleteness Theorem and Löb's theorem**

Let $\tau$ be a finite signature and let $T$ be a recursive $\tau$-theory. Recall that the relations

$$\begin{array}{rcl} \mathsf{Proof}_T(a, b) & \iff & b \text{ is a code of a sentence } a \text{ is a code of a proof of it from } T, \\ \mathsf{Refute}_T(a, b) & \iff & b \text{ is a code of a sentence } a \text{ is a code of a proof of the negation of it from } T, \end{array}$$

are recursive. Let $\mathbf{Proof}_T(x, y)$ and $\mathbf{Refute}_T(x, y)$ be $\tau_{\mathrm{a}}$-formulas representing them in $\mathsf{Q}$.

**Definition 5.37.** *For $T$ as above, we define a $\tau_a$-sentence that expresses the consistency of $T$ as follows:*

$$\mathbf{Con}_T \equiv \neg\exists x\exists y\exists z\mathbf{Proof}_T(x, z) \wedge \mathbf{Refute}_T(y, z).$$

**Lemma 5.38.** *Let $T$ be a recursive $\tau$-theory interpreting* $\mathsf{PA}$ *and let $\pi$ be an interpretation. Also, let $\rho_T$ be the Rosser sentence for $T$ as in the proof of 5.36 above. Then* $\mathsf{PA} \vdash \mathbf{Con}_T \to \rho_T$.

*Proof.* We claim that Rosser's proof of the First Incompleteness theorem can be carried out in $\mathsf{PA}$. It would take too long to actually prove this, but the main point is the following: Rosser's proof is completely syntactic, i.e. playing with formal proofs (we only used models and the Completeness theorem because we were too lazy to do formal proofs, but in principle we could have constructed all necessary formal proofs). Syntactic arguments such as the proofs of the fixed point lemma or Deduction theorem can be expressed and carried through $\mathsf{PA}$ because all they use is induction, which $\mathsf{PA}$ has.

Thus, in particular $\mathsf{PA}$ proves that if $T$ is consistent then $T \nvdash \pi(\rho_T)$:

$$\mathsf{PA} \vdash \mathbf{Con}_T \to \forall x \neg \mathbf{Proof}_{\pi,T}(x, [\rho_T]).$$

On the other hand, it follows from the definition of $\rho_T$ that

$$\mathsf{PA} \vdash \forall x \neg \mathbf{Proof}_{\pi,T}(x, [\rho_T]) \to \rho_T.$$

Therefore, $\mathsf{PA} \vdash \mathbf{Con}_T \to \rho_T$. □

From this we immediately get yet another foundational theorem by Gödel:

**Second Incompleteness Theorem 5.39.** *Let $T$ be a recursive $\tau$-theory interpreting* $\mathsf{PA}$ *and let $\pi$ be an interpretation. Then $T \nvdash \pi(\mathbf{Con}_T)$, i.e. $T$ cannot prove its own consistency.*

*Proof.* By the previous lemma and the fact that $\pi$ is an interpretation of $\mathsf{PA}$ in $T$, we get

$$T \vdash \pi(\mathbf{Con}_T) \to \pi(\rho_T).$$

Thus, if $T \vdash \pi(\mathbf{Con}_T)$ then $T \vdash \pi(\rho_T)$, which is a contradiction. □

For a recursive $\tau$-theory $T$, let $\mathbf{Provable}_T(y) \equiv \exists x \mathbf{Proof}_T(x, y)$.

**Lemma 5.40.** *Let $\phi, \theta$ be $\tau_a$-sentences. The following statements are provable in* $\mathsf{PA}$*:*
*(a) The Deduction theorem:* $\mathbf{Provable}_{\mathsf{PA} \cup \{\theta\}}([\phi]) \leftrightarrow \mathbf{Provable}_{\mathsf{PA}}([\theta \to \phi])$.
*(b) Proof by contradiction:* $\mathbf{Provable}_{\mathsf{PA}}([\neg\theta \to (0 = 1)]) \leftrightarrow \mathbf{Provable}_{\mathsf{PA}}([\theta])$.

*Proof.* To prove this one has to note that the proofs of the corresponding theorems can be formalized in $\mathsf{PA}$ since all they use is syntactic arguments and induction. □

Because $\mathbf{N}$ is a model of $\mathsf{PA}$, we know that whatever $\mathsf{PA}$ proves is true about the natural numbers, in other words, for every $\tau_\mathrm{a}$-sentence $\theta$,

$$\mathbf{N} \vDash \mathbf{Provable}_{\mathsf{PA}}([\theta]) \to \theta.$$

Does $\mathsf{PA}$ know this? That is: does it prove $\mathbf{Provable}_{\mathsf{PA}}([\theta]) \to \theta$ for all $\theta$? Here is the answer:

**Theorem 5.41** (Löb, 1955)**.** *For every $\tau_a$-sentence $\theta$,* $\mathsf{PA}$ *does not prove* $\mathbf{Provable}_{\mathsf{PA}}([\theta]) \to$ *$\theta$ unless it proves $\theta$ itself, i.e.*

$$\mathsf{PA} \vdash \mathbf{Provable}_{\mathsf{PA}}([\theta]) \to \theta \iff \mathsf{PA} \vdash \theta.$$

*Proof.* We prove the left-to-right direction since the other one is trivial. Assume for contradiction that $\mathsf{PA} \vdash \mathbf{Provable}_{\mathsf{PA}}([\theta]) \to \theta$ yet $\mathsf{PA} \nvdash \theta$. Thus the theory $T := \mathsf{PA} \cup \{\neg\theta\}$ is consistent. By contrapositive, $\mathsf{PA} \vdash \neg\theta \to \neg\mathbf{Provable}_{\mathsf{PA}}([\theta])$ and hence,

$$T \vdash \neg\mathbf{Provable}_{\mathsf{PA}}([\theta]). \tag{†}$$

By (a) and (b) of Lemma 5.40, we have

$$\mathsf{PA} \vdash \mathbf{Provable}_T([0 = 1]) \leftrightarrow \mathbf{Provable}_{\mathsf{PA}}([\neg\theta \to (0 = 1)])$$

and

$$\mathsf{PA} \vdash \mathbf{Provable}_{\mathsf{PA}}([\neg\theta \to (0 = 1)]) \leftrightarrow \mathbf{Provable}_{\mathsf{PA}}([\theta]),$$

which yield

$$\mathsf{PA} \vdash \neg\mathbf{Provable}_{\mathsf{PA}}([\theta]) \leftrightarrow \neg\mathbf{Provable}_{T}([0 = 1]).$$

Now by (†), we get

$$T \vdash \neg\mathbf{Provable}_{T}([0 = 1]).$$

But $\mathsf{PA} \vdash \neg\mathbf{Provable}_{T}([0 = 1]) \leftrightarrow \mathbf{Con}_T$ (again by a metamathematical argument that we can formalize this in $\mathsf{PA}$), and hence $T \vdash \mathbf{Con}_T$, contradicting the Second Incompleteness theorem. □

## 6. Undecidable theories

Fix a finite signature $\tau$.

**Definition 6.1.** *For a $\tau$-theory $T$, let $\mathsf{Thm}(T)$ denote the set of its theorems, i.e. $\mathsf{Thm}(T) \coloneqq \{\phi : T \vdash \phi\} \subseteq \mathbb{N}$, where $\phi$ ranges over all $\tau$-sentences. If $\ulcorner\mathsf{Thm}(T)\urcorner$ is recursive, $T$ is called decidable.*

After various incompleteness results, we are now convinced that sufficiently rich recursive theories $T$ such as $\mathsf{PA}$ or $\mathsf{ZFC}$ are incomplete. But maybe we can still write a program that for a given sentence $\phi$ decides whether it is a theorem of $T$ or not? More precisely, is $T$ decidable? (If the answer was yes for example for $\mathsf{ZFC}$, mathematicians would be unemployed and the world would be an uninteresting place to live in.) This section is devoted to answering this question.

### 6.1. $\Sigma^0_1$ sets and Kleene's theorem

**Definition 6.2** ($\Sigma^0_1$ relations)**.** *A relation (set) $Q \subseteq \mathbb{N}^k$ is called $\Sigma^0_1$ if for some recursive relation $R \subseteq \mathbb{N}^{k+1}$,*

$$\vec{a} \in Q \iff \exists x R(\vec{a}, x).$$

*We also denote by $\Sigma^0_1$ the set of all $\Sigma^0_1$ relations.*

Here are some closure properties of $\Sigma^0_1$:

**Lemma 6.3.**

*(1) $\Sigma^0_1$ is closed under finite unions/intersections and taking projections, i.e. if $P, Q \subseteq \mathbb{N}^k$, $R \subseteq \mathbb{N}^{k+1}$ are $\Sigma^0_1$, then so are*

$$P \vee Q,\ P \wedge Q,\ \exists x R(\cdot, x).$$

*(2) $\Sigma^0_1$ is closed under recursive preimages, i.e. if $f : \mathbb{N}^k \to \mathbb{N}$ is recursive and $A \subseteq \mathbb{N}$ is $\Sigma^0_1$, then the relation $B = f^{-1}(A)$ is $\Sigma^0_1$.*

*Proof.* We leave (a) as a homework exercise, and we prove (b). Let $R \subseteq \mathbb{N}^2$ be a recursive relation such that for all $n \in \mathbb{N}$, $n \in A \iff \exists m R(n, m)$. But then the relation $Q \subseteq \mathbb{N}^{k+1}$ defined by

$$(\vec{a}, m) \in Q \iff R(f(\vec{a}), m)$$

is recursive and hence the relation

$$\vec{a} \in B \iff \exists m Q(\vec{a}, m)$$

is $\Sigma^0_1$. □

**Lemma 6.4.** *For a $\tau$-theory $T$, if $T$ is recursive, then $\ulcorner \mathsf{Thm}(T) \urcorner$ is $\Sigma^0_1$.*

*Proof.* If $T$ is recursive, then so is the relation $\mathsf{Proof}_T \subseteq \mathbb{N}^2$ defined in the previous subsection. But then for all $a \in \mathbb{N}$

$$a \in \ulcorner \mathsf{Thm}(T) \urcorner \iff \exists x \mathsf{Proof}_T(x,a).$$

□

Let $\Pi^0_1$ denote the set of complements of $\Sigma^0_1$ relations, i.e. $\Pi^0_1 = \{\neg R : R \in \Sigma^0_1\}$, and let $\Delta^0_1 := \Sigma^0_1 \cap \Pi^0_1$. Also, let $\mathsf{Recursive}$ denote the set of recursive relations.

**Lemma 6.5** (Kleene's theorem)**.** $\Delta^0_1 = \mathsf{Recursive}$.

*Proof.* $\supseteq$: It is clear that $\mathsf{Recursive} \subseteq \Sigma^0_1$ (why?) and since $\mathsf{Recursive}$ is closed under complements, $\mathsf{Recursive} \subseteq \Delta^0_1$.
$\subseteq$: Let $R \subseteq \mathbb{N}^k$ be a $\Delta^0_1$ relation. Hence, there are recursive relations $P, Q \subseteq \mathbb{N}^{k+1}$ such that $\forall \vec{a} \in \mathbb{N}^k$

$$\vec{a} \in R \iff \exists x P(\vec{a}, x), \qquad \vec{a} \in \neg R \iff \exists x Q(\vec{a}, x).$$

But then the function $f : \mathbb{N}^k \to \mathbb{N}$ defined by $f(\vec{a}) = \mu x(P \vee Q(\vec{a}, x))$ is recursive and hence so is $R$ since $\vec{a} \in R \iff f(\vec{a}) \in P$. □

From this we immediately get the following decidability result:

**Proposition 6.6.** *Every complete recursive $\tau$-theory $T$ is decidable.*

*Proof.* Using the fact that for every $\tau$-sentence $\phi$, $\phi \notin \mathsf{Thm}(T) \iff \neg\phi \in \mathsf{Thm}(T)$, we get that for every $a \in \mathbb{N}$,

$$a \notin \ulcorner \mathsf{Thm}(T) \urcorner \iff a \notin \mathsf{Sentence}_\tau \text{ or } <\mathsf{SN}(\neg), a> \in \ulcorner \mathsf{Thm}(T) \urcorner.$$

By Lemma 6.4, $\ulcorner \mathsf{Thm}(T) \urcorner$ is $\Sigma^0_1$. Because $\neg\mathsf{Sentence}_\tau$ is recursive (hence $\Sigma^0_1$) and $\Sigma^0_1$ is closed under recursive preimages and finite unions (6.3), the right hand side is $\Sigma^0_1$ and thus so is $\neg\ulcorner \mathsf{Thm}(T) \urcorner$. Therefore, $\ulcorner \mathsf{Thm}(T) \urcorner$ is $\Delta^0_1$ and hence is recursive (by Kleene's theorem). □

As a corollary, we get that $\mathsf{ACF}_p$, $p = 0$ or prime, and the theory of vector spaces over a countable field[4] are decidable.

## 6.2. Universal $\Sigma^0_1$ relation and Church's theorem

For any sets $A, B$, any relation $R \subseteq A \times B$, and $a \in A$, put $R(a) := \{b \in B : (a,b) \in R\}$. In this subsection we construct a $\Sigma^0_1$ relation $R \subseteq \mathbb{N}^2$ that is universal for recursive relations, i.e. any recursive relation $P \subseteq \mathbb{N}$ is of the form $P = R(a)$, for some $a \in \mathbb{N}$. Using this we prove that any consistent theory interpreting $\mathsf{Q}$ is undecidable. We start with proving the converse of 5.27.

**Proposition 6.7.** *Let $T$ be a recursive consistent $\tau_a$-theory. Then any relation $R \subseteq \mathbb{N}^k$ representable in $T$ is recursive. In particular, any function $f : \mathbb{N}^k \to \mathbb{N}$ representable in $T$ is recursive.*

[4] As it is written, 6.6 applies only to finite signatures and if a countable field $F$ is not finite, the signature $\tau_F$ of the theory of vector spaces over $F$ is infinite. However, we can still assign codes to symbols in $\tau_F$ so that we can decode all the information about the symbol from its code in a primitive recursive way. Thus everything proven above applies to $\tau_F$ as well.

*Proof.* The statement about functions follows from that about relations since if $f$ is representable, then so is its graph (why?), and hence by the first statement the graph is recursive and hence so is $f$, by the graph property (5.10).

Let $R \subseteq \mathbb{N}^k$ be representable in $T$ by a formula $\phi(\vec{x})$. By the definition of representability and because $T$ is consistent, for all $\vec{a} \in \mathbb{N}^k$, we have

$$\vec{a} \in R \iff T \vdash \phi(\Delta(\vec{a})) \iff \ulcorner \phi(\Delta(\vec{a})) \urcorner \in \ulcorner \mathsf{Thm}(T) \urcorner.$$

The function $s : \mathbb{N}^k \to \mathbb{N}$ defined by $\vec{a} \to \ulcorner \phi(\Delta(\vec{a})) \urcorner$ is clearly primitive recursive (just apply the $\mathsf{Sub}$ function for each free variable of $\phi$). By 6.4, $\mathsf{Thm}(T)$ is $\Sigma^0_1$ and hence the right hand side is $\Sigma^0_1$ by (b) of 6.3.

Because the definition of representability is symmetric for $R$ and $\neg R$, we have that $\neg R$ is also representable (by $\neg\phi$) and hence, by what we have already proven, $\neg R$ is $\Sigma^0_1$. Hence, by Kleene's theorem, $R$ is recursive. □

From this and 5.27, we get

**Corollary 6.8.** *Let $f : \mathbb{N}^k \to \mathbb{N}$. $f$ is recursive if and only if it is representable in* $\mathsf{Q}$.

This allows us to construct a relation that enumerates all recursive subsets of $\mathbb{N}$ as follows:

**Definition 6.9.** *Recall the primitive recursive function $\mathsf{Sub}_0(a,n)$ that has the property that for every $\tau_a$-formula $\phi$,*

$$\mathsf{Sub}_0(\ulcorner \phi \urcorner, n) = \ulcorner \phi(\Delta(n)/v_0) \urcorner.$$

*For a $\tau$-theory $T$ that interprets $\mathsf{Q}$ by $\pi$, define a relation $U_T \subseteq \mathbb{N}^2$ by*

$$U_{\pi,T}(a,n) \iff \pi^*(\mathsf{Sub}_0(a,n)) \in \ulcorner \mathsf{Thm}(T) \urcorner.$$

**Proposition 6.10.** *Let $T$ be a consistent $\tau$-theory interpreting $\mathsf{Q}$ by $\pi$. Then for each recursive relation $R \subseteq \mathbb{N}$, there is $e \in \mathbb{N}$ such that $R = U_{\pi,T}(e)$. Furthermore, if $T$ is recursive, then $U_{\pi,T}$ is $\Sigma^0_1$.*

*Proof.* The second statement follows from the definition of $U_{\pi,T}$ and 6.4. For the first statement, let $\phi(v_0)$ be a formula representing $R$ in $\mathsf{Q}$ (there is always one with the free variable being $v_0$), and thus for all $n \in \mathbb{N}$,

$$\begin{array}{ccccc} n \in R & \implies & \mathsf{Q} \vdash \phi(\Delta(n)) & \implies & T \vdash \pi(\phi(\Delta(n))) \\ n \notin R & \implies & \mathsf{Q} \vdash \neg\phi(\Delta(n)) & \implies & T \vdash \neg\pi(\phi(\Delta(n))). \end{array}$$

Since $T$ is consistent, we get

$$n \in R \iff T \vdash \pi(\phi(\Delta(n))),$$

and therefore, letting $e = \ulcorner \phi(v_0) \urcorner$, we have

$$n \in R \iff U_{\pi,T}(e,n).$$

□

If we take $T = \mathsf{Q}$ and $\pi = \mathsf{id}$ in the above proposition, then, denoting $U_{\mathsf{id},\mathsf{Q}}$ by $U_{\mathsf{Q}}$, we get an even stronger result:

**Proposition 6.11.** *The relation $U_{\mathsf{Q}}$ is $\Sigma^0_1$, and for every $\Sigma^0_1$ relation $P \subseteq \mathbb{N}$, there is $e \in \mathbb{N}$ with $P = U_{\mathsf{Q}}(e)$. Thus $U_{\mathsf{Q}}$ is a universal $\Sigma^0_1$ relation.*

*Proof.* This is left as a homework problem. □

If $T$ is recursive, we know that $U_{\pi,T}$ is $\Sigma^0_1$, but is it recursive? The answer is NO, and we show it by the diagonalization method.

**Lemma 6.12** (Cantor). *For a set $A$ and a relation $R \subseteq A^2$, let $P \subseteq A$ be denote its antidiagonal, i.e. $P := \{a : \neg R(a,a)\}$. Then $P$ is not equal to $R(a)$ for any $a \in A$.*

*Proof.* Assume for contradiction that $P = R(a)$, for some $a \in A$. Then we get a contradiction because

$$\neg R(a,a) \iff P(a) \iff R(a,a).$$

□

**Corollary 6.13.** *For every consistent $\tau$-theory $T$ interpreting $\mathsf{Q}$ by $\pi$, the relation $U_{\pi,T}$ is not recursive.*

*Proof.* If $U_{\pi,T}$ were recursive, so would be its antidiagonal $P$ and thus, by 6.10, there is $a \in \mathbb{N}$ such that $P = U_{\pi,T}(a)$, contradicting 6.12. □

As a corollary, we get the following important result:

**Theorem 6.14** (Church, 1936). *Any consistent $\tau$-theory $T$ interpreting $\mathsf{Q}$ is undecidable.*

*Proof.* Let $\pi$ be an interpretation of $\mathsf{Q}$ in $T$. If $T$ were decidable, i.e. $\ulcorner \mathsf{Thm}(T) \urcorner$ were recursive, then $U_{\pi,T}$ would be recursive as well, contradicting 6.13. □

In particular, $\mathsf{Q}$ and $\mathsf{PA}$ are undecidable. Also, $\mathsf{ZFC}$ is undecidable unless it is inconsistent. Church's theorem also has the following rather surprising consequence based on the fact that $\mathsf{Q}$ is finite:

**Corollary 6.15.** *The empty $\tau_a$-theory is undecidable, i.e. $\mathsf{Thm}_{\tau_a}(\varnothing)$ is not recursive.*

*Proof.* Let $\phi_{\mathsf{Q}}$ be the conjunction of the axioms of $\mathsf{Q}$ (here is where we use that $\mathsf{Q}$ is finite!). Then, by the Deduction theorem, for any $\tau_{\mathrm{a}}$-sentence $\theta$,

$$\mathsf{Q} \vdash \theta \iff \varnothing \vdash \phi_{\mathsf{Q}} \to \theta.$$

Thus, letting $e = \ulcorner \phi_{\mathsf{Q}} \urcorner$, we get that for all $a \in \mathbb{N}$,

$$a \in \ulcorner \mathsf{Thm}(\mathsf{Q}) \urcorner \iff \mathsf{<SN(\to)}, e, a\mathsf{>} \in \ulcorner \mathsf{Thm}_{\tau_{\mathrm{a}}}(\varnothing) \urcorner.$$

Hence, $\ulcorner \mathsf{Thm}_{\tau_{\mathrm{a}}}(\varnothing) \urcorner$ cannot be recursive since otherwise $\ulcorner \mathsf{Thm}(\mathsf{Q}) \urcorner$ would also be recursive, contradicting 6.14. □

## 7. Decidable theories and quantifier elimination

Fix a signature $\tau$.

**Definition 7.1** (Quantifier elimination). *We say that a $\tau$-theory $T$ admits quantifier elimination (q.e.), if for every formula $\phi(\vec{x})$, there is a quantifier-free (q.f.) formula $\psi(\vec{x})$ such that*

$$T \vdash \forall \vec{x} (\phi(\vec{x}) \leftrightarrow \psi(x)). \tag{$*$}$$

*Assuming that $\tau$ is finite, we say that $T$ admits effective quantifier elimination if there is recursive function $h : \mathbb{N} \to \mathbb{N}$ such that for every formula $\phi(\vec{x})$, $h(\ulcorner \phi(\vec{x}) \urcorner)$ is a code of a q.f. formula $\psi(\vec{x})$ such that $(*)$ holds.*

We say that a $\tau$-structure $\mathbf{A}$ admits q.e. if so does $\mathsf{Th}(\mathbf{A})$.

Note that for a $\tau$-theory $T$ to even have a chance to admit q.e. $\tau$ must contain at least one constant symbol because there would have to exist a quantifier-free sentence.

There is a deep connection between q.e. and decidability. To see this, consider the set $\mathsf{QFThm}(T) := \{\psi : \psi \text{ is a q.f. sentence and } T \vdash \psi\}$. In most interesting cases, this set (i.e. the set of the codes) is recursive. For example, for $T = \mathsf{Th}(\mathbb{R}, 0, 1, +, -, \cdot, <)$ or $T = \mathsf{ACF}$, a q.f. sentence is just a Boolean combination of (in)equalities about terms made out of $0, 1$ using $+, -, \cdot$, and hence it is (at least intuitively) clear that $\mathsf{QFThm}(T)$ is recursive (in fact primitive recursive).

**Proposition 7.2.** *Let $\tau$ be a finite signature and $T$ a $\tau$-theory such that $\mathsf{QFThm}(T)$ is recursive. If $T$ admits effective q.e. then it is decidable.*

*Proof.* Let $h : \mathbb{N} \to \mathbb{N}$ be a recursive function as in Definition 7.1, then for every $n \in \mathbb{N}$,

$$n \in \ulcorner \mathsf{Thm}(T) \urcorner \iff h(n) \in \mathsf{QFThm}(T).$$

Thus $\ulcorner \mathsf{Thm}(T) \urcorner$ is recursive since so is the right hand side. □

Here are some famous q.e. results.

**Theorem 7.3** (Tarski)**.** *The structure $(\mathbb{R}, 0, 1, +, -, \cdot, <)$ admits effective quantifier elimination and hence its theory is decidable.*

The above result is also known as *the decidability of Euclidean geometry.*

For $p$ prime or 0, because $\mathsf{ACF}_p$ is complete, we know that it is decidable. But here is a stronger result:

**Theorem 7.4** (Robinson, Tarski, possibly others)**.** $\mathsf{ACF}$ *admits effective quantifier elimination.*

To appreciate this theorem, let $A = (a_{ij})_{i,j=1}^2$ and let $\phi(A)$ be a formula in $\tau_{\text{ring}}$ expressing that $A$ has an inverse. Clearly $\phi(A)$ is an existential formula. A q.f. formula equivalent to it is the one expressing that the determinant of $A$ is not 0. But their equivalence is a somewhat nontrivial fact!

Recall the following reduct of $\mathbf{N}$: $\mathbf{N}_+ := (\mathbb{N}, 0, S, +)$. In one of the previous sections, we defined a (primitive recursive) axiomatization $T_+$ for $\mathsf{Th}(\mathbf{N}_+)$ is stated that it is complete (and hence decidable). The completeness of $T_+$ is a consequence of the following

**Theorem 7.5** (Presburger)**.** *$T_+$ admits quantifier elimination.*

To conclude the completeness of $T_+$ from this note that any model $\mathbf{M}$ of $T_+$ has a standard part, i.e. $\mathbf{N} \subseteq \mathbf{M}$. Hence $\mathbf{M}$ and $\mathbf{N}$ believe the same q.f. sentences. But every sentence is equivalent to a q.f. sentence (in $T_+$), and thus $\mathbf{N} \equiv \mathbf{M}$.

For the rest of the section, we will develop a model-theoretic criterion for q.e. using which we will show that $\mathsf{ACF}$ admits q.e. As an application, we will prove Hilbert's Nullstellensatz.

## 7.1. A criterion for quantifier elimination

Let $\tau$ be a signature and $\mathbf{A}$ be a $\tau$-structure. For $B \subseteq A$, put $\tau(B) := \tau \cup B$, where elements of $B$ are treated as new constant symbols. We define the natural expansion of $\mathbf{A}$ to a $\tau(B)$-structure $\mathbf{A}(B)$ by interpreting symbols in $B$ by themselves, i.e. $\forall b \in B$, $b^{\mathbf{A}(B)} = b$.

**Definition 7.6** (Diagram)**.** *For a $\tau$-structure $\mathbf{A}$ and $B \subseteq A$, define* $\mathsf{Diag}(\mathbf{A}, B)$ *as the set of all quantifier free $\tau(B)$-sentences that are true in $\mathbf{A}(B)$, i.e.*

$$\mathsf{Diag}(\mathbf{A}, B) := \{\psi : \psi \textit{ is a q.f. } \tau(B)\textit{-sentence and } \mathbf{A}(B) \vDash \psi\}.$$

*When $B = A$, we simply write* $\mathsf{Diag}(\mathbf{A})$ *instead of* $\mathsf{Diag}(\mathbf{A}, A)$*.*

The following definition gives an equivalent (semantic) condition to quantifier elimination.

**Definition 7.7.** *A $\tau$-theory $T$ is called diagram-complete if for any model $\mathbf{A}$ of $T$ and any $\vec{a} \in A^n$ (for any $n$), the $\tau(\vec{a})$-theory $T \cup \mathsf{Diag}(\mathbf{A}, \vec{a})$ is complete.*

The term was chosen by me since I couldn't find an already existing name (although the notion is equivalent to substructure-completeness).

**Proposition 7.8.** *Suppose $\tau$ has at least one constant symbol $c$. Then a $\tau$-theory $T$ admits q.e. if and only if it is diagram-complete.*

*Proof.* $\Rightarrow$: Put $S := T \cup \mathsf{Diag}(\mathbf{A}, \vec{a})$ and let $\phi(\vec{x})$ be a $\tau$-formula with $\vec{x} = (x_1, ..., x_n)$. We need to show that $S$ proves either $\phi(\vec{a})$ or $\neg\phi(\vec{a})$. By q.e. there is a q.f. formula $\psi(\vec{x})$ such that $T \vdash \phi(\vec{x}) \leftrightarrow \psi(x)$. By definition, $\psi(\vec{a}) \in \mathsf{Diag}(\mathbf{A}, \vec{a})$ or $\neg\psi(\vec{a}) \in \mathsf{Diag}(\mathbf{A}, \vec{a})$, and hence $S \vdash \phi(\vec{a})$ or $S \vdash \neg\phi(\vec{a})$.
$\Leftarrow$: Assume the right hand side and let $\phi(\vec{x})$ be a $\tau$-formula with $\vec{x} = (x_1, ..., x_n)$. Put

$$\Gamma(\vec{x}) := \{\psi(\vec{x}) : \psi \text{ is a q.f. } \tau\text{-formula and } T \vdash \phi(\vec{x}) \to \psi(\vec{x})\}.$$

Take new constant symbols $\vec{d} = (d_1, ..., d_n)$ and consider the $\tau(\vec{d})$-theory $T' = T \cup \Gamma(\vec{d})$. We have three cases:
**Case 1**: $T' \vdash \phi(\vec{d})$. Since proofs are finite, there are $\psi_1(\vec{d}), ..., \psi_k(\vec{d}) \in \Gamma(\vec{d})$ such that $T, \psi_1(\vec{d}), ..., \psi_k(\vec{d}) \vdash \phi(\vec{d})$. By the Deduction theorem, letting $\psi(\vec{x}) \equiv \psi(\vec{x}) \wedge ... \wedge \psi_k(\vec{x})$, we get $T \vdash \psi(\vec{d}) \to \phi(\vec{d})$. By the Constant substitution lemma (2.36), $T \vdash \psi(\vec{d}) \to \phi(\vec{d})$. On the other hand, by the definition of $\Gamma(\vec{x})$, $\psi(\vec{x}) \in \Gamma(\vec{x})$ and hence $T \vdash \phi(\vec{x}) \to \psi(\vec{x})$. Thus $T \vdash \psi(\vec{d}) \leftrightarrow \phi(\vec{d})$, and we are done.
**Case 2**: $T' \vdash \neg\phi(\vec{d})$. By the same argument as above, there is $\psi(\vec{x}) \in \Gamma(\vec{x})$ such that $T \vdash \psi(\vec{x}) \to \neg\phi(\vec{x})$. But by the definition of $\Gamma(\vec{x})$, $T \vdash \phi(\vec{x}) \to \psi(\vec{x})$ and thus $T \vdash \phi(\vec{x}) \to \neg\phi(\vec{x})$, so $T \vdash \neg\phi(\vec{x})$. Therefore, $T \vdash \phi(\vec{x}) \leftrightarrow (c \neq c)$.
**Case 3**: $T' \nvdash \phi(\vec{d})$ and $T' \nvdash \neg\phi(\vec{d})$. Then $T' \cup \{\neg\phi(\vec{d})\}$ is consistent and by the Completeness theorem, has a model $\mathbf{A}(\vec{d})$, where are $\mathbf{A}$ is its reduct to a $\tau$-structure. Since $\mathbf{A} \vDash T$ and $T$ is diagram-complete, $S = T \cup \mathsf{Diag}(\mathbf{A}, \vec{d})$ is a complete $\tau(\vec{d})$-theory and hence proves either $\phi(\vec{d})$ or $\neg\phi(\vec{d})$. But $S$ cannot prove $\phi(\vec{d})$ since $\mathbf{A}(\vec{d}) \vDash \neg\phi(\vec{d})$, so $S \vdash \neg\phi(\vec{d})$. Because proofs are finite and $T \nvdash \neg\phi(\vec{d})$, there is $\psi(\vec{d}) \in \mathsf{Diag}(\mathbf{A}, \vec{d})$ such that $T \vdash \psi(\vec{d}) \to \neg\phi(\vec{d})$. Taking the contrapositive and using the Constant substitution lemma (2.36), $T \vdash \phi(\vec{x}) \to \neg\psi(\vec{x})$ and so $\neg\psi(\vec{x}) \in \Gamma(\vec{x})$. Therefore, $\neg\psi(\vec{d}) \in T'$ and hence $\neg\psi(\vec{d}) \in \mathsf{Diag}(\mathbf{A}, \vec{d})$, contradicting the consistency of $\mathsf{Diag}(\mathbf{A}, \vec{d})$ since $\psi(\vec{d}) \in \mathsf{Diag}(\mathbf{A}, \vec{d})$. □

Note that in the definition of diagram-completeness, the model $\mathbf{A}$ is somewhat irrelevant, it is only there to make sure that $\mathsf{Diag}(\mathbf{A}, \vec{a})$ is consistent and contains $\psi(\vec{a})$ or $\neg\psi(\vec{a})$ for every q.f. formula $\psi(\vec{x})$. We make this precise in the lemma below.

**Definition 7.9.** *Let $\vec{d}$ be a vector of distinct constant symbols that do not occur in $\tau$. A set $\Gamma(\vec{d})$ of quantifier free $\tau(\vec{d})$-sentences is called a $T$-diagram if $T \cup \Gamma(\vec{d})$ is consistent and for every q.f. $\tau(\vec{d})$-sentence $\psi$, $\psi \in \Gamma(\vec{d})$ or $\neg\psi \in \Gamma(\vec{d})$.*

**Lemma 7.10.** *A $\tau$-theory $T$ is diagram-complete if and only if for any $\vec{d}$ (of any length) and any $T$-diagram $\Gamma(\vec{d})$, $T \cup \Gamma(\vec{d})$ is a complete $\tau(\vec{d})$-theory.*

*Proof.* $\Leftarrow$ follows from the Soundness of $\mathbb{FOL}$ and $\Rightarrow$ follows from the Completeness of $\mathbb{FOL}$. □

### 7.2. **Quantifier elimination for** ACF

In this subsection we prove that ACF is diagram-complete. The only method for showing completeness that we have learnt so far is the Łoś-Vaught test, and that is what we will use.

The proof of the following proposition is almost the same as of 4.4.

**Proposition 7.11.** *For every* ACF*-diagram $\Gamma(\vec{d})$,* $\mathsf{ACF} \cup \Gamma(\vec{d})$ *is a $\kappa$-categorical $\tau_{ring}(\vec{d})$-theory, for every uncountable cardinal $\kappa$.*

*Proof.* Let $\mathbf{K}_1, \mathbf{K}_2 \models \mathsf{ACF} \cup \Gamma(\vec{d})$ with $|K_1| = |K_2| = \kappa$. Note that $\mathbf{K}_1, \mathbf{K}_2$ have the same characteristic since it is expressible by a q.f. $\tau_{\text{ring}}$-sentence which must be contained in $\Gamma(\vec{d})$. Let $p$ be the characteristic ($p = 0$ or $p$ is prime).

For $i = 1, 2$, let $F_i$ be the base field of $\mathbf{K}_i$, i.e. the substructures of $\mathbf{K}_i$ generated by $\varnothing$. (If $p = 0$, then $F_i$ is a copy of $\mathbb{Q}$; otherwise it is a copy of $\mathbb{Z}/p\mathbb{Z}$.) Since $F_1$ and $F_2$ are clearly isomorphic (as rings), we can assume without loss of generality that $F_1 = F_2 =: F$. Let $\vec{a} = \vec{d}^{\mathbf{K}_1}$, $\vec{b} = \vec{d}^{\mathbf{K}_2}$, and denote by $F(\vec{a})$, $F(\vec{b})$ the fields inside $K_1, K_2$, generated by $\vec{a}$,$\vec{b}$ over $F$, respectively.

**Claim.** *$F(\vec{a})$ and $F(\vec{b})$ are isomorphic.*

*Proof of Claim.* Elements of $F(\vec{a})$ are of the form $\frac{p(\vec{a})}{q(\vec{a})}$, where $p, q$ are polynomials over $F$ and $q(\vec{a}) \neq 0$. Define $h : F(\vec{a}) \to F(\vec{b})$ by $\frac{p(\vec{a})}{q(\vec{a})} \mapsto \frac{p(\vec{b})}{q(\vec{b})}$. This is well-defined because if $q(\vec{a}) \neq 0$, then $q(\vec{b}) \neq 0$ since $\vec{a}$ and $\vec{b}$ have the same diagram $\Gamma(\vec{d})$ and $q(\vec{d}) \neq 0$ is a q.f. $\tau_{\text{ring}}(\vec{d})$-sentence, which must be in $\Gamma(\vec{d})$ since $\vec{a}$ satisfies it. It is easy to verify that $h$ is a field homomorphism and hence is injective, and it is surjective because elements of $F(\vec{b})$ are of the form $\frac{p(\vec{b})}{q(\vec{b})}$, for some polynomials $p, q$ over $F$. ⊣

Without loss of generality, we can identify $F(\vec{a})$ and $F(\vec{b})$, i.e. assume that $L := F(\vec{a}) = F(\vec{b})$. Let $B_i$ be transcendence base over $L$ in $K_i$. (Transcendence base is a maximal collection of algebraically independent elements over $L$.) Now it is not hard to see that $K_i = \overline{L(B_i)}$, where $L(B_i)$ denotes the field generated by $B_i$ over $L$ and $\overline{L(B_i)}$ denotes its algebraic closure in $K_i$.

Because $L$ is countable, $|K_i| = |B_i| \cdot \aleph_0 + |L|$. If $B_i$ is countable then so is $|B_i| \cdot \aleph_0 + |L|$, but $K_i$ is uncountable, and hence $B_i$ is uncountable. Then, by basic cardinal arithmetic, $|B_i| \cdot \aleph_0 + |L| = |B_i|$ and so $\kappa = |K_i| = |B_i|$. Hence, there is a bijection $f : B_1 \to B_2$, which uniquely extends to an isomorphism of $L(B_1)$ onto $L(B_2)$ by a map similar to the one in the proof of the claim above. This isomorphism in its turn extends (not necessarily uniquely) to an isomorphism of $K_1 = \overline{L(B_1)}$ onto $K_2 = \overline{L(B_2)}$. □

**Corollary 7.12.** ACF *admits quantifier elimination.*

*Proof.* Follows from 7.10 and 7.8. □

**Corollary 7.13.** *The definable subsets of an algebraically closed field are finite or cofinite.*

*Proof.* Let $K$ be an algebraically closed field. By q.e., every definable set $S \subseteq F$ is defined by a q.f. formula $\phi(x)$. For the base case $\phi(x) \equiv (t_1(x) = t_2(x))$, the statement is clear since $t_i(x)$ is a polynomial in $x$ with coefficients in $K$ and the polynomial $t_1(x) - t_2(x)$ has only finitely many roots. The step case is also clear since the set of finite and cofinite subsets of $K$ is closed under finite unions (corresponding to $\wedge$) and complements (corresponding to $\neg$). □

**Remark.** One can also show using a similar argument that the theory of vector spaces over a countable field admits q.e. and conclude that the definable subsets of a vector space are only the finite and cofinite ones. In general, structures with only definable subsets being finite or cofinite are called strongly minimal. It turns out that in those structures one can always define an abstract model-theoretic operation that generalizes algebraic closure (for fields) and span (for vector spaces), and this operation allows to define a notion of a basis such that the rest of the structure is "free" over it in the sense that any bijection between the bases extends to a (not necessarily unique) isomorphism between the structures.

### 7.3. **Model completeness and Hilbert's Nullstellensatz**

The following is a very useful notion that is slightly weaker than quantifier elimination.

**Definition 7.14** (Model-completeness)**.** *A $\tau$-theory $T$ is called model-complete if $\mathbf{A} \subseteq \mathbf{B}$ implies $\mathbf{A} \preceq \mathbf{B}$, for all $\mathbf{A}, \mathbf{B} \vDash T$.*

**Proposition 7.15.** *Quantifier elimination implies model-completeness.*

*Proof.* Suppose $T$ admits q.e. and $\mathbf{A} \subseteq \mathbf{B}$, where $\mathbf{A}, \mathbf{B} \vDash T$. Because $\mathbf{A}$ and $\mathbf{B}$ agree on the q.f. formulas about the elements of $A$, and every formula is equivalent to a q.f. formula (in $T$), $\mathbf{A}$ and $\mathbf{B}$ agree on all formulas about the elements of $A$. □

**Remark.** In fact, model-completeness is equivalent to the statement that every formula is equivalent (in $T$) to an existential and a universal formula (recall that it was a homework problem to show that linear independence was such a formula).

Thus ACF is model-complete since it admits q.e. Note that this fact actually follows from 7.11 directly without using 7.8.

Model-completeness of ACF implies the following famous (basic) theorem of algebraic geometry:

**Hilbert's Nullstellensatz 7.16.** *Let $F$ be an algebraically closed field and $I$ be a proper ideal in the polynomial ring $F[t_1, ..., t_n]$. Then the polynomials in $I$ have a common root in $F$, i.e. there is $\vec{a} \in F^n$ such that $f(\vec{a}) = 0$ for all $f(t_1, ..., t_n) \in F[t_1, ..., t_n]$.*

*Proof.* Take a maximal ideal $M$ containing $I$ (exists by Zorn's lemma) and put

$$K := F[t_1, ..., t_n]/M.$$

Since $M$ is maximal, $K$ is a field. Note that now every polynomial in $M$ has a root in $K$ in the following sense: for $f(t_1, ..., t_n) \in M$, let $f(x_1, ..., x_n)$ be the polynomial obtained from $f(t_1, ..., t_n)$ by replacing $t_i$ with variables $x_i$ of $\mathbb{FOL}(\tau_{\text{ring}})$. Then, by the definition of $K$, for all such $f \in M$, $f(\vec{b}) = 0$, where $\vec{b} = (t_1 + M, ..., t_n + M) \in K$. (This is why we moved from $F$ to $K$: to artificially create a common root).

Let $L$ be an algebraic closure of $K$. Since $K \subseteq L$, there is still a common root in $L$ for all polynomials in $M$. Now we want to use model-completeness of $\mathsf{ACF}$ to transfer this statement down to $F$ to obtain a common root in $F$. However, expressing (in a first-order way) the statement that all polynomials in $M$ have a common root seems to be a problem because there are infinitely many polynomials in $M$ (while formulas are finite). Luckily, Hilbert's basis theorem says that any ideal in $F[t_1, ...t_n]$ is finitely generated, so $M$ is generated by some $f_1, ..., f_m \in F[t_1, ...t_n]$. Thus all polynomials in $M$ having a common root is equivalent to $f_1, ..., f_m$ having a common root. Put

$$\phi(\vec{a}) \equiv \exists \vec{x} \bigwedge_{i=1}^{m} (f_i(\vec{x}) = 0),$$

where $\vec{a} \in F^k$ is a tuple containing all coefficients of $f_1, ..., f_m$. By model-completeness of $\mathsf{ACF}$, because $\mathbf{F} \subseteq \mathbf{L}$ and $\mathbf{F}, \mathbf{L} \vDash \mathsf{ACF}$, we have $\mathbf{F} \preceq \mathbf{L}$. Hence $\mathbf{F} \vDash \phi(\vec{a})$ because $\mathbf{L} \vDash \phi(\vec{a})$, and thus $f_1, ..., f_m$ have a common root in $F$. □

## References

[End01] H. B. Enderton, *A Mathematical Introduction to Logic*, 2nd ed., Academic Press, 2001.

[Mar02] D. Marker, *Model Theory: An Introduction*, Graduate Texts in Mathematics, Springer, 2002.

[Mos08] Y. N. Moschovakis, *Informal notes full of errors*, unpublished, 2008.

[vdD10] L. van den Dries, *Mathematical Logic: Lecture Notes*, unpublished, 2010.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555, USA
*E-mail address*: anush@math.ucla.edu