An introduction to first-order logic
with emphasis on definability and o-minimality

BY VINCENT BAGAYOKO

November 2022

1 First-order logic

1.1 Structures and signatures

A *signature* is a family of *function symbols* $(\underline{f_i})_{i \in I}$ with arities $\alpha_i \in \mathbb{N}$, together with of one of two types together with a family of *relation symbols* $(\underline{R_j})_{j \in J}$ with arities $\beta_i \in \mathbb{N}$. Among the relation symbols, there is always a particular symbol of arity 2 denoted $=$.

Given a signature $\Sigma = ((\underline{f_i})_{i \in I}, (\underline{R_j})_{j \in J})$, a *first-order structure* for Σ is a set M, together with a family $(f_i)_{i \in I}$ of functions $f_i: M^{\alpha_i} \longrightarrow M$ and a family $(R_j)_{j \in J}$ of subsets $R_j \subseteq M^{\beta_j}$ called relations. Each of these defines a specific interpretation of the function and relation symbols. For instance a relation symbol $\underline{R_j}$ of arity $\beta_j = 2$ is interpreted by binary relation $R_j \subseteq M^2$, a function f_i of arity 0 is interpreted as a constant $f_i(\varnothing) \in M$, a relation of arity 1 is interpreted subset of M. The equality symbol is always interpreted as the binary relation of equality on M, i.e. as the diagonal $\{(a, a) \in M^2 : a \in M\}$.

We will consider certain properties of M that pertain to these functions and relations, and which can be stated in a specific language involving symbols for each such function and relation. This language is called a first-order language.

Example 1. For instance, if we want to talk about the properties of the ordered field $\mathbb{Q}$, then we will take two functions $\mathbb{Q}^2 \longrightarrow \mathbb{Q}$, namely the sum and the product, two constants 0 and 1, and one binary relation, namely the standard ordering on $\mathbb{Q}$, seen here as the subset $\{(a, b) \in \mathbb{Q}^2 : a < b\}$ of $\mathbb{Q}^2$.

Example 2. If we want to talk about the properties of a vector field V over $\mathbb{C}$, then we'll have the group operation $+: V^2 \longrightarrow V$ on V and, for each complex number $\lambda \in \mathbb{C}$, the scalar multiplication by λ:

$$V \longrightarrow V; x \mapsto \lambda . x.$$

1.2 First-order language over a signature

Consider a fixed signature $\Sigma = ((\underline{f_i})_{i \in I}, (\underline{R_j})_{j \in J})$. We then define specific types and sets of *words*, i.e. finite strings of symbols, as follows:

- **Variable symbols** are purely formal symbols among $x_0, x_1, \ldots, x_n, \ldots$ or $y_0, \ldots, y_n, \ldots$.
- The set $\mathcal{T}_\Sigma$ of **terms** in the signature Σ is the smallest set of finite strings of symbols among variable symbols, parentheses symbols (and) and function symbols $\underline{f_i}, i \in I$, which contains all variable symbols, and such that for all $i \in I$, if $t_1, \ldots, t_{\alpha_i} \in \mathcal{T}_\Sigma$, then $\underline{f_i}(t_1, \ldots, t_{\alpha_i}) \in \mathcal{T}_\Sigma$.

  For instance, if Σ contains two function symbols $\underline{f}$ and $\underline{g}$ of arities 1 and 2 respectively, then the word

  $$\underline{f}(\underline{g}(\underline{f}(x_2), x_0))$$

  is a term.
- **Atomic formulas** are words of the form $\underline{R_j}(t_1, \ldots, t_{\beta_j})$ where $j \in J$ and $t_1, \ldots, t_{\beta_j}$ are terms and $j \in J$. **Neg-atomic formulas** are words of the form $\neg(\underline{R_j}(t_1, \ldots, t_{\beta_j}))$ where $j \in J$ and $t_1, \ldots, t_{\beta_j}$ are terms and $j \in J$.

- **$\mathcal{L}_\Sigma$-formulas** are *well-written*[1] words involving atomic formulas, parentheses, logical connectives $\neg$ ("not", negation), $\vee$ ("or", disjunction), $\wedge$ ("and", conjunction), and quantifiers $\exists$ ("there exists", existential quantifier), and $\forall$ ("for all", universal quantifier).

The *first-order language* $\mathcal{L}_\Sigma$ over this signature is the set of all $\mathcal{L}_\Sigma$-formulas.

If a symbol of variable occuring in a formula is preceded by a quantifier in one of its occurrences, then we say that it is bound. Otherwise, we say that it is *free*. Usually, we write denote by $\varphi[x_1, \ldots, x_n]$ an $\mathcal{L}_\Sigma$-formula φ whose free variables are among $x_1, \ldots, x_n$.

A formula without free variable is called an $\mathcal{L}_\Sigma$-*sentence*. Those are the formulas which can be interpreted as true or false in structures (whereas formulas with free variables may have a truth value depending on the value of those variables).

1.3 Prenex form and quantifier-free formulas

An $\mathcal{L}$-formula is said *quantifier-free* if it contains no occurrence of $\exists$ or $\forall$. Quantifier-free formulas are thus boolean combinations of atomic formulas, i.e. obtained as conjunctions, disjunctions and negations (and combinations thereof) of atomic formulas.

Proposition 3. *Every quantifier-free formula $\varphi[x_1, \ldots, x_k]$ is is logically equivalent*[2] *to a formula of the form*

$$\bigvee_{i=1}^{m} \bigwedge_{j=1}^{n} \theta_{i,j}[x_1, \ldots, x_k],$$

where $m, n \in \mathbb{N}$, and each $\theta_{i,j}[x_1, \ldots, x_k]$ is either atomic or neg-atomic.

An $\mathcal{L}_\Sigma$-formula is in *prenex normal form* form if it is, up to permutation of the variable symbols, of the form

$$\Box_1 x_1(\Box_2 x_2(\ldots(\Box_n x_n(\theta))))$$

where θ is quantifier-free and $\Box_1, \ldots, \Box_n$ are symbols of quantifiers (i.e. $\exists$ or $\forall$).

Proposition 4. *Every $\mathcal{L}_\Sigma$-formula is logically equivalent to a formula in prenex normal form.*

1.4 Interpretation

We fix a signature Σ and an $\mathcal{L}_\Sigma$-structure $\mathcal{M} = (M, \ldots)$. Given a formula $\varphi[x_1, \ldots, x_n]$ and $a_1, \ldots, a_n \in M$, we say that $\varphi[a_1, \ldots, a_n]$ *holds* in $\mathcal{M}$ if the straightforward interpretation of $\varphi[x_1, \ldots, x_n]$, where

- each variable symbol $x_k, k \in \{1, \ldots, n\}$ is replaced by a_k,
- each function symbol $\underline{f_i}$ is replaced by the function f_i,
- each term $t(x_1, \ldots, x_n)$ is replaced by the element $t(a_1, \ldots, a_n) \in M$ accordingly,
- each atomic formula $\underline{R_j}(t_1, \ldots, t_{\beta_j})$ is replaced by the statement:

$$(t_1(a_1, \ldots, a_n), \ldots, t_{\beta_j}(a_1, \ldots, a_n)) \in R_j,$$

- each logical combination thereof is evaluated following basic logic,

is true[3].

1. Well-written means… well-written, not non-sensical. I leave this undefined and appeal to your experience as mathematicians. For instance, the word

$$\exists((x_1 \wedge \vee x_3 = \forall x_1(\neg\forall$$

is (very) badly written, hence not an $\mathcal{L}_\Sigma$-forumla. On the contrary, the word

$$\exists x_1(x_1 = x_3 \wedge (\forall x_2((x_2 = x_1) \vee \neg(x_2 = x_3))))$$

is well-written. The formal definition of $\mathcal{L}_\Sigma$-formulas is by induction on the length of words.

2. The keyword for more information on this logical equivalence is predicate calculus (Prädikatenlogik erster Stufe).

3. This is a very rough statement of Tarski's definition of satisfaction of formulas in structures. There are many subtleties, and one should study a proper introduction to logic before thinking too hard about this definition.

Remark 5. Why *first*-order? Are there higher orders? First here refers to the fact that quantifiers in the language $\mathcal{L}_\Sigma$ only apply to variables which range in elements, and not subsets of the structure. One could also have symbols of variable $X_0, \ldots, X_n, \ldots$ denoting subsets of M, so that a formula in this higher-order language for the structure $(\mathbb{R}, <)$ could state that $(\mathbb{R}, <)$ has the least upper bound property.

But the first-order language does not allow this. It can be shown that there is no set t of sentences in the first-order language over Σ with one binary relation symbol such that $\mathcal{L}_\Sigma$-structure in which all sentences in T hold are exactly linearly ordered sets with the least upper bound property.

1.5 Quantifier elimination

Consider a first-order signature Σ and an $\mathcal{L}_\Sigma$-structure $\mathcal{M}$. We say that $\mathcal{M}$ *eliminates quantifiers* (or has quantifier elimination) if for every $\mathcal{L}_\Sigma$-formula $\varphi[x_1, \ldots, x_n]$, there is a quantifier-free $\mathcal{L}_\Sigma$-formula $\theta[x_1, \ldots, x_n]$ such that the following sentence holds in $\mathcal{M}$:

$$\forall x_1, \ldots, \forall x_n(\varphi[x_1, \ldots, x_n] \Longleftrightarrow \theta[x_1, \ldots, x_n]).$$

There are many tests in order to show that $\mathcal{M}$ eliminates quantifiers. One of the most basic ones is the following

Proposition 6. *Assume that for each $n \in \mathbb{N}$ and each quantifier-free formula $\theta[x_0, \ldots, x_n, x_{n+1}]$, there is a quantifier-free formula $\psi[x_0, \ldots, x_n]$ such that the following holds in $\mathcal{M}$:*

$$\forall x_0, \ldots, x_n(\exists x_{n+1}(\theta[x_0, \ldots, x_n, x_{n+1}]) \Longleftrightarrow \psi[x_0, \ldots, x_n]).$$

Then $\mathcal{M}$ eliminates quantifiers.

Idea of proof. Prenex normal form + induction. □

2 Definability and o-minimality

In this section, we fix a first-order signature Σ and an $\mathcal{L}_\Sigma$-structure $\mathcal{M} = (M, \ldots)$.

2.1 Definable subsets

Given $n \in \mathbb{N}$, we say that a set X is *definable* in dimension n in $\mathcal{M}$ if $X \subseteq M^n$ and there are an $m \in \mathbb{N}$, a tuple $(a_1, \ldots, a_m) \in M^m$ and a formula $\varphi[x_1, \ldots, x_n, y_1, \ldots, y_n]$ such that

$$X = \varphi[a_1, \ldots, a_n, \mathcal{M}] := \{(b_1, \ldots, b_n) \in M^n : \varphi[a_1, \ldots, a_m, b_1, \ldots, b_n] \text{ holds in } \mathcal{M}\}.$$

We say that X is definable without quantifiers if $\varphi[x_1, \ldots, x_n, y_1, \ldots, y_n]$ can be taken to be quantifier-free.

Example 7. In $(\mathbb{Z}, +, \times)$, the set $\mathbb{N}$ is definable. Indeed, Lagrange's four squares theorem, an integer $n \in \mathbb{Z}$ is positive if and only if it is a sum of for squares of integers. So

$$\mathbb{N} = \{n \in \mathbb{Z} : \exists x_0, x_1, x_2, x_3(n = x_0 \times x_0 + x_1 \times x_1 + x_2 \times x_2 + x_3 \times x_3) \text{ holds}\}.$$

The set $\mathbb{N}$ is *not* definable without quantifiers. Indeed, recall that quantifier-free formulas are equivalent to boolean combinations of atomic formulas. In $(\mathbb{Z}, +, \times)$, an atomic formula in one free variable $\varphi[x_0]$ is equivalent to an equality $P(x_0) = 0$ where $P[X] \in \mathbb{Z}[X]$. Hence it defines either $\mathbb{Z}$ if $P = 0$, or a finite subset otherwise. In particular, it defines a finite or cofinite[4] subset of $\mathbb{Z}$. Since the set of finite or cofinite subsets of $\mathbb{Z}$ is closed under unions and intersections, any boolean combination ot atomic formulas defines a finite or cofinite subset. So $\varphi[x_0]$ defines a finite or cofinite subset. Since $\mathbb{N}$ is neither finite nor cofinite in $\mathbb{Z}$, it cannot be defined by $\varphi[x_0]$.

4. a subset $X \subseteq M^n$ is cofinite if its complement $M^n \setminus X$ in M^n is finite

Remark 8. In fact, in $(\mathbb{Z},+,\times)$, many things are definable. For instance, every recursively enumerable subset of $\mathbb{N}$ is definable in dimension 1 in $(\mathbb{Z},+,\times)$.

2.2 Formal-geometric correspondence

Here we fix an $m\in\mathbb{N}$, an $n\in\mathbb{N}$ with $n>0$, a tuple $a_1,\ldots,a_m\in M$, as well as $\mathcal{L}_\Sigma$-formulas

$$\begin{aligned}\varphi &= \varphi[x_1,\ldots,x_m,y_1,\ldots,y_n] \quad \text{and}\\ \psi &= \psi[x_1,\ldots,x_m,y_1,\ldots,y_n].\end{aligned}$$

We have the following list of correspondences between geometric operations on definable sets

$$\begin{aligned}X &:= \varphi[a_1,\ldots,a_m,\mathcal{M}] \quad \text{and}\\ Y &:= \psi[a_1,\ldots,a_m,\mathcal{M}],\end{aligned}$$

and their defining formulas φ and ψ.

Logical operation on defining formula	**Geometric operation on definable set**
Negation $\neg\varphi$	Complement $M^n\setminus X$
Disjunction: $(\varphi\vee\psi)$	Union $X\cup Y$
Conjunction: $(\varphi\wedge\psi)$	Intersection $X\cap Y$
Existential quantifier: $\exists y_n\,\varphi$	Projection on M^{n-1}.
Universal quantifier: $\forall y_n\,\varphi$	$\bigcup_{Z\times M\subseteq X} Z$

Thus the theory of definability can be stated in purely geometric terms. This is the spirit of [2]. However, many important results regarding definability (in particular in o-minimality) crucially rely on the interplay between intuitions coming from logic and geometry, algebra, analysis, graph theory, and so on…

2.3 O-minimality

Let $(\Gamma,<)$ be a linearly ordered set[5]. An *interval* in $(\Gamma,<)$ is a subset I of one of the following forms for $a,b\in\Gamma$:

- $(a,b):=\{\gamma\in\Gamma:a<\gamma<b\}$,
- $(-\infty,b):=\{\gamma\in\Gamma:\gamma<b\}$,
- $(a,+\infty):=\{\gamma\in\Gamma:a<\gamma\}$,
- $(-\infty,+\infty)=\Gamma$,
- $[a,b):=\{\gamma\in\Gamma:a\leqslant\gamma<b\}$,
- $(a,b]:=\{\gamma\in\Gamma:a<\gamma\leqslant b\}$,
- $[a,b]:=\{\gamma\in\Gamma:a\leqslant\gamma\leqslant b\}$,
- $(-\infty,b]:=\{\gamma\in\Gamma:\gamma\leqslant b\}$,
- $[a,+\infty):=\{\gamma\in\Gamma:a\leqslant\gamma\}$,

The first four types of intervals are called open intervals.

Example 9. The set $C:=\{q\in\mathbb{Q}^{>0}:q^2<2\}$ is not an interval in $(\mathbb{Q},<)$, whereas

$$\{r\in\mathbb{R}^{>0}:r^2<2\}=(0,\sqrt{2})$$

is an interval in $(\mathbb{R},<)$.

Proposition 10. *Let $(\Gamma,<)$ be a linearly ordered set. Then the quantifier-free definable subsets of Γ are exactly the finite unions of intervals in Γ.*

5. it is usual to use strict orderings; this is not crucial for results but it does play a role when manipulating formulas since it affects which formulas are atomic: $x_0\leqslant x_1$ or $x_0<x_1$.

Definition 11. *A first-order structure* $\mathcal{M} = (M, <, \ldots)$ *is said* **o-minimal** *if every definable subset of* $\mathcal{M}$ *in dimension* 1 *is a finite union of intervals in* $(M, <)$.

Corollary 12. *If* $(\Gamma, <)$ *eliminates quantifiers, then it is o-minimal.*

The set C in Example 9 is definable in the ordered group $(\mathbb{Q}^{>0}, \cdot, 1, <)$. This set is not a finite union of intervals, otherwise it would have a least upper bound in $\mathbb{Q}^{>0}$. Therefore $(\mathbb{Q}^{>0}, \cdot, 1, <)$ is *not* o-minimal.

O-minimal ordered fields. In the Fachseminar and in RAG I, we will see examples of o-minimal ordered fields. Here we consider the case of rational numbers. The ordered field $(\mathbb{Q}, +, \times, <)$ is not o-minimal for similar reasons as above. What's more, by a theorem of Julia ROBINSON, the set $\mathbb{Z}$ of integers is definable in $(\mathbb{Q}, +, \times)$. So every definable set in $(\mathbb{N}, +, \times)$ is definable in $(\mathbb{Q}, +, \times, <)$. In view of Remark 8, we see that $(\mathbb{R}, +, \times, <)$ and $(\mathbb{Q}, +, \times, <)$ are at the opposite ends of the spectrum of "tameness" of definable subsets.

Proposition 13. *Assume that* $\mathcal{M} = (M, <, (f_i)_{i \in I}, (R_j)_{j \in J})$ *is o-minimal, and let* $\mathcal{M}' = (M, <, (f_i)_{i \in I'}, (R_j)_{j \in J'})$ *where* $I' \subseteq I$ *and* $J' \subseteq J$, *as a first-oder structure in the language* $\mathcal{L}'$ *with signature* $((\underline{f_i})_{i \in I'}, (\underline{R_j})_{j \in J'})$. *Then* $\mathcal{M}'$ *is o-minimal.*

Proof. $\mathcal{L}'$-formulas are $\mathcal{L}$-formulas, and given an $\mathcal{L}'$-formula $\varphi[x_1, \ldots, x_n, y_0]$ and $a_1, \ldots, a_n \in M$, the definable set $\varphi[a_1, \ldots, a_n, \mathcal{M}']$ is equal to $\varphi[a_1, \ldots, a_n, \mathcal{M}]$, which is a finite union of intervals by o-minimality of $\mathcal{M}$, hence the result. □

3 Quantifier elimination for unary group representations

Let $(G, \cdot, 1, \prec)$ be a bi-ordered group. A *unary representation* of G is a total order $(X, <)$ together with a morphism

$$\begin{aligned} t_{\cdot}\colon G &\longrightarrow \operatorname{Aut}(X, <) \\ g &\longmapsto (t_g\colon x \mapsto t_g(x)) \end{aligned}$$

where $\operatorname{Aut}(X, <)$ is the group under composition of strictly increasing bijections $X \longrightarrow X$, such that for all $g, h \in G$, we have

$$g \prec h \Longrightarrow \forall x \in X, (t_g(x) < t_h(x)).$$

Example 14. We get a unary representation

$$\text{left}_{\cdot}\colon G \longrightarrow \operatorname{Aut}(G, <)$$

of G by letting it act on itself by left translations, i.e. by setting $t_g = \text{left}_g := h \mapsto g\,h$ for each $g \in G$.

Example 15. Consider the group $(G, \cdot, 1, \succ)$ with the reverse ordering $\succ$ defined by

$$g \succ h \Longleftrightarrow h \prec g.$$

This is still a bi-ordered group. We have a unary representation

$$\text{right}_{\cdot}\colon G \longrightarrow \operatorname{Aut}(G, \prec)$$

of G on its underlying linear ordering $(G, \prec)$ given by right translations: $\text{right}_g := h \mapsto h\,g^{-1}$ for each $g \in G$.

We now fix a unary representation $(X, <, t_{\cdot})$ of $(G, \cdot, 1, \prec)$, and we consider the first-order structure $(X, (t_g)_{g \in G}, <)$ in the first-order language $\mathcal{L}_G$ with unary function symbols $\underline{t_g}$ for each $g \in G$ and a binary relation symbol $<$. We then have

Theorem 16. (adapted from [1, Theorem 8]) *Assume that* $(X, <)$ *is dense or that* $(X, <, t_{\cdot}) = (G, \prec, \text{left})$. *Then* $(X, (t_g)_{g \in G}, <)$ *has quantifier elimination in* $\mathcal{L}_G$.

Proof. We first note that we have the following equivalences for $g,h\in G$ and $x,y\in X$:

$$t_g(x)<t_h(y) \iff x<t_{g^{-1}h}(y) \tag{1}$$
$$t_g(x)=t_g(y) \iff x=t_{g^{-1}h}(y) \tag{2}$$
$$\neg(t_g(x)<t_g(y)) \iff (t_g(y)<t_g(x))\vee(t_g(x)=t_g(y)) \tag{3}$$
$$\neg(t_g(x)=t_g(y)) \iff (t_g(y)<t_g(x))\vee(t_g(x)<t_g(y)). \tag{4}$$

Now consider an existential formula $\exists x\,(\varphi[x,y_1,\ldots,y_n])$ for $n\in\mathbb{N}$. So there are atomic and neg-atomic formulas $\theta_{i,j}[x,y_1,\ldots,y_n]$ such that

$$\varphi[x,y_1,\ldots,y_n]\equiv\bigvee_i\bigwedge_j\theta_{i,j}[x,y_1,\ldots,y_n].$$

Each neg-atomic formula among the $\theta_{i,j}[x,y_1,\ldots,y_n]$'s may be replaced by a disjunction of atomic formulas as in (3-4). Then as in (1-2), we may replace all atomic formulas by formulas of the form $x<t_h(y_i)$ or $x=t_h(y_i)$ for $h\in G$ and $i\in\{1,\ldots,n\}$. So each $\bigwedge_j\theta_{i,j}[x,y_1,\ldots,y_n]$ is equivalent to a formula

$$\bigvee_l\bigwedge_k\mu_{k,l}[x,y_1,\ldots,y_n]$$

where each $\bigwedge_k\mu_{k,l}[x,y_1,\ldots,y_n]$ says that

$$t_{g_1}(y_{i_1})<\cdots<t_{g_k}(y_{i_k})<x<t_{g_{k+1}}(y_{i_{k+1}})<\cdots<t_{g_m}(y_{i_m})$$

for some $m\in\{1,\ldots,n\}$, $g_1,\ldots,g_m\in G$ and $i\colon\{1,\ldots,m\}\longrightarrow\{1,\ldots,n\}$. Note that the formula $\exists x(\bigwedge_k\mu_{k,l}[x,y_1,\ldots,y_n])$ is false (hence equivalent to a quantifer-free formula) if $t_{g_1}(y_{i_1})<\cdots<t_{g_k}(y_{i_k})<t_{g_{k+1}}(y_{i_{k+1}})<\cdots<t_{g_m}(y_{i_m})$ is false. If this last formula is true, then $\exists x(\bigwedge_k\mu_{k,l}[x,y_1,\ldots,y_n])$ is equivalent to

$$t_{g_k}(y_{i_k})<x<t_{g_{k+1}}(y_{i_{k+1}}).$$

If X is densely ordered, then this last formula is always true.

If $(X,<,t.)=(G,\prec,\text{left})$ and $(G,\prec)$ is not densely ordered, then consider the least element f of $\{g\in G:g\succ 1\}$. The formula $\exists x(\bigwedge_k\mu_{k,l}[x,y_1,\ldots,y_n])$ is equivalent to

$$t_{fg_k}(y_{i_k})<t_{g_{k+1}}(y_{i_{k+1}}).$$

So in any case, the formula $\varphi[x,y_1,\ldots,y_n]$ is equivalent to a quantifier-free formula. □

Corollary 17. *Any dense total order has quantifier elimination in $\mathcal{L}_o$, and is thus o-minimal.*

Corollary 18. *The structure $(G,(\text{left}_g)_{g\in G},\prec)$ is o-minimal.*

Bibliography

[1] B. Baizhanov, J. Baldwin, and V. Verbovskiy. Cayley's theorem for ordered groups: o-minimality. *Sibirskie Èlektronnye Matematicheskie Izvestiya [electronic only]*, 4:278–281, 2007.

[2] L. van den Dries. *Tame topology and o-minimal structures*, volume 248 of *London Math. Soc. Lect. Note*. Cambridge University Press, 1998.

Index

Glossary

$\mathcal{L}_\Sigma$	first-order language over Σ	2
$\varphi[x_1, \ldots, x_n]$	formula with free variables among $x_1, \ldots, x_n$	2
left	representation of a group by left translations on itself	5