

# Order-Invariance of Two-Variable Logic is coNEXPTIME-complete

Bartosz Bednarczyk

Computational Logic Group, TU Dresden & Institute of Computer Science, University of Wrocław

---

## Abstract

We establish coNEXPTIME-completeness of the problem of deciding order-invariance of a given two variable first-order formula, improving and significantly simplifying coN2EXPTIME bound by Zeume and Harwath.

*Keywords:* satisfiability, complexity, order-invariance, classical decision problem, two-variable logic

---

## 1. Introduction

The main goal of finite model theory is to understand formal languages describing finite structures: their complexity and their expressive power. Such languages are ubiquitous in computer science, starting from descriptive complexity, where they are used to provide machine-independent characterisations of complexity classes, and ending up on database theory and knowledge-representation, where formal languages serve as fundamental querying formalism. A classical idea in finite model theory is to employ invariantly-used relations, capturing data-independence principle in databases: it makes sense to give queries the ability to exploit the presence of the order in which the data is stored in the memory but at the same time we would like to make query results independent of this specific ordering. **It is not immediately clear that the use of invariantly-used linear order in first-order logic (FO) allow us to gain anything expressivity-wise. And as soon as we stick to the arbitrary (*i.e.* not necessarily finite) structures it does not, which is a direct consequence of FO having the Craig Interpolation Property.** However, as it was first shown by Gurevich [7, Thm. 5.3], **the claim holds true over finite structures: order-invariant FO is more expressive than plain FO.**

Unfortunately, order-invariant FO is poorly understood. As stated in [1] one of the reasons why the progress in understanding order-invariance is rather slow is the lack of logical toolkit. The classical model-theoretic methods based on types were

proposed only recently [1], and the **order-invariant FO is not even a logic in the classical sense: its syntax is undecidable.** Moreover, the availability of locality-based methods is limited: order-invariant FO is known to be Gaifman-local [6, Thm. 2] but the status of an analogous of Hanf-locality from FO is open. This suggest that a good way to understand order-invariant FO is to first look at its fragments, *e.g.* the fragments with a limited number of variables.

*Our contribution.* We continue the research on the two-variable fragment ( $\text{FO}^2$ ) of order-invariant FO, initiated in [11]. It was shown, in contrast to the full FO, that checking order-invariance of an input  $\text{FO}^2$  is decidable in coN2EXPTIME, see: [11, Thm. 12]. We provide a tight bound for the mentioned problem, showing that deciding order-invariance for  $\text{FO}^2$  is coNEXPTIME-complete. **Our proof method relies on establishing exponential-size counter-example for order-invariance and is surprisingly easy.**

## 2. Preliminaries

We employ standard terminology from finite model theory, assuming that the reader is familiar with the syntax and the semantics of first-order logic (FO) [7, Sec. 2.1], basics on computability and complexity [7, Secs. 2.2–2.3], and order-invariant queries [7, Secs. 5.1–5.2]. In what follows,  $\text{FO}^2$  denotes the set of all FO sentences employing only the variables  $x, y$ .

*Structures.* Structures are denoted with fraktur letters  $\mathfrak{A}, \mathfrak{B}$  and their domains are denoted with the corresponding Roman letters  $A, B$ . We assume that structures have non-empty, *finite* domains, and are

---

*Email address:* bartosz.bednarczyk@cs.uni.wroc.pl  
(Bartosz Bednarczyk)

over some purely-relational vocabulary. For paper-specific reasons we employ three distinguished symbols  $\preceq, \preceq_0, \preceq_1$  that are interpreted as linear orders (*i.e.* a reflexive, antisymmetric, transitive and total relation). We write  $\text{FO}[\Theta]$  for  $\Theta \subseteq \{\preceq, \preceq_0, \preceq_1\}$  to indicate that only distinguished symbols from  $\Theta$  may appear in  $\varphi$ , and  $\varphi[\preceq/\preceq_i]$  to denote the formula obtained from  $\varphi$  by replacing each occurrence of  $\preceq$  with  $\preceq_i$ . We write  $\varphi(\vec{x})$  to indicate that all free variables of  $\varphi$  are in  $\vec{x}$ . A sentence is a formula without free variables. With  $\mathfrak{A}|_S$  we denote the substructure of the structure  $\mathfrak{A}$  restricted to the set  $S \subseteq A$ .

*Types.* An (atomic) 1-type over  $\tau$  is a maximal satisfiable set of atoms or negated atoms from  $\tau$  with a free variable  $x$ . Similarly, an (atomic) 2-type over  $\tau$  is a maximal satisfiable set of atoms or negated atoms with free variables  $x, y$ . Note that the total number of atomic 1- and 2-types over  $\tau$  is bounded exponentially in  $|\tau|$ . We often identify a type with the conjunction of all its elements. The set of 1-types and 2-types over the signature consisting of symbols appearing in  $\varphi$  is denoted with  $\alpha_\varphi$  and  $\beta_\varphi$ . Given a structure  $\mathfrak{A}$  and its element  $d \in A$  we say that  $d$  *realises* a 1-type  $\alpha$  if  $\alpha$  is the unique 1-type such that  $\mathfrak{A} \models \alpha[d]$ . Similarly, for distinct  $d, e \in A$ , we denote by  $\text{tp}^\mathfrak{A}(d, e)$  the unique 2-type *realised* by the pair  $(d, e)$ , *i.e.* the 2-type  $\beta$  such that  $\mathfrak{A} \models \beta[d, e]$ .

*Decision problems.* The *finite satisfiability* (resp. *validity*) *problem* for a logic  $L$  is the problem of deciding whether an input sentence  $\varphi$  from  $L$  is satisfied in some (resp. all) finite structures. Recall that the finite satisfiability and validity for  $\text{FO}$  are undecidable [10, 2], while for  $\text{FO}^2$  are respectively  $\text{NEXPTIME}$ -complete and  $\text{coNEXPTIME}$ -complete [4, Thm. 5.3][3, Thm. 3]. Note that  $\varphi$  is finitely valid iff  $\neg\varphi$  is finitely unsatisfiable.

*Order-invariance.* A sentence  $\varphi$  is *order-invariant* (or:  $\preceq$ -invariant) if for all finite  $\tau$ -structures  $\mathfrak{A}$  and all (linear-order!) interpretations  $\preceq^\mathfrak{A}, (\preceq^\mathfrak{A})'$  of  $\preceq$  over  $A$  we have that  $(\mathfrak{A}, \preceq^\mathfrak{A}) \models \varphi$  iff  $(\mathfrak{A}, (\preceq^\mathfrak{A})') \models \varphi$ . With  $\preceq\text{-inv FO}^2$  we denote the set of all  $\preceq$ -invariant  $\text{FO}^2$  sentences. Note that  $\varphi$  is *not* order-invariant if there is an  $\mathfrak{A}$  and two linear orders  $\preceq_0, \preceq_1$  on  $A$  such that  $(\mathfrak{A}, \preceq_0) \models \varphi$  and  $(\mathfrak{A}, \preceq_1) \not\models \varphi$ . Deciding if an  $\text{FO}$  sentence is order-invariant is undecidable [7, Ex. 9.3]. Checking order-invariance for  $\text{FO}^2$  formulae was shown to be in  $\text{coN2EXPTIME}$  in [11, Thm. 12].<sup>1</sup>

<sup>1</sup>The authors of [11] incorrectly stated the complexity in their Thm. 12, mistaking “invariance” with “non-invariance”.

### 3. Deciding Order-Invariance

We study the complexity of the problem of deciding if an input formula  $\varphi[\preceq] \in \text{FO}^2$  is order-invariant. We start from the lower bound first. We consider the following program, inspired by [9, Slide 9].

---

#### Procedure 1: From validity to $\preceq$ -invariance

---

**Input:** An  $\text{FO}^2[\emptyset]$ -formula  $\varphi$ .  
1 If  $\neg\varphi$  has a model with a single-element domain, **return False**. // A corner case  
2 Let  $\psi_\preceq := \exists x (P(x) \wedge \forall y (y \preceq x))$  for  $P \notin \tau$ .  
// not  $\preceq$ -inv. on structs. with  $\geq 2$  elements!  
3 **Return True** if  $(\neg\varphi) \rightarrow \psi_\preceq$  is  $\preceq$ -invariant and **False** otherwise. // Actual reduction.

---

The above program Turing-reduces finite  $\text{FO}^2$ -validity to testing order-invariance of  $\text{FO}^2$ -sentences. It is straightforward to see that the presented reduction is correct, so we left the details for the reader. From the complexity bounds on the finite validity problem for  $\text{FO}^2$  [3, Thm. 3] we conclude:

**Corollary 1.** *Procedure 1 returns **True** iff its input is finitely valid. Hence, testing whether an  $\text{FO}^2$  formula is order-invariant is  $\text{coNEXPTIME}$ -hard.*

The upper bound relies on the following fact that follows directly from the definition of order-invariance.

**Fact 2.** *An  $\text{FO}^2[\preceq]$  formula  $\varphi$  is not order-invariant iff the  $\text{FO}^2[\preceq_0, \preceq_1]$  formula  $\varphi[\preceq/\preceq_0] \wedge \neg\varphi[\preceq/\preceq_1]$  is finitely-satisfiable over structures interpreting  $\preceq_0$  and  $\preceq_1$  as linear orders over the domain.*

Let  $\text{FO}^2[\preceq_0, \preceq_1]$  be composed of all sentences of the shape  $\varphi[\preceq/\preceq_0] \wedge \neg\varphi[\preceq/\preceq_1]$  with  $\varphi \in \text{FO}^2[\preceq]$ . We stress that we always assume that  $\preceq$  symbols are interpreted as linear orders over the domain. To simplify the reasoning about such formulae, we first reduce them to Scott-like *normal forms*, cf. [4, §4], [8, Sec. 3.1]. By applying [11, Lemma 1] to  $\varphi[\preceq/\preceq_0]$  and  $\neg\varphi[\preceq/\preceq_1]$ , and taking their conjunction, we infer:

**Corollary 3.** *For any  $\text{FO}^2[\preceq_0, \preceq_1]$  formula there is an equi-satisfiable, linear-time computable formula (over an extended signature) having the form:*

$$\bigwedge_{i=0}^1 \left( \forall x \forall y \chi_i(x, y) \wedge \bigwedge_{j=1}^{m_i} \forall x \exists y \gamma_i^j(x, y) \right),$$

where the decorated  $\chi$  and  $\gamma$  are quantifier-free and the symbols  $\preceq_i$  do not appear in  $\chi_{1-i}$  and  $\gamma_{1-i}^j$ .

Given a model  $\mathfrak{A} \models \varphi$  of  $\varphi$  in normal form and elements  $d, e \in A$  witnessing  $\mathfrak{A} \models \gamma_i^j(d, e)$ , we call  $e$  a  $\gamma_i^j$ -witness for  $d$  (or simply a witness).

The core of the paper is the following small model theorem, with a proof relying on the circular witnessing scheme by Grädel-Kolaitis-Vardi [4, Thm. 4.3].

**Lemma 4.** *Any finitely satisfiable  $\varphi \in \text{FO}^2_{\leq}[\leq_0, \leq_1]$  has a model with  $\mathcal{O}(|\varphi|^3 \cdot 2^{|\varphi|})$  elements.*

*Proof.* W.l.o.g. we assume that  $\varphi$  is in the normal form from Corollary 3 and put  $M := \max(m_0, m_1)$ . Let  $\mathfrak{A} \models \varphi$ . If  $|A| \leq 224 |\varphi|^3 \cdot 2^{|\varphi|}$  then we are done, so assume otherwise. We are going to construct a new model  $\mathfrak{B}$  having the domain  $W_0 \cup W_1 \cup W_2 \cup W_3$ , where the sets  $W_i$  are constructed below.

Call 1-type *rare* if it has at most  $32M$  realisations in  $\mathfrak{A}$ . Let  $S$  be composed of all elements of  $\mathfrak{A}$  of rare 1-types, and of the  $8M$  minimal and  $8M$  maximal (w.r.t. each  $\leq_0^{\mathfrak{A}}, \leq_1^{\mathfrak{A}}$ ) realisations of each non-rare 1-type in  $\mathfrak{A}$ . We make  $W_0$  to be composed of all elements realising rare-types, as well as  $M$  minimal and  $M$  maximal (w.r.t.  $\leq_0^{\mathfrak{A}}$  and  $\leq_1^{\mathfrak{A}}$ ) realisations of each non-rare 1-type in  $\mathfrak{A}$ . Put the rest of elements of  $S$  to  $W_1$ . We have  $|W_0 \cup W_1| \leq 32M \cdot |\alpha_\varphi|$ .

The idea behind  $W_0$  is that this set contains “dangerous” elements, i.e. the ones for which  $\mathfrak{A}|_{W_0}$  may be uniquely determined by  $\varphi$ . Elements from  $W_1$  will help to restore the satisfaction of  $\forall\exists$  conjuncts.

Finally, we close  $W_0 \cup W_1$  twice under witnesses. More precisely, let  $W_2$  be any  $\subseteq$ -minimal subset of  $A$  so that all elements from  $W_0 \cup W_1$  have all the required  $\gamma_i^j$ -witnesses in  $W_0 \cup W_1 \cup W_2$ . Similarly, we define  $W_3$  to be any  $\subseteq$ -minimal subset of  $A$  so that all elements from  $W_0 \cup W_1 \cup W_2$  have all the required  $\gamma_i^j$ -witnesses in  $W_0 \cup W_1 \cup W_2 \cup W_3$ . Observe that  $|W_2| \leq 2M|W_0 \cup W_1| \leq 2M \cdot 32M|\alpha_\varphi| = 64M^2|\alpha_\varphi|$  and  $|W_3| \leq 2M|W_2| \leq 128M^3|\alpha_\varphi|$  hold.

Let  $\mathfrak{B} := \mathfrak{A}|_{W_0 \cup W_1 \cup W_2 \cup W_3}$ . We see that:

$$|B| \leq |W_0 \cup W_1| + |W_2| + |W_3| \leq (32M + 64M^2 + 128M^3)|\alpha_\varphi| \leq 224M^3|\alpha_\varphi| \leq 224 |\varphi|^3 \cdot 2^{|\varphi|}.$$

Universal formulae are preserved under substructures, thus  $\leq_1^{\mathfrak{B}}, \leq_2^{\mathfrak{B}}$  are linear orders over  $B$  and  $\mathfrak{B}$  satisfies  $\forall\forall$ -conjuncts of  $\varphi$ . Hence, the only reason for  $\mathfrak{B}$  to not be a model of  $\varphi$  is the lack of required  $\gamma_i^j$ -witnesses for elements from  $W_3$ . We fix this by reinterpreting relations between  $W_3$  and  $W_1$ .

Before we start, we are going to collect per each non-rare 1-type  $\alpha$ , pairwise-disjoint sets of  $M$  minimal and  $M$  maximal (w.r.t. each  $\leq_0^{\mathfrak{A}}, \leq_1^{\mathfrak{A}}$ ) realisations of  $\alpha$  from  $W_1$ . Formally: Fix a non-rare  $\alpha$ .

Let  $V_\alpha^0$  be composed of the first  $M$   $\leq_0$ -minimal elements from  $\mathfrak{A}|_{W_1}$ . Next, let  $V_\alpha^1$  be composed of the last  $M$   $\leq_0$ -maximal elements from  $\mathfrak{A}|_{W_1 \setminus V_\alpha^0}$ . Similarly, let  $V_\alpha^2$  be composed of the first  $M$   $\leq_1$ -minimal elements from  $\mathfrak{A}|_{W_1 \setminus (V_\alpha^0 \cup V_\alpha^1)}$ . Finally let  $V_\alpha^3$  be composed of the last  $M$   $\leq_1$ -maximal elements from  $\mathfrak{A}|_{W_1 \setminus (V_\alpha^0 \cup V_\alpha^1 \cup V_\alpha^2)}$ . Put  $V_\alpha := \bigcup_{k=0}^3 V_\alpha^k$ . Notice that all the components of  $V_\alpha$  are pairwise disjoint (by construction), and they are well-defined since we included sufficiently many elements in  $W_1$ .

Going back to the proof, we fix any element  $d$  from  $W_3$  that violate some of  $\forall\exists$ -conjuncts of  $\varphi$ . Next, fix any  $\forall\exists$ -conjunct  $\psi := \forall x \exists y \gamma_i^j(x, y)$ , whose satisfaction is violated by  $d$ . Since  $\mathfrak{A} \models \varphi$  we know that there is an element  $e \in A$  such that  $e$  is a  $\gamma_i^j$ -witness for  $d$  and  $\gamma_i^j$  in  $\mathfrak{A}$  and let  $\alpha$  be the 1-type of  $e$  in  $\mathfrak{A}$ . Observe that  $\alpha$  is not rare (otherwise  $e \in W_0$ , and hence  $e \in B$ ), and  $d \neq e$ . Moreover either  $e \leq_i^{\mathfrak{A}} d$  or  $d \leq_i^{\mathfrak{A}} e$  holds. Thus, we take  $V_\alpha^{2i+k}$  (where  $k$  equals 0 if  $e \leq_i^{\mathfrak{A}} d$  and 1 otherwise) to be the corresponding set of  $M$  minimal/maximal  $\leq_i$  realisations of  $\alpha$  in the same direction to  $d$  as  $e$  is.

Now it suffices to take the  $j$ -th element  $e_j$  from  $V_\alpha^{2i+k}$  and change the binary relations between  $d$  and  $e_j$  in  $\mathfrak{B}$  so that the equality holds  $\text{tp}^{\mathfrak{A}}(d, e) = \text{tp}^{\mathfrak{B}}(d, e_j)$  holds (which can be done as  $e$  and  $e_j$  have equal 1-types). We repeat the process for all remaining  $\gamma_i^j$  formulae violated by  $d$ . We stress that it is not a coincidence that we use the  $j$ -th element  $e_j$  from the corresponding set  $V_\alpha^{2i+k}$  to be a fresh  $\gamma_i^j$ -witness for  $d$ : this guarantees that we never redefine connection between  $d$  and some element twice.

Observe that all elements from  $B$  that had  $\gamma_i^j$ -witnesses before our redefinition of certain 2-types, still do have them (as we did not touch 2-types between them and their witnesses),  $\mathfrak{B}$  still satisfies the  $\forall\forall$ -component of  $\varphi$  (since the modified 2-type does not violate  $\varphi$  in  $\mathfrak{B}$  it does not violate  $\varphi$  in  $\mathfrak{A}$ ) and  $d$  has all required witnesses. By repeating the strategy for all the other elements from  $W_3$  violating  $\varphi$ , we obtain a desired “small” model of  $\varphi$ .  $\square$

Lemma 4 yields an NEXPTIME algorithm for deciding satisfiability of  $\text{FO}^2_{\leq}[\leq_0, \leq_1]$  formulae: convert an input into normal form, guess its exponential size model and verify the modelhood with a standard model-checking algorithm (in PTIME [5, Prop. 4.1]). After applying Proposition 2 we conclude:

**Theorem 5.** *Checking if an  $\text{FO}^2[\leq]$ -formula is  $\leq$ -invariant is coNEXPTIME-complete.*

## Acknowledgements

This work was supported by the ERC through the Consolidator Grant No. 771779 (DeciGUT). I thank Antti Kuusisto and Anna Karykowska for inspiring discussions, as well as Reijo Jaakkola, Julien Grange and Emanuel Kieroński for proofreading.

## References

- [1] Pablo Barceló and Leonid Libkin. Order-Invariant Types and Their Applications. *Log. Methods Comput. Sci.*, 12(1), 2016.
- [2] Alonzo Church. A note on the Entscheidungsproblem. *The journal of symbolic logic*, 1(1):40–41, 1936.
- [3] Martin Fürer. The Computational Complexity of the Unconstrained Limited Domino Problem (with Implications for Logical Decision Problems). *Logic and Machines: Decision Problems and Complexity*, 1983.
- [4] Erich Grädel, Phokion G. Kolaitis, and Moshe Y. Vardi. On the Decision Problem for Two-Variable First-Order Logic. *Bull. Symb. Log.*, 1997.
- [5] Erich Grädel and Martin Otto. On Logics with Two Variables. *Theor. Comput. Sci.*, 224(1-2):73–113, 1999.
- [6] Martin Grohe and Thomas Schwentick. Locality of Order-Invariant First-Order Formulas. *ACM Trans. Comput. Log.*, 1(1):112–130, 2000.
- [7] Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- [8] Martin Otto. Two Variable First-Order Logic over Ordered Domains. *J. Symb. Log.*, 66(2):685–702, 2001.
- [9] Nicole Schweikardt. A Tutorial on Order- and Arb-Invariant Logics. <https://www.irif.fr/~steiner/jifp/schweikardt.pdf>, 2013.
- [10] Alan Mathison Turing. On computable numbers, with an application to the Entscheidungsproblem. A correction. *Proceedings of the London Mathematical Society*, 2(1):544–546, 1938.
- [11] Thomas Zeume and Frederik Harwath. Order-Invariance of Two-Variable Logic is Decidable. *LICS*, 2016.