An Extension of Trakhtenbrot’s Theorem

Reijo Jaakkola
Tampere University, Finland

Abstract
The celebrated Trakhtenbrot’s theorem states that the set of finitely valid sentences of first-order logic is not computably enumerable. In this note we will extend this theorem by proving that the finite satisfiability problem of any fragment of first-order logic is RE-complete, as long as it has an effective syntax, it is equi-expressive with first-order logic over finite models and it is effectively closed under conjunction.

1 Introduction
It is well-known that both the general satisfiability problem and the finite satisfiability problem of first-order logic FO are undecidable. The general satisfiability problem was proved to be undecidable independently by Church [2] and Turing [8], while the finite satisfiability problem was proved to be undecidable by Trakhtenbrot [7], a result which implied the celebrated “Trakhtenbrot’s theorem” which states that the set of finitely valid sentences of FO is not effectively enumerable. Given these negative results, research efforts have been focused on fragments of FO, which are essentially subsets of FO obtained by imposing further syntactical restrictions on the sentences of FO. The main idea behind this type of research is that there is a fundamental trade-off between expressive power and tractability, and the goal is to find fragments which have both high expressive power and tractable satisfiability problem. For a survey of this research area we direct the reader to [1] and to the introduction of [3].

Given this goal, it is natural ask the following naïve question: is there a fragment of FO which has the same expressive power as FO and which furthermore has a decidable satisfiability problem? If we impose no restrictions on what counts as a fragment, then the answer is trivially true: one can, for instance, take all the satisfiable sentences of FO together with ⊥ (to make the resulting fragment equi-expressive with FO). However, if we make the very minimal requirement that the fragment should at least have an effective syntax, by which we mean that we can effectively determine whether a given sentence of FO belongs to this fragment, then one can show quite easily that such a fragment can not have a decidable satisfiability problem. This result follows from the fact that the set of valid sentences of FO is RE-complete, which allows one to translate FO effectively into any equi-expressive fragment with an effective syntax.

Thus in the context of general satisfiability problem there is indeed a necessary trade-off between expressive power and tractability. Next we move our attention
to the finite satisfiability problem, which has been gaining more and more interest in the recent years due to its applications in computer science. For example, if one views a sentence as a specification, then one might be only interested in violations of this specification which are finite. For several fragments these two questions coincide, since they have the finite model property, but there are also several interesting fragments which do not have this property, such as the two-variable logic with counting and the extension of Ackermann fragment with a single unary function.

Now we can ask an analogous question: is there a fragment of FO which has the same expressive power as FO and which furthermore has a decidable finite satisfiability problem? Again, to avoid banal counterexamples, we restrict our attention to fragments that have an effective syntax. By Trakhtenbrot’s theorem, the set of finitely valid sentences of FO is coRE-complete, and so we can not in general translate effectively sentences of FO into sentences of an arbitrary equi-expressive fragment with an effective syntax. In fact, it is easy to show that in the finite case the undecidability of the fragment does not follow from the fact that it has an effective syntax and it is equi-expressive with FO over finite models. Indeed, the following set

\[ \{ \bot \} \cup \{ \phi^n \mid \phi \in \text{FO and } \phi \text{ has a model of size } n \}, \]

where

\[ \phi^n := \phi \land \cdots \land \phi, \]

is a fragment that has an effective syntax, is equi-expressive with FO over finite models and it has a decidable finite satisfiability problem.

The main purpose of this note is to show that we can nevertheless prove the following.

**Theorem 1.1.** If \( \mathcal{L} \subseteq \text{FO} \) is a computable set which has the same expressive power as FO over finite models and is effectively closed under conjunction, then its finite satisfiability problem is RE-complete.

The assumptions that we are working with here are still very minimal, since practically all the well-studied decidable fragments of FO are effectively closed under conjunction. Furthermore, one could argue that at least from the point of view of applications, a decidable fragment is not very useful if it is not effectively closed under conjunction, since then one can not write down constraints in a modular fashion.

### 2 Preliminaries

We fix some reasonable enumeration \((M_x)_{x \in \{0,1\}^*}\) of all Turing machines. In particular, the binary string \(x\) encodes the Turing machine \(M_x\). We also fix some computable set \(\mathcal{V}\) of relation symbols, which will serve us as a background vocabulary. We will use FO to denote the set of all sentences of first-order logic over the vocabulary \(\mathcal{V}\).

---

\[ \text{Every satisfiable sentence has a finite model.} \]
A computable subset \( L \subseteq \text{FO} \) is a fragment of FO. Most of them are closed under conjunction syntactically, by which we mean that if \( \phi \) and \( \psi \) are sentences of a fragment then so is \( \phi \land \psi \), but there are also those fragments which are only closed under conjunction with respect to expressive power (for instance some of the prefix fragments). This motivates the following definition: we say that a fragment \( L \) of FO is effectively closed under conjunction, if there exists a computable mapping \( f : L \times L \to L \) so that \( f(\phi, \psi) \) is equivalent with \( \phi \land \psi \).

Word is a finite structure \( \mathfrak{A} \) over a vocabulary consisting of unary relation symbols and a single binary relation symbol \(<\) which satisfies the following two requirements.

1. \(<^{\mathfrak{A}}\) is a linear ordering of \( A \).
2. For every \( a \in A \) there exists precisely one unary relation symbol \( P \) in the underlying vocabulary for which \( a \in P^{\mathfrak{A}} \).

The following well-known fact is proved in [6].

**Proposition 2.1.** The satisfiability problem of FO over words is decidable.

### 3 Proof of the main result

Let \( L \) be a fragment of FO that has the same expressive power as FO over finite models and is effectively closed under conjunction. Our goal is to show that we can effectively reduce halting problem to the finite satisfiability problem of \( L \).

The chief technical obstacle that we need to bypass is that over finite models we cannot in general translate effectively FO-sentences to sentences of \( L \). However, by Proposition [2.1] we know that we can effectively determine whether two sentences FO are equivalent over words, and hence any such FO-sentence \( \phi \) can be effectively translated to a sentence \( \phi' \in L \) which is equivalent with \( \phi \) over words. Our general strategy is now that we will use these sentences to encode inputs to a universal Turing machine, which in turn will be encoded with a single fixed sentence of \( L \).

Now we will proceed with the formal proof. Let \( U \) be a Turing machine, which when given as an input a binary string \( x \), simulates the run of \( M_x \) on empty input. Thus \( U \) halts on \( x \) if and only if \( M_x \) halts on empty input. We will assume that the vocabulary of \( U \) is \( \{ 0, 1 \} \) and that \( U \) has a single one-way infinite memory tape. We will next describe sentences \( \phi_U \) and \( \phi_x \) which have the following properties.

1. \( \phi_U \land \phi_x \) is finitely satisfiable iff \( U \) halts on input \( x \).
2. For every model \( \mathfrak{A} \) of \( \phi_U \) the restriction of \( \mathfrak{A} \) to the vocabulary of \( \phi_x \) is a word.

We start with the sentence \( \phi_x \). Consider the vocabulary \( \{ E, Z, O, < \} \), where \( E, Z, O \) are unary while \( < \) is binary. Intuitively \( Z \) and \( O \) are used to indicate that an element is labelled with zero and one respectively, while \( E \) is used to indicate that the element is labelled with a “blank” symbol, i.e., it correspondence to an empty square on the tape of \( U \). \( \phi_x \) will be the following sentence:

\[
\exists x_1 \ldots \exists x_n [\text{first}_<(x_1) \land \bigwedge_{1 \leq i < n} \text{succ}(x_i, x_{i+1})]
\]
Here first$_<^\prime(z)$ describes the fact that $z$ is the smallest element in $<$, while succ$(z, w)$ describes the fact that $w$ is the immediate successor of $z$ in $<$. Thus, over words, the sentence $\phi_x$ expresses the fact that the underlying word consists of $x$ which is followed by a sequence of “blank” symbols.

Next we will consider the sentence $\phi_U$. Before describing it, we need to discuss how we are going to formalize the computations of $U$. In brief, we will describe the computation of $U$ as a finite grid, where the $i$:th row describes a sufficiently large prefix of the memory tape of $U$ at stage $i$. In addition to $\tau_0$, the vocabulary of $\phi_U$ will contain in addition at least three binary relation symbols $P_{\text{zero}}, P_{\text{one}}, P_{\text{empty}}$, which can be used to describe the content of the memory tape of $U$; and a binary relation symbol $<^\prime$ which will be used to describe a second linear ordering on the models domain. Roughly speaking, in the finite grid the element at $(i, j)$ correspondence to a pair of elements in the domain where the first element is the $i$:th element in $<$ while the second element is the $j$:th element in $<^\prime$.

Given this intuition, it is mostly routine to write down $\phi_U$. The only non-standard part of $\phi_U$ is the manner in which we encode the input received by $U$, since we want $\phi_U$ to be independent of $x$. The following sentence $\exists x[\text{first}_{<^\prime}(x) \land \forall y((Z(y) \rightarrow P_{\text{zero}}(x, y)) \land (O(y) \rightarrow P_{\text{one}}(x, y)) \land (E(y) \rightarrow P_{\text{empty}}(x, y)))], where first$_{<^\prime}(x)$ is a formula specifying that $x$ is the first element in the linear order $<^\prime$, will take care of this, since together with the sentence $\phi_x$ it encodes the fact that the bottom row in the finite grid – essentially the input tape of $U$ – starts with the binary word $x$ followed by a sequence of empty squares.

To complete the proof, we will describe how a sentence of $L$ equivalent with $\phi_U \land \phi_x$ can be obtained effectively from the input string $x$. First, since the sentence $\phi_U$ does not depend on the word $x$, we can hardcode an equivalent sentence $\phi_U^\prime \in L$ into our procedure. Secondly, since the satisfiability problem of FO over words is decidable, we can effectively find a sentence of $L$ which is equivalent with the sentence $\phi_x$ over words; let $\phi_x^\prime \in L$ denote this sentence. Since $L$ is effectively closed under conjunction, we can effectively compute a third sentence $\chi_x$ which is equivalent with $\phi_U^\prime \land \phi_x^\prime$.

4 Conclusions

In this note we have proved that there exists no computable subset of the set of sentences of FO which is equi-expressive with FO over finite models, is effectively closed under conjunction and furthermore has a decidable finite satisfiability problem. This result extends the celebrated Trakhtenbrot’s theorem, since it essentially shows that the undecidability of the finite satisfiability problem of FO can not be avoided by simply fixing the syntax of FO in an appropriate way, if we require that the resulting syntax should satisfy some very minimal requirements. More generally speaking, our result shows that the the tight connection between expressive power and decidability, which in the case of general models follows from the fact that valid sentences of FO can be enumerated by a Turing machine, extends also to the case of finite models.
Acknowledgement

This note is based on ideas that I developed while I was writing my master’s thesis. I would like to thank my advisors Antti Kuusisto and Lauri Hella for their useful and encouraging comments on these ideas.

References

[1] Egon Börger, Erich Grädel, and Yuri Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997.

[2] Alonzo Church. A note on the entscheidungsproblem. *J. Symb. Log.*, 1:40–41, 1936.

[3] Reijo Jaakkola and Antti Kuusisto. Algebraic classifications for fragments of first-order logic and beyond. *arXiv Preprint*, arXiv:2005.01184v2, 2021.

[4] Ian Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *Journal of Logic, Language and Information*, 14(3):369–395, 2005.

[5] Saharon Shelah. Decidability of a portion of the predicate calculus. *Israel Journal of Mathematics*, 28:32–44, 1977.

[6] Larry Stockmeyer. *The Complexity of Decision Problems in Automata Theory and Logic*. Massachusetts Institute of Technology, Project MAC, 1974.

[7] Boris Trakhtenbrot. Impossibility of an algorithm for the decision problem on finite classes (in russian). *Doklady Akademii Nauk SSSR*, 70:569–572, 1950.

[8] Alan Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of The London Mathematical Society*, 41:230–265, 1937.