ISSN 1064-5624, Doklady Mathematics, 2022, Vol. 106, No. 3, pp. 458–461. © The Author(s), 2022. This article is an open access publication.
ISSN 1064-5624, Doklady Mathematics, 2022. © The Author(s), 2022. This article is an open access publication.
Russian Text © The Author(s), 2022, published in Doklady Rossiiskoi Akademii Nauk. Matematika, Informatika, Protsessy Upravleniya, 2022, Vol. 507, pp. 61–65.

MATHEMATICS

Computational Complexity of Theories of a Binary Predicate with a Small Number of Variables

M. Rybakov[a,*]

Presented by Academician of the RAS A.L. Semenov

Received August 7, 2022; revised September 19, 2022; accepted September 23, 2022

Abstract—We prove Σ^0_1-hardness of a number of theories of a binary predicate with three individual variables (in languages without constants or equality). We also show that, in languages with equality and the operators of composition and of transitive closure, theories of a binary predicate are Π^1_1-hard with only two individual variables.

Keywords: first-order logic, dyadic logic, undecidability, Church's theorem, Trakhtenbrot's theorem

DOI: 10.1134/S1064562422700053

INTRODUCTION

The paper concerns the so-called classical decision problem [1], in particular, the algorithmic complexity of fragments of elementary theories [2, 3]. Below we deal with theories of a binary predicate, both in languages with a binary predicate letter only, and those enriched with some additional tools.

The classical logic of a binary predicate is undecidable [4, Chapter 21]. Note that the corresponding proof [4] uses infinitely many individual variables; at the same time, to prove that the first-order logic is undecidable, it suffices to use only three individual variables and, alongside a binary predicate letter, infinitely many unary letters [5]. If we only allow one of the following: only two variables [6, 7], only unary predicated letters and equality [4, Chapter 21], or only guarded formulas (where binary letters are allowed only with certain limitations) [8], then we obtain decidable fragments. A natural question, thus, arises concerning decidability of the logic and theories of a single binary predicate in languages with a finite number (at least, three) of variables.

A similar situation applies to first-order languages enriched with operators motivated by applications. For example, the first-order logic with equality and the transitive closure operator is Π^1_1-hard in languages with two variables, but the proof uses several binary predicate letters and infinitely many unary ones [9]; again, the question arises concerning the computational complexity of theories of a binary predicate with a finite number (in this case, at least two) of variables.

The answer to the question about the decidability of the logic of a binary predicate in languages with at least three variables (and without additional operators) follows from [10]: it is undecidable (see [10, clause (ii) of Section 4.8]). Similar results for the classical logic of a binary predicate in languages with two variables, enriched with additional operators, are unknown to the author.

We shall describe a construction providing us with a short proof, firstly, of the undecidability of many theories of a binary predicate in languages with three individual variables (in particular, Σ^0_1-completeness of the logic of a binary predicate and Π^0_1-completeness of the theory of finite models of a binary predicate), and secondly, Π^1_1-hardness of the validity problem for languages with a binary predicate letter, equality, and two individual variables, enriched with the operators of composition and transitive closure. The construction will consist of modelling tiling problems [11, 12]; note that this is a well-known method, with applications both in algebra [13, 14] and in logic [1, 15–18].

An undecidable tiling problem. *Tiles* are squares, all of the same size; *type* t of a tile is determined by the *colors* $\mathtt{l}t$, $\mathtt{r}t$, $\mathtt{u}t$, and $\mathtt{d}t$ of its edges. The following tiling problem is Π^0_1-complete [11]: given a set $T = \{t_0, \ldots, t_n\}$ of tile types, we are to determine if there exists a T-*tiling*, i.e., a map $f : \mathbb{N}\times\mathbb{N} \to T$ such that $\mathtt{r}f(i, j) = \mathtt{l}f(i+1, j)$ and $\mathtt{u}f(i, j) = \mathtt{d}f(i, j+1)$, for every $i, j \in \mathbb{N}$.

[a] *Institute for Information Transmission Problems (Kharkevich Institute) of Russian Academy of Sciences, Moscow, Russia*
**e-mail: m_rybakov@mail.ru*

Modelling a tiling. Fix a binary predicate letter P and introduce the following abbreviations: $rx = xPx$, $\overline{r}x = \neg rx$, $gx = \exists y(ry \wedge xPy)$, $\overline{g}x = \neg gx$, and

$$xEy = \forall z((xPz \leftrightarrow yPz) \wedge (zPx \leftrightarrow zPy));$$
$$xSy = xPy \wedge \neg xEy \wedge \forall z(xPz \wedge zPy \rightarrow zEx \vee zEy);$$
$$xHy = xSy \wedge (rx \leftrightarrow ry);$$
$$xVy = xSy \wedge (rx \leftrightarrow \overline{r}y).$$

If rx, then x is reflexive, and if $\overline{r}x$, then x is irreflexive; gx means that x belongs to an $\mathbb{N}\times\mathbb{N}$ grid, i.e., x is a tile-holder; E induces a congruence; S means a single P-step; and H and V are understood as moving, respectively, rightwards and upwards. To obtain a definition of xSz, replace, in the definition of xSy, both y with z and z with y; analogously for ySz, zSy, xEz, etc. For a property u, let $\forall_u xA = \forall x(ux \rightarrow A)$ and $\exists_u xA = \exists x(ux \wedge A)$.

Define G be the conjunction of the formulas $\forall x\forall y\forall z(xPy \wedge yPz \rightarrow xPz)$, $\forall_g x\exists_g y xHy$, $\forall_g x\exists_g y xVy$, $\forall_g x\forall_g y\ (\exists z(xHz \wedge zVy) \leftrightarrow \exists z(xVz \wedge zHy))$, and $\exists_g x\overline{r}x$. It should be clear that if G is true in a model M, then M contains an $\mathbb{N}\times\mathbb{N}$ grid whose rows are determined by H and columns by V; the elements of the first row are irreflexive, the elements of the second row are reflexive, the the elements of the third row are irreflexive, etc.

Let $h^y x = \neg\exists y xPy$, $t_0^y x = \exists_{\overline{g}} y(xSy \wedge h^x y)$, $t_{k+1}^y x = \exists_{\overline{g}} y(xSy \wedge t_k^x y)$, where $k \in \mathbb{N}$. Formula $t_m^y x$ says that x sees an S-successor outside the grid, and that the said successor sees a dead end in m S-steps. We write $t_m x$ and $t_m y$ instead of $t_m^y x$ and $t_m^x y$. The intended meaning of $t_m x$ is that x contains a tile of type t_m.

Define t_T be the conjunction of the following formulas:

$$\forall_g x \bigvee_{i=0}^{n} (t_i x \wedge \bigwedge_{j\neq i} \neg t_j x);$$
$$\forall_g x \bigvee_{i=0}^{n} (t_i x \rightarrow \forall_g y(xHy \rightarrow \bigvee_{\mathrm{r}t_i = \mathrm{l}t_j} t_j y));$$
$$\forall_g x \bigvee_{i=0}^{n} (t_i x \rightarrow \forall_g y(xVy \rightarrow \bigvee_{\mathrm{u}t_i = \mathrm{d}t_j} t_j y)).$$

Lemma 1. *$G \wedge t_T$ is satisfiable $\Leftrightarrow$ there exists a T-tiling.*

Observe that, in the arguments given above, we can make do with *positive* formulas only: first, replace $G \wedge t_T$ with its negation and the satisfiability problem with the refutability problem; then, replace every negation with implication to the formula $\forall x\forall y xPy$.

Theories of a binary predicate. By Lemma 1 and the observation concerning positive formulas, we obtain the following refinement of the Church's theorem [19]:

- the positive fragment of the classical predicate logic is Σ_1^0-complete in the language with a single binary predicate letter and three individual variables.

Theorem 1 can be extended to theories of a binary predicate: it suffices to make minor adjustments in the above-given encoding. For a class C of models of a binary predicate, let us denote by $Th(C)$ the elementary theory of the class C. Let F, If, R, Ir, S, As, T, and It be the classes of, respectively, finite, infinite, reflexive, irreflexive, symmetric, antisymmetric, transitive, and intransitive models of a binary predicate. If X and Y are classes of models, then we will write XY rather than $X \cap Y$. So, $IfRST$ is the class of infinite models that are reflexive, symmetric, and transitive.

Theorem 1. *Let C be a class of models of a binary predicate containing at least one of the following classes: $IfIrT$, $IfIt$, $IfRS$, $IfIrS$. Then the positive fragment of $Th(C)$ is Σ_1^0-hard in the language with three individual variables.*

As a corollary, the elementary theories of the classes If, R, Ir, S, As, T, and It are not decidable in languages with three individual variables.

Using a similar argument, we obtain the following refinement of Trakhtenbrot's theorem [20, 22]:

- the positive fragment of the theory of finite models is Π_1^0-complete in languages with a single binary predicate letter and three individual variables.

To prove this statement, it is enough to notice that one can encode with a suitable tiling problem the problem of non-termination of Turing machines on the empty tape, and then to adjust the formulas given above so that they say that the first row of a tiling corresponds to the initial configuration of a Turing machine on the empty tape and that the initial tiling segment does not contain a tile corresponding to a final state of the machine; note that another possible way is to use the effective inseparability [21] and Theorem 1.

A more general statement is also true.

Theorem 2. *Let C be a class of finite models of a binary predicate containing FR or $FIrAsT$. Then the positive fragment of $Th(C)$ is Π_1^0-hard in languages with three individual variables.*

Enriching the language. We shall now show that logics and theories in enriched languages can be highly undecidable, even with only two variables. Consider the tiling problem where the T-tiling f is required to additionally satisfy the condition that the set $\{j \in \mathbb{N} : f(0, j) = t_0\}$ is infinite. This problem is known to be Σ_1^1-complete [12, Theorem 6.4].

Extend the language of first-order logic with the transitive closure operator and denote by P^+ the transitive closure of P. For a formula A, define A' to be the formula obtained from A by uniformly replacing P with P^+. Let $lx = \neg\exists y\, yH'x$. Define G^* be the conjunction of the formulas $\forall x \forall y(xPy \wedge \neg yPx \to xS'z)$, $\forall_{g'} x \exists_{g'} y\, xH'y$, $\forall_{g'} x \exists_{g'} y\, xV'y$, $\forall_{g'} x \exists_{g'} y(\exists z(xH'z \wedge zV'y) \leftrightarrow \exists z(xV'z \wedge zH'y))$, $\exists_{g'} x(\bar{r}x \wedge lx)$, $\forall_{g'} x(lx \to \forall_{g'} y(xVy \to ly))$, and t_T^* to be the conjunction of t_T' and the formula $\forall_{g'} x(lx \to \exists_{g'} y(\neg xE'y \wedge xP^+y \wedge ly \wedge t_0'y))$. Then, $G^* \wedge t_T^*$ is satisfiable if, and only if, there exists a T-tiling with the extra condition. This proves Π_1^1-hardness of a number of theories of a binary predicate letter with three variables; note that the transitive closure operator was applied only to atomic formulas.

If, in addition, we have equality and composition in the language, then, using ideas from [9], we can describe the same tiling problem with formulas containing only two variables and only one binary predicate letter. The third variable is used in the definitions of E and S, as well as in one of the conjuncts of the formula G. Replace E with equality. Using equality and the transitive closure operator, one can define the following properties of a binary relation: functionality, surjectivity, and disjointness of its domain and its range, see [9]. This allows us to divide steps alongside H and V into "even" steps H_0 and V_0 and "odd" steps H_0, H_1, V_0, and V_1; also, we can use P instead of S in $t_m x$; then, there is no need for S any more. The conjunct of G with three variables can be replaced with $\forall_{g'} x \exists_{g'} y(x[H \circ V]y \wedge x[V \circ H]y)$, where $xHy = xH_0y \vee xH_1y$ and $xVy = xV_0y \vee xV_1y$. Again, we obtain Π_1^1-hardness of theories, but this time with two variables.

As a result, we obtain the following theorem.

Theorem 3. *The validity problem for languages with two individual language, binary predicate letter, equality and the operators of composition and transitive closure is* Π_1^1*-hard.*

Note that the validity problem for such languages in the class of all finite models is in the class Π_1^0, since it is possible to effectively enumerate both all formulas and all finite models (up to isomorphism), which makes it possible to construct an effective enumeration of the set of refutable formulas.

DISCUSSION

Note that slightly weakened versions of Theorems 1 and 2 can be obtained from [10] in view of the results presented in [2, 3, 22] and in other papers (see, for example, [23; 24, Appendix]). Thus, in [10], the undecidability of the logic of a binary predicate in the language with three individual variables is proved, and using the translations from [2, 3], one can obtain undecidability (and, in view of [22], Σ_1^0-hardness or Π_1^0-hardness) of various theories of a binary predicate in languages with a finite (sometimes, perhaps, quite large) number of individual variables; the idea is to eliminate some variables in translations when nested quantifiers appear in formulas, roughly as in the definition of formulas $t_m(x)$ above, i.e., to reuse a variable x in the recursive clause for a formula φ whenever x does not occur freely in φ, rather than introducing a new variable.

We also note that a lot of attention has been devoted (see, for example, [1, 3]) to the study of the computational properties of elementary fragments defined by quantifier prefixes from some regular set; if the said set is infinite, then it contains arbitrarily long quantifier prefixes; hence, the corresponding fragment of the language contains infinitely many individual variables. Sets of quantifier prefixes leading to undecidability of the logic of a binary predicate determined are infinite. Thus, a natural question arises: is it possible to extract from the above construction a proof of the undecidability of some fragment of the logic of a binary predicate defined by a finite set of quantifier prefixes? The answer is negative: the formulas $t_m^y(x)$ use nested quantifiers for the variables x and y, and their quantifier depth depends on m; therefore, converting formulas of the form t_T to prefix normal form and increasing the number of elements in T, we get an increase in the length of the quantifier prefix, which leads to an increase in the number of variables in the resultant formula.

ACKNOWLEDGMENTS

The author expresses sincere gratitude to anonymous reviewers for their comments and suggestions, as well as for discussing the topic of the work and the results.

FUNDING

The research was carried out at the IITP RAS with the financial support of the Russian Science Foundation, project no. 21-18-00195.

CONFLICT OF INTEREST

The author declares that he has no conflicts of interest.

OPEN ACCESS

REFERENCES

1. E. Börger, E. Grädel, and Yu. Gurevich, *The Classical Decision Problem* (Springer, Berlin, 1997).
2. Yu. L. Ershov, I. A. Lavrov, A. D. Taimanov, and M. A. Taitslin, "Elementary theories," Russ. Math. Surv. **20** (4), 35–105 (1965).
3. A. Nies, "Undecidable fragments of elementary theories," Algebra Univers. **35**, 8–33 (1996).
4. G. S. Boolos, J. P. Burgess, and R. C. Jeffrey, *Computability and Logic*, 5th ed. (Cambridge Univ. Press, Cambridge, 2007).
5. J. Surányi, "Zur Reduktion des Entscheidungsproblems des logischen Funktioskalküls," Math. Fiz. Lapok **50**, 51–74 (1943).
6. M. Mortimer, "On languages with two variables," Z. Math. Logik Grundlagen Math. **21**, 135–140 (1975).
7. E. Grädel, P. G. Kolaitis, and M. Y. Vardi, "On the decision problem for two-variable first-order logic," Bull. Symb. Logic **3** (1), 53–69 (1997).
8. E. Grädel, "On the restraining power of guards," J. Symb. Logic **64** (4), 1719–1742 (1999).
9. E. Grädel, M. Otto, and E. Rosen, "Undecidability results on two-variable logics," Arch. Math. Logic **38**, 313–354 (1999).
10. A. Tarski and S. Givant, *A Formalization of Set Theory without Variables* (Am. Math. Soc., Providence, R.I., 1987).
11. R. Berger, *The Undecidability of the Domino Problem* (Am. Math. Soc., Providence, R.I., 1966).
12. D. Harel, "Effective transformations on infinite trees, with applications to high undecidability, dominoes, and fairness," J. ACM **33**, 224–248 (1986).
13. J. H. Conway and J. C. Lagarias, "Tiling with polyominoes and combinatorial group theory," J. Combin. Theory Ser. A **53** (2), 183–208 (1990).
14. J. Kari and P. Papasoglu, "Deterministic aperiodic tile sets," Geom. Funct. Anal. **9**, 353–369 (1999).
15. M. Reynolds and M. Zakharyaschev, "On the products of linear modal logics," J. Logic Comput. **11**, 909–931 (2001).
16. R. Kontchakov, A. Kurucz, and M. Zakharyaschev, "Undecidability of first-order intuitionistic and modal logics with two variables," Bull. Symb. Logic **11**, 428–438 (2005).
17. M. Rybakov and D. Shkatov, "Algorithmic properties of first-order modal logics of linear Kripke frames in restricted languages," J. Logic Comput. **31** (5), 1266–1288 (2021).
18. M. Rybakov and D. Shkatov, "Undecidability of QLTL and QCTL with two variables and one monadic predicate letter," Logical Invest. **27** (2), 93–120 (2021).
19. A. Church, "A note on the 'Entscheidungsproblem'," J. Symb. Logic **1**, 40–41 (1936).
20. B. A. Trakhtenbrot, "Impossibility of an algorithm for the decision problem in finite classes," Am. Math. Soc. Transl. **23**, 1–5 (1963).
21. B. A. Trakhtenbrot, "On recursive separability," Dokl. Akad. Nauk SSSR **88** (6), 953–956 (1953).
22. S. Speranski, "A note on hereditarily Π_1^0- and Σ_1^0-complete sets of sentences," J. Logic Comput. **26** (5), 1729–1741 (2016).
23. A. Nerode and R. A. Shore, "Second order logic and first order theories of reducibility orderings," *The Kleene Symposium*, Ed. by J. Barwise, H. J. Keisler, and K. Kunen (North-Holland, Amsterdam, 1980), pp. 181–200.
24. P. Kremer, "On the complexity of propositional quantification in intuitionistic logic," J. Symb. Logic **62** (2), 529–544 (1997).