HAL open science

Ehrenfeucht-Fraïssé Goes Automatic for Real Addition

Felix Klaedtke

▶ To cite this version:

Felix Klaedtke. Ehrenfeucht-Fraïssé Goes Automatic for Real Addition. STACS 2008, Feb 2008, Bordeaux, France. pp.445-456. hal-00232963

HAL Id: hal-00232963

https://hal.science/hal-00232963v1

Submitted on 3 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Symposium on Theoretical Aspects of Computer Science 2008 (Bordeaux), pp. 445-456
www.stacs-conf.org

EHRENFEUCHT-FRAÏSSÉ GOES AUTOMATIC FOR REAL ADDITION

FELIX KLAEDTKE

ETH Zurich, Department of Computer Science, Switzerland
E-mail address: felixkl@inf.ethz.ch

Abstract. Various logical theories can be decided by automata-theoretic methods. Notable examples are Presburger arithmetic $\mathrm{FO}(\mathbb{Z}, +, <)$ and the linear arithmetic over the reals $\mathrm{FO}(\mathbb{R}, +, <)$, for which effective decision procedures can be built using automata. Despite the practical use of automata to decide logical theories, many research questions are still only partly answered in this area. One of these questions is the complexity of such decision procedures and the related question about the minimal size of the automata of the languages that can be described by formulas in the respective logic. In this paper, we establish a double exponential upper bound on the automata size for $\mathrm{FO}(\mathbb{R}, +, <)$ and an exponential upper bound for the discrete order over the integers $\mathrm{FO}(\mathbb{Z}, <)$. The proofs of these upper bounds are based on Ehrenfeucht-Fraïssé games. The application of this mathematical tool has a similar flavor as in computational complexity theory, where it can often be used to establish tight upper bounds of the decision problem for logical theories.

1. Introduction

Various logical theories admit automata-based decision procedures. The idea of using automata-theoretic methods to decide logical theories goes at least back to Büchi [7]. The elements of the domain of the logical theory are encoded by words over some alphabet in such a way that equality and the relations of the logical theory correspond to regular languages. In order to decide whether a formula is satisfiable, one constructs an automaton that precisely accepts the representatives of the elements that satisfy the formula. This automaton can be constructed by recursion over the formula structure, where standard automata constructions handle the boolean connectives and quantifiers. The satisfiability problem is thus reduced to the emptiness problem for automata.

The logical theories that admit such automata-based decision procedures are often called automatic and they have been systematically studied, e.g., in [4, 12, 13]. Prominent and practically relevant examples are the weak monadic second-order theory of one successor WS1S, Presburger arithmetic $\mathrm{FO}(\mathbb{Z}, +, <)$, and the linear arithmetic over the reals $\mathrm{FO}(\mathbb{R}, +, <)$, see, e.g., [5–7]. Tools like MONA [15] and LIRA [3], which have been applied to various verification problems, implement such automata-based decision procedures for

1998 ACM Subject Classification: F.1.1,F.4.1.

Key words and phrases: automata theory, automata-based decision procedures for logical theories, upper bounds, minimal sizes of automata, linear arithmetic over the reals, first-order equivalence, complexity.

This work was supported by the Swiss National Science Foundation (SNF).

SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

© F. Klaedtke
Creative Commons Attribution-NoDerivs License

logical theories such as WS1S, Presburger arithmetic, and the linear arithmetic over the reals. Furthermore, model checkers for counter systems like FAST [1, 2] use an automata-based representation of sets definable in Presburger arithmetic.

A crude complexity analysis of an automata-based decision procedure leads to a non-elementary worst-case complexity. Namely, for every quantifier alternation there is a potential exponential blow-up in the state space of the automaton. For WS1S, this wost-case scenario actually exists, since the decision problem for WS1S has a non-elementary worst-case complexity [20, 23]. However, for many other automatic logical theories, the non-elementary complexity upper bounds of automata-based decision procedures often contrasts with the known computational complexity upper bounds on the decision problems for the logical theories. Moreover, such exponential blow-ups in the state spaces of the automata are rarely observed in practice in automata-based decision procedures for Presburger arithmetic and the linear arithmetic over the reals. In fact, in many cases, one obtains a smaller automaton after eliminating a quantifier. However, only partial answers exist that explain this phenomenon.

In [14], it is shown that the size of the minimal deterministic automaton that represents a Presburger definable set is triply exponentially bounded with respect to the formula length. This upper bound is established by comparing the automata for Presburger arithmetic formulas with the formulas produced by Reddy and Loveland's quantifier-elimination method for Presburger arithmetic [22]. The proof on the upper bound in [14] is rather tedious in the sense that several auxiliary upper bounds on the formulas that are generated by the quantifier-elimination method need to be established. These additional upper bounds depend on Reddy and Loveland's quantifier-elimination method. With the slightly different quantifier-elimination method by Cooper [8], we obtain an upper bound on the automata size that has at least one additional exponent.

For the linear arithmetic over the reals, the approach of using quantifier-elimination methods to establish upper bounds on the automata sizes does not lead to a satisfactory result: an application of this approach establishes only a triple exponential upper bound on the automata size when using the quantifier-elimination method for the linear arithmetic over the reals described in [10]. The author is not aware of any quantifier-elimination method for the linear arithmetic over the reals that would lead to a upper bound on the automata size that is smaller than triple exponential. However, since there are decision procedures for the linear arithmetic over the reals that run in double exponential deterministic time [10], one might conjecture that the automata size is also doubly exponentially bounded.

The main result of this paper proves this conjecture. The presented proof of the double exponential upper bound is based on Ehrenfeucht-Fraïssé games (EF-games, for short from now on). It relates the states of a minimal automaton for a formula and the equivalence classes of a refinement of the equivalence relation determined by EF-games played over $(\mathbb{R}, +, <)$. This proof technique can also be used for other automatic logical theories to establish tight upper bounds on the automata sizes. As another example, we establish an exponential upper bound on the automata size for $\mathrm{FO}(\mathbb{Z}, <)$. Note that the best known deterministic algorithms that decide $\mathrm{FO}(\mathbb{Z}, <)$ run in exponential time [11]. In summary, the results presented in this paper shed light on the complexity of automata-based decision procedures for logical theories by identifying a relationship to EF-games.

It is worth pointing out that EF-games have already been used in similar contexts. Closely related to our work is Ladner's work [17]. He uses EF-games to show decidability of monadic second-order theories of one successor and first fragments of it. Similar to

this paper, he relates the equivalence classes determined by EF-games to automata states. However, Ladner does not focus on the automata sizes and he does not consider $\mathrm{FO}(\mathbb{R}, +, <)$.

The use of EF-games in computational complexity theory [11] and constraint databases [21] is reminiscent of their use in this paper by partitioning the domain and connecting such a partition to the definable sets. Roughly speaking, the use EF-games for establish upper bounds on the decision problem for logical theories is as follows: The key ingredient for obtaining an upper bound for the respective logical theory is to show that the quantifiers, which can range over an infinite domain, can be relativized to a finite subset. Usually, one uses EF-games here to establish upper bounds on the sizes of such sets by analyzing the information that the formulas of a certain quantifier depth can convey. Given such a result on relativizing the quantifiers, satisfiability of a formula can be checked by an exhaustive search. The upper bounds on the sizes of the sets over which the relativized quantifiers range in turn yield upper bounds on the time and space that is needed to perform this search. For several logical theories, this use of EF-games yield tight upper bounds on the computational complexity for their decision problem.

The remainder of the paper is organized as follows. In §2, we give preliminaries. In §3, we illustrate our method by analyzing the languages that are $\mathrm{FO}(\mathbb{Z}, <)$-definable. In §4, we analyze the languages that are $\mathrm{FO}(\mathbb{R}, +, <)$-definable and establish the double exponential upper bound on the automata size. Finally, in §5, we draw conclusions. Due to space restrictions some proofs are omitted or sketched. They can be found in the full version of the paper, which is available from the author's web-page.

2. Preliminaries

We assume that the reader is familiar with first-order logic and automata theory over finite and infinite words. Here, we recall the needed background in these areas and fix the notation and terminology that we use in the remainder of the text.

2.1. Words and Languages

Let Σ be an alphabet. We denote the set of all finite words over Σ by Σ^* and Σ^+ denotes the set $\Sigma^* \setminus \{\varepsilon\}$, where ε is the empty word. Σ^ω is the set of all ω-words over Σ. The *concatenation* of words is written as juxtaposition. We write $|w|$ for the *length* of $w \in \Sigma^*$. We often write a word $w \in \Sigma^*$ of length $\ell \geq 0$ as $w(0) \dots w(\ell - 1)$ and an ω-word $\alpha \in \Sigma^\omega$ as $\alpha(0)\alpha(1)\alpha(2)\dots$, where $w(i)$ and $\alpha(i)$ denote the ith letter of w and α, respectively.

For a language $L \subseteq \Sigma^*$, the Nerode relation $\sim_L \subseteq \Sigma^* \times \Sigma^*$ is defined as $u \sim_L v$ iff for all $w \in \Sigma^*$, it holds that $uw \in L \Leftrightarrow vw \in L$. Analogously, for an ω-language $L \subseteq \Sigma^\omega$, we define $\sim_L \subseteq \Sigma^* \times \Sigma^*$ as $u \sim_L v$ iff for all $\gamma \in \Sigma^\omega$, it holds that $u\gamma \in L \Leftrightarrow v\gamma \in L$.

2.2. First-order Logic

The (first-order) formulas over a signature are defined as usual: they are built from variables $v_0, v_1, \dots$, the symbol $\approx$ for equality, the atomic formulas over the signature, the boolean connectives $\neg$ and $\vee$, and the quantifier $\exists$. In this paper, we only consider signatures that consist of relation symbols. The signature, its relation symbols, and the arities of its relation symbols are always clear from the context. We write $\varphi(x_1, \dots, x_r)$ when at most

the variables $x_1, \ldots, x_r$ occur free in the formula φ. The *quantifier depth* of a formula φ is recursively defined as

$$\mathrm{qd}(\varphi) := \begin{cases} \mathrm{qd}(\psi) & \text{if } \varphi = \neg\psi, \\ \max\{\mathrm{qd}(\psi), \mathrm{qd}(\psi')\} & \text{if } \varphi = \psi \vee \psi', \\ 1 + \mathrm{qd}(\psi) & \text{if } \varphi = \exists x \psi, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

A (first-order) structure over a signature consists of a nonempty universe U and it associates with each relation symbol in the signature a relation over U^r, where r is the arity of the relation symbol. We use $\mathfrak{R}$ and $\mathfrak{Z}$ to denote the structures $(\mathbb{R}, +, <)$ and $(\mathbb{Z}, <)$, respectively, where $+$ is the ternary addition relation and $<$ is the ordering relation over the reals or the integers, respectively.

Let $\mathfrak{A}$ be a structure over some signature and with the universe A. For $a_1, \ldots, a_r \in A$ and a formula $\varphi(x_1, \ldots, x_r)$, we write $\mathfrak{A} \models \varphi[a_1 \ldots, a_r]$ if φ is satisfied in $\mathfrak{A}$ when the variable x_i is interpreted as a_i, for all $1 \leq i \leq r$. For the sake of brevity, we often write $\bar{x}$ and $\bar{a}$ instead of $x_1, \ldots, x_r$ and $a_1, \ldots, a_r$, respectively.

Let $m, r \in \mathbb{N}$, $\bar{a} \in A^r$ and $\bar{b} \in A^r$. We write $\bar{a} \equiv^r_m \bar{b}$ if for all formulas $\varphi(x_1, \ldots, x_r)$ with $\mathrm{qd}(\varphi) \leq m$, it holds that $\mathfrak{A} \models \varphi[\bar{a}] \Leftrightarrow \mathfrak{A} \models \varphi[\bar{b}]$. Note that the relation $\equiv^r_m$ partitions the elements of A^r. The equivalence classes of $\equiv^r_m$ can be game-theoretically characterized by so-called Ehrenfeucht-Fraïssé games. For details on these games, see, for instance, [9]. Instead of working directly with $\equiv^r_m$, we work with refinements of it, since the reasoning about a well-chosen refinement of $\equiv^r_m$ simplifies matters. In particular, it might be difficult for $\equiv^r_m$ to directly establish an upper bound on the index of $\equiv^r_m$, to identify elements $\bar{a}, \bar{b} \in A^r$ that are in the same equivalence class, and to find a representative of an equivalence class.

2.3. Representation of Sets Definable in Real Addition

Boigelot, Jodogne, and Wolper have shown in [5] that every first-order definable set $X \subseteq \mathbb{R}^r$ in $\mathfrak{R}$ determines an ω-language L that is in the Borel class $F_\sigma \cap G_\delta$. In other words, L can be accepted by a so-called weak deterministic Büchi automaton. In fact, Boigelot, Jodogne, and Wolper have established in [5] a stronger result. First, they have proved the result for an extension of $\mathfrak{R}$ with the additional predicate $\mathbb{Z}$. Second, for a formula $\varphi(x_1, \ldots, x_r)$ over this extended structure, they have shown how to effectively construct a weak deterministic Büchi automaton that represents the set $\{\bar{a} \in \mathbb{R}^r : \mathfrak{R} \models \varphi[\bar{a}]\}$.

We recall the representation of subsets of $\mathbb{R}^r$ by ω-languages from [5]. In the remainder of the text, let $\varrho > 1$ and $\Sigma := \{0, \ldots, \varrho - 1\}$ be fixed. ϱ is called the *base*. Let $r \geq 1$.

(a) V_r denotes the set of all ω-words over the alphabet $\Sigma^r \cup \{\star\}$ of the form $v \star \gamma$, where $v \in (\Sigma^r)^+$ and $\gamma \in (\Sigma^r)^\omega$.

(b) Let $v \star \gamma$ be an ω-word in V_r with $v(0) = (v_1, \ldots, v_r)$. The ω-word $v \star \gamma$ represents the vector of real numbers with r components

$$\langle v \star \gamma \rangle := -\varrho^{|v|-1} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_r \end{pmatrix} + \textstyle\sum_{0<i<|v|} \varrho^{|v|-i-1} \cdot v(i) + \sum_{i \geq 0} \varrho^{-i-1} \cdot \gamma(i)\,,$$

where $b_i := \left\lceil \frac{v_i}{\varrho} \right\rceil$, for $1 \leq i \leq r$. Observe that $b_i = 0$ if $v_i = 0$, and $b_i = 1$, otherwise. Here, scalar multiplication is as usual and vector addition is componentwise. Note that we do not distinguish between vectors and tuples.

(c) For a formula $\varphi(x_1, \ldots, x_r)$, we define $L(\varphi) := \{\alpha \in \mathsf{V}_r : \mathfrak{R} \models \varphi[\langle \alpha \rangle]\}$.

Note that the encoding $v \star \gamma \in \mathsf{V}_1$ of a real number is based on the ϱ's complement representation. The symbol $\star$ plays the role of a decimal point, separating the integer part v from the fractional part γ. Furthermore, the first letter determines whether a "track" represents a number that is greater than or equal to 0, or a number that is less than or equal to 0. Note that the ω-words $0 \star 0^\omega$ and $(\varrho - 1) \star (\varrho - 1)^\omega$ both represent the number 0, where b^ω denotes the infinite repetition of the letter $b \in \Sigma$.

We overload the notation $\langle \cdot \rangle$ by using it also for finite nonempty prefixes in V_r. For $v \in (\Sigma^r)^+$ and $v' \in (\Sigma^r)^*$, we write $\langle v \rangle$ and $\langle v \star v' \rangle$ for $\langle v \star \bar{0}^\omega \rangle$ and $\langle v \star v' \bar{0}^\omega \rangle$, respectively, where $\bar{0}$ denotes the vector $(0, \ldots, 0) \in \Sigma^r$.

3. Automata Upper Bound for the Ordering over the Integers

Before looking at the ω-languages that can be described by the first-order logic over $\mathfrak{R}$, we look at a simpler case. Namely, we investigate the languages that can be described by formulas over $\mathfrak{Z}$. We establish an exponential upper bound on the automata size for these languages. The purpose of investigating this simpler case first is twofold. First, it introduces the main concepts, which we also use in §4 for the ω-languages definable in the first-order logic over $\mathfrak{R}$. Second, it demonstrates the generality of the approach. The results in this section illustrate the relationship between the equivalence classes of a refinement of the equivalence relation $\equiv_m^r$ and the equivalence classes of the Nerode relation of a language described by a formula $\varphi(x_1, \ldots, x_r)$ over $\mathfrak{Z}$ with $\mathrm{qd}(\varphi) \leq m$.

Throughout this section, formulas are over $\mathfrak{Z}$'s signature, and m and r range over the natural numbers. We start with some definitions. For a formula $\varphi(x_1, \ldots, x_r)$, we define the language

$$K(\varphi) := \{v \in (\Sigma^r)^+ \,:\, \mathfrak{Z} \models \varphi[\langle v \rangle]\}\,.$$

We partition $\mathbb{Z}^r$ by the equivalence relation E_m^r that is defined as

$$\bar{a}\, E_m^r\, \bar{b} \qquad \text{iff} \qquad \begin{array}{l} \mathrm{sign}(a_i - a_j - c) = \mathrm{sign}(b_i - b_j - c), \\ \text{for all } c, i, j \in \mathbb{N} \text{ with } c \leq m \text{ and } 1 \leq i, j \leq r\,, \end{array}$$

where $\bar{a}, \bar{b} \in \mathbb{Z}^r$, and $\mathrm{sign}(x) := 0$ if $x < 0$ and $\mathrm{sign}(x) := 1$, otherwise, for $x \in \mathbb{R}$. Intuitively speaking, $\bar{a}, \bar{b} \in \mathbb{Z}^r$ are in the same equivalence class of E_m^r if the distances between their components are equal up to the threshold m.

Before we launch into the proof of establishing an upper bound on the size of the minimal deterministic automaton for a formula $\varphi(x_1, \ldots, x_r)$, we give an outline: (i) We show that $E_{2^{\mathrm{qd}(\varphi)}}^r$ refines $\equiv_{\mathrm{qd}(\varphi)}^r$. (ii) We establish an upper bound on the index of $E_{2^{\mathrm{qd}(\varphi)}}^r$. (iii) We show that $E_{2^{\mathrm{qd}(\varphi)}}^r$ has a congruence property with respect to word concatenation. (iv) By using (i) and (iii), we show that $E_{2^{\mathrm{qd}(\varphi)}}^r$ determines an equivalence relation on $(\Sigma^r)^+$ that refines the Nerode relation $\sim_{K(\varphi)}$. Finally, from (ii) we derive an upper bound on the index of $\sim_{K(\varphi)}$. Note that the equivalence classes of $\sim_{K(\varphi)}$ can be viewed as the states of the minimal deterministic finite automaton that accepts $K(\varphi)$. The properties (i) to (iv) correspond to the Lemmas 3.1 to 3.4, respectively, which are given below.

Lemma 3.1. *The equivalence relation $E_{2^m}^r$ refines the equivalence relation $\equiv_m^r$. That means, $\bar{a}\, E_{2^m}^r\, \bar{b}$ implies $\bar{a} \equiv_m^r \bar{b}$, for all $\bar{a}, \bar{b} \in \mathbb{Z}^r$.*

To prove Lemma 3.1, we apply a standard technique from model theory. First, we show that the family $(E_n^s)_{s,n\in\mathbb{N}}$ of equivalence relations has the following property:

$$\text{If } \bar{a}\ E_{2^{m+1}}^r\ \bar{b} \text{ then for every } a' \in \mathbb{Z}, \text{ there is some } b' \in \mathbb{Z} \text{ such that } (\bar{a}, a')\ E_{2^m}^{r+1}\ (\bar{b}, b'). \quad (3.1)$$

Properties of this kind are often called back-and-forth properties in the literature. Note that $E_{2^{m+1}}^r$ is symmetric. Second, we complete the proof by an induction over m, where we use the property (3.1) in the induction step for the existential quantifier.

Lemma 3.2. *The index of E_m^r is at most $r! \cdot (m+1)^r$.*

Proof. There are at most $r!$ many possibilities to order the r elements increasingly. If in such an ordering the distance between the ith element x and the $(i+1)$st element y is greater than or equal to m, we have that $\text{sign}(y - x - c) = 1$, for all $c \in \mathbb{N}$ with $c \leq m$. We obtain that the index is at most $r! \cdot (m+1)^r$. ∎

Lemma 3.3. *Let $u, v \in (\Sigma^r)^+$. If $\langle u\rangle\ E_m^r\ \langle v\rangle$ then $\langle uw\rangle\ E_m^r\ \langle vw\rangle$, for all $w \in (\Sigma^r)^*$.*

Proof. Let $n := |w|$, $\bar{a} := (a_1, \ldots, a_r) := \langle u\rangle$, $\bar{b} := (b_1, \ldots, b_r) := \langle v\rangle$, and $\bar{d} := (d_1 \ldots, d_r) := \langle \bar{0}w\rangle$. We have that $\langle uw\rangle = \varrho^n\bar{a} + \bar{d}$ and $\langle vw\rangle = \varrho^n\bar{b} + \bar{d}$. Furthermore, it holds that $d_i < \varrho^n$, for all $i \in \{1, \ldots, r\}$. Let $i, j, c \in \mathbb{N}$ with $1 \leq i, j \leq r$ and $c \leq m$. We have to show that

$$\text{sign}\left(\varrho^n(a_i - a_j) + d_i - d_j - c\right) = \text{sign}\left(\varrho^n(b_i - b_j) + d_i - d_j - c\right). \quad (3.2)$$

Case $a_i - a_j = 0$. We have that $\text{sign}(a_i - a_j) = 1 = \text{sign}(a_j - a_i)$. From the assumption $\bar{a}\ E_m^r\ \bar{b}$, it follows that $\text{sign}(a_i - a_j) = \text{sign}(b_i - b_j)$ and $\text{sign}(a_j - a_i) = \text{sign}(b_j - b_i)$, and hence, $b_i - b_j = 0$. Obviously, the equality (3.2) holds.

Case $b_i - b_j = 0$. This case is symmetric to the case $a_i - a_j = 0$ above.

Case $a_i - a_j \neq 0$ and $b_i - b_j \neq 0$. For showing (3.2), it suffices to show the equality

$$\text{sign}\left(a_i - a_j + \tfrac{d_i - d_j - c}{\varrho^n}\right) \quad = \quad \text{sign}\left(b_i - b_j + \tfrac{d_i - d_j - c}{\varrho^n}\right). \quad (3.3)$$

- If $m = 0$, we have that $c = 0$ and thus $\left|\frac{d_i - d_j - c}{\varrho^n}\right| \leq \frac{|d_i - d_j|}{\varrho^n} \leq \frac{\varrho^n - 1}{\varrho^n} < 1$. Since $a_i - a_j \neq 0$ and $b_i - b_j \neq 0$ and by the assumption $\bar{a}\ E_0^r\ \bar{b}$, we conclude that the equality (3.3) holds.
- If $m > 0$, we have that $\left|\frac{d_i - d_j - c}{\varrho^n}\right| \leq \frac{|d_i - d_j| + |c|}{\varrho^n} \leq \frac{\varrho^n - 1 + |c|}{\varrho^n} \leq \frac{m(\varrho^n - 1) + m}{\varrho^n} = m$. The equality (3.3) follows from the assumption $\bar{a}\ E_m^r\ \bar{b}$. ∎

Lemma 3.4. *Let φ be a formula with at most r free variables and with quantifier depth at most m. If $\langle u\rangle\ E_{2^m}^r\ \langle v\rangle$ then $u \sim_{K(\varphi)} v$, for all $u, v \in (\Sigma^r)^+$.*

Proof. We prove the lemma by contraposition. Assume that $u \not\sim_{K(\varphi)} v$, i.e., there is a word $w \in \Sigma^*$ such that $uw \in K(\varphi) \not\Leftrightarrow vw \in K(\varphi)$. It follows that $\langle uw\rangle \not\equiv_m^r \langle vw\rangle$. By Lemma 3.1, we conclude that $\langle uw\rangle\ E_{2^m}^r\ \langle vw\rangle$ does not hold. By Lemma 3.3, we obtain that $\langle u\rangle\ E_{2^m}^r\ \langle v\rangle$ does not hold. ∎

Theorem 3.5. *Let φ be a formula. The index of $\sim_{K(\varphi)}$ is at most $1 + 2^{n^2}$, where n is the length of the formula φ, i.e., φ consists of n symbols.*

Proof. Let r be the number of free variables of φ and $m := \text{qd}(\varphi)$. Note that $n \geq r + m + 1$. Without loss of generality, we assume that $r > 0$. By Lemma 3.2, we have that the index of $E_{2^m}^r$ is at most $r! \cdot (2^m + 1)^r \leq 2^{r^2 + rm + r} \leq 2^{rn} \leq 2^{n^2}$. From Lemma 3.4, it follows that $\sim_{K(\varphi)}$ partitions $(\Sigma^r)^+$ in at most 2^{n^2} equivalence classes. Note that the empty word can be in an equivalence class that is distinct from all the others. ∎

4. Automata Upper Bound for Real Addition

In this section, we establish an upper bound on the automata size for the first-order logic over $\mathfrak{R}$. The proof has a similar structure as the proof in the previous section §3. However, it is more involved. In §4.1, we define a family $(F_n^s)_{s,n\in\mathbb{N}}$ of equivalence relations. In §4.1 and §4.2, we show that $(F_n^s)_{s,n\in\mathbb{N}}$ has similar properties as the family $(E_n^s)_{s,n\in\mathbb{N}}$ defined in §3. Namely, (1) we show that each $F^r_{2^{2m+2}}$ refines $\equiv^r_m$ and (2) we establish a relationship between the equivalence classes of the congruence relations determined by the definable ω-languages and equivalences classes of refinements of the equivalence relations $(F_n^s)_{s,n\in\mathbb{N}}$. Finally, in §4.3, we derive the double exponential upper bound on the size of a minimal Büchi automaton that accepts the ω-language of a formula of the first-order logic over $\mathfrak{R}$.

In the following, formulas are always over $\mathfrak{R}$'s signature, and r and m range over the natural numbers.

4.1. Partitioning the Reals by First-order Formulas

The results, which we use later, and their presentation of this subsection are based on Chapter 22 of Kozen's book [16]. Since subtle modifications are made, we provide proofs in the full version of the paper. At the end of this subsection, we comment on these modifications and their implications.

An *integer affine function* of arity r is a function $f : \mathbb{R}^r \to \mathbb{R}$ defined by a linear polynomial with integer coefficients, i.e., there are $c_0, \ldots, c_r \in \mathbb{Z}$ such that for all $x_1, \ldots, x_r \in \mathbb{R}$, it holds that $f(x_1, \ldots, x_r) = c_0 + \sum_{1\le i\le r} c_i x_i$. For such a function, f^* denotes the function with $f^*(x_1, \ldots, x_r) = \sum_{1\le i\le r} c_i x_i$, for all $x_1, \ldots, x_r \in \mathbb{R}^r$. We define $||f|| := \max\{0, |c_1|, \ldots, |c_r|\}$. Let A^r be the set of all integer affine functions of arity r and

$$B^r_m := \left\{ f \in A^r \,:\, ||f|| \le m \text{ and } |f(\bar{0})| \le rm \right\}.$$

Definition 4.1. We partition $\mathbb{R}^r$ by the equivalence relation F^r_m that is defined as

$$\bar{a}\; F^r_m\; \bar{b} \qquad \text{iff} \qquad \text{for all } f \in B^r_m,\ \operatorname{sign}(f(\bar{a})) = \operatorname{sign}(f(\bar{b}))\,,$$

where $\bar{a}, \bar{b} \in \mathbb{R}^r$.

Note that F^r_m decomposes $\mathbb{R}^r$ into cells. Each such cell is described by a conjunction of linear inequations, where the absolute values of the coefficients of the inequations are bounded. Moreover, we remark that the technique that we present in the following by connecting such partitions to first-order logic and Ehrenfeucht-Fraïssé games is reminiscent of techniques in computational complexity (see [11]) and constraint databases (see [21]). A novel insight is that these partitions are also connected to the relation $\sim_{L(\varphi)}$ for a formula φ. We start with some properties about the family $(F_n^s)_{s,n\in\mathbb{N}}$ of equivalence relations.

Lemma 4.2. *Let $\bar{a}, \bar{b} \in \mathbb{R}^r$. If $\bar{a}\; F^r_{4m^2}\; \bar{b}$ then for all $a' \in \mathbb{R}$, there is some $b' \in \mathbb{R}$ such that $(\bar{a}, a')\; F^{r+1}_m\; (\bar{b}, b')$.*

Similar to Lemma 3.1, we obtain the following lemma by using Lemma 4.2.

Lemma 4.3. *For all $\bar{a}, \bar{b} \in \mathbb{R}^r$, it holds that if $\bar{a}\; F^r_{2^{2m+2}}\; \bar{b}$ then $\bar{a} \equiv^r_m \bar{b}$.*

The following two lemmas show how to obtain a set $R \subseteq \mathbb{R}^r$ such that each equivalence class of F^r_m has at least one representative in R. Let σ be an equivalence class of F^r_m and let σ' be an equivalence class of F^{r+1}_n, where $n \in \mathbb{N}$. We say that σ' is *consistent* with σ if $(\sigma \times \mathbb{R}) \cap \sigma' \neq \emptyset$.

Lemma 4.4. *For each equivalence class σ of F^1_m, we have that*

$$\sigma \cap \left\{ \tfrac{d}{c} \,:\, c, d \in \mathbb{Z} \text{ with } c \neq 0 \text{ and } |c|, |d| \leq 2m^2 \right\} \neq \emptyset\,.$$

Lemma 4.5. *Let $r > 1$, $\bar{a} \in \mathbb{R}^r$, where σ is the equivalence class of $\bar{a}$ with respect to $F^r_{2m^2}$. For every equivalence class σ' of F^{r+1}_m that is consistent with σ, we have that*

$$\sigma' \cap \left\{ \left(\bar{a}, \tfrac{f(\bar{a})+d}{c}\right) \,:\, f \in B^r_2 \text{ and } c, d \in \mathbb{Z} \setminus \{0\} \text{ with } |c| \leq 2, \text{ and } |d| < 2 \right\} \neq \emptyset$$

if $m = 1$, and, for $m \neq 1$, we have that

$$\sigma' \cap \left\{ \left(\bar{a}, \tfrac{f(\bar{a})}{c}\right) \,:\, f \in B^r_{2m^2} \text{ and } c \in \mathbb{Z} \setminus \{0\} \text{ with } |c| \leq 2m^2 \right\} \neq \emptyset\,.$$

Remark 4.6. Before we proceed to establish the upper bound on the size of the minimal automata representation for the set defined by a formula φ, we point out the differences between the family $(F^s_n)_{s,n\in\mathbb{N}}$ of equivalence relations and the family of equivalence relations defined in Kozen's book [16] in Chapter 22.

In Kozen's book, two elements $\bar{a}, \bar{b} \in \mathbb{R}^r$ are related iff $\mathrm{sgn}(f(\bar{a})) = \mathrm{sgn}(f(\bar{b}))$, for all integer affine function $f \in A^r$ with $||f|| \leq m$ and $|f(\bar{0})| \leq m$. Here, sgn denotes the signum function that is defined as $\mathrm{sgn}(x) := -1$ if $x < 0$, $\mathrm{sgn}(x) := 1$ if $x > 0$, and $\mathrm{sgn}(0) := 0$.

There are two differences to our definition. First, we use the function sign instead of the function sgn. This difference is actually irrelevant. Using sign instead of sgn in the definition in Kozen's book would not change the equivalence relations. However, we found the reasoning in the proofs when using the function sign slightly simpler. Second and more relevant, we require $|f(\bar{0})| \leq rm$ instead of $|f(\bar{0})| \leq m$. The proofs of the Lemmas 4.2 to 4.5 follow the lines of the proofs of the corresponding lemmas in Kozen's book. However, there are subtle differences, e.g., in Lemma 4.5, we have the special case for $m = 1$, which is not needed in the corresponding lemma in Kozen's book.

An immediate consequence of only requiring this weaker restriction on the functions $f \in A^r$ is that the equivalence relation F^r_m refines the corresponding equivalence relation as defined in Kozen's book. The purpose for having finer equivalence relations is the following: For a formula $\varphi(x_1, \ldots, x_r)$, we show in §4.2 that the equivalence classes of $\sim_{L(\varphi)}$ are related to the equivalence classes of a certain relation in the family $(F^s_n)_{s,n\in\mathbb{N}}$. Without the weaker requirement we were not able to establish a similar relationship. The problem can be pinpointed to Lemma 4.8, which is crucial in relating the equivalence relations. The corresponding statement of Lemma 4.8 would not be correct when using the equivalence relations as defined in Kozen's book.

4.2. Relationship to Languages

In this subsection, we establish a relationship between the equivalence relation $F^r_{2^{2^{m+2}}+1}$ and the congruence relation $\sim_{L(\varphi)}$, where $\varphi(x_1, \ldots, x_r)$ is a formula with $\mathrm{qd}(\varphi) \leq m$. Namely, we show that $F^r_{2^{2^{m+2}}+1}$ determines a refinement of the congruence relation $\sim_{L(\varphi)}$.

We start with a technical lemma. Its proof is straightforward and we therefore omit it. In the following, we will use it without explicitly referring to it.

Lemma 4.7. *For $f \in A^r$, $u \in (\Sigma^r)^+$, $u' \in (\Sigma^r)^*$, and $\gamma \in (\Sigma^r)^\omega$, the following facts hold:*
(1) $f(\langle uu' \rangle) = f(\bar{0}) + \varrho^{|u'|} f^*(\langle u \rangle) + f^*(\langle \bar{0} u' \rangle)$, and
(2) $f(\langle u \star u' \gamma \rangle) = f(\bar{0}) + f^*(\langle u \star u' \rangle) + \varrho^{-|u'|} f^*(\langle \bar{0} \star \gamma \rangle)$.

The next two lemmas show that the equivalence relations in the family $(F_n^s)_{s,n\in\mathbb{N}}$ have congruence properties on words with respect to word concatenation and show how their equivalence classes relate to the equivalence classes of the congruence relation $\sim_{L(\varphi)}$. We want to point out a technical detail, which is reflected in the (b)-parts of the lemmas, is illustrated by the following example. The words $u\star u'$ and $u\star u'\bar{0}$ represent the same vector of real numbers, i.e., $\langle u\star u'\rangle = \langle u\star u'\bar{0}\rangle$. Therefore, $u\star u'$ and $u\star u'\bar{0}$ represent the same equivalence class in F_m^r. However, $u\star u'$ and $u\star u'\bar{0}$ might not be in the same equivalence class with respect to $\sim_{L(\varphi)}$. Observe that appending an ω-word $\gamma\in(\Sigma^r)^\omega$ to $u\star u'$ and $u\star u'\bar{0}$ may yield representations of different vectors of real numbers, i.e., $\langle u\star u'\gamma\rangle \neq \langle u\star u'\bar{0}\gamma\rangle$, and $u\star u'\gamma$ and $u\star u'\bar{0}\gamma$ may represent different equivalence classes in F_m^r.

Lemma 4.8. *For all $u,v\in(\Sigma^r)^+$ and $u',v'\in(\Sigma^r)^*$, the following two facts hold:*
(a) If $\langle u\rangle\, F_m^r\, \langle v\rangle$ then for all $w\in(\Sigma^r)^*$, $\langle uw\rangle\, F_m^r\, \langle vw\rangle$.
(b) If $\langle u\star u'\rangle\, F_{2m}^r\, \langle v\star v'\rangle$ and $|u'|\geq|v'|$ then for all $\gamma\in(\Sigma^r)^\omega$, $\langle u\star u'\gamma\rangle\, F_m^r\, \langle v\star v'\bar{0}^k\gamma\rangle$ with $k=\min\{|u'|-|v'|\}\cup\{k\in\mathbb{Z}\,:\,\varrho^k\geq rm\}$.

Proof. For $r=0$, there is nothing to prove. In the following, we assume that $r>0$.
(a) We prove (a) by contraposition. Assume that for some $w\in(\Sigma^r)^*$, it is not the case that $\langle uw\rangle\, F_m^r\, \langle vw\rangle$, i.e., there is some $f\in B_m^r$ with $\mathrm{sign}(f(\langle uw\rangle))\neq\mathrm{sign}(f(\langle vw\rangle))$. Without loss of generality, we assume that $f(\langle uw\rangle)<0$ and hence $f(\langle vw\rangle)\geq 0$. The other cases can be reduced to this case by using the function $g\in B_m^r$ with $g(\bar{x})=-f(\bar{x})$, for all $\bar{x}\in\mathbb{R}^r$.

We have that $\varrho^{|w|}f^*(\langle u\rangle)+f(\langle\bar{0}w\rangle)<0$ and $\varrho^{|w|}f^*(\langle v\rangle)+f(\langle\bar{0}w\rangle)\geq 0$. Obviously, it must hold that $f^*(\langle u\rangle)\neq f^*(\langle v\rangle)$. If $\mathrm{sign}(f^*(\langle u\rangle))\neq\mathrm{sign}(f^*(\langle v\rangle))$ then $\langle u\rangle\, F_m^r\, \langle v\rangle$ does not hold and we are done. So, assume that $\mathrm{sign}(f^*(\langle u\rangle))=\mathrm{sign}(f^*(\langle v\rangle))$. If $|f^*(\langle u\rangle)|\leq rm$ or $|f^*(\langle v\rangle)|\leq rm$ then we are also done by choosing an appropriate function $g\in B_m^r$ with $\mathrm{sign}(g(\langle u\rangle))\neq\mathrm{sign}(g(\langle v\rangle))$. So, assume that $|f^*(\langle u\rangle)|,|f^*(\langle v\rangle)|>rm$. Note that $|f^*(\langle\bar{0}w\rangle)|\leq(\varrho^{|w|}-1)rm$.

– If $f^*(\langle v\rangle)<-rm$, we obtain a contradiction to the assumption $f(\langle vw\rangle)\geq 0$, since

$$\begin{aligned}\varrho^{|w|}f^*(\langle v\rangle)+f(\langle\bar{0}w\rangle) &= \varrho^{|w|}f^*(\langle v\rangle)+f^*(\langle\bar{0}w\rangle)+f(\bar{0})\\ &< -\varrho^{|w|}rm+(\varrho^{|w|}-1)rm+rm\leq 0\,.\end{aligned}$$

– If $f^*(\langle v\rangle)>rm$, we conclude that $f^*(\langle u\rangle)>rm$. Analogously, as in the above case, we obtain a contradiction to the assumption $f(\langle uw\rangle)<0$.

(b) Let f be an arbitrary function in B_m^r and $\gamma\in(\Sigma^r)^\omega$. We have to show that $\mathrm{sign}(f(\langle u\star u'\gamma\rangle))=\mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle))$. Since $B_m^r\subseteq B_{2m}^r$, it follows from the assumption $\langle u\star u'\rangle F_{2m}^r\langle v\star v'\rangle$ that $\mathrm{sign}(f(\langle u\star u'\rangle))=\mathrm{sign}(f(\langle v\star v'\rangle))$. That means, either (1) $f(\langle u\star u'\rangle),f(\langle v\star v'\rangle)<0$ or (2) $f(\langle u\star u'\rangle),f(\langle v\star v'\rangle)\geq 0$ holds. Since the case (1) can be reduced to the case (2) by considering the function $g(\bar{x})=-f(\bar{x})$, for all $\bar{x}\in\mathbb{R}^r$, we restrict ourselves to (2).

For the sake of readability, we use the abbreviations $a:=f^*(\langle u\star u'\rangle)$, $b:=f^*(\langle v\star v'\rangle)$, and $c:=f^*(\langle\bar{0}\star\gamma\rangle)$. Note that

$$f(\langle u\star u'\gamma\rangle)=f(\bar{0})+a+c\varrho^{-|u'|}\qquad\text{and}\qquad f(\langle v\star v'\bar{0}^k\gamma\rangle)=f(\bar{0})+b+c\varrho^{-|v'|-k}\,. \tag{4.1}$$

If $c\geq 0$ then $\mathrm{sign}(f(\langle u\star u'\gamma\rangle))=\mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle))=1$. In the following, assume $c<0$.

Case $a\neq b$. With the assumption $\langle u\star u'\rangle F_{2m}^r\langle v\star v'\rangle$ we conclude that $a,b>2rm$. Note that $|f^*(\langle\bar{0}\star\alpha\rangle)|\leq rm$, for all $\alpha\in(\Sigma^r)^\omega$. It follows that

$$f^*(\langle u\star u'\gamma\rangle)=a+c\varrho^{-|u'|}>2rm-rm\geq rm\,.$$

The reasoning for $f^*(\langle v\star v'\bar{0}^k\gamma\rangle) > rm$ is similar. Since $|f(\bar{0})| \leq rm$, we have that $\mathrm{sign}(f(\langle u\star u'\gamma\rangle)) = \mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle)) = 1$.

Case $a = b$. For $k = |u'| - |v'|$, it immediately follows from the equalities in (4.1) that $f(\langle u\star u'\gamma\rangle) = f(\langle v\star v'\bar{0}^k\gamma\rangle)$, and hence $\mathrm{sign}(f(\langle u\star u'\gamma\rangle)) = \mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle))$. For $a = b = -f(\bar{0})$, it is also straightforward to see from the two equalities in (4.1) that $\mathrm{sign}(f(\langle u\star u'\gamma\rangle)) = \mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle))$. For the rest of the proof, assume $k = \min\{k \in \mathbb{Z} : \varrho^k \geq rm\}$ and $b \neq -f(\bar{0})$. Moreover, for $|c|\cdot\varrho^{-|u'|} > f(\bar{0}) + a$, it follows directly from the equalities (4.1) that $\mathrm{sign}(f(\langle u\star u'\gamma\rangle)) = \mathrm{sign}(f(\langle v\star v'\bar{0}^k\gamma\rangle)) = 0$. So, we also assume that $|c|\cdot\varrho^{-|u'|} \leq f(\bar{0}) + a$. Observe that $f(\langle u\star u'\gamma\rangle) \geq 0$. Furthermore, observe that $f(\bar{0}) + b \geq \varrho^{-|v'|}$. We have that $f(\langle v\star v'\bar{0}^k\gamma\rangle) \geq \frac{1}{\varrho^{|v'|}} + \frac{c}{\varrho^{|v'|+k}} = \frac{\varrho^k - |c|}{\varrho^{|v'|+k}} \geq \frac{rm - rm}{\varrho^{|v'|+k}} \geq 0$. ∎

Lemma 4.9. *Let $\varphi(x_1,\ldots,x_r)$ be a formula with $\mathrm{qd}(\varphi) \leq m$. For all $u, v \in (\Sigma^r)^+$ and $u', v' \in (\Sigma^r)^*$, the following two facts hold:*
(a) If $\langle u\rangle\ F^r_{2^{2^{m+2}+1}}\ \langle v\rangle$ then $u \sim_{L(\varphi)} v$.
(b) If $\langle u\star u'\rangle\ F^r_{2^{2^{m+2}+1}}\ \langle v\star v'\rangle$ and $|u'| \geq |v'|$ then $u\star u' \sim_{L(\varphi)} v\star v'\bar{0}^k$ with $k = \min\{|u'| - |v'|\} \cup \{k \in \mathbb{Z} : \varrho^k \geq rm\}$.

Proof. We only show (a). The proof for (b) is analogous and we omit it. From Lemma 4.8(a), it follows that $\langle uw\rangle\ F^r_{2^{2^{m+2}+1}}\ \langle vw\rangle$, for all $w \in (\Sigma^r)^*$. With Lemma 4.8(b), we obtain that $\langle uw\star\gamma\rangle\ F^r_{2^{2^{m+2}}}\ \langle vw\star\gamma\rangle$, for all $w \in (\Sigma^r)^*$ and $\gamma \in (\Sigma^r)^\omega$. By Lemma 4.3, we conclude that $\langle uw\star\gamma\rangle \equiv^r_m \langle vw\star\gamma\rangle$, for all $w \in (\Sigma^r)^*$ and $\gamma \in (\Sigma^r)^\omega$. In particular, we have that $uw\star\gamma \in L(\varphi) \Leftrightarrow vw\star\gamma \in L(\varphi)$, for all $w \in (\Sigma^r)^*$ and $\gamma \in (\Sigma^r)^\omega$. From this it follows that $u \sim_{L(\varphi)} v$, since for any ω-word α not in V_r, we have that $u\alpha, v\alpha \notin L(\varphi)$. ∎

4.3. Upper Bounds

We establish an upper bound on the index of F^r_m, from which we then derive an upper bound on the automata size. We start with a simple lemma.

Lemma 4.10. *The cardinality of B^r_m is at most $(2rm+1)(2m+1)^r$.*

Using the Lemmas 4.4, 4.5, and 4.10, we establish an upper bound on the index of F^r_m.

Lemma 4.11. *The index of F^r_m is at most $\max\{1, m^{2^{3+r}}\cdot 2^{2^{3+r}}\}$.*

Theorem 4.12. *Let φ be a formula. The index of $\sim_{L(\varphi)}$ is at most $2^{2^{8+n}}$, where n is the length of the formula φ, i.e., the number of symbols of φ.*

Proof. Let r be the number of free variables in φ and $m := \mathrm{qd}(\varphi)$. We use $F^r_{2^{2^{m+2}+1}}$ to define a refinement R of $\sim_{L(\varphi)}$. First, the singleton $\{\varepsilon\}$ is an equivalence class of R. Second, the set of words with at least two occurrences of the letter $\star$ is another equivalence class of R. The equivalence class of a word $v \in (\Sigma^r)^+$ of R is $\{u \in (\Sigma^r)^+ : \langle v\rangle\ F^r_{2^{2^{m+2}+1}}\ \langle u\rangle\}$.

It remains to define the equivalence classes of R on $F := \{v\star v' : v \in (\Sigma^r)^+ \text{ and } v' \in (\Sigma^r)^*\}$. For $v\star v' \in F$, let $S := \{u\star u' \in F : \langle v\star v'\rangle\ F^r_{2^{2^{m+2}+1}}\ \langle u\star u'\rangle\}$. R chops S into equivalence classes, assuming $|v'| \leq |u'|$, for all $u\star u' \in S$:

- For $k \in \{0,\ldots,\lceil\log_\varrho r2^{2^{m+2}+1}\rceil - 1\}$, the equivalence class of $v\star v'\bar{0}^k$ of R is $\{u\star u' \in S : |u'| = |v'| + k\}$.
- For $k = \lceil\log_\varrho r2^{2^{m+2}+1}\rceil$, the equivalence class of $v\star v'\bar{0}^k$ of R is $\{u\star u' \in S : |u'| \geq |v'| + k\}$.

Note that any word $u \star u' \in S$ relates to exactly one word $v \star v'\bar{0}^k$.

With Lemma 4.9 at hand, it is easy to see that R refines $\sim_{L(\varphi)}$. It remains to prove an upper bound on the index of R. Note that $n \geq m + r \geq 1$. By Lemma 4.11, an upper bound on the index of $F^r_{2^{2^{m+2}+1}}$ is

$$\left(2^{2^{m+2}+1}\right)^{2^{3+r}} \cdot 2^{2^{3+r}} = 2^{2^{m+2} \cdot 2^{3+r} + 2^{3+r} + 2^{3+r}} = 2^{2^{5+r+m} + 2^{4+r}} \leq 2^{2^{6+n}} .$$

Hence, R partitions $(\Sigma^r)^+$ into at most $2^{2^{6+n}}$ equivalence classes and F is partitioned into at most $2^{2^{6+n}} \cdot \lceil \log_\varrho r 2^{2^{m+2}+1} \rceil \geq 2^{2^{6+n}} \cdot r(2^{m+2} + 1) \geq 2^{2^{6+n}} \cdot 2^{3+n} \geq 2^{2^{7+n}}$ equivalence classes. From this, we derive the upper bound $2^{2^{8+n}}$ on R's index. ∎

Remark 4.13. Since for any formula φ, $L(\varphi)$ is an ω-language in the Borel class $F_\sigma \cap G_\delta$ [5], we can—similar to deterministic finite automata—view the equivalence classes of $\sim_{L(\varphi)}$ as the states of a minimal deterministic Büchi automaton that accepts $L(\varphi)$. For further details, see [19] and [18]. Thus, Theorem 4.12 establishes a double exponential upper bound with respect to the formula length on the size of the minimal number of states of any Büchi automaton that accepts $L(\varphi)$.

Remark 4.14. The double exponential upper bound on the automata size is tight, i.e., there is a family of formulas $(\varphi_n)_{n \in \mathbb{N}}$ such that for each $n \in \mathbb{N}$, the length of φ_n is linear in n and the index of $\sim_{L(\varphi_n)}$ is double exponential in n. An analogous result with a similar proof has already been shown in [14] for Presburger arithmetic.

5. Conclusion

This papers presented a new method to reason about the sizes of automata that represent first-order definable sets of automatic structures. The method consists of identifying a relationship between the states of a minimal deterministic automaton for a formula and the equivalence classes of a refinement of the equivalence relation determined by Ehrenfeucht-Fraïssé games. We applied the presented method to establish tight upper bounds on the minimal sizes of automata that represent sets definable in $\mathrm{FO}(\mathbb{Z}, <)$ and $\mathrm{FO}(\mathbb{R}, +, <)$. For $\mathrm{FO}(\mathbb{R}, +, <)$, previously proposed techniques based on quantifier-elimination methods [14] failed to establish a double exponential upper bound on the automata size.

As future work, we want to investigate how, and to what extent, the upper bounds on the automata sizes depend on how elements of a structure are encoded as words. The word encoding of integers and reals that we have used in this paper is based on the ϱ's complement representation, for some $\varrho \in \mathbb{N}$ with $\varrho \geq 2$. There are various other word encodings of numbers so that, e.g., $\mathrm{FO}(\mathbb{Z}, <)$ admits an automata-based decision procedure. For a study on the impact of encodings in automatic structures, see, e.g., [13]. We also plan to apply the presented technique to establish further upper bounds on automata sizes for other automatic structures and use it to simplify the proofs of previously established upper bounds. For instance, for Presburger arithmetic, we expect that we can use equivalence relations similar to the ones used in this paper for $\mathrm{FO}(\mathbb{R}, +, <)$. However, we have to adjust the bounds on the coefficients and take the definable divisibility relations into account.

Acknowledgments. The author thanks David Basin, Cas Cremers, Matthias Schmalz, and the anonymous reviewers for their comments on earlier versions of the paper.

References

[1] S. Bardin, A. Finkel, J. Leroux, and L. Petrucci, *FAST: Fast acceleration of symbolic transition systems*, in Proc. of the 15th Int. Conf. on Computer Aided Verification (CAV), vol. 2725 of Lect. Notes Comput. Sci., Springer, 2003, pp. 118–121.

[2] S. Bardin, J. Leroux, and G. Point, *FAST extended release*, in Proc. of the 18th Int. Conf. on Computer Aided Verification (CAV), vol. 4144 of Lect. Notes Comput. Sci., Springer, 2006, pp. 63–66.

[3] B. Becker, C. Dax, J. Eisinger, and F. Klaedtke, *LIRA: Handling constraints of linear arithmetics over the integers and the reals*, in Proc. of the 19th Int. Conf. on Computer Aided Verification (CAV), vol. 4590 of Lect. Notes Comput. Sci., Springer, 2007, pp. 307–310.

[4] A. Blumensath and E. Grädel, *Finite presentations of infinite structures: Automata and interpretations*, Theory Comput. Syst., 37 (2004), pp. 641–674.

[5] B. Boigelot, S. Jodogne, and P. Wolper, *An effective decision procedure for linear arithmetic over the integers and reals*, ACM Trans. Comput. Log., 6 (2005), pp. 614–633.

[6] B. Boigelot and P. Wolper, *Representing arithmetic constraints with finite automata: an overview*, in Proc. of the 18th Int. Conf. on Logic Programming (ICLP), vol. 2401 of Lect. Notes Comput. Sci., Springer, 2002, pp. 1–19.

[7] J. Büchi, *Weak second-order arithmetic and finite automata*, Z. Math. Logik Grundlagen Math., 6 (1960), pp. 66–92.

[8] D. C. Cooper, *Theorem proving in arithmetic without multiplication*, in Proc. 7th Annual Machine Intelligence Workshop, B. Meltzer and D. Michie, eds., Edinburgh Univ. Press, 1972, pp. 91–100.

[9] H.-D. Ebbinghaus, J. Flum, and W. Thomas, *Mathematical Logic*, Springer, 2nd ed., 1994.

[10] J. Ferrante and C. Rackoff, *A decision procedure for the first order theory of real addition with order*, SIAM J. Comput., 4 (1975), pp. 69–76.

[11] J. Ferrante and C. W. Rackoff, *The Computational Complexity of Logical Theories*, vol. 718 of Lect. Notes Math., Springer, 1979.

[12] B. Khoussainov and A. Nerode, *Automatic presentations of structures*, in Proc. of the Int. Workshop on Logical and Computational Complexity (LCC), vol. 960 of Lect. Notes Comput. Sci., Springer, 1995, pp. 367–392.

[13] B. Khoussainov, S. Rubin, and F. Stephan, *Definability and regularity in automatic structures*, in Proc. of the 21st Annual Symp. on Theoretical Aspects of Computer Science (STACS), vol. 2996 of Lect. Notes Comput. Sci., Springer, 2004, pp. 440–451.

[14] F. Klaedtke, *On the automata size for Presburger arithmetic*, in Proc. of the 19th Annual IEEE Symp. on Logic in Computer Science (LICS), IEEE Computer Society Press, 2004, pp. 110–119. Accepted for publication in ACM Trans. Comput. Log..

[15] N. Klarlund, A. Møller, and M. I. Schwartzbach, *MONA implementation secrets*, Int. J. Found. Comput. Sci., 13 (2002), pp. 571–586.

[16] D. Kozen, *Theory of Computation*, Springer, 2006.

[17] R. E. Ladner, *Application of model theoretic games to discrete linear orders and finite automata*, Inform. and Control, 33 (1977), pp. 281–303.

[18] C. Löding, *Efficient minimization of deterministic weak ω-automata*, Inform. Process. Lett., 79 (2001), pp. 105–109.

[19] O. Maler and L. Staiger, *On syntactic congruences for omega-languages*, Theoret. Comput. Sci., 181 (1997), pp. 93–112.

[20] A. R. Meyer, *Weak monadic second-order theory of successor is not elementary-recursive*, in Logic Colloquium, vol. 453 of Lect. Notes Math., Springer, 1975, pp. 132–154.

[21] J. Paredaens, J. Van den Bussche, and D. Van Gucht, *First-order queries on finite structures over the reals*, SIAM J. Comput., 27 (1998), pp. 1747–1763.

[22] C. Reddy and D. W. Loveland, *Presburger arithmetic with bounded quantifier alternation*, in Proc. of the 10th Annual ACM Symp. on Theory of Computing (STOC), ACM Press, 1978, pp. 320–325.

[23] L. Stockmeyer, *The complexity of decision problems in automata theory and logic*, PhD thesis, Department of Electrical Engineering, MIT, Boston, MA, USA, 1974.

This work is licensed under the Creative Commons Attribution-NoDerivs License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/3.0/.