

Ehrenfeucht-Fraïssé Goes Automatic for Real Addition

Felix Klaedtke

ETH Zurich, Department of Computer Science, Switzerland

Abstract

The decision problem of various logical theories can be decided by automata-theoretic methods. Notable examples are Presburger arithmetic $\text{FO}(\mathbb{Z}, +, <)$ and the linear arithmetic over the reals $\text{FO}(\mathbb{R}, +, <)$. Despite the practical use of automata to solve the decision problem of such logical theories, many research questions are still only partly answered in this area. One of these questions is the complexity of the automata-based 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 article, we establish a double exponential upper bound on the automata size for $\text{FO}(\mathbb{R}, +, <)$ and an exponential upper bound for the first-order theory of the discrete order over the integers $\text{FO}(\mathbb{Z}, <)$. 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.

Keywords: automata theory, automata-based decision procedures for logical theories, automatic structures, upper bounds, minimal sizes of automata, linear arithmetic over the reals, first-order equivalence, complexity

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]. In a nutshell, the decision problem for a logical theory of a given relational structure is solved as follows in an automata-theoretic setting: The elements of the domain of the structure are encoded by words over some alphabet in such a way that equality and the relations of the structure correspond to regular languages. In order to decide whether a formula belongs to the corresponding logical theory, one constructs an automaton that precisely accepts the representatives of the assignments 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

Email address: `felixkl@inf.ethz.ch` (Felix Klaedtke)

decision problem of the logical theory 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, for instance, in [4, 15, 16, 17, 22]. Prominent and practically relevant examples of automatic logical theories are the weak monadic second-order theory of one successor WS1S, Presburger arithmetic $\text{FO}(\mathbb{Z}, +, <)$, and the linear arithmetic over the reals $\text{FO}(\mathbb{R}, +, <)$, see, for instance, [5, 6, 7]. Tools like MONA [20] and LIRA [3], which have been applied to various verification problems, implement such automata-based decision procedures for 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 worst-case scenario actually exists, since the decision problem for WS1S has a non-elementary worst-case complexity [25, 29]. 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. Only partial answers exist that explain this phenomenon.

In [18], 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 [27]. The proof on the upper bound in [18] 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 [9], one obtains 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 [13]. The author is not aware of any quantifier-elimination method for the linear arithmetic over the reals that would lead to an 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 [13], one might conjecture that the automata size is also doubly exponentially bounded.

The main result of this article proves this conjecture. The presented proof of the double exponential upper bound is based on Ehrenfeucht-Fraïssé games. It relates the states of a minimal automaton for a formula and the equivalence classes of a refinement of the equivalence relation determined by Ehrenfeucht-Fraïssé games played over the structure $(\mathbb{R}, +, <)$. 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 $\text{FO}(\mathbb{Z}, <)$. Note that the best known deterministic algorithms that decide $\text{FO}(\mathbb{Z}, <)$ run in exponential time [14]. In summary, the results presented in this article shed some light on the complexity of automata-based decision procedures for logical theories by identifying a relationship to Ehrenfeucht-Fraïssé games.

Since the appearance of the extended abstract of this article as the conference paper [19], the described technique based on Ehrenfeucht-Fraïssé games has also been used by Eisinger in [12] to establish upper bounds on the automata size for other logical theories. Namely, Eisinger gives a simpler proof of the triple exponential upper bound on the automata size for Presburger arithmetic as the one in [18] and he establishes a triple exponential upper bound on the automata size of the mixed linear arithmetic over the integers and the reals $\text{FO}(\mathbb{R}, \mathbb{Z}, +, <)$.

It is worth pointing out that the use of Ehrenfeucht-Fraïssé games in computational complexity theory [14] and constraint databases [26] is reminiscent of their use in this article and in [12] by partitioning the domain and connecting such a partition to the definable sets. Roughly speaking, the use Ehrenfeucht-Fraïssé games for establishing 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 Ehrenfeucht-Fraïssé 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 Ehrenfeucht-Fraïssé games yield tight upper bounds on the computational complexity for their decision problem.

In the context of automatic structures, such decision procedures for logical theories that are based on Ehrenfeucht-Fraïssé games to relativize quantified variables to a finite subset of the domain have, for example, been presented in [23] and more recently in [22]. Ladner [23] presents such decision procedures for monadic second-order theories of one successor and first-order fragments of them. Kuske and Lohrey [22] establish tight upper bounds on the decision problem for the first-order theories for a class of automatic structures, namely, the automatic structures of bounded degree. An example in this class is the structure over the integers with the successor relation. However, neither the

structure $(\mathbb{Z}, <)$ nor the structure $(\mathbb{R}, +, <)$, which we consider in this article, are of bounded degree. We point out that these decision procedures in [23] and [22] do not construct automata recursively over the formula structure as the standard automata-theoretic approach of deciding logical theories of automatic structures.

Ladner [23] also relates the equivalence classes determined by Ehrenfeucht-Fraïssé games to automata states for obtaining alternative proofs for the logical characterization of the regular languages [7, 8] and star-free languages [24]. Analogously to this article, Ladner establishes this relationship between the equivalence classes and the automata states by showing that the equivalence relations determined by the Ehrenfeucht-Fraïssé games are congruence relations with respect to the word encoding of the elements of the domain of the structures. However, Ladner does not focus on the automata sizes and he does not consider the structures $(\mathbb{Z}, <)$ and $(\mathbb{R}, +, <)$. Furthermore, the domains and the word encodings of the domain elements of the considered structures in Ladner's article and in this article are different. Namely, in [23], the domain is always an initial subset of \mathbb{N} and the domain elements are encoded unary. We use a binary encoding under which both relations $<$ and $+$ describe languages that can be recognized by automata.

Finally, we remark that the decision procedures for the first-order theories of automatic structures of bounded degree by Kuske and Lohrey [22] take the word encoding of the domain elements into account. However, Kuske and Lohrey do not establish a relationship between the equivalence classes and the automata states along the lines of Ladner [23] and this article. From such a relationship, we would obtain upper bounds on the automata sizes for this class of logical theories. However, it is not obvious under which conditions such a relationship exists for a word encoding of the domain elements of an automatic structure.

The remainder of this article is organized as follows. In Section 2, we give preliminaries. In Section 3, we illustrate our method by analyzing the languages that are $\text{FO}(\mathbb{Z}, <)$ -definable. In Section 4, we analyze the languages that are $\text{FO}(\mathbb{R}, +, <)$ -definable and establish the double exponential upper bound on the automata size. Finally, in Section 5, we draw conclusions. The appendix contains further proof details.

2. Preliminaries

We assume that the reader is familiar with first-order logic and automata theory over finite and infinite words. Here, we recapitulate the needed background in these areas and fix the notation and the 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 $(i + 1)$ st 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 article, 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, \dots, x_r occur free in the formula φ . The *length* of a formula φ , written as $\text{len}(\varphi)$, is recursively defined as $\text{len}(x \approx y) := 3$, $\text{len}(Rx_1 \dots x_n) := 1 + n$, where R is a relation symbol of arity n , $\text{len}(\neg\psi) := 1 + \text{len}(\psi)$, $\text{len}(\psi \vee \psi') := 1 + \text{len}(\psi) + \text{len}(\psi')$, and $\text{len}(\exists x\psi) := 1 + \text{len}(\psi)$. Note that a variable has always length 1, the size of its index is irrelevant. The *quantifier depth* of a formula φ is recursively defined as

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

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{J} 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 $a_1, \dots, a_r \in A$ and a formula $\varphi(x_1, \dots, x_r)$, we write $\mathfrak{A} \models \varphi[a_1 \dots a_r]$ 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, \dots, x_r and a_1, \dots, a_r , respectively.

Let $m, r \in \mathbb{N}$, $\bar{a} \in A^r$ and $\bar{b} \in A^r$. We write $\bar{a} \equiv_m^r \bar{b}$ if $\mathfrak{A} \models \varphi[\bar{a}] \Leftrightarrow \mathfrak{A} \models \varphi[\bar{b}]$, for all formulas $\varphi(x_1, \dots, x_r)$ with $\text{qd}(\varphi) \leq m$. Note that the relation \equiv_m^r partitions the elements of A^r . The equivalence classes of \equiv_m^r can be game-theoretically characterized by so-called Ehrenfeucht-Fraïssé games. For details on these games, see, for instance, [11]. Instead of working with \equiv_m^r of the structures \mathfrak{J} and \mathfrak{R} , respectively, we work in the following with well-chosen refinements of it. For these refinements, we establish upper bounds on their indexes. Furthermore, we relate the equivalence classes of these refinements to the states of the minimal deterministic finite state automaton that recognizes the satisfying assignments of a given formula.

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 $\varphi(x_1, \dots, x_r)$ 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 $\Sigma := \{0, \dots, \varrho - 1\}$ be fixed. ϱ is called the *base*. Let $r \geq 1$.

- (a) \mathbf{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 \mathbf{V}_r with $v(0) = (v_1, \dots, v_r)$. The ω -word $v \star \gamma$ represents the vector of real numbers with r components

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

where $b_i := \lceil \frac{v_i}{\varrho} \rceil$, 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 $\varphi(x_1, \dots, x_r)$, we define $L(\varphi) := \{\alpha \in \mathbf{V}_r : \mathfrak{R} \models \varphi[\langle \alpha \rangle]\}$.

Note that the encoding $v \star \gamma \in \mathbf{V}_1$ 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 \mathbf{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 $(0, \dots, 0) \in \Sigma^r$.

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{J} . 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 Section 4 for the ω -languages definable in the first-order logic over \mathfrak{A} . 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 $\varphi(x_1, \dots, x_r)$ over \mathfrak{Z} with $\text{qd}(\varphi) \leq m$.

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 $\varphi(x_1, \dots, x_r)$, 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

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

where $\bar{a}, \bar{b} \in \mathbb{Z}^r$, and $\text{sign}(x) := 0$ if $x < 0$ and $\text{sign}(x) := 1$, 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 $\varphi(x_1, \dots, x_r)$, we give an outline: (i) We show that $E_{2^{\text{qd}(\varphi)}}^r$ refines $\equiv_{\text{qd}(\varphi)}^r$. (ii) We establish an upper bound on the index of $E_{2^{\text{qd}(\varphi)}}^r$. (iii) We show that $E_{2^{\text{qd}(\varphi)}}^r$ has a congruence property with respect to word concatenation. (iv) By using (i) and (iii), we show that $E_{2^{\text{qd}(\varphi)}}^r$ 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 1 to 4, respectively, which are given below.

Lemma 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 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:

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

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 (1) in the induction step. Further proof details are given in the appendix.

Lemma 2. *The index of E_m^r is at most $r! \cdot (m+1)^{r-1}$.*

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

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

$$\text{sign}(\varrho^n(a_i - a_j) + d_i - d_j - c) = \text{sign}(\varrho^n(b_i - b_j) + d_i - d_j - c). \quad (2)$$

Case $a_i - a_j = 0$. We must $\text{sign}(a_i - a_j) = 1 = \text{sign}(a_j - a_i)$. 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 (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 (2), it suffices to show the equality

$$\text{sign}\left(a_i - a_j + \frac{d_i - d_j - c}{\varrho^n}\right) = \text{sign}\left(b_i - b_j + \frac{d_i - d_j - c}{\varrho^n}\right). \quad (3)$$

– If $m = 0$, we have $c = 0$ 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) holds.

– If $m > 0$, we have $\left|\frac{d_i - d_j - c}{\varrho^n}\right| \leq \frac{|d_i - d_j| + |c|}{\varrho^n} \leq \frac{\varrho^n - 1 + |c|}{\varrho^n} \leq \frac{m(\varrho^n - 1) + m}{\varrho^n} = m$. The equality (3) follows from the assumption $\bar{a} E_m^r \bar{b}$. \square

Lemma 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 $u \not\sim_{K(\varphi)} v$, that means, there is a word $w \in \Sigma^*$ such that $uw \in K(\varphi) \not\equiv vw \in K(\varphi)$. It follows that $\langle uw \rangle \not\equiv_m^r \langle vw \rangle$. By Lemma 1, we conclude that $\langle uw \rangle E_{2^m}^r \langle vw \rangle$ does not hold. By Lemma 3, we obtain that $\langle u \rangle E_{2^m}^r \langle v \rangle$ does not hold. \square

Theorem 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 .*

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 2, we have that the index of $E_{2^m}^r$ is at most $r! \cdot (2^m + 1)^{r-1} \leq 2^{r^2 + rm + r} \leq 2^{rn} \leq 2^{n^2}$. From Lemma 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. \square

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 structure of the proof is similar to the proof in the previous Section 3. However, it is more involved. In Section 4.1, we define a family $(F_n^s)_{s,n \in \mathbb{N}}$ of equivalence relations. In Section 4.1 and Section 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 Section 3. Namely, (1) we show that each $F_{2^{2m+2}}^r$ refines \equiv_m^r and (2) we establish a relationship between the equivalence classes of the congruence relations determined by the definable ω -languages and equivalence classes of refinements of the equivalence relations $(F_n^s)_{s,n \in \mathbb{N}}$. Finally, in Section 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 [21]. We point out that subtle modifications to the statements in [21] are necessary. In Remark 10, at the end of this subsection, we comment on these modifications and their implications.

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

$$B_m^r := \{f \in A^r : \|f\| \leq m \text{ and } |f(\bar{0})| \leq rm\}.$$

Definition 6. We partition \mathbb{R}^r by the equivalence relation F_m^r that is defined as

$$\bar{a} F_m^r \bar{b} \quad \text{iff} \quad \text{for all } f \in B_m^r, \text{ sign}(f(\bar{a})) = \text{sign}(f(\bar{b})),$$

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

Note that equivalence relation F_m^r 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. The conjunction of such a cell is defined as follows. For $\bar{a} \in \mathbb{R}^r$, we partition B_m^r into the sets S and T , where $S := \{f \in B_m^r : \text{sign}(f(\bar{a})) = 0\}$ and $T := \{f \in B_m^r : \text{sign}(f(\bar{a})) = 1\}$. The sets S and T are finite, since B_m^r is finite. For each $\bar{b} \in \mathbb{R}^r$, we have $\bar{a} F_m^r \bar{b}$ iff $\mathfrak{R} \models \Phi_{\bar{a}}^{r,m}[\bar{b}]$, where $\Phi_{\bar{a}}^{r,m} := \bigwedge_{f \in S} f(\bar{x}) < 0 \wedge \bigwedge_{f \in T} \neg f(\bar{x}) < 0$. Obviously, since B_m^r is finite, there are only finitely many cells. In Section 4.3, we establish an upper bound on the number of cells with respect to r and m that allows us to prove a double exponential upper bound on the automata size for $\text{FO}(\mathbb{R}, +, <)$.

Moreover, we remark that the technique that we present in the following by connecting such decompositions of \mathbb{R}^r to first-order logic and Ehrenfeucht-Fraïssé games is reminiscent of techniques in computational complexity (see [14]) and constraint databases (see [26]). A novel insight is the relationship of the decompositions obtained from the family $(F_n^s)_{s,n \in \mathbb{N}}$ to the relation $\sim_{L(\varphi)}$ for a formula φ . See Section 4.2.

We start with some properties about the family $(F_n^s)_{s,n \in \mathbb{N}}$ of equivalence relations.

Lemma 7. *Let $\bar{a}, \bar{b} \in \mathbb{R}^r$. If $\bar{a} F_{4m^2}^r \bar{b}$ then for all $a' \in \mathbb{R}$, there is some $b' \in \mathbb{R}$ such that $(\bar{a}, a') F_m^{r+1} (\bar{b}, b')$.*

PROOF. For $m = 0$, there is nothing to prove. For $r = 0$, we must show that for every $a' \in \mathbb{R}$, there is some $b' \in \mathbb{R}$ such that $a' F_m^1 b'$. Obviously, $b' := a'$ works. In the following, we assume that $r, m > 0$.

Assume that $\bar{a} F_{4m^2}^r \bar{b}$ and let $a' \in \mathbb{R}$. We need to show the existence of some $b' \in \mathbb{R}$ such that

$$\text{sign}(f(\bar{a}) + ca' + p) = \text{sign}(f(\bar{b}) + cb' + p), \quad (4)$$

for all $f \in B_m^r$ and $c, p \in \mathbb{Z}$ with $|c|, |p| \leq m$.

Case $c = 0$. The equality (4) simplifies to $\text{sign}(f(\bar{a}) + p) = \text{sign}(f(\bar{b}) + p)$. We must show that the latter equality holds. Let $h(\bar{x}) := f(\bar{x}) + p$, for all $\bar{x} \in \mathbb{R}^r$. We have $h \in B_{4m^2}^r$, since $\|h\| \leq m$ and $|h(\bar{0})| \leq rm + m \leq 2rm \leq 4rm^2$. By the assumption $\bar{a} F_{4m^2}^r \bar{b}$, we conclude that $\text{sign}(f(\bar{a}) + p) = \text{sign}(f(\bar{b}) + p)$.

Case $c \neq 0$. The equality (4) can be rewritten to

$$a' < -\frac{f(\bar{a})+p}{c} \quad \text{iff} \quad b' < -\frac{f(\bar{b})+p}{c}.$$

To show the existence of b' , it suffices to show that the numbers $\frac{f(\bar{a})+p}{c}$, for all $f \in B_m^r$ and $c, p \in \mathbb{Z}$ with $c \neq 0$ and $|c|, |p| \leq m$ lie in the same order on the real line as the corresponding numbers $\frac{f(\bar{b})+p}{c}$. This is the case iff for all $f, g \in B_m^r$ and $c, d, p, q \in \mathbb{Z}$ with $c, d \neq 0$ and $|c|, |d|, |p|, |q| \leq m$, we have

$$\frac{f(\bar{a})+p}{c} < \frac{g(\bar{a})+q}{d} \quad \text{iff} \quad \frac{f(\bar{b})+p}{c} < \frac{g(\bar{b})+q}{d},$$

or in other words,

$$\text{sign}(df(\bar{a}) + dp - cg(\bar{a}) - cq) = \text{sign}(df(\bar{b}) + dp - cg(\bar{b}) - cq). \quad (5)$$

For the function $h(\bar{x}) := df(\bar{x}) + dp - cg(\bar{x}) - cq$, for all $\bar{x} \in \mathbb{R}^r$, we have $\|h\| \leq 2m^2$ and $|h(\bar{0})| \leq 2rm^2 + 2m^2 \leq 4rm^2$, that means, $h \in B_{4m^2}^r$. The equality (5) holds because of the assumption $\bar{a} F_{4m^2}^r \bar{b}$. \square

Similar to Lemma 1, we obtain the following lemma by using Lemma 7.

Lemma 8. *For all $\bar{a}, \bar{b} \in \mathbb{R}^r$, it holds that if $\bar{a} F_{2^{2m+2}}^r \bar{b}$ then $\bar{a} \equiv_m^r \bar{b}$.*

PROOF. It suffices to prove that $F_{2^3 \cdot 2^{m-2}}^r$ refines \equiv_m^r , for every $m, r \in \mathbb{N}$, since $2^{3 \cdot 2^m - 2} \leq 2^{4 \cdot 2^m} = 2^{2^{m+2}}$. We prove the claim by induction over $m \in \mathbb{N}$. The base case $m = 0$ is straightforward. Note that $2^{3 \cdot 2^0 - 2} = 2$ and recall that the atomic formulas are of the form $x \approx y$, $x < y$, and $x + y \approx z$. For the step case, we assume that the claim is true for some $m \geq 0$, that means, $F_{2^3 \cdot 2^m - 2}^r$ refines \equiv_m^r , where $r \in \mathbb{N}$. We must show that if $\bar{a} F_{2^3 \cdot 2^{m+1} - 2}^r \bar{b}$ then $\bar{a} \equiv_{m+1}^r \bar{b}$. In the following, we assume that $\bar{a} F_{2^3 \cdot 2^{m+1} - 2}^r \bar{b}$.

Since every formula with at most quantifier depth $m+1$ is logically equivalent to a Boolean combination of formulas of the form $\exists y \varphi$ with $\text{qd}(\varphi) \leq m$, it suffices to show that $\mathfrak{R} \models \exists y \varphi[\bar{a}] \Leftrightarrow \mathfrak{R} \models \exists y \varphi[\bar{b}]$, where φ is a formula with $\text{qd}(\varphi) \leq m$. By symmetry, we only need to prove the direction from left to right, that means, $\mathfrak{R} \models \exists y \varphi[\bar{a}]$ implies $\mathfrak{R} \models \exists y \varphi[\bar{b}]$. If $\mathfrak{R} \models \varphi[\bar{a}, a']$, for some $a' \in \mathbb{R}$, then by Lemma 7, there is some $b' \in \mathbb{R}$ such that $(\bar{a}, a') F_{2^3 \cdot 2^m - 2}^{r+1} (\bar{b}, b')$. Note that $4 \cdot (2^{3 \cdot 2^m - 2})^2 = 4 \cdot 2^{3 \cdot 2^{m+1} - 4} = 2^{3 \cdot 2^{m+1} - 2}$. By the induction hypothesis, we have $(\bar{a}, a') \equiv_{m+1}^{r+1} (\bar{b}, b')$. We conclude that $\mathfrak{R} \models \varphi[\bar{b}, b']$ and hence, $\mathfrak{R} \models \exists x \varphi[\bar{b}]$. \square

The following lemma shows how to obtain a set $R \subseteq \mathbb{R}^r$ such that each equivalence class of F_m^r has at least one representative in R . Note that the cases where $m = 0$ or $r = 0$ are easy since F_m^r has only a single equivalence class. Hence, $\bar{a} \in \sigma$, for every $\bar{a} \in \mathbb{R}^r$ and every equivalence class σ of F_m^r if $m = 0$ or $r = 0$. Let σ be an equivalence class of F_m^r and let σ' be an equivalence class of F_n^{r+1} , where $n \in \mathbb{N}$. We say that σ' is *consistent* with σ if $(\sigma \times \mathbb{R}) \cap \sigma' \neq \emptyset$.

Lemma 9. *Let $m, r > 0$, $\bar{a} \in \mathbb{R}^r$, and σ the equivalence class of \bar{a} with respect to $F_{2m^2}^r$. For every equivalence class σ' of F_m^{r+1} that is consistent with σ , we have*

$$\sigma' \cap \left\{ \left(\bar{a}, \frac{h(\bar{a})+d}{c} \right) : h \in B_{2m^2}^r \text{ and } c, d \in \mathbb{Z} \text{ with } c \neq 0 \text{ and } |c|, |d| \leq 2m^2 \right\} \neq \emptyset.$$

PROOF. Let $a', b' \in \mathbb{R}$. We have $(\bar{a}, a') F_m^{r+1} (\bar{a}, b')$ iff for all $f \in B_m^r$ and $c, c' \in \mathbb{Z}$ with $c' \neq 0$ and $|c|, |c'| \leq m$, we have

$$\text{sign}(c + f(\bar{a}) + c'a') = \text{sign}(c + f(\bar{a}) + c'b').$$

This equality is equivalent to

$$a' < -\frac{c+f(\bar{a})}{c'} \quad \text{iff} \quad b' < -\frac{c+f(\bar{a})}{c'}.$$

That means, all equivalence classes of F_m^{r+1} that are consistent with σ can be represented by a pair (\bar{a}, a') , where a' is a rational number that satisfies one of the following properties:

- (i) a' is the rational number $\frac{c+f(\bar{a})}{c'}$, where $f \in B_m^r$, and $c, c' \in \mathbb{Z}$ with $c' \neq 0$ and $|c|, |c'| \leq m$.
- (ii) a' is a rational number in an interval strictly between the adjacent rational numbers $\frac{c+f(\bar{a})}{c'}$ and $\frac{d+g(\bar{a})}{d'}$, where $f, g \in B_m^r$, and $c, c', d, d' \in \mathbb{Z}$ with $c', d' \neq 0$ and $|c|, |c'|, |d|, |d'| \leq m$.

- (iii) a' is a rational number strictly less than the smallest rational number of the form $\frac{c+f(\bar{a})}{c'}$, where $f \in B_m^r$, $c, c' \in \mathbb{Z}$ with $c' \neq 0$ and $|c|, |c'| \leq m$.
- (iv) a' is a rational number strictly greater than the largest rational number of the form $\frac{c+f(\bar{a})}{c'}$, where $f \in B_m^r$, $c, c' \in \mathbb{Z}$ with $c' \neq 0$ and $|c|, |c'| \leq m$.

For (i), we choose $a' := \frac{h(\bar{a})}{c'}$, where h is the integer affine function $h(\bar{x}) := f(\bar{x}) + c$, for $\bar{x} \in \mathbb{R}^r$. Note that $|c'| \leq 2m^2$ and $h \in B_{2m^2}^r$ since $\|h^*\| = \|f^*\|$ and $|h(\bar{0})| \leq rm + m \leq 2rm^2$. For (iii), we chose $a' := \frac{c+f(\bar{a})}{c'} - 1 = \frac{c+f(\bar{a})-c'}{c'}$ and analogously, for (iv), we choose $a' := \frac{c+f(\bar{a})}{c'} + 1 = \frac{c+f(\bar{a})+c'}{c'}$. Since $|c + f(\bar{0}) \pm c'| \leq m + rm + m = rm + 2m \leq 2rm^2$ in both cases, there is an appropriate integer affine function $h \in B_{2m^2}^r$. For (ii), we choose the midpoint of the interval, that means,

$$a' := \frac{1}{2} \left(\frac{c+f(\bar{a})}{c'} + \frac{d+g(\bar{a})}{d'} \right) = \frac{d'c + d'f(\bar{a}) + c'd + c'g(\bar{a})}{2c'd'}.$$

Note that $|2c'd'| \leq 2m^2$. Let $h(\bar{x}) := d'c + d'f(\bar{x}) + c'd + c'g(\bar{x})$, for $\bar{x} \in \mathbb{R}^r$. We have $\|h^*\| \leq 2m^2$ and $|h(\bar{0})| \leq rm^2 + rm^2 + m^2 + m^2 = 2rm + m^2$. For $|h(\bar{0})| \leq 2rm^2$, we have $h \in B_{2m^2}^r$ and we are done. For $|h(\bar{0})| > 2rm^2$, we define $h'(\bar{x}) := h(\bar{x}) - \ell$, for $\bar{x} \in \mathbb{R}^r$ and some appropriate $\ell \in \mathbb{Z}$ with $|\ell| \leq 2m^2$ such that $h' \in B_{2m^2}^r$ and $h'(\bar{a}) + \ell = h(\bar{a})$. \square

Remark 10. 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 $(F_n^s)_{s,n \in \mathbb{N}}$ of equivalence relations and the family of equivalence relations defined in Kozen's book [21] in Chapter 22.

Kozen is not using the function $\text{sign} : \mathbb{R} \rightarrow \{0, 1\}$ but the standard signum function $\text{sgn} : \mathbb{R} \rightarrow \{-1, 0, 1\}$ that is defined as $\text{sgn}(x) := -1$ if $x < 0$, $\text{sgn}(x) := 1$ if $x > 0$, and $\text{sgn}(0) := 0$. In Kozen's book, two elements $\bar{a}, \bar{b} \in \mathbb{R}^r$ are related iff $\text{sgn}(f(\bar{a})) = \text{sgn}(f(\bar{b}))$, for all integer affine function $f \in A^r$ with $\|f\| \leq m$ and $|f(\bar{0})| \leq m$.

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 slightly simpler when using the function sign . Second and more relevant, we require $|f(\bar{0})| \leq rm$ instead of $|f(\bar{0})| \leq m$. An immediate consequence of only requiring this weaker restriction on the functions $f \in A^r$ is that the equivalence relation F_m^r refines the corresponding equivalence relation as defined in Kozen's book. The purpose for having finer equivalence relations is the following: For a formula $\varphi(x_1, \dots, x_r)$, we show in Section 4.2 that the equivalence classes of $\sim_{L(\varphi)}$ are related to the equivalence classes of a certain relation in the family $(F_n^s)_{s,n \in \mathbb{N}}$. Without the weaker requirement we were not able to establish a similar relationship. The problem can be pinpointed to Lemma 12, which is crucial in relating the equivalence relations. The corresponding statement of Lemma 12 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 $F_{2^{2m+2+1}}^r$ and the congruence relation $\sim_{L(\varphi)}$, where $\varphi(x_1, \dots, x_r)$ is a formula with $\text{qd}(\varphi) \leq m$. Namely, we show that $F_{2^{2m+2+1}}^r$ 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 11. *For $f \in A^r$, $u \in (\Sigma^r)^+$, $u' \in (\Sigma^r)^*$, and $\gamma \in (\Sigma^r)^\omega$, the following properties 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 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 subtle technical detail, which is reflected in the (b)-parts of the lemmas. The words $u \star u'$ and $u \star u' \bar{0}$ represent the same vector of real numbers, that means, $\langle u \star u' \rangle = \langle u \star u' \bar{0} \rangle$. 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, that means, $\langle u \star u' \gamma \rangle \neq \langle u \star u' \bar{0} \gamma \rangle$ whenever $\gamma \neq \bar{0}^\omega$. In particular, $u \star u' \gamma$ and $u \star u' \bar{0} \gamma$ may represent different equivalence classes in F_m^r .

Lemma 12. *For all $u, v \in (\Sigma^r)^+$ and $u', v' \in (\Sigma^r)^*$, the following two properties hold:*

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

PROOF. For $r = 0$, there is nothing to prove. In the following, we assume that $r > 0$.

We prove (a) by contraposition. Assume that for some $w \in (\Sigma^r)^*$, it is not the case that $\langle uw \rangle F_m^r \langle vw \rangle$, that means, there is some $f \in B_m^r$ with $\text{sign}(f(\langle uw \rangle)) \neq \text{sign}(f(\langle vw \rangle))$. Without loss of generality, we assume that $f(\langle uw \rangle) < 0$ and $f(\langle vw \rangle) \geq 0$.

We have $\varrho^{|w|} f^*(\langle u \rangle) + f(\langle \bar{0}w \rangle) < 0$ 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 $\text{sign}(f^*(\langle u \rangle)) \neq \text{sign}(f^*(\langle v \rangle))$ then $\langle u \rangle F_m^r \langle v \rangle$ does not hold and we are done. So, assume that $\text{sign}(f^*(\langle u \rangle)) = \text{sign}(f^*(\langle v \rangle))$. 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 $\text{sign}(g(\langle u \rangle)) \neq \text{sign}(g(\langle v \rangle))$. 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 $f^*(\langle u \rangle) > rm$. Note that $|f^*(\langle u \rangle)|, |f^*(\langle v \rangle)| > rm$ and $\text{sign}(f^*(\langle u \rangle)) = \text{sign}(f^*(\langle v \rangle))$ by assumption. Analogously, as in the above case, we obtain a contradiction to the assumption $f(\langle uw \rangle) < 0$.

For proving (b), let f be an arbitrary function in B_m^r and $\gamma \in (\Sigma^r)^\omega$. We must show that $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle))$. 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 $\text{sign}(f(\langle u \star u' \rangle)) = \text{sign}(f(\langle v \star v' \rangle))$. 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

$$f(\langle u \star u' \gamma \rangle) = f(\bar{0}) + a + c\varrho^{-|u'|} \quad (6)$$

and

$$f(\langle v \star v' \bar{0}^k \gamma \rangle) = f(\bar{0}) + b + c\varrho^{-|v'|-k}. \quad (7)$$

If $c \geq 0$ then $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle)) = 1$. 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 $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle)) = 1$.

Case $a = b$. For $k = |u'| - |v'|$, it immediately follows from the equalities (6) and (7) that $f(\langle u \star u' \gamma \rangle) = f(\langle v \star v' \bar{0}^k \gamma \rangle)$, and hence $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle))$. For $a = b = -f(\bar{0})$, it is also straightforward to see from the two equalities (6) and (7) that $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle))$. 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 (6) and (7) that $\text{sign}(f(\langle u \star u' \gamma \rangle)) = \text{sign}(f(\langle v \star v' \bar{0}^k \gamma \rangle)) = 0$. 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 $f(\langle v \star v' \bar{0}^k \gamma \rangle) \geq \frac{1}{\varrho^{|v'|}} + \frac{c}{\varrho^{|v'|+k}} = \frac{\varrho^k - |c|}{\varrho^{|v'|+k}} \geq \frac{rm - rm}{\varrho^{|v'|+k}} \geq 0$. \square

Lemma 13. *Let $\varphi(x_1, \dots, x_r)$ be a formula with $\text{qd}(\varphi) \leq m$. For all $u, v \in (\Sigma^r)^+$ and $u', v' \in (\Sigma^r)^*$, the following two properties hold:*

(a) If $\langle u \rangle F_{2^{2m+2+1}}^r \langle v \rangle$ then $u \sim_{L(\varphi)} v$.

(b) If $\langle u \star u' \rangle F_{2^{2m+2+1}}^r \langle v \star v' \rangle$ 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 12(a), it follows that $\langle uw \rangle F_{2^{2m+2+1}}^r \langle vw \rangle$, for all $w \in (\Sigma^r)^*$. With Lemma 12(b), we obtain that $\langle uw \star \gamma \rangle F_{2^{2m+2}}^r \langle vw \star \gamma \rangle$, for all $w \in (\Sigma^r)^*$ and $\gamma \in (\Sigma^r)^\omega$. By Lemma 8, we conclude that $\langle uw \star \gamma \rangle \equiv_m^r \langle vw \star \gamma \rangle$, for all $w \in (\Sigma^r)^*$ and $\gamma \in (\Sigma^r)^\omega$. In particular, we have $uw \star \gamma \in L(\varphi) \Leftrightarrow vw \star \gamma \in L(\varphi)$, 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 $u\alpha, v\alpha \notin L(\varphi)$. \square

4.3. Upper Bounds

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

Lemma 14. *The cardinality of B_m^r is at most $(2rm + 1)(2m + 1)^r$.*

Using the Lemmas 9 and 14, we establish an upper bound on the index of F_m^r .

Lemma 15. *The index of F_m^r is at most $\max\{1, 2^{2r+4} \cdot m^{2r+3}\}$.*

PROOF. The index of F_m^r is 1 if $r = 0$ or $m = 0$. In the following, we assume that $r > 0$ and $m > 0$.

For establishing the upper bound on the index of F_m^r , we consider the sequence $(F_{2^{2r-\ell-1}, m^{2r-\ell}}^\ell)_{0 \leq \ell \leq r}$. Note that the last element in the sequence is the equivalence relation F_m^r and $F_{2^{2r-1}, m^{2r}}^0$ is the first element in the sequence. For the integer ℓ with $0 \leq \ell \leq r$, let I^ℓ denote the index of the equivalence relation $F_{2^{2r-\ell-1}, m^{2r-\ell}}^\ell$. We have $I^0 = 1$ and by Lemma 9, we know that for the integer ℓ with $0 \leq \ell < r$,

$$I^{\ell+1} \leq I^\ell \cdot |B_{2^{2r-\ell-1}, m^{2r-\ell}}^\ell| \cdot (2^{2r-\ell-1} \cdot m^{2r-\ell})^2.$$

We obtain that

$$I^r \leq \prod_{0 \leq \ell < r} |B_{2^{2r-\ell-1}, m^{2r-\ell}}^\ell| \cdot (2^{2r-\ell-1} \cdot m^{2r-\ell})^2.$$

With Lemma 14, it follows that

$$\begin{aligned} I^r &\leq \prod_{0 \leq \ell < r} ((\ell + 1) \cdot 2^{2r-\ell} \cdot m^{2r-\ell}) \cdot (2 \cdot 2^{2r-\ell} \cdot m^{2r-\ell})^\ell \cdot (2^{2r-\ell+1} \cdot m^{2r-\ell+1}) \\ &= 2^{(\sum_{0 \leq \ell < r} (\ell+1)2^{r-\ell}) + 2(\sum_{0 \leq \ell < r} 2^{r-\ell})} \cdot m^{(\sum_{0 \leq \ell < r} (\ell+1)2^{r-\ell}) + 2(\sum_{0 \leq \ell < r} 2^{r-\ell})} \\ &\quad \left(\prod_{0 \leq \ell < r} (\ell + 1) \cdot 2^\ell \right) \\ &\leq 2^{2^{r+2} + 2^{r+2}} \cdot m^{2^{r+2} + 2^{r+2}} \cdot 2^{\sum_{0 \leq \ell \leq r} 2^\ell} \\ &\leq 2^{2^{r+4}} \cdot m^{2^{r+3}}. \end{aligned}$$

Note that $\sum_{0 \leq \ell < r} 2^{r-\ell} \leq 2^{r+1}$, $\sum_{0 \leq \ell < r} (\ell+1) \cdot 2^{r-\ell} \leq 2^{r+2}$, and $\prod_{0 \leq \ell < r} (\ell+1) \cdot 2^\ell \leq 2^{\sum_{0 \leq \ell < r} 2^\ell}$. Proofs for the latter two inequalities are given in the appendix. \square

Theorem 16. *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 .*

PROOF. Let r be the number of free variables in φ and $m := \text{qd}(\varphi)$. We use $F_{2^{2^{m+2}+1}}^r$ 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 $\{u \in (\Sigma^r)^+ : \langle v \rangle F_{2^{2^{m+2}+1}}^r \langle u \rangle\}$.

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 $S := \{u \star u' \in F : \langle v \star v' \rangle F_{2^{2^{m+2}+1}}^r \langle u \star u' \rangle\}$. R chops S into equivalence classes, assuming $|v'| \leq |u'|$, for all $u \star u' \in S$:

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

We now show that R refines $\sim_{L(\varphi)}$. Assume that $x R y$, for $x, y \in (\Sigma^r \cup \{\star\})^*$. For $x = y = \varepsilon$, or the letter \star occurs at least twice in x and y , then obviously $x \sim_{L(\varphi)} y$. For $x, y \in (\Sigma^r)^+$, it follows from Lemma 13(a) that $x \sim_{L(\varphi)} y$. For $x, y \in F$, assume that x and y have the form $u \star u'$ and $v \star v'$, respectively with $|u'| \geq |v'|$. It follows from Lemma 13(b) that $u \star u' \sim_{L(\varphi)} v \star v'$.

It remains to prove an upper bound on the index of R . Note that $n \geq m + r \geq 1$. By Lemma 15, an upper bound on the index of $F_{2^{2^{m+2}+1}}^r$ is

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

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

Remark 17. 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 $L(\varphi)$ [28]. Thus, Theorem 16 establishes a double exponential upper bound with respect to the formula length on the size of the minimal number of states of a Büchi automaton that accepts $L(\varphi)$.

We also remark that this double exponential upper bound on the automata size is tight, that means, 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 . We omit the details, since with only minor modifications the definition of the formulas φ_n and the proof of the upper bound on the index of $\sim_{L(\varphi_n)}$ are similar to the corresponding definition and proof for the tightness of the upper bound on the automaton size for Presburger arithmetic [18].

Last but not least, we point out that the double exponential upper bound on the automata size for $\text{FO}(\mathbb{R}, +, <)$ as well as the exponential upper bound on the automata size for $\text{FO}(\mathbb{Z}, <)$ in Section 3 is with respect to the used word encoding of the reals and the integers, respectively. Our word encodings are based on the ϱ 's big-endian complement representation, for some base $\varrho \in \mathbb{N}$ with $\varrho \geq 2$ (see Section 2.3). It is not obvious whether these upper bounds carry over to other word encodings, for example, word encodings based on the little-endian complement representation. For obtaining such upper bounds for another word encoding, it would suffice to find families of equivalence relations $(\tilde{E}_n^s)_{s,n \in \mathbb{N}}$ and $(\tilde{E}_n^s)_{s,n \in \mathbb{N}}$ that have similar properties as $(E_n^s)_{s,n \in \mathbb{N}}$ and $(F_n^s)_{s,n \in \mathbb{N}}$, respectively. In particular, these families must have similar congruence properties as the ones stated in Lemma 3 and Lemma 12 for $(E_n^s)_{s,n \in \mathbb{N}}$ and $(F_n^s)_{s,n \in \mathbb{N}}$, respectively as these congruence properties are crucial for establishing the relationship between their equivalence classes and automata states.

5. Conclusion

This article 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 upper bounds on the minimal sizes of automata that represent sets definable in $\text{FO}(\mathbb{Z}, <)$ and $\text{FO}(\mathbb{R}, +, <)$. For $\text{FO}(\mathbb{R}, +, <)$, the presented double exponential upper bound on the automata size is tight and previously proposed techniques based on quantifier-elimination methods [18] failed to establish a double exponential upper bound on the automata size. We hope that the new insights will eventually lead to more efficient automata constructions that can be used to decide $\text{FO}(\mathbb{R}, +, <)$ more efficiently.

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

Appendix A. Additional Proof Details

Appendix A.1. Proof of Lemma 1

We first show the following back-and-forth property about the family $(E_n^s)_{s,n \in \mathbb{N}}$ of equivalence relations:

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

Assume $\bar{a} E_{2^{m+1}}^r \bar{b}$. For $a' \in \mathbb{Z}$, let $k \in \{1, \dots, r\}$ be an index so that $d := |a_k - a'|$ is minimal. We choose

$$b' := \begin{cases} b_k + \min\{d, 2^m\} & \text{if } a_k \leq a', \\ b_k - \min\{d, 2^m\} & \text{otherwise.} \end{cases}$$

For $d = 0$, the proof is straightforward, since $a' = a_k$ and also $b' = b_k$ by definition. Assume $d > 0$. By definition, we have $a' < a_k \Leftrightarrow b' < b_k$. Without loss of generality, we assume that $a' < a_k$ and $b' < b_k$. The other case is symmetric. It remains to show that $\text{sign}(a_i - a' - c) = \text{sign}(b_i - b' - c)$ and $\text{sign}(a' - a_i - c) = \text{sign}(b' - b_i - c)$, for all $c, i \in \mathbb{N}$ with $c \leq 2^m$ and $1 \leq i \leq r$. We show that $\text{sign}(a' - a_i - c) = \text{sign}(b' - b_i - c)$. The case $\text{sign}(a_i - a' - c) = \text{sign}(b_i - b' - c)$ can be proved analogously and we omit it.

Case $a' - a_i < 0$. We have $a_k \leq a_i$, otherwise k is not minimal. From the assumption $\bar{a} E_{2^{m+1}}^r \bar{b}$, it follows that $b_k \leq b_i$. By definition, we have $b' - b_i = b_k - b_i - \min\{d, 2^m\} < 0$. So, it holds that $a' - a_i - c < 0$ and $b' - b_i - c < 0$.

Case $a' - a_i \geq 0$. If $d \leq 2^m$, we have

$$a' - a_i = a_k - d - a_i = a_k - a_i - d$$

and

$$b' - b_i = b_k - d - b_i = b_k - b_i - d.$$

From the assumption $\bar{a} E_{2^{m+1}}^r \bar{b}$, it follows that $\text{sign}(a' - a_i - c) = \text{sign}(b' - b_i - c)$. Note that $c + d \leq 2^{m+1}$.

If $d > 2^m$ then $a' - a_i > 2^m$ because of the choice of k . It follows that $\text{sign}(a' - a_i - c) = 1$. From the choice of k , we conclude that $a_k - a_i \geq 2^{m+1}$. From the assumption $\bar{a} E_{2^{m+1}}^r \bar{b}$, it follows that $b_k - b_i \geq 2^{m+1}$, and thus,

$$b' - b_i = b_k - 2^m - b_i = b_k - b_i - 2^m \geq 2^{m+1} - 2^m = 2^m.$$

That means, we have $\text{sign}(b' - b_i - c) = 1$. This completes the proof of the back-and-forth property (1).

Now, we prove the lemma by induction over $m \geq 0$. For $m = 0$, it is straightforward to see that if $\bar{a} E_1^r \bar{b}$ then \bar{a} and \bar{b} satisfy the same formulas φ with $\text{qd}(\varphi) = 0$. Note that an atomic formula is of the form $x < y$ or $x \approx y$. So, we have $\bar{a} \equiv_0^r \bar{b}$.

Assume that the claim is true for $m \geq 0$. Furthermore, assume that $\bar{a} E_{2^{m+1}}^r \bar{b}$. We must show that $\bar{a} \equiv_{m+1}^r \bar{b}$. Since every formula with at most quantifier depth $m + 1$ is logically equivalent to a Boolean combination of formulas of the form $\exists y \varphi$ with $\text{qd}(\varphi) \leq m$, it suffices to show that $\mathfrak{Z} \models \exists y \varphi[\bar{a}] \Leftrightarrow \mathfrak{Z} \models \exists y \varphi[\bar{b}]$, where φ is a formula with $\text{qd}(\varphi) \leq m$. By symmetry, we only need to prove the direction from left to right, that means, $\mathfrak{Z} \models \exists y \varphi[\bar{a}]$ implies $\mathfrak{Z} \models \exists y \varphi[\bar{b}]$. If $\mathfrak{Z} \models \varphi[\bar{a}, a']$, for some $a' \in \mathbb{Z}$, then by the property (1), there is some $b' \in \mathbb{Z}$ such that $(\bar{a}, a') E_{2^m}^{r+1} (\bar{b}, b')$. By the induction hypothesis, we have $(\bar{a}, a') \equiv_m^{r+1} (\bar{b}, b')$. We conclude that $\mathfrak{Z} \models \varphi[\bar{b}, b']$ and hence, $\mathfrak{Z} \models \exists y \varphi[\bar{b}]$.

Appendix A.2. Additional Proof Details of Lemma 15

Let $r > 0$. In the following, we provide proof details for the inequalities $\sum_{0 \leq \ell < r} (\ell + 1) \cdot 2^{r-\ell} \leq 2^{r+2}$ and $\prod_{0 \leq \ell < r} (\ell + 1) \cdot 2^\ell \leq 2^{\sum_{0 \leq \ell < r} 2^\ell}$, which we used to prove Lemma 15.

We prove the first inequality by induction over r . For $r = 1$, the inequality is obviously true. For the step case, assume that $r > 1$. We have

$$\begin{aligned} \sum_{0 \leq \ell < r} (\ell + 1) \cdot 2^{r-\ell} &= 2^r + \left(\sum_{1 \leq \ell < r} 2^{r-\ell} \right) + 1 - 1 + \left(\sum_{1 \leq \ell < r} \ell \cdot 2^{r-\ell} \right) \\ &= 2^r + 2^r - 2 + \sum_{0 \leq \ell < r-1} (\ell + 1) \cdot 2^{r-1-\ell}. \end{aligned}$$

With the induction hypothesis, we obtain

$$2^{r+1} - 2 + \sum_{0 \leq \ell < r-1} (\ell + 1) \cdot 2^{r-1-\ell} \leq 2^{r+1} - 2 + 2^{r+1} \leq 2^{r+2}.$$

The second inequality follows easily from $\prod_{0 \leq \ell < r} (\ell + 1) \cdot 2^\ell \leq \prod_{0 \leq \ell < r} 2^{\ell+1}$. $2^\ell = 2^{\sum_{0 \leq \ell < r} (2^{\ell+1})} = 2^{r^2}$.

References

- [1] S. Bardin, A. Finkel, J. Leroux, and L. Petrucci. FAST: Fast acceleration of symbolic transition systems. In *Proceedings of the 15th International Conference on Computer Aided Verification (CAV'03)*, volume 2725 of *Lect. Notes Comput. Sci.*, pages 118–121. Springer-Verlag, 2003.
- [2] S. Bardin, J. Leroux, and G. Point. FAST extended release. In *Proceedings of the 18th International Conference on Computer Aided Verification (CAV'06)*, volume 4144 of *Lect. Notes Comput. Sci.*, pages 63–66. Springer-Verlag, 2006.
- [3] B. Becker, C. Dax, J. Eisinger, and F. Klaedtke. LIRA: Handling constraints of linear arithmetics over the integers and the reals. In *Proceedings the 19th International Conference on Computer Aided Verification (CAV'07)*, volume 4590 of *Lect. Notes Comput. Sci.*, pages 307–310. Springer-Verlag, 2007.
- [4] A. Blumensath and E. Grädel. Finite presentations of infinite structures: Automata and interpretations. *Theory Comput. Syst.*, 37(6):641–674, 2004.
- [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(3):614–633, 2005.
- [6] B. Boigelot and P. Wolper. Representing arithmetic constraints with finite automata: an overview. In *Proceedings of the 18th International Conference on Logic Programming (ICLP'02)*, volume 2401 of *Lect. Notes Comput. Sci.*, pages 1–19. Springer-Verlag, 2002.

- [7] J. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6:66–92, 1960.
- [8] J. Büchi. On a decision methods in restricted second order arithmetic. In *Proceedings of the 1960 International Congress on Logic, Methodology and Philosophy of Science*, pages 1–11. Stanford University Press, 1962.
- [9] D. C. Cooper. Theorem proving in arithmetic without multiplication. In B. Meltzer and D. Michie, editors, *Proceedings of the 7th Annual Machine Intelligence Workshop*, pages 91–100. Edinburgh University Press, 1972.
- [11] H.-D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic*. Springer-Verlag, 2nd edition, 1994.
- [12] J. Eisinger. Upper bounds on the automata size for integer and mixed real and integer linear arithmetic (extended abstract). In *Proceedings of the 17th Annual Conference of the EACSL on Computer Science Logic (CSL'08)*, volume 5213 of *Lect. Notes Comput. Sci.*, pages 431–445. Springer-Verlag, 2008.
- [13] J. Ferrante and C. W. Rackoff. A decision procedure for the first order theory of real addition with order. *SIAM J. Comput.*, 4(1):69–76, 1975.
- [14] J. Ferrante and C. W. Rackoff. *The Computational Complexity of Logical Theories*, volume 718 of *Lect. Notes Math.*. Springer-Verlag, 1979.
- [15] B. Khossainov and A. Nerode. Automatic presentations of structures. In *Proceedings of the International Workshop on Logical and Computational Complexity (LCC'94)*, volume 960 of *Lect. Notes Comput. Sci.*, pages 367–392. Springer-Verlag, 1995.
- [16] B. Khossainov, A. Nies, S. Rubin, and F. Stephan. Automatic structures: Richness and limitations. *Log. Methods Comput. Sci.*, 3(2), 2007.
- [17] B. Khossainov, S. Rubin, and F. Stephan. Definability and regularity in automatic structures. In *Proceedings of the 21st Annual Symposium on Theoretical Aspects of Computer Science (STACS'04)*, volume 2996 of *Lect. Notes Comput. Sci.*, pages 440–451. Springer-Verlag, 2004.
- [18] F. Klaedtke. Bounds on the automata size for Presburger arithmetic. *ACM Trans. Comput. Log.*, 9(2), 2008.
- [19] F. Klaedtke. Ehrenfeucht-Fraïssé goes automatic for real addition. In *Proceedings of the 25th International Symposium on Theoretical Aspects of Computer Science (STACS'08)*, pages 445–456. IBFI Schloss Dagstuhl, 2008. Available online from <http://stacs-conf.org>.
- [20] N. Klarlund, A. Møller, and M. I. Schwartzbach. MONA implementation secrets. *Int. J. Found. Comput. Sci.*, 13(4):571–586, 2002.

- [21] D. Kozen. *Theory of Computation*. Springer-Verlag, 2006.
- [22] D. Kuske and M. Lohrey. Automatic structures of bounded degree revisited. In *Proceedings of the 18th Annual Conference of the EACSL on Computer Science Logic (CSL'09)*, volume 5771 of *Lect. Notes Comput. Sci.*, pages 364–378. Springer-Verlag, 2009.
- [23] R. E. Ladner. Application of model theoretic games to discrete linear orders and finite automata. *Information and Control*, 33:281–303, 1977.
- [24] R. McNaughton and S. A. Papert. *Counter Free Automata*. The MIT Press, 1971.
- [25] A. R. Meyer. Weak monadic second-order theory of successor is not elementary-recursive. In *Logic Colloquium*, volume 453 of *Lect. Notes Math.*, pages 132–154. Springer-Verlag, 1975.
- [26] J. Paredaens, J. Van den Bussche, and D. Van Gucht. First-order queries on finite structures over the reals. *SIAM J. Comput.*, 27(6):1747–1763, 1998.
- [27] C. Reddy and D. W. Loveland. Presburger arithmetic with bounded quantifier alternation. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing (STOC'78)*, pages 320–325. ACM Press, 1978.
- [28] L. Staiger. Finite-state ω -languages. *J. Comput. Syst. Sci.*, 27(3):434–448, 1983.
- [29] L. Stockmeyer. *The complexity of decision problems in automata theory and logic*. PhD thesis, Department of Electrical Engineering, MIT, Boston, MA, USA, 1974.