Second-order finite automata: expressive power and simple proofs using automatic structures

Dietrich Kuske

Technische Universität Ilmenau, Germany
dietrich.kuske@tu-ilmenau.de

Abstract. Second-order finite automata, introduced recently by Andrade de Melo and de Oliveira Oliveira, represent classes of languages. Since their semantics is defined by a synchronized rational relation, they can be studied using the theory of automatic structures. We exploit this connection to uniformly reprove and strengthen known and new results regarding closure and decidability properties concerning these automata. We then proceed to characterize their expressive power in terms of automatic classes of languages studied by Jain, Luo, and Stephan.

Keywords: classes of languages · automatic classes.

1 Introduction

Andrade de Melo and de Oliveira Oliveira [1] propose a mechanism to represent possibly infinite classes of regular languages by a single finite automaton $\mathcal{A}$. The idea is to start with an alphabet of simple automata that can make only one step. A word W over this alphabet is understood as a concatenation of such small automata, and therefore as an automaton $\mathcal{A}_W$. Consequently, the "second-order finite automaton" $\mathcal{A}$ describes a class of languages: the class of languages accepted by these finite automata $\mathcal{A}_W$ for W accepted by $\mathcal{A}$. We call such a class "full-length regular". The central result in [1] is an effective canonisation procedure for second-order finite automata. Then, the authors derive effective closure and decidability results for the collection of all full-length regular classes.

Recall that at the basis of the definition of second-order finite automata and their language class lies the interpretation of a word W from $L(\mathcal{A})$ as an automaton $\mathcal{A}_W$. We consider the natural binary relation of all pairs (W, w) where the NFA $\mathcal{A}_W$ accepts the word w. Since this relation is synchronized rational (a basic observation not made explicte in [1]), we can use automatic structures [7, 9, 4] as a tool to reason about second-order finite automata – and this is the core of the current paper's first part. This approach gives a uniform and simple way

- to build several normalized second-order finite automata (e.g., saturated),
- to uniformly prove closure properties (e.g., intersection and difference) shown in [1] and to improve them partly, and
- to prove decidability of inclusion, equality, and disjointness uniformly (the results are known from [1]).

We demonstrate that it also allows to prove new closure properties (e.g., the class of differences of languages from two classes) and new decidabilities (e.g., whether the intervals of languages in a full-length regular class, ordered by inclusion, are of bounded size). In a nutshell, all these results hold since they amount to the evaluation of some formula (from an appropriate and proper extension of first-order logic) in some automatic structure.

The second part of this paper is devoted to the expressiveness of second-order finite automata. The definition of full-length regular classes of languages via a rational relation is very similar to that of automatic classes of languages from [8] (that studies the learnability of such classes). Fernau (discussion at "Computer Science in Russia 2020") conjectured the two concepts to be closely related; this paper's second part details and confirms his conjecture. At this point, it is only important that an automatic class is given by a regular language and a synchronized rational relation. We show that a class of languages is full-length regular iff it is automatic with a length-preserving synchronized rational relation.

This characterization allows us to reduce the isomorphism problem for automatic equivalence structures to that of full-length regular classes ordered by inclusion. As a consequence, this latter problem is undecidable.

A limitation of full-length regular classes is that all languages in such a class are sets of words of equal length. In this paper, we extend the definition from [1] to regular and to ω-regular classes (that can contain arbitrary finite and regular languages, resp.). We actually prove the above mentioned closure and decidability results for regular classes, but the proofs can be transfered to full-length regular and partly to ω-regular classes of languages. We also present characterisations of these classes in terms of automatic classes.

In summary, we investigate classes of languages presented by finite automata and we demonstrate that the established theory of automatic structures can be useful in this study.

2 Second-order finite automata and regular classes of languages

For an alphabet Σ, let Σ^*, Σ^+, and Σ^ω denote the set of finite, finite nonempty, and ω-words, resp. A language $L \subseteq \Sigma^*$ is *single-length* if all its words have the same length. A relation $R \subseteq \Gamma^* \times \Sigma^*$ is *length-reducing* (*length-increasing*, *length-preserving*, resp.) if $(u, v) \in R$ implies $|u| \geq |v|$ ($|u| \leq |v|$, $|u| = |v|$, resp.).

Definition. Let A and B be sets, $W \in A$, $L \subseteq A$, and $R \subseteq A \times B$ a relation. Then we set $W^R = \big\{w \in B \mid (W, w) \in R\big\}$ and $L^R = \big\{W^R \mid W \in L\big\}$.

Intuitively, we consider the relation R as a function $R\colon A \to \mathcal{P}(B)$. Then W^R is the image of W under this mapping and L^R is the class of images of elements of the set L. We apply these constructions mainly for $A = \Gamma^+$ and $B = \Sigma^+$.

Definition. Let Σ be some alphabet and S be some finite set. A (Σ, S)-*block* is a tuple $B = (I, T, F)$ where $I, F \subseteq S$ and $T \subseteq S \times \Sigma \times S$; $\mathcal{B}(\Sigma, S)$ denotes the set of all (Σ, S)-blocks.

A block is an NFA over the alphabet Σ with set of states S. We will consider sequences of such blocks as a single NFA and run a word by chosing a transition from the i^{th} block for its i^{th} letter. We found it convenient to think of a block as consisting of two copies of the set of *locations* S where the *transition* $(s_1, a, s_2) \in T$ connects the location s_1 from the first copy to the location s_2 from the second copy. The *initial locations* $\iota \in I$ are considered as elements of the first copy, the *final locations* $f \in F$ as belonging to the second copy.

Definition. An NFA over $\mathcal{B}(\Sigma, S)$ is called *second-order* or *SO automaton over Σ and S.*

For an NFA $M = (Q, \Gamma, I, \Delta, F)$ (with sets of initial states I, of transitions $\Delta \subseteq Q \times \Gamma \times Q$, and of final states F) over Γ, we write $L^+(M) \subseteq \Gamma^+$ for the set of nonempty words accepted by M.

We will define the second-order language of the SO automaton $\mathcal{A}$ which will be a class of languages over Σ. To this aim, we need the following relation.

Definition. The relation $\mathrm{Acc}_{\Sigma,S}$ consists of all pairs $(B_1\, B_2 \cdots B_m, c_1\, c_2 \cdots c_n)$ with $m \geq n \geq 1$, $B_i = (I_i, T_i, F_i) \in \mathcal{B}(\Sigma, S)$ for all $i \in [m]$, and $c_i \in \Sigma$ for all $i \in [n]$ such that there exist locations $s_1, s_2, \ldots, s_{n+1} \in S$ with

$$(1)\ s_1 \in I_1,\ (2)\ (s_i, c_i, s_{i+1}) \in T_i \text{ for all } i \in [n], \text{ and } (3)\ s_{n+1} \in F_n.$$

Intuitively, we understand the word W as an NFA over Σ. Its state space consists of $m+1$ layers of the set of locations S. The transitions from B_i connect the locations from layer i to those of layer $i+1$. The initial states of the NFA are the initial states of B_1 in the first layer, the final states are those of B_i in layer $i+1$ (for any $i \in [m]$). Then $(W, w) \in \mathrm{Acc}_{\Sigma,S}$ iff the word $w \in \Sigma^+$ is accepted by the NFA described by the word $W \in \mathcal{B}(\Sigma, S)^+$.

Definition. Let $\mathcal{A}$ be an SO automaton over Σ and S. Then the *second-order language of* $\mathcal{A}$ is the class $\mathcal{L}_2(\mathcal{A}) = L^+(\mathcal{A})^{\mathrm{Acc}_{\Sigma,S}} = \{W^{\mathrm{Acc}_{\Sigma,S}} \mid W \in L^+(\mathcal{A})\}$.

By the definition, $L^+(\mathcal{A})$ is a language over $\mathcal{B}(\Sigma, S)$, but $\mathcal{L}_2(\mathcal{A})$ is a class of ε-free languages over Σ, i.e., a subset of $\mathcal{P}(\Sigma^+)$.

Definition. A class of languages $\mathcal{C} \subseteq \mathcal{P}(\Sigma^+)$ is *regular* if there exists an SO automaton $\mathcal{A}$ over Σ and some finite set of locations S such that $\mathcal{C} = \mathcal{L}_2(\mathcal{A})$.

Since $\mathrm{Acc}_{\Sigma,S}$ is length-reducing, the class $\mathcal{L}_2(\mathcal{A})$ consists of finite languages, only. Hence regular classes of languages are classes of finite languages.

In [1], the authors consider words W only where all words in $W^{\mathrm{Acc}_{\Sigma,S}}$ are of length $|W|$. We capture this by the following concept.

Definition. A word over $\mathcal{B}(\Sigma, S)$ is *full-length* if at most its last block has a non-empty set of accepting states. An SO automaton $\mathcal{A}$ is *full-length* iff all words from $L^+(\mathcal{A})$ are full-length. A class $\mathcal{C}$ of languages is *full-length regular* if there exists a full-length SO automaton $\mathcal{A}$ with $\mathcal{C} = \mathcal{L}_2(\mathcal{A})$.

To overcome the limitation to classes of finite languages, we will now consider infinite words $\alpha \in \mathcal{B}(\Sigma, S)^\omega$ and understand them as "infinite NFAs" M over Σ that can accept some infinite language (of finite words).

Definition. The binary relation $\mathrm{Acc}^\omega_{\Sigma,S}$ consists of all pairs $(B_1\, B_2 \cdots, w)$ with $B_i \in \mathcal{B}(\Sigma, S)$ for all $i \geq 1$ and $w \in \Sigma^+$ such that $(B_1\, B_2 \cdots B_{|w|}, w) \in \mathrm{Acc}_{\Sigma,S}$.

Example 2.1. Let $M = (S, \Sigma, I, \Delta, F)$ be some NFA over Σ. We consider it as block $B = (I, \Delta, F) \in \mathcal{B}(\Sigma, S)$ and set $\alpha = B^\omega$. Then $\alpha^{\mathrm{Acc}^\omega_{\Sigma,S}} = L^+(M)$. Hence, all regular languages $K \subseteq \Sigma^+$ are of the form $\alpha^{\mathrm{Acc}^\omega_{\Sigma,S}}$ for some ω-word α.

Definition. A Büchi-automaton over $\mathcal{B}(\Sigma, S)$ is called an *SO Büchi-automaton* over Σ and S. Let $\mathcal{A}$ be an SO Büchi-automaton over Σ and S. Then the *second-order language of* $\mathcal{A}$ is the class $\mathcal{L}^\omega_2(\mathcal{A}) = \left\{\alpha^{\mathrm{Acc}^\omega_{\Sigma,S}} \mid \alpha \in L^\omega(\mathcal{A})\right\} \subseteq \mathcal{P}(\Sigma^+)$. A class $\mathcal{C} \subseteq \mathcal{P}(\Sigma^+)$ of languages is *ω-regular* if there exists an SO Büchi-automaton $\mathcal{A}$ over Σ and some finite set of locations S such that $\mathcal{C} = \mathcal{L}^\omega_2(\mathcal{A})$.

Note that, for any block $B \in \mathcal{B}(\Sigma, S)$, the ω-language $\{B^\omega\}$ is ω-regular. Hence, in view of Example 2.1, any class $\{K\}$ with $K \subseteq \Sigma^+$ regular is ω-regular.

Example 2.2. For $c \in \Sigma$, consider the block $B_c = \big(\{s\}, \{(s, c, s)\}, \{s\}\big)$ and let $L = \{B_c \mid c \in \Sigma\}^\omega$. For any $\alpha \in \Sigma^\omega$, the ω-regular class $L^{\mathrm{Acc}^\omega_{\Sigma,S}}$ contains the language of all prefixes of α, i.e., is uncountable and contains non-regular languages.

3 Closure properties and special representations of regular classes of languages

From the canonisation result in [1], the authors infer closure properties of the collection of all full-length regular classes of languages. This section is devoted to alternative proofs and strengthenings (e.g., by providing much smaller automata) of these results. For notational simplicity, we only give our proofs for the collection of all regular classes, the results as well as the proofs all carry over to full-length regular classes and to ω-regular classes (if not stated otherwise). Since the main tool in our proofs are automatic structures, we first sketch their definition and their relation to SO automata.

3.1 Automatic structures

Basically, automatic structures are relational structures whose universe and relations can be accepted by finite automata. This is rather straightforward for the universe and unary relations: they have to form regular languages. Relations of larger arity are required to be synchronized rational [6], i.e., accepted by a synchronous multi-head automaton.

Definition ([7, 9]). A relational structure $\mathcal{S} = \big(U, (R_i)_{i\in[k]}\big)$ with $R_i \subseteq U^{n_i}$ for $i \in [k]$ is *automatic* if there is an alphabet Σ such that $U \subseteq \Sigma^*$ is regular and R_i is synchronized rational for all $i \in [k]$.

For the current paper, the following is the most interesting example.

Theorem 3.1. *The relation* $\mathrm{Acc}_{\Sigma,S}$ *is effectively synchronized rational.*

Consequently, given SO automata $\mathcal{A}_i$ *over* Σ_i *and* S_i *(for* $i \in [n]$*), the following structure* $\mathcal{S}\big((\mathcal{A}_i)_{i\in[n]}\big)$ *is effectively automatic:*

- *Its universe is* $\bigcup_{i\in[n]}\big(\mathcal{B}(\Sigma_i, S_i)^+ \cup \Sigma_i^+\big)$.
- *Its relations are* $\mathcal{B}(\Sigma_i, S_i)^+$, Σ_i^+, $L^+(\mathcal{A}_i)$, *and* $\mathrm{Acc}_{\Sigma_i,S_i}$ *for* $i \in [n]$.

The proofs in this section are all based on relations in this structure that are defined by logical formulas. As an example, consider the formula

$$\forall w\colon \Big((W_1, w) \in \mathrm{Acc}_{\Sigma,S_1} \leftrightarrow \big((W_2, w) \in \mathrm{Acc}_{\Sigma,S_2} \wedge (W_3, w) \notin \mathrm{Acc}_{\Sigma,S_3}\big)\Big)$$

with three free variables W_1, W_2, and W_3. In $\mathcal{S}(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$, it expresses that

$$W_1^{\mathrm{Acc}_{\Sigma,S_1}} = W_2^{\mathrm{Acc}_{\Sigma,S_2}} \setminus W_3^{\mathrm{Acc}_{\Sigma,S_3}}$$

holds. We will therefore allow to write such Boolean combinations in formulas.

Furthermore, our formulas allow not only the classical first-order quantifiers $\exists$ and $\forall$, but also the following:

- *infinity quantifier* $\exists^\infty$ [3]: For instance, $\forall x \neg\exists^\infty y\colon E(x, y)$ holds in a directed graph iff the graph has finite out-degree.
- *boundedness quantifer* Ǝ [11]: For instance, the number of paths of length two between any two nodes of a possibly infinite directed graph is uniformly bounded iff the directed graph satisfies Ǝ$(x_1, x_2; y)\colon \big(E(x_1, y) \wedge E(y, x_2)\big)$.
- *Ramsey quantifier* Ɐ [14]: Let $k \in \mathbb{N}$ and let $\overline{x_i}$ be mutually disjoint k-tuples of variables (for $1 \le i \le n$). The formula Ɐ$(\overline{x_1}, \ldots, \overline{x_n})\colon \varphi(\overline{x_1}, \ldots, \overline{x_n})$ holds in a structure $\mathcal{S}$ if there exists an infinite k-ary relation R such that any n tuples from R satisfy φ. For instance, with $k = 2$, the formula Ɐ$(\overline{x_1}, \overline{x_2})\colon E(\overline{x_1}) \wedge E(\overline{x_2}) \wedge \big(\overline{x_1} = \overline{x_2} \vee \{x_{1,1}, x_{1,2}\} \cap \{x_{2,1}, x_{2,2}\} = \emptyset\big)$ expresses of a graph that it contains infinitely many mutually disjoint edges.

We denote the extension of first-order logic by these quantifiers by FO^+.

Theorem 3.2 ([9, 3, 11, 14]). *Let* $\mathcal{S} = \big(U, (R_i)_{i\in[n]}\big)$ *be an automatic structure and* $\varphi(x_1, \ldots, x_n)$ *a formula from* FO^+. *Then the relation* $\varphi^{\mathcal{S}} = \big\{\overline{u} \in U^n \bigm| \mathcal{S} \models \varphi(\overline{u})\big\}$ *of all witnesses for* φ *in* $\mathcal{S}$ *is effectively synchronized rational (uniformly in the automatic structure* $\mathcal{S}$ *given by a tuple of finite automata).*

The proof of this theorem proceeds by induction on the construction of the formula φ. Standard constructions on NFAs allow to handle Boolean operations and classical quantification. The infinity quantifier can be reduced to existential quantification (using the synchronized rational relation $|u| \le |v|$) [3]. For the boundedness quantifier, one resorts to [15]; the Ramsey quantifier requires new automata constructions [14].

Blumensath and Grädel [3, 4] introduced the more general notion of an ω-automatic structure that is based on Büchi-automata instead of NFAs. The relation $\mathrm{Acc}^{\omega}_{\Sigma,S}$ is synchronized ω-rational such that Theorem 3.1 also holds: for SO Büchi-automata $\mathcal{A}_i$, the analogous structure is ω-automatic.

Theorem 3.2 holds for ω-automatic structures and for the extension of first-order logic with the quantifiers $\exists^{\aleph_0}$ and $\exists^{>\aleph_0}$ [4, 13], but not for the quantifier ꓤ [10]; the status of the quantifier ꓭ is not known. Consequently, whenever the following proofs use at most the existential and the cardinality quantifiers, they carry over to the case of ω-regular classes of languages.

3.2 Special representations of regular classes

Let $\mathcal{A}$ be some SO automaton and let, intuitively, $\mathcal{N}$ denote the class of NFAs represented by words $W \in L^+(\mathcal{A})$. We show that every regular class $\mathcal{C}$ of languages can be represented by some SO automaton such that $\mathcal{N}$ is a class of deterministic finite automata. Alternatively, we can require $\mathcal{N}$ to consist of all NFAs that accept some language from $\mathcal{C}$ and can be represented by some word over $\mathcal{B}(\Sigma, S)$. In the other extreme, we can require that every language from $\mathcal{C}$ is accepted by only one NFA from $\mathcal{N}$.

A block $B = (I, T, F) \in \mathcal{B}(\Sigma, S)$ is *deterministic*, i.e., belongs to $\mathrm{det}\mathcal{B}(\Sigma, S)$, if $|I| = 1$ and, for every $s \in S$ and $a \in \Sigma$, there is precisely one location $s' \in S$ with $(s, a, s') \in T$. Then any word from $\mathrm{det}\mathcal{B}(\Sigma, S)^+$ describes a DFA.

Theorem 3.3 (cf. [1, Theorem 4(4)]). *From an SO automaton $\mathcal{A}$ over Σ and S, one can construct an SO automaton $\mathcal{A}'$ over Σ and $\mathcal{P}(S)$ such that $\mathcal{L}_2(\mathcal{A}) = \mathcal{L}_2(\mathcal{A}')$ and $L^+(\mathcal{A}') \subseteq \mathrm{det}\mathcal{B}\big(\Sigma, \mathcal{P}(S)\big)^+$.*

Proof. We extend the universe of the automatic structure $\mathcal{S}(\mathcal{A})$ by the set $\mathrm{det}\mathcal{B}\big(\Sigma, \mathcal{P}(S)\big)^+$ and consider this set as an additional unary relation.

Now consider the following formula $\varphi(W')$ with free variable W':

$$W' \in \mathrm{det}\mathcal{B}\big(\Sigma, \mathcal{P}(S)\big)^+ \wedge \exists W \in L^+(\mathcal{A})\colon W^{\mathrm{Acc}_{\Sigma,S}} = W'^{\mathrm{Acc}_{\Sigma,\mathcal{P}(S)}}$$

It expresses that W' describes a DFA that accepts some language from $\mathcal{L}_2(\mathcal{A})$.

Since the structure $\mathcal{S}$ is effectively automatic, the set L_φ of words W' satisfying this formula is effectively regular, i.e., we can construct an NFA $\mathcal{A}'$ over $\mathrm{det}\mathcal{B}\big(\Sigma, \mathcal{P}(S)\big)$ with $L^+(\mathcal{A}') = L_\varphi$. Then $L_\varphi^{\mathrm{Acc}_{\Sigma,\mathcal{P}(S)}} \subseteq \mathcal{L}_2(\mathcal{A})$ by the construction of the language L_φ. For the converse inclusion, one shows that any word W over $\mathcal{B}(\Sigma, S)$ has a word $W' \in \mathrm{det}\mathcal{B}\big(\Sigma, \mathcal{P}(S)\big)^+$ with $W^{\mathrm{Acc}_{\Sigma,S}} = W'^{\mathrm{Acc}_{\Sigma,\mathcal{P}(S)}}$. The idea is to first apply the powerset construction to all blocks from W and then concatenate the resulting deterministic blocks to obtain W'. $\square$

Any word $W \in \mathcal{B}(\Sigma, S)^+$ (considered as NFA) has infinitely many equivalent words over $\mathcal{B}(\Sigma, S)$, e.g., all those from $W\,(S, T, \emptyset)^*$ (where T is an arbitrary set of transitions). Consequently, any language in the regular class of languages $\mathcal{L}_2(\mathcal{A})$ can have more than one representing word in $L^+(\mathcal{A})$. But this number of representing words can be controlled:

Theorem 3.4. *From an SO automaton $\mathcal{A}$ over Σ and S, one can construct SO automata $\mathcal{A}_{\min}$ and $\mathcal{A}_{\max}$ over Σ and S with $\mathcal{L}_2(\mathcal{A}) = \mathcal{L}_2(\mathcal{A}_{\min}) = \mathcal{L}_2(\mathcal{A}_{\max})$ such that the following hold:*

(1) Any word $W \in \mathcal{B}(\Sigma, S)^+$ with $W^{\mathrm{Acc}_{\Sigma,S}} \in \mathcal{L}_2(\mathcal{A})$ belongs to $L^+(\mathcal{A}_{\max})$.
(2) For any language $K \in \mathcal{L}_2(\mathcal{A})$, there exists a unique word $W \in L^+(\mathcal{A}_{\min})$ with $K = W^{\mathrm{Acc}_{\Sigma,S}}$.

Proof. Let $\sqsubseteq$ be a length-lexicographic order on the set $\mathcal{B}(\Sigma, S)^+$. The extension of the structure $\mathcal{S}(\mathcal{A})$ from Theorem 3.1 with $\sqsubseteq$ is automatic. The following formula $\varphi_{\max}(W)$ with free variable W expresses $W^{\mathrm{Acc}_{\Sigma,S}} \in \mathcal{L}_2(\mathcal{A})$:

$$W \in \mathcal{B}(\Sigma, S)^+ \wedge \exists W' \in L^+(\mathcal{A})\colon W^{\mathrm{Acc}_{\Sigma,S}} = W'^{\mathrm{Acc}_{\Sigma,S}}$$

Similarly, the formula $\varphi_{\min}(W)$

$$\varphi_{\max}(W) \wedge \forall W' \in \mathcal{B}(\Sigma, S)^+\colon \left(W^{\mathrm{Acc}_{\Sigma,S}} = W'^{\mathrm{Acc}_{\Sigma,S}} \to W \sqsubseteq W'\right)$$

expresses $W^{\mathrm{Acc}_{\Sigma,S}} \in \mathcal{L}_2(\mathcal{A})$ and that it is the length-lexicographically minimal representative of this language. In both cases, we can continue as in the proof of Theorem 3.3. $\square$

Remark. Since no synchronized rational well-order exists on the set $\mathcal{B}(\Sigma, S)^\omega$ [5], the above construction of $\mathcal{A}_{\min}$ does not transfer to SO Büchi-automata.

3.3 Decidable properties of regular classes

Since emptiness of regular languages is decidable, it follows from Theorem 3.2 that the FO^+-theory of every automatic structure is decidable (even if the automatic structure is part of the input). This classical result immediately gives the following from [1].

Theorem 3.5 ([1, Theorem 4(6,7)]). *For SO automata $\mathcal{A}_1$ and $\mathcal{A}_2$, inclusion and disjointness of $\mathcal{L}_2(\mathcal{A}_1)$ and $\mathcal{L}_2(\mathcal{A}_2)$ are decidable.*

Let $\mathcal{A}$ be an SO automaton. Then, by Theorem 3.4, we can construct an "unambiguous" SO automaton $\mathcal{A}_{\min}$ with $\mathcal{L}_2(\mathcal{A}) = \mathcal{L}_2(\mathcal{A}_{\min})$. Consequently, the class $\mathcal{L}_2(\mathcal{A})$ is finite iff $\mathcal{A}_{\min}$ accepts a finite language. Since this is decidable, we obtain that finiteness of $\mathcal{L}_2(\mathcal{A})$ is decidable for any SO automaton $\mathcal{A}$.

Apart from this, we can also decide further properties of the class $\mathcal{L}_2(\mathcal{A})$:

Theorem 3.6. *The following problems are decidable:*
input: an SO automaton $\mathcal{A}$ over Σ and S
question 1: Do all words over Σ belong to some language from $\mathcal{L}_2(\mathcal{A})$?
question 2: Do all $w \in \Sigma^+$ belong to only finitely many languages from $\mathcal{L}_2(\mathcal{A})$?
question 3: Do all $w \in \Sigma^+$ belong to a bounded number of languages from $\mathcal{L}_2(\mathcal{A})$?
question 4: Are the languages from $\mathcal{L}_2(\mathcal{A})$ of bounded size?

Proof. By Theorem 3.4, we can assume $\mathcal{A}$ to be "unambiguous". The formulas

1. $\forall w \in \Sigma^+ \, \exists W \in L^+(\mathcal{A})\colon (W, w) \in \mathrm{Acc}_{\Sigma,S}$
2. $\neg \exists w \in \Sigma^+ \, \exists^\infty W \in L^+(\mathcal{A})\colon (W, w) \in \mathrm{Acc}_{\Sigma,S}$
3. $\exists\!\!\!\exists (w, W)\colon (W, w) \in \mathrm{Acc}_{\Sigma,S} \wedge W \in L^+(\mathcal{A})$
4. $\exists\!\!\!\exists (W, w)\colon (W, w) \in \mathrm{Acc}_{\Sigma,S} \wedge W \in L^+(\mathcal{A})$

express the four properties such that the claims follow from Theorem 3.2. □

3.4 Closure properties of the collection of regular classes

We now strengthen some results from [1] that concern Boolean combinations of regular classes of languages. The corresponding constructions in [1] increase the number of locations exponentially. Our proofs are analogous to the proof of Theorem 3.3.

Theorem 3.7 (cf. [1, Theorem 4(1-3)]). *From SO automata $\mathcal{A}_i$ over Σ_i and S_i (for $i \in \{1, 2\}$), one can construct SO automata $\mathcal{A}'_1$ over $\Sigma_1 \cup \Sigma_2$ and $S_1 \cup S_2$ and $\mathcal{A}'_2$, $\mathcal{A}'_3$ over Σ_1 and S_1 such that $\mathcal{L}_2(\mathcal{A}'_1) = \mathcal{L}_2(\mathcal{A}_1) \cup \mathcal{L}_2(\mathcal{A}_2)$, $\mathcal{L}_2(\mathcal{A}'_2) = \mathcal{L}_2(\mathcal{A}_1) \cap \mathcal{L}_2(\mathcal{A}_2)$, and $\mathcal{L}_2(\mathcal{A}'_3) = \mathcal{L}_2(\mathcal{A}_1) \setminus \mathcal{L}_2(\mathcal{A}_2)$.*

So far, we considered, e.g., the intersection of two regular classes $\mathcal{C}_1$ and $\mathcal{C}_2$ of languages. Now, we will, e.g., consider the class of all intersections of languages in $\mathcal{C}_1$ and $\mathcal{C}_2$.

Theorem 3.8. *From SO automata $\mathcal{A}_i$ over Σ_i and S_i (for $i \in [2]$), one can construct SO automata $\mathcal{A}'_i$ such that*

1. $\mathcal{L}_2(\mathcal{A}'_1) = \{K_1 \cup K_2 \mid K_i \in \mathcal{L}_2(\mathcal{A}_i)\}$ *and $\mathcal{A}'_1$ is over $\Sigma_1 \cup \Sigma_2$ and $S_1 \uplus S_2$,*
2. $\mathcal{L}_2(\mathcal{A}'_2) = \{K_1 \cap K_2 \mid K_i \in \mathcal{L}_2(\mathcal{A}_i)\}$ *and $\mathcal{A}'_2$ is over $\Sigma_1 \cup \Sigma_2$ and $S_1 \times S_2$,*
3. $\mathcal{L}_2(\mathcal{A}'_3) = \{K_1 \setminus K_2 \mid K_i \in \mathcal{L}_2(\mathcal{A}_i)\}$ *and $\mathcal{A}'_3$ is over Σ_1 and $S_1 \times \mathcal{P}(S_2)$.*

Proof. One first proceeds analogously to the proof of Theorem 3.3. In the final step, one adapts the corresponding constructions for union, intersection, and difference of NFAs to blocks. □

Note that, for any regular class of languages $\mathcal{L}_2(\mathcal{A})$, the union $\bigcup_{L \in \mathcal{L}_2(\mathcal{A})} L$ is regular since it is the image of the regular language $L^+(\mathcal{A})$ under the rational relation $\mathrm{Acc}_{\Sigma,S}$. Using automatic structures, we can show that also the *limit inferior* and the *limit superior* is effectively regular since both these languages can be defined (using the quantifier $\exists^\infty$) in the automatic structure $\mathcal{S}(\mathcal{A}_{\min})$ from Theorem 3.1 (where $\mathcal{A}_{\min}$ is the "unambiguous" automaton from Theorem 3.4).

Theorem 3.9. *From an SO automaton $\mathcal{A}$ over Σ and S, one can construct NFAs accepting the languages*

$$\liminf \mathcal{L}_2(\mathcal{A}) = \bigcup_{\substack{\mathcal{C} \subseteq \mathcal{L}_2(\mathcal{A}) \\ \text{finite}}} \; \bigcap_{K \in \mathcal{L}_2(\mathcal{A}) \setminus \mathcal{C}} K \quad \text{and} \quad \limsup \mathcal{L}_2(\mathcal{A}) = \bigcap_{\substack{\mathcal{C} \subseteq \mathcal{L}_2(\mathcal{A}) \\ \text{finite}}} \; \bigcup_{K \in \mathcal{L}_2(\mathcal{A}) \setminus \mathcal{C}} K\,.$$

4 Expressiveness of second-order finite automata

In this section, we determine what classes of languages can be described by SO automata, i.e., are regular. We obtain a close relation to so-called automatic classes of languages as defined by Jain *et al.* in [8].

Definition. A class of languages $\mathcal{C} \subseteq \mathcal{P}(\Sigma^+)$ is *automatic* if there are a regular language $L \subseteq \Gamma^+$ over some alphabet Γ and a synchronized rational relation $R \subseteq \Gamma^+ \times \Sigma^+$ with $\mathcal{C} = L^R$.

Example (from [8]). For any alphabet Σ, the following classes $\mathcal{C} \subseteq \mathcal{P}(\Sigma^+)$ are automatic:

- The class of finite languages with at most k elements (for any $k \in \mathbb{N}$).
- The class of all finite and cofinite subsets of $\{a\}^+$.
- The class of all intervals of $(\Sigma^+, \leq)$ where $\leq$ is the lexicographic order.
- Let U be the universe of any automatic structure $\mathcal{S}$ and let $\varphi(x, y)$ be any formula from FO^+. For $w \in U$, $\mathcal{S}^{\varphi(x,w)} \subseteq U$ is a language. The class of all these languages $\mathcal{S}^{\varphi(x,w)}$ with $w \in U$ is automatic.

4.1 Regular and automatic classes of languages

From the very definition, we obtain that every regular class $\mathcal{C}$ of languages is effectively automatic and contains only finite languages.

For the converse implication, one first shows that automatic classes of finite languages can be represented by length-reducing rational relations:

Lemma 4.1. *Let $L \subseteq \Gamma^+$ be regular and $R \subseteq \Gamma^+ \times \Sigma^+$ be a synchronized rational relation with W^R finite for all $W \in L$. There effectively exist an alphabet Γ', a regular language $L' \subseteq \Gamma'^+$, and a synchronized rational and length-reducing relation $R' \subseteq \Gamma'^+ \times \Sigma^+$ such that $L'^{R'} = \mathcal{C} \setminus \{\emptyset\}$.*

If R is length-increasing and all languages in L^R are single-length, then R' can be chosen length-preserving.

Proof. The relation R' consists of all pairs $(W\,\$^n, w)$ with $(W, w) \in R$ such that $|W| + n$ is the maximal length of words from $W^R \cup \{W\}$. □

Then one proves that, indeed, any length-reducing synchronized rational relation R gives rise to a regular class of languages:

Proposition 4.2. *Let $L \subseteq \Gamma^+$ be regular and $R \subseteq \Gamma^+ \times \Sigma^+$ be length-reducing and synchronized rational. Then L^R is, effectively, a regular class of languages.*

Proof. One starts with a synchronous 2-head automaton M accepting R. For any input letter $A \in \Gamma$, one restricts the behavior of M to its output behavior when A is input. In addition, depending on the remaining input word V, one defines a state to be accepting if, from that state, M can read V with empty output. This defines a block $B_{A,V} \in \mathcal{B}(\Sigma, S)$ as well as a sequence of blocks for every input word W). One then obtains an automaton $\mathcal{A}$ that accepts the set of block sequences for all valid input words. Then $L^R = \mathcal{L}_2(\mathcal{A})$. □

The following theorem summarises the work reported in this section.

Theorem 4.3. *The following are effectively equivalent for any class $\mathcal{C}$ of ε-free languages:*

(a) $\mathcal{C}$ is a regular class of languages.
(b) $\mathcal{C}$ is an automatic class of finite languages.
(c) $\mathcal{C} = L^R$ for some regular language L and some length-reducing synchronized rational relation R.

4.2 Regular and automatic classes of single-length languages

In this section, we want to characterise, similarly to Theorem 4.3, the full-length regular classes of languages, i.e., the language classes considered in [1].

Theorem 4.4. *The following are effectively equivalent for any class $\mathcal{C}$ of ε-free languages:*

(a) $\mathcal{C}$ is a full-length regular class of languages.
(b) $\mathcal{C}$ is a regular class of single-length languages.
(c) $\mathcal{C}$ is an automatic class of single-length languages.
(d) $\mathcal{C} = L^R$ for some regular language L and some length-preserving synchronized rational relation R.

The implication (a)$\Rightarrow$(b) is clear by the definition of full-length words, the implication (b)$\Rightarrow$(c) is an immediate consequence of Theorem 4.3. The implication (d)$\Rightarrow$(a) can be shown as Proposition 4.2. For the remaining implication (c) $\Rightarrow$ (d), one first splits R into its length-increasing and its length-reducing parts $R_{\leq}$ and $R_{\geq}$. Since L^R is a class of single-length languages, it equals $L^{R_{\leq}} \cup L^{R_{\geq}}$. The final claim of Lemma 4.1 allows to replace $R_{\leq}$ by some length-preserving relation. For the length-reducing part $R_{\geq}$, one then proves a slightly weaker fact:

Lemma 4.5. *Let $L \subseteq \Gamma^+$ be regular and $R \subseteq \Gamma^+ \times \Sigma^+$ be synchronized rational and length-reducing such that L^R is a class of single-length languages. Then there exist, effectively, $k \in \mathbb{N}$, regular languages $L_1, \ldots, L_k \subseteq \Gamma^+$, and synchronized rational length-preserving relations $R_1, \ldots, R_k \subseteq \Gamma^+ \times \Sigma^+$ with $\bigcup_{1 \leq i \leq k} {L_i}^{R_i} = L^R \setminus \{\emptyset\}$.*

Proof. Since L is regular, we can assume $R \subseteq L \times \Sigma^+$. Let M be some synchronous 2-head automaton accepting R. For a set X of states, we define relations $R_X, S_X \subseteq \Gamma^* \times \Sigma^*$ as follows: R_X is the set of pairs (W_1, w_1) of nonempty words of equal length such that (W_1, w_1) allows to reach (from some initial state) *some* state in X. Further, S_X is the set of pairs (W_2, w_2) such that X *equals* the set of states that allow to reach some accepting state via (W_2, w_2).

The crucial point is that for $(W_1, w) \in R_X$ and $(W_2, \varepsilon) \in S_X$, one has $\emptyset \neq W_1^{R_X} = (W_1 W_2)^R$ and $W_1 W_2 \in L$. It can be infered that $L^R \setminus \{\emptyset\}$ is the union of the classes $\mathrm{proj}_1(R_X)^{R_X}$ for X such that $S_X \cap \big(\Gamma^* \times \{\varepsilon\}\big) \neq \emptyset$. $\square$

It follows that any automatic class of single-length languages is the union of finitely many classes L^R with L regular and R length-preserving. Considering copies of the languages L over mutually disjoint alphabets allows to infer the missing implication (c) $\Rightarrow$ (d) in Theorem 4.4

4.3 ω-regular and automatic classes of languages

An automatic class of languages is always *countable* and consists of *regular languages*, only. By Example 2.2, both these properties may fail for ω-regular classes. The main result of this section states that these are the two only (and equivalent) reasons for an ω-regular class not to be automatic.

Theorem 4.6. *The following are effectively equivalent for any class $\mathcal{C}$ of ε-free languages:*

(a) $\mathcal{C}$ is an ω-regular class of regular *languages.*
(b) $\mathcal{C}$ is a countable *ω-regular class of languages.*
(c) $\mathcal{C}$ is an automatic class of languages.

The implication (a)$\Rightarrow$(b) is trivial since there are only countably many regular languages. The proof of the implication (c)$\Rightarrow$(a) is based on the idea of Example 2.1. For the implication (b)$\Rightarrow$(c), one considers the ω-automatic structure $\big(\mathcal{B}(\Sigma,S)^\omega \cup \Sigma^+, L^\omega(\mathcal{A}), \mathcal{B}(\Sigma,S)^\omega, \mathrm{Acc}^\omega_{\Sigma,S}\big)$. Identifying pairs of ω-words over $\mathcal{B}(\Sigma,S)$ that represent the same language over Σ gives rise to a countable quotient that is ω-automatic [2] and therefore automatically representable [3]. This automatic structure then allows to prove that the class is automatic.

5 Regular classes of languages, ordered by inclusion

In this final section, we consider regular classes $\mathcal{L}_2(\mathcal{A})$ of languages under inclusion, i.e., the structure $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$. Note that the universe of $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ is not a language, but a class of languages. Hence this structure cannot be automatic. The first result shows that $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ is effectively isomorphic to some automatic structure, i.e., is *automatically representable.*

Lemma 5.1. *Let $\mathcal{A}$ be some SO automaton over Σ and S. Then $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ is effectively automatically representable.*

Now we have, again, the theory of automatic structures at our disposal. In particular, Theorem 3.2 allows to infer the following decidabilities.

Theorem 5.2. *The following problems are decidable:*
input: an SO automaton $\mathcal{A}$
question 1: Is $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ a lattice?
question 2: Does $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ contain some infinite antichain or some infinite chain, resp.?
question 3: Are intervals in $\big(\mathcal{L}_2(\mathcal{A}),\subseteq\big)$ of bounded finite size?

By Lemma 5.1, regular classes of languages (ordered by inclusion) can be understood as automatic partial order. By the theorems from Section 4, one can conversely understand automatic partial orders as regular classes of languages. This allows to infer results concerning the isomorphism problem from [12].

Theorem 5.3. *There exist partial orders $\mathcal{P}_1$ and $\mathcal{P}_2$ such that the set of*

1. *SO automata $\mathcal{A}$ with $\mathcal{P}_1 \cong \big(\mathcal{L}_2(\mathcal{A}), \subseteq\big)$ is Σ^1_1-hard and*
2. *full-length SO automata $\mathcal{A}$ with $\mathcal{P}_2 \cong \big(\mathcal{L}_2(\mathcal{A}), \subseteq\big)$ is Π^0_1-hard.*

In particular, the isomorphism problem for structures $\big(\mathcal{L}_2(\mathcal{A}), \subseteq\big)$ is (highly) undecidable.

References

1. Andrade de Melo, A., de Oliveira Oliveira, M.: Second-order finite automata. In: CSR'20. pp. 46–63. Lecture Notes in Comp. Science vol. 12159, Springer (2020)
2. Bárány, V., Kaiser, Ł., Rubin, S.: Cardinality and counting quantifiers on omega-automatic structures. In: STACS'08. pp. 385–396. IFIB Schloss Dagstuhl (2008)
3. Blumensath, A.: Automatic structures (1999), Diplomarbeit, RWTH Aachen
4. Blumensath, A., Grädel, E.: Automatic Structures. In: LICS'00. pp. 51–62. IEEE Computer Society Press (2000)
5. Carayol, A., Löding, C.: MSO on the infinite binary tree: Choice and order. In: CSL'07. Lecture Notes in Comp. Science vol. 4646, vol. 4646, pp. 161–176. Springer (2007)
6. Frougny, C., Sakarovitch, J.: Synchronized rational relations of finite and infinite words. Theoretical Computer Science **108**, 45–82 (1993)
7. Hodgson, B.: On direct products of automaton decidable theories. Theoretical Computer Science **19**, 331–335 (1982)
8. Jain, S., Luo, Q., Stephan, F.: Learnability of automatic classes. J. Comput. Syst. Sci. **78**(6), 1910–1927 (2012)
9. Khoussainov, B., Nerode, A.: Automatic presentations of structures. In: Logic and Computational Complexity. pp. 367–392. Lecture Notes in Comp. Science vol. 960, Springer (1995)
10. Kuske, D.: Is Ramsey's theorem ω-automatic? In: STACS 2010. pp. 537–548. Leibniz International Proceedings in Informatics (LIPIcs) vol. 5, Schloss Dagstuhl–Leibniz-Zentrum für Informatik (2010)
11. Kuske, D.: Where automatic structures benefit from weighted automata. In: Algebraic Foundations of Computer Science (Bozapalidis Festschrift). pp. 257–271. Lecture Notes in Comp. Science vol. 7020, Springer (2011)
12. Kuske, D., Liu, J., Lohrey, M.: The isomorphism problem on classes of automatic structures with transitive relations. Transactions of the AMS **365**, 5103–5151 (2013)
13. Kuske, D., Lohrey, M.: First-order and counting theories of ω-automatic structures. In: FoSSaCS 2006. pp. 322–336. Lecture Notes in Comp. Science vol. 3921, Springer (2006)
14. Rubin, S.: Automata presenting structures: A survey of the finite string case. Bulletin of Symbolic Logic **14**, 169–209 (2008)
15. Weber, A.: On the valuedness of finite transducers. Acta Informatica **27**, 749–780 (1990)