AUTOMATIC STRUCTURES AND THE THEORY OF LISTS

N.A. BAZHENOV

Abstract. Goncharov constructed the axiomatic theory of linear lists over the elements of a given data type. We study algorithmic complexity for models of this theory. We prove that the enriched list structure over a finite set of atoms is automatically presentable, and the enriched list structure over an infinite set of atoms has no automatic presentations.

Keywords: automatic structure, linear list, theory of lists, decidable model, list superstructure.

1. Introduction

Moore and Russell [1] introduced the axiomatic specification of linear lists. Their formal theory of lists was given as a first-order two-sorted theory. Goncharov [2] constructed the generalization of the theory from [1]. He built the axiomatic theory of lists over the elements of a given abstract data type. Moreover, Goncharov proved that the theory of lists over models of a decidable theory is decidable. In particular, this result implies that the list structure over a finite set of atoms is decidable.

In this short note, we study algorithmic complexity for models of the theory of lists. In our study, we use the framework of automatic structures. One of the main motivations for choosing this framework is that automatic structures are uniformly decidable (i.e. their complete diagrams can be decided uniformly in an automatic presentation and a first-order formula).

The idea of using automata to study structures goes back to Büchi [3, 4] and Rabin [5] who used automata to prove the decidability of some weak second-order theories. The systematic study of automatic structures was initiated by Khoussainov and Nerode [6]. In recent years, the theory of automatic structures has become
increasingly popular. For further bibliography and background on automatic structures, we refer the reader to the surveys [7, 8, 9] and the recent papers [10, 11].

In this paper, we prove that the enriched list structure over a finite set of atoms is automatically presentable, and the enriched list structure over an infinite set of atoms has no automatic presentations.

2. Preliminaries

Let $\sigma$ be a computable signature. A computable structure $M$ of the signature $\sigma$ is decidable if its complete diagram $D^c(M)$ is computable. We refer the reader to [12, 13] for further background on computable and decidable structures.

Following [1, 2], we define the first-order two-sorted theories $LL$ and $LL(T)$, where $T$ is a first-order theory of a signature $\sigma$.

The language $L_\sigma$ contains the following symbols:

- parentheses: $(, )$
- comma: ,
- logical connectives: $\&$, $\lor$, $\neg$, $\rightarrow$
- variables of two sorts: $b, b_0, b_1, \ldots$ have the sort $\text{atom}$, and $x, x_0, x_1, \ldots$ have the sort $\text{list}$
- equality symbol: $=$
- quantifier symbols: $\forall, \exists$
- constant symbol $\text{nil}$ of sort $\text{list}$
- function symbol $\text{cons}$: $\text{list} \times \text{atom} \rightarrow \text{list}$
- all the symbols from $\sigma$.

Suppose that $\beta \in \{\text{atom, list}\}$. Terms of sort $\beta$ are defined inductively:

1. Any variable of sort $\beta$ or constant symbol of sort $\beta$ is a term of sort $\beta$.
2. If $t$ is a term of sort $\text{list}$ and $q$ is a term of sort $\text{atom}$, then $\text{cons}(t, q)$ is a term of sort $\text{list}$.
3. Suppose that $f$ is an $n$-ary function symbol from $\sigma$, and $t_1, \ldots, t_n$ are terms of sort $\text{atom}$. Then $f(t_1, \ldots, t_n)$ is a term of sort $\text{atom}$.
4. There are no other terms.

$L_\sigma$-formulas are also defined inductively:

1. If $t$ and $q$ are terms of sort $\beta$, then $(t = q)$ is an $L_\sigma$-formula.
2. Suppose that $P$ is an $m$-ary predicate symbol from $\sigma$, and $t_1, \ldots, t_m$ are terms of sort $\text{atom}$. Then $P(t_1, \ldots, t_m)$ is an $L_\sigma$-formula.
3. Suppose that $\phi$ and $\psi$ are $L_\sigma$-formulas, and $v$ is a variable. Then $(\phi \lor \psi)$, $(\phi \land \psi)$, $\neg \phi$, $(\phi \rightarrow \psi)$, $\forall v \phi$, and $\exists v \phi$ are $L_\sigma$-formulas.
4. There are no other $L_\sigma$-formulas.

An $L_\sigma$-formula $\phi$ is an $L_\sigma$-formula of sort $\beta$ if all the variables and constants from $\phi$ have the sort $\beta$. A set $T$ of $L_\sigma$-sentences is a theory of sort $\beta$ if every formula from $T$ is a formula of sort $\beta$. Logical axioms and inference rules for $L_\sigma$-formulas are the same as for standard first-order formulas (see, e.g., [14, §18]).

Suppose that $\phi$ is an $L_\sigma$-formula, $v$ is a variable, and $t$ is a term such that for any variable $y$ from $t$, the formula $\phi$ does not contain quantifiers $\exists y$ and $\forall y$. We denote by $\phi^v_t$ the formula $\phi$ with all free occurrences of the variable $v$ replaced by the term $t$.

**Definition 1** ([1, 2]). The theory $LL$ is defined by the following axioms:
function symbols, predicate symbols, all the symbols from \( L \) are axiomatized by the three axioms above and all the formulas from \( T \).

Suppose that \( T \) is a theory of sort atom. The theory \( LL(T) \) is the theory that is axiomatized by the three axioms above and all the formulas from \( T \).

Goncharov [2, Proposition 3] proved that for a complete theory \( T \), the theory \( LL(T) \) is also complete.

For a signature \( \sigma \), the language \( \tilde{L}_\sigma \) contains the following symbols:

- all the symbols from \( L_\sigma \),
- function symbols \( head : list \to atom \cup list \) and \( tail : list \to list \),
- predicate symbols \( \subseteq \subseteq list \times list \) and \( \in \subseteq atom \times list \).

**Definition 2.** Suppose that \( \mathcal{M} \) is a structure of a signature \( \sigma \), and \( \Lambda \notin \mathcal{M} \). The list structure over the structure \( \mathcal{M} \) is a two-sorted structure \( LS(\mathcal{M}) \) of the language \( L_\sigma \) such that

\[
\text{atom}^{LS(\mathcal{M})} = [\mathcal{M}],
\]
\[
\text{list}^{LS(\mathcal{M})} = \{\Lambda\} \cup \{(a_0, \ldots, a_n) : n \in \omega, a_i \in \mathcal{M}\},
\]
\[
\sigma^{LS(\mathcal{M})} = \sigma^\mathcal{M},
\]
\[
\text{nil}^{LS(\mathcal{M})} = \Lambda,
\]
\[
\text{cons}^{LS(\mathcal{M})}(\Lambda, a) = \langle a \rangle, \ a \in \mathcal{M},
\]
\[
\text{cons}^{LS(\mathcal{M})}(\langle a_0, \ldots, a_n \rangle, b) = \langle a_0, \ldots, a_n, b \rangle, \ a_0, \ldots, a_n \in \mathcal{M}.
\]

The enriched list structure over the structure \( \mathcal{M} \) is a two-sorted structure \( ELS(\mathcal{M}) \) of the language \( \tilde{L}_\sigma \) such that

\[
ELS(\mathcal{M}) | L_\sigma = LS(\mathcal{M});
\]
\[
\text{head}^{ELS(\mathcal{M})}(\Lambda) = \Lambda,
\]
\[
\text{head}^{ELS(\mathcal{M})}(\langle a_0, \ldots, a_n \rangle) = a_n;
\]
\[
\text{tail}^{ELS(\mathcal{M})}(\Lambda) = \text{tail}^{ELS(\mathcal{M})}(\langle a \rangle) = \Lambda,
\]
\[
\text{tail}^{ELS(\mathcal{M})}(\langle a_0, \ldots, a_n, a_{n+1} \rangle) = \langle a_0, \ldots, a_n \rangle;
\]
\[
(x, y) \in ELS(\mathcal{M}) \iff x, y \in \text{list}^{ELS(\mathcal{M})} \text{ and the list } x \text{ is an initial segment of the list } y;
\]
\[
(a, x) \in ELS(\mathcal{M}) \iff \text{there exist } a_0, \ldots, a_n \in \text{atom}^{ELS(\mathcal{M})} \text{ and } m \leq n \text{ such that } x = \langle a_0, \ldots, a_n \rangle \text{ and } a = a_m.
\]

If \( A \) is a set, one can treat \( A \) as a structure of the empty signature. In that case, the structure \( LS(A) \) (\( ELS(A) \)) is called the (enriched) list structure over the set of atoms \( A \).

Definition 2 is motivated by the following considerations. One can think of the structure \( LS(\mathcal{M}) \) as a “natural” model of the theory \( LL \). This is justified by the result below.
Suppose that a structure \( M \) is a model of a theory \( T \) of sort \( \text{atom} \). Then \( LS(M) \) is a model of the theory \( LL(T) \).

The enriched list structure \( ELS(M) \) is a substructure of the (hereditarily finite) list superstructure \( HWM(M) \). List superstructures play important role in theoretical computer science [15, 16, 17] and the theory of generalized computability, especially computability over the field of real numbers [18, 19, 20].

We refer the reader to [21] for basic terminology and notations from automata theory. Suppose that \( \Sigma \) is a finite alphabet. Let \( \Sigma_0 = \Sigma \cup \{\diamond\} \), where \( \diamond \notin \Sigma \). As usual, \( \Sigma^* \) denotes the set of all words of the alphabet \( \Sigma \). For \( w \in \Sigma^* \), we denote by \( |w| \) the length of the word \( w \).

A (nondeterministic) finite automaton is a tuple \( A = (S, \Delta, i, F) \), where \( S \) is the finite set of states, \( \Delta \subseteq S \times \Sigma \times S \) is the transition relation, \( i \in S \) is the initial state, and \( F \subseteq S \) is the set of accepting states. The language \( L(A) \) accepted by the automaton \( A \) is defined in a standard way (see, e.g., [21, § 2.3]). A language \( L \subseteq \Sigma^* \) is regular if \( L = L(A) \) for some automaton \( A \).

For words \( w_1, \ldots, w_n \in \Sigma^* \), their convolution \( \otimes (w_1, \ldots, w_n) \) is a word from \( (\Sigma^*)^* \) of length \( max\{|w_1|, \ldots, |w_n|\} \), and the \( k \)-th symbol of \( \otimes (w_1, \ldots, w_n) \) is \( (a_1, \ldots, a_n) \), where \( a_k \) is the \( k \)-th symbol of \( w_i \) if \( k \leq |w_i| \), and \( a_i = \diamond \) otherwise. For an \( n \)-ary relation \( R \subseteq (\Sigma^*)^n \), \( R^\otimes \) denotes the set of convolutions of tuples from \( R \). A relation \( R \) is FA recognizable if \( R^\otimes \) is a regular language.

Let \( \sigma \) be a finite predicate signature. A structure \( M \) of the signature \( \sigma \) is automatic over \( \Sigma \) if the universe of \( M \) is a regular subset of \( \Sigma^* \) and each of the basic predicates of \( M \) is FA recognizable. We also consider automatic structures for an arbitrary finite signature (proceeding, in a standard way, from the basic functions to their graphs, if necessary). A structure \( M \) is automatically presentable if there exists an automatic structure \( A \) isomorphic to \( M \).

For many familiar classes \( K \), there are nice structure theorems that give a characterization of automatically presentable members of \( K \). Blumensath and Grädel [22] proved that a structure is automatically presentable if and only if it is first-order definable in the structure \( (\{0, 1\}^*, \preceq, \text{Left}, \text{Right}, \text{EqL}) \), where \( \preceq \) is the prefix relation, \( \text{Left}(\tau) = \tau 0 \), \( \text{Right}(\tau) = \tau 1 \), and \( \text{EqL} \) is the equal length relation. Delhomme [23] proved that an ordinal \( \alpha \) is automatically presentable if and only if \( \alpha < \omega^\omega \). Khoussainov, Nies, Rubin, and Stephan [24] showed that an infinite Boolean algebra is automatically presentable if and only if it is isomorphic to \( B_n^\omega \) for some \( n \geq 1 \), where \( B_\omega \) is the Boolean algebra of finite and co-finite subsets of \( \omega \).

Suppose that \( 0 \leq t < k \) and \( k \geq 2 \). The quantifier \( \exists^{(t,k)} \) is defined as follows: for a structure \( M \) and a formula \( \psi(x) \), \( M \models (\exists^{(t,k)}x) \psi(x) \) if and only if the set \( \psi^{\exists^t} = \{ a : M \models \psi(a) \} \) is finite and \( t = |\psi^{\exists^t}| \mod k \) (i.e., the cardinality of \( \psi^{\exists^t} \) is equal to \( kz + t \) for some integer \( z \)).

We denote by \( FO_\sigma(\exists^{(t,k)}) \) the set of all formulas of the signature \( \sigma \) in the first-order logic extended by the quantifiers \( \exists^{(t,k)} \), \( t, k \in \omega \).

Khoussainov and Nerode [6] proved that every automatic structure is decidable. We will use the following generalization of their result.

**Proposition 2** (Khoussainov, Rubin, and Stephan [25]). Suppose that \( M \) is an automatic structure of a signature \( \sigma \). There is an algorithm that, given a formula \( \psi(\bar{x}) \) from \( FO_\sigma(\exists^{(t,k)}) \), produces a finite automaton recognizing the set \( (\psi^{\exists^t})^\otimes \).
3. Results

Proposition 3. Suppose that $A = \{a_0, \ldots, a_n\}$ is a finite set. Then the enriched list structure $ELS(A)$ is automatically presentable.

Proof. We define the finite alphabet $\Sigma = A \cup \{\langle, \rangle\}$. If $f$ is a function, then $\Gamma_f$ denotes the graph of $f$.

The automatic copy $\mathfrak{A}$ of the structure $ELS(A)$ is defined as follows:

- $\text{atom}^\mathfrak{A} = A$;
- $\text{list}^\mathfrak{A} = \langle A^* \rangle$;
- $\text{nil}^\mathfrak{A} = \langle \rangle$;
- $(\Gamma_{\text{cons}})^\mathfrak{A} = \bigcup_{i=0}^{n} \langle a_i \rangle \left( \bigcup_{j=0}^{n} (a_j \diamond a_i) \right)^* \langle \rangle$;
- $(\Gamma_{\text{head}})^\mathfrak{A} = \bigcup_{i=0}^{n} \langle a_i \rangle \left( \bigcup_{j=0}^{n} (a_j \diamond a_i) \right)^* \langle a_i \rangle$;
- $(\Gamma_{\text{tail}})^\mathfrak{A} = \bigcup_{i=0}^{n} \langle a_i \rangle \left( \bigcup_{j=0}^{n} (a_j \diamond a_i) \right)^* \langle \rangle$

It is straightforward to prove that the languages above are regular, and the structure $\mathfrak{A}$ is isomorphic to $ELS(A)$. 

For $n \geq 1$, we denote by $\sigma^n$ the signature $\{a_0, \ldots, a_{n-1}\}$, where $a_i$ is a constant symbol of sort $\text{atom}$. Let $LL_n = LL(T_n)$, where

$T_n = \{a_i \neq a_j : i < j < n\} \cup \{\neg \exists b (b \neq a_0 \& \ldots \& b \neq a_{n-1})\}$.

Goncharov [2] showed that the theory $LL_n$ is complete. Note that Propositions 2 and 3 give the alternative proof of the following result:

Corollary 1 (Goncharov [2]). The theory $LL_n$ is decidable.

The rest of the section deals with structures that have no automatic presentations.

Theorem 1. Let $A$ be a countably infinite set. Then the enriched list structure $ELS(A)$ has no automatic copies.

Proof. We show that the existence of an automatic copy of $ELS(A)$ implies the existence of an automatic copy of the ordinal $\omega^\omega$. This contradicts the characterization of automatically presentable ordinals [23].

Suppose that $\mathfrak{A}$ is an automatic copy of $ELS(A)$. We define the automatic copy $\mathfrak{B}$ of the structure $ELS(\omega, \leq)$ as follows. For a symbol $s \in \{\text{atom}, \text{list}, \text{nil}, \text{cons}, \text{head}, \ldots\}$
If $x \preceq y$, then $x$ is lexicographically less than or equal to $y$. This is defined by the formula

$$x \preceq y \iff (\exists \phi \in \varphi^a \phi(x) \in \varphi^a \phi(y) \in \varphi^a \phi(x) \neq \phi(y) \vee \exists \psi \in \varphi^a \psi(x) \in \varphi^a \psi(y) \in \varphi^a \psi(x) = \psi(y)) \vee \exists \tau \in \varphi^a \tau(x) \in \varphi^a \tau(y) \in \varphi^a \tau(x) \neq \tau(y)).$$

We describe the idea behind the construction of the automatic copy $L$ of the ordinal $\omega^\omega$. Suppose that $F$ is an isomorphism from $(\omega, \preceq)$ onto $(\varphi^a, \preceq_a)$. For a non-zero ordinal $\alpha < \omega^\omega$, take the Cantor normal form of $\alpha$:

$$\alpha = \omega^{n_0} \cdot k_0 + \omega^{n_1} \cdot k_1 + \ldots + \omega^{n_t} \cdot k_t,$$

where $\omega > n_0 > n_1 > \ldots > n_t \geq 0$ and $0 < k_i < \omega$ for all $i$. Then the ordinal $\alpha$ corresponds to the list $(F(n_0), F(k_0), F(n_1), F(k_1), \ldots, F(n_t), F(k_t))$ from the structure $B$.

First, we define auxiliary formulas $l_0(x)$ and $l_1(x)$.

$$l_0(x) := \exists \{1, 2\} y (y \subseteq x),$$

$$l_1(x) := \neg l_0(x),$$

where $x$ is a variable of sort $\text{list}$. Note that $B \models l_0(a)$ if $a$ is a list.

The universe of the structure $L$ is given by the formula

$$D(x) = l_0(x) \land \forall y \forall z ((y \subseteq z \subseteq x \land l_1(y) \land l_1(z)) \rightarrow \text{head}(z) < \text{head}(y)) \land \forall y ((y \subseteq x \land l_0(y) \land y \neq \text{nil}) \rightarrow \exists b (b < \text{head}(y))).$$

The order $\preceq$ is defined by the formula

$$(x \preceq y) \equiv D(x) \land D(y) \land [(x \subseteq y) \lor \exists z \exists b (a < b \land \text{cons}(z, a) \subseteq x \land \text{cons}(z, b) \subseteq y)].$$

It is not difficult to show that the structure $L = (D^a, \preceq_a)$ is a linear order isomorphic to the ordinal $\omega^\omega$. On the other hand, Proposition 2 implies that the order $L$ is automatic. This is a contradiction. Therefore, the structure $ELS(A)$ is not automatically presentable.

**Corollary 2.** If $M$ is an infinite structure, then the structure $ELS(M)$ has no automatic copies.

**Corollary 3.** Suppose that $M$ is a structure. Then the list superstructure $\mathbb{H}(M)$ is not automatically presentable.

**Proof.** Suppose that $\mathbb{H}(M)$ has an automatic presentation. One can define a copy of the structure $ELS(\omega, \preceq)$ in $\mathbb{H}(M)$ using first-order formulas:

$$\text{atom}_1(x) := (x \neq \text{nil}) \land (\forall y \in x)(y = \text{nil}),$$

$$\text{list}_1(x) := (\forall y \in x)\text{atom}_1(y),$$

$$\text{head}_1(x) = a \iff \text{cons}_1(x, a) = \text{list}_1(x) \land \text{atom}_1(a) \land (\text{cons}(x, a) = y),$$

$$\text{tail}_1(x) = y \iff \text{cons}_1(x, a) = \text{list}_1(x) \land (\text{head}(x) = a),$$

$$\text{nil}_1 = \text{nil},$$

$$x \subseteq 1 \iff \text{list}_1(y) \land (x \subseteq y),$$

$$x \subseteq 1 \iff \text{list}_1(y) \land (x \subseteq y),$$

$$a \in 1 = \text{list}_1(x) \land (a \in x).$$
Hence, \( EL S(\omega, \leq) \) is automatically presentable. This and the proof of Theorem 1 yield a contradiction.

We also give an example of an automatic structure \( \mathfrak{A} \) such that the list structure \( LS(\mathfrak{A}) \) is not automatically presentable. It is well-known (see, e.g., [9, Example 8]) that the commutative semigroup \((\omega, +)\) has an automatic presentation.

**Theorem 2.** The structure \( LS(\omega, +) \) has no automatic copies.

**Proof.** Recall the following useful result.

**Lemma 1** (Khoussainov and Nerode [6, Corollary 4.5]; Blumensath and Grädel [22, Proposition 4.4]). Suppose that \( \mathfrak{A} = (A, f_0^{a_0}, f_1^{a_1}, \ldots, f_k^{a_k}) \) is an automatic structure, and \( X \) is a finite subset of \( A \). For \( n \in \omega \), the set \( G^n_\mathfrak{A}(X) \) is defined as follows.

\[
G^0_\mathfrak{A}(X) = X, \\
G^n_{\mathfrak{A}+1}(X) = G^n_\mathfrak{A}(X) \cup \{f_i(a_1, \ldots, a_m) : i \leq k, a_j \in G^n_\mathfrak{A}(x)\}. 
\]

Then there exists a constant \( C \in \omega \) such that for any \( n \), \( |G^n_\mathfrak{A}(X)| \leq 2^{C(n+1)} \).

Suppose that \( \mathfrak{A} \) is an automatic copy of the structure \( LS(\omega, +) \). For the sake of convenience, we assume that \( \text{atom}^\mathfrak{A} = \omega \). Define the set

\[
X = \{0, 1, \text{null}^\mathfrak{A}\}.
\]

For \( n \in \omega \), \( G_n \) denotes the set \( G^n_\mathfrak{A}(X) \).

**Lemma 2.** For \( n \geq 1 \),

\[
|G_n| \geq 1 + 2^n + 3 \cdot \prod_{i=0}^{n-2} (1 + 2^i).
\]

(Here we assume that \( \prod_{i=0}^{-1} (1 + 2^i) \) is equal to 1.)

**Proof.** Let \( G^0_n = G_n \cap \text{atom}^\mathfrak{A} \), and \( G^1_n = G_n \cap \text{list}^\mathfrak{A} \). It is easy to show, by induction on \( n \), that \( G^0_n = \{a \in \omega : a \leq 2^n\} \) for any \( n \). Therefore, \( |G^0_n| = 1 + 2^n \).

We prove that for any \( n \geq 1 \), \(|G^1_n| \geq 3 \cdot \prod_{i=0}^{n-2} (1 + 2^i) \). It is obvious that \( |G^1_1| = 3 \).

Suppose that \( n \geq 2 \) and our claim is true for every non-zero \( k < n \). Note that \( G^1_n \supseteq G^1_{n-1} \cup \{\text{cons}(z, a) : z \in G^0_{n-1}, a \in G^0_{n-1} \setminus G^0_{n-2}\} \).

Therefore,

\[
|G^1_n| \geq |G^1_{n-1}| \cdot (1 + |G^0_{n-1}| - |G^0_{n-2}|) = |G^1_{n-1}| \cdot (1 + 2^{n-2}) \geq \prod_{i=0}^{n-3} (1 + 2^i) \cdot 3 \prod_{i=0}^{n-2} (1 + 2^i) = 3 \prod_{i=0}^{n-2} (1 + 2^i).
\]

Lemma 2 implies that for any \( n \geq 6 \),

\[
|G_n| \geq \prod_{i=0}^{n-2} (1 + 2^i) \geq 3^{n-2} = 2^{(n-1)(n-2)/2} > 2^{n^2/4}.
\]

On the other hand, by Lemma 1, there exists a constant \( C \in \omega \) such that \( n^2 < 4C(n+1) \) for any \( n \geq 6 \), a contradiction. Therefore, \( LS(\omega, +) \) is not automatically presentable. 

In conclusion, we formulate an open question related to our results.

**Question.** Suppose that $A$ is a countably infinite set. We know that the enriched list structure $\text{ELS}(A)$ has no automatic copies. Is the list structure $\text{LS}(A)$ automatically presentable?

**Acknowledgements.** The author is grateful to Sergey Goncharov for fruitful discussions on the subject.

**References**


Nikolay Alekseyevich Bazhenov
Sobolev Institute of Mathematics,
pr. Koptyuga, 4,
630090, Novosibirsk, Russia
Institute of Mathematics and Mechanics, Kazan Federal University,
Kremlevskaya ul., 35,
420008, Kazan, Russia
E-mail address: bazhenov@math.nsc.ru