Languages, automata and computation II
Tutorial 4

Winter semester 2023/2024

In this tutorial we study automata theory in sets with atoms.

Sets with atoms

Exercise 1. Consider an orbit-finite set X and an equivariant relation $R \subseteq X \times X$. For every $n \in \mathbb{N}$, let $R_n = R^0 \cup R^1 \cup R^2 \cup \cdots \cup R^n$. Show that the following chain computing the reflexive-transitive closure of R terminates:

$$R_0 \subseteq R_1 \subseteq \cdots \subseteq X \times X.$$

Solution: Each R_n is an equivariant subset of $X \times X$: R_0 is just the identity relation, which is equivariant, and equivariant relations are closed under composition and union. Since $X \times X$ is orbit-finite (X being orbit-finite and the equality atoms being oligomorphic), R_n is a union of finitely many orbits of $X \times X$. One can then show that there is some $n \leq |\mathsf{Orbits}(X \times X)|$ s.t. $R_n = R_{n+1} = \cdots$. $\square$

Orbit-finite automata

Fix an oligomorphic atom structure $\mathbb{A}$, which will usually consist of a countable set with equality $(\mathbb{A}, =)$. A *orbit-finite automaton* (OFA) is a tuple $A = (\Sigma, Q, I, F, \Delta)$ where Σ is a orbit-finite *input alphabet* (often $\Sigma = \mathbb{A}$), Q is a orbit-finite set of *states*, $I, F \subseteq Q$ are equivariant subsets of Q (thus orbit-finite), called *initial*, resp., *final* states, and $\Delta \subseteq Q \times \Sigma\mathbb{Q}$ is an equivariant set of *transitions* (thus orbit-finite).

Exercise 2. Consider an orbit-finite automaton with input alphabet $\hat{\Sigma} := \Sigma \times \mathbb{A}$ where Σ is finite. Consider the following *projection* mapping $\pi : \hat{\Sigma}^* \to \Sigma^*$ which forgets the data part of a word:

$$\pi : (\sigma_1, a_1) \cdots (\sigma_n, a_n) \mapsto \sigma_1 \cdots \sigma_n.$$

Show that the projection $\pi L \subseteq \Sigma^*$ of a data language $L \subseteq \hat{\Sigma}^*$ recognised by an orbit-finite automaton is a regular language.

Solution: Let $A = \left(\hat{\Sigma}, Q, I, F, \Delta\right)$ be a OFA. Build a $\mathsf{NBA}B$ whose states are orbits of Q, initial states are orbits of I, and final states are orbits of F. A transition $(p, (\sigma, a), q) \in \Delta$ of A induces a transition $\mathsf{orbit}(p) \xrightarrow{\sigma} \mathsf{orbit}(q)$ of B. One then shows that $L(B) = \pi L(A)$. $\square$

The following is a summary of (non)-closure properties of languages of finite data words recognised by OFA and its deterministic variant.

	$\cup$	$\cap$	$\_^R$	$\Sigma^* \setminus \_$
Deterministic OFA	✓	✓	×	✓
Nondeterministic OFA	✓	✓	✓	×

Exercise 3. Show a nondeterministic OFA language which is not recognised by a deterministic OFA.

Solution: Consider the language of all words $w \in \mathbb{A}^*$ s.t. the last letter appears at least twice:

$$L = \{a_1 \cdots a_n \in \mathbb{A}^* \mid \text{there is } 1 \leq i < n \text{ s.t. } a_i = a_n\}.$$

This language is OFA recognisable, in dimension one: The automaton guesses the occurrence of a_i and checks that it appears at the end of the word.

This language is not recognisable by a deterministic OFA. By way of contradiction, let A be a deterministic OFA recognising L. Build a long word of pair-wise distinct letters $w = a_1 \cdots a_n \notin L$ and look at the corresponding run of the automaton

$$I \ni q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n.$$

There is some $n \in \mathbb{N}$ s.t. some a_i is not in the support of the last state,

$$a_i \notin \mathsf{supp}\, q_n.$$

Let $b \in \mathbb{A}$ be a fresh input symbol. Since $w \cdot b \notin L$, the extended run is rejecting:

$$I \ni q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n \xrightarrow{b} q \notin F.$$

Let α be any atom automorphism fixing $\mathsf{supp}\, q_n$ s.t. $\alpha(b) = a$. In particular, $\alpha(q_n) = q_n$. The following modified run

$$I \ni q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n = \alpha(q_n) \xrightarrow{\alpha(b)} \alpha(q) \notin F.$$

shows $w \cdot a \notin L(A)$ since F is equivariant (and thus the same applies to its complement) and the automaton is deterministic. Since $w \cdot a \in L$, this contradicts $L(A) = L$. $\square$

Exercise 4. Show that the class of nondeterministic OFA languages is not closed under complement.

Solution: Consider the language $L \subseteq \mathbb{A}^*$ containing all words where a data value appears at least twice, which is easily seen to be OFA-recognisable. By way of contradiction, assume that its complement is recognised by some OFA A. Consider a very long word $w = a_1 \cdots a_n \in \Sigma^*$ of pairwise distinct data values, and look at some accepting run of A when reading it

$$I \ni q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n \in F.$$

For n sufficiently large there are indices $1 \leq i \leq j < k \leq n$ s.t. $a_i, a_k \notin \mathsf{supp}\, q_j$. There exists an atom automorphism α which fixes $\mathsf{supp}\, q_j$ (and thus $\alpha(q_j) = q_j$) s.t. $\alpha(a_i) = a_k$. The following run is also accepting

$$I \ni \alpha(q_0) \xrightarrow{\alpha(a_1)} \cdots \xrightarrow{\alpha(a_j)} \alpha(q_j) = a_i \xrightarrow{a_{j+1}} \cdots \xrightarrow{a_n} q_n \in F.$$

and thus $w' = \alpha(a_1) \cdots \alpha(a_j) a_{j+1} \cdots a_n \in L(A)$. However the data value a_k appears at least twice in w', thus $w' \in L$, which is a contradiction. □

Exercise 5. Show that the class of deterministic OFA languages is not closed under reversal.

Solution: The language "the last letter appears at least twice" from the solution of Exercise 3 cannot be recognised by a deterministic OFA, however its reversal can. □

Exercise 6. Show that the class of non-guessing OFA languages is not closed under reversal.

Solution: Consider the language L of all words $w \in \mathbb{A}^*$ s.t. "the first letter appears exactly once". This can be recognised by a determinstic OFA, which is non-guessing. Its reversal L^R contains all words where "the last letter appears exactly once". We show that it cannot be recognised without guessing. By way of contradiction let A be a non-guessing OFA recognising L^R. Consider a long word $w = a_1 \cdots a_n \in \mathbb{A}^*$ with pairwise distinct data values. Since $w \in L(A)$, there is an accepting run

$$I \ni q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_{n-1}} q_{n-1} \xrightarrow{a_n} q_n \in F.$$

There is $a_i \notin \mathsf{supp}\, q_{n-1}$. Since the automaton is without guessing $\mathsf{supp}\, q_{n-1} \subseteq a_1, \ldots, a_{n-1}$, and since a_n is fresh (i.e., not in the latter set), also $a_n \notin \mathsf{supp}\, q_{n-1}$. There is an automorphism α that 1) fixes all elements in $\mathsf{supp}\, q_{n-1}$ (and in particular $\alpha(q_{n-1}) = q_{n-1}$), and 2) maps a_i to $\alpha(a_i) = a_n$. The following run is also accepting

$$I \ni \alpha(q_0) \xrightarrow{\alpha(a_1)} \cdots \xrightarrow{\alpha(a_{n-1})} \alpha(q_{n-1}) \xrightarrow{a_n} q_n \in F,$$

however it accepts a word where the last letter a_n appears twice, which is a contradiction. □

Exercise 7 (Universality is undecidable for nondeterministic OFA)**.** Consider the data alphabet $\hat{\Sigma} = \Sigma \times \mathbb{A} \cup \{\$\}$ with Σ finite and $\$ \notin \Sigma$. Consider the following data language

$$L = \{w\$w \mid w = (b_1, a_1) \cdots (b_n, a_n) \text{ and the } a_i\text{'s are pairwise distinct}\}.$$

Show that the complement $\hat{\Sigma}^* \setminus L$ of L can be recognised by

1. A nonguessing OFA of dimension two (two registers in the sense of register automata).

2. A nondeterministic OFA of dimension one, which uses guessing.

Conclude that the universality problem is undecidable for nondeterministic OFA.

Solution: TODO. □