Languages, automata and computation II
Tutorial 6

Winter semester 2023/2024

In this tutorial we explore ω-regular languages $L \subseteq \Sigma^*$. By definition, these are languages recognised by ω-regular expressions:

$$e, f ::= \varepsilon \mid a \mid e + f \mid e \cdot f \mid e^*,$$
$$g, h ::= g + h \mid e \cdot g \mid e^\omega.$$

In the construction of e^ω we require that $\varepsilon \notin L(e)$.

Exercise 1. Show that a language $L \subseteq \Sigma^\omega$ is ω-regular iff it is recognised by a nondeterministic Büchi automaton (NBA).

Solution: For the "only if" direction, assume $L = L(g)$ for some ω-regular expression g. In fact, one can see that g can be taken of the form $g = \sum_{i=1}^{n} e_i \cdot f_i^\omega$, without loss of generality. Since NBA-recognisable languages are closed under union, it suffices to build a NBA for an expression of the form $e \cdot f^\omega$. This is easily done starting from NFA's A, B recognising $L(A) = L(e)$, resp., $L(B) = L(f)$: From each transition $p \xrightarrow{a} q$ of A or B with q accepting, add a transition $p \xrightarrow{a} p_I^B$ to the initial state of B.

For the "if" direction, assume $L = L(A)$ is recognised by a NBA. For every two states p, q of A, consider the language $L(p, q) = \left\{ w \in \Sigma^* \mid p \xrightarrow{w} q \right\}$ of all words labelling a run from p to q. Then L equals

$$L = \bigcup \{ L(p, q)(L(q, q) \setminus \{\varepsilon\})^\omega \mid p \text{ initial}, q \text{ accepting} \}.$$

The writing above clearly shows that the language on the right is ω-regular. Both inclusions are easy to prove. $\square$

Exercise 2 (1)**.** Are the following languages ω-regular?

1. ω-words with infinitely many prefixes in a fixed regular language of finite words $L \subseteq \Sigma^*$.
2. ω-words with infinitely many infixes of the form $ab^p a$ with $p \in \mathbb{N}$ prime.
3. ω-words with infinitely many infixes of the form $ab^n a$ with $n \in \mathbb{N}_{>0}$ even.

Solution: 1. Yes, even deterministically. Take a deterministic finite automaton A recognising L. Now with the same syntax we interpret A as a deterministic Büchi automaton. By definition, the latter automaton recognises the required language.

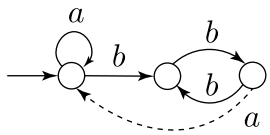

Figure 1: A deterministic Büchi automaton.

2. No. By way of contradiction assume there is an NBA A recognising ω-words with infinitely many infixes of the form ab^pa with $p \in \mathbb{N}$ prime. One such word is $w = (b^pa)^\omega$ where p is a prime chosen to be larger than the number of states of A. Since w is accepted and p is large, while reading the i-th occurrence of the infix b^pa the automaton visits some state q_i twice within the b^p part (possibly a different one across different infixes). Let k_i be the distance between the two occurrences of q_i in the i-th infix. We now pump this segment by repeating it p times, so that the i-th infix is now $b^{p+p\cdot k_i}$, certainly not of prime length. The resulting word $b^{p+p\cdot k_1}a \cdot b^{p+p\cdot k_2}a \cdots$ is still accepted by the automaton, however it does not have *any* infix ab^pa with $p \in \mathbb{N}$ prime, let alone infinitely many of them, which is a contradiction.

   We have the following generalisation: If an NBA A of k states recognises an infinite word $w = u_0u_1 \cdots \in \Sigma^\omega$ with the length of each infix $u_i \in \Sigma^*$ larger than k, then for every $i \in \mathbb{N}$ we can write $u_i = x_iy_iz_i$ with y_i nonempty, x_iz_i of length at most k, and s.t. for every sequence of numbers $n_0, n_1, \cdots \in \mathbb{N}$ (including the case $n_i = 0$) the automaton A also accepts the ω-word

$$x_0y_0^{n_0}z_0 \cdot x_1y_1^{n_1}z_1 \cdots .$$

3. Yes, even deterministically. The automaton keeps track of the parity of k in blocks of the form b^ka and accepts at the end of such a block with k even; cf. fig. 1.

□

Exercise 3 (2)**.** Show that a non-empty ω-regular language contains an ultimately periodic word.

Solution: Let A be an NBA recognising the language. Since the language is not empty, there exists an infinite accepting run. The run visits the set of accepting states infinitely often, and since this set is finite, some particular accepting state $q \in F$ is visited infinitely often. In particular, q is visited twice. We now split the accepting run as $\alpha_u\beta_v\gamma$ where α_u is a finite run over some word $u \in \Sigma^*$ ending in q, β_v over some $v \in \Sigma^*$ starts and ends in q, and γ is the rest of the run. The run $\alpha_u\beta_v^\omega$ witnesses acceptance of the ultimately periodic word uv^ω. The same argument works for acceptance on transitions. □

Exercise 4 (Ultimately periodic words vs. runs)**.** An ultimately periodic run is labelled by an ultimately periodic word. Is it the case that if an NBA accepts an ultimately periodic word, then it has an accepting ultimately periodic run over this word?

Solution: Yes. Let uv^ω be accepted by the automaton and fix an accepting run witnessing this. This run can be factored as $\alpha\beta_1\beta_2\cdots$ where $\alpha\beta_1\cdots\beta_n$ is the run reading prefix uv^n. Infinitely many β_i's visit an accepting state, but not all of them. Refactor the run by combining several β_i's into one in such a way that they always visit an accepting state. The new refactoring is $\alpha'\beta'_1\beta'_2\cdots$ where α' is labelled by some word in uv^* and β'_i by v^+. The run $\alpha'\beta'_1\cdots\beta'_n$ ends in some state, call it q_n. For n larger than the number of states of the automaton, there is a repetition of states $q_i = q_j$, for some $0 \leq i < j \leq n$. The run $\alpha'\beta'_1\cdots\beta'_{i-1}(\beta'_i\cdots\beta'_{j-1})^\omega$ is ultimately periodic and accepts uv^ω. □

Exercise 5 (4)**.** Are the following languages ω-regular?

1. ω-words with arbitrarily long infixes from the regular language ab^*a.
2. ω-words with infinitely many prefixes from a fixed language $L \subseteq \Sigma^*$ (not necessarily regular).

Solution:

1. No, the set of infinite words with arbitrarily long infixes from ab^*a contains no ultimately periodic word.
2. No. Fix any non-ultimately periodic word $u \in \Sigma^\omega$ and let L be the set of its prefixes. Then our language is $M = \{u\}$, which cannot be ω-regular since it is nonempty and does not contain any ω-regular word.

□

Exercise 6 (5)**.** Show that "there exists a letter b" cannot be accepted by an NBA where all states are accepting (but some transitions may be missing).

Solution: By way of contradiction, assume such an automaton exists. The word a^ω is rejected, therefore the automaton cannot read already some finite prefix a^k. In particular a^kb^ω is rejected as well, however it should have been accepted, which is a contradiction. □

Exercise 7 (6)**.** Show that the language "finitely many a's" cannot be accepted by a deterministic Büchi automaton.

Solution: By way of contradiction, assume such an automaton exists. Since the automaton accepts b^ω, it visits an accepting state after some number n_1 of b's. Since the automaton accepts $b^{n_1}ab^\omega$, it visits an accepting state after some number n_2 of b's at the end. Continuing this reasoning, in the limit construct an infinite word $b^{n_1}ab^{n_2}a\cdots$ accepted by the automaton, however it contains infinitely many a's. □

Exercise 8. Are deterministic Büchi languages closed under union? intersection? complement?

Solution: They are closed under union and intersection, as shown by suitable product constructions. They are not closed under complement, however a language is deterministic Büchi iff its complement if deterministic coBüchi. Recall that "finitely many a's" is not deterministic Büchi, however it is deterministic coBüchi. If they were closed under complement then it would be deterministic Büchi, a contradiction. □

Exercise 9. Show that nonemptiness is decidable for nondeterministic Büchi automata. Is the problem in PTIME? Can one do better in terms of computational complexity?

Solution: The problem is complete for nondeterministic logarithmic space (NL), thus in PTIME. For the upper bound, we observe that the language is nonempty iff there is an accepting state p s.t.1) p is reachable from the initial state, and 2) p is reachable from itself via a path visiting an accepting state. A NL algorithm can guess such a p and corresponding witnessing paths.

For the lower bound, notice that s, t-reachability in directed graphs (which is NL-hard) reduces to nonemptiness of Büchi automata where s is the initial state and t is the unique accepting state, augmented with a self-loop (sink state). □

Exercise 10 (7)**.** Show that every language accepted by a nondeterministic Muller automaton is also accepted by some nondeterministic Büchi automaton.

Solution: Initially, the Büchi automaton simulates the Muller automaton, while visiting only rejecting states. At some point, it guesses a Muller set $F \in \mathcal{F}$ and ensures that from this point on *all and only* states from F can be visited (thus F is the *infinity set* of the run). To do this, it uses a break-point construction with the help of an auxiliary set G. Initially, G is the empty set $\emptyset$. Every time a state from F is visited, it adds this to an auxiliary set G. Transitions to states not in F are removed. The first time $G = F$, the automaton declares the current state as accepting and resets G to $\emptyset$. □

Exercise 11 (3)**.** Show that if two ω-regular languages agree over the set of all ultimately periodic words, then they are equal.

Solution: Assume $L, M \subseteq \Sigma^\omega$ contain the same ultimately periodic words. If they are not equal, then the language $L \setminus M \cup M \setminus L$ is nonempty. Since the latter language is ω-regular (determinise Büchi automata into Muller automata, perform Boolean operations, and then convert back to a nondeterministic Büchi automaton), it contains an ultimately periodic word, which is a contradiction. □

Exercise 12. Is the set of all ultimately periodic words ω-regular?

Solution: No, its complement is nonempty and contains no ω-regular word. □

Exercise 13 (8)**.** Show that nonemptiness is decidable for nondeterministic Muller automata.

Solution: Transform a Muller into a Büchi automaton and solve emptiness for the latter. □

Exercise 14 (9)**.** Consider the following binary operation on infinite words:

$$d : \Sigma^\omega \times \Sigma^\omega \to \mathbb{Q}$$
$$d(u, v) := 2^{-n},$$

where n is the length of the longest common prefix of u, v.

- Show that d satisfies the axioms of a *metric*, and thus (Σ^ω, d) is a metric space.

- An *open set* in this space is a language $L \subseteq \Sigma^\omega$ s.t. for all $w \in L$ there is a radius $r > 0$ s.t. the open ball of radius r centered at w is also in L:

$$\{u \in \Sigma^\omega \mid d(w, u) < r\} \subseteq L.$$

  Find an open language which is not ω-regular and viceversa.

- Show that L is open iff $L = K \cdot \Sigma^\omega$ for some $K \subseteq \Sigma^*$.

- If additionally L is ω-regular, then show that K can be chosen to be regular. Conclude that an open ω-regular language is DBA recognisable.

Solution: The verification of d to be a metric (in fact, even an ultrametric) is a matter of expanding the definition.

Regarding the second point, any nonregular language K gives rise to an open non-ω-regular language $K \cdot \{\$\} \cdot \Sigma^\omega$ where \$ is not in the alphabet of K. For instance, this shows that $\{a^n b^n \mid n \in \mathbb{N}\}\,\$\,\{a, b, \$\}^\omega$ is not ω-regular. Alternatively, one can use a direct pumping argument. The language (recognised by) $a^\omega + b^\omega$ is ω-regular but not open.

We now address the third point. Assume that L is of the form $L = K \cdot \Sigma^\omega$. Then $w \in L$ implies that we can write $w = u \cdot v$ for some $u = u_w \in K$ and $v \in \Sigma^\omega$. Moreover, $L_w := u_w \cdot \Sigma^\omega \subseteq L$. This shows that the whole ball of radius $|u_w|$ is in L, and thus L is open. Assume now that L is open. In fact L is covered by languages of the form L_w with $w \in L$:

$$L = \bigcup_{w \in L} L_w.$$

Take K to be $\{u_w \in \Sigma^*\}\, w \in L$.

Regarding the third point, assume L is recognised by a finite deterministic automaton A with any acceptance condition (can be Muller, parity, ...). By the previous point we can write $L = K \cdot \Sigma^\omega$ for some $K \subseteq \Sigma^*$ (not necessarily regular). Let Q be the set of states of A reached by reading some word from K. Then necessarily every $q \in Q$ recognises Σ^ω. From A construct a DFA B with the same states, initial state, and transitions, where additionally all states in Q are accepting. Then $K \subseteq L(B)$ and $L = L(B) \cdot \Sigma^\omega$. In fact, $L(B) \cdot \Sigma^\omega$ is even DBA-recognisable: Replace all accepting states of B with a single universal accepting state. □

Exercise 15 (10)**.** Reminiscent of the Myhill-Nerode characterisation of regular languages via congruences of finite index, we would like to characterise ω-regular languages L via equivalences $\sim_L$ s.t., for every language $L \subseteq \Sigma^\omega$,

$$L \text{ is } \omega\text{-regular iff } \sim_L \text{ has finite index.} \tag{1}$$

Which of the following candidates yields an equivalence satisfying (1)?

1. The equivalence $\sim_L$ on Σ^* defined as

$$u \sim_L v \quad \text{iff} \quad \forall w \in \Sigma^\omega : uw \in L \text{ iff } vw \in L.$$

2. The equivalence $\sim_L$ on Σ^ω defined as

$$u \sim_L v \quad \text{iff} \quad \forall w \in \Sigma^* : wu \in L \text{ iff } wv \in L.$$

3. The equivalence $\sim_L$ on Σ^+ defined as

L: In the book it is on Σ^*, however in that case it is not clear what $s(ut)^\omega$ means when $ut = \varepsilon$.

$$u \sim_L v \quad \text{iff} \quad \begin{cases} \forall w \in \Sigma^\omega : uw \in L \text{ iff } vw \in L, \text{ and} \\ \forall s, t \in \Sigma^* : s(ut)^\omega \in L \text{ iff } s(vt)^\omega \in L. \end{cases}$$

Solution: First we notice that if L is ω-regular, then all candidate equivalences have finite index.

1. We show that $\sim_L$ is refined by an equivalence $\sim_A$ of finite index, and thus itself of finite index. Let $u \sim_A v$ if the NBA A reaches the same set of states when reading u and v. This clearly refines $\sim_L$ and it is of finite index since there are finitely many subsets of states that can possibly be reached.

2. As above, we show that $\sim_L$ is refined by an equivalence $\sim_A$ of finite index. For an NBA A and an ω-word u, let Q_u be the set of states of A from which an accepting run over u can be started. Let $u \sim_A v$ if $Q_u = Q_v$. This is clearly of finite index; it refines $\sim_L$ since $Q_u = Q_v$ and $wu \in L$ implies that from initial state q we can read w and reach a state in Q_u, and thus $wv \in L$ as well.

3. As above, given an NBA A we define an equivalence $\sim_A$ of finite index on Σ^* refining $\sim_L$. Consider the three-valued commutative ring

$$S = (\{-\infty, 0, 1\}, \max, +)$$

where the addition is max, the multiplication is $+$, with $-\infty$ absorbing for $+$ and $1 + 1 = 1$. The *characteristic matrix* of a finite nonempty word $u \in \Sigma^+$ is the matrix $M_u^A \in \{S\}^{Q \times Q}$ s.t., for states $p, q \in Q$, $M_u^A(p, q)$ is 1 if there is a run $p \xrightarrow{u} q$ visiting an accepting state, 0 if there is a run $p \xrightarrow{u} q$ but not one visiting an accepting state, and $\perp$ if there is no run $p \xrightarrow{u} q$. Let $u \sim_A v$ if $M_u = M_v$. This equivalence has finite index since there are finitely many characteristic matrices. We show that this equivalence refines $\sim_L$. Assume $u \sim_A v$ and $s(ut)^\omega \in L$. There is an accepting run $\alpha\beta_0\beta_1 \cdots$ where α reads s, every β_i's reads ut, and for infinitely many i's it does so while visiting an accepting state. Since $M_u = M_v$ we have $M_{ut} = M_u M_t = M_v M_t = M_{vt}$ (the multiplications are taken in the matrix semiring $S^{Q \times Q}$). We can construct a run $\alpha\beta'_0\beta'_1 \cdots$ where β'_i reads vt and visits an accepting state whenever β_i does so. It follows that $s(vt)^\omega \in L$, as required.

We show that for each candidate there is a nonregular language L s.t. $\sim_L$ has finite index.

1. If $L \subseteq \Sigma^\omega$ is prefix-independent, i.e., $uv \in L$ iff $v \in L$ for every finite word $u \in \Sigma^*$, then all finite words u, v are $\sim_L$-equivalent: $uw \in L$ iff $w \in L$ iff $vw \in L$. It thus suffices to find a prefix-independent nonregular language, for example the one from 2. in Exercise 2 would do.

2. If $L \subseteq \Sigma^\omega$ is prefix-independent, then there are two equivalence classes only (and this is the best possible if L is nontrivial): L and its complement $\Sigma^\omega \setminus L$. We can conclude with the same example as in the previous point.

3. Take the example from 1. of Exercise 5, "arbitrarily long infixes in ab^*a". It is prefix independent, therefore whether $uw \in L$ does not depend on u. Moreover, $s(ut)^\omega \in L$ is always false since L does not contain any ultimately periodic word. Thus $\sim_L$ has only one equivalence class.

   L: Check this argument

□

Exercise 16. A language of ω-words is *prefix-independent* if

$$uv \in L \quad \text{iff} \quad v \in L, \qquad \text{for all } u \in \Sigma^*, v \in \Sigma^\omega.$$

1. Consider the definition of prefix-independence for finite words. Describe all the prefix-independent languages of finite words.

2. Describe all prefix-independent ω-regular languages.

Solution: TODO. □

Exercise 17 (Infinite Ramsey theorem)**.** Show that a infinite finitely coloured clique contains an infinite monochromatic clique. (The perhaps more familiar finite Ramsey theorem—any sufficiently large finitely colored clique contains a monochromatic clique—is implied by the infinite version by a compactness argument.)

Solution: We address the countable case. Order the vertices of the graph as an infinite sequence $\pi_0 := v_0 v_1 \cdots$. Construct an infinite sequence of infinite sequences $\pi_0, \pi_1, \ldots$ s.t.

1. sequence π_{i+1} agrees with π_i on the first i vertices, and

2. for all $i \in \mathbb{N}$ there is a color c_i s.t. the i-th vertex of π_i has a c_i-edge to all vertices to its right in the sequence π_i.

The construction is inductive. We obtain a limit sequence π_ω thanks to the first condition, and thanks to the second condition each vertex u_i in it has a color c_i s.t. it has c_i-edges to all vertices to its right. Some color c_i appears infinitely often, inducing the required monochromatic infinite clique. □