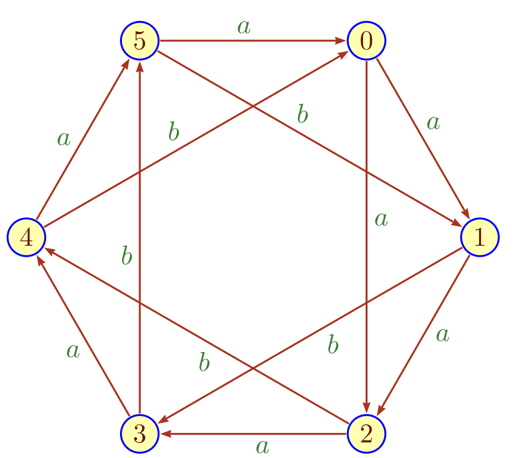
15-354: CDM	**K. Sutner**
Assignment 5	Due: **Oct. 6, 2023, 24:00**.

1. Blow-Up (40)

Background

Let n be a positive integer and write $\mathcal{A}$ for the semi-automaton on n states whose diagram is the circulant with n nodes and strides 1 and 2. The edges with stride 1 are labeled a and the edges with stride 2 are labeled b, except for one. For example, the following picture shows $\mathcal{A}$ for $n = 6$.



Note the a-transition from 0 to 2, if this were a b-transition, the automaton would be completely boring (recall that in a semi-automaton $I = F = Q$, so the language would simply be $\Sigma^\star$). Write K for the acceptance language of $\mathcal{A}$. Call a set $P \subseteq Q$ persistent if $\bigcap_{p \in P} [\![p]\!]_{\mathcal{A}} - [\![Q - P]\!]_{\mathcal{A}} \neq \emptyset$. A string in the non-empty set is a witness for P.

Task

A. Find the length-lex minimal witness for $P = \emptyset$.

B. Repeat for $P = Q$, but with the extra condition that the witness must contain a letter b.

C. Show that all $P \subseteq Q$ are persistent.

D. Show that determinization of $\mathcal{A}$ produces an accessible DFA $\mathcal{B}$ with 2^n states.

E. Argue that $\mathcal{B}$ is already reduced and conclude that K has state complexity 2^n.

Comment

In particular for part (C), arguing in terms of pebbles is probably the best approach.

2. Window Languages (30)

Background

In this problem we only consider languages in Σ^+, so the empty word causes no technical problems. A language L is a window language if membership in L can be tested by sliding a window of size 2 across the word and observing the 2-factors of the word.

Here is a formalization of this idea. Define $\mathsf{fact}_2(x) = \{\, ab \in \Sigma^2 \mid x = uabv, u, v \in \Sigma^\star \,\}$ to be the set of all 2-factors of x. Define an equivalence relation $\approx$ on Σ^+ as follows:

$$x \approx y \iff x_1 = y_1 \wedge \mathsf{fact}_2(x) = \mathsf{fact}_2(y) \wedge x_{-1} = y_{-1}$$

Then L is a window language if it saturates $\approx$: $L = \bigcup_{x \in L} [x]$. Write Σ^{++} for all words of length at least 2. Given $F \subseteq \Sigma^2$ let $L_F = \{\, x \in \Sigma^{++} \mid \mathsf{fact}_2(x) \subseteq F \,\}$.

Task

A. Find a fast algorithm to check whether L_F is finite. What is the running time of your algorithm?

B. Find a simple, infinite language $L \subseteq \Sigma^{++}$ that is not of the form L_F.

C. Show that $\approx$ is a congruence: $x \approx y$ and $u \approx v$ implies $xu \approx yv$.

D. Show that every window language is regular.

E. Show that every regular language is a homomorphic image of a window language.

Comment

For the last part you need to produce a regular language $R \subseteq \Gamma^\star$ for some alphabet Γ and a homomorphism $\Phi : \Gamma^\star \to \Sigma^\star$ such that $\Phi(R) = L$. Γ will depend strongly on a DFA for the given regular language.

This is perhaps a little counterintuitive: the window seems to be too narrow for arbitrary regular languages with long-distance constraints (say, every a is followed by a b after at most 123 letters).

3. The Un-Equal Language (30)

Background

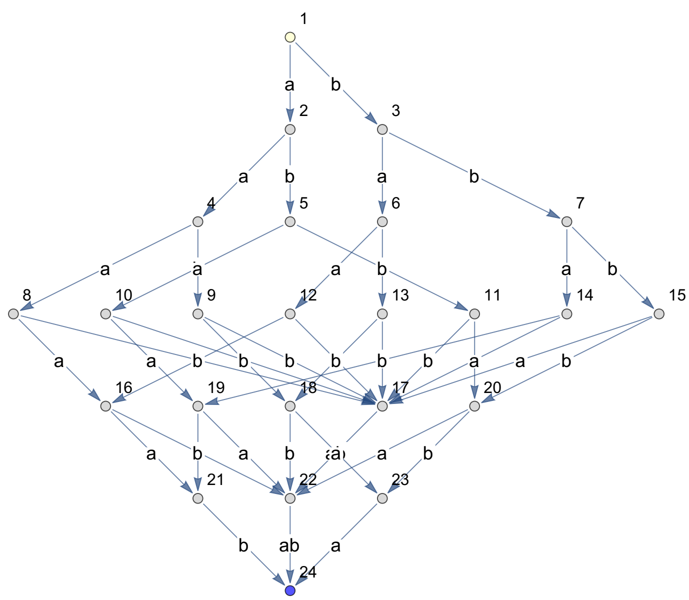
Consider the language of all strings of length $2k$ that are not of the form uu:

$$L_k = \{\, uv \in \{a,b\}^\star \mid |u| = |v| = k, u \neq v \,\}.$$

These languages are finite, hence trivially regular. The following table shows the state complexity of L_k up to $k = 6$.

k	1	2	3	4	5	6
sc	5	12	25	50	99	196

The minimal DFA for L_3 looks like so (the layout algorithm is not too great):



This is the partial DFA without sink, the top state is initial and the bottom state final.

Task

A. What happens when you run Moore's algorithm on this DFA? How many rounds are there?

B. Determine all quotients for L_2.

C. Generalize. In particular explain the diagram for L_3.

D. Determine the state complexity of L_k.

E. Determine the state complexity of $K_k = \{\, uu \mid u \in \{a,b\}^k \,\}$.

Comment For part (A), and in the interest of maintaining TA sanity, use the state numbering from above (sink is state 25).

From the diagram and the table it is not hard to conjecture a reasonable closed form for the state complexity. For a proof one can exploit the description of the minimal DFA in terms of quotients.