Assignment 3

Naman Mishra

16 January, 2024

Problem 3.1. *Give a language $L \subseteq \{a, b\}^*$ such that neither L nor $\{a, b\}^* \setminus L$ contains an infinite regular set.*

We will give a construction by exploiting ultimate periodicity.

Lemma 3.1. *Let $S \subseteq \mathbb{N}$ be given by*

$$S = \bigl\{n \in \mathbb{N} \mid \exists\, k \in \mathbb{N}(2^{2k} \leq n < 2^{2k+1})\bigr\}.$$

There neither S nor $\mathbb{N} \setminus S$ contains an infinite ultimately periodic set.

Proof. We have

$$\mathbb{N} \setminus S = \bigl\{n \in \mathbb{N} \mid \exists\, k \in \mathbb{N}(2^{2k+1} \leq n < 2^{2k+2})\bigr\}.$$

This is easy to see if one notices that S is the set of all n such that $\lfloor \log_2 n \rfloor$ is even.

Claim: S does not contain an infinite ultimately periodic set.

Suppose S contains an infinite ultimately periodic set S', with period $p > 0$ and starting from n_0, *i.e.*, for all $n \geq n_0$, $n \in S'$ iff $n + p \in S'$. Since S is infinite, it is unbounded, so there exists some $m \in S'$ larger that n_0. Then $m + np \in S'$ for all $n \in \mathbb{N}$. By the well-ordering principle, let

$$n = \min\bigl\{n \in \mathbb{N} \mid m + np \geq 2^{2(m+p)+1}\bigr\}.$$

Then $n \neq 0$ since $m < 2^m < 2^{2(m+p)+1}$, and $m + (n-1)p < 2^{2(m+p)+1}$ by minimality of n. So

$$\begin{aligned} m + np &= m + (n-1)p + p \\ &< 2^{2(m+p)+1} + p \\ &< 2^{(m+p)+1} + 2^{(m+p)+1} \\ &= 2^{(m+p)+2} \end{aligned}$$

which gives that $m + np \in \mathbb{N} \setminus S$. But then $m + np$ cannot be in S', a contradiction.

Claim: $\mathbb{N} \setminus S$ does not contain an infinite ultimately periodic set.

This proof is almost identical, ~~so we leave it as an exercise for the grader~~. Let S' be an infinite ultimately periodic subset of $\mathbb{N} \setminus S$, and let p, n_0 and m be as above. Let $n = \min\{n \in \mathbb{N} \mid m + np \geq 2^{2(m+p)}\}$. Again $n \neq 0$ since $m < 2^m < 2^{2(m+p)}$, and $m + (n-1)p < 2^{2(m+p)}$ by minimality of n. Then $m + np < 2^{2(m+p)} + p < 2^{2(m+p)+1}$, so $m + np \in S$. Thus it cannot be in S', a contradiction.

Alternatively, one can show that if $S' \subseteq \mathbb{N} \setminus S$ is infinite and ultimately periodic, then $S'' = \{2n \mid n \in S'\}$ is an infinite and ultimately periodic subset of S, which does not exist by the first claim. □

We will now use this lemma to construct the language L.

Solution. Let S be as in the lemma. Define $L = \{w \in A^* \mid \#w \in S\}$, where $A = \{a, b\}$. Then lengths$(L) = S$ and lengths$(A^* \setminus L) = \mathbb{N} \setminus S$. Let L' be an infinite subset of L. Then lengths$(L') \subseteq S$. If lengths(L') were finite, then L' would be finite, since there are finitely many strings in A^* of each length (to be precise, $|A^n| = 2^n$). But then lengths(L') is infinite, so it cannot be ultimately periodic and hence L' is not regular.

Similarly, if L' is an infinite subset of $A^* \setminus L$, then lengths$(L') \subseteq \mathbb{N} \setminus S$ is infinite, and hence not ultimately periodic, so L' is not regular. ■

Problem 3.2. *For a language L over an alphabet A define*

$$\textit{first-halves}(L) = \{x \in A^* \mid \exists\, y(|x| = |y| \text{ and } xy \in L)\}$$

Prove or disprove: if L is regular, then so is first-halves(L).

Solution. Let $L \subseteq A^*$ be regular with DFA $\mathcal{A} = (Q, s, \delta, F)$ accepting it, with no unreachable states.

For each state $q \in Q$, define

$$\ell(q) := \Big\{ n \in \mathbb{N} \mid \exists\, t \in A^n (\widehat{\delta}(q, t) \in F) \Big\}.$$

For any string w such that $\widehat{\delta}(s,w)=q$, we have $\widehat{\delta}(q,t)=\widehat{\delta}(s,wt)$ for all $t\in A^*$.[1]So $\ell(q)=\{n-|w| : n\in \text{lengths}(L_w)\}$, where L_w is the intersection of L with the set of all strings beginning with w. By closure, this is regular, so $\ell(q)$ is ultimately periodic.[2]Thus we define

$$n(q),p(q) \coloneqq (n,p) \text{ such that } \forall\, m\geq n(m\in l(q) \iff m+p\in l(q)).$$

where $n(q)\in\mathbb{N}$ and $p(q)\in\mathbb{N}\setminus\{0\}$. The particular choice of $n(q)$ and $p(q)$ does not matter, but one can still prescribe a scheme such as the following: Among all such pairs (n,p), choose those with the smallest p, and among those, choose the one with the smallest n.

Now let $P=\prod_{q\in Q} p(q)$ and $N=P\cdot\max_{q\in Q} n(q)$. Then for each state q, we have

$$\forall\, m\geq N(m\in\ell(q) \iff m+P\in\ell(q)).$$

Let $A^{<N}$ and $A^{\geq N}$ be the sets of all strings of length less than N and at least N respectively. We will show that

$$\textit{first-halves}(L)\cap A^{\geq N}$$

is regular. Let

$$\begin{aligned}
Q' &= Q\times\{0,1,\ldots,P-1\}\\
s' &= (s,0)\\
\delta'((q,r),a) &= (\delta(q,a),(r+1) \bmod P)\\
F' &= \{(q,r)\in Q' \mid N+r\in\ell(q)\}.
\end{aligned}$$

Let $\mathcal{A}'=(Q',s',\delta',F')$. We first show that for any $w\in A^*$,

$$\widehat{\delta'}(s',w)=(\widehat{\delta}(s,w),|w| \bmod P).$$

The base case $w=\epsilon$ is direct substitution. For the inductive step, suppose this is true for some w. Then

$$\begin{aligned}
\widehat{\delta'}(s',wa) &= \delta'(\widehat{\delta'}(s',w),a)\\
&= \delta'((\widehat{\delta}(s,w),|w| \bmod P),a)\\
&= (\delta(\widehat{\delta}(s,w),a),(|w|+1) \bmod P)\\
&= (\widehat{\delta}(s,wa),|wa| \bmod P).
\end{aligned}$$

This closes the induction.

[1]This follows from $\widehat{\delta}(q_1,xy)=\widehat{\delta}(\widehat{\delta}(q_1,x),y)$, proved in the first quiz.

[2]The set of all strings with prefix w is the concatenation of $\{w\}$ with A^*, which are both regular.

We claim that

$$L(\mathcal{A}') \cap A^{\geq N} = \textit{first-halves}(L) \cap A^{\geq N}.$$

Proof. Let $|w| \geq N$, so we can write $|w| = N + mP + r$, where $m \in \mathbb{N}$ and $r \in \{0, 1, \ldots, P-1\}$. Since N is a multiple of P, $r = |w| \bmod P$. Let $q = \widehat{\delta}(s, w)$. w is said to be in *first-halves*(L) iff there exists an $x \in A^{N+mP+r}$ such that $wx \in L$. But for any x, this is the same as saying

$$\widehat{\delta}(s, wx) \in F \quad \text{or} \quad \widehat{\delta}(q, x) \in F.$$

So the existence of such an x is equivalent to

$$N + mP + r \in \ell(q)$$

But by the construction of N and P, this is equivalent to

$$N + r \in \ell(q)$$

by the periodicity of $\ell(q)$ with period $p(q) \mid P$ and starting from $n(q) \leq N$. But $\widehat{\delta'}(s, w) = (q, r)$, so this is in turn equivalent to

$$\widehat{\delta'}(s', w) \in F'.$$

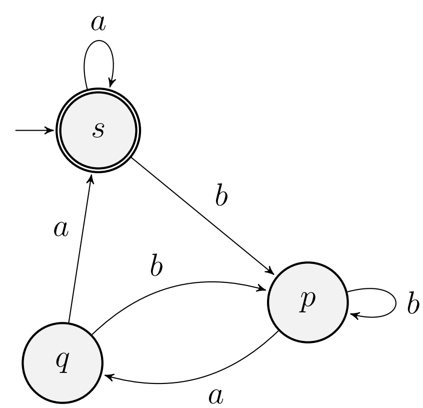
Since each step was an equivalence, we have that for any string w of length at least N, $w \in \textit{first-halves}(L)$ iff $w \in L(\mathcal{A}')$. This proves the claim. □

Finally,

$$\begin{aligned} \textit{first-halves}(L) &= \textit{first-halves}(L) \cap A^{<N} \cup \textit{first-halves}(L) \cap A^{\geq N} \\ &= \textit{first-halves}(L) \cap A^{<N} \cup L(\mathcal{A}') \cap A^{\geq N}. \end{aligned}$$

first-halves$(L) \cap A^{<N}$ is regular since it is finite, and $A^{\geq N}$ is regular, since it is the complement of $A^{<N}$, which is regular by finiteness. Regularity of *first-halves*(L) follows from the closure properties. ■

Problem 3.3. *Use the McNaughton-Yamada construction done in class to construct a regular expression corresponding to the language accepted by the DFA below (i.e. the expression corresponding to $L_{ss}^{\{s,p,q\}}$).*



Solution. We wish to compute the regular expression for $L_{ss}^{\{s,p,q\}}$, since the only start and final states are s. We will write $L \to r$ to mean that the regular expression for L is r.

We will use the equation

$$\begin{aligned} L_{ss}^{\{s,p,q\}} &= L_{ss}^{\{p,q\}} \cup L_{ss}^{\{p,q\}}(L_{ss}^{\{p,q\}})^* L_{ss}^{\{p,q\}} \\ &= L_{ss}^{\{p,q\}}(\epsilon + (L_{ss}^{\{p,q\}})^+) \\ &= L_{ss}^{\{p,q\}}(L_{ss}^{\{p,q\}})^* \\ &= (L_{ss}^{\{p,q\}})^* \quad \text{since } \epsilon \in L_{ss}^{\{p,q\}}. \end{aligned}$$

We have

$$\begin{array}{lll} L_{ss}^{\varnothing} = \{\epsilon, a\} & L_{sp}^{\varnothing} = \{b\} & L_{sq}^{\varnothing} = \varnothing \\ L_{ps}^{\varnothing} = \varnothing & L_{pp}^{\varnothing} = \{\epsilon, b\} & L_{pq}^{\varnothing} = \{a\} \\ L_{qs}^{\varnothing} = \{a\} & L_{qp}^{\varnothing} = \{b\} & L_{qq}^{\varnothing} = \{\epsilon\} \end{array}$$

First note that $(L_{qq}^{\varnothing})^* = \epsilon^* = \epsilon$. So

$$\begin{aligned} L_{sq}^{\varnothing}(L_{qq}^{\varnothing})^* &\to \varnothing \\ L_{pq}^{\varnothing}(L_{qq}^{\varnothing})^* &\to a \end{aligned}$$

Now

$$\begin{aligned} L_{ss}^{\{q\}} &= L_{ss}^{\varnothing} \cup L_{sq}^{\varnothing}(L_{qq}^{\varnothing})^* L_{qs}^{\varnothing} \\ &\to (\epsilon + a) + \varnothing a \\ &= \epsilon + a \end{aligned} \qquad \begin{aligned} L_{sp}^{\{q\}} &= L_{sp}^{\varnothing} \cup L_{sq}^{\varnothing}(L_{qq}^{\varnothing})^* L_{qp}^{\varnothing} \\ &\to b + \varnothing b \\ &= b \end{aligned}$$

$$\begin{aligned} L_{ps}^{\{q\}} &= L_{ps}^{\varnothing} \cup L_{pq}^{\varnothing}(L_{qq}^{\varnothing})^* L_{qs}^{\varnothing} \\ &\to \varnothing + aa \\ &= aa \end{aligned} \qquad \begin{aligned} L_{pp}^{\{q\}} &= L_{pp}^{\varnothing} \cup L_{pq}^{\varnothing}(L_{qq}^{\varnothing})^* L_{qp}^{\varnothing} \\ &\to (\epsilon + b) + ab \end{aligned}$$

And we can now write

$$\begin{aligned} L_{ss}^{\{p,q\}} &= L_{ss}^{\{q\}} \cup L_{sp}^{\{q\}}(L_{pp}^{\{q\}})^* L_{ps}^{\{q\}} \\ &\to (\epsilon + a) + b(\epsilon + b + ab)^* aa \\ &= \epsilon + a + b(b + ab)^* aa \end{aligned}$$

and so

$$\begin{aligned} L_{ss}^{\{s,p,q\}} &\to (\epsilon + a + b(b + ab)^* aa)^* \\ &= (a + b(b + ab)^* aa)^* \end{aligned}$$

which is the regular expression for the language accepted by the given DFA. ■

Problem 3.4. *In the McNaughton-Yamada construction of an RE from an NFA, we inductively define $L(p, X, q)$ to be the words accepted by paths from state p to state q possibly using intermediate states in the set of states X. Inductively define $LA(p, Y, q)$, the words accepted by paths from state p to state q, but avoiding using intermediate states in Y. What would be the base case?*

Solution. Let the NFA be $\mathcal{A} = (Q, S, \Delta, F)$. The base case is $Y = Q$, and $LA(p, Q, q)$ is given by

$$LA(p, Q, q) = \{a \in A \cup \{\epsilon\} \mid q \in \Delta(p, a)\}.$$

The inductive step becomes

$$LA(p, Y \setminus \{y\}, q) = LA(p, Y, q) \cup LA(p, Y, y) \cdot (LA(y, Y, y))^* \cdot LA(y, Y, q)$$

whenever $y \in Y$. This can be seen simply by noticing that

$$L(p, X, q) = LA(p, Q \setminus X, q)$$

and using the inductive definition of $L(p, X, q)$. ■

Problem 3.5. *Consider the languages L and M below over the alphabet $\{a, b\}$.*

- *L is the language of all strings in which the difference between the number of a's and b's is at most 2. That is:*

$$L = \{w \in \{a, b\}^* : |\#_a(w) - \#_b(w)| \leq 2\}.$$

- *M is the language of all strings which satisfy the property that in every prefix the difference between the number of a's and b's is at most 2. That is:*

$$M = \{w \in \{a, b\}^* \mid \textit{for all prefixes } u \textit{ of } w,\ |\#_a(u) - \#_b(u)| \leq 2\}.$$

Describe the classes of the canonical MN relation $\equiv_L$ for L, and similarly for M. Finally, conclude whether L and M are regular or not.

Solution. Let $A = \{a, b\}$. Also define the operator $\delta = \#_a - \#_b$. We claim that for all $v, w \in A^*$,

$$v \equiv_L w \iff \delta(v) = \delta(w).$$

and therefore,

$$A^*/{\equiv_L} = \{\{w \in A^* : \delta(w) = k\} \mid k \in \mathbb{Z}\}.$$

Proof. Let $\sim$ be the relation on A^* defined by

$$v \sim w \iff \delta(v) = \delta(w).$$

We wish to prove that $\sim\, = \,\equiv_L$. Let $v, w \in A^*$ with $v \sim w$. Then for any $z \in A^*$,

$$\begin{aligned} vz \in L &\iff |\#_a(vz) - \#_b(vz)| \leq 2 \\ &\iff |\#_a(v) - \#_b(v) + \#_a(z) - \#_b(z)| \leq 2 \\ &\iff |\#_a(w) - \#_b(w) + \#_a(z) - \#_b(z)| \leq 2 \\ &\iff |\#_a(wz) - \#_b(wz)| \leq 2 \\ &\iff wz \in L. \end{aligned}$$

Thus $v \equiv_L w$.

For the converse, let $a^{-n} = b^n$ for $n \in \mathbb{Z}^+$. Note that $\delta(a^n) = n$ for every $n \in \mathbb{Z}$, and so for every $x \in A^*$,

$$\begin{aligned} \delta(xa^n) &= \delta(x) + \delta(a^n) \\ &= \delta(x) + n. \end{aligned}$$

Suppose $v \equiv_L w$. Let $\delta(v) = k$. Then va^{-k-2} and va^{-k+2} are in L by $(*)$. But since $v \equiv_L w$, wa^{-k-2} and wa^{-k+2} are also in L. Thus

$$\begin{aligned} |\delta(w) - k - 2| &\leq 2 & \qquad |\delta(w) - k + 2| &\leq 2 \\ \delta(w) - k - 2 &\geq -2 & \qquad \delta(w) - k + 2 &\leq 2 \\ \delta(w) &\geq k & \qquad \delta(w) &\leq k \end{aligned}$$

so $\delta(w) = k$, which means $v \sim w$.

Thus $v \equiv_L w \iff v \sim w$. $\square$

Now for M, we first note that for any $x, y \in A^*$,

$$xy \in M \implies x \in M, \tag{$\dagger$}$$

since each prefix of x is also a prefix of xy.

Let $\sim$ be the relation on A^* defined by

$$v \sim w \iff v, w \notin M \text{ or } v, w \in M \wedge (\delta(v) = \delta(w)).$$

We claim that $\sim = \equiv_M$.

Proof. Let $v, w \in A^*$ with $v \sim w$. We have two cases, either $v, w \notin M$, or $v, w \in M$ and $\delta(v) = \delta(w)$. In the first case, $vz, wz \notin M$ for any $z \in A^*$, because of ($\dagger$). So $vz \in M \iff wz \in M$.

In the second case, each prefix u of v or w has $|\delta(u)| \leq 2$ (since $v, w \in M$). Thus for any $z \in A^*$, we only need to consider prefixes of vz that are longer than v, and similarly for wz. That is,

$$\begin{aligned} vz \in M &\iff \text{for all prefixes } u \text{ of } z,\ |\delta(vu)| \leq 2 \\ &\iff \text{for all prefixes } u \text{ of } z,\ |\delta(v) + \delta(u)| \leq 2 \\ &\iff \text{for all prefixes } u \text{ of } z,\ |\delta(w) + \delta(u)| \leq 2 \\ &\iff \text{for all prefixes } u \text{ of } z,\ |\delta(wu)| \leq 2 \\ &\iff wz \in M. \end{aligned}$$

In either case, $v \equiv_M w$.

Now suppose $v \equiv_M w$, *i.e.*, for all $z \in A^*$, $vz \in M \iff wz \in M$. Then $v \in M \iff w \in M$ (take $z = \epsilon$). If $v, w \notin M$, then $v \sim w$ by definition.

Otherwise, $v, w \in M$. We need to show $\delta(v) = \delta(w)$. Let $\delta(v) = k$. Since $v \in M$, $-2 \leq k \leq 2$ Then $k = \delta(v) < \delta(va) < \cdots < \delta(va^{2-k}) = 2$, and $k = \delta(v) > \delta(vb) > \cdots > \delta(vb^{k+2}) = -2$. Since $v \in M$, these are all the prefixes we need to consider to conclude that $va^{2-k} \in M$ and $vb^{k+2} \in M$.

But since $v \equiv_M w$,

$$\begin{aligned} wa^{2-k} &\in M \\ \delta(wa^{2-k}) &\le 2 \\ \delta(w) + 2 - k &\le 2 \\ \delta(w) &\le k \end{aligned} \qquad\qquad \begin{aligned} wb^{k+2} &\in M \\ \delta(wb^{k+2}) &\ge -2 \\ \delta(w) - k - 2 &\ge -2 \\ \delta(w) &\ge k \end{aligned}$$

so $\delta(w) = k = \delta(v)$, which gives $v \sim w$.

Thus $v \equiv_M w \iff v \sim w$. $\square$

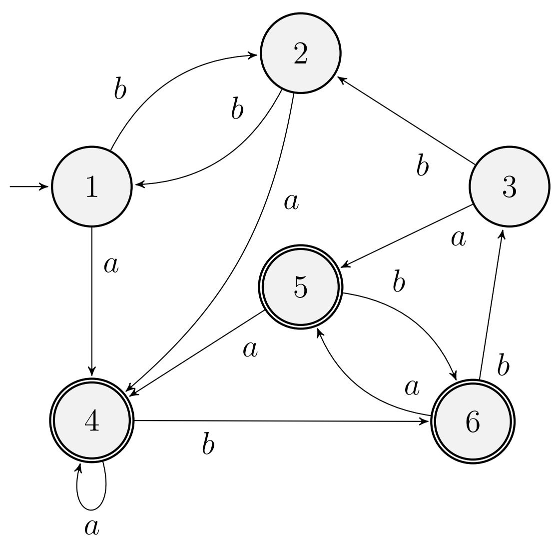
Thus the equivalence classes of $\equiv_M$ are

$$A^*/{\equiv_M} = \{M \cap \{w \in A^* : \delta(w) = k\} \mid k \in \{-2, \dots, 2\}\} \cup \{A^* \setminus M\}.$$

Since $A^*/{\equiv_M}$ is finite but $A^*/{\equiv_L}$ is not,

- L is not regular.
- M is regular. $\blacksquare$

Problem 3.6. *Minimize the DFA below using the algorithm done in class:*



Solution. We start with marking all pairs of nodes that contain an accepting state and a non-accepting state.

	1	2	3	4	5	6
1				✓	✓	✓
2				✓	✓	✓
3				✓	✓	✓
4	✓	✓	✓			
5	✓	✓	✓			
6	✓	✓	✓			

Now $\delta(4, b) = \delta(5, b) = 6$, but $\delta(6, b) = 3$. Thus we can mark $(4, 6)$ and $(5, 6)$.

	1	2	3	4	5	6
1				✓	✓	✓
2				✓	✓	✓
3				✓	✓	✓
4	✓	✓	✓			✓
5	✓	✓	✓			✓
6	✓	✓	✓	✓	✓	

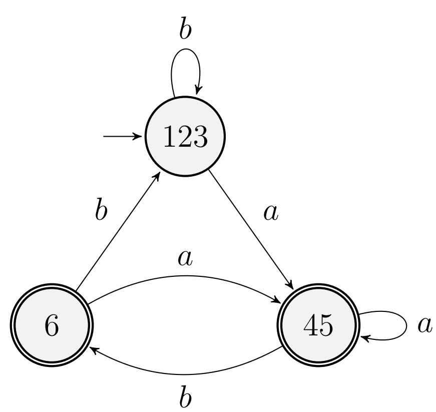
Now $\{\delta(1, a), \delta(2, a), \delta(3, a)\} = \{4, 5\}$, but no pair from $\{4, 5\}$ is marked. $\{\delta(1, b), \delta(2, b), \delta(3, b)\} = \{1, 2\}$, and again no pair from $\{1, 2\}$ is marked.

Finally, $\delta(4, a) = \delta(5, a)$ and $\delta(4, b) = \delta(5, b)$, but obviously no pair (q, q) is ever marked.

Thus there are no more pairs to mark, and we get equivalence classes

$$\{1, 2, 3\} \qquad \{4, 5\} \qquad \{6\}.$$

This gives the minimized DFA



■