COL352
Problem Sheet 2

Himanshi Ghai (2019CS50433)
Mallika Prabhakar (2019CS50440)
Sayam Sethi (2019CS10399)

February 2022

Contents

1 Question 1

Question 1

Question. *Prove that $L_1 = \{bin(p) : p \text{ is a prime number}\}$ is not a regular language.*

Solution. Proof by contradiction using Pumping lemma.
We use the following corollary (contrapositive) of the pumping lemma-

$$(\forall n \ \exists w \in L \text{ where } |w| \geq n, \forall w = xyz \text{ such that } y \neq \epsilon \text{ and } |xy| \leq n \implies \exists k \geq 0 \text{ s.t. } xy^k z \notin L)$$
$$\implies L \text{ is not regular} \tag{1}$$

Let word $w \in L_1$ be a binary representation of a prime number p which can be represented as xyz such that $|xy| \leq n \ \forall n$.
Given the binary representation xyz, the value of prime p in decimal is:

$$p = 2^{k+m}x + 2^m y + z$$

After pumping y, k times, value of p changes to p'-

$$p' = 2^{ki+m}x + 2^m y \frac{2^{ki}-1}{2^k-1} + z$$

Where $\frac{2^{ki}-1}{2^k-1}$ comes from summing the geometric progression made by terms containing y
Let i = p in this case-

$$p' = 2^{kp+m}x + 2^m y \frac{2^{kp}-1}{2^k-1} + z$$

By Fermat's little theorem, we know that if p is prime then $a^p \equiv a(\text{mod p})$.
Taking modulo p-

$$p' = 2^{k+m}x + 2^m y \frac{2^k-1}{2^k-1} + z$$

which is same as

$$2^{k+m}x + 2^m y + z$$

Since bin(p') is longer than bin(p), it implies that p is a factor of p', hence p' is not a prime.
Therefore we can say that $\exists k$ such that $xy^k z \notin L_1$.
Hence proved that L_1 is not a regular language. $\square$

2 Question 2

Question 2

Question. *The n-th Fibonacci number is defined as $F_1 = 1, F_2 = 1$, and for all $n \geq 3, F_n = F_{n-1} + F_{n-2}$.*
Consider the language over $\Sigma = \{a\}$

$$L_2 = \{a^m | m = F_n\} \tag{2}$$

Is L_2 regular? Justify your answer.

Solution. Let's consider an equivalence relation $\equiv$ on $\{a\}^*$ and suppose it is right congruence and refines given language L_2. Suppose that it has finite index i.e it has finite equivalence classes.

So, there exist n and m such that $1 \leq n < m$ and $[a^n] = [a^m]$ as the language is infinite. As the relation is right congruence then

$$[a^n][a^{fib_m - n}] = [a^m][a^{fib_m - n}]$$

where fib_m is the m^{th} Fibonacci number.
So,

$$[a^{fib_m}] = [a^{fib_m + (m-n)}]$$

We know that $LHS \in L_2$ and, as $fib_m + (m-n) < fib_m + m < fib_m + fib_m < fib_m + fib_{m+1}$ we can say that $RHS \notin L_2$.
This contradicts our assumption. Hence there doesn't exist a Myhill Nerode Relation for language L_2. So L_2 is not regular.

□

3 Question 3

Question 3

Question. *If A is any language, let $A_{\frac{1}{2}-}$ denote the set of all first halves of strings in A so that*

$$A_{\frac{1}{2}-} = \{x| \text{ for some } y,\ |x| = |y| \text{ and } xy \in A\} \tag{3}$$

Show that if A is regular, then so is $A_{\frac{1}{2}-}$.

Solution. Let the DFA of A be $D = (Q, \Sigma, \delta, q_0, F)$. We propose the following DFA $D_{\frac{1}{2}-} = (Q_{\frac{1}{2}-}, \Sigma, \delta_{\frac{1}{2}-}, q_{0\frac{1}{2}-}, F_{\frac{1}{2}-})$ as the DFA which accepts $A_{\frac{1}{2}-}$:

$$\begin{aligned}
Q_{\frac{1}{2}-} &= Q \times 2^Q \\
q_{0\frac{1}{2}-} &= (q_0, F) \\
\delta_{\frac{1}{2}-}((q, R), a) &= (\delta(q, a), \{s \in Q | \exists \alpha \in Q : \delta(s, \alpha) \in R\}) \\
F &= \{(q, R) | q \in R, R \in 2^Q\}
\end{aligned} \tag{4}$$

Intuitively, $A_{\frac{1}{2}-}$ is the DFA which begins with the start symbol and the set of final states. Each time it processes a symbol, it advances the state and moves backwards on the transition for the set of states. Finally, if we arrive at a state which is a subset of the reverse traversal states, we know that we have arrived at a state which has a path using $|w|$ transitions which arrives at a final state in D.
We now formally prove the correctness using induction.

Claim 3.1. *After processing i symbols, let the state be (q_i, R_i). Then from all states $s \in R_i$ there exists a walk to F in i transitions. Additionally, each such state is present in R_i.*

Proof. The base case is true ($i = 0$) trivially. Now assume that the claim is true for i, consider the claim for $i + 1$.
Let the sequence of states be $(q_0, R_0), (q_1, R_1), \ldots, (q_i, R_i), (q_{i+1}, R_{i+1})$. We know that there is a path of i transitions from all states in R_i. From the construction of R_{i+1} we know that there is a transition from each state of R_{i+1} to a state of R_i. Therefore, adding this transition to the walk of i transitions in R_i, we get that there exists a walk to a F in $i + 1$ transitions. Also, each state that has a walk of $i + 1$ transitions is present in R_{i+1} since otherwise, there would be a state corresponding to that walk which won't be present in R_i. This would lead to a contradiction to our inductive hypothesis. Therefore, each such state having a walk of $i + 1$ transitions to F is present in R_{i+1}. ☐

We now use the above claim to prove the following:

1. For every word $w \in A$, $w_{\frac{1}{2}-}$ is recognised by $D_{\frac{1}{2}-}$
2. For every word w which is recognised by $D_{\frac{1}{2}-}$, there exists a word $w_0 \in A$.

To prove the first statement, we use the result of the previous claim. Let $|w| = 2n$. Now, after $D_{\frac{1}{2}-}$ processes n symbols, let the state be (q_n, R_n). Now, since w is accepted by D, there exists a walk of length n from q_n to a state in F. Therefore, $q_n \in R_n$. Thus, (q_n, R_n)

is an accepting state and it accepts $w_{\frac{1}{2}-}$.
For the second statement, we know that there exists a path of n transitions from any accepting state (q_n, R_n). We can construct a word w_0 by appending the transition symbols which will be arrive at a state in F. Therefore, for each word which is accepted by $D_{\frac{1}{2}-}$, we have a corresponding word that is accepted by D. $\square$

4 Question 4

Question 4

Question. *If A is any language, let $A_{\frac{1}{3}-\frac{1}{3}}$ denote the set of strings in A with the middle third removed so that*

$$A_{\frac{1}{3}-\frac{1}{3}} = \{xz | \text{ for some } y, |x| = |y| = |z| \text{ and } xyz \in A\} \tag{5}$$

Show that if A is regular, then $A_{\frac{1}{3}-\frac{1}{3}}$ is not necessarily regular.

Solution. We will show this by contradiction. Consider the language $A = 0^+1^+2^+$. A is regular. Now assume that $A_{\frac{1}{3}-\frac{1}{3}}$ is regular.
We also know that $B = 0^+2^+$ is regular. Consider the intersection of $A_{\frac{1}{3}-\frac{1}{3}}$ and B. We know that this language, say C, will only have 0 and 2. We will now show that the number of 0 and number of 2 in each word of C will be equal.
Consider any word w in C. Consider the corresponding word in A from which we got w in $A_{\frac{1}{3}-\frac{1}{3}}$, call it $w_0 = 0^a1^b2^c$. For the case $a = b = c$, it is trivial that the number of 0 and 2 in w will be equal. Otherwise, we know that at least one of a, b, c will be less than $\frac{a+b+c}{3}$. We will now show that if $w \in C$, then $b < \frac{a+b+c}{3}$. Assume that it is not the case. Then, wlog assume that a is the smallest number. Then, in w, we will also have 1 at some positions. However since $w \in C$, such a situation is not possible. Therefore, $\#0 = \#2$ in w.
Thus, we have proved that $C \subseteq \{0^n2^n, n > 0\}$. We now show that any word w in $\{0^n2^n, n > 0\}$ will be in C. This is easy to see since $0^n1^n2^n$ is in A and thus 0^n2^n will be in C. Therefore, $C = \{0^n2^n, n > 0\}$
However, we know that C is non regular. However, from our assumption that $A_{\frac{1}{3}-\frac{1}{3}}$ is regular and B is regular, C must be regular since it is the intersection of $A_{\frac{1}{3}-\frac{1}{3}}$ and B. This is a contradiction. Therefore, $A_{\frac{1}{3}-\frac{1}{3}}$ is irregular. Hence, proved. $\square$

5 Question 5

Question 5

Question. *A 2-NFA A is a 5-tuples $A = (Q, S, t, \Sigma, \Delta)$ where Q is the set of states, S the set of start states, t is an accept state, the transition function*

$$\Delta : Q \times (\Sigma \cup \{\#, \$\}) \rightarrow 2^{Q \times (\{L,R\})}$$

Assume that whenever M accepts, it does so by moving the head (pointer) all the way to the right end marker \$ and entering accept state t. In the subsequent two questions, we will try to prove 2-NFAs accept only regular languages.

1. *Let $x = a_1 \ldots a_n \in \Sigma^*, a_i \in \Sigma, 1 \leq i \leq n$. Let $a_0 = \#, a_{n+1} = \$$. Argue that x is not accepted by A if and only if there exists sets $W_i \subseteq Q, 0 \leq i \leq n+1$ such that the following hold:*
   - $S \subseteq W_0$
   - *If $u \in W_i, 0 \leq i \leq n$, and $(v, R) \in \Delta(u, a_i)$, then $v \in W_{i+1}$*
   - *If $u \in W_i, 1 \leq i \leq n+1$, and $(v, L) \in \Delta(u, a_i)$, then $v \in W_{i-1}$*
   - $t \notin W_{n+1}$
2. *Using the previous part, show that $L(A)$ is regular*

Solution: Part 1A. ($\implies$) We assume that x is not accepted
For each run of A, we keep a track of the states that we end up in before reading each symbol in x. Let us denote the set of states reached before reaching symbol a_i for run j as $(S_j)_i$. There are a finite number of runs for each x and therefore j is bounded by say k. We now define the sets W_i as:

$$W_i = \bigcup_{j=1}^{k} (S_j)_i \tag{6}$$

Now, we prove all the four conditions:

1. $S \subseteq W_0$: This is trivially true since at least one run exists where A begins in each start state.

2. If $u \in W_i, 0 \leq i \leq n$, and $(v, R) \in \Delta(u, a_i)$, then $v \in W_{i+1}$: If for a run S_j of x, we get the transition (v, R), we then read a_{i+1}. Therefore, v will be in the set $(S_j)_{i+1}$ and thus in the state W_{i+1}

3. If $u \in W_i, 1 \leq i \leq n+1$, and $(v, L) \in \Delta(u, a_i)$, then $v \in W_{i-1}$: Similar to the previous part, this will be true.

4. $t \notin W_{n+1}$: Since x is not accepted by A, therefore, no run of A on reading x will end on t on reaching $a_{n+1} = \$$ (just before reading \$). Therefore, $t \notin W_{n+1}$

□

Solution: Part 1B. ($\Longleftarrow$) all conditions are true, show x is rejected.

First condition says that S is an improper subset of W_0. Second and third condition set up the sets $W_i s$ using R and L on given states. Since the number of runs for x is finite, let's say that x has k runs. Let the set of states reached just before reaching symbol a_i in run j be $(S_j)_i$.
Using second and third, we observe that W_i is the set of all such states we can be present in before reading the i-th symbol of x, i.e., a_i. After reading the i-th symbol, we can go to a state which is in either of W_{i-1} or W_{i+1}. Therefore, W_i is exactly equal to:

$$W_i = \bigcup_{j=1}^{k} (S_j)_i \tag{7}$$

Following this formula, W_{n+1} contains all the possible states where x can end up in on arriving at $a_{n+1} = \$$ in its runs. Since $t \notin W_{n+1}$, x has never ended up in state t (when at \$) which is the only accepting state. Hence x is rejected. $\square$

Solution: Part 2. We construct an NFA $N = (Q', \Sigma, \delta, q_0, F)$ for $\overline{L(A)}$ as follows:

$$\begin{aligned} Q' &= 2^Q \cup \{q_0\} \\ \delta(S, a) &= \{T \mid \exists\, x \in \Sigma^* : \exists\, 0 \le i \le |x| + 1 : S = W_i(x) \wedge T = W_{i+1}(x)\} \\ \delta(q_0, \epsilon) &= \{S \mid \exists\, x \in \Sigma^* : S = W_0(x)\} \\ F &= \{S \mid \exists\, x \in \Sigma^* : t \notin S \wedge S = W_{|x|+1}(x)\} \end{aligned} \tag{8}$$

It is easy to see that the above definition defined an NFA. In words, it effectively defines outward transitions for the right closure and incoming transitions for the left closure of the sets W_i defined in the previous part. We now prove the correctness formally.
On reading any word x, if it is not accepted by A, i.e., $x \notin L(A) \implies x \in \overline{L(A)}$, we can generate a sequence of states $W_0, W_1, \ldots, W_{|x|+1}$ such that $t \notin W_{|x|+1}$. From the definition of N, it is also easy to see that a run of N exists where we end up at state $W_{|x|+1}$ and this state $W_{|x|+1}$ will be in F (since $t \notin W_{|x|+1}$). Therefore, $x \in L(N)$. Also it is easy to see that any word x accepted by N will be in $\overline{L(A)}$.
Therefore, we have constructed an NFA N for $\overline{L(A)}$. Therefore, $\overline{L(A)}$. From the closure of regular languages under complement, $L(A)$ is also regular. Hence, proved. $\square$

6 Question 6

Question 6

Question. *Let $M = (Q, \Sigma, q_0, \delta, F)$ be a DFA and let h be a state of M called its "home". A synchronizing sequence for M and h is a string $s \in \Sigma^*$ where $\hat{\delta}(q, s) = h$ for every $q \in Q$. Say that M is synchronizable if it has a synchronizing sequence for some state h. Prove that if M is a k-state synchronizable DFA, then it has a synchronizing sequence of length at most k^3. Can you improve upon this bound?*

Solution. Given a DFA M such that M is a k-state synchronizable DFA i.e there exist a sequence s and state $h \in Q$ such that $\hat{\delta}(q, s) = h \ \forall q \in Q$.
Let's define a set

$$Q(v) = \{\hat{\delta}(q, v) \mid \forall \ q \ in \ Q\}$$

We need to find a sequence s to reduce the set states of size $|Q|$ to $\{h\}$ for some $h \in Q$.

Hence we will find word w such that $|Q(w)| < |Q|$ and for any s of length $< |w|$ $|Q(s)| = |Q|$. In worst case 1 pair at a time is synchronised then length of such a string can be at most $k(k-1)$ (maximum different pairs of state in a transition). Any string having length greater than $k(k-1)$ will have pair of states repeated and can be removed to find other string which will synchronize the pair of state, and hence the second condition will not hold true.

Starting from Q, there can be at most $k - 1$ such reductions. Thus length of sequence can be at most $k(k-1) * (k-1)$ i.e at most k^3. □

7 Question 7

Question 7

Question. *For every string $x \in \{0,1\}^+$ consider the number*

$$0.x = x[1] \cdot \frac{1}{2} + x[2] \cdot \frac{1}{2^2} + \cdots + x[|x|] \cdot \frac{1}{2^{|x|}}$$

where $|x|$ is the length of x. For a real number $\theta \in [0,1]$ let

$$L_\theta = \{x : 0.x \leq \theta\}$$

Prove that L_θ is regular if and only if θ is rational.

Solution. ($\Longleftarrow$) We construct the following DFA for any rational number $0.\alpha_1\alpha_2\ldots\alpha_k\overline{\beta_1\beta_2\ldots\beta_m}$. We know that every rational number can be represented this way since its binary representation is finite or infinte but recurring.
We define $D = (Q, \{0,1\}, \delta, q_1, F)$ as:

$$Q = \{q_1, q_2, \ldots, q_k, q_{k+1}, \ldots, q_{k+m}, q_{rej}, q_{acc}\}$$
$$\delta(q_i, a) = \begin{cases} q_{i+1}, & \text{if } a = \alpha_i, i \leq k \\ q_{acc}, & \text{if } a < \alpha_i, i \leq k \\ q_{rej}, & \text{if } a > \alpha_i, i \leq k \\ q_{k+1+(i-k-1 \mod m)}, & \text{if } a = \beta_{(i-k-1 \mod m)} \\ q_{acc}, & \text{if } a < \beta_{(i-k-1 \mod m)} \\ q_{rej}, & \text{if } a > \beta_{(i-k-1 \mod m)} \\ q_{acc}, & \text{if } i = acc \\ q_{rej}, & \text{if } i = rej \end{cases} \tag{9}$$
$$F = Q \setminus \{q_{rej}, q_1\}$$

It is easy to see that any number $0.x$ such that $0.x \leq \theta$ is accepted by D. This is because we move to the accepted as soon as any comparison is smaller and we move to the reject state as soon as any comparison is greater. For equality, we simply move to the next state and all such states are accepting.
Therefore, for any rational θ, we have constructed a DFA for L_θ. Therefore, L_θ is regular if theta is rational.
($\Longrightarrow$) We will assume that L_θ is regular for any irrational number θ. Therefore, by Myhil-Nehrode theorem, L_θ has finite equivalence classes. And since there are infinite words in L_θ, at least one of the equivalence class will have more than one word. Let those two words be x and y such that x is a prefix of y and $y = \theta_1\theta_2\ldots\theta_{|y|}$. Since θ is irrational, we know that there exists an index i such that $\theta_{|x|+i} \neq \theta_{|y|+i}$.
We prove the above claim by contradiction. Assume that $\theta_{|x|+i} = \theta_{|y|+i}$ for all i. Then, θ has a recurring expansion with the recurring part being $\theta_{|x|}\theta_{|x|+1}\theta_{|y|-1}$. This will be a contradiction to the irrationality of θ.

We now consider the smallest such i in the following equation:

$$\begin{aligned} &[x] \equiv [y] \\ &\implies [x\theta_{|x|+1}\theta_{|x|+2}\dots\theta_{|x|+i-1}1] \equiv [y\theta_{|x|+1}\theta_{|x|+2}\dots\theta_{|x|+i-1}1] \\ &= s_x \equiv s_y \text{ (say)} \end{aligned} \tag{10}$$

Now, since $\theta_{|x|+i} \neq \theta_{|y|+i}$, exactly one of them will be equal to 1. Therefore, one of $x\theta_{|x|+1}\theta_{|x|+2}\dots\theta_{|x|+i-1}1$ and $y\theta_{|x|+1}\theta_{|x|+2}\dots\theta_{|x|+i-1}1$ will be greater than θ and the other will be accepted and be present in L_θ. However, both of s_x and s_y are in the same equivalence classes. This is a contradiction the the Myhil-Nehrode theorem. Therefore, L_θ cannot be regular for an irrational θ. Hence, proved. $\square$