Answer key to Mid-Semester Examination - Even Semester 2020-2021
CSO 322 - Theory of Computation
Department of Mathematical Sciences
Indian Institute of Technology (BHU) Varanasi

1. Let us say that a string x is obtained from a string w by ***deleting symbols*** if it is possible to remove zero or more symbols from w so that just the string x remains. For example, the following strings can all be obtained from 0110 by deleting symbols: $\epsilon, 0, 1, 00, 01, 10, 11, 010, 011, 110$, and 0110.

   Let $\Sigma = \{0,1\}$, let $A \subseteq \Sigma^*$ be a regular language, and define

   $$B = \{x \in \Sigma^* : \text{ there exists a string } w \in A \text{ such that } x \text{ is obtained from } w \text{ by deleting symbols}\}.$$

   Prove or Disprove: B is regular. **(3 marks)**

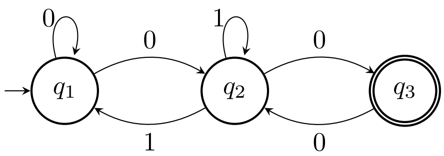
   Solution. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA such that $L(M) = A$. Define an NFA $N = (Q, \Sigma, \eta, q_0, F)$ as follows:

   $$\eta(q,a) = \{\delta(q,a)\} \quad \text{and} \quad \eta(q,\varepsilon) = \{\delta(q,b) : b \in \Sigma\} \quad (\text{for all } q \in Q \text{ and } a \in \Sigma).$$

   In words, N is similar to M, but for every transition in M we include in N the same transition as well as an ε-transition between the same pair of states.

   It is the case that $L(N) = B$, and therefore B is regular.

2. Solve by Arden's method to find a regular expression for the FA: **(3 marks)**



   Solution: By Arden's theorem,

   $$\begin{aligned} q_1 &= q_1 \cdot 0 + q_2 \cdot 1 + \epsilon \\ q_2 &= q_1 \cdot 0 + q_2 \cdot 1 + q_3 \cdot 0 \\ q_3 &= q_2 \cdot 0 \end{aligned}$$

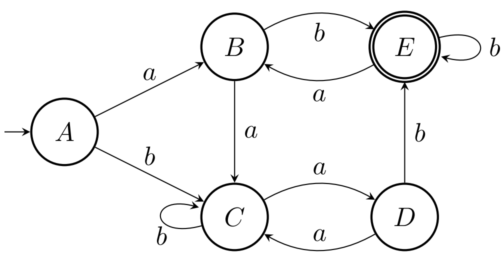
   Upon solving the equations,

   $$\begin{aligned} q_1 &= [0 + 0(1+00)^*1]^* \\ q_2 &= [0 + 0(1+00)^*1]^*0(1+00)^* \\ q_3 &= [0 + 0(1+00)^*1]^*0(1+00)^*0 \end{aligned}$$

   Solution is $[0 + 0(1+00)^*1]^*0(1+00)^*0$.

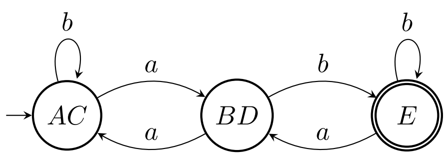
   (1 mark for the equations, 1 mark for solving them and 1 for the final answer.)

3. Minimize using Myhill Nerode theorem. Also, find the language recognized by the FA. **(3 marks)**



B	x	-	-	-
C	o	x	-	-
D	x	o	x	-
E	X	X	X	X
	A	B	C	D

The equivalence classes are $\{A, C\}$, $\{B, D\}$ and $\{E\}$



(1 mark for the matrix table, 1 mark for identifying the equivalence classes and 1 mark for the minimized DFA)

4. Construct a a regular grammar for the automata recognizing the regular expression $R = 1(1+10)^* + 10(0+01)^*$ over $\Sigma = \{0, 1\}$. **(3 marks)**

Solution:

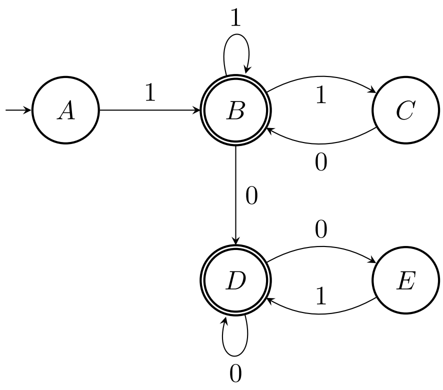


Figure 1: DFA for R given in Qn. 4

The grammar with production rules for R is

$$\begin{aligned} A &\rightarrow 1B \mid 1; \\ B &\rightarrow 1B \mid 1C \mid 0D \mid 1 \mid 0; \\ C &\rightarrow 0B \mid 0; \\ D &\rightarrow 0D \mid 0E \mid 0; \\ E &\rightarrow 1D \mid 1; \end{aligned}$$

(1.5 for the DFA and 1.5 for the grammar)

5. Using Kleene's theorem construct an ϵ-NFA to accept the regular expression $R = 0(11 + 0(00+1)^*)^*$ and find an equivalent NFA/DFA by eliminating ϵ-transitions. **(3 marks)**

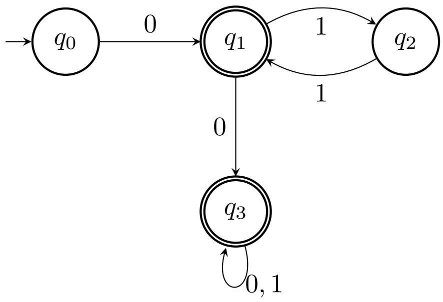


Figure 2: DFA/NFA for R given in Qn. 5

(1 mark for the intermediate steps in applying Klein's theorem and 2marks for eliminating ϵ and the final DFA/NFA without ϵ moves.)

6. Let $\Sigma = \{0, 1\}$, and define a language

$$MIDDLE = \{u0v : u, v \in \Sigma^* \text{ and } |u| = |v|\}.$$

In words, MIDDLE is the language of all binary strings of odd length whose middle symbol is 0. Prove that MIDDLE is not regular. **(3 marks)**

Pumping lemma. For every alphabet Σ and every regular language $A \subseteq \Sigma^*$, there exists a pumping length $n \geq 1$ for A that satisfies the following property. For every string $w \in A$ with $|w| \geq n$, it is possible to write $w = xyz$ for strings $x, y, z \in \Sigma^*$ such that

1. $y \neq \varepsilon$,
2. $|xy| \leq n$, and
3. $xy^iz \in A$ for every $i \in \mathbb{N}$.

Solution. Assume toward contradiction that MIDDLE is regular. By the pumping lemma, there exists a pumping length $n \geq 1$ for MIDDLE satisfying the property expressed in that lemma.

Define $w = 1^n01^n$. It is the case that $w \in$ MIDDLE and $|w| = 2n + 1 \geq n$. Therefore, by the pumping lemma, it is possible to write $w = xyz$ for strings $x, y, z \in \Sigma^*$ such that (i) $y \neq \varepsilon$, (ii) $|xy| \leq n$, and (iii) $xy^iz \in$ MIDDLE for all $i \in \mathbb{N}$.

Because $w = xyz = 1^n01^n$ and $|xy| \leq n$, it must be the case that $y = 1^k$ for some natural number $k \in \mathbb{N}$, as the prefix xy of w is not long enough to reach the single 0 that string contains. Moreover, because $y \neq \varepsilon$ we conclude that $k \geq 1$. Setting $i = 2$, we find from the third condition on x, y, and z that

$$xy^2z = xyyz = 1^{n+k}01^n \in \text{MIDDLE}.$$

However, because $k \geq 1$, the string $1^{n+k}01^n$ is clearly not in MIDDLE; either its length is even (in case k is odd) or it has a 1 rather than a 0 in its middle position (in case k is even).

Having obtained a contradiction, we conclude that MIDDLE is not regular.

7. Give a context-free grammar (in its Chomsky normal form) for each of the following languages: **(4 marks)**

(a) $A = \{w \in \{0,1\}^* : w = w^R \text{ and } |w| \text{ is divisible by } 3\}$.
Solution: Here is a CFG for A:

$$\begin{aligned} S_0 &\rightarrow 0S_00 \mid 1S_11 \mid 0 \mid \epsilon \\ S_1 &\rightarrow 0S_10 \mid 1S_21 \mid 1 \\ S_2 &\rightarrow 0S_20 \mid 1S_01 \end{aligned}$$

Converting this to CNF, we have: S_0 is nullable.

$$\begin{aligned} S_0 &\rightarrow X_0Y_0 \mid X_1Y_1 \mid X_0X_0 \mid 0 \\ S_1 &\rightarrow X_0Y_2 \mid X_1Y_3 \mid 1 \\ S_2 &\rightarrow X_0Y_4 \mid X_1Y_5 \mid X_1X_1 \\ Y_0 &\rightarrow S_0X_0; \\ Y_1 &\rightarrow S_1X_1; \\ Y_2 &\rightarrow S_1X_0; \\ Y_3 &\rightarrow S_2X_1; \\ Y_4 &\rightarrow S_2X_0; \\ Y_5 &\rightarrow S_0X_1; \\ X_0 &\rightarrow 0; \\ X_1 &\rightarrow 1; \end{aligned}$$

(1 mark for the CFG and 1 for the CNF)

(b) $C = \{0^n1^m : m \leq n \leq 3m\}$
Solution: The CFG for the language C is:

$$\begin{aligned} S &\rightarrow XS1 \mid \epsilon \\ X &\rightarrow 0 \mid 00 \mid 000 \end{aligned}$$

Converting this to a CNF, we have: S is nullable.

$$\begin{aligned} S &\rightarrow XU \mid XZ; \\ U &\rightarrow SZ; \\ X &\rightarrow 0 \mid YY \mid YW; \\ W &\rightarrow YY; \\ Y &\rightarrow 0; \\ Z &\rightarrow 1; \end{aligned}$$

(1 mark for the CFG and 1 for the CNF)

8. Let $\Sigma = \{0, 1\}$ and let $C \subseteq \Sigma^*$ be a given context-free language.

   Prove that the language **(3 marks)**

$$A = \{w : 0w0 \in C\}$$

is context-free. In words, A is the language of all strings that can be obtained by taking any string in C that starts and ends with 0 (and has length at least 2), and then deleting the starting and ending 0.

Solution. If the language C is empty, then A is also empty, and therefore context-free. We will therefore assume that C is non-empty for the remainder of the solution. Let G be a context-free grammar in Chomsky normal form that generates C, and assume without loss of generality that G has no useless variables (i.e., variables that generate no strings). We will refer to the grammar G in both parts of the solution.

We will first define a context-free grammar H that generates A as follows:

- For each variable X that appears in G, we will have four variables in H:

$$X, X_L, X_R, \text{ and } X_{LR}.$$

  (Naturally, this assumes that none of the variables X_L, X_R, and X_{LR} appears in G.)

  The meaning of these variables is as follows:

  - X will generate exactly the same strings in H that it generates in G.
  - X_L will generate those strings that result from taking any string that could be generated from X in G, except that a 0 is removed from the left-hand side of that string.
  - X_R is similar to X_L, except that a 0 is removed from the right-hand side of the string rather than the left.
  - X_{LR} generates strings that can be obtained by removing a 0 from both the left-hand side and right-hand side of a string generated by X in G.

  The start variable of H is S_{LR}, for S being the start variable of G.

- For every rule of the form $X \rightarrow YZ$ in G, include these rules in H:

$$\begin{gathered} X \rightarrow YZ \\ X_L \rightarrow Y_L Z \\ X_R \rightarrow Y Z_R \\ X_{LR} \rightarrow Y_L Z_R \end{gathered}$$

- For every rule of the form $X \dashrightarrow 0$ in G, include these rules in H:

$$
\begin{aligned}
X &\rightarrow 0 \\
X_L &\rightarrow \varepsilon \\
X_R &\rightarrow \varepsilon
\end{aligned}
$$

- For every rule of the form $X \dashrightarrow 1$ in G, include just this rule in H:

$$X \dashrightarrow 1.$$

- If the rule $S \dashrightarrow \varepsilon$ appears in G, ignore it.

It is the case that $L(H) = A$, and therefore A is context-free.

(1 mark for the grammar and 2 marks for the proof that it is CFG)