1. **[1 mark]** Consider the following statement about a language $L \subseteq \Sigma^*$.

   If $\forall w \in L$, $\exists$ a DFA M accepting w, then L is regular.

   Is the statement correct? Justify your answer.

   Solution: The statement is incorrect.

   This statement holds for every language L (including for non-regular languages) since for each $w \in L$, we can construct a DFA that accepts only w and nothing else.

2. **[1 mark]** Let $L \subseteq \Sigma^*$ be a regular language for some finite alphabet Σ. Give one condition that is *both necessary and sufficient* for L to **not contain** a non-regular language as a subset. Justify your answer.

   Solution: The condition is that L **should be finite**.

   If L is finite, then every subset is finite and hence all subsets are regular.

   If L is infinite, then since L is countable, the power set of L is uncountable. But, there are only countably many regular languages.

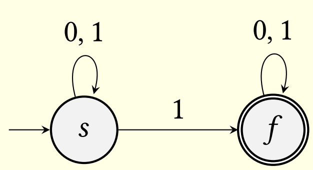
3. **[2 marks]** While writing his examination for CS2200, a student made the following statement in one of his answers.

   If $N = (Q, \Sigma, \Delta, q_0, F)$ is an NFA accepting L, then $N' = (Q, \Sigma, \Delta, q_0, Q - F)$ is an NFA accepting $\overline{L}$.

   Help the student write a formal proof for the statement. If you think the statement is incorrect, give a concrete justification to convince the student why the statement is false.

   Solution: The statement is incorrect.

   Consider the following NFA.

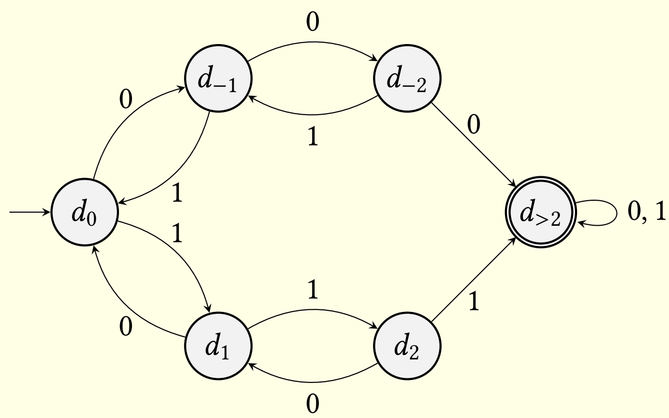


   This NFA accepts the string $w = 1$. If you complement the set of final states, then s becomes a final state, and the NFA N' still accepts $w = 1$. Thus $L(N') \neq \overline{L(N)}$.

4. **[3 marks]** Draw a DFA/NFA accepting $L = \{w \in \{0,1\}^* \mid \forall x, y \in \{0,1\}^* \text{ s.t } w = xy, |\#1(x) - \#0(x)| \leq 2\}$, where $\#1(w)$ and $\#0(w)$ are the number of 1s and 0s in w, respectively. Justify your answer.

   Solution: Let's look at the complement of this language - $\overline{L} = \{w \in \{0,1\}^* \mid \exists x, y \in \{0,1\}^* \text{ s.t } w = xy, |\#1(x) - \#0(x)| > 2\}$.

   These are precisely the set of strings w such that w has a prefix that has $|\#1(x) - \#0(x)| > 2$. The following DFA accepts precisely these strings. The states d_i is reached if for the current string x, $\#1(x) - \#0(x) = i$. The state $d_{>2}$ is the final state if the current string x has at least two 0s more than 1s, or two 1s more than 0s.



5. **[4 marks]** Consider the following languages over $\Sigma = \{0, 1\}$.

$$L_1 = \{accb \mid a, b, c \text{ are non-empty strings}\},$$
$$L_2 = \{cabc \mid a, b, c \text{ are non-empty strings}\}$$

Write down the correct option from the choices below and justify your answer.

A. Both L_1 and L_2 are regular

B. L_1 is regular and L_2 is non-regular

C. L_1 is non-regular and L_2 is regular

D. Both L_1 and L_2 are non-regular

Solution: The correct option is B.

L_1 is regular: If the length of a string w is less than 4, then $w \notin L_1$. For every string of length 4 or 5, we can explicitly verify the ones that are in L_1, and there are only finitely many of them.

We will now show that every string of length ≥ 6 must be in L_1. Firstly, if the string w contains 00 or 11 as a substring, we can choose $c = 0$ or $c = 1$, respectively and the corresponding prefix and suffix as a and b.

Otherwise w has length ≥ 6 and is an alternating string of 0s and 1. Thus, w will contain a substring of the form 0101 or 1010 and a non-zero length prefix and suffix which correspond to a and b. The string 0101 can be thought of as cc where $c = 01$. Similarly for 1010. Thus L_1 is precisely the set of all strings of length at least 6, and some fixed strings of length 4 and 5. Hence L_1 is regular.

L_2 is not regular: Suppose that the prover chooses $k > 0$. The spoiler will choose $x = 10^{k+1}111$, $y = 0^{k+1}$, and $z = \varepsilon$. Now, $xyz = 10^{k+1}1110^{k+1} \in L$. Now, for every choice of $u = 0^l$, $v = 0^m$, and $w = 0^n$ such that $l + m + n = k + 1$ and $m > 0$, choose $i = 2$. This gives $xuwz = 10^{k+1}1110^{k+1+m}$, and $10^{k+1}1110^{k+1+m} \notin L_2$.

6. For a string string w, $|w|$ denotes the number of characters in the string w. For $w = \varepsilon$, $|w| = 0$. Define the language $\text{count}(L) = \{0^{|w|} \mid w \in L\}$.

   (a) **[3 marks]** Show that if L is regular, then $\text{count}(L)$ is regular. Give a **clear description of your NFA/DFA** and **prove the correctness of the construction**.

   Solution: The idea is to take the DFA for M and convert all the edge labels to 0.

   Mention clearly the NFA corresponding to $\text{count}(L)$ and a proof of correctness of the construction, as shown below.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA accepting L. Consider the NFA $N = (Q, \{0\}, \Delta, q_0, F)$ defined as follows - note that the set of the states, start state, and the set of final states of N are same as M. The function Δ is defined as follows: $\Delta(q, 0) = \{\delta(q, \sigma) \mid \sigma \in \Sigma\}$.

Now, we need to show that $L(N) = \text{count}(L)$.

Suppose that $z \in L(N)$. Then $z = 0^k$ for some k, and $\widehat{\Delta}(q_0, z) \cap F \neq \emptyset$. Thus, there is a path of length k in the NFA from q_0 to some $q_f \in F$ where all the edges are labelled 0. From the definition of Δ, there exists symbols $\sigma_1, \sigma_2, \ldots, \sigma_k$ such that $\widehat{\delta}(q_0, \sigma_1\sigma_2 \ldots \sigma_k) = q_f$. Thus, $\exists w \in L$ such that $|w| = k$. Therefore $0^k \in \text{count}(L)$.

If $z = 0^k \in \text{count}(L)$. Then $\exists w$ such that $|w| = |z|$ and $w \in L$. Thus there is a path in the DFA for L that is labeled by the symbols of w such that $\widehat{\delta}(q_0, w) \in F$. From the construction, this path remains in the NFA and is labeled by 0s. Hence $\widehat{\delta}(q_0, 0^k) \cap F \neq \emptyset$ and therefore, $z = 0^k \in L(N)$.

This next part is not necessary to be written in the exam. It is merely for illustrating the induction proof.

We will prove the following stronger claim (like we did in class).

$$q' \in \widehat{\Delta}(q_0, 0^k) \Leftrightarrow \exists w \text{ s.t } |w| = k, \text{ and } \widehat{\delta}(q_0, w) = q'.$$

Observe that this suffices to prove the correctness of the construction since this statement implies that

$$\begin{aligned} 0^k \in L(N) &\Leftrightarrow \exists q' \in F \text{ s.t } q' \in \widehat{\Delta}(q_0, 0^k), \text{ (defn of NFA acceptance)} \\ &\Leftrightarrow \exists w \text{ s.t } |w| = k, \text{ and } \widehat{\delta}(q_0, w) = q' \in F, \text{ (from the claim above)} \\ &\Leftrightarrow w \in L \Leftrightarrow 0^{|w|} = 0^k \in \text{count}(L). \end{aligned}$$

The statement can be proved by induction on k. For $k = 1$, the statement follows from the definition of Δ. Suppose that the statement is true for $k = n$. Let $k = n + 1$. Then we can write

$$\widehat{\Delta}(q_0, 0^{n+1}) = \bigcup_{q \in \widehat{\Delta}(q_0, 0^n)} \Delta(q, 0)$$

Thus, we can say that

$$\begin{aligned} q' \in \widehat{\Delta}(q_0, 0^{n+1}) &\Leftrightarrow \exists q \in \widehat{\Delta}(q_0, 0^n), \text{ and } q' \in \Delta(q, 0), \text{ (follows from the previous equation)} \\ &\Leftrightarrow \exists w \text{ s.t } |w| = n, \text{ and } \widehat{\delta}(q_0, w) = q, \text{ and } \exists \sigma \in \Sigma, q' = \delta(q, \sigma) \text{ (induction hypothesis)} \\ &\Leftrightarrow \exists w\sigma \text{ s.t } |w| = n, \sigma \in \Sigma, \text{ and } \delta(\widehat{\delta}(q_0, w), \sigma) = \widehat{\delta}(q_0, w\sigma) = q' \end{aligned}$$

(b) **[1 mark]** Is the converse of the statement in Part (a) also true? Justify your answer.

Solution: The converse is false.

Let $L = \{0^n 1^n \mid n \geq 0\}$. We know that L is non-regular. But, $\text{count}(L) = \{0^{2n} \mid n \geq 0\}$ which is regular.