Midterm Exam 1 Solutions

1. (a) Give a context-free grammar for the language

$$A = \{0^n 1^m \,:\, n, m \in \mathbb{N},\ n \neq m\}\,.$$

Solution. There are many correct grammars, including this one:

$$\begin{aligned} S &\rightarrow 0\,S\,1 \mid 0\,X \mid 1\,Y \\ X &\rightarrow 0\,X \mid \varepsilon \\ Y &\rightarrow 1\,Y \mid \varepsilon \end{aligned}$$

2. (b) Let

$$B = \{0^n 1^n 2^n \,:\, n \in \mathbb{N}\}.$$

We proved in lecture that B is not context-free. Prove that $\overline{B}$ (i.e., the complement of the language B over the alphabet $\Sigma = \{0, 1, 2\}$) is context-free.

If your answer includes the specification of one or more context-free grammars or PDAs, you do not need to include formal proofs that the grammars (or PDAs) generate (or accept) the languages you claim they do.

Solution. First let us prove that the language

$$C = \left\{0^k 1^l 2^m \,:\, k \neq l \text{ or } l \neq m\right\}$$

is context-free. This follows from the fact that C is generated by the following CFG:

$$\begin{aligned} S &\rightarrow A\,Z \mid X\,B \\ A &\rightarrow 0\,A\,1 \mid 0\,X \mid 1\,Y \\ B &\rightarrow 1\,B\,2 \mid 1\,Y \mid 2\,Z \\ X &\rightarrow 0\,X \mid \varepsilon \\ Y &\rightarrow 1\,Y \mid \varepsilon \\ Z &\rightarrow 2\,Z \mid \varepsilon \end{aligned}$$

The fact that $\overline{B}$ is context-free now follows from the expression

$$\overline{B} = C \cup L\left((0 \cup 1 \cup 2)^*(10 \cup 20 \cup 21)(0 \cup 1 \cup 2)^*\right),$$

along with the fact that (1) every regular language is context-free, and (2) the union of any two context-free languages is context-free.

3. Suppose that an infinite subset $S \subseteq \mathbb{N}$ of the natural numbers is given, and define

$$C = \{0^n 1^m \,:\, n \in S \text{ and } m \leq n\}\,.$$

Prove that C is not regular.

Solution. Assume toward contradiction that C is regular. By the pumping lemma for regular languages, there exists a pumping length $p \geq 1$ for C. As S is infinite, we may choose a natural number $n \in S$ with $n \geq p$. Now define

$$w = 0^n 1^n.$$

It holds that $w \in C$ and $|w| = 2n \geq p$, so by the pumping lemma we may write $w = xyz$ for $x, y, z \in \{0,1\}^*$ such that $y \neq \varepsilon$, $|xy| \leq p$, and $xy^iz \in C$ for all $i \in \mathbb{N}$. As $y \neq \varepsilon$ and $|xy| \leq p \leq n$ it follows that $y = 0^k$ for $1 \leq k \leq p$. Taking $i = 0$ we have

$$xy^iz = xz = 0^{n-k}1^n,$$

which is not an element of C. We have reached a contradiction, which implies that C is not regular.

4. Let $\Sigma = \{0,1\}$ and let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA, where $Q = \{q_0, \ldots, q_{n-1}\}$ for some positive integer n. Suppose further that an NFA N is formed by adding a (non-accepting) state q_n to M, along with any number of transitions to or from q_n.

   In more formal terms, you are to assume that $N = (Q \cup \{q_n\}, \Sigma, \mu, q_0, F)$ for some choice of a transition function

   $$\mu : (Q \cup \{q_n\}) \times (\Sigma \cup \{\varepsilon\}) \to \mathcal{P}(Q \cup \{q_n\})$$

   that satisfies these properties:

   (a) $\mu(q_j, \sigma) = \{\delta(q_j, \sigma)\}$ or $\mu(q_j, \sigma) = \{\delta(q_j, \sigma), q_n\}$ for each $j \in \{0, \ldots, n-1\}$ and $\sigma \in \Sigma$,
   (b) $\mu(q_j, \varepsilon) = \varnothing$ or $\mu(q_j, \varepsilon) = \{q_n\}$ for each $j \in \{0, \ldots, n-1\}$, and
   (c) $\mu(q_n, \sigma)$ is an arbitrary subset of $Q \cup \{q_n\}$ for each $\sigma \in \Sigma \cup \{\varepsilon\}$.

   Decide whether each of the following statements is **true** or **false**, and defend each answer with a short, high-level proof sketch or a counter-example.

   (a) $L(M) \subseteq L(N)$.

Solution. True. For any $x \in L(M)$, we have that x is accepted by N by precisely the same sequence of states that accepts x in M.

   (b) The language of all strings that are accepted by N, but not accepted by M, is regular.

Solution. True. The languages $L(M)$ and $L(N)$ are regular, and the regular languages are closed under intersection and complementation. Therefore the language of all strings that are accepted by N, but not accepted by M, i.e.,

$$L(N) \cap \overline{L(M)}$$

is regular.

   (c) The language of all strings x that are accepted by a computation of N that visits the state q_n at least once (i.e., for which there exists a sequence of states that includes q_n and satisfies the definition for acceptance of x by N) is regular.

Solution. True. Define two NFAs as follows:

$$K_1 = (Q \cup \{q_n\}, \Sigma, \mu, q_0, \{q_n\})$$
$$K_2 = (Q \cup \{q_n\}, \Sigma, \mu, q_n, F).$$

The NFA K_1 accepts all strings that take N from its start state to the state q_n, while K_2 accepts all strings that take N from q_n to an accept state. The language described in the question is

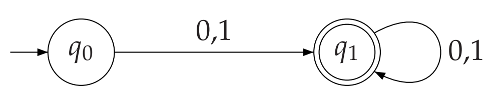
$$L(K_1)L(K_2),$$

which is regular given that $L(K_1)$ and $L(K_2)$ are regular and the regular languages are closed under concatenation.

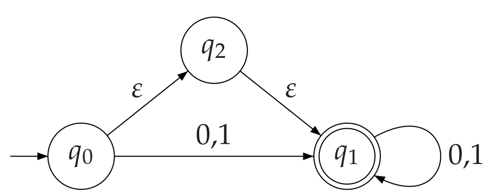
(d) The minimal DFA for $L(N)$ is at least as large as the minimal DFA for $L(M)$.

(Recall that the minimal DFA for any regular language is the unique DFA that has the smallest possible number of states among all the DFAs for that language.)

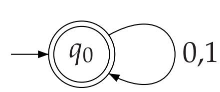
Solution. False. Let M be the following DFA, which is obviously minimal:



Now take N as follows, which is a valid choice of N with respect to the constraints imposed by the problem:



Then the minimal DFA for $L(N)$ has just one state, which is smaller than M:



5. Recall that a context-free grammar G over an alphabet Σ is in *Chomsky Normal Form* if each of its rules has one of the following forms:

   (a) $S \to \varepsilon$ (where S is the start variable)
   (b) $X \to YZ$ (where X, Y and Z are variables, and neither Y nor Z is the start variable)
   (c) $X \to \sigma$ (where X is a variable and $\sigma \in \Sigma$)

Let $\Sigma = \{0,1\}$, suppose that $D \subseteq \Sigma^*$ is a context-free language, and assume G is a context-free grammar for D that is in Chomsky Normal Form.

Describe how you could obtain from G a new grammar H (not necessarily in Chomsky Normal Form) for the language

$$E = \{x \in \Sigma^* \,:\, xy \in D \text{ for some } y \in \Sigma^*\}\,.$$

You do not need to prove that the grammar H generates E—just describe how to obtain H from G.

Solution. For every variable X in G, the grammar H will have two variables: X and X_0. We view that X_0 generates all the prefixes of strings generated by X. Now, for every rule of the form $X \to YZ$ in G, add these rules to H:

$$\begin{aligned} X &\to YZ \\ X_0 &\to YZ_0 \mid Y_0 \end{aligned}$$

For every rule $X \to \sigma$ in G, add these rules to H:

$$\begin{aligned} X &\to \sigma \\ X_0 &\to \sigma \mid \varepsilon \end{aligned}$$

If the rule $S \to \varepsilon$ appears in G, then add

$$S_0 \to \varepsilon$$

to H. (This is really only necessary to handle certain grammars that generate the language $\{\varepsilon\}$.) Finally, let S_0 be the start variable of H.