**CS 341-007, Fall 2022, Face-to-Face Section**
**Solutions for Midterm 1**

1. Multiple choice.

1.1. Answer: (c).

- HW 6, problem 2a, shows that the class of CFLs is not closed under intersection, so (a) is incorrect.
- HW 6, problem 2b, shows that the class of CFLs is not closed under complementation, so (b) is incorrect.
- HW 5, problem 3b, shows that (c) is correct.
- The language $\{ a^n b^n c^n \mid n \geq 0 \}$ is not context-free by slide 2-96, so not all languages are context-free, so (d) is incorrect.
- By Corollary 2.32, every regular language is also context-free, so (e) is incorrect.

1.2. Answer: (a).

- By Corollary 2.32, every regular language (even if it is infinite) is also context-free, so (a) is correct, and (b) and (c) are incorrect.
- Theorem 1.25 implies that $L_1 \cup L_2$ must be regular, so (d) is incorrect.

1.3. Answer: (c).

- The languages $L_1 = \{ a^n b^n c^n \mid n \geq 0 \}$ and $L_2 = \{ b^n a^n c^n \mid n \geq 0 \}$ are infinite, non-regular languages, with $L_1 \cap L_2 = \{\varepsilon\}$, which is regular because it is finite (slide 1-95), making (a) incorrect.
- If $L_1 = L_2 = \{ a^n b^n c^n \mid n \geq 0 \}$, then $L_1 \cap L_2 = L_1$, which is non-regular and non-context-free, so (b) and (d) are incorrect.

1.4. Answer: (e).

- The language $L_1 = \{ a^n b^n c^n \mid n \geq 0 \}$ is infinite, not regular, and not context-free, so (a), (b), and (c) are incorrect.
- For the same language $L_1$, note that $x = abc \in L_1$ and $y = aabbcc \in L_1$, but $xy = abcaabbcc \notin L_1$, so $L_1$ is not closed under Kleene star, making (d) incorrect.

1.5. Answer: (e).

- The regular expression $b^*a^*$ generates the string $bba \notin A$, so (a) is incorrect.
- The regular expression $(ba)^*$ generates the string $baba \notin A$, so (b) is incorrect.
- The given CFG $G$ in part (c) has language $L(G) = \emptyset$ (i.e., no strings at all) because derivations can never terminate: $S \Rightarrow bSa \Rightarrow bbSaa \Rightarrow bbbSaaa \Rightarrow \cdots$, so (c) is incorrect.
- The language $A$ has CFG with rules $S \rightarrow bSa \mid \varepsilon$, so (d) is incorrect.

1.6. Answer: (d).

- The language $A = \{ a^n b^n c^n \mid n \geq 0 \}$ is non-context-free and infinite, so (a) is incorrect. In fact, if $A$ is non-context-free language, $A$ must be infinite.
- $A = \{ a^n b^n c^n \mid n \geq 0 \}$ is also nonregular, so (b) is incorrect.
- $A = \{ a^n b^n c^n \mid n \geq 0 \}$ is non-context-free, and $abc \in A$ but $(abc)^{\mathcal{R}} = cba \notin A$, so $A$ is not closed under reversals, making (c) incorrect.

1.7. Answer: (c).

- By Theorem 2.20, a language is context-free if and only if some PDA recognizes it, so we can answer the question by considering CFLs. The language $\{ a^n b^n \mid n \geq 0 \}$ is context-free but infinite, so (a) is incorrect.
- The language $\{\varepsilon\}$ is finite, so it is regular (slide 1-95), and Corollary 2.32 ensures it is also context-free, so (b) is incorrect.
- Theorem 2.9 guarantees that a context-free language has a CFG in Chomsky normal form, so (d) is incorrect.

1.8. Answer: (c).

- Kleene's Theorem (Theorem 1.54) implies that $L$ must be regular, so (a) is incorrect.
- Because $L$ must be regular, Corollary 2.32 ensures $L$ is also context-free, so (c) is correct and (b) is incorrect.
- For the language $L$ with regular expression $ab^*$, we have that $x = ab \in L$ and $y = abb \in L$, but $xy = ababb \notin L$, so $L$ is not closed under concatenation, making (d) incorrect.

1.9. Answer: (b).

- HW 4, problem 5c, shows that (a) is incorrect, and that (b) is correct.
- Slightly modifying the proof on slide 1-105 shows that the language $L_1 = \{ a^n b^n \mid n \geq 1 \}$ is non-regular. Adding $\varepsilon$ to $L_1$ leads to the language $L_2 = \{ a^n b^n \mid n \geq 0 \}$, which is context-free (with CFG having rules $S \to aSb \mid \varepsilon$), so (c) is incorrect.
- Slightly modifying the proof on slide 2-96 shows that the language $L_1 = \{ a^n b^n c^n \mid n \geq 1 \}$ is non-context-free, so it is also non-regular by Corollary 2.32. Adding $\varepsilon$ to $L_1$ leads to the language $L_2 = \{ a^n b^n c^n \mid n \geq 0 \}$, which is non-context-free by the proof on slide 2-96, so (d) is incorrect.

1.10. Answer: (h).

- The regular expression $01^*0$ cannot generate the string $11 \in L$, so (i) is incorrect.
- The regular expression $0(0\cup 1)^*0 \cup 1(0\cup 1)^*1$ cannot generate the string $1 \in L$, so (ii) is incorrect.
- The regular expression $(0\cup 1)(0\cup 1)^*(0\cup 1)$ cannot generate the string $1 \in L$, so (iii) is incorrect.

2. (a) $(a \cup b)^*(a \cup b)(a \cup b)a(a \cup b)$

   There are infinitely many other correct regular expressions for this language, such

as $(a \cup b)(a \cup b)^*(a \cup b)a(a \cup b)$
or $(a \cup b)(a \cup b)(a \cup b)^*a(a \cup b)$
or $a(a \cup b)(a \cup b)^*a(a \cup b) \cup b(a \cup b)(a \cup b)^*a(a \cup b)$ or ...

(b) $(aa \cup b)b^*aa^*$. Another regular expression is $(aab^* \cup bb^*)aa^*$. There are infinitely many other correct regular expressions for this language.

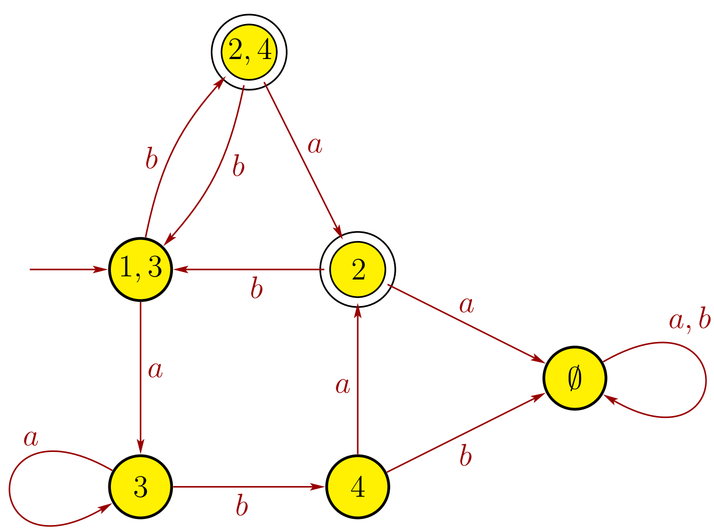
(c) After one step, the CFG is then

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow 1SA0A \mid 1S0A \mid 1SA0 \mid 1S0 \mid 0AS1S \mid 0S1S \mid \varepsilon \\ A &\rightarrow 10S1 \end{aligned}$$

(d) (Homework 2, problem 5.) Given a DFA $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ for language $A_1$ and a DFA $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$ for language $A_2$, the language $A_3 = A_1 \cap A_2$ is recognized by the DFA $M_3 = (Q_3, \Sigma, \delta_3, q_3, F_3)$, with

- $Q_3 = Q_1 \times Q_2$,
- $\delta_3((x, y), \ell) = (\delta_1(x, \ell), \delta_2(y, \ell))$ for $(x, y) \in Q_3$ and $\ell \in \Sigma$,
- $q_3 = (q_1, q_2)$, and
- $F_3 = F_1 \times F_2$.

3. A DFA for $C$ is below:

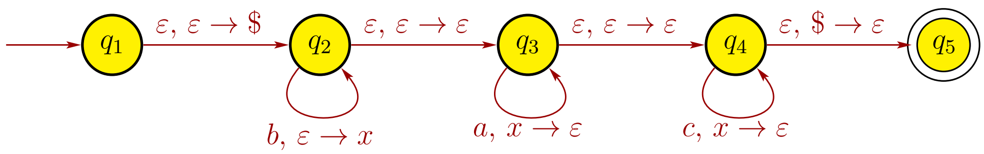



4. (a) $G = (V, \Sigma, R, S)$ with set of variables $V = \{S, X\}$, where $S$ is the start variable; set of terminals $\Sigma = \{a, b, c\}$; and rules

$$\begin{aligned} S &\rightarrow bSc \mid X \\ X &\rightarrow bXa \mid \varepsilon \end{aligned}$$

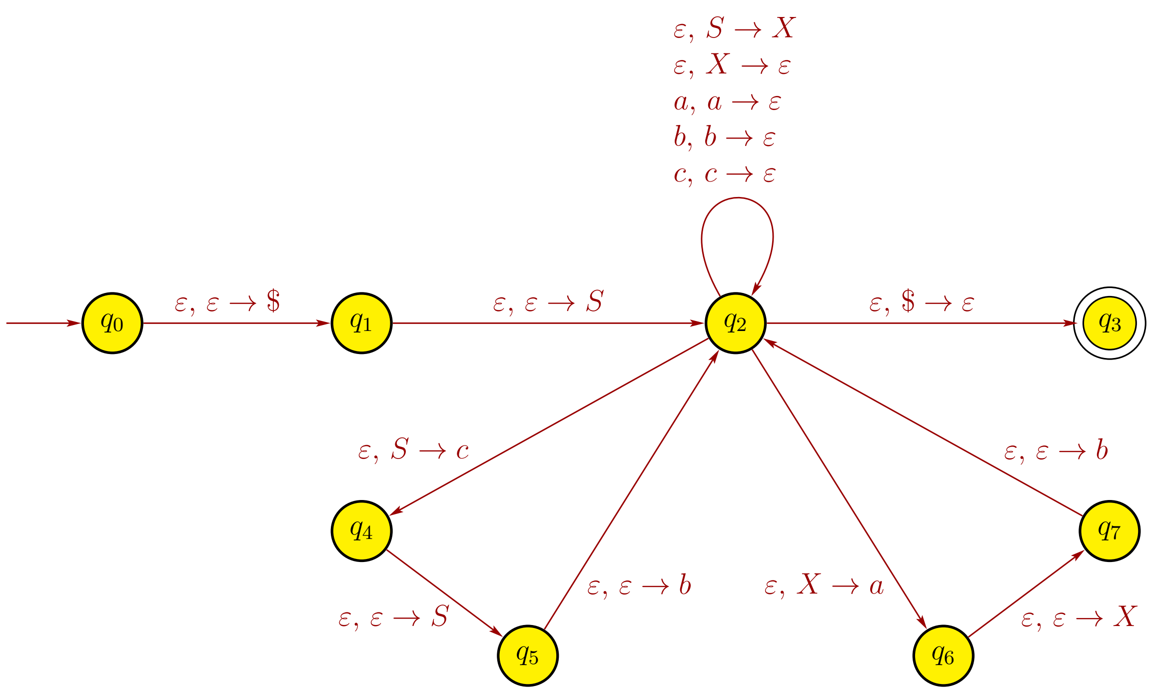
There are infinitely many other correct CFGs for $L$.

(b) There are infinitely many correct PDAs for $L$. Here is one:




In the above PDA, state $q_2$ pushes an $x$ for each $b$ read, state $q_3$ pops an $x$ for each $a$ read, and state $q_4$ pops an $x$ for each $c$ read to match the $c$'s and $a$'s with the $b$'s.

Another approach uses the algorithm from Lemma 2.21 to convert the CFG in part (a) into a PDA.




In the second PDA,

- The path $q_2 \to q_4 \to q_5 \to q_2$ corresponds to the rule $S \to bSc$, where the symbols on the right side of the rule are pushed in reverse order.
- The path $q_2 \to q_6 \to q_7 \to q_2$ corresponds to the rule $X \to bXa$, where the symbols on the right side of the rule are pushed in reverse order.

5. Language $A = \{\, b^i a^j c^k \mid i, j, k \geq 0 \text{ and } i = j + k \,\}$ is nonregular. We prove this by contradiction. Suppose that $A$ is a regular language. Let $p$ be the "pumping length"

of the Pumping Lemma. Consider the string $s = b^p a^p$. Note that $s \in A$ because $s = b^i a^j c^k$ with $i = j = p$ and $k = 0$. Also, we have that $|s| = 2p > p$, so the Pumping Lemma will hold. Thus, there exist strings $x$, $y$, and $z$ such that $s = xyz$ and

(a) $xy^iz \in A$ for each $i \geq 0$,

(b) $|y| > 0$,

(c) $|xy| \leq p$.

Because the first $p$ symbols of $s$ are all $b$'s, the third property implies that $x$ and $y$ consist only of $b$'s. So $z$ will be the rest of the first set of $b$'s (possibly none), followed by $a^p$. The second property states that $|y| > 0$, so $y$ has at least one $b$. More precisely, we can then say that

$$\begin{aligned} x &= b^j \text{ for some } j \geq 0, \\ y &= b^k \text{ for some } k \geq 1, \\ z &= b^m a^p \text{ for some } m \geq 0. \end{aligned}$$

Because

$$b^p a^p = s = xyz = b^j b^k b^m a^p = b^{j+k+m} a^p,$$

we must have that

$$j + k + m = p \quad \text{and} \quad k \geq 1.$$

The first property implies that the pumped string $xy^2z \in A$, but

$$\begin{aligned} xy^2z &= b^j b^k b^k b^m a^p \\ &= b^{p+k} a^p \notin A \end{aligned}$$

because the number of $b$'s does not equal the sum of the number of $a$'s and $c$'s. This contradicts the first property of the pumping lemma. Therefore, $A$ is a nonregular language.

Another possible string that will result in a contradiction is $s = b^{2p} a^p c^p \in A$, where $|s| = 4p > p$. Then splitting $s = xyz$ satisfying properties (ii) and (iii) of the pumping lemma will lead to

$$\begin{aligned} x &= b^j \text{ for some } j \geq 0, \\ y &= b^k \text{ for some } k \geq 1, \\ z &= b^{m+p} a^p c^p \text{ for some } m \geq 0, \end{aligned}$$

where $j + k + m = p$. Property (i) of the pumping lemma states that $xyyz \in A$, but $xyyz = b^{2p+k} a^p c^p \notin A$ because $k \geq 1$, giving a contradiction.