PDAs, Chomsky's Hierarchy

In this assignment, **"construct directly a PDA"** means that no auxiliary stack symbols are used other than the bottom-marker. This implies that you cannot construct your PDA by converting a CFG to it.

Unless otherwise stated, Σ will denote the alphabet $\{\mathtt{a}, \mathtt{b}\}$.

1. (20%) For each of the following statements about languages $L, M, R \subseteq \Sigma^*$ determine whether it is always true, and justify your answer.

(i) If L is CF and R regular, then $L - R$ is CF.

Solution. True. The complement $\bar{R}$ of R must be regular, and so $L - R = L \cap \bar{R}$ must be CF, as the intersection of a CFL and a regular language.

(a) If R is regular and L CF, then $R - L$ is regular.

Solution. False. Σ^* is regular, but $\Sigma^* - L$, i.e. the complement of the CFL L, need not be CF.

(b) If R and R' are regular and L is CF then $(R \cdot L) \cup (R' \cdot L)$ is CF.

Solution. True, since every regular language is CF, and the collection of CFLs is closed under concatenation and union.

Note: The initial phrasing of this problem asked about the ***intersection*** $(R \cdot L) \cap (R' \cdot L)$, with the intent that the presumably equivalent languagae $(R \cap R') \cdot L)$ is CF, since it is the intersection of a CF and a refular language, However, that premise is false, since concatenation does ***not*** in general distribute over intersection. Consider $R = \{\mathtt{a}\}$, $R' = \{\mathtt{aa}\}$ and $L = \{\mathtt{a}, \mathtt{aa}\}$. More generally, if $u \neq v$ while $u \cdot v = v \cdot u$ then $\{u\} \cdot \{v\} = \emptyset\}$ but $u \cdot v$ is in both $\{u\} \cdot \{u, v\}$ and $\{v\} \cdot \{u, v\}$.

A. For a string $w \in \{\mathtt{a}, \mathtt{b}, \#\}^*$ let $w^\#$ be the initial substring of w through the first occurrence of $\#$, if there is one. For example, if $w = \mathtt{baa\#aba\#a}$ then $w^\# = \mathtt{baa\#}$ and if $w = \mathtt{abb}$ then $w^\# = \mathtt{abb}$. For a language $L \subseteq \{\mathtt{a}, \mathtt{b}, \#\}^*$ define $L^\# =_{\text{df}} \{w^\# \mid w \in \{\mathtt{a}, \mathtt{b}, \#\}^* \}$.

Show that if L is a CFL then so is $L^\#$.

Solution. Given a CFL L consider a PDA P recognizing L, with f as accepting state. We may assume that f is a terminal state, i.e. there are no transitions of the form $f \xrightarrow{\sigma(\alpha \to \beta)} q$. Let $P^\#$ be the PDA P with each transition of the form $q \xrightarrow{\#(\alpha \to \beta)} p$ replaced by $q \xrightarrow{\#(\alpha \to \beta)} f$. Then $P^\#$ is a PDA that recognizes $L^\#$. So $L^\#$ is a CFL.

B. **(i)** Construct directly a PDA that recognizes the language $L = \{\mathtt{a}^i\mathtt{c}(\mathtt{ab})^i \mid i \geqslant 0\}$.

Solution. L is recognized by the following PDA M. M pushes $\mathtt{ab}$'s on the for each consecutive $\mathtt{a}$ it reads. On reading $\mathtt{c}$ M switches to a phase that pops an $\mathtt{a}$ for every $\mathtt{a}$ read and a $\mathtt{b}$ for each $\mathtt{b}$ read.
With states $\{s, q, p_0, p_1, f\}$, s being initial and f accepting, the transition rules are:

$$\begin{array}{lll} s \xrightarrow{\epsilon\,(\epsilon\to\$)} q & \qquad & p_0 \xrightarrow{\mathtt{a}\,(\mathtt{a}\to\epsilon)} p_1 \\ q \xrightarrow{\mathtt{a}\,(\epsilon\to\mathtt{a})} q & & p_1 \xrightarrow{\mathtt{b}\,(\epsilon\to\epsilon)} p_0 \\ q \xrightarrow{\mathtt{c}\,(\epsilon\to\epsilon)} p_0 & & p_0 \xrightarrow{\epsilon\,(\$\to\$)} f \end{array}$$

(ii) The obvious CFG G to generate L is $S \to \mathtt{a}\,S\,\mathtt{ab} \mid \mathtt{c}$. Convert G into another PDA that recognizes L.

Solution. With initial state s and accepting state f:

$$\begin{array}{lllll} s \xrightarrow{\epsilon\,(\epsilon\to S\$)} q & \qquad & q \xrightarrow{\epsilon\,(S\to\mathtt{c})} q & \qquad & q \xrightarrow{\mathtt{b}\,(\mathtt{b}\to\epsilon)} q \\ q \xrightarrow{\epsilon\,(S\to\mathtt{a}S\mathtt{ab})} q & & q \xrightarrow{\mathtt{a}\,(\mathtt{a}\to\epsilon)} q & & q \xrightarrow{\epsilon\,(\$\to\epsilon)} f \end{array}$$

2. (20+20%)

(a) Construct directly a PDA that recognizes the language
$L = \{\mathtt{a}^{p+q}\mathtt{b}^q\mathtt{c}^p \mid p, q \geqslant 0\}$.

Solution. $Q = \{s, q, r, f\}$, $\Gamma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \$\}$; initial state s; accepting states $A = \{s, f\}$.

$$\begin{array}{lll} s \xrightarrow{\epsilon\,(\epsilon\to\$)} q & \qquad & p \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} r \\ q \xrightarrow{\mathtt{a}\,(\epsilon\to\mathtt{a})} q & & r \xrightarrow{\mathtt{c}\,(\mathtt{a}\to\epsilon)} r \\ q \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} p & & r \xrightarrow{\epsilon\,(\$\to\epsilon)} f \\ p \xrightarrow{\mathtt{a}\,(\mathtt{a}\to\epsilon)} p & & \end{array}$$

(b) Define a CFG that generates L, and then convert it into another PDA N that recognizes L, different from the one in (a).

Solution. Let G have the productions $S \to \mathtt{a}S\mathtt{c} \mid T$, $T \to \mathtt{a}T\mathtt{b} \mid \varepsilon$.
The initial nonterminal is S.

$$\begin{array}{l} s \xrightarrow{\epsilon\,(\epsilon\to S\$)} q \\ q \xrightarrow{\epsilon\,(S\to\mathtt{a}S\mathtt{c})} q \\ q \xrightarrow{\epsilon\,(S\to T)} q \\ q \xrightarrow{\epsilon\,(T\to\mathtt{a}T\mathtt{b})} q \\ q \xrightarrow{\mathtt{a}\,(\mathtt{a}\to\epsilon)} q \\ q \xrightarrow{\mathtt{b}\,(\mathtt{b}\to\epsilon)} q \\ q \xrightarrow{\epsilon\,(\$\to\epsilon)} f \end{array}$$

3. (20%) For this problem read the revised slides for Chomsky's Hierarchy.
Let L be the language $L = \{\, \mathtt{a}^p \mathtt{b}^q \mathtt{c}^p \mathtt{d}^q \mid p, q > 0 \,\}$.
Show that L is context-sensitive.
[Hint: Define a non-contracting grammar that generates L, using the same approach as the one used for $\mathtt{a}^n \mathtt{b}^n \mathtt{c}^n$ in the slides.]

Solution. Nonterminals: S (initial), X, Y, B, C

$$
\begin{aligned}
&S \to XY \\
&X \to \mathtt{a}XC \mid \mathtt{a}C, \qquad Y \to BY\mathtt{d} \mid B\mathtt{d} \\
&CB \to BC \\
&\mathtt{a}B \to \mathtt{ab}, \qquad \mathtt{b}B \to \mathtt{bb} \\
&C\mathtt{d} \to \mathtt{cd}, \qquad C\mathtt{c} \to \mathtt{cc}
\end{aligned}
$$