Assignment 7: Dual-Clipping, PDAs

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.

1. (15+15%) Show that the following languages over $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ are not CF.

i. $L = \{\mathtt{a}^i\,\mathtt{b}\,\mathtt{a}^i\,\mathtt{b}\,\mathtt{a}^i \mid i \geqslant 0\}$

Solution. *Solution phrased as failure of the dual-clipping property:*
Given $k > 0$, let $w = \mathtt{b}^k\mathtt{a}\mathtt{b}^k\mathtt{a}\mathtt{b}^k$. We have $w \in L$ and $|w| > k$.

If p is a substring of w of length $\leqslant k$ then p intersects at most two of the three substrings $\mathtt{b}^k$, so clipping letters our of p either removes an $\mathtt{a}$, and the resulting string would not be in L, or removes at least one b, but not from all three blocks $\mathtt{b}^k$, yielding also a string not in L. Thus L fails the dual-clipping property, and cannot be CF.

Same solution via the dual-clipping theorem:
Suppose L is generated by a CFG G. Let m be the number of variables in G, d the degree of G, and $k = d^m$.
Take $w = \mathtt{b}^k\mathtt{a}\mathtt{b}^k\mathtt{a}\mathtt{b}^k$. Since $w \in L$ and $|w| > k$ it follows by the Dual-Clipping Theorem that there is a substring p of w, of length $\leqslant k$, so that the string w' obtained from w by removing some letters in p is also in L.
But since p has length $\leqslant k$ it intersects at most two of the blocks $\mathtt{b}^k$. So clipping letters out of p either removes an $\mathtt{a}$, and the resulting string would not be in L, or removes some b's from some but not all three blocks $\mathtt{b}^k$, yielding also a string not in L. Thus $w' \notin L$, contradicting the assumption that G is a CFG generating L.

(a) $L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{a}^i \mid j > i\}$.
Solution stated as failure of dual-clipping:
Given $k > 0$,
let $w = \mathtt{b}^k\mathtt{a}\mathtt{b}^k\mathtt{a}\mathtt{b}^k$. We have $w \in L$ and $|w| > k$.

If p is any substring of w of length $\leqslant k$ then p intersects at most two of the blocks $\mathtt{b}^k$. So clipping letters our of p either removes an $\mathtt{a}$, and the resulting string would not be in L, or removes at least one b, but not from all three blocks $\mathtt{b}^k$, yielding also a string not in L. Thus L fails the dual-clipping property, and cannot be CF.

Same solution stated as a contradiction derived from the existence of a CFG generating L:

Suppose L is generated by some CFG G. Let m be the number of G's variables, and d the maximal degree of G's productions.

Let $w = \mathtt{b}^k\mathtt{ab}^k\mathtt{ab}^k$. We have $w \in L$ and $|w| > k$. So by the Dual-Clipping Theorem there is a substring p of w of length $\leqslant k$, so that the string w' obtained from w by removing some letters in p is also in L.

But such a p, having length $\leqslant k$, intersects at most two of the blocks $\mathtt{b}^k$. So clipping letters out of p either removes an $\mathtt{a}$, and the resulting string would not be in L, or removes at least one b, but not from all three blocks $\mathtt{b}^k$, yielding also a string not in L. Thus $w' \notin L$, contradicting the conclusion above.

Thus there is no CFG G that generates L, and L cannot be a CFL.

(b) $L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{a}^i\mathtt{b}^j \mid i,j \geqslant 0\}$.

Suppose $L = \mathcal{L}(G)$, G a CFG, and let k be G's clipping constant. Consider the string $w = \mathtt{a}^k\mathtt{b}^k\mathtt{a}^k$. Since $w \in L$ and $|w| \geqslant k$ w has a substring p of length $\leqslant k$, such that w' obtained from W by removing certain letters from w is in L. Since $|p| \leqslant k$ it spans at most two of the substrings $\mathtt{a}^k$, $\mathtt{b}^k$ and $\mathtt{a}^k$. If an occurrence of $\mathtt{a}$ is removed in one block of tta's then no instance is removed in the other. And similarly for $\mathtt{b}$. Thus $w' \notin L$, a contradiction.

A.

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 a's on the stack as long as it reads a's. On reading c M switches to a pair of states that pops an a for every ab read:

- States: $\{s, q, p_0, p_1, f\}$, initial state s, accepting state f.
- Transition rules:

$$s \xrightarrow{\epsilon\,(\epsilon\to\$)} q \qquad p_0 \xrightarrow{a\,(a\to\epsilon)} p_1$$
$$q \xrightarrow{a\,(\epsilon\to a)} q \qquad p_1 \xrightarrow{b\,(\epsilon\to\epsilon)} p_0$$
$$q \xrightarrow{c\,(\epsilon\to\epsilon)} p_0 \qquad p_0 \xrightarrow{\epsilon\,(\$\to\$)} f$$

ii. Give an accepting computation trace of your PDA for acabab.

Solution.

$$s \xrightarrow{\epsilon\,(\epsilon\to\$)} q \qquad p \xrightarrow{\mathtt{b}\,(\epsilon\to\mathtt{b})} p \qquad r \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} t$$
$$q \xrightarrow{\mathtt{a}\,(\epsilon\to\mathtt{a})} q \qquad p \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} r \qquad t \xrightarrow{\mathtt{b}\,(\mathtt{a}\to\epsilon)} t$$
$$q \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} p \qquad r \xrightarrow{\mathtt{a}\,(\mathtt{b}\to\epsilon)} r \qquad t \xrightarrow{\epsilon\,(\$\to\epsilon)} f$$

iii. Give a CFG G that generates L.

Solution. $S \to \mathtt{a}\,S\,\mathtt{ab} \mid \mathtt{c}$

iv. Convert G into another PDA that recognizes L.

Solution. With initial state s and accepting state f:

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

2. (10+5+10%)

(a) Construct directly a PDA that recognizes the language

$L = \{\mathtt{a}^{p+q}\mathtt{b}^q\mathtt{c}^p \mid p,q \geqslant 0\}$.

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

$$s \xrightarrow{\epsilon\,(\epsilon\to\$)} q \qquad p \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} r$$
$$q \xrightarrow{\mathtt{a}\,(\epsilon\to\mathtt{a})} q \qquad r \xrightarrow{\mathtt{c}\,(\mathtt{a}\to\epsilon)} r$$
$$q \xrightarrow{\epsilon\,(\epsilon\to\epsilon)} p \qquad r \xrightarrow{\epsilon\,(\$\to\epsilon)} f$$
$$p \xrightarrow{\mathtt{b}\,(\mathtt{a}\to\epsilon)} p$$

(b) Give an (accepting) computation trace of your PDA for **aaabcc** , and a (non-accepting) trace for **abc**.

$$\begin{aligned}(s,\mathtt{aaabcc},\varepsilon) &\Rightarrow (q,\mathtt{aaabcc},\$)\\ &\Rightarrow (q,\mathtt{abcc},\mathtt{aa}\$)\\ &\Rightarrow (q,\mathtt{bcc},\mathtt{aaa}\$)\\ &\Rightarrow (p,\mathtt{bcc},\mathtt{aaa}\$)\\ &\Rightarrow (p,\mathtt{cc},\mathtt{aa}\$)\\ &\Rightarrow (r,\mathtt{cc},\mathtt{aa}\$)\\ &\Rightarrow (r,\mathtt{c},\mathtt{a}\$)\\ &\Rightarrow (r,\varepsilon,\$)\\ &\Rightarrow (f,\varepsilon,\$)\end{aligned}$$

$$\begin{aligned}(s,\mathtt{abc},\varepsilon) &\Rightarrow (q,\mathtt{abc},\$)\\ &\Rightarrow (q,\mathtt{bc},\mathtt{a}\$)\\ &\Rightarrow (p,\mathtt{bc},\mathtt{a}\$)\\ &\Rightarrow (p,\mathtt{c},\mathtt{aa}\$)\\ &\Rightarrow (r,\mathtt{c},\mathtt{aa}\$)\\ &\Rightarrow (r,\varepsilon,\mathtt{a}\$)\end{aligned}$$

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

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

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

3. (25%) Construct directly a PDA recognizing the language

$L = \{\mathtt{a}^{i+j}\mathtt{b}^{j+k}\mathtt{c}^{k+i} \mid i,j,k \geqslant 0\}$.

L is recognized by a PDA that

(a) pushes x's while reading a's, then

(b) pops x's while reading b's, then switching on a whim to

(c) pushing x's while reading b's, then

(d) popping x's while reading c's.

An accepting trace would have stage (a) push $i+j$ x's, then, in phase (b), would pop j of these, thus leaving i x's on the stack. Stage (c) then reads the remaining k b's, ending with $i+k$ x's on the stack, which stage (d) pops off.

With s the start state and f the accepting state, the PDA is:

$$s \xrightarrow{(\epsilon \to \$)} q \qquad r \xrightarrow{\mathtt{b}\,(\epsilon \to \mathtt{x})} r$$

$$q \xrightarrow{\mathtt{a}\,(\epsilon \to \mathtt{x})} q \qquad r \xrightarrow{\epsilon\,(\epsilon \to \epsilon)} t$$

$$q \xrightarrow{\epsilon\,(\epsilon \to \epsilon)} p \qquad t \xrightarrow{\mathtt{c}\,(\mathtt{x} \to \epsilon)} t$$

$$p \xrightarrow{\mathtt{b}\,(\mathtt{x} \to \epsilon)} p \qquad t \xrightarrow{\epsilon\,(\$ \to \epsilon)} f$$

$$p \xrightarrow{\epsilon\,(\epsilon \to \epsilon)} r$$

If the input is $w = \mathtt{a}^{i+j}\mathtt{b}^{j+k}\mathtt{c}^{k+i}$ then it is accepted by a trace that switches states from q to p after $i+j$ pushes, from p to r after j pops, and from r to t after k pushes.

Conversely, if w is accepted where
m is the number of pushes in state q,
j is the number of pops in state p,
and k the number of pushes in state r,
then $m \geqslant j$, and writing i for $m-j$ we have $i+k$ x's on the stack when state t takes over, leading to acceptance.

4. (20%) Let $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, and let $X, Y \subseteq \Sigma^*$ be CFL's.

Prove that the following languages are also CF.

i. $L = \{\, u \cdot v \mid u \in A,\ v \in B,\ u$ without a's and v without b's $\}$.

Solution. Let $X' = X \cap \{\mathtt{b}, \mathtt{c}\}^*$ and $Y' = Y \cap \{\mathtt{a}, \mathtt{c}\}^*$. Thus $L = X' \cdot Y'$. X' is a CFL as the intersection of a CFL and a regular language, and similarly for Y'. So L is a CFL as the concatenation of two CFL's.

ii. $L = \{\, x \cdot y \cdot x' \mid x, x' \in X,\ y \in Y \,\}$.

Solution. $L = X \cdot Y \cdot X$, a concatenation of CFL's, and therefore a CFL.

(a) Let R be a regular language and K a CFL. Show that the following is a CFL.

$L = \{ x \cdot \# \cdot x' \mid x, x' \in R, |x| = |x'| \}$ (where $\#$ is a fresh symbol).

L is generated by the following CFG:

$S \to XSX \mid \#, \quad X \to \mathtt{a} \mid \mathtt{b}.$

(b) For R and K as above, let $L' = \{ x \cdot y \cdot x' \mid x, x' \in R, y \in K, |x| = |x'| \}$.

L' is the intersection of CFL language L and the language $R \cdot \{\#\} \cdot R$, which is regular as the concatenation of regular languages. The intersection of a CFL with a regular language is a CFL.