Assignment 8: Parse-trees and Dual-Clipping
Solutions

Practice problems and problem-parts are labeled in red.
Assigned problems and problem-parts are labeled in green.

1. (30%) For each of the following languages L construct a CFG G that generates it. Then give a parse tree for the string indicated, as well as the left-most derivation for that parse-tree.

(i) $L = \{(\mathtt{ab})^n(\mathtt{ba})^n \mid n \geqslant 0\}$ **ababbaba**.

Solution. (without the parse-tree.) $S \to \mathtt{ab}S\mathtt{ba} \mid \varepsilon$

$S \Rightarrow \mathtt{ab}S\mathtt{ba} \Rightarrow \mathtt{abab}S\mathtt{baba} \Rightarrow \mathtt{ababbaba}$

(a) $L = \{\mathtt{a}^n\mathtt{b}^k \mid n \leqslant k\}$ **aabbb**

Solution. $S \to \mathtt{a}S\mathtt{b} \mid S\mathtt{b} \mid \varepsilon$

Derivation: $S \Rightarrow S\mathtt{b} \Rightarrow \mathtt{a}S\mathtt{bb} \Rightarrow \mathtt{aa}S\mathtt{bbb}$

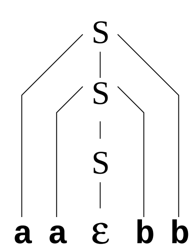
(b) $L = \{\mathtt{a}^n\mathtt{b}^k \mid k \leqslant n \leqslant 2k\}$ **aaabb**

Solution. $S \to \mathtt{a}S\mathtt{b} \mid \mathtt{aa}S\mathtt{b} \mid \varepsilon$

$S \Rightarrow \mathtt{a}S\mathtt{b} \Rightarrow \mathtt{aaa}S\mathtt{bb} \Rightarrow \mathtt{aaabb}$

A. Consider the CFG $S \mid \mathtt{a}S\mathtt{b} \mid \varepsilon$. So the degree is $d = 3$ and the number of nonterminals is m = 1. Take some string in $\mathcal{L}(G)$ of length $\geqslant d^m = 3$, exhibit its parse-tree, and extract the strings v_0, x, y, z, v_1 predicted by the Dual-Clipping Theorem.

Solution.



Taking the two lower occurrences of S, we have $v_0 = \mathtt{a}$. $x = \mathtt{a}$, $y = \varepsilon$, $z = \mathtt{b}$, and $v_1 = \mathtt{b}$.

2. (20%) Consider the following CFG G. (It generates the language $\{\mathtt{c}^i\mathtt{a}^n\mathtt{c}^k\mathtt{d}^k\mathtt{b}^{i+n}\}$.)

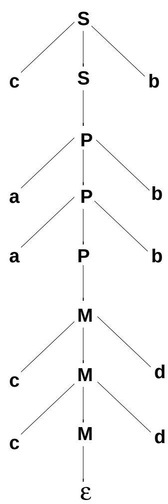
$$
\begin{aligned}
S &\rightarrow \mathtt{c}S\mathtt{b} \mid P \\
P &\rightarrow \mathtt{a}P\mathtt{b} \mid M \\
M &\rightarrow \mathtt{c}M\mathtt{d} \mid \varepsilon
\end{aligned}
$$

(a) Identify the clipping constant $k = d^m$ for G.

Solution. $d = 3,\ m = 3,\ k = 3^3 = 27$.

(b) Construct a parse-tree of G for the string **caaccddbbb**.

Solution.



(c) For the top two P's identify the partition $v_0 \cdot x \cdot y \cdot z \cdot v_1$ stated in the Dual-Clipping Theorem.

Solution. $\mathtt{c} \cdot \mathtt{a} \cdot \mathtt{accddb} \cdot \mathtt{b} \cdot \mathtt{b}$.

(d) Repeat for the bottom two occurrences of M.

Solution. $\mathtt{caac} \cdot \mathtt{c} \cdot \varepsilon \cdot \mathtt{d} \cdot \mathtt{dbbb}$.

3. (20%) Show that $L = \{\mathsf{a}^i\mathsf{b}^{2i}\mathsf{a}^i \mid i \geqslant 0\}$ is not a CFL.

Solution. Suppose L were CF, with clipping constant k. Take $w = \mathsf{a}^k\mathsf{b}^{2k}\mathsf{a}^k \in L$. We have $w \in L$ and $|w| > k$. By the Dual-Clipping Theorem there is a substring $p = y_0 \cdot x \cdot y_1$ of w with $y_0y_1 \neq \varepsilon$ and $|p| < k$ such that the string w' obtained by removing y_0 and y_1 from w is in L.

The string p cannot straddle all three blocks, since $|p| \leqslant k$. So either w' misses letters of w in $\mathsf{a}^k\mathsf{b}^{2k}$ but not in the last block a^k, or w' misses letters in $\mathsf{b}^{2k}\mathsf{a}^k$ but not in the initial block a^k. In either case w' cannot be of the form $\mathsf{a}^i\mathsf{b}^{2i}\mathsf{a}^i$ and so is not in L, a contradiction. It follows that no CFG generating L exists.

4. (20%) Show that $L = \{\mathsf{a}^i\mathsf{b}^j\mathsf{a}^i \mid i \geqslant j \geqslant 0\}$ is not CF. [Hint: Use dual-pumping]

Solution. Let $k > 0$. Choose $w = a^kb^ka^k$, which is in L. Any substring $p = y_0xy_1$ of w with $y_0y_1 \neq \varepsilon$ and $|p| \leqslant k$ intersects at most two of the k-long blocks.

Consider the result w' of removing from w the substrings y_0 and y_1. If p does not intersect the block b^k then w' has the form $\mathsf{a}^p\mathsf{b}^k\mathsf{b}^q$ where $p+q < 2k$, which is not in L

Otherwise the second pumping instance of w over y_0 and y_1 has the form $\mathsf{a}^p\mathsf{b}^\ell\mathsf{a}^q$ where $\ell > k$ and $p+q \leqslant 2k$, a string which is, again, not in L.

Thus L fails the Dual-Pumping property of CFLs, and cannot be CF.

5. (10%) The *reverse* of a language L is $L^R = \{w^R \mid w \in L\}$.
Give an algorithm that converts a CFG G generating L to a CFG H generating L^R. Your algorithm should produce the desired conversion, but you need not prove that it does.

Solution. Given a CFG G generating L, let $\tilde{G}$ be the CFG whose productions are $A \to w^R$ for each production $A \to w$ of G. Then $\tilde{L} = \mathcal{L}(\tilde{G})$.

In proof (not required) we show that

$$S \Rightarrow_G^n \quad \textbf{IFF} \quad S \Rightarrow_{\tilde{G}}^n w^R$$

for any n and any string w (of terminals and nonterminals of G). This is established by induction on n.