# Solution to COMP4141 Homework 3

Kai Engelhardt

**Solution to Exercise 1** If $A$ and $B$ are languages, define

$$A \diamond B = \{\ xy \mid x \in A, y \in B, |x| = |y|\ \} \quad .$$

If $A$ and $B$ are regular languages, then $A \diamond B$ is a CFL.

**Proof with license for brevity:** Let $A, B \subseteq \Sigma^*$ be regular languages. Let $D_A$ and $D_B$ be DFAs with $L(D_A) = A$ and $L(D_B) = B$. Construct a PDA $P$ that non-deterministically jumps once from parsing input like $D_A$ while pushing $\mathtt{1}$'s onto the stack to parsing like $D_B$ while popping $\mathtt{1}$'s from the stack. The jump can only occur in accepting states of $D_A$. The PDA accepts if it is in an accepting state of $D_B$ when hitting the empty stack. It follows that $L(P) = A \diamond B$. ■

**Proof without the license:** Let $A, B \subseteq \Sigma^*$ be regular languages. Let $D_A = (Q_A, \Sigma, \delta_A, q_0^A, F_A)$ and $D_B = (Q_B, \Sigma, \delta_B, q_0^B, F_B)$ be DFAs with $L(D_A) = A$ and $L(D_B) = B$ and, w.l.o.g., disjoint state spaces containing neither $q_{\text{start}}$ nor $q_{\text{accept}}$. Construct a PDA $P = (Q, \Sigma, \Gamma, \delta, q_{\text{start}}, \{q_{\text{accept}}\})$ where

$$\begin{aligned}
Q &= Q_A \cup Q_B \cup \{q_{\text{start}}, q_{\text{accept}}\} \\
\Gamma &= \{\mathtt{1}, \$\} \\
\delta(q_{\text{start}}, \epsilon, \epsilon) &= \{(q_0^A, \$)\} \\
\delta(q, a, \epsilon) &= \{(q', \mathtt{1})\} && \text{if } q \in Q_A \wedge \delta_A(q, a) = q' \notin F_A \\
\delta(q, \epsilon, \epsilon) &= \{(q_0^B, \epsilon)\} && \text{if } q \in F_A \\
\delta(q, a, \mathtt{1}) &= \{(q', \epsilon)\} && \text{if } q \in Q_B \wedge \delta_B(q, a) = q' \\
\delta(q, \epsilon, \$) &= \{(q_{\text{accept}}, \epsilon)\} && \text{if } q \in F_B
\end{aligned}$$

that non-deterministically jumps once from parsing input like $D_A$ while pushing $\mathtt{1}$'s onto the stack to parsing like $D_B$ while popping $\mathtt{1}$'s from the stack. The jump can only occur in accepting states of $D_A$. The PDA accepts if it is in an accepting state of $D_B$ when hitting the empty stack. We finally show that $L(P) = L(D_A) \diamond L(D_B)$. Let $w \in \Sigma^*$.

$$\begin{aligned}
w \in L(D_A) \diamond L(D_B) &\Leftrightarrow w = xy \wedge x \in L(D_A) \wedge y \in L(D_B) \wedge |x| = |y| \\
&\Leftrightarrow w = xy \wedge \hat{\delta_A}(q_0^A, x) \in F_A \wedge \hat{\delta_B}(q_0^B, y) \in F_B \wedge |x| = |y| \\
&\Leftrightarrow w = xy \wedge (q_0^A, xy, \$) \overset{*}{\rightsquigarrow} (q, y, \$\mathtt{1}^{|x|}) \wedge (q_0^B, y, \$\mathtt{1}^{|y|}) \overset{*}{\rightsquigarrow} (q', \epsilon, \$) \wedge \\
&\qquad |x| = |y| \wedge q \in F_A \wedge q' \in F_B \\
&\Leftrightarrow w = xy \wedge (q_{\text{start}}, xy, \epsilon) \overset{*}{\rightsquigarrow} (q, y, \$\mathtt{1}^{|x|}) \wedge (q_0^B, y, \$\mathtt{1}^{|y|}) \overset{*}{\rightsquigarrow} (q_{\text{accept}}, \epsilon, \epsilon) \wedge \\
&\qquad |x| = |y| \wedge q \in F_A \\
&\Leftrightarrow (q_{\text{start}}, w, \epsilon) \overset{*}{\rightsquigarrow} (q_{\text{accept}}, \epsilon, \epsilon) \\
&\Leftrightarrow w \in L(P)
\end{aligned}$$

■

Let, as usual, $\|w\|_v$ denote the number of occurrences of the substring (or letter) $v$ in string $w$. Let $\sqsubseteq$ denote the non-strict prefix order on strings: $x \sqsubseteq y$ iff $\exists z\,(xz = y)$. Write $x \sqsubset y$ for the strict version, that is, $x \sqsubset y$ iff $x \sqsubseteq y$ and $x \neq y$.

**Solution to Exercise 2** Let $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ and

$$L = \{\ w \in \Sigma^* \mid \forall v \sqsubseteq w\,(\|v\|_{\mathtt{a}} \geq \|v\|_{\mathtt{b}})\ \}\quad .$$

We claim that $G = (\{S, A\}, \Sigma, R, S)$ where

$$S \to A\mathtt{a}S \mid A \tag{1}$$
$$A \to \mathtt{a}A\mathtt{b}A \mid \epsilon \tag{2}$$

is an unambiguous CFG for $L$.

**Proof sketch:** First we claim that $A$ generates all balanced strings[1] in $L$ unambiguously: Let $w = w_1 \ldots w_n \in \Sigma^*$. Let $c_i = \|w_1 \ldots w_i\|_{\mathtt{a}} - \|w_1 \ldots w_i\|_{\mathtt{b}}$. The *mate* of $\mathtt{a}$ at position $i$ in $w$ is the $\mathtt{b}$ at the lowest position $j > i$ where $c_j > c_i$. It is easy to show inductively that for any balanced $w \in L$ and any parse tree for $w$ generated from $A$, the first rule in (2) generates the mated pairs at the same time, hence it divides $w$ in a unique way and consequently the grammar is unambiguous for the balanced strings.

Next we claim that $S$ generates also all unbalanced strings in $L$. We can show inductively that for any $w \in L$, the first rule of (1) generates the unmated $\mathtt{a}$'s and the $A$ rules generate the mated pairs. The generation can be done in only one way so the grammar is unambigious. ■

**Solution to Exercise 3** Define

$$\mathit{NOPREFIX}(A) = \{\ w \in A \mid \forall x \sqsubset w\,(x \notin A)\ \}\quad .$$

The CFLs are not closed under the *NOPREFIX* operation.

**Proof:** Let

$$A = \left\{\ \mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid j > 0 \text{ and } (i = j \text{ or } j = k)\ \right\}\quad .$$

It is context free a.o. by[2] lecture 4. Next observe that

$$\mathit{NOPREFIX}(A) = \{\ \mathtt{a}^i\mathtt{b}^i \mid i > 0\ \} \cup \left\{\ \mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i > j > 0 \text{ and } j = k\ \right\}\quad .$$

Assume $L = \mathit{NOPREFIX}(A)$ is a CFL and let $p$ be its pumping length. Consider the word $w = \mathtt{a}^{p+1}\mathtt{b}^p\mathtt{c}^p \in L$. Let $uvxyz = w$ be a partition satisfying the 3 conditions of the pumping lemma for CFLs.

**Case $vy = \mathtt{a}^j$:** Note that $p + 1 > j$. Hence pumping down results in $uxz = \mathtt{a}^{p+1-j}\mathtt{b}^p\mathtt{c}^p \in A$ which has the prefix $\mathtt{a}^{p+1-j}\mathtt{b}^{p+1-j} \in A$ and is thus not in $L$.

**Case $v = \mathtt{a}^j$ and $y = \mathtt{b}^k$ for some $j, k > 0$:** pumping down leads outside $L$.

**Case $vy = \mathtt{b}^j$:** pumping up once leads outside $A$ and hence $L$.

**Case $v = \mathtt{b}^j$ and $y = \mathtt{c}^k$ for some $j, k > 0$:** pumping up once results in $uv^2xy^2z = \mathtt{a}^{p+1}\mathtt{b}^{p+j}\mathtt{c}^{p+k}$ which has the prefix $\mathtt{a}^{p+1}\mathtt{b}^{p+1} \in A$ and is thus not in $L$.

**Case $vy = \mathtt{c}^j$:** pumping up once leads outside $A$ and hence $L$.

Since $w$ cannot be pumped $L$ is not context free. ■

[1] A string $w \in \{\mathtt{a}, \mathtt{b}\}^*$ is *balanced* if $\|w\|_{\mathtt{a}} = \|w\|_{\mathtt{b}}$.

[2] Technically, we showed context freedom of $A \cup \{\epsilon\}$ but finitely many differences don't make a difference since one could run a DFA checking for the excluded words in parallel and only accept if that DFA does not spot any of the differences.