Equivalence of PDAs and CFGs

In the last lecture, we proved the following theorem by starting with a CFG in Greibach Normal Form and constructing an equivalent PDA.

Theorem 1 If L is a CFL then $L = S(M)$ for some PDA M.

The converse is also true.

Theorem 2 If $L = S(M)$ for some PDA M, then $L = L(G)$ for some CFG G.

Proof Sketch: Let $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$.

The idea is to build a grammar $G = (V, \Sigma, P, S)$ such that $X \overset{*}{\Rightarrow} w$ iff on reading string $w \in \Sigma^*$, X is popped off the stack in M.

Case 1: M has one state

Let the state be q. Define the variable set V of the grammar to be Γ. The above condition then translates to:

$$(q, w, X) \overset{*}{\vdash} (q, \varepsilon, \varepsilon) \text{ for any } w \in \Sigma^*. \tag{1}$$

Therefore, $V = \Gamma$, $S = Z_0$, and Σ is of course the same for M and G. The set of productions $P = \{A \to a\alpha \mid (q, \alpha) \in \delta(q, a, A)\}$, for all $a \in \Sigma \cup \{\varepsilon\}$, $\alpha \in \Gamma^*$ and $A \in \Gamma$. Note that G is not in GNF, to allow for ε-moves in the PDA.

Case 2: M has multiple states

Define the set of variables of G to be $V = \{\langle p, X, q\rangle \mid p, q \in Q, X \in \Gamma\}$, the goal being to satisfy the following condition:

$$\langle p, X, q\rangle \overset{*}{\Rightarrow} w \text{ iff} (p, w, X) \overset{*}{\vdash} (q, \varepsilon, \varepsilon) \text{ for any } w \in \Sigma^*. \tag{2}$$

So what should the start symbol S of G be? With the above goal in mind, we would need the start symbol to be $\langle q_0, Z_0, q\rangle$, for each $q \in Q$. Instead we pick a new variable S (not in V) and add the following productions to the production set P: $S \to \langle q_0, Z_0, q\rangle$, for all $q \in Q$. Thus the variable set of G is $V \cup \{S\}$.

Now we need to consider each transition in the δ function to complete the set of production rules P.

We add productions to P as indicated below, for $p, q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$, $X \in \Gamma$. For each case, keep in mind the above condition (2) that we want to satisfy.

- If $(q, \varepsilon) \in \delta(p, a, X)$, then add the production $\langle p, X, q\rangle \rightarrow a$.
- If $(q, \gamma) \in \delta(p, a, X)$, then add the productions $\langle p, X, r\rangle \rightarrow a\langle q, \gamma, r\rangle$, for all $r \in Q$.
- If $(q, \gamma_1\gamma_2) \in \delta(p, a, X)$, then add the productions $\langle p, X, r\rangle \rightarrow a\langle q, \gamma_1, s\rangle\langle s, \gamma_2, r\rangle$, for all $r, s \in Q$.
- $\cdots$
- If $(q, \gamma_1\gamma_2 \ldots \gamma_m) \in \delta(p, a, X)$, then add the productions $\langle p, X, r\rangle \rightarrow a\langle q, \gamma_1, s_1\rangle\langle s_1, \gamma_2, s_2\rangle \ldots \langle s_{m-1}, \gamma_m, s_m\rangle$, for all $s_1, s_2, \ldots s_m \in Q$.

$\square$

To solidify your understanding of the above construction, do the following exercise:

Give a grammar for the language $S(M)$ where $M = (\{q_0, q_1\}, \{0, 1\}, \{Z_0, X\}, \delta, q_0, Z_0, \emptyset)$ and δ is given by:

$$\begin{aligned}
\delta(q_0, 1, Z_0) &= \{(q_0, XZ_0)\} \\
\delta(q_0, 1, X) &= \{(q_0, XX)\} \\
\delta(q_0, 0, X) &= \{(q_1, X)\} \\
\delta(q_0, \varepsilon, Z_0) &= \{(q_0, \varepsilon)\} \\
\delta(q_1, 1, X) &= \{(q_1, \varepsilon)\} \\
\delta(q_1, 0, Z_0) &= \{(q_0, Z_0)\}
\end{aligned}$$