Solution to Exercise 7: Proving that a language cannot have an LL(1) grammar

October 27, 2014

This exercise is quite difficult. It is completely optional for you to read and understand the following proof. We would certainly not be asking questions as difficult as this in the quiz.

1 Exercise 7

Show that the language $L = \{a^l b^m \mid l > m\}$ which is defined by the grammar

$$S \rightarrow aS \mid P$$
$$P \rightarrow aPb \mid a$$

cannot have an LL(1) grammar.

1.1 Solution

Say we have an LL(1) grammar G recognizing L. Without loss of generality, assume that G only has reachable and productive non-terminals. Since the language is infinite, the grammar has at least one "recursive" non-terminal N, i.e, $N \Rightarrow^* \alpha N \beta$, where α and β are sentential forms which is a (possibly empty) sequence of terminals and non-terminals. Moreover, there exists a recursive non-terminal A such that $A \Rightarrow^* \alpha A \beta$ and $\alpha \Rightarrow^* a^k$ for some $k > 0$. Otherwise, it is easy to show that the number of a's has to be bounded in every string generated by the grammar.

Case (i): β is empty i.e, $A \Rightarrow^* \alpha A$, or β only derives empty string i.e, $\beta \Rightarrow^* w$ implies $w = \epsilon$.

Consider a derivation D of a string $a^l b^m$, $l > 0$ that uses the reduction $A \Rightarrow^* \alpha A \beta$. Note that there has to be one such derivation since A is a reachable non-terminal. Let ρ be the prefix of A in the derivation before the application of $A \Rightarrow^* \alpha A \beta$. That is, let D be $S \Rightarrow^* \rho A \delta \Rightarrow^* \rho \alpha A \beta \delta \Rightarrow^* a^l b^m$. By assumption, β is empty or it can derive only empty string. Therefore, D is of the form $S \Rightarrow^* \rho \alpha A \delta \Rightarrow^* a^l b^m$.

We know that α derives a non-empty sequence of a's i.e, $\alpha \Rightarrow a^k$, $k > 0$. Hence, ρ can only derive (a possibly empty) sequence of a's. Otherwise, if $\rho \Rightarrow^* a^l b^i$, $i > 0$ then we can derive a string that does not belong to the language as $\rho\alpha \Rightarrow^* a^l b^i a^k$, where i, k

are positive integers. Therefore, D has to be of the form $S \Rightarrow^* a^j A\delta \Rightarrow^* a^l b^m$, for some $j > 0$.

Now, consider the (partial) derivation $D' : S \Rightarrow^* a^j A\delta \Rightarrow a^j \alpha A\delta \Rightarrow^* a^{j+k} A\delta$, where $k > 0$. (The sentential form β is omitted in D' as it either empty or it can only derive ϵ). Since $a^{j+k} b^j \in L$ and the grammar G is LL(1), $a^{j+k} b^j$ has to be derivable through $a^{j+k} A\delta$. That is, $S \Rightarrow^* a^{j+k} A\delta \Rightarrow^* a^{j+k} b^j$. Hence, $A\delta \Rightarrow^* b^j$.

Using this fact in derivation D, we get $S \Rightarrow^* a^j A\delta \Rightarrow^* a^j b^j$. But, $a^j b^j \notin L$ (note that $j > 0$). Hence, when β is empty or when it can only derive ϵ, we obtain a contradiction.

Case (ii): β is non-empty and it derives a non-empty string. That is, $\beta = N_1 N_2 \cdots N_n$ and $\beta \Rightarrow^* w$ s.t. $|w| > 0$.

Claim 1: Both A and β are nullable i.e, $\beta \Rightarrow^* \epsilon$ and $A \Rightarrow^* \epsilon$.

As in the previous case, consider a derivation D of a string $a^l b^m$ that uses the reduction $A \Rightarrow^* \alpha A \beta$. By the same argument presented earlier, we can deduce that D has to be of the form $S \Rightarrow^* a^j A\beta\delta \Rightarrow^* a^l b^m$, where $j > 0$. Since $a^j \in L$ and the grammar G is LL(1), a^j has to be derivable through $a^j A\beta\delta$. Therefore, $S \Rightarrow^* a^j A\beta\delta \Rightarrow^* a^j$. This implies that both A and β are nullable.

Claim 2: $first(\beta) = \{b\}$.

By the above claim, A is nullable. If $a \in first(\beta)$ then $first(A) \cap follow(A) \neq \emptyset$ which violates the LL(1) property. Therefore, $first(\beta) \subseteq \{b\}$. By assumption, β can derive a non-empty string. Hence, $first(\beta) \neq \emptyset$. Therefore, $first(\beta) = \{b\}$.

Now, let's come back to the proof of the main statement. Given $\beta = N_1 N_2 \cdots N_n$. Since β is nullable, each of the N_i's are nullable. By the definition of $follow$, $follow(A) \subseteq follow(N_i)$ for each $1 \leq i \leq n$ as every N_i is nullable. Hence, $b \in follow(N_i)$ for all $1 \leq i \leq n$ as $b \in follow(A)$. Since $b \in first(\beta)$, there exists a j such that $b \in first(N_j)$. Therefore, $first(N_j) \cap follow(N_j) = \{b\}$ and N_j is nullable. This violates the LL(1) property and hence is a contradiction.

Since we get a contradiction in both cases where β is empty and is non-empty, there cannot exist an LL(1) grammar G for L