
Problem 5.35. Say that a variable A in CFG G is necessary if it appears in every derivation of some string $w \in G$. Let $NECESSARY_{CFG} = \{\langle G, A\rangle \mid A \text{ is a necessary variable in } G\}$.

Part a. Show that $NECESSARY_{CFG}$ is Turing-recognizable.

Proof. We construct a ***TM*** N which recognizes $NECESSARY_{CFG}$.

$N =$ "On input $\langle G, A\rangle$, where G is a CFG and A is a symbol:

1. If A is not a variable in G, then *reject*.
2. Repeat for each $i = 0, 1, 2, \cdots$.
3. Generate all strings $s_1, s_2, s_3, \cdots$ of length i, such that each $s_i \in L(G)$.
4. If for some s_i and s_j, where $s_i = s_j$ and $i \neq j$ and A is included in the derivation of both s_i and s_j, then *accept*."

□

Part b. Show that $NECESSARY_{CFG}$ is undecidable.

Proof. This proof is similar to the proof of Problem 5.21, except one minor change. The rules of the CFG G are modified to use new variables $A_1 \ldots A_k$, where each rule A_i has the form $A_i \to a_i$.

To show that $NECESSARY_{CFG}$ is undecidable, we give a reduction from PCP to $NECESSARY_{CFG}$. Given an instance

$$P = \left\{ \left[\frac{t_1}{b_1}\right], \left[\frac{t_2}{b_2}\right], \ldots, \left[\frac{t_k}{b_k}\right] \right\}$$

of the Post Correspondence Problem, construct a CFG G with the rules

$$\begin{aligned}
S &\to T \mid B \\
T &\to t_1 T A_1 \mid \cdots \mid t_k T A_k \mid t_1 A_1 \mid \cdots \mid t_k A_k \\
B &\to b_1 B A_1 \mid \cdots \mid b_k B A_k \mid b_1 A_1 \mid \cdots \mid b_k A_k \\
A_1 &\to a_1 \\
&\ \vdots \\
A_k &\to a_k,
\end{aligned}$$

where $a_1, a_2, \cdots, a_k$ are new terminal symbols.

If P is an instance of Post Correspondence Problem, then for every possible arrangement of the dominoes $i_1, i_2, \ldots, i_l$, where $t_{i_1} t_{i_2} \ldots t_{i_l}$ is the top string and $b_{i_1} t_{b_2} \ldots b_{i_l}$ the bottom string, there exists exactly one derivation for the top and bottom strings in G. The terminal symbols $a_1, a_2, \cdots, a_k$, make sure that every string generated using the variable T has at most one left-most derivation even if some t_i and t_j are same. Same argument can be said for the variable B.

If $P \in PCP$, then there exists a match $i_1, i_2, \ldots, i_l$, where $t_{i_1} t_{i_2} \ldots t_{i_l} = b_{i_1} t_{b_2} \ldots b_{i_l}$. In this case, the CFG G is guaranteed to have following two left-most derivations that produce the same string and some variable A_i appears in both of these derivations.

$$
\begin{aligned}
S &\to T \\
&\to t_{i_1} T A_{i_1} \\
&\to t_{i_1} t_{i_2} T A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots T \ldots A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} A_{i_l} \ldots A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots a_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots a_{i_2} a_{i_1}
\end{aligned}
$$

$$
\begin{aligned}
S &\to B \\
&\to b_{i_1} B A_{i_1} \\
&\to b_{i_1} b_{i_2} B A_{i_2} A_{i_1} \\
&\to b_{i_1} b_{i_2} \ldots B \ldots A_{i_2} A_{i_1} \\
&\to b_{i_1} b_{i_2} \ldots b_{i_l} A_{i_l} \ldots A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots A_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots a_{i_2} A_{i_1} \\
&\to t_{i_1} t_{i_2} \ldots t_{i_l} a_{i_l} \ldots a_{i_2} a_{i_1}
\end{aligned}
$$

□