Post's Correspondence Problem PCP is about context-free grammars

Wim H. Hesselink (whh471)
w.h.hesselink@rug.nl

University of Groningen, The Netherlands

Post's correspondence problem PCP is defined in the book [1] as follows. An instance of the problem consists of two lists of strings over some alphabet Σ. The two lists, say A and B, must be of equal length, say k. Assume $A = w_1, \ldots, w_k$ and $B = x_1, \ldots, x_k$. A *solution* of an instance (A, B) is a nonempty finite sequence of indices $i_1,\ i_2,\ \ldots,\ i_m$ such that the concatenations $w_{i_1} w_{i_2} \ldots w_{i_m}$ and $x_{i_1} x_{i_2} \ldots x_{i_m}$ are equal. The book [1] proceeds to prove that problem PCP is undecidable, by reducing it to Turing's Halting problem.

As I wanted to eliminate the finite nonempty sequence of indices, I came to the following reformulation. Recall that w^R is the reversal of string w. The equality $w_{i_1} w_{i_2} \ldots w_{i_m} = x_{i_1} x_{i_2} \ldots x_{i_m}$ is equivalent with $(w_{i_1} w_{i_2} \ldots w_{i_m})^R = y_{i_m} y_{1_{m-1}} \ldots y_{i_1}$ where $y_i = x_i^R$. Let us assume that # is a symbol not in Σ. The above equality is then equivalent to the condition that the concatenation

$$w_{i_1} w_{i_2} \ldots w_{i_m} \# y_{i_m} y_{1_{m-1}} \ldots y_{i_1}$$

is a palindrome over the alphabet $\Sigma \cup \{\#\}$. The candidate strings here are the strings generated by the grammar with a single nonterminal S and the $2k$ production rules

$$(0) \qquad G_p : \qquad S \quad \to \quad w_i S y_i \mid w_i \# y_i \ , \quad i \in \{1 \ldots k\}.$$

An instance of the reformulated PCP is thus a context-free grammar over $\Sigma \cup \{\#\}$ of the form (0) with strings $w_i,\ y_i \in \Sigma^*$. The question to be answered is whether the language generated by grammar G contains a palindrome, or more precisely a palindrome generated by the grammar

$$G_0 : \qquad S \quad \to \quad aSa \mid a\#a \ , \quad a \in \Sigma.$$

The reformulated problem is thus whether the intersection of the languages generated by two rather specific context-free grammars is nonempty.

If one knows context-free grammars, the reformulation is easier to understand than the original version of PCP, but in the time of Post [2] context-free grammars had not yet been defined. It may well be that the reformulation is known but, if so, it did not reach the standard sources about the subject.

Given that PCP is undecidable, the reformulation of PCP directly implies indecidability of the question whether two context-free languages given by their grammars have a nonempty intersection. The proofs of this fact in the books [1,3] are more complicated.

References

1. J.E. Hopcroft and J.D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.
2. E. Post. A variant of a recursively unsolvable problem. *Bulletin of the AMS*, 52:264–268, 1946.
3. Th.A. Sudkamp. *Languages and Machines.* Pearson, 3rd edition, 2006.