

POST CORRESPONDENCE PROBLEM (PCP): Given a finite set of ordered pairs $(x_1, y_1), \dots, (x_n, y_n)$ of strings over Σ , determine whether there is a finite sequence of integers (i_1, i_2, \dots, i_m) , with each $i_j \in \{1, \dots, n\}$, such that

$$x_{i_1}x_{i_2}\cdots x_{i_m} = y_{i_1}y_{i_2}\cdots y_{i_m}. \quad (1)$$

For a particular instance $\{(x_1, y_1), \dots, (x_n, y_n)\}$ of the problem PCP, if there exists a sequence (i_1, i_2, \dots, i_m) satisfying (1), then we say the string $x_{i_1}x_{i_2}\cdots x_{i_m}$ is a *solution* to this instance.

An easy way to understand the problem PCP is to treat each pair (x_i, y_i) as a domino with string x_i at the top and string y_i at the bottom: $\begin{array}{|c|} \hline x_i \\ \hline y_i \\ \hline \end{array}$. The question here then is to select, from the given dominoes

$$\begin{array}{|c|} \hline x_1 \\ \hline y_1 \\ \hline \end{array}, \begin{array}{|c|} \hline x_2 \\ \hline y_2 \\ \hline \end{array}, \dots, \begin{array}{|c|} \hline x_n \\ \hline y_n \\ \hline \end{array},$$

with unlimited supply for each type, some dominoes and arrange them into a row so that the top part of the dominoes spells the same word as the bottom part of the dominoes. For instance, we can obtain a solution $baaaaa$ from the following given dominoes

$$\begin{array}{|c|} \hline aa \\ \hline a \\ \hline \end{array}, \begin{array}{|c|} \hline ba \\ \hline baaa \\ \hline \end{array}$$

as follows:

$$\begin{array}{|c|c|c|} \hline ba & aa & aa \\ \hline baaa & a & a \\ \hline \end{array}.$$

*** Example 0.1** Prove that the problem PCP is undecidable (with respect to some alphabet Σ).

Proof. Let M be a fixed DTM, with a *one-way tape* (i.e., the original one-tape DTM defined in Section 4.1), such that the problem of determining whether M halts on a given string $x \in \{0, 1\}^*$ is undecidable. (I.e., $L(M)$ is a nonrecursive set.) We construct a reduction from the halting problem of this fixed DTM M to the problem PCP. That is, for each string x , we need to produce an instance $P_x = \{(x_1, y_1), \dots, (x_m, y_m)\}$ such that M halts on x if and only if P_x has a solution z .

Assume that the set of states in M is $Q = \{q_1, q_2, \dots, q_n\}$, where q_1 is the initial state and q_n is the halting state, and that the set of tape symbols of M is $\Gamma = \{s_1, s_2, \dots, s_k\}$. Also assume that $Q \cap \Gamma = \emptyset$. Let $Q' = \{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n\}$ and $\Gamma' = \{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_k\}$. Then, we fix the alphabet of our PCP problem as

$$\Sigma = Q \cup Q' \cup \Gamma \cup \Gamma' \cup \{q_{n+1}, \bar{q}_{n+1}, *, \bar{*}, [,]\}.$$

The pairs of strings in P_x consists of the following groups (for the sake of clarity, we show them as dominoes):

$$(1) \frac{[}{[Bxq_1B*}.$$

$$(2) \frac{a}{\bar{a}}, \frac{\bar{a}}{a}, \text{ for each } a \in \Gamma \cup \{*\}.$$

$$(3) \frac{q_i a c}{\bar{b} \bar{q}_j \bar{c}}, \frac{\bar{q}_i \bar{a} \bar{c}}{b q_j c}, \frac{q_i a^*}{\bar{b} \bar{q}_j \bar{B}^*}, \frac{\bar{q}_i \bar{a} \bar{*}}{b q_j B^*}, \text{ for each instruction } \delta(q_i, a) = (q_j, b, R) \text{ of } M \text{ and for each } c \in \Gamma.$$

$$(4) \frac{c q_i a}{\bar{q}_j \bar{c} \bar{b}}, \frac{\bar{c} \bar{q}_i \bar{a}}{q_j c b}, \text{ for each instruction } \delta(q_i, a) = (q_j, b, L) \text{ of } M \text{ and for each } c \in \Gamma.$$

$$(5) \frac{q_n a}{\bar{q}_n}, \frac{\bar{q}_n \bar{a}}{q_n}, \frac{q_n^*}{\bar{q}_{n+1} \bar{*}}, \frac{\bar{q}_n \bar{*}}{q_{n+1}^*}, \text{ for each } a \in \Gamma.$$

$$(6) \frac{a q_{n+1}}{\bar{q}_{n+1}}, \frac{\bar{a} \bar{q}_{n+1}}{q_{n+1}}, \text{ for each } a \in \Gamma.$$

$$(7) \frac{q_{n+1}^*]}{]}, \frac{\bar{q}_{n+1} \bar{*}]}{]}.$$

We need to prove that M halts on x if and only if P_x has a solution. First, assume that M halts on x . We modify the definition of the configuration of M to include all tape symbols in all cells that have ever been visited by the tape head in the computation of M on input x . That is, the rightmost blank symbols in a configuration are not removed in the representation of the configuration. Assume that the computation of M on x consists of the following configurations:

$$\alpha_0 \vdash \alpha_1 \vdash \cdots \vdash \alpha_\ell,$$

where $\alpha_0 = Bxq_1B$ and $\alpha_\ell = y_1 \cdots y_p q_n z_1 z_2 \cdots z_q$, where each y_i and each z_j is a single symbol in Γ . Define $\alpha_{\ell+i} = y_1 y_2 \cdots y_p q_n z_{i+1} \cdots z_q$, for $1 \leq i \leq q$, $\alpha_{\ell+q+1} = y_1 y_2 \cdots y_p q_{n+1}$, and $\alpha_{\ell+q+j} = y_1 \cdots y_{p+1-j} q_{n+1}$, for $2 \leq j \leq p+1$. We claim that if $\ell + q + p + 1$ is odd, then

$$[\alpha_0 * \bar{\alpha}_1 \bar{*} \alpha_2 * \cdots * \bar{\alpha}_{\ell+q+p+1} \bar{*}]$$

is a solution to P_x , and if $\ell + q + p + 1$ is even, then

$$[\alpha_0 * \bar{\alpha}_1 \bar{*} \alpha_2 * \cdots \bar{*} \alpha_{\ell+q+p+1}^*]$$

is a solution to P_x . (In the above, we write $\bar{\alpha}_i$ to denote the string obtained from α_i with each symbol a in α_i replaced by the symbol \bar{a} .)

To prove this claim, the critical observation is that if we use top strings x_i 's in P_x to form a string α_k^* , $0 \leq k \leq \ell + q + p$, that is, if

$$x_{i_1}x_{i_2} \cdots x_{i_t} = \alpha_k^*,$$

then the corresponding bottom strings must form the string $\bar{\alpha}_{k+1}^{\bar{*}}$, that is,

$$y_{i_1}y_{i_2} \cdots y_{i_t} = \bar{\alpha}_{k+1}^{\bar{*}}.$$

Similarly, if $x_{i_1}x_{i_2} \cdots x_{i_t} = \bar{\alpha}_k^{\bar{*}}$, then $y_{i_1}y_{i_2} \cdots y_{i_t} = \alpha_{k+1}^*$. For instance, if $k < \ell$, then these x_i 's must be from groups 2, 3 or 4 and it is easy to see that the strings in groups 3 and 4 force the corresponding y_i 's to form the successor configuration.

From this observation, we can prove the claim by induction. More precisely, we first define a pair of strings $\begin{bmatrix} u \\ v \end{bmatrix}$ to be a *partial solution* to P_x if there exists a sequence (i_1, \dots, i_t) , with each $i_j \in \{1, \dots, n\}$, such that $u = x_{i_1}x_{i_2} \cdots x_{i_t}$, $v = y_{i_1}y_{i_2} \cdots y_{i_t}$ and u is a prefix of v . We can then prove by induction that for each odd $i \leq \ell + q + p$,

$$\begin{bmatrix} \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} \\ \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} \alpha_{i+1}^* \end{bmatrix}$$

is a partial solution, and for each even $i \leq \ell + q + p$,

$$\begin{bmatrix} \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} \\ \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} * \bar{\alpha}_{i+1}^{\bar{*}} \end{bmatrix}$$

is a partial solution. First, we observe that the first pair $\begin{bmatrix} \alpha_0 \\ \mathbf{B}xq_1\mathbf{B}^* \end{bmatrix}$ is a partial solution. Next, assume that for some even $i < \ell + q + p$,

$$\begin{bmatrix} \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} \\ \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_i^{\bar{*}} * \bar{\alpha}_{i+1}^{\bar{*}} \end{bmatrix}$$

is a partial solution. Then, the only way for the top part of the partial solution to match the bottom part is to attach a string $\bar{\alpha}_{i+1}^{\bar{*}}$ to it. From the observation we made above, we know that the corresponding bottom part must be α_{i+2}^* , and so

$$\begin{bmatrix} \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_{i+1}^{\bar{*}} \\ \alpha_0 * \bar{\alpha}_1^{\bar{*}} \cdots * \bar{\alpha}_{i+1}^{\bar{*}} \alpha_{i+2}^* \end{bmatrix}$$

is also a partial solution. The other case of odd i is similar.

4

So, from the above induction proof, we know that either

$$\begin{array}{|l} \hline [\alpha_0 * \cdots * \bar{*}\alpha_{\ell+q+p} * \\ \hline [\alpha_0 * \cdots * \bar{*}\alpha_{\ell+q+p} * \bar{*}\alpha_{\ell+q+p+1} * \\ \hline \end{array}$$

or

$$\begin{array}{|l} \hline [\alpha_0 * \cdots * \bar{*}\alpha_{\ell+q+p} * \\ \hline [\alpha_0 * \cdots * \bar{*}\alpha_{\ell+q+p} * \bar{*}\alpha_{\ell+q+p+1} * \\ \hline \end{array}$$

is a partial solution. Now, we can attach one of the pairs in group 7 to them to obtain the desired solution.

Conversely, assume that P_x has a solution z . We note that it must begin with $[\alpha_0*$, since the pair in group 1 is the only one whose top string and bottom string begin with the same symbol. The only way to extend this partial solution to z is to add α_0* to the top part and, from the observation above, we know that this will add to the bottom part an extra string $\bar{\alpha}_1\bar{*}$. Continuing this argument, we can see that the solution must contain prefixes of the form

$$[\alpha_0 * \bar{\alpha}_1 \bar{*} \cdots \bar{*} \alpha_i *$$

for an even i , or

$$[\alpha_0 * \bar{\alpha}_1 \bar{*} \cdots * \bar{\alpha}_i \bar{*}$$

for an odd i . Now, suppose M does not halt on x , then the computation of $M(x)$ never enters the state q_n , and so these partial solutions do not contain q_n or \bar{q}_n and, hence, do not contain q_{n+1} or \bar{q}_{n+1} . However, the two pairs of group 7 are the only ones whose two strings have the same ending symbol, and they contain either q_{n+1} or \bar{q}_{n+1} . That is a contradiction. We conclude that M must halt on x . \square