# Theory of Computation
# Undecidability of
# Post Correspondence Problem

# The Problem

Input: Two lists $A = \{w_1, w_2, \ldots, w_k\}$ and $B = \{x_1, x_2, \ldots, x_k\}$ of strings over $\Sigma$.
Solution: A sequence of integers $i_1, i_2, \ldots i_m$ (multiplicity allowed, no ordering on the integers) for some $m \geq 1$ such that $w_{i_1} w_{i_2} \ldots w_{i_m} = x_{i_1} x_{i_2} \ldots x_{i_m}$.
*A solution may not always exist.

# Example 1

Input: $\Sigma = \{0, 1\}$, $A = \{1, 10111, 10\}$, $B = \{111, 10, 0\}$
Solution: $m = 4$, $i_1 = 2, i_2 = 1, i_3 = 1, i_4 = 3$.
The string $1011110$

# Example 2

Input: $\Sigma = \{0, 1\}$, $A = \{10, 011, 101\}$, $B = \{101, 11, 011\}$
No Solution: The strings are such that the alphabets will match only if $i_1 = 1, i_2 = i_3 = i_4 = \cdots = 3$. But for any $i_j$, $|w_{i_1} w_{i_2} \ldots w_{i_j}| < |x_{i_1} x_{i_2} \ldots x_{i_j}|$.

# Modified PCP (MPCP)

Input: Two lists $A = \{w_1, w_2, \ldots, w_k\}$ and $B = \{x_1, x_2, \ldots, x_k\}$ of strings over $\Sigma$.
Solution: A sequence of integers $i_1, i_2, \ldots i_r$ (multiplicity allowed, no ordering on the integers) for some $r \geq 0$ such that
$w_1 w_{i_1} w_{i_2} \ldots w_{i_r} = x_1 x_{i_1} x_{i_2} \ldots x_{i_r}$.
*Solution required to start with the first strings of $A$ And $B$. A solution may not always exist.

# MPCP and PCP

Lemma: MPCP is Turing reducible to PCP.
Thus, If MPCP is undecidable then so is PCP.

# MPCP and PCP contd.

Proof:

- $A = \{w_1, w_2, \ldots, w_k\}$ and $B = \{x_1, x_2, \ldots, x_k\}$ of strings over $\Sigma$ are an input instance of MPCP.
- We construct an instance of PCP such that the given instance of MPCP has a solution if and only if the constructed instance of PCP has a solution.

# MPCP and PCP contd.

Proof contd.:

- Let symbols $\vdash$ and $\$$ not be in $\Sigma$. Construct new alphabet $\Sigma' = \Sigma \cup \{\vdash, \$\}$.
- Over $\Sigma'$, construct set $C = \{y_0, y_1, y_2, \ldots, y_{k+1}\}$, where for $1 \leq i \leq k$, $y_i$ is obtained from $w_i$ by inserting the symbol $\vdash$ after each alphabet,
  $y_0 = \vdash y_1$ and $y_{k+1} = \$$.
- Over $\Sigma'$, construct set $D = \{z_0, z_1, z_2, \ldots, z_{k+1}\}$, where for $1 \leq i \leq k$, $z_i$ is obtained from $x_i$ by inserting the symbol $\vdash$ before each alphabet,
  $z_0 = z_1$ and $z_{k+1} = \vdash \$$.

# MPCP and PCP contd.

Proof contd.

- If $\{1, i_1, i_2, \ldots, i_r\}$ is a solution of instance $(\Sigma, A, B)$ of MPCP, then $\{0, i_1, i_2, \ldots, i_r, k+1\}$ is a solution of instance $(\Sigma', C, D)$ of PCP.
- If $\{i_1, i_2, \ldots i_r\}$, $r \geq 1$ is a solution for PCP then $i_1 = 0$ and $i_r = k+1$ as $y_0$ and $z_0$ are the only words with the same index that start with the same symbol, and $y_{k+1}$ and $z_{k+1}$ are the only words with same index that end with the same symbol. Let $i_j$ be the smallest integer where $i_j = k+1$. Then $\{i_1, i_2, \ldots, i_j\}$ is also a solution of the PCP instance: The symbol $\$$ only occurs as last symbol of $y_{k+1}$ and $z_{k+1}$ and for no $1 \leq \ell < j$ is $i_\ell = k+1$.
  $\{1, i_2, i_3, \ldots, i_{j-1}\}$ is a solution for the instance $(\Sigma, A, B)$ of MPCP.

# MPCP is undecidable

Proof:

- We reduce Membership problem MP ($\{M\#x | x \in L(M)\}$) to MPCP.
- Recall that the current configuration of a Turing machine can be denoted as $\alpha p \beta$, where $\alpha\beta$ is the current tape content, $p$ is the current state, and the current position of the tape head is at the first alphabet of $\beta$.
- If $M\#x$ is a yes instance of MP, there it can be captured by a finite sequence of configurations $(q_0 w, \alpha_1 q_1 \beta_1, \alpha_2 q_2 \beta_2, \ldots, \alpha_k q_k \beta_k)$, where $q_0$ is the start state, $w$ is the input string and $q_k$ is a final state.
- Our constructed instance of MPCP will be such that there will be a solution if and only if $M\#x \in MP$ and the solution will create the string $\#q_0 w \# \alpha_1 q_1 \beta_1 \# \alpha_2 q_2 \beta_2 \# \ldots \# \alpha_k q_k \beta_k \#$.

# MPCP is undecidable contd.

Proof contd.:

- We describe the sets $A$ and $B$ of strings for the MPCP instance constructed by the reduction.
- The first pair for $A$ and $B$ will be $\#$ and $\#q_0w\#$ respectively (starting off according to accepting sequence of $M\#x$).
- Group I (matching tape content and accepting sequence separator):
  $X \in A$ and $X \in B$ for each $X \in \Gamma$
  $\# \in A$ and $\# \in B$.

# MPCP is undecidable contd.

Proof contd.:

- Group II (copying non-final state transitions): For each $q \in Q - F$ (non-final) and $X, Y, Z \in \Gamma$:
  $qX \in A$ and $Yp \in B$ if $\delta(q, X) = (p, Y, R)$
  $ZqX \in A$ and $pZY \in B$ if $\delta(q, X) = (p, Y, L)$
  $q\# \in A$ and $Yp\# \in B$ if $\delta(q, B) = (p, Y, R)$
  $Zq\# \in A$ and $pZY\# \in B$ if $\delta(q, B) = (p, Y, L)$

# MPCP is undecidable contd.

Proof contd.:

- Group III (clearing out the tape contents after final state): For each $q \in F$ (final state, which does not change once entered) and $X, Y \in \Gamma$:
  $XqY \in A$ and $q \in B$
  $Xq \in A$ and $q \in B$
  $qY \in A$ and $q\# \in B$

# MPCP is undecidable contd.

Proof contd.:

- Group IV (final matching):
  $q\#\# \in A$ and $\# \in B$ for each $q \in F$.

# MPCP is undecidable contd.

Proof contd.:

- $(x, y)$ is a partial MPCP solution if $x$ is a prefix of $y$, and $x, y$ are concatenations of corresponding strings from $A$ and $B$. If $xz = y$ then $z$ is called the remainder.
- Suppose the accepting sequence of configurations starting with $q_0w$ has $k$ configurations then there is a partial solution $(x, y) = (\#q_0w\#\alpha_1q_1\beta_1\#\alpha_2q_2\beta_2\#\ldots\#\alpha_{k-1}q_{k-1}\beta_{k-1}\#,$ $\#q_0w\#\alpha_1q_1\beta_1\#\alpha_2q_2\beta_2\#\ldots\#\alpha_kq_k\beta_k\#)$
  such that this is the only partial solution whose larger string is as long as $|y|$ (Can prove by induction on $k$, using the description of first pair of strings $\#$ and $\#q_0w\#$, and the strings in Groups I - III).

# MPCP is undecidable contd.

Proof contd.:

- Suppose $q_k \in F$ then it is possible to derive a solution from the partial solution, by first using pairs from Groups I and III, and then using the pair in Group IV once. Note that this happens when $M\#x$ is in MP.
- If $M\#x$ is not in MP, then in the partial solution $(x, y)$ at any stage no pairs from Groups III and IV can be used. Therefore the partial solution will always have $|y| > |x|$ and cannot be converted into a solution.
- Reduction from MP to MPCP $\implies$ MPCP undecidable $\implies$ PCP undecidable.

# Practice Problems

1. Provide a solution for the following instance of PCP over $\Sigma = \{0, 1\}$:
   $A = \{110, 0, 01\}, B = \{11, 100, 00\}$
2. Show that PCP is decidable over the unary alphabet $\{1\}$.
3. Show that the language $\text{PF} = \{G | G \text{ is a CFG and L(G) is prefix-free}\}$ is undecidable. The language $L(G)$ is such that for no $u, v \in L(G)$ is $u$ a prefix of $v$ or vice-versa. Hint: Reduce from $\neg PCP$.

(Please try the problems first. Solutions in next slide.)

# Problem 1

Solution: $m = 4$, $\{1, 3, 1, 2\}$.

# Problem 2

- Let $A = \{w_1, w_2, \ldots, w_n\}, B = \{x_1, x_2, \ldots, x_n\}$ be an instance of PCP over $\Sigma = \{1\}$.
- for any $i$, if $|w_i| = |x_i|$ then as $\Sigma = \{1\}$, $w_i = x_i$ and we have found a solution.
- If for each $i$, $|w_i| > |x_i|$, then there is no solution. Similarly, if for each $i$, $|x_i| > |w_i|$
- Now, there is an $i$ where $|w_i| - |x_i| = a > 0$ and a $j$ where $|x_j| - |w_j| = b > 0$ : Solution is $m = b + a$, a sequence where $i$ appears $b$ times and $j$ appears $a$ times.

# Problem 3

- If PCP is undecidable then so is $\neg PCP$. We give a reduction from $\neg PCP$ to PF.
- $A = \{w_1, \ldots, w_k\}, B = \{x_1, x_2, \ldots, x_k\}$ is an instance of $\neg PCP$ over alphabet $\Sigma$.
- New distinct symbols introduced to form $\Sigma'$: $a_1, a_2, \ldots, a_k, \#, \dashv$.

# Problem 3 contd.

- Define CFG $G = (N = \{S, S_A, S_B\}, \Sigma', P, S)$ with the following productions:
  $S \to S_A \# \dashv | S + B\#$,
  $S_A \to w_i S_A a_i | w_i a_i$ for $1 \leq i \leq k$,
  $S_B \to x_i S_A a_i | x_i a_i$ for $1 \leq i \leq k$.

# Problem 3 contd.

- All strings derived from $S \to S_A\# \dashv$ end with $\# \dashv$ and all strings derived from $S \to S_B\#$ end with $\#$.
- Suppose for contradiction, there are distinct strings $u, v \in L(G)$ such that $u$ is a prefix of $v$.
- All symbols in $u$ and $v$ upto and including $\#$ must match. It must be that $u = u'\#, v = v'\# \dashv$ such that $u' = v'$ , $S_A \to^* v'$ and $S_B \to^* u'$.

# Problem 3 contd.

- We show that $(A, B) \in \neg PCP$ iff $L(G)$ is prefix-free.
- One direction: Suppose that $(A, B) \notin \neg PCP$. For a solution $\{i_1, i_2, \ldots, i_m\}$ , we have $w_{i_1} w_{i_2} \ldots w_{i_m} = x_{i_1} x_{i_2} \ldots x_{i_m}$.
- Let $z = w_{i_1} w_{i_2} \ldots w_{i_m} a_{i_m} \ldots a_{i_2} a_{i_1} = x_{i_1} x_{i_2} \ldots x_{i_m} a_{i_m} \ldots a_{i_2} a_{i_1}$.
- By the grammar $G$, $L(G)$ contains both $z\#a$ and $z\#$ and hence is not prefix-free.

# Problem 3 contd.

- Other direction: Suppose there are $u, v \in L(G)$ such that $u$ is a prefix of $v$. Then, $u = u'\#$, $v = v'\# \dashv$ such that $u' = v'$, $S_A \rightarrow^* v'$ and $S_B \rightarrow^* u'$.
- The string $v'$ , derived from $S_A$ , must be of the form $w_{i_1} w_{i_2} \ldots w_{i_m} a_{i_m} \ldots a_{i_2} a_{i_1}$. Similarly, $u'$ has the form $x_{i_1} x_{i_2} \ldots x_{i_m} a_{i_m} \ldots a_{i_2} a_{i_1}$.
- The $a_i$'s at the end must all match since $u' = v'$.
- So, $w_{i_1} w_{i_2} \ldots w_{i_m} = x_{i_1} x_{i_2} \ldots x_{i_m}$, implying that $\{i_1, i_2, \ldots, i_m\}$ is a solution for $(A, B)$ i.e., $(A, B) \notin \neg PCP$.
- Therefore PF is undecidable.