Improved Algorithms for the Boxed-Mesh Permutation Pattern Matching Problem

Sukhyeun Cho, Joong Chae Na,
and Jeong Seop Sim

TCLAB Theory of Computation Laboratory in INHA University

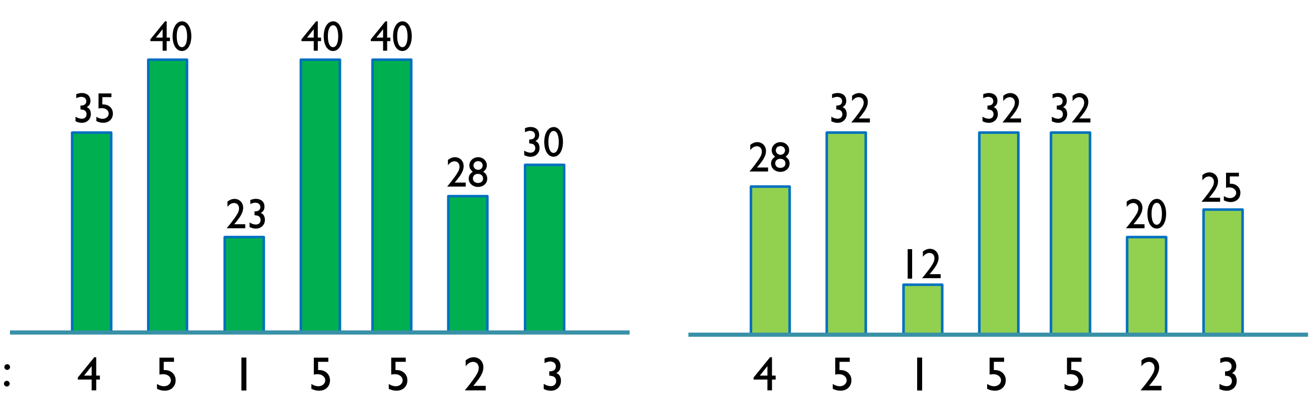
Outline

- Introduction
- Main Idea
- $O(n^2 \log m)$ Algorithm

Definitions

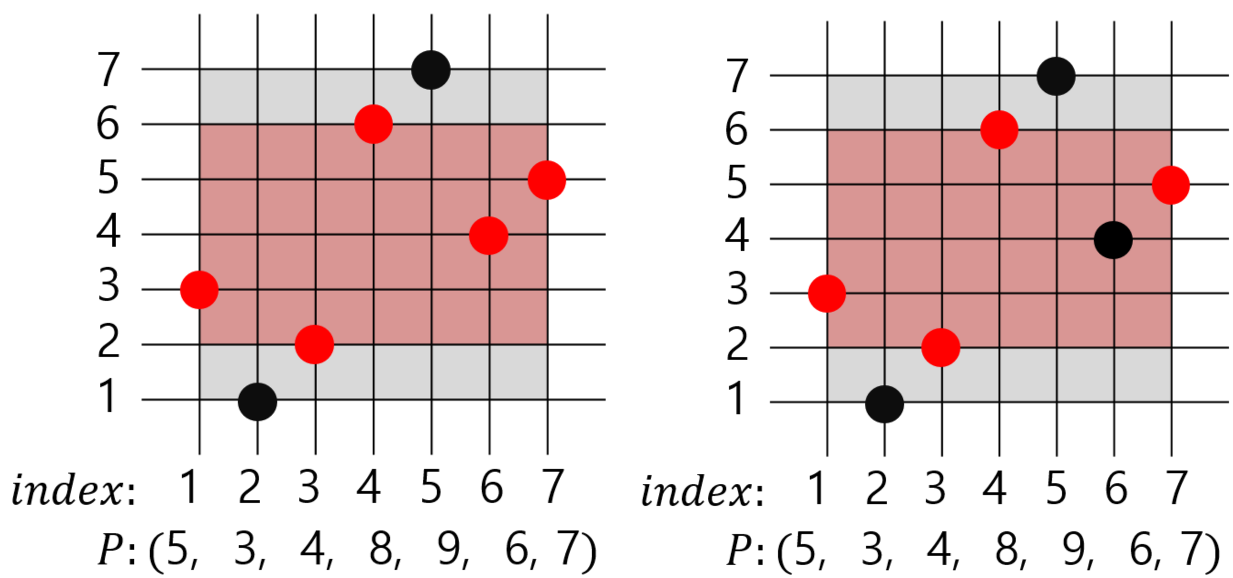
- Order-Isomorphism
  - $x,\ y$: Two (numeric) strings
  - $x \approx y$: x and y are order-isomorphic
    - $|x| = |y|$
    - $x[i] \leq x[j] \Leftrightarrow y[i] \leq y[j]$ for $0 \leq i, j < |x|$
  - x and y have the same relative orders.
    - ex) $x = (35, 40, 23, 40, 40, 28, 30)$, $y = (28, 32, 12, 32, 32, 20, 25)$

Orders :

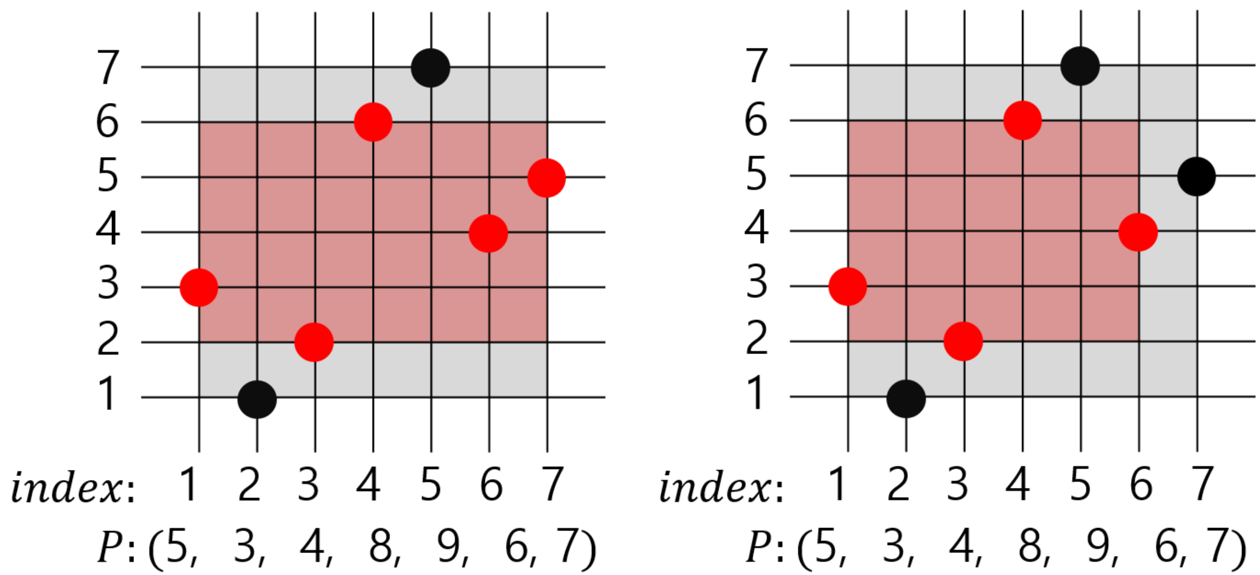


- Two-dimensional representation of string x
  - Mark a point of coordinate $(i, rank(x, c))$ for the i-th character c of x (cf. $rank(x, c)$: rank of c on x)

- Boxed Subsequence
  - If all the character in a subsequence y of x are included in a rectangle without any other character, then y is a boxed subsequence of x
  - Ex) (5, 4, 8, 6, 7) => boxed sub., (5, 4, 8, 7) => no boxed sub.



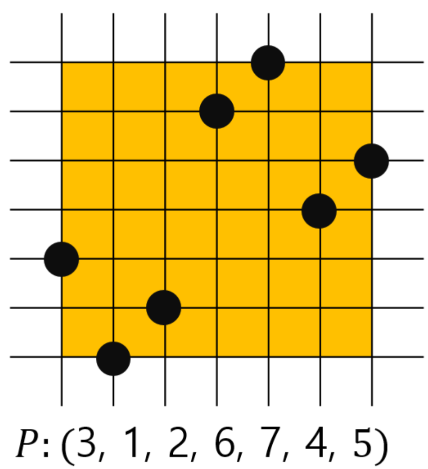
- Full-Width Boxed Subsequence
  - A subsequence y of string x is called a full-width boxed subsequence of x if y includes only and all the characters c of x such that $\min(y) \le c \le \max(y)$
  - Ex) (5, 4, 8, 6, 7) => full-width boxed sub. of P,
    (5, 4, 8, 6) => no full-width boxed sub. of P since $4 \le 7 \le 8,\ 7 \in P$

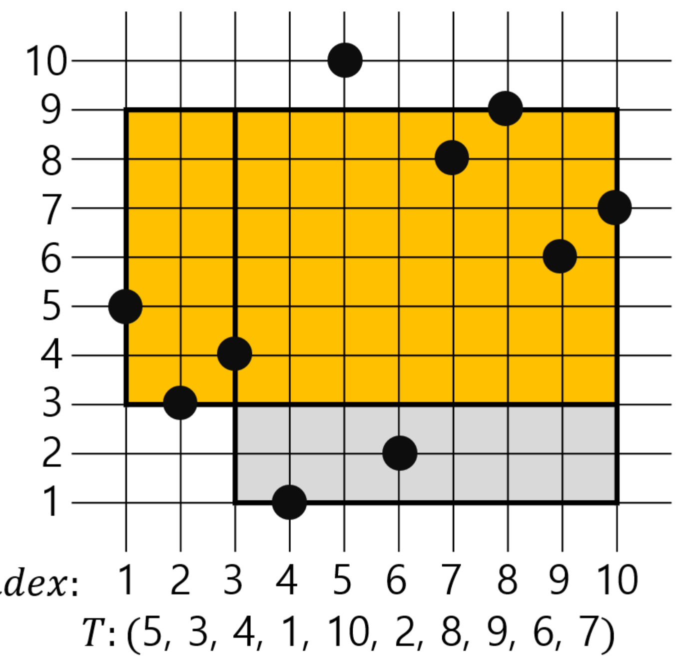


Boxed-Mesh Permutation Pattern Matching

- Problem Definition
  - **Input:** A text $T[1..n]$ and a pattern $P[1..m]$
  - **Output:** All positions of T' such that $T' \approx P$ and T' is a boxed subsequence of T
  - Ex) Two occurrences : $T' = (5, 3, 4, 8, 9, 6, 7), (4, 1, 2, 8, 9, 6, 7)$

Pattern P





Text T

Contributions

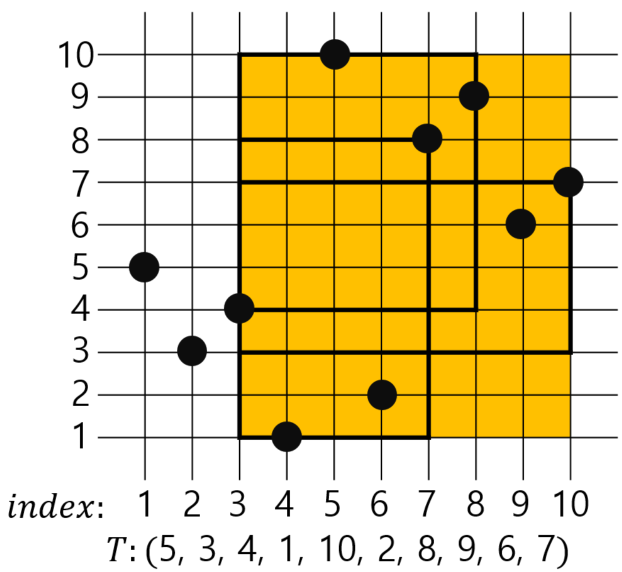
- Improved algorithms for the boxed-mesh permutation pattern matching problem
  - Previous work : $O(n^3)$ algorithm by Bruner et al.
  - $O(n^2\, m)$ time algorithm
  - $O(n^2 \log m)$ time algorithm
    - Remove duplicated comparisons

Outline

- Introduction
- **Main Idea**
- $O(n^2 \log m)$ Algorithm

Algorithm Outline

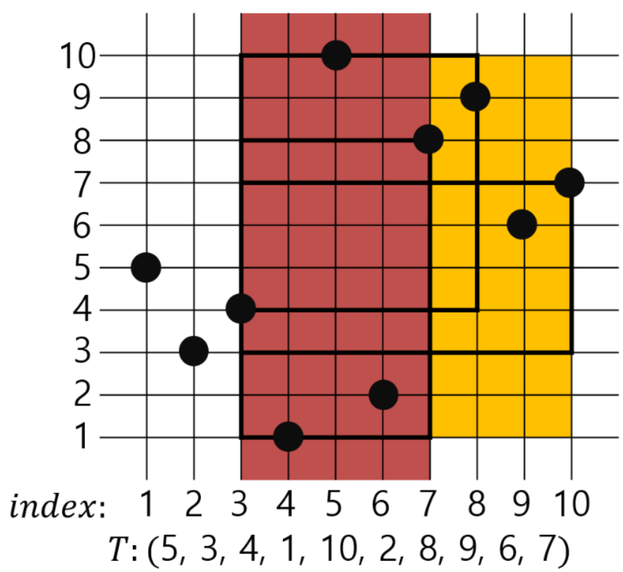
- Algorithm Outline : $Phase\ i$
  - Our algorithm consists of $n - m + 1$ phases.
  - In each $Phase\ i$, we find all the occurrences of P whose first character is $T[i]$
  - Each phase is completely independent of the other phases



Ex) $Phase\ 3$,
Find all occurrences
such that $T' = (T[3], \dots)$,
T' is boxed sub. and $T' \approx P$

Algorithm Outline

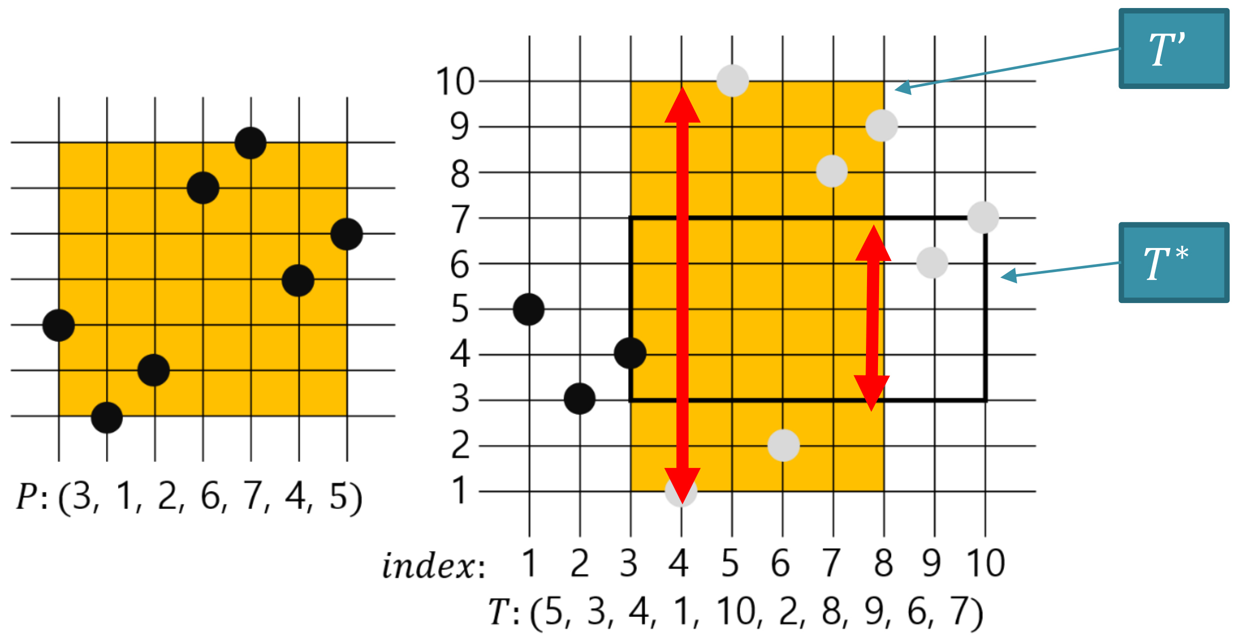
- Algorithm Outline : $Step\ j$
  - In $Step\ j$ on $Phase\ i$, we check that a subsequence $T' = (T[i], \ldots, T[j])$ is an occurrence (from $j = i$ to $j = n$)
  - Naive approach produces duplicated comparisons between $Step\ j - 1$ and $Step\ j$



Ex) $Step$ 7 in $Phase$ 3,
Check $T' = (T[3], \ldots, T[7])$
is an occurrence

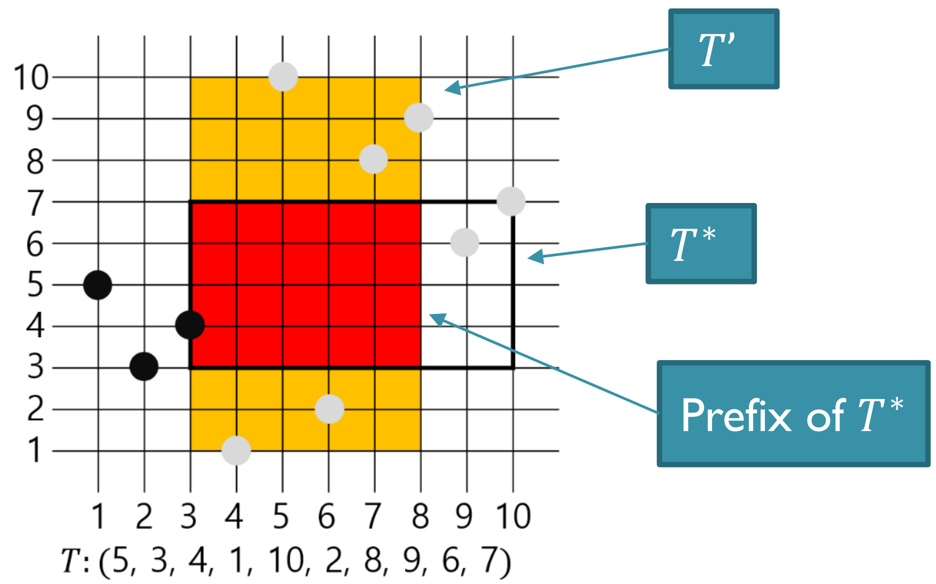
Properties

- Vertical size of the box always decreases
  - Assume the orange box (i.e., $T' = (T[3], \dots, T[8])$ is an occurrence
  - Note that T' is unique since the number of characters larger than (resp. smaller than) $T[3]$ must be 4 (resp. 2) to be order-isomorphic to P
  - If $T^* = (T[3], \dots, T[10])$ is an occurrence, the vertical size of box must be smaller than the box of T' since $T^* \approx P$



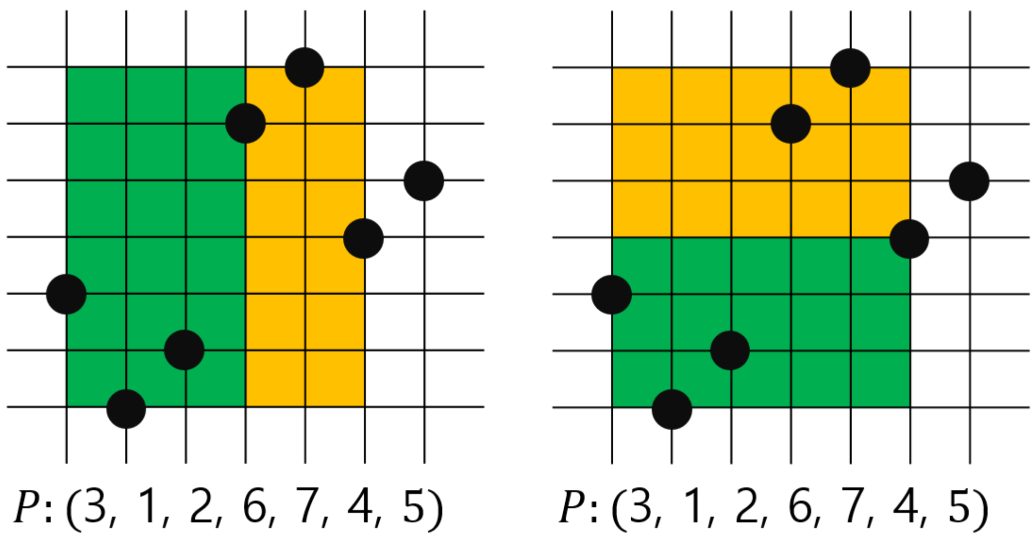
Properties

- The box of prefix of T^* is always included in box of T'
  - It comes from the fact "vertical size of the box always decreases"



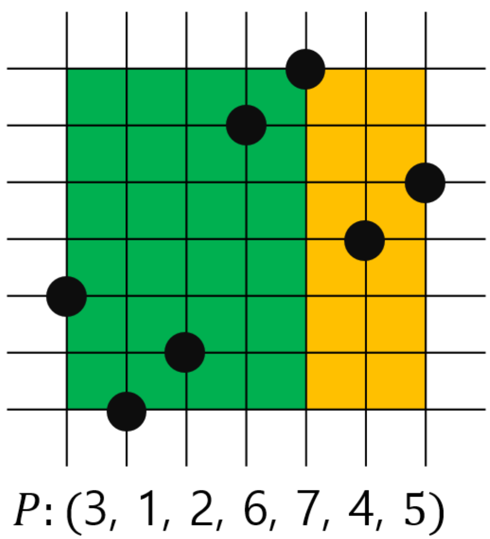
π function

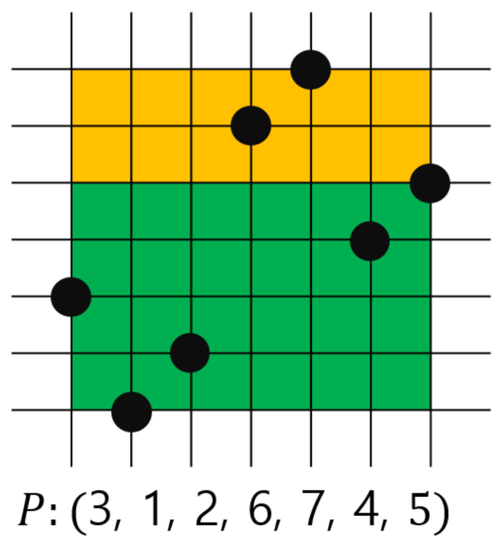
- π function
  - Let z be a full-width boxed subsequence of $P[1..i]$ such that $z[1] = P[1]$ and $|z| < i$
  - Then, $\pi[i]$ is the length of longest subsequence z such that $z \approx P[1..|z|]$
  - Ex) $\pi[5] = 4$



π function

- π function
  - Let z be a full-width boxed subsequence of $P[1..i]$ such that $z[1] = P[1]$ and $|z| < i$
  - Then, $\pi[i]$ is the length of longest subsequence z such that $z \approx P[1..|z|]$
  - Ex) $\pi[7] = 5$



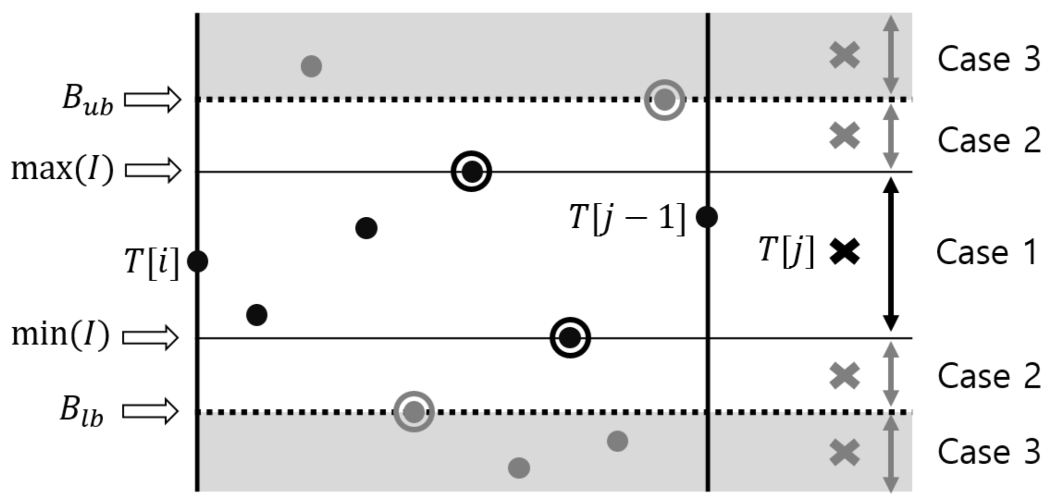


Outline

- Introduction
- Main Idea
- **$O(n^2 \log m)$ Algorithm**

Main logic in *Step j* on *Phase i*

- Let I be the longest full-width boxed subsequence of $T[i..j-1]$ such that $I[1] = T[i]$ and $|I| < m$
- If $(I, T[j])$ is full-width boxed sub. : Case 1, 2
  - If $(I, T[j]) \approx P[1..|I|+1]$, then set $I = (I, T[j])$: Case 1, 2
  - If not $(I, T[j]) \approx P[1..|I|+1]$
    - If Case 1, then calculate next I^* using π function such that $|I^*| < |I|$
    - If Case 2, then go to next step
- If $(I, T[j])$ is not full-width boxed sub., then go to next step : Case 3



Thank You