KMP – prefix suffix table

Source: http://mathcs.emory.edu/~cheung/Courses/323/Syllabus/Text/Matching-KMP2.html

pattern p: $p_0 \; p_1 \; \cdots \; p_{m-1}$

$PS(k) =$ longest proper prefix of $p_0 p_1 \cdots p_k$ which is also a suffix of $p_0 p_1 \cdots p_k$

Goal: Compute $PS(i) \;\; \forall \; i < m$

$PS(0) = 0$ (there are no proper prefixes of p_0)

Suppose you know $PS(k-1) = x$

$$\overbrace{p_0 \, p_1 \cdots p_{x-1}}^{x} \; p_x \; \cdots \cdots p_{k-2} \, p_{k-1} \; p_k$$

match

$$p_0 \, p_1 \cdots\cdots p_{k-x-1} \; \underbrace{p_{k-x} \cdots p_{k-2} \, p_{k-1}}_{x} \; p_k$$

$p_x = p_k$,

then $PS(k) \geq x+1$

(because $p_0 p_1 \cdots p_x$ will be a proper prefix which works)

$p_x \neq p_k$

??

Claim: $PS(k) \leq PS(k-1)+1$ (in any scenario)

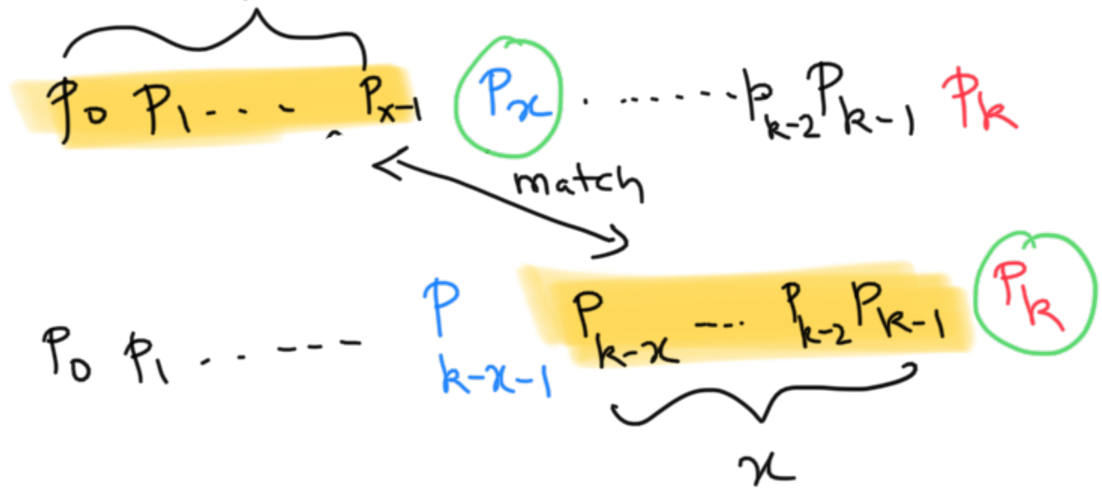
Pf: Let $PS(k-1) = x$. Suppose $PS(k) > x+1$.

let $\underbrace{p_i\, p_{i+1} \cdots p_{k-1}\, p_k}_{\geq x+2}$ be the longest suffix of $p_0 \cdots p_k$

which is also a prefix of it. Then $\underbrace{p_i \cdots p_{k-1}}_{\geq x+1}$ will a suffix of $p_0 \cdots p_{k-1}$ which is also a prefix of it.

But $PS(k-1) = x$! $\rightarrow \leftarrow$

Suppose you know $PS(k-1) = x$



if $P_x = P_k$,

then $PS(k) \geq x+1$

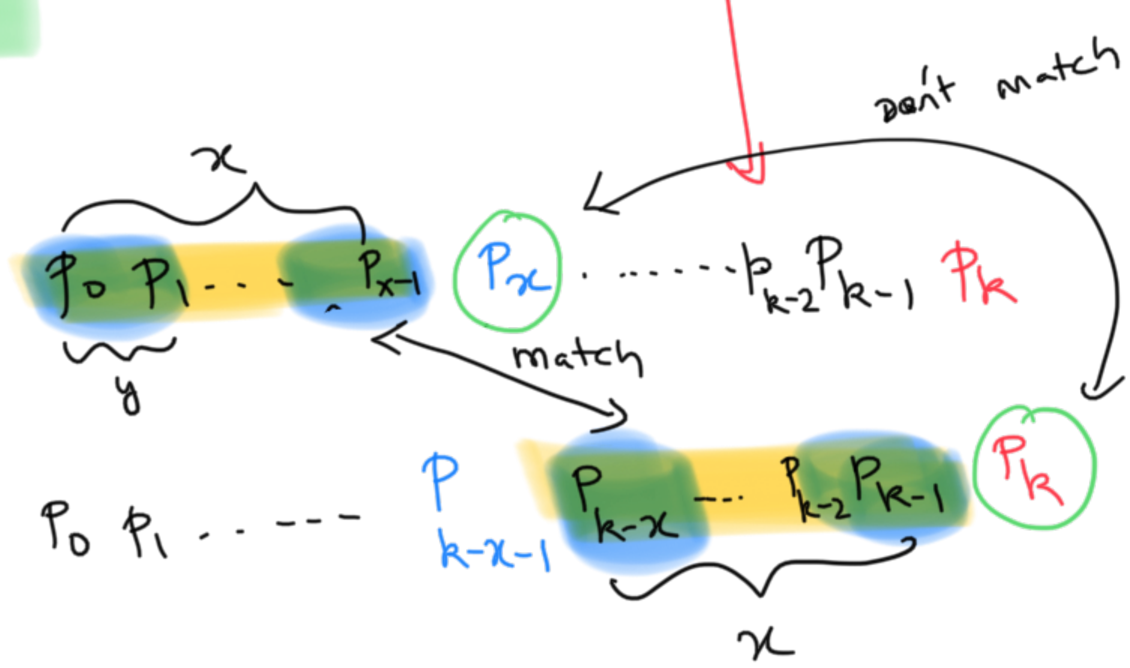
and $PS(k) \leq x+1$

so $PS(k) = x+1$

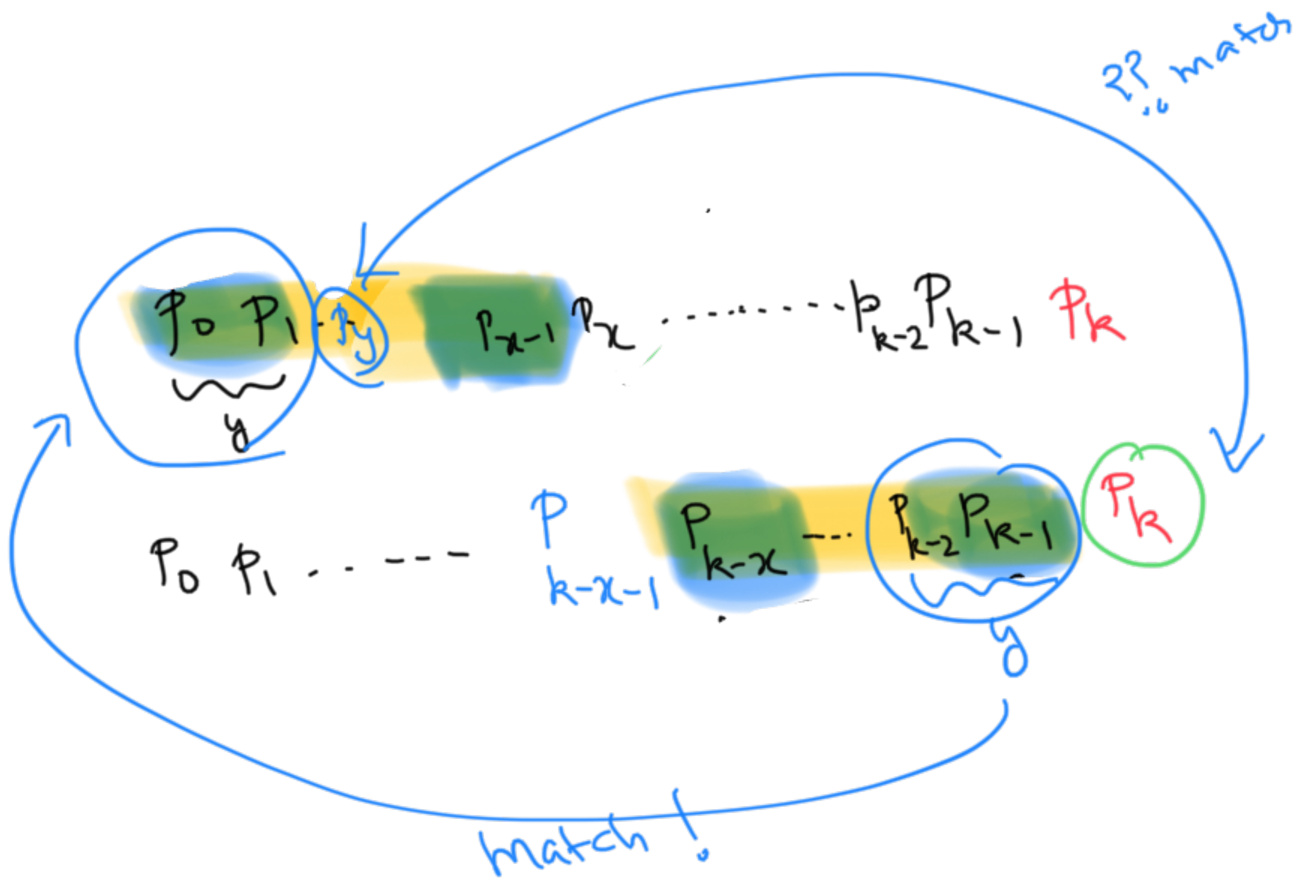
(ie) $PS(k) = PS(k-1)+1$

if $P_x \neq P_k$

Let $PS(x-1) = y$



Next valid prefix to try and match a suffix to $p_0 p_1 - p_k$



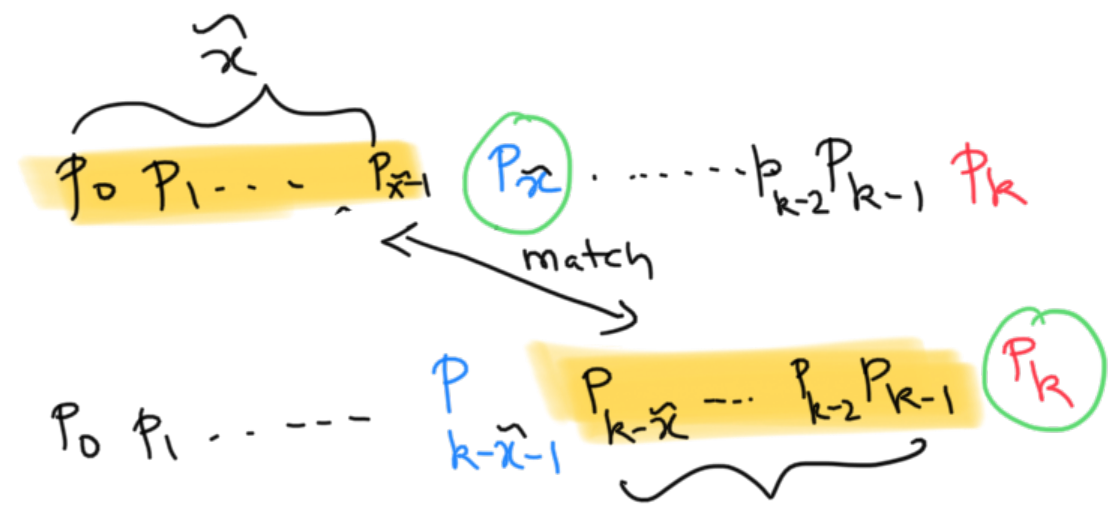
→ if $p_y = p_k$, $PS(k) = y+1$

↓ if $p_y \neq p_k$, repeat this process!

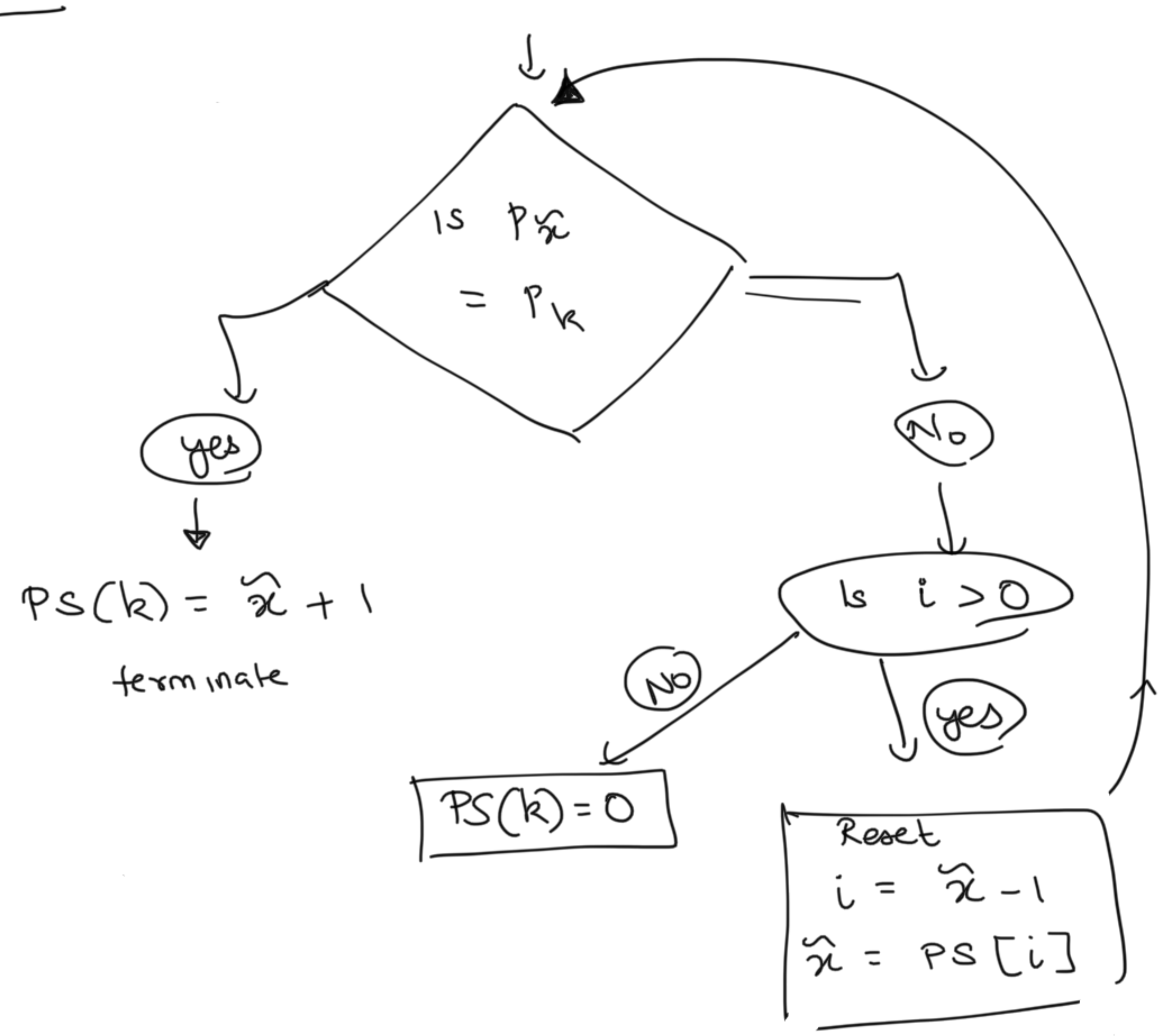
So $PS(0) = 0$

$PS(k-1) = \hat{x}$

$PS(k) = ?$



<u>Define</u> $i = k-1$ $\quad \hat{x} = PS[i]$



Dry Run

P : P_0 P_1 P_2 P_3 P_4 P_5

a b a b a a

$PS(0) = 0$

$PS(1) =$

$i = 1-1 = 0$ $\hat{x} = PS[0] = 0$

Is $P_{\hat{x}} = P_1$?

Is $P_0 = P_1$?

No

$i = 0$,

so $PS(1) = 0$

ab|ab aa
ab|ab aa

$PS(2) =$

$i = 2-1 = 1$ $\hat{x} = PS[1] = 0$

Is $P_{\hat{x}} = P_2$?

Is $P_0 = P_2$?

yes!

$PS(2) = PS(1) + 1$
$= 1$

ab ab aa
ab ab aa

$PS(3) =$

$i = 3-1 = 2$ $\hat{x} = PS[2] = 1$

Is $P_{\hat{x}} = P_3$?

Is $P_1 = P_3$?

yes

$PS(3) = PS(2) + 1 = 2$

abab|aa
abab|aa

$PS(4) =$

$i = 4-1 = 3$ $\quad \hat{x} = PS[3] = 2$

Is $P_{\hat{x}} = P_4$?

Is $P_2 = P_4$?

ababaa

ababaa

yes

$PS(4) = PS(3)+1$

$= 3$

$PS(5) =$

$i = 5-1 = 4$ $\quad \hat{x} = PS[4] = 3$

Is $P_{\hat{x}} = P_5$?

Is $P_3 = P_5$?

Is $P_1 = P_5$?

ababaa

ababaa

No

Is $4 > 0$

yes

$i = 3-1 = 2$

$\hat{x} = PS[i] = PS[2]$

$= 1$

$\hat{x} = 1$

$i = 2$

No

is $2 > 0$

yes

$i = 1-1 = 0$

$\hat{x} = ps[i] = 0$

$\hat{x} = 0,\ i = 0$

Is $P_0 = P_5$?

yes

$PS(5) = PS(0)+1$

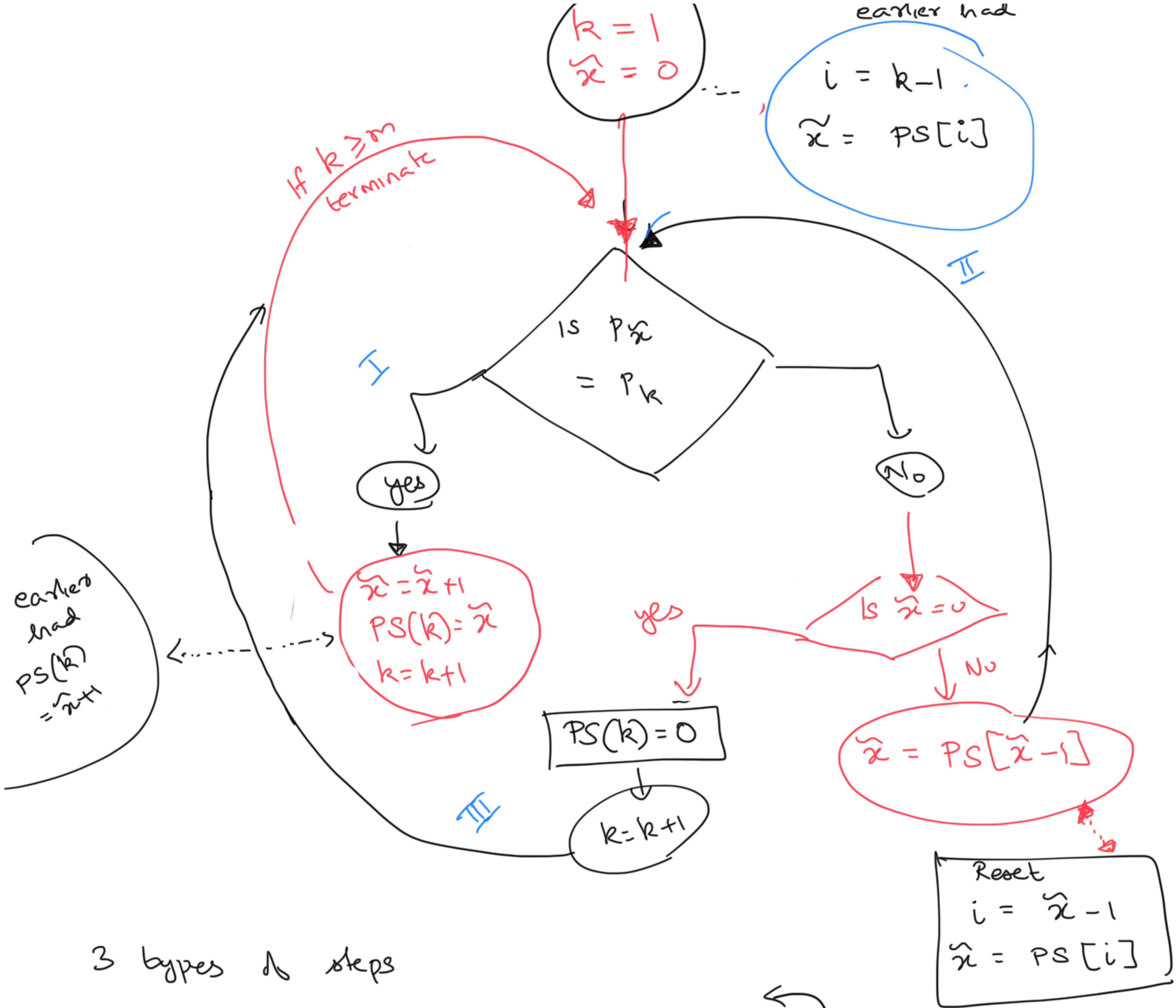
$= 1$

Let us modify the flowchart to make it simpler

* Goal: Find $PS(k)$

$\forall \quad k \leq m$

* start with $PS(0) = 0$



3 types of steps

I

$k = k+1$

$\hat{x} = \hat{x}+1$

II

$\hat{x}$ dec by atleast 2.

$(\hat{x} > 0)$

k same

III

$\hat{x} = 0$

$k = k+1$

$$\begin{array}{ll} a & \mathrm{I} \\ b & \mathrm{II} \\ c & \mathrm{III} \end{array} \Rightarrow \begin{array}{l} a+c \le m \\ a-b \ge 0 \end{array}$$

$\Rightarrow \quad a \ge b$

$\Rightarrow \quad a+b+c$

$\le a+a+c$

$\le a+a+c+c$

$= 2m.$ ∎