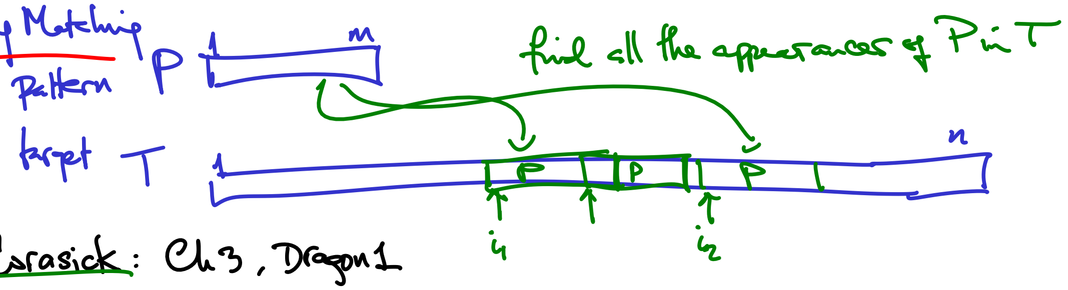
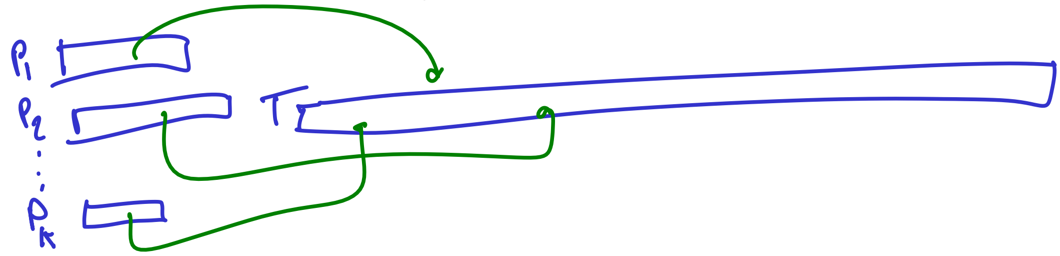
CS 445 2/14/19 ① Knuth-Morris-Pratt : Ch 32 Cormen et al "Intro to Algorithms"

String Matching



find all the appearances of P in T

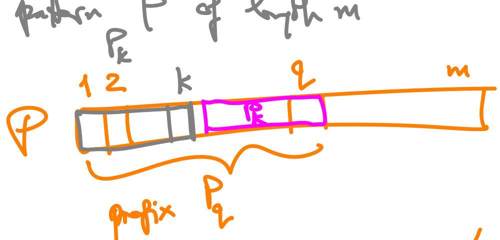
② Aho-Corasick : Ch 3, Dragon 1

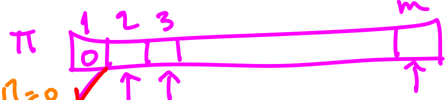


Knuth-Morris-Pratt: prefix function

$$\boxed{\pi(q) \doteq \max\{k : k < q,\ P_k \sqsupset P_q\}}$$

fix pattern P of length m





π[1]=0 ✓

```
π[1] ← 0 ✓
k ← 0 ✓
for q = (2) to m : /* compute π[q] */
    /* invariant: k = π[q-1] */
    while k>0 & P[k+1] ≠ P[q] :   } k∈π*[q-1]
        (k ← π[k])
    if P[k+1] == P[q]:
        k ← k+1
    π[q] ← k ✓
```

$$\pi^*(q) = \{\underline{\pi(q)}, \pi^{(2)}(q), \pi^{(3)}(q), \dots\}$$

Fact 1: $\pi^*(q) = \{k : k < q,\ P_k \sqsupset P_q\}$

Fact 2: If $\pi(q) > 0$, then $\pi(q) - 1 \in \pi^*(q-1)$

Fact 3 If $\mathcal{E}_{q-1} = \phi$ then $\pi(q) = 0$

else $\pi(q) = 1 + \max\{k : k \in \mathcal{E}_{q-1}\}$

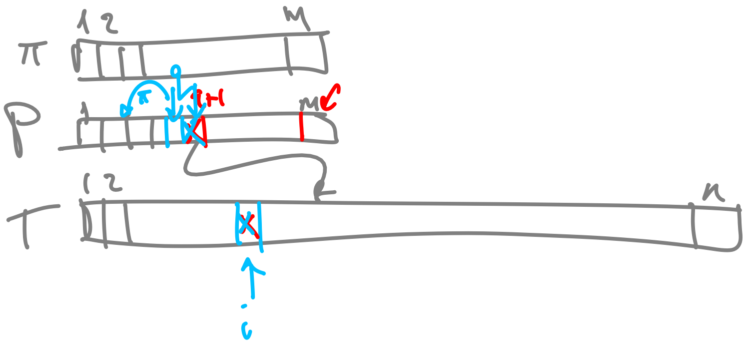
$$\mathcal{E}_{q-1} = \{k \in \pi^*(q-1) : P[k+1] = P[q]\}$$

$$k = \pi[q-1] \xrightarrow{\pi} \pi[k] \rightarrow \pi(\pi[k]) \rightarrow \rightarrow \rightarrow$$

Knuth-Morris-Pratt (KMP): P, T

pattern, target

```
q ← 0
for i = 1 to n:
    while q > 0 & P[q+1] ≠ T[i]:
        (q ← π[q])
    if (P[q+1] == T[i]):
        q ← q+1
    if (q == m):
        Match
```



Fact 1: $\pi^*(q) = \{k : k < q, P_k \sqsupset P_q\}$

$$\pi^*(q) \overset{\text{def}}{\doteq} \{\underbrace{\pi(q)}_{\pi(q)<q}, \underbrace{\pi(\pi(q))}_{\pi^{(2)}(q)}, \underbrace{\pi(\pi(\pi(q)))}_{\pi^{(3)}(q)}, \dots, 0\}$$

$u=1$ ✓, $u=2$ ✓, $u=3$ ✓ u, $u+1$

proof: ($\subseteq$) Show $\pi^*(q) \subseteq \{k : k<q, P_k \sqsupset P_q\}$ ✓

Let $i \in \pi^*(q)$. This means $i = \pi^{(u)}(q)$. Must show $\boxed{P_i \sqsupset P_q}$ and $\underbrace{i < q}_{\text{obvious}}$.

Use induction on u:

Base case: $u=1$ $\quad i = \pi(q) = \max\{k : k<q, P_k \sqsupset P_q\}$ ✓

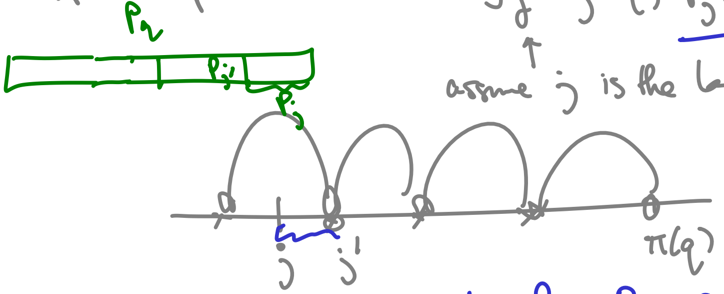
Inductive case: Say claim holds for some u, now show it is true for $u+1$

$$\underset{\pi^{(u+1)}(q)}{\pi(i)} < \underset{\pi^{(u)}(q)}{i} < q, \quad P_{\pi(i)} \sqsupset P_i \sqsupset P_q \quad ✓$$

($\supseteq$) Must show $\{k : k<q, P_k \sqsupset P_q\} \subseteq \pi^*(q) = \{\pi(q), \pi(\pi(q)), \dots, 0\}$ ✓

Assume for contradiction $\exists j : j<q, \underline{P_j \sqsupset P_q}$ and $j \notin \pi^*(q)$.

assume j is the largest such integer



$j' = \underline{\text{smallest integer in } \pi^*(q)}$ that is greater than j.

$$P_j \sqsupset P_q,\ P_{j'} \sqsupset P_q \Rightarrow \underline{P_j \sqsupset P_{j'}} \Rightarrow \underbrace{\pi(j')}_{\max\{k : k<j', P_k \sqsupset P_{j'}\}} = j$$

Q1 How do we know j' exists?

Q2 How do we know $\pi[q] > j$?

□