CSL 630 Lecture 15 Sept 18

$$S \rightarrow bA \mid aB$$
$$A \rightarrow bAA \mid aS \mid a$$
$$B \rightarrow aBB \mid bS \mid b$$

CFG

NT. $\{S, A, B\}$

Terminals $\{a, b\}$

Start S

$$S \rightarrow bA \rightarrow baS \rightarrow baaB \rightarrow baab$$

$$S \xrightarrow{*} baab$$

Membership problem: Given a string

$$s = x_1 \, x_2 \, x_3 \, \ldots \, x_n$$

$x_i \in$ Terminal symbol

$$S \overset{?}{\underset{*}{\longrightarrow}} s$$

Arbitrary CFG can be transformed into Chomsky Normal Form (CNF)

$$A \rightarrow BC \mid a$$

$$S \rightarrow C_b A \mid C_a B$$
$$A \rightarrow C_a S \mid C_b D_1 \mid a$$
$$B \rightarrow C_b S \mid C_a D_2 \mid b$$
$$D_1 \rightarrow AA$$

$$D_2 \rightarrow BB$$
$$C_a \rightarrow a$$
$$C_b \rightarrow b$$

Given $x_1\ x_2 \cdots x_n$

Can we derive this from S

length $n=1$ trivial

Just check the unit productions

$$S \to a \mid b \mid c \mid$$

$$S_1 \xrightarrow{*} x_1\ x_2 \cdots x_n$$

$$S_1 \to A'B' \xrightarrow{*} \underbrace{x_1 \cdots x_i}_{A',\ A'',}\ \underbrace{x_{i+1}\ x_n}_{B',\ C,\ D}$$

$$A' \xrightarrow{*} \underbrace{x_1 \cdots x_i}_{S_{1,i}} \qquad B' \xrightarrow{*} \underbrace{x_{i+1} \cdot x_n}_{S_{i+1,\,n-i}}$$

$\boxed{S_{ij}}$: substring starting from x_i of length j

We want to address a more general problem, viz.

For any N.T. $A \xrightarrow{*} S_{ij}$

Given a substring S_{ij} which N.T.

$$A \xrightarrow{*} S_{ij}$$

For all s_{ij} $\quad 1 \le i \le n-1$

$\qquad j \le n-i+1$

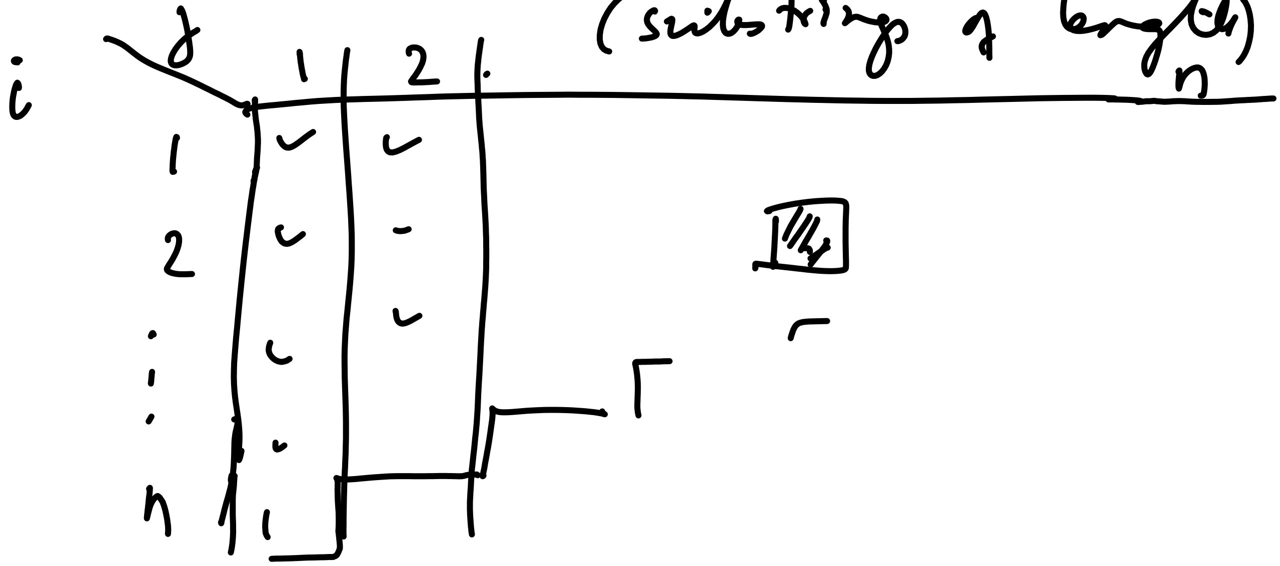
we want to know if $A \xrightarrow{*} s_{ij}$?

$$P_{ij} = \{ A \mid A \xrightarrow{*} s_{ij} \}$$

If $S \in P_{1n}$ then the string can be derived

$$P_{ij} = \{ A \mid A \to BC \text{ and } B \in P_{ik} \text{ and } C \in P_{i+k,\, j-k} \text{ for some } 1 \le k < j \}$$

We know how to compute P_{i1} (substrings of length 1)



Running time :

entries × time to compute each entry

In the j^{th} col we are filling up $n-j$ entries and we have to do j look ups

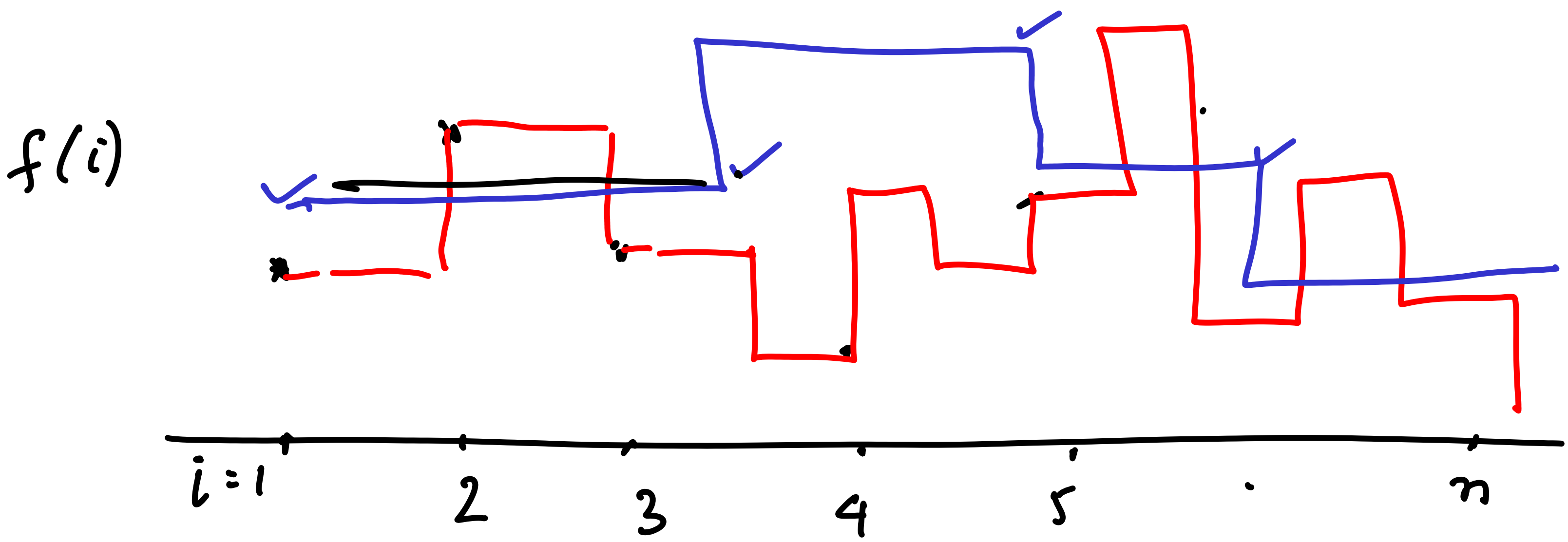
. Each lookup yields a set of non terminals. Suppose # NT = m

We could be considering m^2 combinations

$$\sum_{j=1}^{n} (n-j)(j) \cdot m^2 = m^2 \sum (n-j)(j)$$

$$\leq O(m^2 n^3)$$

Cocke - Young - Kasami (CYK)



f(i) is an integral function defined at $i = 1, 2, \ldots n$

$\{i, f(i)\}_{1 \le i \le n}$

$(i, g(i))$ where $g(i)$ is defined by K points $K \ll n$

We want to achieve a "good approximation" of f using g

L_∞ min $\max_i |g(i) - f(i)|$

L_1 min $\sum_i (g(i) - f(i))$

L_2 min $\sum_i (g(i) - f(i))^2$

Eg. of different metrics