COL 702 Lecture 17 Sept 25

$S \to bA \mid aB$

$A \to bAA \mid aS \mid a$

$B \to aBB \mid bS \mid b$

Given a context free grammar $G = (V, T, P)$

Start symbol : S — V: set of variables T: terminals; P: production rules

For given string $x_1\, x_2\, x_3 \ldots x_n$

$x_i \in T$

we want to determine if $S \xrightarrow{*} x_1 x_2 \ldots x_n$

The above grammar generate all strings over a, b s.t. $\#a\text{'s} = \#b\text{'s}$

<u>Possible approach</u>: Generate all strings of length n and check if the given string S is among them

If the grammar doesn't contain ϵ productions then this can be done in finite # steps. Potentially exponential

The derivation would depend on the way G is presented. The complexity of parsing would depend on the way a grammar is described.

Assume that G is given in CNF

Rule $A \to BC \mid a$

Given G in CNF and a length n string s $\quad S \xrightarrow{*} s$?

If $S \xrightarrow{*} x_1 x_2 \quad x_n$

Then $S \to AB \xrightarrow{*} x_1 x_2 \;.\;.\; x_n$

for some i $\quad A \xrightarrow{*} x_1 x_2 \cdots x_i$

$\qquad\qquad\qquad\quad B \xrightarrow{*} x_{i+1} x_{i+2} \cdots x_n$

unknown

$S \xrightarrow{*} x_1 \cdots x_n \iff S \to AB$, $A \xrightarrow{*} x_1 \cdots x_i$, $B \xrightarrow{*} x_i \cdots x_n$

The base case $A \to a$ is easy to check from the given rules

Since i is unknown we must check for all values of i $\quad 1 \le i \le n-1$

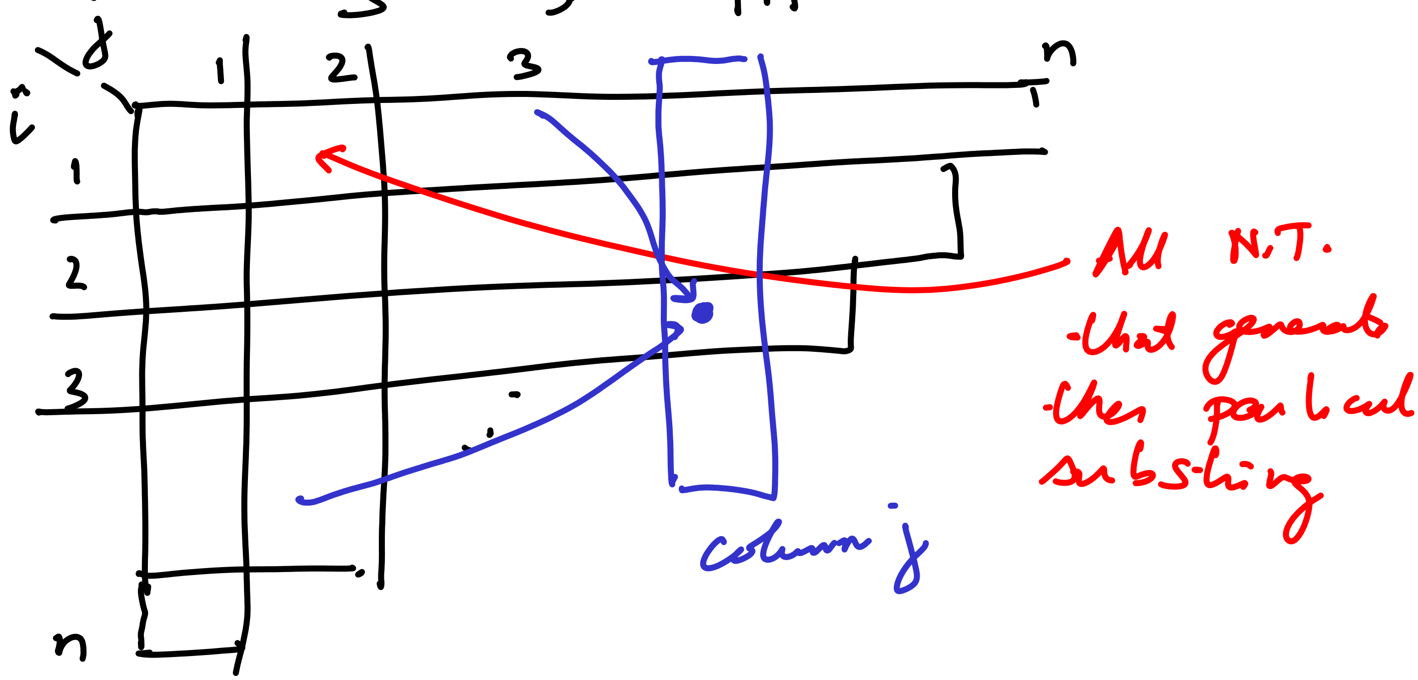
The general problem is that for some variable A and a substring

$$s_{ij} \quad \big(\underbrace{x_i x_{i+1} \cdots x_{i+j-1}}_{j}\big)$$

$$A \xrightarrow{\overset{?}{*}} s_{ij}$$ Note there are $O(n^2)$ s_{ij}'s

Original problem

$$S \xrightarrow{\overset{?}{*}} s_{1n}$$



In column j we are considering substrings of length j

$A' \xrightarrow{*} S_{ij}$ iff there is a rule of the form $A' \to B'C'$

and $B' \xrightarrow{*}$ a prefix of the string

$C' \xrightarrow{*}$ the remaining tail

For every fixed split we have to consult two entries in the table to see if they contain B' and C' respectively. If so then A' is entered in the table

How many splits? $j-1$ (all possible lengths)

To fill every entry in j^{th} col we are consulting $2 \cdot (j-1)$ entries

So fill j^{th} column we need $\leq 2 \cdot n(j-1)$

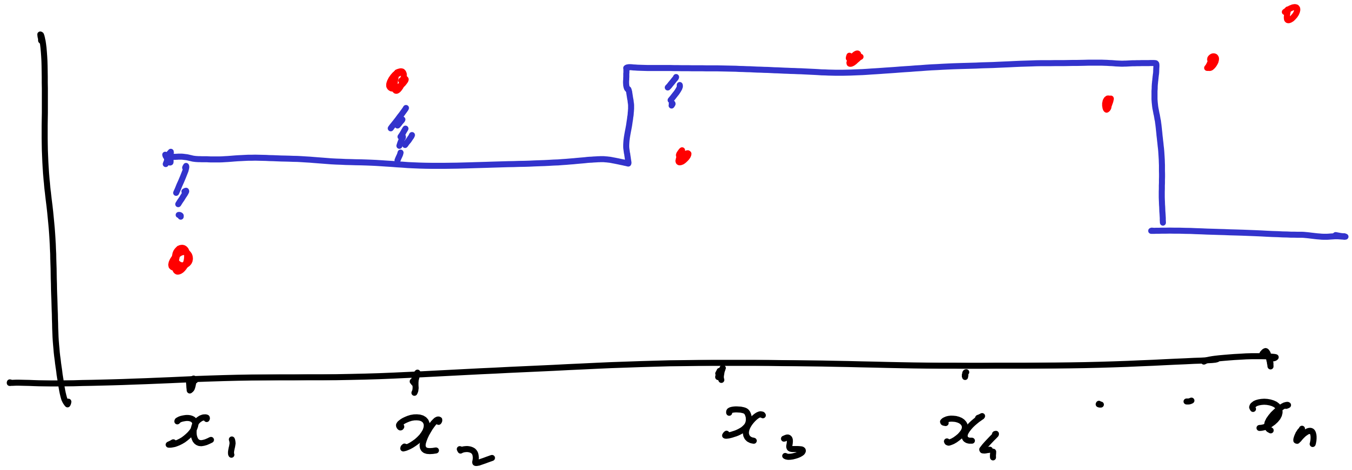
Total work in filling the entire table : $\sum_{j=1}^{n} 2n(j-1)$ which is $O(n^3)$

CYK (Cocke-Young-Kasami)

Suppose there are m productions in the grammar, then we can bound by $O(m^2 n^3)$

Function simplification

We have a integer valued function $x_i, f(x_i)$ given as tuples x_i are



we want to "compress" the representation using only k values $k \leq n$, s.t. the difference between the given function and the approximation is minimized according to some metric

Possible metrics : minimize maximum difference

: minimize average diff

: minimize sum of squares