Reading for pda & cfg

- Sipser: Chapter 2
- Hopcroft et al.: 5.1, 6.1-3

Assignment 1: Moodle 5 pm

Seek help from the tutor, if necessary.

Languages

Let L be a language over a certain alphabet.

- L is said to be a context free language if L = L(G) for some context free grammar G (or equivalently, L = L(A) for some pda A).
- L is said to be a regular language if L = L(G) for some right linear grammar G (or equivalently, L = L(M) for some dfa/nfa M).

Pumping Lemma for regular languages

Theorem Let L be a regular language. There exists a constant n such that for **any** string $w \in L$ of length at least n, w can be divided into three pieces, $w = xyz$ such that

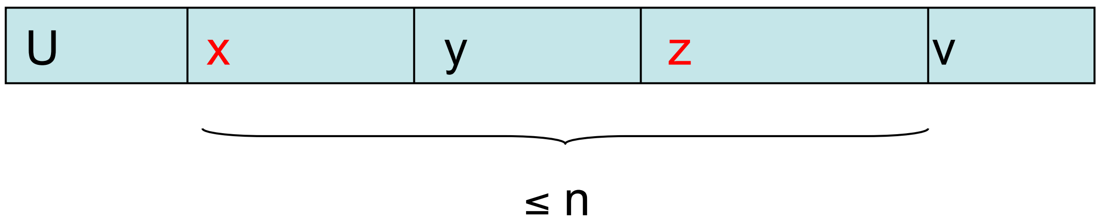
- $|y| > 0$,
- $|xy| \leq n$, and
- for all $i \geq 0$, xy^iz is in L.

$$\underbrace{\boxed{\;x\;}\boxed{\;y\;}}_{\leq n}\boxed{\;z\;}$$

Pumping Lemma for context free languages (cfl)

Theorem. Let L be a cfl. Then there exists a constant n such that for all $w \in L$ and $|w| \geq n$, w can be divided into five pieces $uxyzv$ such that

- $|xz| \geq 1$,
- $|xyz| \leq n$, and
- $ux^iyz^iv \in L$ for all $i \geq 0$.



Application of pumping lemma

Lemma $L = \{a^i b^i c^i \mid i \geq 0\}$ is not a cfl.

Proof Suppose, for the sake of contradiction, that L is a cfl.

Let n be constant generated by the Pumping Lemma.

Consider $w = a^n b^n c^n$

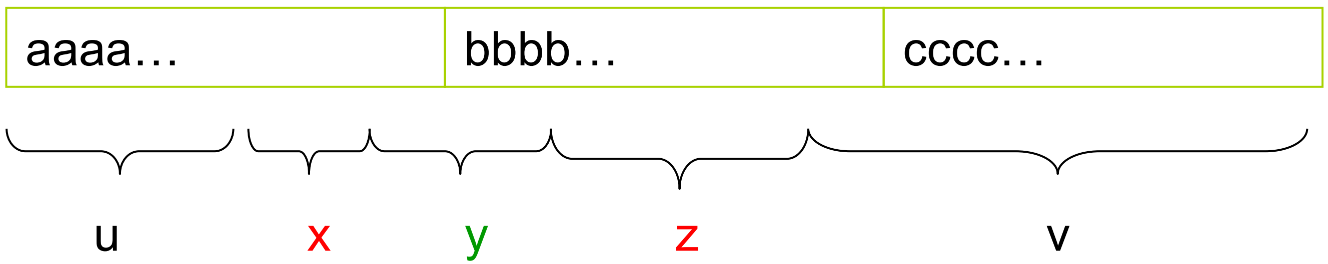
By Pumping Lemma, $w = uxyzv$ with $|xz| \geq 1$, $|xyz| \leq n$, $ux^i yz^i v \in L$ for all $i \geq 0$.

aaaa… (a^n)	bbbb… (b^n)	cccc… (c^n)

xyz

Since $|xyz| \leq n$, xyz can't contain both a's and c's.

Suppose xyz contain no c's (the other case of containing no a's is symmetric).



By Pumping Lemma, uyv is in L.

Since $|xz| \geq 1$, the number of a's and b's in uyv $< 2n$

Therefore, uyv does not have the same number of a's, b's, and c's.

A contradiction occurs. Thus, L is not a cfl.

Proof of Pumping Lemma

Let L be a cfl, and let $G = (V, \Sigma, R, S)$ be a context free grammar such that $L(G) = L$.

Without loss of generality, we can assume that every rule in R is in the form*

- $A \rightarrow BC$; or
- $A \rightarrow a$

Let $|V|$ be the number of non-terminals, and let $n = 2^{|V|}$.

- Consider any $w \in L$ with length at least n.

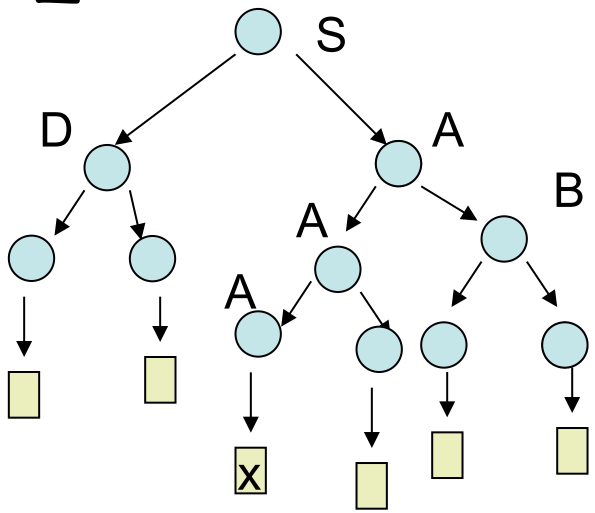
* Such a grammar is said to be in Chomsky Normal Form. For example, if there is a rule $A \rightarrow BdCF$, we can convert it to 4 rules:

$$A \rightarrow BA_1,\ A_1 \rightarrow A_2 A_3,\ A_2 \rightarrow d,\ A_3 \rightarrow CF.$$

Parse tree of w

- Let T be a parse tree of w.
- S is the root. Every internal node is a variable in V, and each leaf is a terminal in ∑.

$S \to DA$
$A \to AB$
$A \to x$

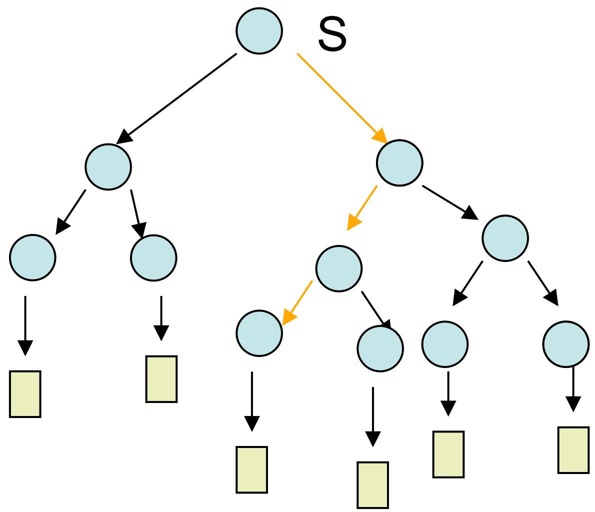


Recall that $|w| \geq n = 2^{|V|}$.

- T has exactly |w| leaves.
- T is **binary**, i.e, every internal node has <u>either</u> two children labeled with variables (non-terminals), <u>or</u> one child labeled with a terminal.

Height of T

- Define **Height**(T) = the number of internal nodes on a longest path from the root to a leaf.

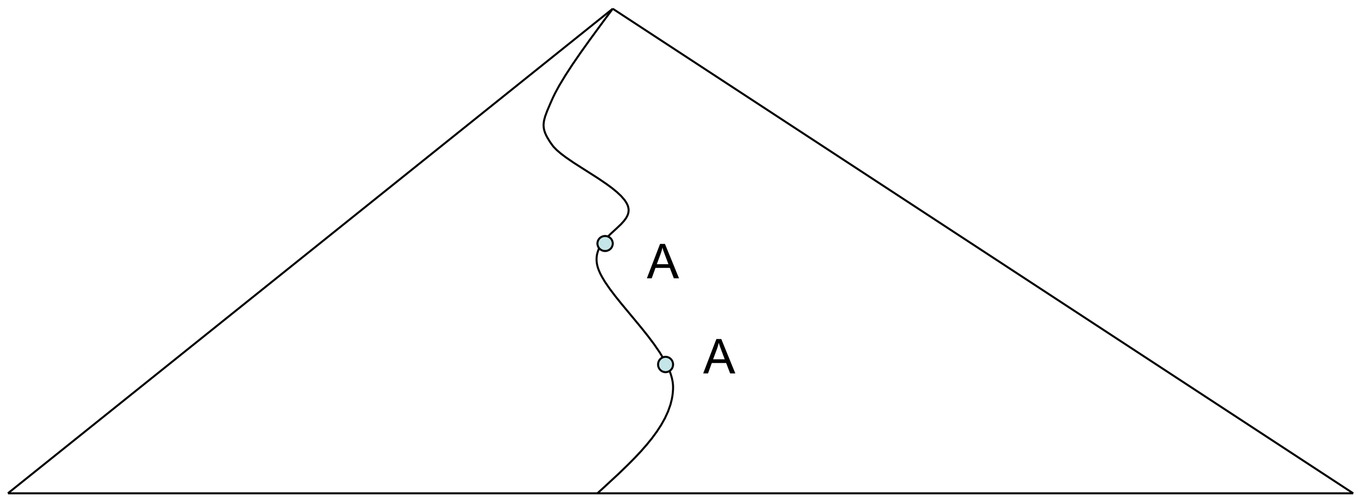


Recall that $|w| \geq n = 2^{|V|}$.

- In the above example, **Height**(T) = 4; Number of leaves $= 6 \leq 2^{4-1} = 8$;
- Number of leaves in T = $|w| \leq 2^{\text{Height}(T) - 1}$;

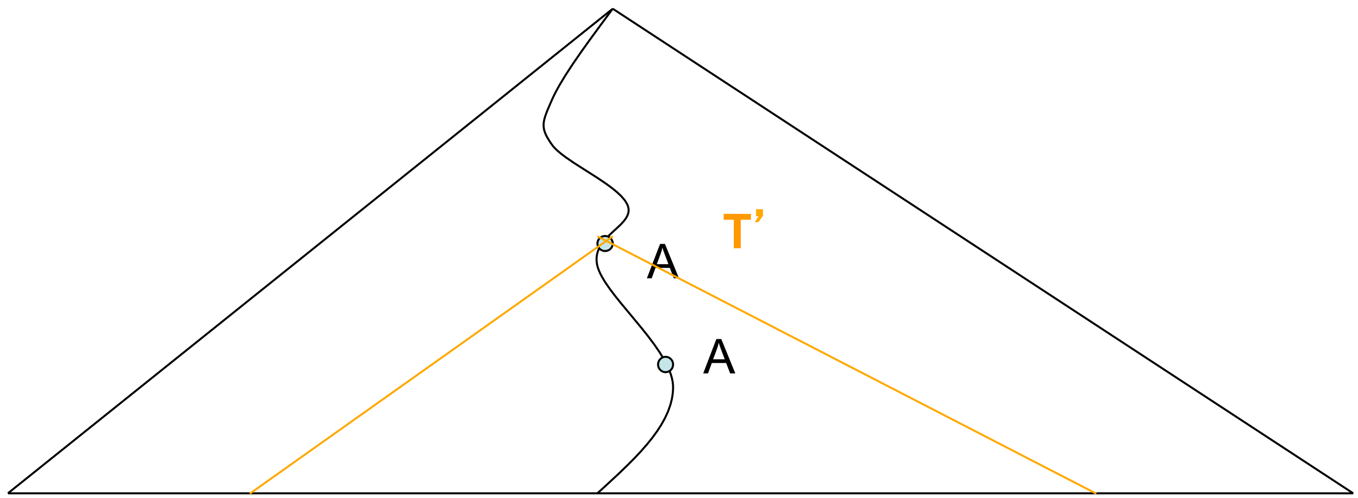
  thus, $\text{Height}(T) \geq \log_2 |w| + 1 \geq \log_2 n + 1 = |V| + 1$.

- Consider the longest path in T, which contains at least |V| +1 internal nodes.

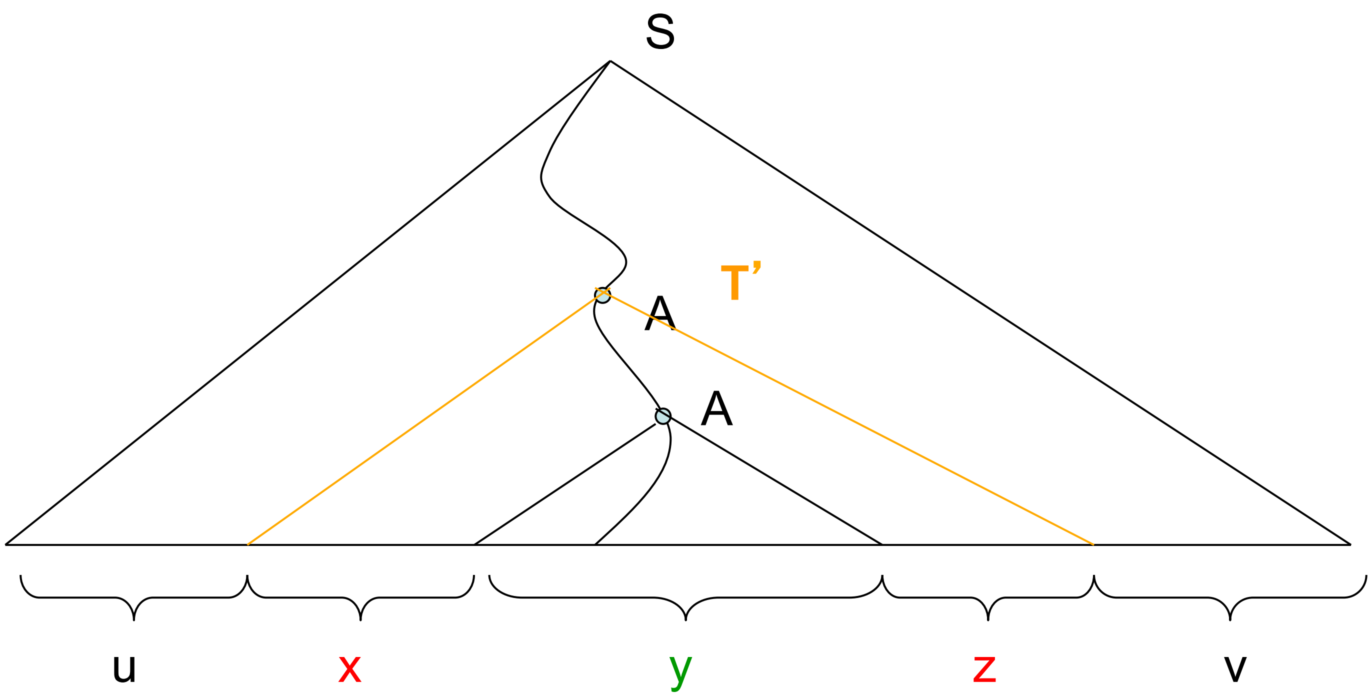


- There are two internal nodes labeled the same variable. Let A be the ***first*** duplicate variable encountered when we walk up the path.

- Consider the longest path in T, which contains at least |V| +1 internal nodes.



- There are two internal nodes labeled the same variable. Let *A* be the ***first*** duplicate variable encountered when we walk up the path.
  - Let **T"** be the subtree rooted at the lowest occurrence of *A*;
  - let **T'** be the subtree rooted at 2nd lowest occurrence of *A*.
- Height(**T'**) $\le$ |V|+1. Why ?



- $A \overset{*}{\Rightarrow} x\,A\,z$; and $A \overset{*}{\Rightarrow} y$
- $S \overset{*}{\Rightarrow} u\,A\,v \overset{*}{\Rightarrow} ux\,A\,zv \overset{*}{\Rightarrow} uxyzv$
- By induction on i, $S \Rightarrow ux^i yz^i v$

- x and z can't be both null strings. Why ?
- Height(T') ≤ |V|+1. Thus, T's has ≤ $2^{|V|}$ leaves. Thus $|xyz| \le 2^{|V|} = n$.

Context free languages are not closed under complementation

- Define $\overline{L} = \{ x \in \Sigma^* \mid x \notin L\}$.
- If L is regular, then L can be accepted by a dfa, and so does L. I.e., L is also regular. (dfa-based argument)
- Question: If L is a cfl, is L is a cfl?

No.

Counter example: $L = \{ a^i b^j c^k \mid i \neq j \text{ or } j \neq k\}$.

- L is context free.
- But $\overline{L} = \{a^i b^i c^i \mid i \geq 0\}$ is not context free.

Union, intersection

- Question: If L_1 and L_2 are cfl, is $L = L_1 \cup L_2 = \{w \mid w$ in L_1 or w in $L_2\}$ a cfl?
  - Yes (proof); No (counter example)

- Question: If L_1 and L_2 are cfl, is $L = L_1 \cap L_2 = \{w \mid w$ in L_1 and w in $L_2\}$ a cfl?
  - Yes (proof); No (counter example)

Deterministic pushdown automata

A dpda is a pda such that for any state q, input symbol a, and stack symbol s,

- $| f(q, a, s) | \leq 1$; and
- if $f(q, \varepsilon, s) \neq \varnothing$, then $f(q, a, s) = \varnothing$.

That is, at any time, the next move of a dpda is uniquely defined.

A language L is said to be a deterministic context free language (dcfl) if L can be accepted by a deterministic dpda.

dpda versus pda

Recall that dfa and nfa have the same power.

Yet, dpda are not as powerful as pda. There exist languages L such that L can be accepted by some pda but not by any dpda.

Example:

- $\{a^n b^n \mid n \geq 1\} \cup \{a^n b^{2n} \mid n \geq 1\}$ is a cfl but not dcfl.
- $\{ ww^T \mid w \text{ in } \{0,1\}^*\}$

In fact, dpda is closed under complementation.