

CSCE 355
3/20/2023

CFG \rightarrow PDA

①

Recall: CFG $G = \langle \underbrace{V, \Sigma}_{\text{grammar symbols}}, S, P \rangle$

One state PDA

$$P = \langle \{q\}, \Sigma', V \cup \Sigma', \delta, \{q\}, S, \emptyset \rangle$$

so that $N(P) = L(G)$. For every $a \in \Sigma'$,

$$\delta(q, a, a) = \{(q, \epsilon)\} \quad \text{"matching a"}$$

and for every $A \in V$,

$$\delta(q, \epsilon, A) = \{(q, \alpha) : A \rightarrow \alpha \text{ is a production in } P\}$$

Ex: $E \rightarrow E + T \mid T$
 $T \rightarrow T * F \mid F$
 $F \rightarrow c \mid ('E')$

no other transitions allowed

Input $c*(c+c)$

$$(q, "c*(c+c)", \underline{E}) \vdash (q, "c*(c+c)", T) \vdash (q, "c*(c+c)", T * F)$$

$$\vdash(q, "c*(c+c)", F*F) \vdash(q, "c*(c+c)", ~~c~~ c*F) \quad (2)$$

$$\vdash(q, "*(c+c)", *F) \vdash(q, (c+c), F)$$

$$\vdash(q, "(c+c)", "(E)') \vdash(q, "(c+c)", E')')$$

$$\vdash(q, "c+c)", E+T')')$$

$$\vdash(q, "c+c)", T+T')')$$

$$\vdash(q, "c+c)", F+T')')$$

$$\vdash(q, "c+c)", c+T')')$$

$$\vdash(q, "+c)", +T')') \vdash(q, ~~"c+c)"~~ "c)", T')')$$

$$\vdash(q, "c)", F')') \vdash(q, "c)", c')')$$

$$\vdash(q, ')', ')') \vdash(q, \epsilon, \epsilon) \text{ accept}$$

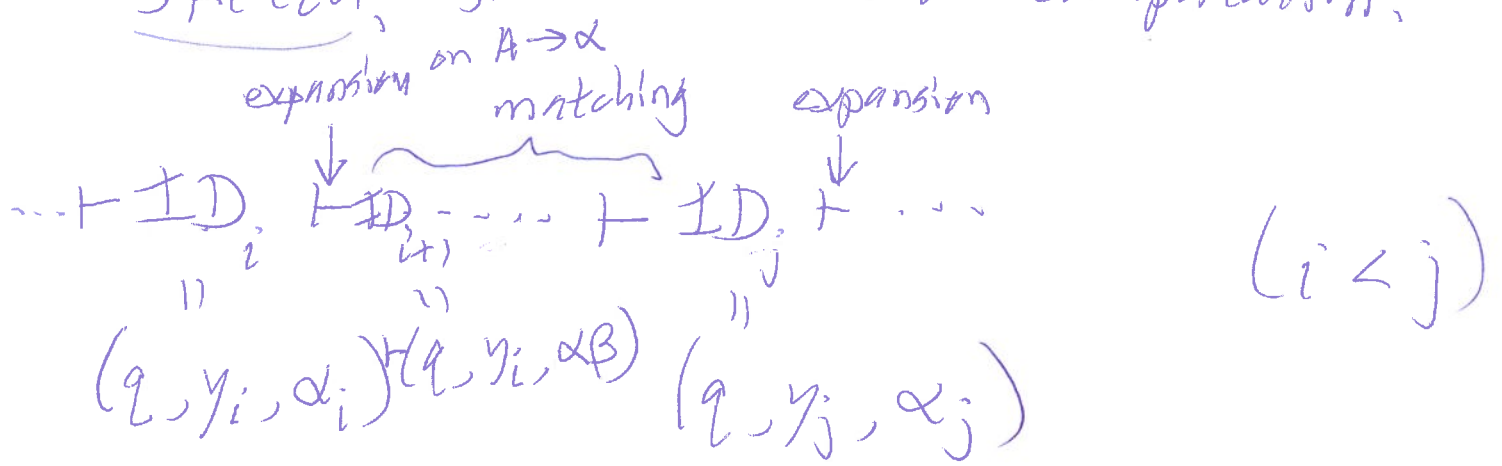
observe { The expansion steps (i.e., the non-matching steps) use the same sequence of productions used in a leftmost derivation of the input string.

Proof of correctness (in general, arbitrary CFG G): (3)

Part 1: If P accepts a string $w \in \Sigma^*$ via empty stack, then there exists a leftmost derivation of w in G .

Proof: (Formally by induction on the length of the sequence of ~~the~~ ID's in the accepting computation of P on input w).

Sketch: somewhere in the computation:



$\alpha_i = A\beta$ (some β)

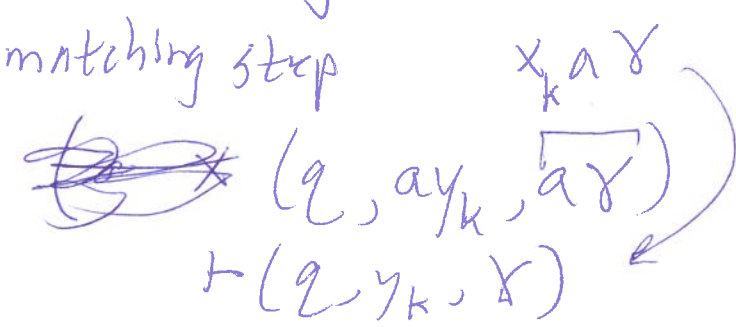
let x_i be the input consumed already. So:

$w = x_i y_i$

Note: $x_i A\beta \Rightarrow x_i \alpha\beta$

$= x_j \alpha_j$

matching step



$$(q, \underbrace{ay, a} \gamma) + (q, \gamma, r)$$

~~x~~
x consumed
so far

↑
xa consumed
so far

(concatenate prev consumed symbols
with stack contents)

does not change in a matching step.

So "nothing happens to the concatenation
~~during~~ matching steps,

but an expansion step corresponds to
a single step in a leftmost derivation.

Part 2: If $w \in L(G)$, then P accepts
w via empty stack:

Sketch: Given $S \Rightarrow \alpha_1 \Rightarrow \dots \Rightarrow \alpha_k = w$
leftmost derivation, then $\alpha_i \Rightarrow \alpha_{i+1}$

~~α~~ corresponds to an expansion step of P
preceded by matching steps

Ex: $(q, \gamma, \dots A \dots) \xrightarrow{\text{matching}} (q, \gamma', A \dots)$

$\vdash_{\text{expand}} (q, y', \alpha \dots)$

(A $\rightarrow \alpha$ production) ⁽⁵⁾

Next up: PDA \Rightarrow CFG

Convert an arbitrary PDA P into a CFG G
such that $L(G) = N(P)$.

Plan:

PDA \Rightarrow ~~resp~~ restricted PDA \Rightarrow CFG

Def: A PDA is restricted if the
only allowed transitions are of the
following 2 forms:

$\delta(q, a, X)$ contains (r, ϵ)

for $q, r \in Q$

$a \in \Sigma \cup \{\epsilon\}$

$X \in \Gamma$

"pop X"

or

"pop"

or

$\delta(q, a, X)$ contains (r, \underline{YX})

q, r, a as above, $X, Y \in \Gamma$

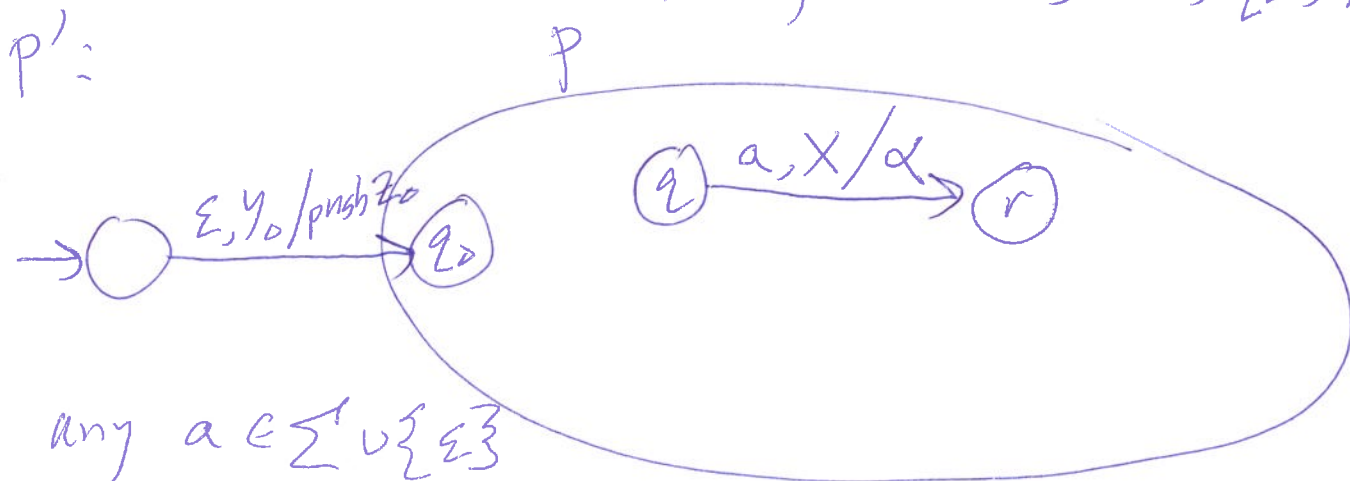
"push Y"

Lemma: For every PDA P there exists a restricted PDA P' such that $N(P') \subseteq N(P)$.

Proof: By construction. Let

$$P = \langle Q, \Sigma, \Gamma, \delta, q_0, Z_0, - \rangle$$

$$P' = \langle Q \cup \{\text{more states}\}, \Sigma, \Gamma \cup \{Y_0\}, \delta', q'_0, Y_0, \emptyset \rangle$$



any $a \in \Sigma \cup \{\epsilon\}$

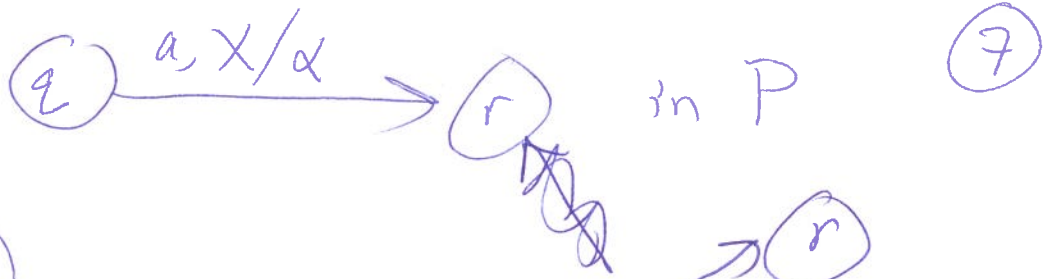
$X \in \Gamma$

$\alpha \in \Gamma^*$

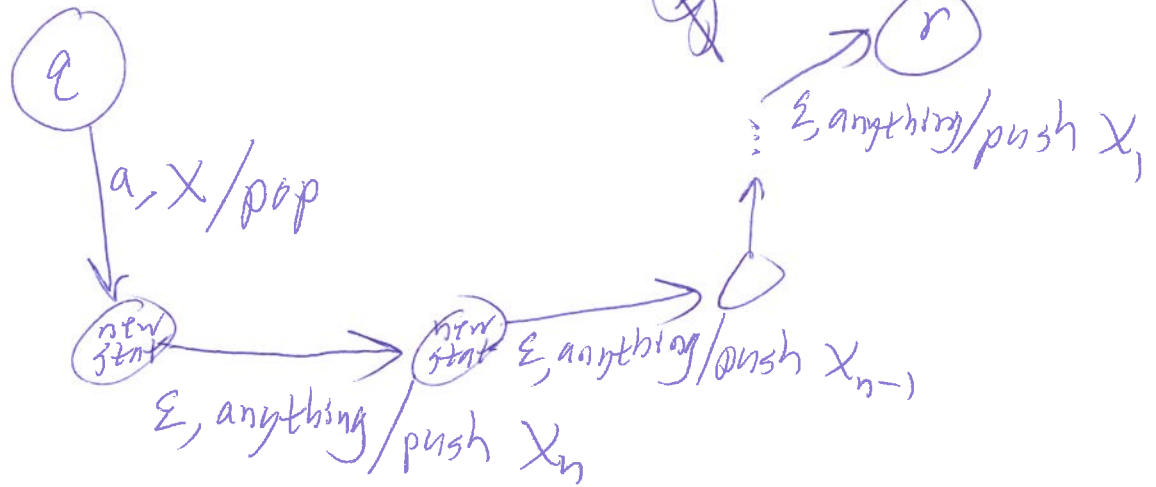
Let $\alpha = X_1 X_2 \dots X_n$

for some $n \geq 0$

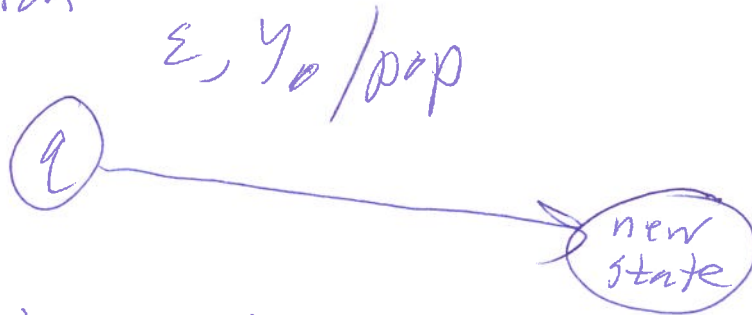
replace



with



Finally, for every state q of P ,
add a transition



$N(P') = N(P)$ by construction.

What remains: convert a restricted PDA into an equivalent CFG.

CSCE 355
3/27/2023

Today: PDA \Rightarrow CFG

①

$P \mapsto G$

Last time: PDA \Rightarrow restricted PDA

A restricted PDA is a PDA that allows only two types of actions

~~$\delta(q, a)$ contains~~

$\delta(q, a, x)$ contains (r, ε)
 (r, pop)

or

$\delta(q, a, x)$ contains $(r, \gamma x)$
 $(r, \text{push } \gamma)$

Today: restricted PDA \Rightarrow CFG

Given a restricted PDA $P = \langle Q, \Sigma, \Gamma, \delta, q_0, z_0 \rangle$
Construct a CFG G such that $L(G) = N(P)$.

$$G = \langle V, \Sigma, S, P \rangle$$

(2)

V contains symbols of the form,

$$\underline{[qXr]}$$

for all states $q, r \in Q$

and ~~the~~ stack symbols $X \in \Gamma$,

as well as S .

Want productions so that

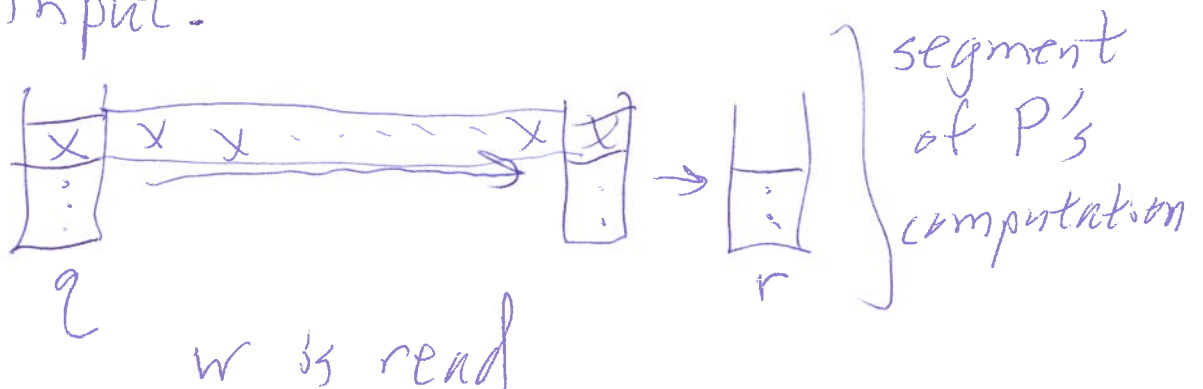
$$\underline{[qXr]} \Rightarrow^* w \quad (w \in \Sigma^*)$$

iff P , starting with X on top of stack,

and in state q can reach state r

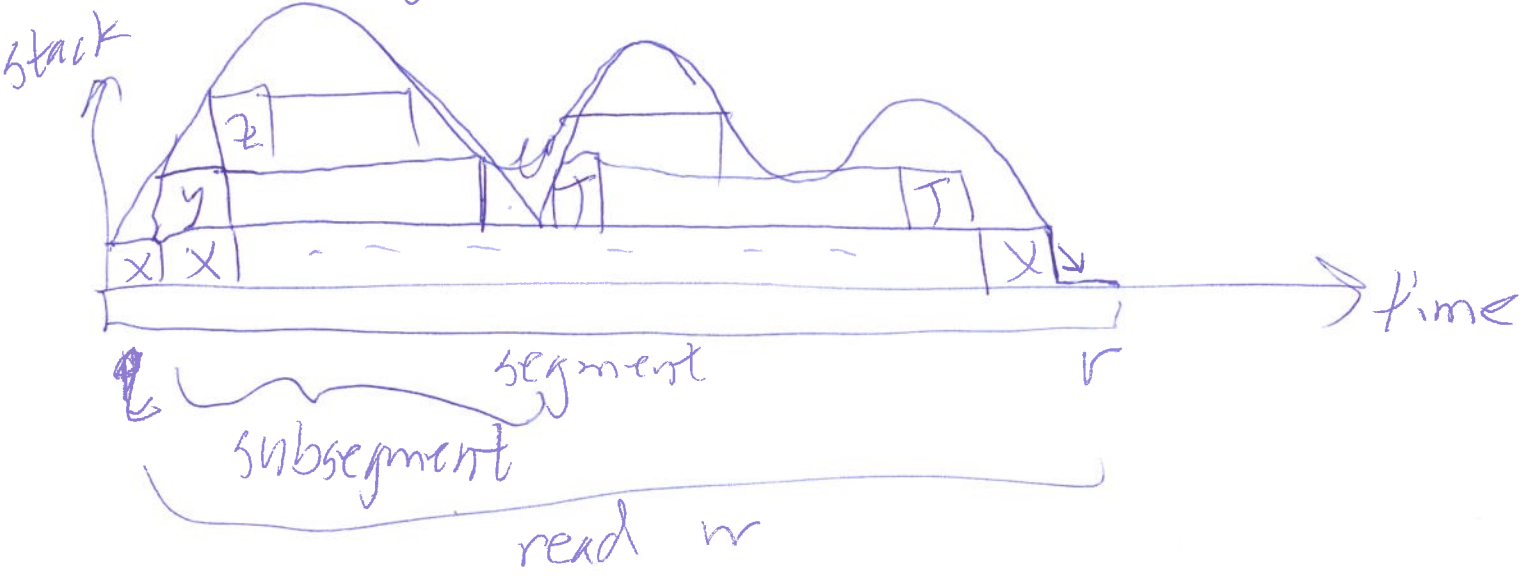
and X stays on top of the stack until

~~the~~ it gets popped, and w is read from the input.



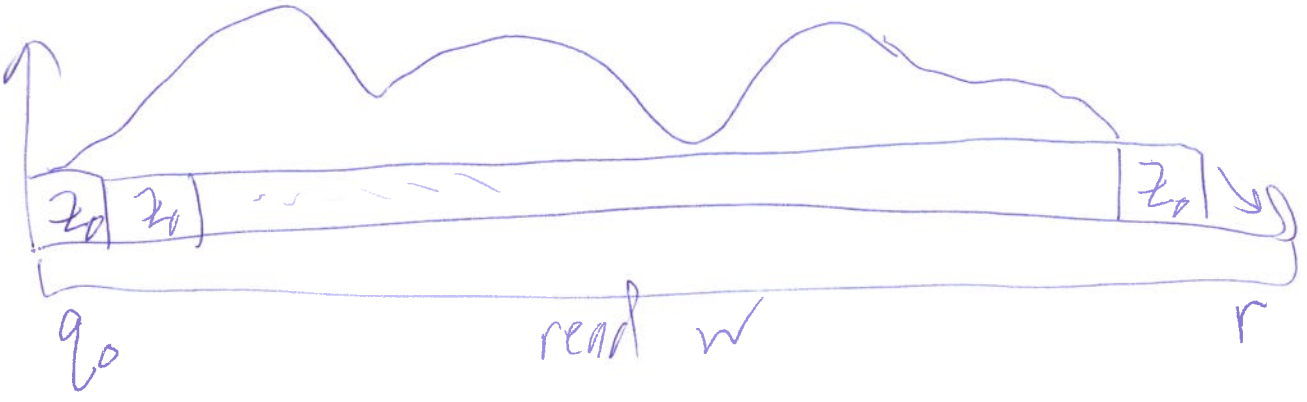
3

A segment of P's comp starts with some X on top of the stack, and in some state q and ends in some state r, where X stays on the stack until the last step, when it gets popped



want $[q, X, r] \Rightarrow^* w$

Whole computation (assuming w is accepted)



S-productions (S is the start symbol) ④

For every state $r \in Q$, add the production

$$S \rightarrow [q_0 z_0 r]$$

Poping productions: For any $a \in \Sigma \cup \{\epsilon\}$ and state $q \in Q$ and stack symbol $x \in \Gamma$,

if $\delta(q, a, x)$ contains (r, pop)

(for any $r \in Q$), add the production

$$[q x r] \rightarrow a \quad (\text{shortest segment})$$

Pushing productions: suppose

$\delta(q, a, x)$ contains $(\overset{S}{\cancel{S}}, \text{push } y)$

(some state $s \in Q$ and $y \in \Gamma$)



Then for every state $t \in Q$ add
the production

(5)

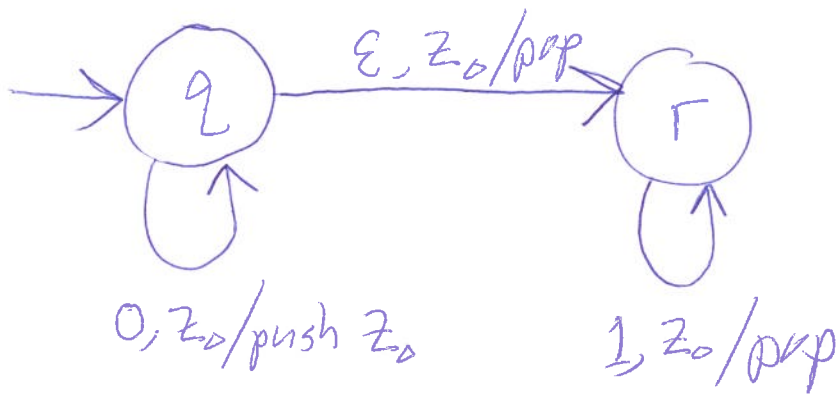
$$[qXr] \rightarrow a[syt][tXr]$$

No other productions.

~~End~~ Completes the construction of G .

EX: $L = \{0^n 1^n : n \geq 0\}$

$\Gamma = \{z_0\}$



$G:$

$$V = \{S, [qz_0q], [qz_0r], [rz_0q], [rz_0r]\}$$

S-productions

$$S \rightarrow [qz_0q] \mid [qz_0r]$$

Popping productions:

$$\begin{aligned} [qz_r] &\rightarrow \epsilon \\ [rz_r] &\rightarrow 1 \end{aligned}$$

Push productions

$$\begin{aligned} [qz_0q] &\rightarrow 0 [qz_0q] [qz_0q] \\ &| 0 [qz_0r] [rz_0q] \end{aligned}$$

- A := [qz_0q]
- B := [qz_r]
- C := [rz_0q]
- D := [rz_r]

$S \rightarrow A \mid B$
 $B \rightarrow \epsilon \mid 0AB \mid 0BD$
 $A \rightarrow 0AA \mid 0BC$
 $D \rightarrow 1$
~~⊥~~

bypass

$$\begin{aligned} [qz_0r] &\rightarrow 0 [qz_0q] [qz_0r] \\ &| 0 [qz_0r] [rz_0r] \end{aligned}$$

$$S \rightarrow A \mid B$$

$$B \rightarrow \epsilon \mid OAB \mid OB1$$

$$A \rightarrow OAA \mid \underline{OBC}$$

useless



$$S \rightarrow \overline{A} \mid B$$

useless

$$B \rightarrow \epsilon \mid \overline{OAB} \mid OB1$$

useless

$$A \rightarrow OAA \} \text{useless}$$



$$S \rightarrow B$$

$$B \rightarrow \epsilon \mid OB1$$

⇓ make B the start symbol:

$$B \rightarrow OB1 \mid \epsilon$$

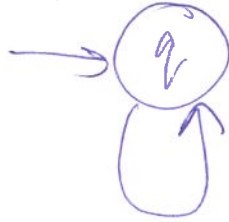
Proof of correctness omitted.

Proof idea: $N(P) \subseteq L(G)$ by induction on the length of an ~~of a derivat~~ accepting path of P
 $L(G) \subseteq N(P)$ by induction on the length of a derivation of G

Ex:

Properly nested parentheses

(8)



0 = (

1 =)

0, z₀ / push +

0, + / push +

1, + / pop

→ ε, z₀ / pop

S → [qz₀q]

[qz₀q] → ε

[q+q] → 1

A := [qz₀q]

[qz₀q] → 0 [q+q] [qz₀q]

B := [q+q]

[q+q] → 0 [q+q] [q+q]

~~S → A~~

A → ε | 0 B A

B → 1 | 0 B B

Applications: $L := \{a^i b^i c^i : i \geq 0\}$ ①

is not CFL-pumpable (hence not a CFL by the Lemma)

What "not CFL-pumpable" means:

$\forall p > 0$

$\exists s \in L, |s| \geq p$

$\forall u, v, \overline{x}, \overline{y}, z$ such that

$s = uv\overline{wxy}$

$\rightarrow |v\overline{wx}| \leq p$ *

$\rightarrow |vx| > 0$

$\exists i, uv^i \overline{w} x^i \overline{y} \notin L.$

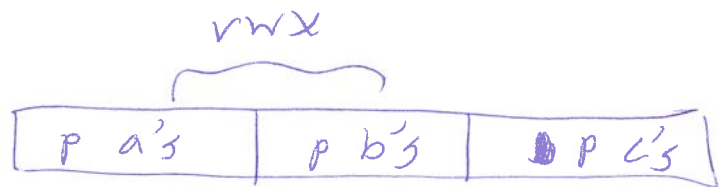
L is not CFL-pumpable:

Given $p > 0$, $s = a^p b^p c^p$. ($s \in L$ & $|s| = 3p \geq p$)

Given u, v, \overline{w}, x, y as above,

let $i := 0$.

This works. Why?



Since $|v\overline{wx}| \leq p$, $v\overline{wx}$ cannot contain both some a's and some c's. So

$u\overline{w}y = uv^0 \overline{w} x^0 y$ has either the same

CSCE 355
3/29/2023

Prog project handout:

(2)

1) ϵ -NFA \Rightarrow NFA

2) simulate an NFA on input strings

Pumping Lemma for CFLs

~~Used~~ Used to show a lang is not a CFL.

Lemma (Pumping Lemma for CFLs):

Every context-free language is CFL-pumpable.

Def: A lang. L is CFL-pumpable if

$\exists p > 0$, ("pumping length")

$\forall s \in L$ such that $|s| \geq p$,

$\exists u, v, w, x, y$ strings such that

1) $s = uvwx^i y$

2) $|vwx| \leq p$

3) $|v| + |x| > 0$ (i.e., v, x cannot both be ϵ)

and

$\forall i \geq 0$, $uv^iwx^i y \in L$.

Proof (later).

(3)

number of c & fewer a's or b's,
or the same # of a's and fewer b's
or fewer c's (or both)

$\therefore uv^i y \notin L \quad //$

Similarly:

$$L := \{ a^i b^j c^i d^j : i, j \geq 0 \}$$

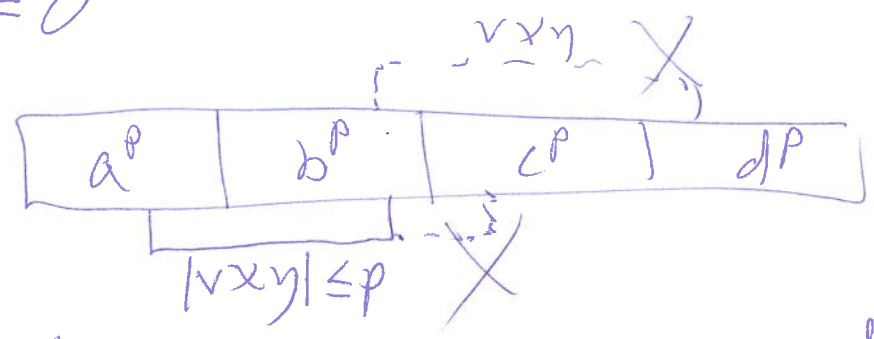
is not CFL-pumpable.

$$\forall p > 0, \text{ let } s := a^p b^p c^p d^p$$

Given u, v, x, y, z s.t. $s = uvxy^i z$,

$$|vxy| \leq p, |vy| > 0,$$

Let $i := 0$



vxy ~~can't~~ can't have both a's & c's, and
can't have both b's & d's. (too far apart).

So any $i \neq 1$ will work, adjust # of one
of a, c's with out the other, or one of b's &
d's & not the other, leaving \neq numbers of
a's & c's or b's and d's.

So $uv^i xy^i z \notin L$ for any $i \neq 1$.

```
int f(a, b x, y, z) {
    ...
}
```

```
int g(a, b) {
    ...
}
```

```
f(2, 3, 4)
:
g(5, 6)
```

OK b/c
 # actual args
 for f, g
 matches the number
 of formal params
 for each function

A CFG alone cannot check this — not a context-free part of the language (C++)
 Need the symbol table — a global data struct that remembers this info (# formal params)

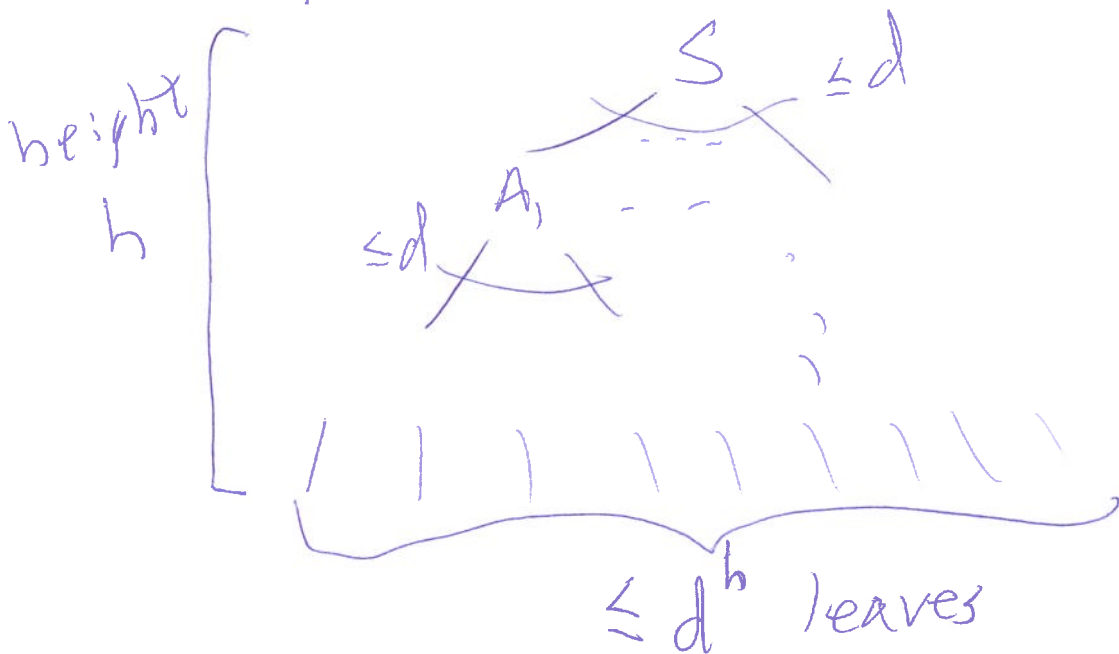
Proof (of the Pumping Lemma for CFLs) ⁽⁵⁾

[Every CFL is CFL-pumpable]

Let L be any CFL. Fix a CFG G such that $L = L(G)$.

~~Let G~~ Let (n) be the number of nonterminals of G , and let (d) be the max length of any body of a production of G .

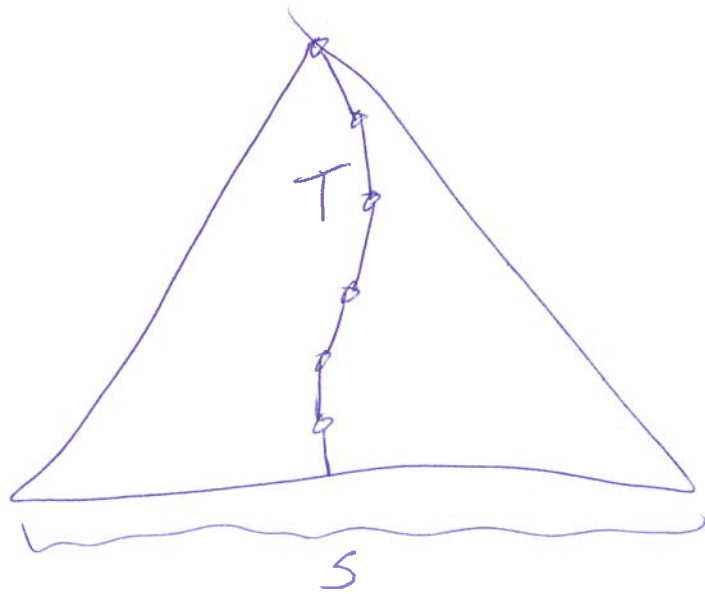
Any ~~complete~~ parse tree of G has branching $\leq d$



Let $p := d^{n+1}$

Let sel be any string of length $\geq p$.

Since $s \in L$, there is a parse tree (6) yielding s . Let T be a min-size (min # of nodes) parse tree of G yielding s .



$|s| \geq p = d^{n+1}$. What is the height of T ?

Letting h be the height of T ,

T has $\leq d^h$ many leaves, so

~~#~~ # leaves = $|s| \geq d^{n+1}$, so

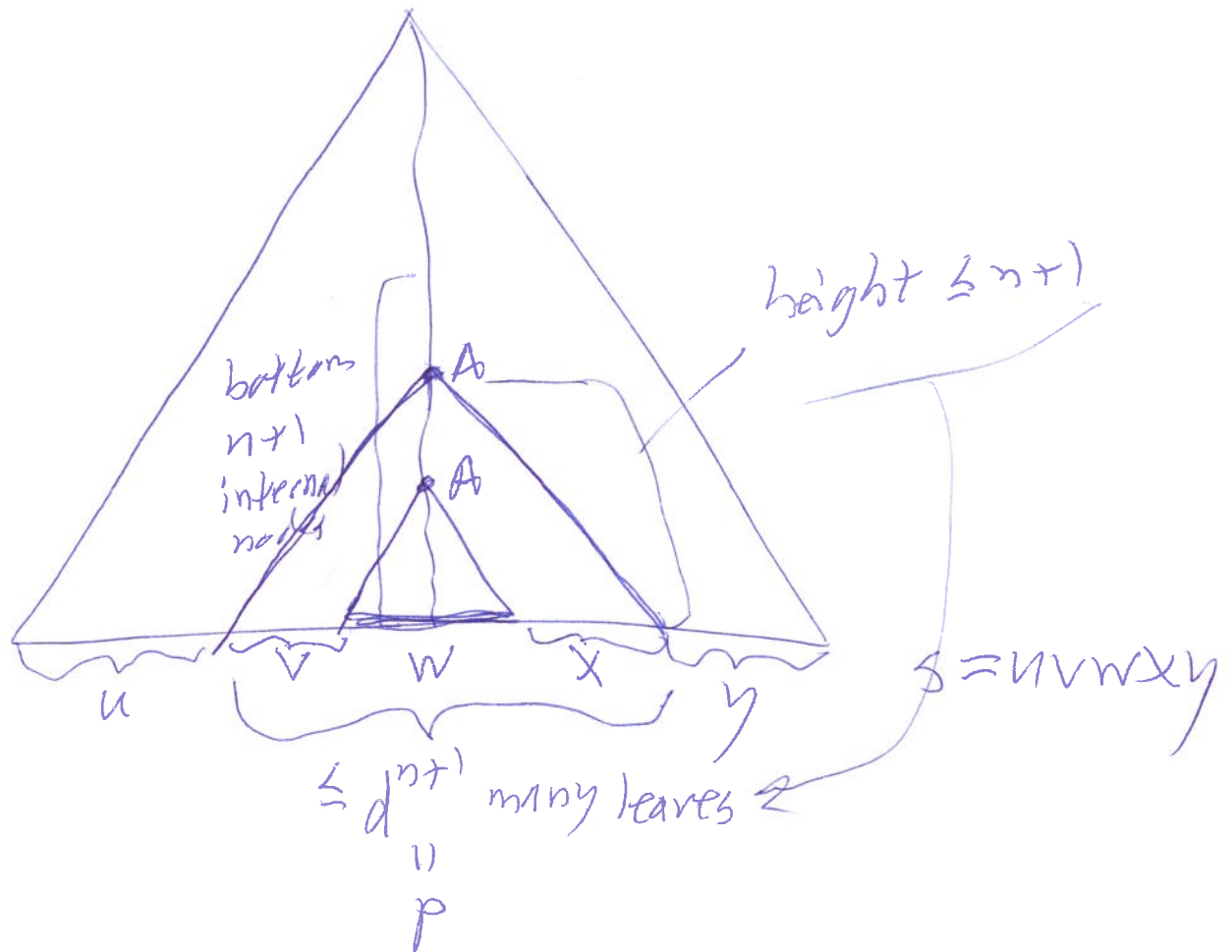
$$d^h \geq d^{n+1}$$

$$\therefore h \geq n+1$$

So T has height $\geq n+1$

\therefore there is a path in T with $\geq n+1$ many internal nodes, each labeled by a non-terminal. But only n nonterminals

i. (pigeonhole principle) there is some (7)
 nonterminal (A , say) that is repeated
 (occurs \geq twice) among the bottom $n+1$
 internal nodes along the path.



$$\therefore |vwx| \leq p$$

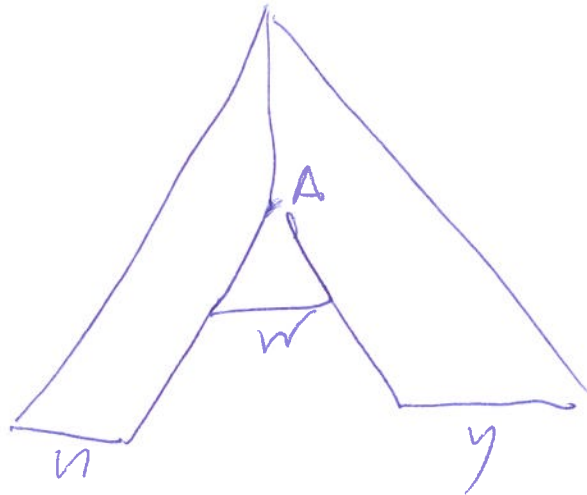
What's left: show that $|vx| > 0$

and $uv^iwx^iy \in L$ for all $i \geq 0$.

Show the 2nd one first.

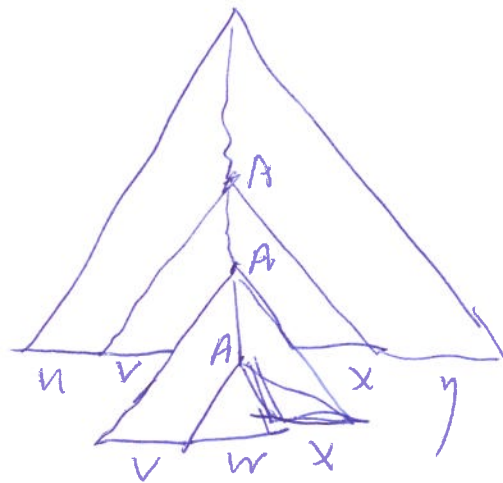
$i := 0$: Parse tree for $uv^0wx^0y = uwy$:
 delete the "wedge" and merge the
 lower A with the upper A;

(8)



$i := 2$. Duplicate the wedge:

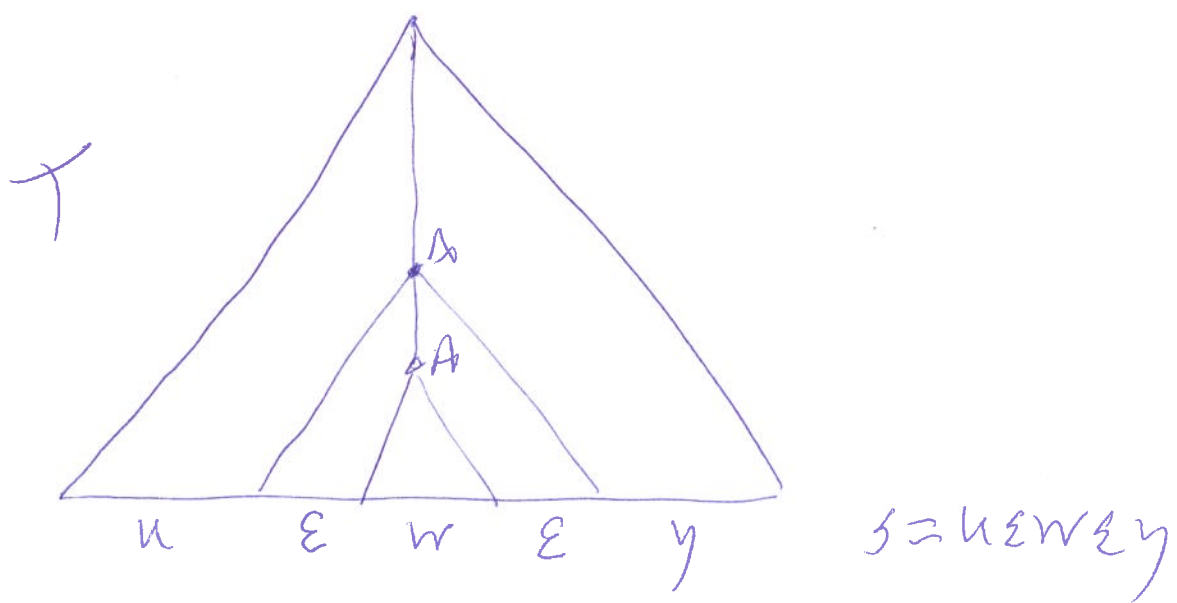
parse tree
 for
 uv^2wx^2y



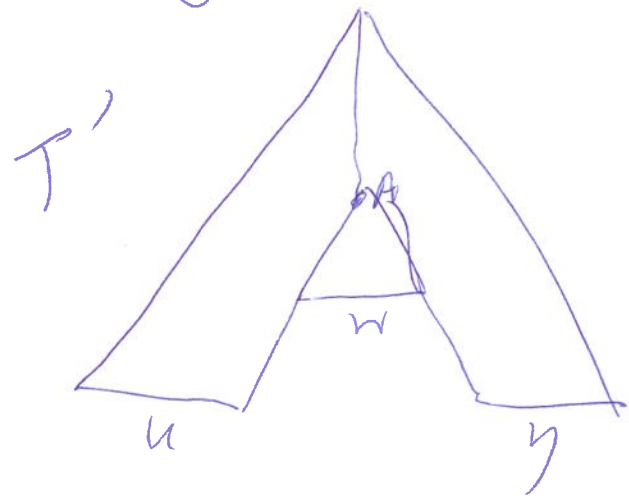
For any i , have i copies of the wedge,
 one on top of the other to get a parse
 tree for uv^iwx^iy

$\therefore uv^iwx^iy \in L$ for all i .

Last thing: show that v and x can't both be ϵ . Suppose otherwise. (9)



Remove the wedge & merge the two A's: get another parse tree T' also yielding s :



but T' is smaller than T . But we picked T to be min size yielding s . \downarrow so $|vx| > 0$. □

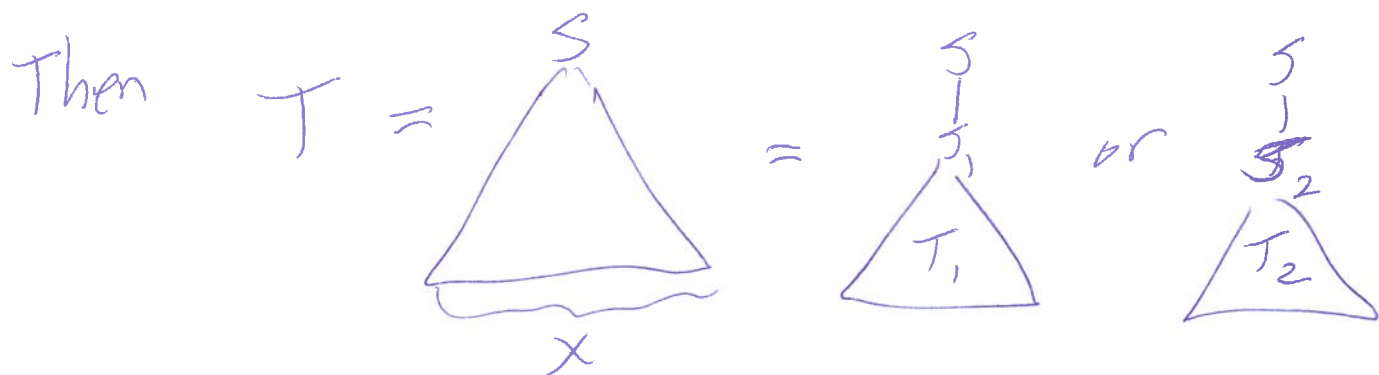
Therefore, $w \in L(G)$.

Similarly, any string in L_2 is also in G .

Thus $L_1 \cup L_2 \subseteq L(G)$

want $=$

Let x be any string in $L(G)$, and let T be a complete parse tree yielding x of G



Then T_1 (or T_2 , whichever) is a complete parse tree of either G_1 or G_2 yielding x .

$\therefore x \in L(G_1)$ or $x \in L(G_2)$

$\therefore L(G) \subseteq L_1 \cup L_2$

$\therefore L(G) = L_1 \cup L_2$ 

Prop: The concatenation of two CFLs is a CFL.

Proof: Given ~~L_1~~ $L_1 = L(G_1)$, $L_2 = L(G_2)$ as in the previous proof. It suffices to find a grammar G such that $L(G) = L_1 L_2$.

CSCE 355
4/3/23

Closure Properties of CFLs

Closure under union, concat, *

Prop: The union of any two CFLs is a CFL.

Proof: Let $L_1 = L(G_1)$ and $L_2 = L(G_2)$

for CFGs $G_1 = \langle V_1, \Sigma, S_1, P_1 \rangle$

and $G_2 = \langle V_2, \Sigma, S_2, P_2 \rangle$

WLOG, $V_1 \cap V_2 = \emptyset$ (by renaming nonterminals).

Let $G := \langle V_1 \cup V_2 \cup \underbrace{\{S\}}_{\substack{\text{new} \\ \text{symbol} \\ \text{not in} \\ V_1 \cup V_2}}, \Sigma, S, P \rangle$

where $P = P_1 \cup P_2 \cup \{ \underbrace{S \rightarrow S_1, S \rightarrow S_2} \}$

Given $w \in L_1$, let T_1 be a complete parse tree of G_1 yielding w . Then

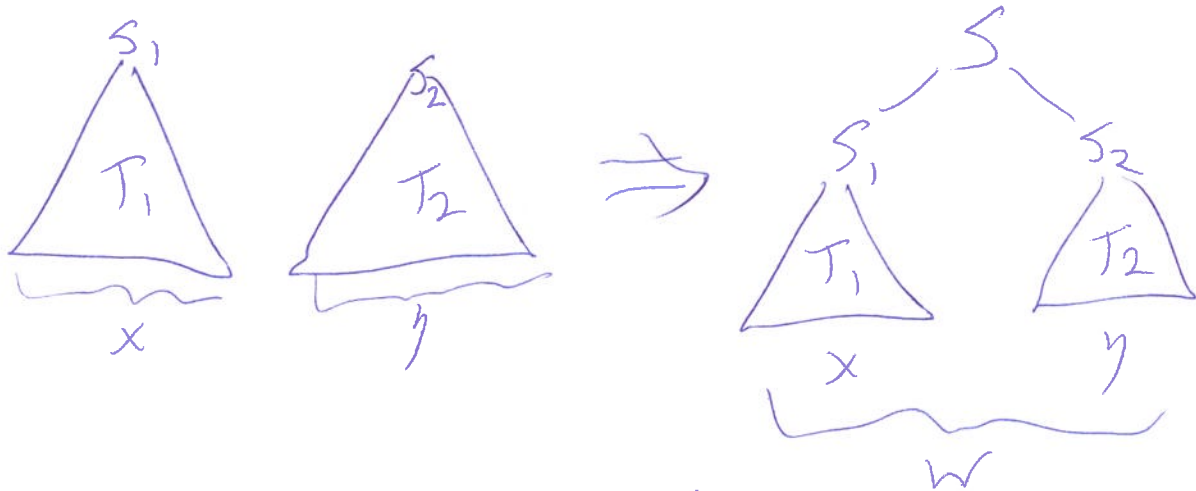


is a complete parse tree in G yielding w .

$$G := \langle V, \cup V_2 \cup \{S\}, \Sigma, S, P \rangle$$

where $P = P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}$.

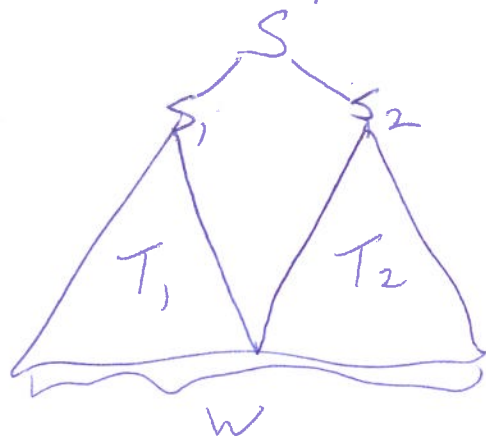
Let w be any string. If $w = xy$ for $x \in L_1, y \in L_2$, then there are parse trees



so \exists parse tree of G yielding $w \quad \therefore w \in L(G)$

$\therefore L_1 L_2 \subseteq L(G)$. Conversely, if $w \in L(G)$,

then there is a parse tree of G yielding w



T_1 is a parse tree of G_1 yielding some string x ;

T_2 is a parse tree of G_2 yielding some string y , and $w = xy$.

$$\therefore w \in L_1 L_2 \quad \therefore L(G) \subseteq L_1 L_2 \quad \therefore L_1 L_2 = L(G) //$$

Kleene closure ($*$ -operator)

Prop. If L is a CFL then L^* is a CFL.

[recall: $L^* = \{\epsilon\} \cup L \cup LL \cup LLL \cup \dots$]

Let $L = L(G_1)$ where

$G_1 = \langle V_1, \Sigma, S_1, P_1 \rangle$ is a CFG.

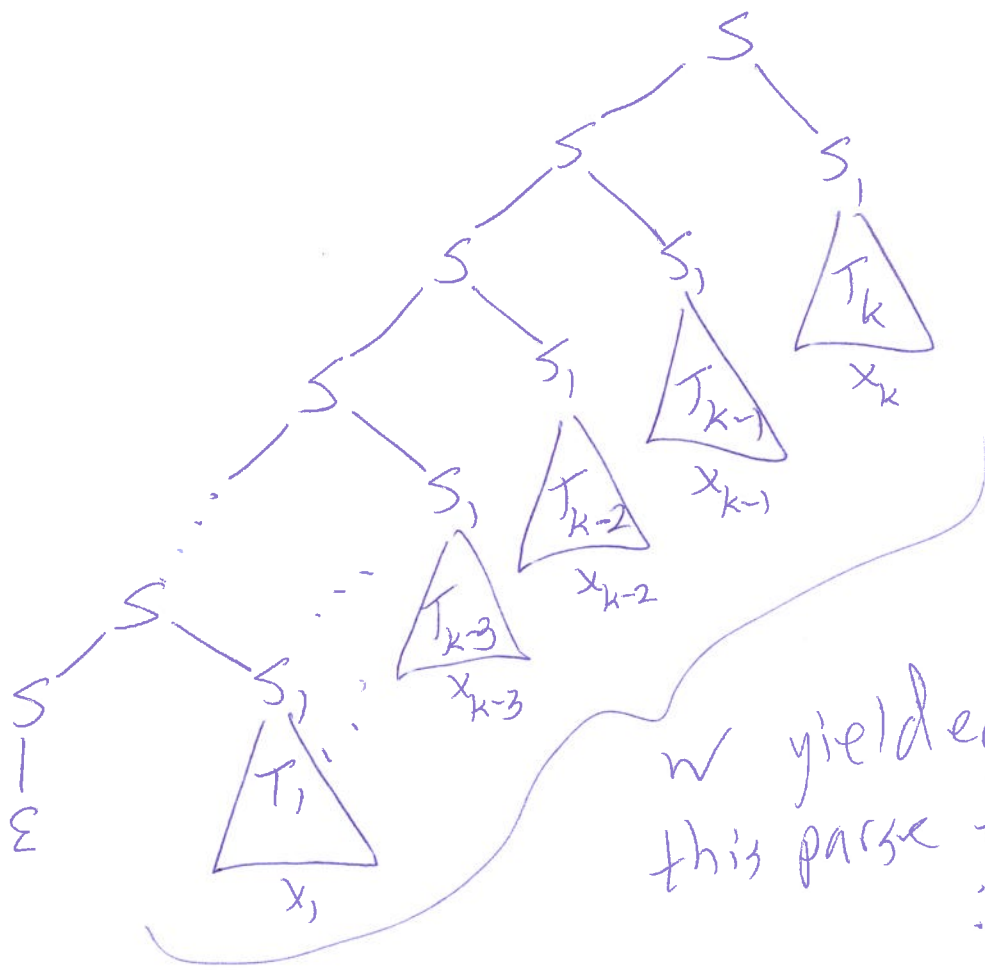
Define $G := \langle V_1 \cup \{S\}, \Sigma, S,$

$P := P_1 \cup \{S \rightarrow SS_1, S \rightarrow \epsilon\}$

Let $w \in L^*$. Show that $w \in L(G)$.

Suppose $w = x_1 x_2 \dots x_k$ for some $k \geq 0$ and
each $x_i \in L$

For $1 \leq i \leq k$, let T_i be a complete parse tree of G_1 yielding x_i . Then the ~~the~~ following complete parse tree of G yields w :



w yielded by
this parse tree of G .
 $\therefore w \in L(G)$.

Conversely, if $w \in L(G)$, then any parse tree of G yielding w must look like the one drawn above, for some k , where the T_i are all parse trees of G . $\therefore w = x_1 \cdots x_k$, where x_i ~~is~~ is the yield of T_i , and so is in L .

$$\therefore w \in L^k \subseteq L^*$$

$$\therefore w \in L^* \iff w \in L(G) \quad (\text{any string } w \in \Sigma^*)$$

$$\therefore L(G) = L^* //$$

Cor: Every regular language is context-free

Pf: Only thing left is to find grammars for the atomic regexes, corresp to languages \emptyset and $\{a\}$ for all $a \in \Sigma$.

$$L(G) = \emptyset \text{ for } G := \langle \{S\}, \Sigma, S, \emptyset \rangle$$

or
 $\{S \rightarrow S\}$

$$\forall a \in \Sigma, L(G_a) = \{a\} \text{ for } G_a := \langle \{S\}, \Sigma, S, \{S \rightarrow a\} \rangle //$$

Alternate proof of the corollary:

Any DFA can be simulate by a PDA (i.e. equivalent) that ignores its stack. //

Prop: If L is a CFL, then L^R is a CFL.

$$\left[\text{recall: } L^R := \{w^R : w \in L\} \right] \xrightarrow{G} \langle V, \Sigma, S, P \rangle$$

Proof: Let G be a CFG s.t. $L = L(G)$.

Want a CFG for L^R :

$$L^R = L(G^R), \text{ where } G^R := \langle V, \Sigma, S,$$

[proof of correctness omitted]

$$\{A \rightarrow \alpha^R : A \rightarrow \alpha \in P\} //$$

Prop: There exist CFLs L_1 and L_2 such that $L_1 \cap L_2$ is not context-free.

Proof: Recall that we showed that the

language $L := \{a^i b^i c^i : i \geq 0\}$

is not a CFL (by the pumping lemma for CFLs). But $L = L_1 \cap L_2$ for CFLs L_1, L_2 where

$$L_1 := \{a^i b^i c^k : i, k \geq 0\}$$

$$L_2 := \{a^k b^i c^i : i, k \geq 0\}$$

$L_1 = L(G_1)$, $L_2 = L(G_2)$, where

$$G_1: S \rightarrow Sc \mid T$$

$$T \rightarrow aTb \mid \epsilon$$

$$G_2: S \rightarrow aS \mid T$$

$$T \rightarrow bTc \mid \epsilon$$

Corollary: ~~There~~ exists a CFL L such that \bar{L} is not a CFL.

Proof: Suppose CFLs are closed under complements. Know (proved) that CFLs under union. But by De Morgan's laws, for any L_1 and L_2

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

~~The CFLs are~~ The right-hand side is a CFL by assumption, for any CFLs L_1, L_2

$\therefore L_1 \cap L_2$ is a CFL for any CFLs L_1 & L_2 , but we just saw a counterexample to this $\Downarrow \therefore$ CFLs not closed under complement.

Ex: Find an explicit CFL L such that \overline{L} is not a CFL.

Ex: Show that $\{ww : w \in \Sigma^{*}\}$ is not pumpable \therefore not a CFL, for $|\Sigma| \geq 2$.

Ex: Show that $\{x : x \text{ is not of the form } ww\}$ is a CFL.

Useful precomputation step:
find the reversals of all the ϵ -moves (back edges)

In steps 1 & 2; given a state q , search for all states reachable from q following back edges only (reverse ϵ -moves).
Can use BFS or DFS for this.

Pass: do this ~~to~~ starting at each state in sequence, for all states.

Repeat passes until nothing changes during a complete pass.

Turing Machines (TMs)

"Turing machine model captures the (informal) notion of computation"

Church-Turing thesis

Next few lectures will convince you of this.

Def. A Turing machine (TM) is a 7-tuple

$$\langle Q, \Sigma, \Gamma, \delta, q_0, B, F \rangle$$
 where

Q is a finite set (elements are states)

Σ, Γ are alphabets:

Σ is the <u>input alphabet</u>	} $\Sigma \subseteq \Gamma$
Γ is the <u>tape alphabet</u>	

$q_0 \in Q$ is the start state

$B \in \Gamma \setminus \Sigma$ is the blank symbol

$F \subseteq Q$ ~~is~~ elements are accepting states



New things a TM can do:

head can move in both directions
(but only one cell at a time)

and can move off the input.

cell contents can be altered by the TM.

In one time step: TM can

- change state
- overwrite the scanned cell
- move head one cell left or right

Finally $\delta : Q \times \Gamma \rightarrow Q \times \Gamma \times \{\leftarrow, \rightarrow\}$ ⁽⁴⁾

is a partial function (not necessarily ~~not~~ defined for all combos of state, tape symbol).

Informally: $q \in Q, a \in \Gamma$

$$\delta(q, a) = (r, b, d) \quad \begin{array}{l} r \in Q \\ b \in \Gamma \\ d \text{ is either} \\ \leftarrow \text{ or } \rightarrow \end{array}$$

means: If current state is q and a is in the currently scanned cell, then in the next time step, the state becomes r , the scanned a is changed to b , and then the head moves one cell to the left or right, depending on d .

Initially, given an input string $w \in \Sigma^*$, the tape has w in contiguous cells, ~~with~~ with the rest of the tape blank (cells contain B). The head scans the leftmost symbol of w , if there is one.

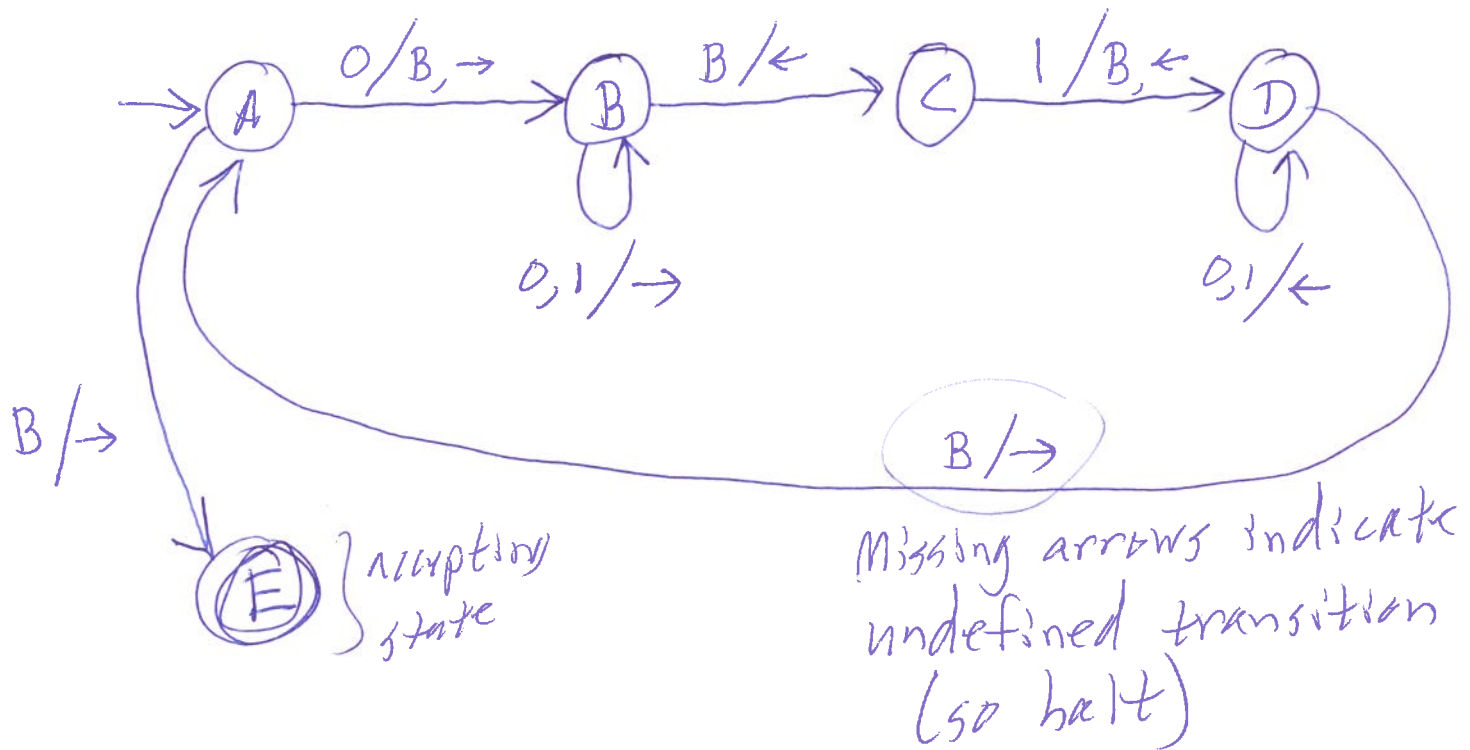
At any time, if the state is q and scanned symbol is a such that $\delta(q, a)$ is undefined then the computation halts (does not proceed).

~~The~~ The TM accepts iff it halts in an accepting state.

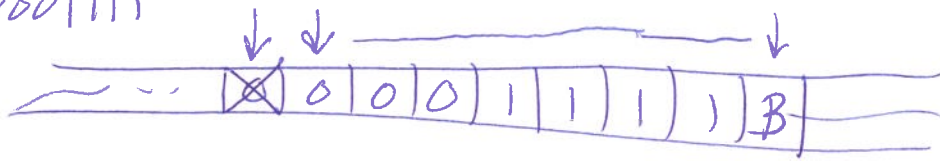
The TM rejects iff it halts in a rejecting state (not in F)

Third possibility: the computation goes on forever without halting. (It "loops")

Ex TM as a transition diagram:



Input: 00001111



$\Sigma = \{0, 1\}$
 $\Gamma = \{0, 1, B\}$
(6)

