Intro to Theory of Computation

CS 464

LECTURE 9

Last time:

- Converting a PDA to a CFG
- Pumping Lemma for CFLs

Today:

- Pumping Lemma for CFLs
- Review of CFGs/PDAs

Sofya Raskhodnikova

I-clicker problem (frequency: AC)

To prove that a language L is not context free, we can

A. **argue that a PDA cannot remember enough information to recognize L;**

B. **use pumping lemma;**

C. **give a CFG and show that it does not generate L.**

D. **None of the above.**

E. **More than one choice above works.**

Pumping lemma for CFLs

Let L be a context-free language

Then there exists P such that
For every $w \in L$ with $|w| \geq P$

there exist $uvxyz = w$, where:

1. $|vy| > 0$
2. $|vxy| \leq P$
3. $uv^ixy^iz \in L$ for all $i \geq 0$

CS 464

Negating the pumping lemma

L is not context-free

If for every P,

there is a w $\in$ L with $|w| \geq P$

for every uvxyz=w, where:

1. $|vy| > 0$
2. $|vxy| \leq P$

there is an $i \geq 0$, $uv^ixy^iz \notin L$

CS 464

Pumping lemma as a game

1. **YOU pick the language L to be proved not CFL.**
2. **ADVERSARY picks p, but doesn't reveal to YOU what p is; YOU must devise a play for all possible p's.**
3. **YOU pick w ∈ L, which should depend on p and which must be of length at least p.**
4. **ADVERSARY divides w into u, v, x, y, z, obeying PL conditions: vy is not empty and vxy has length $\leq p$. Again, ADVERSARY does not tell YOU what u, v, x, y, z are.**
5. **YOU win by picking i, which may be a function of $p, u, v, x, y,$ z, such that $uv^i xy^i z$ is not in L.**

I-clicker problem (frequency: AC)

Prove {ww | w ∈ {0,1}*} is not context free.

Proof: Assume … pumping length p.

What string can you choose in your next move?

A. 0101

B. $0^p 1^p$

C. $(01)^p$

D. $0^p 1^p 0^p 1^p$

E. More than one choice above works.

Using the pumping lemma

Prove L = {ww | w ∈ {0,1}*} is not context-free

Assume L is context-free.

Then there is a pumping length P.

No matter what P is, the string $s = 0^P1^P0^P1^P$ has $|s| \geq P$ and $s \in L$.

So there should be uvxyz=s with:

1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall i, uv^ixy^iz \in L$.

P P P P

s = 00...0011...1100...0011...11

{ww | w∈{0,1}*} is not a CFL: proof

No matter what P is, the string $s = 0^P1^P0^P1^P$ has $|s| \geq P$ and $s \in L$

vxy cannot be only in the first half since pumping down would move a 0 to the end of the first half.

Then pumping down must remove at least one 1 or one zero; e.g. $uxz = 0^P1^i0^j1^P$ where $i \neq P$ or $j \neq P$

vxy cannot be only in the last half since pumping up would move a 0 to the end of the first half

Both here should be u,v,x,y,z how they are chosen, (i = 0, 2, |vy| > 0, |vxy| ≤ P), $uv^ixy^iz \in L$; L cannot be context-free!

s = 00...0011...1100...0011...11

(braces over each block labelled P, P, P, P; brace under the middle labelled vxy)

CS 464

Using the pumping lemma

Prove L = $\{w\#w^R \mid w \in \{0,1\}^*$ and #1s = #0s$\}$ is not context-free

Assume L is context-free.

Then there is a pumping length P.

No matter what P is, the string $s = 0^P1^P\#1^P0^P$ has $|s| \geq P$ and $s \in L$.

So there should be uvxyz=s with:

1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall i$, $uv^ixy^iz \in L$.

P P P P

s = 00...0011...11#11...1100...00

$\{w\#w^R \mid w \in \{0,1\}^*$ and $\#1s = \#0s\}$ is not a CFL: proof

Assume L is context-free.

Then there is a pumping length P.

No matter what P is, the string $s = 0^P1^P\#1^P0^P$ has $|s| \leq P$ and $s \in L$.

So there should be uvxyz=s with:

1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall i,\ uv^ixy^iz \in L$.

vxy cannot be only in either half, since pumping would make one side of # longer than the other.

 # cannot be in v or y since pumping would add too many #s. So it must be in x.

since # is in x and $|vxy| \leq P$, neither v nor y can have 0s, and at least one must have a 1.

Then uv^2xy^2z has more 1s than 0s and is not in L.

P P P P

s = 00...0011...11#11...1100...00

vxy

CS 464

I-clicker problem (frequency: AC)

$$S \rightarrow aSb \mid bSa \mid SS \mid \varepsilon$$

What is the language of this CFG?

A. $\{a^n b^n | n \geq 0\} \cup \{b^n a^n | n \geq 0\}$

B. **The set of strings over alphabet $\{a, b\}^*$**

C. **The set of strings over alphabet $\{a, b\}^*$ with the same number of a's and b's**

D. **None of the above**

Give the language of this CFG

$$S \rightarrow aSb \mid bSa \mid SS \mid \varepsilon$$

Answer:

the set of all strings over alphabet {a,b} that have the same number of a's and b's

Prove that your answer is correct

CS 464

Proof of correctness $\Rightarrow$

$$S \to aSb \mid bSa \mid SS \mid \varepsilon$$

1. **Every generated string has the same number of a's and b's**

Proof:

Every rule that generates terminals, generates the same number of a's and b's

Proof of correctness $\Leftarrow$

$$S \to aSb \mid bSa \mid SS \mid \varepsilon$$

2. Every string that has the same number of a's and b's is generated.

Proof idea: Consider a string $w_1 \dots w_n$ that has the same number of a's and b's. Plot # of a's$-$# of b's in the first i characters of w as a function of i.

2. Every string that has the same number of a's and b's is generated.

2. Every string that has the same number of a's and b's is generated.

Claim. Every string of length $2m$ that has the same number of a's and b's is generated, $\forall$ integer $m \geq 0$.

Proof by strong induction on m.

Base case: $m = 0$. String ε is generated by the rule $S \rightarrow \varepsilon$.

IH: Suppose the claim holds $\forall m = 0, 1, \ldots, k$ for some k.

IS: We prove it for $m = k + 1$.

Each string of length $2(k+1)$ with the same number of a's and b's falls in one of the 3 cases on previous slide.

2. Every string that has the same number of a's and b's is generated.

Case 1: $w = aw'b$, where w' is a string of length $2k$ with the same number of a''s and b's.
Then, by IH, we can derive w' from S:

$$S \rightarrow \cdots \rightarrow w' \qquad (1)$$

To get w, we use $S \rightarrow aSb$ and then (1):

$$S \rightarrow aSb \rightarrow \cdots \rightarrow aw'b.$$

We proved that w is generated.

2. Every string that has the same number of a's and b's is generated.

Case 2: $w = bw'a,$ where w' is a string of length $2k$ with the same number of a''s and b's.

Then, by IH, we can derive w' from S:

$$S \to \cdots \to w' \qquad (2)$$

To get w, we use $S \to bSa$ and then (2):

$$S \to bSa \to \cdots \to bw'a.$$

We proved that w is generated.

CS 464

2. Every string that has the same number of a's and b's is generated.

Case 3: $w = w_s w_e,$ where w_s and w_e are strings of length $\leq 2k$ with the same number of a's and b's.

Then, by IH, we can derive w_s and w_e from S:

$$S \to \cdots \to w_s \qquad (3)$$

$$S \to \cdots \to w_e \qquad (4)$$

To get w, we use $S \to SS,$ then (3) and then (4):

$$S \to SS \to \cdots \to w_s S \to \cdots \to w_s w_e.$$

We proved that w is generated.

To conclude, all strings of length $2(k+1)$ with the same number of a's and b's are generated.

Context-free or not?

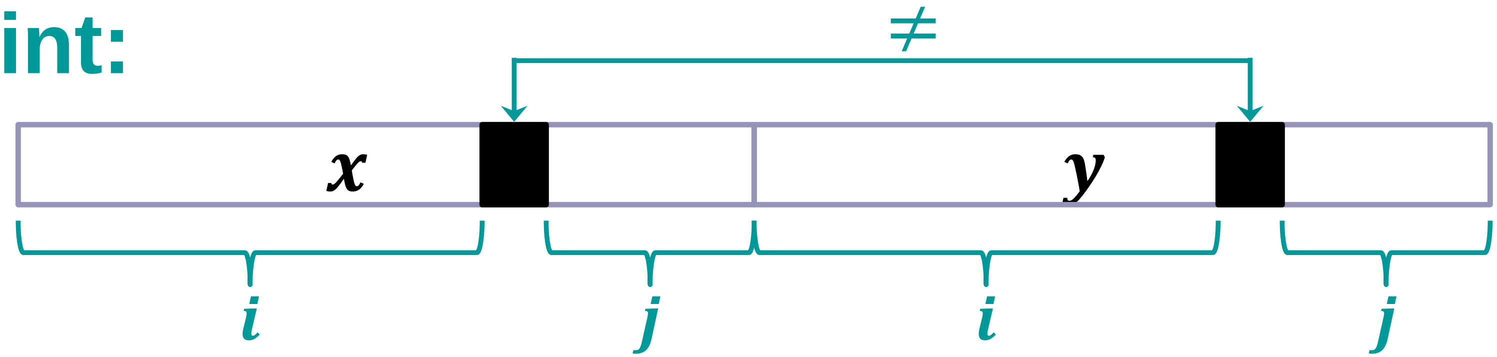
NOT $L_1 = \{ xy \mid x,y \in \{0,1\}^* \text{ and } x=y\}$

YES $L_2 = \{xy \mid x,y \in \{0,1\}^*, |x|=|y| \text{ and } x \neq y\}$

Give an algorithmic description of a PDA for

$\{xy \mid x, y \in \{0,1\}^*$ **and** $|x| = |y|$ **but** $x \neq y\}$.

Hint:



Give an algorithmic description of a PDA for

$\{xy \mid x, y \in \{0,1\}^* \text{ and } |x| = |y| \text{ but } x \neq y\}$.

Hint: