1 Pumping Lemma

1.1 Non-context-free languages

Non-Context Free Languages

Question
Are there languages that are not context-free? What about $L = \{a^n b^n c^n \mid n \geq 0\}$?

Answer
L is not context-free, because

- Recognizing if $w \in L$ requires remembering the number of as seen, bs seen and cs seen
- We can remember one of them on the stack (say as) , and compare them to another (say bs) by popping, but not to both bs and cs

The precise way to capture this intuition is through the pumping lemma

---

1.2 The Statement

Pumping Lemma for CFLs
Informal Statement

For all sufficiently long strings z in a context free language L, it is possible to find *two* substrings, not too far apart, that can be *simultaneously* pumped to obtain more words in L.

Lemma 1. *If L is a CFL, then $\exists p$ (pumping length) such that $\forall z \in L$, if $|z| \geq p$ then $\exists u, v, w, x, y$ such that $z = uvwxy$*

1. $|vwx| \leq p$
2. $|vx| > 0$
3. $\forall i \geq 0.\ uv^i wx^i y \in L$

Proof
Defered to later in the lecture.

Two Pumping Lemmas side-by-side

Context-Free Languages
If L is a CFL, then $\exists p$ (pumping length) such that $\forall z \in L$, if $|z| \geq p$ then $\exists u, v, w, x, y$ such that $z = uvwxy$

1. $|vwx| \leq p$
2. $|vx| > 0$

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

Regular Languages
If L is a regular language, then $\exists p$ (pumping length) such that $\forall z \in L$, if $|z| \geq p$ then $\exists u, v, w$ such that $z = uvw$

1. $|uv| \leq p$
2. $|v| > 0$
3. $\forall i \geq 0.\, uv^iw \in L$

Pumping Lemma for CFLs
Game View Game between *Defender*, who claims L satisfies the pumping condition, and Challenger, who claims L does not.

Defender		**Challenger**
Pick pumping length p	$\xrightarrow{p}$	
	$\xleftarrow{z}$	Pick $z \in L$ s.t. $|z| \geq p$
Divide z into u, v, w, x, y s.t. $|vwx| \leq p$, and $|vx| > 0$	$\xrightarrow{u,v,w,x,y}$	
	$\xleftarrow{i}$	Pick i, s.t. $uv^iwx^iy \notin L$

Pumping Lemma: If L is CFL, then there is always a winning strategy for the defender (i.e., challenger will get stuck).

Pumping Lemma (in contrapositive): If there is a winning strategy for the challenger, then L is not CFL.

Consequences of Pumping Lemma

- If L is context-free then L satisfies the pumping lemma.
- If L satisfies the pumping lemma that *does not* mean L is context-free
- If L does not satisfy the pumping lemma (i.e., challenger can win the game, *no matter* what the defender does) then L is not context-free.

1.3 Examples

Example I

Proposition 2. $L_{anbncn} = \{a^nb^nc^n \mid n \geq 0\}$ *is not a CFL.*

Proof. Suppose L_{anbncn} is context-free. *Let p be the pumping length.*

- Consider $z = a^p b^p c^p \in L_{anbncn}$.
- *Since $|z| > p$, there are u, v, w, x, y such that $z = uvwxy$, $|vwx| \leq p$, $|vx| > 0$ and $uv^i wx^i y \in L$ for all $i \geq 0$.*
- Since $|vwx| \leq p$, vwx cannot contain all three of the symbols a, b, c, because there are p bs. So vwx either does not have any as or does not have any bs or does not have any cs. Suppose, (wlog) vwx does have any as. Then $uv^0wx^0y = uwy$ contains more as than either bs or cs. Hence $uwy \notin L$. □

Example II

Proposition 3. $L_{a=c \wedge b=d} = \{a^i b^j c^i d^j \mid i, j \geq 0\}$ *is not a CFL.*

Proof. Suppose $L_{a=c \wedge b=d}$ is context-free. *Let p be the pumping length.*

- Consider $z = a^p b^p c^p d^p \in L$.
- *Since $|z| > p$, there are u, v, w, x, y such that $z = uvwxy$, $|vwx| \leq p$, $|vx| > 0$ and $uv^i wx^i y \in L$ for all $i \geq 0$.*
- Since $|vwx| \leq p$, v, x cannot contain both as and cs, nor can it contain both bs and ds. Further $|vx| > 0$. Now $uv^0wx^0y = uwy \notin L$, because it either contains fewer as than cs, or fewer cs than as, or fewer bs than ds, or fewer ds than bs. □

Example III
Wrong Proof

Proposition 4. $E = \{ww \mid w \in \{0, 1\}^*\}$ *is not a CFL.*

Proof. Suppose E is context-free. *Let p be the pumping length.*

- Consider $z = 0^p 1 0^p 1 \in L$.
- *z can be pumped if we make the following division.*

$$\overbrace{\underbrace{00\cdots00}_{u}\,\underbrace{0}_{v}\,\underbrace{1}_{w}}^{0^p1}\,\overbrace{\underbrace{0}_{x}\,\underbrace{00\cdots001}_{y}}^{0^p1}$$

- So is E CFL? No! Does E satisfy the pumping lemma? No!

□

Example III
Corrected Proof

Proposition 5. $E = \{ww \mid w \in \{0, 1\}^*\}$ *is not a CFL.*

Proof. Suppose E is context-free. *Let p be the pumping length.*

- Consider $z = 0^p1^p0^p1^p \in L$.
- *Since $|z| > p$, there are u, v, w, x, y such that $z = uvwxy$, $|vwx| \leq p$, $|vx| > 0$ and $uv^iwx^iy \in L$ for all $i \geq 0$.*
- vwx must straddle the midpoint of z.
  - Suppose vwx is only in the first half. Then in uv^2wx^2y the second half starts with 1. Thus, it is not of the form ww.
  - Case when vwx is only in the second half. Then in uv^2wx^2y the first half ends in a 0. Thus, it is not of the form ww.
  - Suppose vwx straddles the middle. Then uv^0wx^0y must be of the form $0^p1^i0^j1^p$, where either i or j is not p. Thus, $uv^0wx^0y \notin E$. □

1.4 Proof of the Pumping Lemma: Informal Ideas

Proof of Pumping Lemma

Lemma 6. *If L is a CFL, then $\exists p$ (pumping length) such that $\forall z \in L$, if $|z| \geq p$ then $\exists u, v, w, x, y$ such that $z = uvwxy$*

1. $|vwx| \leq p$
2. $|vx| > 0$
3. $\forall i \geq 0.\ uv^iwx^iy \in L$

Let G be a CFG in *Chomsky Normal Form* such that $L(G) = L$. Let z be a "very long" string in L ("very long" made precise later).

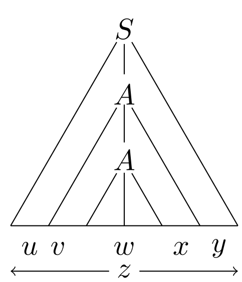

Figure 1: Parse Tree for z

- Since $z \in L$ there is a parse tree for z
- Since z is very long, the parse tree (which is a binary tree) must be "very tall"

- The longest path in the tree, by pigeon hole principle, must have some variable (say) A repeat. Let u, v, w, x, y be as shown.

Pumping down and up

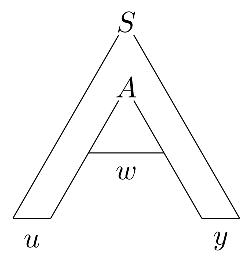

Figure 2: Pumping zero times

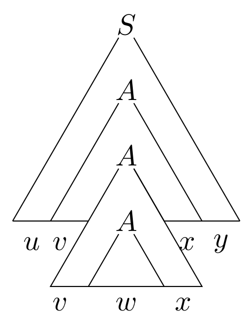

Figure 3: Pumping two times

- Thus, uv^iwx^iy has a parse tree, for any i.

1.5 Formal Proof

Proof of Pumping Lemma

Proof. Let G be a grammar in *Chomsky Normal Form* with k variables such that $L(G) = L$. Take $p = 2^k$. Consider $z \in L$ such that $|z| \geq p = 2^k$.

- Consider a parse tree for z. Height of this tree is at least $k + 1$. For proof see Homework 6, problem 3.
- Repeated Variables:

- A parse tree for z has a path of length $k+1$
- A path of length $k+1$ has $k+2$ vertices, out of which the last one is leaf that is labelled by a terminal; thus, there are at least $k+1$ internal vertices on path.
- Thus, there must be two vertices n_1 and n_2 on this path such that n_1 and n_2 have the same label (say A) and n_1 is an ancestor of n_2.
- Let the yield of tree rooted at n_2 be w, and yield of n_1 be vwx. Yield of the root = z is say $uvwxy$.

- Properties of u, v, w, x, y
  - Height of n_1 can be assumed to be at most $k+1$; thus, the yield of n_1 (vwx) is at most $2^k = p$.
  - $n_1 \neq n_2$. Since the grammar has no ϵ-productions and no unit-productions, $vwx \neq w$. i.e., $|vx| > 0$.
- Pumping the strings: Based on the parse tree for z, and definitions of u, v, w, x, y, we have
  - There is a parse tree with yield uAy and root S, obtained by not expanding n_1. Thus, $S \stackrel{*}{\Rightarrow} uAy$.
  - There is a parse tree with yield vAx and root A, obtained from n_1 and not expanding n_2. Thus, $A \stackrel{*}{\Rightarrow} vAx$.
  - There is a parse tree with yield w and root A; this is the tree rooted at n_2. Thus, $A \stackrel{*}{\Rightarrow} w$.

  Putting it together, we have

$$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uvAxy \stackrel{*}{\Rightarrow} uvvAxxy \stackrel{*}{\Rightarrow} \cdots \stackrel{*}{\Rightarrow} uv^iAx^iy \stackrel{*}{\Rightarrow} uv^iwx^iy \qquad \square$$

---

2 Closure Properties

Proving Non-context-freeness

Like in the case of regular languages, one can use closure properties to show that a language is non-contextfree. To prove L is not context-free, we construct a language L' from L using only operations under which context-free languages are known to be closed. If L' is known to be not context free (like L_{anbncn}) then one can conclude that L is not context-free.

Example 7. Here is a proof that $E = \{ww \mid w \in \{0,1\}^*\}$ is not a CFL using closure properties.

Consider $h : \{a,b,c,d\}^* \to \{0,1\}*$ such that $h(a) = 0$, $h(b) = 1$, $h(c) = 0$, $h(d) = 1$. Observe that

$$h^{-1}(E) \cap \mathbf{L}(a^*b^*c^*d^*) = L_{a=c \wedge b=d}$$

Thus, E is not context-free.
