Lecture 8

Properties of Context-free Languages

COSE215: Theory of Computation

Seunghoon Woo

Fall 2023

Contents

- **Pumping lemma for context-free languages**

Pumping lemma for context-free languages

- **Revisit: pumping lemma for regular languages**

Identifying non-regular languages

- **Pumping lemma**
  - Let L be a regular language
  - There exists a positive integer m such that any $w \in L$, if $|w| \geq m$, there exists $w = xyz$ such that
    - ❖ $|xy| \leq m$
    - ❖ $|y| \geq 1$
    - ❖ For all $i \geq 0,\ xy^iz \in L$
  - A string of sufficiently large length ($|w| \geq m$) can be represented in the form of xyz, and xy^iz, which is pumped "y" i times, can also always belong to this language

Pumping lemma for context-free languages

- **Similar approach**
  - Even if we pump a portion for a long string, it should be contained in the CFL!

Pumping lemma for context-free languages

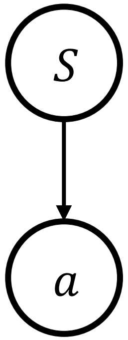
- **Pumping lemma for CFL**
  - Let L be a context-free language (infinite)
  - There exists a positive integer m such that any $w \in L$, if $|w| \geq m$, there exists $w = uvxyz$ such that
    - ❖ $|vxy| \leq m$
    - ❖ $|vy| \geq 1$
    - ❖ For all $i \geq 0,\ uv^ixy^iz \in L$

Pumping lemma for context-free languages

- **The size of parse tree in CNF**
  - For a given CFG G in CNF
    - For all $w \in L(G)$
      - If the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$

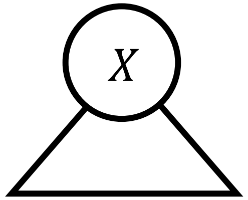
Pumping lemma for context-free languages

- **The size of parse tree in CNF**
  - For a given CFG G in CNF
    - ❖For all $w \in L(G)$
      - If the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$
  - Proof by induction
    - ❖ If $n = 1$
      - $|w| = 1$
      - $2^{n-1} = 2^0 = 1$
      - $\therefore |w| = 2^{n-1}$



Pumping lemma for context-free languages

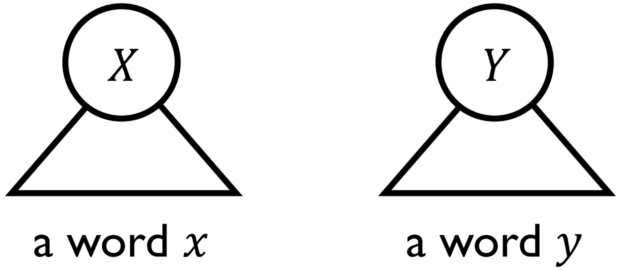
- **The size of parse tree in CNF**
  - For a given CFG G in CNF
    - ❖For all $w \in L(G)$
      - If the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$
  - Proof by induction
    - ❖ Let $n = k - 1$ and $X \overset{*}{\Rightarrow} x$
    - ❖ By the inductive hypothesis, $|x| \leq 2^{k-2}$



a word x

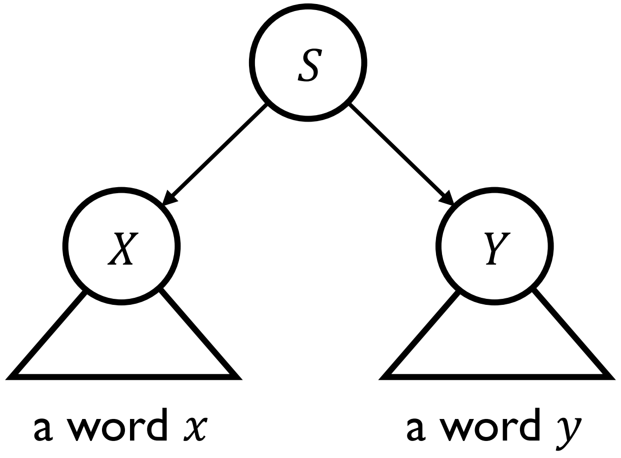
Pumping lemma for context-free languages

- **The size of parse tree in CNF**
  - For a given CFG G in CNF
    - For all $w \in L(G)$
      - If the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$
  - Proof by induction
    - Let $n = k - 1$ and $X \overset{*}{\Rightarrow} x$
    - By the inductive hypothesis, $|x| \leq 2^{k-2}$
    - Similarly, consider $Y \overset{*}{\Rightarrow} y$ ($|y| \leq 2^{k-2}$)



Pumping lemma for context-free languages

- **The size of parse tree in CNF**
  - For a given CFG G in CNF
    - ❖ For all $w \in L(G)$
      - If the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$
  - Proof by induction
    - ❖ Let n = $k - 1$ and $X \overset{*}{\Rightarrow} x$
    - ❖ By the inductive hypothesis, $|x| \leq 2^{k-2}$
    - ❖ Similarly, consider $Y \overset{*}{\Rightarrow} y$ ($|y| \leq 2^{k-2}$)
    - ❖ If there is a rule $S \rightarrow XY$ exists ($S \overset{*}{\Rightarrow} w$) where $w = xy$, then $|w| = |x| + |y| \leq 2^{k-2} + 2^{k-2} = 2^{k-1}$
      - Inductive step is proved



Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - Let L be a context-free language

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - Let L be a context-free language
  - Starting with a CNF grammar $G = (V, T, S, P)$ such that $L(G) = L$

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - Let L be a context-free language
  - Starting with a CNF grammar $G = (V, T, S, P)$ such that $L(G) = L$
  - Let $n \geq 0$ be the number of variables in G and choose $m = 2^n \geq 1$

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - Let L be a context-free language
  - Starting with a CNF grammar $G = (V, T, S, P)$ such that $L(G) = L$
  - Let $n \geq 0$ be the number of variables in G and choose $m = 2^n \geq 1$
  - Take any $w = a_1 a_2 \ldots a_k \in L$ such that $|w| = k \geq m$

Pumping lemma for context-free languages

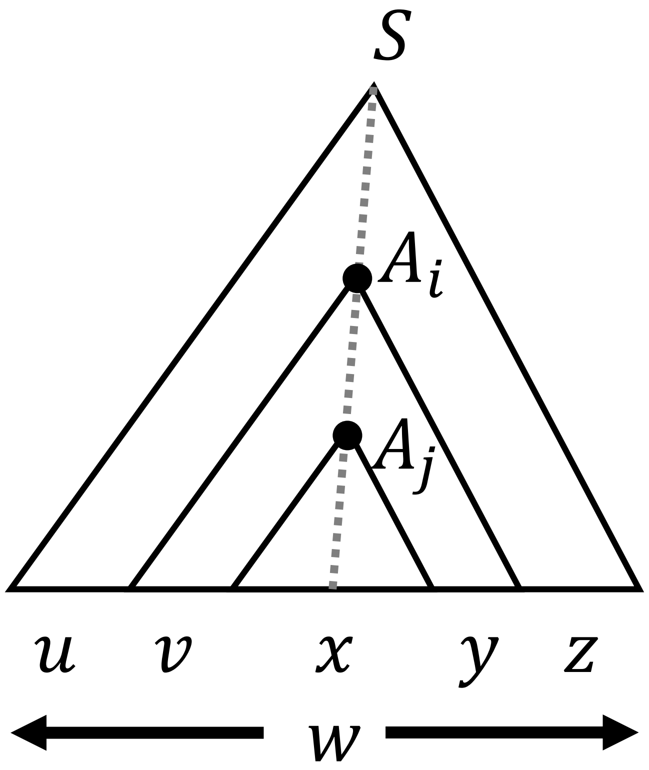
- **Proof of pumping lemma**
  - Let L be a context-free language
  - Starting with a CNF grammar $G = (V, T, S, P)$ such that $L(G) = L$
  - Let $n \geq 0$ be the number of variables in G and choose $m = 2^n \geq 1$
  - Take any $w = a_1 a_2 \ldots a_k \in L$ such that $|w| = k \geq m$
  - Consider the longest path $(S, A_1, A_2, \ldots, A_{p-1})$ in the parse tree of w
    - ❖ $|w| = k \leq 2^{p-1}$
    - ❖ Because $k \geq m = 2^n$, then $2^{p-1} \geq 2^n$, therefore, $p \geq n + 1$
    - ❖ This means that there are at least $n + 1$ occurrences of variables on the path

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - Let L be a context-free language
  - Starting with a CNF grammar $G = (V, T, S, P)$ such that $L(G) = L$
  - Let $n \geq 0$ be the number of variables in G and choose $m = 2^n \geq 1$
  - Take any $w = a_1 a_2 \dots a_k \in L$ such that $|w| = k \geq m$
  - Consider the longest path $(S, A_1, A_2, \dots, A_{p-1})$ in the parse tree of w
    - ❖ $|w| = k \leq 2^{p-1}$
    - ❖ Because $k \geq m = 2^n$, then $2^{p-1} \geq 2^n$, therefore, $p \geq n + 1$
    - ❖ This means that there are at least $n + 1$ occurrences of variables on the path
  - As there are only n different variables, at least two of the variables should be the same
    - ❖ Suppose $A_i = A_j$ where $p - n \leq i < j \leq p$

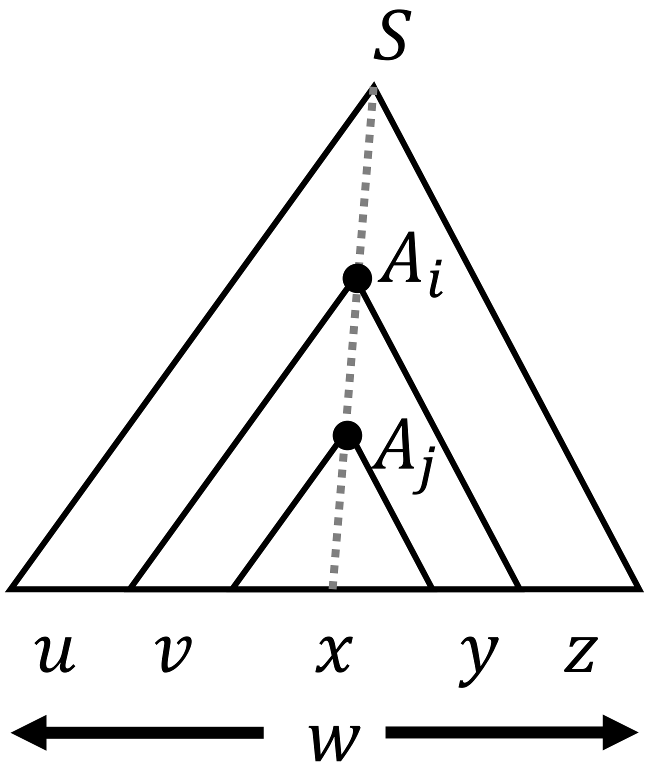
Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - It is possible to divide the parse tree of $w = uvxyz$ as follows



Pumping lemma for context-free languages

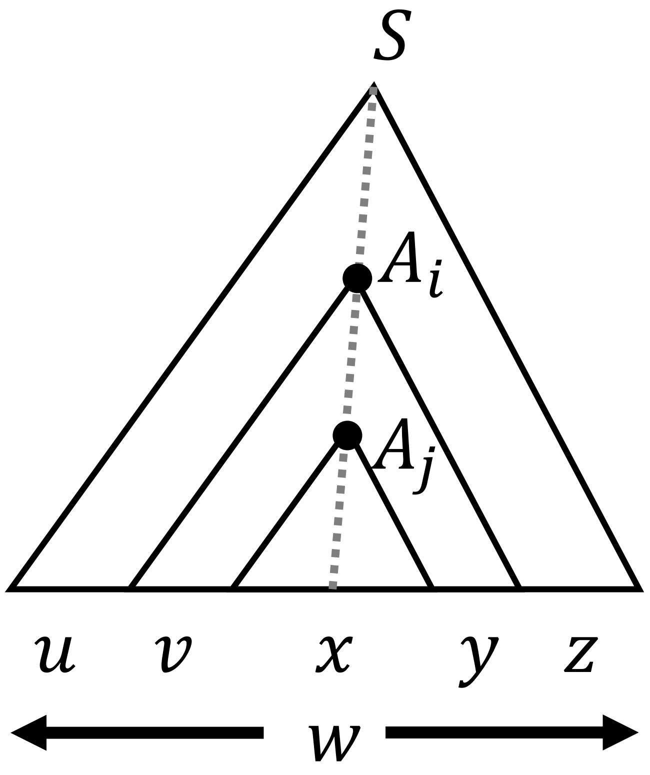
- **Proof of pumping lemma**
  - It is possible to divide the parse tree of $w = uvxyz$ as follows



① Since there are no unit or λ productions, v and y could not both be λ
$\therefore |vy| \geq 1$

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - It is possible to divide the parse tree of $w = uvxyz$ as follows



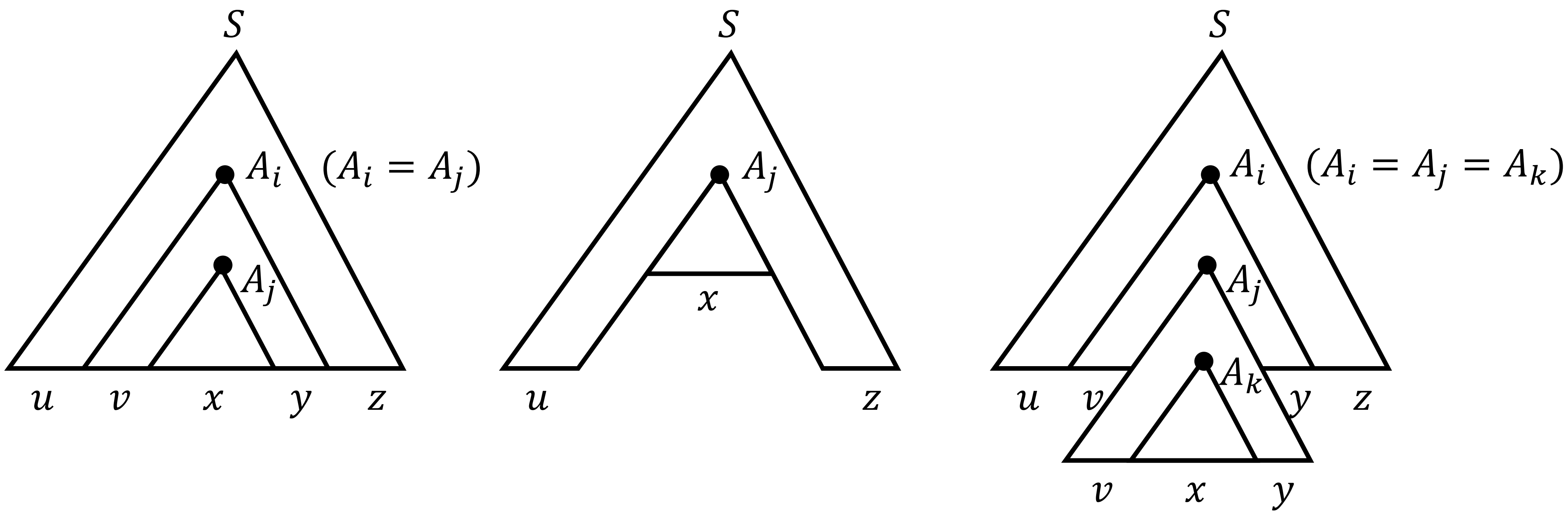
① Since there are no unit or λ productions,
v and y could not both be λ
$\therefore |vy| \geq 1$

② The length of the longest path from A_i
in the parse tree of w is $p - i + 1$
=> Since $p - n \leq i$, then $p - i + 1 \leq n + 1$
$\therefore |vxy| \leq 2^{n+1-1} = 2^n = m$

Pumping lemma for context-free languages

- **Proof of pumping lemma**
  - It is possible to divide the parse tree of $w = uvxyz$ as follows

③ For all $i \geq 0, uv^ixy^iz \in L$



Pumping lemma for context-free languages

- **Pumping lemma for CFL**
  - Adversary game
    1. We pick a language L that we want to show is not a CFL
    2. Our "adversary" gets to pick m, which we do not know, and we therefore must plan for any possible m
    3. We get to pick w, and may use m as a parameter when we do so
    4. Our adversary gets to break w into $uvxyz$, subject only to the constraints that $|vxy| \leq m$ and $|vy| \geq 1$
    5. We win the game, if we can, by picking i and showing that uv^ixy^iz is not in L

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example**
  - Show that the following language is not context free
    - $L = \{a^n b^n c^n : n \geq 0\}$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example**
  - $L = \{a^n b^n c^n : n \geq 0\}$
    - ❖ The adversary picks m, and suppose we pick $w = a^m b^m c^m$
    - ❖ The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example**
  - $L = \{a^n b^n c^n : n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m c^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Case 1: The adversary picks vxy consist entirely of a's (or b's or c's)

      - Let $|vy| = k$
      - If we pump v and y i times ($i \neq 1$), the result string $w' = a^{m+(i-1)k} b^m c^m \notin L$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example**
  - $L = \{a^n b^n c^n : n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m c^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Case 2: The adversary picks vxy consist of a's and b's (or b's and c's)

      - Then, the pumped string $a^k b^k c^m \notin L$ $(k \neq m)$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example**
  - $L = \{a^n b^n c^n : n \geq 0\}$
    - ❖ The adversary picks m, and suppose we pick $w = a^m b^m c^m$
    - ❖ The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$
    - ❖ The only way is to pick vy with the same number of 'a's, 'b's, and 'c's
      - Impossible: because of the condition $|vxy| \leq m$
      - L is not context free!

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example2**
  - $L = \{ww : w \in \{a, b\}^*\}$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example2**
  - $L = \{ww: w \in \{a,b\}^*\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m a^m b^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example2**
  - $L = \{ww : w \in \{a, b\}^*\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m a^m b^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Case 1: The adversary picks vxy consist entirely of a's
      - Let $|vy| = k$
      - If we pump v and y i times, the result string $w' = a^{m+(i-1)k} b^m a^m b^m \notin L$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example2**
  - $L = \{ww: w \in \{a, b\}^*\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m a^m b^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Case 2: The adversary picks v consists of a's and y consists of b's

      - Then the pumped string $a^k b^j a^m b^m \notin L$ can be generated $(k \neq m$ and $j \neq m)$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example2**
  - $L = \{ww : w \in \{a, b\}^*\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m a^m b^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$
    - ❖ vy cannot cover both two different blocks consisting of a's (or b's)
      - Impossible: because of the condition $|vxy| \leq m$
      - L is not context free!

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example3**
  - $L = \{a^n b^n : n \geq 0\}$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example3**
  - $L = \{a^n b^n : n \geq 0\}$
    - ❖ The adversary picks m, and suppose we pick $w = a^m b^m$
    - ❖ The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example3**
  - $L = \{a^n b^n : n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^m b^m$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$
    - ❖The adversary picks v consists of a's ($v = a^k$) and y consists of b's ($y = b^k$)
      - No matter how what i we choose, the resulting pumped string w_i is in L
      - Unable to get any conclusion from the pumping lemma in this case

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example4**
  - $L = \{a^{n!}: n \geq 0\}$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example4**
  - $L = \{a^{n!} : n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^{m!}$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example4**
  - $L = \{a^{n!}: n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^{m!}$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Only one case: The adversary picks $v = a^k$ and $y = a^l$

      - $w_0 = uxz = a^{m!-(k+l)}$
      - $m! - (k + l) = j!$ ⁇

Pumping lemma for context-free languages

- **Pumping lemma for CFL: example4**
  - $L = \{a^{n!}: n \geq 0\}$
    - ❖The adversary picks m, and suppose we pick $w = a^{m!}$
    - ❖The adversary breaks $w = uvxyz$, where $|vxy| \leq m$ and $|vy| \geq 1$

    Only one case: The adversary picks $v = a^k$ and $y = a^l$

    - $w_{m!+1} = a^{m!+m!(k+l)} = a^{m!(1+k+l)}$
    - $m! < m! \times (1 + k + l) \leq m! \times (1 + m) = (m + 1)!$ (because $k + l < m$)
    - L is not context free!

Pumping lemma for context-free languages

- **Pumping lemma for CFL: practice**
  - $L = \{a^n b^j : n = j^2\}$

Pumping lemma for context-free languages

- **Pumping lemma for CFL**
  - Pumping Lemma for CFL is violated => not a CFL
  - Pumping Lemma for CFL is not violated => do not know if it is context free or not

Next Lecture

- **Closure properties of CFL**