Lecture 9 – The Pumping Lemma for Regular Languages

COSE215: Theory of Computation

Jihyeok Park

PLRG

2026 Spring

Recall

The class of regular languages is closed under the following operations:

- Union
- Concatenation
- Kleene Star
- Complement
- Intersection
- Difference
- Reversal
- Homomorphism
- Inverse Homomorphism

But, it is **NOT** closed under the following operations:

- Infinite Union
- Infinite Intersection

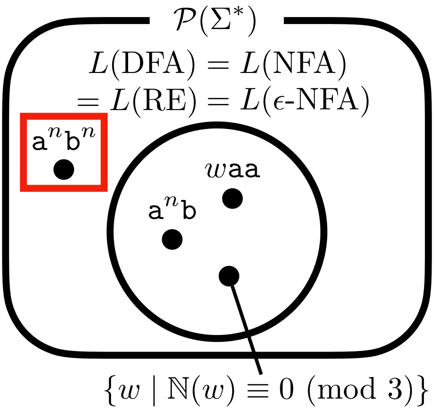
For example, consider $L_i = \{\mathtt{a}^i\mathtt{b}^i\}$ for $i \geq 0$. Then, L_i and its complement L_i^c are regular languages. But, their infinite union and intersection are not regular languages:

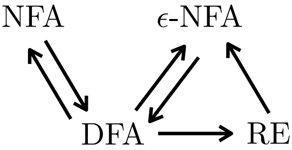
$$\bigcup_{i \geq 0} L_i = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\} = L \qquad \bigcap_{i \geq 0} L_i^c = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}^c = L^c$$

because $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$ is not regular.

Recall

- Not all languages are regular: e.g., $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$.





- How to prove that a language is **NOT** regular? **Pumping Lemma!**

Contents

Contents

Intuition

Let's consider a **regular language** L.

Then, there exists a **regular expression** R such that $L(R) = L$.

Then, there are two possibilities:

1. R does **not contain** the **Kleene star** operator $(^*)$.

   `abc` `ab|(bc`ϵ`)|d` `(a|bcd)`ϵ`|`$\varnothing$`bca`

   Then, the language L should be **finite**.

2. R **contains** at least one Kleene star operator $(^*)$.

   $$R_1 \, R_2{}^* \, R_3$$

   Roughly speaking, we can **repeat** the middle part R_2 as many times as we want (including 0 times) when we generate a string.

   The **Pumping Lemma** formally captures this intuition.

Lemma (Pumping Lemma for Regular Languages)

For a given regular language L, **there exists** *a positive integer n such that* **for all** $w \in L$, *if* $|w| \geq n$, **there exists** $w = xyz$ *such that*

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^iz \in L$

$$A \quad = \quad L \text{ is regular}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ \text{①} \land \text{②} \land \text{③}$$

Proof of Pumping Lemma

- Let L be a regular language.
- Then, $\exists$ DFA $D = (Q, \Sigma, \delta, q_0, F)$. s.t. $L(D) = L$. Let $n = |Q| > 0$.
- Take any $w = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_m \in L$ s.t. $|w| = m \geq n$.
- Let $p_i = \delta^*(q_0, \mathtt{a}_1 \cdots \mathtt{a}_i)$ for all $0 \leq i \leq m$. Then, $p_0 = q_0 \land p_m \in F$.

$$q_0 = p_0 \xrightarrow{\mathtt{a}_1} p_1 \xrightarrow{\mathtt{a}_2} \cdots \xrightarrow{\mathtt{a}_n} p_n \xrightarrow{\mathtt{a}_{n+1}} \cdots \xrightarrow{\mathtt{a}_m} p_m \in F$$

Proof of Pumping Lemma

- Let L be a regular language.
- Then, $\exists$ DFA $D = (Q, \Sigma, \delta, q_0, F)$. s.t. $L(D) = L$. Let $n = |Q| > 0$.
- Take any $w = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_m \in L$ s.t. $|w| = m \geq n$.
- Let $p_i = \delta^*(q_0, \mathtt{a}_1 \cdots \mathtt{a}_i)$ for all $0 \leq i \leq m$. Then, $p_0 = q_0 \wedge p_m \in F$.

$$q_0 = \underbrace{p_0 \xrightarrow{\mathtt{a}_1} p_1 \xrightarrow{\mathtt{a}_2} \cdots \xrightarrow{\mathtt{a}_n} p_n}_{n+1 \text{ states}} \xrightarrow{\mathtt{a}_{n+1}} \cdots \xrightarrow{\mathtt{a}_m} p_m \in F$$

- By Pigeonhole Principle, there exists $i < j \leq n$ s.t. $p_i = p_j$.

$$q_0 = p_0 \overbrace{\xrightarrow{\mathtt{a}_1} p_1 \xrightarrow{\mathtt{a}_2} \cdots \xrightarrow{\mathtt{a}_i}}^{x} p_i \overbrace{\xrightarrow{\mathtt{a}_{i+1}} \cdots \xrightarrow{\mathtt{a}_j}}^{y} p_j \overbrace{\xrightarrow{\mathtt{a}_{j+1}} \cdots \xrightarrow{\mathtt{a}_n} p_n \xrightarrow{\mathtt{a}_{n+1}} \cdots \xrightarrow{\mathtt{a}_m}}^{z} p_m \in F$$

- We can split $w = xyz$ as above. Then,

$$|y| = j - i > 0 \qquad\qquad |xy| = j \leq n$$

- Since y represents a **cycle** from p_i to p_i, $\forall i \geq 0.\ \delta^*(q_0, xy^iz) = p_m$.
- It means that $\forall i \geq 0.\ xy^iz \in L$.

Proof of Pumping Lemma

PLRG

- Let L be a regular language.
- Then, $\exists$ DFA $D = (Q, \Sigma, \delta, q_0, F)$. s.t. $L(D) = L$. Let $n = |Q| > 0$.
- Take any $w = a_1 a_2 \cdots a_m \in L$ s.t. $|w| = m \geq n$.
- Let $p_i = \delta^*(q_0, a_1 \cdots a_i)$ for all $0 \leq i \leq m$. Then, $p_0 = q_0 \wedge p_m \in F$.

$$q_0 = \underbrace{p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} p_n}_{n+1 \text{ states}} \xrightarrow{a_{n+1}} \cdots \xrightarrow{a_m} p_m \in F$$

- By Pigeonhole Principle, there exists $i < j \leq n$ s.t. $p_i = p_j$.

$$q_0 = p_0 \overbrace{\xrightarrow{a_1} p_1 \xrightarrow{a_2} \cdots \xrightarrow{a_i}}^{x} p_i \overbrace{\xrightarrow{a_{i+1}} \cdots \xrightarrow{a_j}}^{y} p_j \overbrace{\xrightarrow{a_{j+1}} \cdots \xrightarrow{a_n} p_n \xrightarrow{a_{n+1}} \cdots \xrightarrow{a_m}}^{z} p_m \in F$$

- We can split $w = xyz$ as above. Then,

$$\text{①} - |y| = j - i > 0 \qquad\qquad |xy| = j \leq n - \text{②}$$

- Since y represents a **cycle** from p_i to p_i, $\forall i \geq 0.\ \delta^*(q_0, xy^iz) = p_m$.
- It means that $\forall i \geq 0.\ xy^iz \in L$. – ③

Proof of Pumping Lemma

Lemma (Pumping Lemma for Regular Languages)

For a given regular language L, **there exists** *a positive integer n such that* **for all** $w \in L$, *if* $|w| \geq n$, **there exists** $w = xyz$ *such that*

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^i z \in L$

$$A \quad = \quad L \text{ is regular}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③$$

Application: Proving Languages are Not Regular

Lemma (Pumping Lemma for Regular Languages)

$$A \quad = \quad L \textit{ is regular}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③$$

$$A \Longrightarrow B \quad \text{(O)}$$

$$B \Longrightarrow A \quad \text{(X)}$$

$$\neg B \Longrightarrow \neg A \quad \text{(O)}$$

$$\begin{aligned}
\neg B &= \neg(\exists n > 0.\ \forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③) \\
&= \forall n > 0.\ \neg(\forall w \in L.\ |w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③) \\
&= \forall n > 0.\ \exists w \in L.\ \neg(|w| \geq n \Rightarrow \exists w = xyz.\ ① \land ② \land ③) \\
&= \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \neg(\exists w = xyz.\ ① \land ② \land ③) \\
&= \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ \neg(① \land ② \land ③) \\
&= \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ \neg(① \land ②) \lor \neg ③ \\
&= \forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ (① \land ②) \Rightarrow \neg ③
\end{aligned}$$

Application: Proving Languages are Not Regular

To prove a language L is **NOT** regular, we need to show that

$$\forall n > 0.\ \exists w \in L.\ |w| \geq n \land \forall w = xyz.\ (①\land ②) \Rightarrow \neg ③$$

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0.\ xy^i z \in L$

Note that $\neg ③ = \exists i \geq 0.\ xy^i z \notin L$.

We can prove this by following the steps below:

1. Assume any positive integer n is given.
2. Pick a word $w \in L$.
3. Show that $|w| \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \land$ ② $|xy| \leq n$.
5. $\neg$③ Pick $i \geq 0$, and show that $xy^i z \notin L$ using ① and ②.

Contents

Example 1

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{\mathrm{a}^n\mathrm{b}^n \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $w = \mathrm{a}^n\mathrm{b}^n \in L$.
3. $|w| = n + n = 2n \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \wedge$ ② $|xy| \leq n$.
5. Let $i = 0$. We need to show that $\neg$③ $xy^0z \notin L$:
   - Since ② $|xy| \leq n$,

     $$x = \mathrm{a}^p \qquad y = \mathrm{a}^q \qquad z = \mathrm{a}^{n-p-q}\mathrm{b}^n$$

     for some $0 \leq p, q \leq n$ such that $p + q \leq n$.
   - Since ① $|y| > 0$, we know $q > 0$.
   - Finally, $xy^0z = xz = \mathrm{a}^p\mathrm{a}^{n-p-q}\mathrm{b}^n = \mathrm{a}^{n-q}\mathrm{b}^n \notin L$ ($\because q > 0$). □

Example 2

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

1. Assume any positive integer n is given.
2. Let $w = \mathtt{a}^n\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n \in L$.
3. $|w| = n + n + n + n = 4n \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \land$ ② $|xy| \leq n$.
5. Let $i = 0$. We need to show that ¬③ $xy^0z \notin L$:
   - Since ② $|xy| \leq n$,

     $$x = \mathtt{a}^p \qquad y = \mathtt{a}^q \qquad z = \mathtt{a}^{n-p-q}\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n$$

     for some $0 \leq p, q \leq n$ such that $p + q \leq n$.
   - Since ① $|y| > 0$, we know $q > 0$.
   - Finally, $xy^0z = xz = \mathtt{a}^p\mathtt{a}^{n-p-q}\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n = \mathtt{a}^{n-q}\mathtt{b}^n\mathtt{b}^n\mathtt{a}^n \notin L$ $(\because q > 0)$.

□

Example 3

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{a^l b^m c^n \mid l + m \leq n\}$$

1. Assume any positive integer n is given.
2. Let $w = a^n b^n c^{2n} \in L$.
3. $|w| = n + n + 2n = 4n \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \wedge$ ② $|xy| \leq n$.
5. Let $i = $ 2 . We need to show that $\neg$③ $xy^2z \notin L$:
   - Since ② $|xy| \leq n$,

     $$x = a^p \qquad y = a^q \qquad z = a^{n-p-q} b^n c^{2n}$$

     for some $0 \leq p, q \leq n$ such that $p + q \leq n$.
   - Since ① $|y| > 0$, we know $q > 0$.
   - Finally, $xy^2z = xyyz = a^{n+q} b^n c^{2n} \notin L$
     ($\because q > 0$. Thus, $(n + q) + n = 2n + q > 2n$). □

Example 4

Let's prove that L is **NOT** regular using the Pumping Lemma:

$$L = \{a^{n^2} \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $w = a^{n^2} \in L$.
3. $|w| = n^2 \geq n$.
4. Assume any split $w = xyz$ is given, and ① $|y| > 0 \land$ ② $|xy| \leq n$.
5. Let $i = 2$. We need to show that ¬③ $xy^2z \notin L$:
   - Since ① $|y| > 0$ and ② $|xy| \leq n$,

     $$y = a^k$$

     where $0 < k \leq n$. Then,

     $$n^2 < n^2 + k \ (\because 0 < k) \qquad n^2 + k < (n+1)^2 \ (\because k \leq n)$$

   - Finally, $xy^2z = xyyz = a^{n^2+k} \notin L$ □

Example 5

Let's prove that L is **NOT** regular:

$$L = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^{n+k} \mid n, k \geq 0\}$$

- It is much easier to use **closure properties** under **homomorphisms**.
- Consider a homomorphism $h : \{\mathtt{a}, \mathtt{b}, \mathtt{c}\} \to \{\mathtt{a}, \mathtt{b}\}^*$:

$$h(\mathtt{a}) = \mathtt{a} \qquad h(\mathtt{b}) = \mathtt{a} \qquad h(\mathtt{c}) = \mathtt{b}$$

- Then,

$$h(L) = \{\mathtt{a}^{n+k}\mathtt{b}^{n+k} \mid n, k \geq 0\} = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$$

- If L is regular, then $h(L)$ must be regular as well.
- However, we know $h(L)$ is **NOT** regular.
- Therefore, L is **NOT** regular. $\square$

Summary

1. Pumping Lemma for Regular Languages
   - Intuition
   - Pumping Lemma
   - Proof of Pumping Lemma
   - Application: Proving Languages are Not Regular

2. Examples
   - Example 1: $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 0\}$
   - Example 2: $L = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
   - Example 3: $L = \{\mathtt{a}^l\mathtt{b}^m\mathtt{c}^n \mid l + m \leq n\}$
   - Example 4: $L = \{\mathtt{a}^{n^2} \mid n \geq 0\}$
   - Example 5: $L = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^{n+k} \mid n, k \geq 0\}$

Next Lecture

- Equivalence and Minimization of Finite Automata

Jihyeok Park
jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr