Theory of Formal Languages and Automata
Lecture 7

Mahdi Dolati

Sharif University of Technology

Fall 2024

March 2, 2024


Pumping Lemma

Nonregular Languages

- Limitations of finite automata
- Example:

$$B = \{0^n 1^n | n \geq 0\}. \quad (1)$$

- The machine has to remember the number of zeros, which is unlimited.
- It is impossible using any finite number of states.
- How to formally prove it? Do we need a proof?

Example

Consider the following languages:

$$C = \{w | w \text{ has an equal number of zeros and ones }\}, \text{ and} \quad (2)$$

$$D = \{w | w \text{ has an equal number of substrings } 01 \text{ and } 10\}. \quad (3)$$

Nonregular Languages

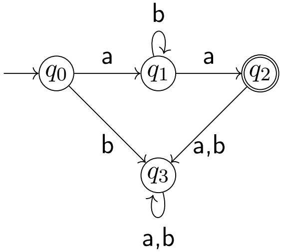
- Pumping lemma: The proof technique.
- According to pumping lemma, all regular languages have a special property.

Property

If a string in a regular language is longer than the pumping length, then, it contains a section that can be repeated indefinitely and still remain in the language.

Nonregular Languages

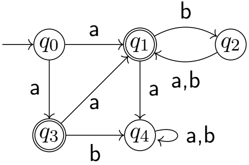
- Some strings:
  - aa
  - aba
  - abba
  - abbba
- Interestingly, for all strings of length at least three:



$$\underbrace{a}_{x}\underbrace{bbbb}_{y>0}\underbrace{a}_{z} \in L \rightarrow xy^i z \in L, \quad \forall i \geq 0$$

- Some strings:

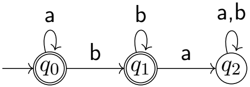
- a
- aba
- ababa
- abb
- abbbb
- abbbbbb
- ab[ab]
- ab[ab]b[ab]
- ab[ab]b[ab]b[ab]



- Interestingly, for all strings of length at least three:

$$\underbrace{a}_{x}\underbrace{b[ab]}_{y>0}\underbrace{\varepsilon}_{z} \in L \rightarrow xy^iz \in L, \quad \forall i \geq 0$$

- Some strings:
  - a
  - aa
  - aaa
  - aabbb
  - aaabbbb
  - aaaabbbbb
- Interestingly, for all strings of length at least one:



$$\underbrace{\varepsilon}_{x}\underbrace{a}_{y>0}\underbrace{\varepsilon}_{z} \in L \rightarrow xy^{i}z \in L, \quad \forall i \geq 0$$

$$\underbrace{\varepsilon}_{x}\underbrace{b}_{y>0}\underbrace{\varepsilon}_{z} \in L \rightarrow xy^{i}z \in L, \quad \forall i \geq 0$$

Theorem (Pumping lemma)

There is a number p (the pumping length) for any regular language A such that any string $s \in A$ of length at least p may be written as $s = xyz$, satisfying the following conditions:

1. $\forall i \geq 0 \quad xy^iz \in A$,
2. $|y| > 0$, *and*
3. $|xy| \leq p$.

Nonregular Languages

Proof idea:

- Let $M = (Q, \Sigma, \delta, q_q, F)$ be a DFA recognizing A and $|Q| = 5$.
- Let $s = s_1 s_2 \ldots s_n$ a string in A with $n = 7$.
- Let $q_1, q_3, \ldots, q_5$ be 8 states entered during processing of s.

$$\begin{array}{ccccccccccccccccc} s = & & s_1 & & s_2 & & s_3 & & s_4 & & s_5 & & s_6 & & s_7 & \\ & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\ & q_1 & & {\color{red}q_3} & & q_2 & & q_4 & & q_5 & & {\color{red}q_3} & & q_5 & & q_2 \end{array}$$

- Observed a repeated state in the first 6 states, i.e., q_3, (pigenhole principle).
- Set:
  1. $x = s_1$,
  2. $y = s_2 \ldots s_5$, and
  3. $z = s_6 s_7$.

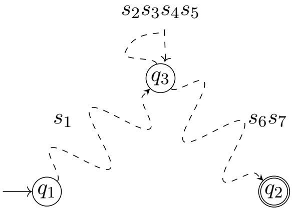
Nonregular Languages

Proof idea (cont.):

$$
\begin{array}{cccccccccccccccc}
s = & & s_1 & & s_2 & & s_3 & & s_4 & & s_5 & & s_6 & & s_7 & \\
& \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\
& q_1 & & \color{red}{q_3} & & q_2 & & q_4 & & q_5 & & \color{red}{q_3} & & q_5 & & q_2
\end{array}
$$

1. $x = s_1$,
2. $y = s_2 \ldots s_5$, and
3. $z = s_6 s_7$.

1. $|y| > 0$,
2. $|xy| \leq 5$, and
3. $xy^i z \in A$.



Proof.

- Let $M = (Q, \Sigma, \delta, q_q, F)$ be a DFA recognizing A and $p = |Q|$.
- Let $s = s_1 s_2 \dots s_n$ a string in A with $n \geq p$.
- Let $r_1, \dots, r_{n+1}$ be the sequence of states entered during processing of s, i.e., $\delta(r_i, s_i) = r_{i+1}$.
- $n + 1$ is at least $p + 1$. Thus, there is at least one repeated state (pigenhole principle).
- Let j and l be the first and second indices of the repeated state. Note that $l \leq p + 1$. Let,

$$x = s_1 \dots s_{j-1}, \quad y = s_j \dots s_{l-1}, \quad z = s_l \dots s_n. \tag{4}$$

- Thus,
  1. $r_j = r_l$ and $r_{n+1} \in F \quad \rightarrow \quad xy^i z \in A$,
  2. $j \neq l \quad \rightarrow \quad |y| > 0$, and
  3. $l \leq p + 1 \quad \rightarrow \quad |xy| \leq p$.

□

Nonregular Languages

- Proof B is not regular:
  - Assume B is regular,
  - Find string $s \in B$ such that for all divisions $s = xyz$ (respecting conditions in pumping lemma) there is i that $xy^iz \notin B$,
  - This contradicts with the pumping lemma,
  - Thus, B is not regular.
- Requires creativity!

Note

While the pumping lemma states that all regular languages satisfy the conditions described above, the converse of this statement is not true: a language that satisfies these conditions may still be non-regular.

Example

$B = \{0^n 1^n | n \geq 0\}$.

- Consider $s = 0^p 1^p$ for arbitrary p.
- Assume B is regular. Consider three cases for $s = xyz$:
  1. If y consists only of zeros, string $xyyz$ has more zeros than ones. Thus, it is not in B.
  2. If y consists only of ones, string $xyyz$ has more ones than zeros. Thus, it is not in B.
  3. If y consists of both zeros and ones, string $xyyz$ has correct number of zeros and ones but not in correct order.
- We considered all possible cases for y and concluded that xy^2z can not be in B. Therefore, the assumption of B being regular is not correct.

Example

$C = \{xx \mid x \in \{0,1\}^*\}$.

- Consider $s = 0^p 1^p 0^p 1^p$ for pumping length p.
- Let $s = xyz$, where
  1. $|xy| \leq p$, and
  2. $|y| > 0$.
- Thus, $x = 0^m$ and y^n where $m + n \leq p$,
- Thus, $z = 0^{p-m-n} 1^p 0^p 1^p$.
- However, $xy^0 z = 0^{p-m} 1^p 0^p 1^p$ which is not in C.
- Therefore, C can not be regular.

Nonregular Languages

Example

$C = \{xx^{\mathcal{R}} \mid x \in \{0,1\}^*\}$.

- Consider $s = 0^p1^p1^p0^p$ for pumping length p.
- Let $s = xyz$, where
  1. $|xy| \leq p$, and
  2. $|y| > 0$.
- Thus, $x = 0^m$ and y^n where $m + n \leq p$,
- Thus, $z = 0^{p-m-n}1^p1^p0^p$.
- However, $xy^0z = 0^{p-m}1^p0^p1^p$ which is not in C.
- Therefore, C can not be regular.

- Another way of proving that a language is not regular:
  - Assume that the given language is regular,
  - Apply an operator (maybe, together with another known regular language) that regular languages are closed under it:
    - Union
    - Concatenation
    - Star
    - Complement
    - Intersection
  - Obtain the resulting language,
  - Show that the resulting language is not regular,
  - This is a contradiction, which proves that the original assumption can not be true.

Example (Method 1: Pumping Lemma)

$C = \{w|w$ has an equal number of zeros and ones $\}$.

- Consider $s = 0^p1^p \in C$.
- For all $s = xyz$, applying $|xy| \leq p$, substring y consists only of zeros.
- Thus, $xyyz$ is not in C.

However, we know that:

- Regular languages are closed under intersection,
- 0^*1^* is regular, and
- $B = \{0^n 1^n | n \geq 0\}$ is not regular.

Example (Method 2: Closure Properties)

$C = \{w | w$ has an equal number of zeros and ones $\}$.

- Assume C is regular.
- So, $B = C \cap 0^*1^*$ should be regular, which is a contradiction.

Example (Not Repeated)

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

- Regular languages are closed under intersection,
- $(\Sigma\Sigma)^*$ is regular, and
- $C = \{xx \mid x \in \{0,1\}^*\}$ is not regular.
- Assume D is regular.
- So, $\overline{D} = \{xx \mid x \in \{0,1\}^*\} \cup \{y \mid |y|$ is odd $\}$[a] should be regular.
- Therefore, $\overline{D} \cap (\Sigma\Sigma)^*$[b] $= \{xx \mid x \in \{0,1\}^*\}$ should be regular.
- A contradiction.

[a] An odd-length string can not be in D
[b] Even-length strings

Example (A Pumpable Language)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

- Let the pumping length $p = 2$.
- We show for all $s \in E$ ($|s| \geq 2$) it is possible to:
  - Partition $s = xyz$, where $|xy| \leq 2$, and $|y| > 0$,
  - $xy^i z \in E$ for all $i \geq 0$.
- $k = 0$ and $m = 0$:
  $s = c^n = \varepsilon c c^{n-1} = xyz$. Then, $xy^i z = c^i c^{n-1} = c^{n+i-1} \in E$.
- $k = 0$ and $m > 0$:
  $s = b^m c^n = \varepsilon b b^{m-1} c^n = xyz$. Then,
  $xy^i z = b^i b^{m-1} c^n = b^{m+i-1} c^n \in E$.

Example (A Pumpable Language Cont.)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

- Let the pumping length $p = 2$.
- We show for all $s \in E$ ($|s| \geq 2$) it is possible to:
  - Partition $s = xyz$, where $|xy| \leq 2$, and $|y| > 0$,
  - $xy^iz \in E$ for all $i \geq 0$.
- $k = 1$ and $m = n$:
  $s = ab^nc^n = \varepsilon ab^nc^n = xyz$. Then, $xy^iz = a^ib^nc^n =\in E$.

Example (A Pumpable Language Cont.)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

- Let the pumping length $p = 2$.
- We show for all $s \in E$ ($|s| \geq 2$) it is possible to:
  - Partition $s = xyz$, where $|xy| \leq 2$, and $|y| > 0$,
  - $xy^i z \in E$ for all $i \geq 0$.
- $k = 2$:
  $s = aab^m c^n = \varepsilon aab^m c^n = xyz$. Then, $xy^i z = a^{2i} b^m c^n \in E$.[a]

[a] $2i$ is never equal to 1.

Example (A Pumpable Language Cont.)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

- Let the pumping length $p = 2$.
- We show for all $s \in E$ ($|s| \geq 2$) it is possible to:
  - Partition $s = xyz$, where $|xy| \leq 2$, and $|y| > 0$,
  - $xy^i z \in E$ for all $i \geq 0$.
- $k \geq 3$:
  $s = a^k b^m c^n = \varepsilon a a^{k-1} b^m c^n = xyz$. Then,
  $xy^i z = a^i a^{k-1} b^m c^n = a^{k+i-1} b^m c^n \in E$.[a]

[a] $k \geq 3 \rightarrow k + i - 1 \geq 2$ is never equal to 1.

Example (A Pumpable Language Cont.)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

- Let the pumping length $p = 2$.
- We show for all $s \in E$ ($|s| \geq 2$) it is possible to:
    - Partition $s = xyz$, where $|xy| \leq 2$, and $|y| > 0$,
    - $xy^i z \in E$ for all $i \geq 0$.
- Thus, all strings of length 2 and longer are pumpable.

Example (A Nonregular Lang.)

$E = \{a^k b^m c^n \mid \text{ if } k = 1, \text{ then } m = n\}$.

Assume E is regular:

- We know that ab^*c^* is regular,
- We know that regulars are closed under intersection,
- Thus, $F = \{ab^n c^n \mid n \geq 0\}$ should be regular with pumping length p.
- Let $s = ab^p c^p$ and all partitions $s = xyz$ with $|xy| \leq p$ and $|y| > 0$:
    - If a is in y, when xy^0z does not start with a!
    - If a is not in y, due to $|xy| \leq p$ and $|y| > 0$, then y only contains b. Thus, xy^0z does not have equal numbers of b and c.
- Thus, F is not pumpable and not regular, which is a contradiction.

Example

$G = \{0^m 1^n \mid m \neq n\}$.

- $\overline{G} \cap a^* b^* = \{0^n 1^n \mid n \geq 0\}$,
- We proved that this is not regular.

Example

$H = \{w \mid w \in \{0,1\}^*$ and w has an unequal number of 0s and 1s $\}$.

- $\overline{H} \cap a^* b^* = \{0^n 1^n \mid n \geq 0\}$,
- We proved that this is not regular.

Theorem

If L is a nonregular language, then $\overline{L}$ is not regular.

Proof.

- Proof by contradiction:
- Assume L is nonregular and $\overline{L}$ is regular,
- We know that class of regular languages is closed under complement,
- Thus, $\overline{(\overline{L})} = L$ should be regular,
- Which is a contradiction.

□

Theorem

If L is a nonregular language, then $L^{\mathcal{R}}$ is not regular.

Proof.

- Proof by contradiction:
- Assume L is nonregular and $L^{\mathcal{R}}$ is regular,
- We know that class of regular languages is closed under reversal,
- Thus, $(L^{\mathcal{R}})^{\mathcal{R}} = L$ should be regular,
- Which is a contradiction.

□

Theorem

The class of nonregular languages is not closed under union.

Proof.

We find two nonregular languages and show that their union is regular:

- $A = \{0^n 1^n \mid n \geq 0\}$ is not regular,
- We saw that nonregulars are closed under complement,
- Thus, $\overline{A}$ is also nonregular,
- However, $\overline{A} \cup A = \Sigma^*$ is regular.

□

Note: Not being closed means that the result of the operation may or may not be nonregular. In above, $A \cup A = A$ shows that the union of two nonregulars is still nonregular.

Theorem

The class of nonregular languages is not closed under intersection.

Theorem

The class of nonregular languages is not closed under Kleene star.