IN2080

4. Pumping Lemma for Regular Languages

Einar Broch Johnsen

University of Oslo

2nd February 2022



Department of
Informatics



University of
Oslo

Last lecture

- Regular expressions
- Regular expressions are shorthand notations for languages
- RE = GNFA = DFA = NFA
- Proof involved transforming a DFA to a GNFA, then reducing the number of states to 2 while accepting the same language
- $\rightarrow$ the regular expressions describe the paths in the DFA

- How can we prove that a given language L is regular?

- When is a language *nonregular*? How can we check that?
  $\Rightarrow$ Pumping Lemma!

Intuition for the Pumping Lemma

- Long words in regular languages must have repeating parts. Why is that?

Pigeonhole principle

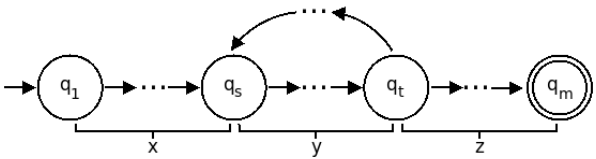


You want to place n pigeons in m pigeonholes. If $m < n$, then at least two items must be in the same hole.

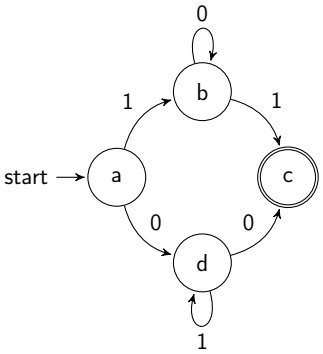
- A DFA has only *finite* memory, aka Q states.
- Assume that the DFA accepts a very long word...
- Then there must be an accepting sequence $s_0, s_1, \ldots s_n$ such that $\delta(s_i, w_{i+1}) = s_{i+1}$
- What happens if $i > |Q|$?

Pumping Lemma

- Method to recognize languages that are nonregular
- We know that RLs can be recognized by DFAs. These have a finite number of states.
- If a DFA with p states accepts a string w such that $|w| \geq p$, then we *must* visit some state(s) more than once in the accepting path.
- This means that there must be a *loop* in the DFA than can be repeated many times!

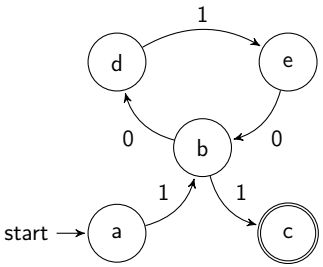


Pumping Lemma - Example



- Language $(10^*1) \cup (01^*0)$
- DFA has 4 states
- Consider the string 10001, length 5
- Is $10001 \in (10^*1) \cup (01^*0)$?
- Path to the accepting state: *a b b b b a*
  $\Rightarrow$ The path contains a loop!

Pumping Lemma - Example



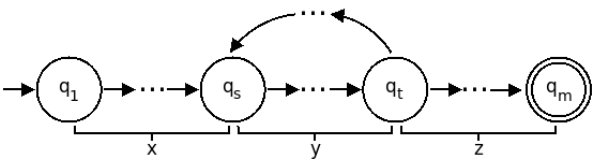
- Which language do we have here? $1(010)^*1$
- DFA has 5 states
- Now consider the string 10101, length 5
  Is $10101 \in 1(010)^*1$?
  $\Rightarrow$ The path to the accepting state must have length 6
  $\Rightarrow$ Then the path must contain a loop
  (in this case, with nodes b, d, e)
- Which word do we get if we repeat the loop two times?
  $\Rightarrow$ 10100101
- Which word do we get if we repeat the loop zero times?
  $\Rightarrow$ 11

Pumping Lemma

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

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



Pumping Lemma

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

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

Proof: We formalize our intuition.

- Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA, $A = L(M)$ the language accepted by M, and p be the number of states in M.
- Let $w = w_1 \cdots w_n$ be a word in A of length $n \geq p$. Since $w \in A$, it is accepted by M and there exists a sequence of states $s_0, s_1, \ldots s_n$ of length $n + 1$, where s_n is an accept state and $\delta(s_i, w_{i+1}) = s_{i+1}$.

Pumping Lemma

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

1. $xy^i z \in A$ *for every* $i \geq 0$,
2. $|y| > 0$,
3. $|xy| \leq p$.

- Let $w = w_1 \cdots w_n$ be a word in A of length $n \geq p$. Since $w \in A$, it is accepted by M and there exists an accepting sequence of states $s_0, s_1, \ldots s_n$ of length $n + 1$.
- Since $n + 1 \geq p + 1$, one state must occur twice within the first $p + 1$ elements of the accepting sequence (pigeonhole principle).
- Let these occurrences be s_j and s_l. Since these occur in the first $p + 1$ elements of the accepting sequence, we have $j < l \leq p + 1$.

Pumping Lemma

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

1. $xy^i z \in A$ *for every* $i \geq 0$,
2. $|y| > 0$,
3. $|xy| \leq p$.

- Let these occurrences be s_j and s_l. Since these occur in the first $p+1$ elements of the accepting sequence, we have $j < l \leq p+1$.
- Define $x = w_1 \cdots w_j, \quad y = w_{j+1} \cdots w_l, \quad z = w_{l+1} \cdots w_n$.
- Then $|y| > 0$ and $|xy| \leq p$.

Pumping Lemma

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

1. $xy^i z \in A$ *for every* $i \geq 0$,
2. $|y| > 0$,
3. $|xy| \leq p$.

- Define $x = w_1 \cdots w_{j-1}, \quad y = w_j \cdots w_{l-1}, \quad z = w_l \cdots w_n$.
- x takes M from s_1 to s_j, y takes M from s_j to s_l, z takes M from s_l to s_n.
- Thus the word $xy^i z$ takes M from the start state s_1 to s_j, follows the path from s_j to s_l i times (recall that $s_j = s_l$), then takes M from s_l to s_n.
- Thus M accepts any word $xy^i z$ for $i \geq 0$. □

Pumping Lemma

- Very useful to determine if a language is nonregular

  $\rightarrow$ Find a string with length $\geq$ p such that
  the pumping lemma does not hold

- *Not* very useful for proving a language is regular

  $\rightarrow$ Not an if and only if statement!
  $\rightarrow$ We need to find the counterexample!

Pumping Lemma - Applied

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

1. $xy^i z \in A$ *for every* $i \geq 0$,
2. $|y| > 0$,
3. $|xy| \leq p$.

- Let $A = \{0^n 1^n \mid n \geq 0\}$.
- Is A regular? We proceed by a proof by contradiction.
- *If A is regular*, then the pumping lemma gives us a pumping length p.
- Let $w = 0^p 1^p$.

Pumping Lemma - Applied

Lemma (Pumping Lemma)

If A is a regular language, then there is a number p, called the pumping length, where if w is a word in A of length $\geq p$ then w can be divided into three parts, $w = xyz$, such that

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

- Let $A = \{0^n1^n \mid n \geq 0\}$. Assume A regular, and apply thePL
- Let $w = 0^p1^p$.
- Need to divide w into parts x, y and z such that the PL conditions hold
- Condition 3 tells us that y consists of only 0s.
  $\Rightarrow$ Then xy^iz for $i \geq 2$ has more 0s than 1s. Contradiction! $\Rightarrow$ A cannot be regular!

Pumping Lemma - Applied

- Even if a language is nonregular, it might contain strings for which the pumping lemma is true!
- We have to be careful!

Pumping Lemma - Applied

Lemma (Pumping Lemma)

$|w| \geq p, w = xyz$, *s.t.*

1. $xy^i z \in A$ *for every* $i \geq 0$,
2. $|y| > 0$,
3. $|xy| \leq p$.

- Let $B = \{\omega \mid \omega \text{ contains an equal number of 0s and 1s}\}$. Assume B regular, apply PL.
- Let $w = (01)^p$.
- $x = \varepsilon, y = 01, z = (01)^{p-1}$
- All conditions are met!
- Did we just prove that B is regular?

Pumping Lemma - Applied

Lemma (Pumping Lemma)

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

- Let $B = \{\omega \mid \omega$ contains an equal number of 0s and 1s$\}$. Assume B regular, apply PL.
- Let $w = 0^p1^p$.
- $x = \varepsilon, y = 0^p1^p, z = \varepsilon$
- Looks like it can be pumped, but are all conditions met?
- Condition 3 $\Rightarrow$ y must contain only 0s, so it cannot be pumped

$\Rightarrow$ Contradiction, so B cannot be regular!

Pumping Lemma - Applied

- $A = \{0^n 1^n \mid n \geq 0\}$.
- $B = \{\omega \mid \omega$ contains an equal number of 0s and 1s$\}$

- Another way of showing B is nonregular is to reduce it to the nonregularity of A:
- Regular languages are closed under intersection
  and $A = B \cap 0^*1^*$

- If B is regular and since 0^*1^* is regular, then A must be as well.
- But we know that A is nonregular. Contradiction!

Summary

- Every regular language has a pumping length
- Pumping lemma is useful for determining if a language is *nonregular*
- Argument about non-regularity by contradiction!