

Theory of Computer Science

B6. Regular Languages: Pumping Lemma

Gabriele Röger

University of Basel

March 30, 2022

Theory of Computer Science

March 30, 2022 — B6. Regular Languages: Pumping Lemma

B6.1 Pumping Lemma

B6.1 Pumping Lemma

Pumping Lemma: Motivation



You can show that a language is regular by specifying an appropriate grammar, finite automaton, or regular expression. How can you show that a language is **not** regular?

- ▶ Direct proof that no regular grammar exists that generates the language
~~ difficult in general
- ▶ **Pumping lemma:** use a necessary property that holds for all regular languages.

Pumping Lemma

Theorem (Pumping Lemma)

If L is a regular language then there is a number $p \in \mathbb{N}$ (a *pumping number* for L) such that all words $x \in L$ with $|x| \geq p$ can be split into $x = uvw$ with the following properties:

- ① $|v| \geq 1$,
- ② $|uv| \leq p$, and
- ③ $uv^i w \in L$ for all $i = 0, 1, 2, \dots$

Question: what if L is finite?

Pumping Lemma: Proof

Theorem (Pumping Lemma)

If L is a regular language then there is a number $p \in \mathbb{N}$ (a *pumping number* for L) such that all words $x \in L$ with $|x| \geq p$ can be split into $x = uvw$ with the following properties:

- ① $|v| \geq 1$,
- ② $|uv| \leq p$, and
- ③ $uv^i w \in L$ for all $i = 0, 1, 2, \dots$

Proof.

For regular L there exists a DFA $M = \langle Q, \Sigma, \delta, q_0, E \rangle$ with $\mathcal{L}(M) = L$. We show that $p = |Q|$ has the desired properties.

Consider an arbitrary $x \in \mathcal{L}(M)$ with length $|x| \geq |Q|$. Including the start state, M visits $|x| + 1$ states while reading x . Because of $|x| \geq |Q|$ at least one state has to be visited twice. \dots

Pumping Lemma: Proof

Theorem (Pumping Lemma)

If L is a regular language then there is a number $p \in \mathbb{N}$ (a *pumping number* for L) such that all words $x \in L$ with $|x| \geq p$ can be split into $x = uvw$ with the following properties:

- ① $|v| \geq 1$,
- ② $|uv| \leq p$, and
- ③ $uv^i w \in L$ for all $i = 0, 1, 2, \dots$

Proof (continued).

Choose a split $x = uvw$ so M is in the same state after reading u and after reading uv . Obviously, we can choose the split in a way that $|v| \geq 1$ and $|uv| \leq |Q|$ are satisfied. \dots

Pumping Lemma: Proof

Theorem (Pumping Lemma)

If L is a regular language then there is a number $p \in \mathbb{N}$ (a *pumping number* for L) such that all words $x \in L$ with $|x| \geq p$ can be split into $x = uvw$ with the following properties:

- ① $|v| \geq 1$,
- ② $|uv| \leq p$, and
- ③ $uv^i w \in L$ for all $i = 0, 1, 2, \dots$

Proof (continued).

The word v corresponds to a loop in the DFA after reading u and can thus be repeated arbitrarily often. Every subsequent continuation with w ends in the same end state as reading x . Therefore $uv^i w \in \mathcal{L}(M) = L$ is satisfied for all $i = 0, 1, 2, \dots$ \square

Pumping Lemma: Application

Using the pumping lemma (PL):

Proof of Nonregularity

- ▶ If L is regular, then the pumping lemma holds for L .
- ▶ By contraposition: if the PL does **not** hold for L , then L **cannot** be regular.
- ▶ That is: if there is no $p \in \mathbb{N}$ with the properties of the PL, then L cannot be regular.

Pumping Lemma: Example

Example

The language $L = \{a^n b^n \mid n \in \mathbb{N}\}$ is not regular.

Proof.

Assume L is regular. Then let p be a pumping number for L .

The word $x = a^p b^p$ is in L and has length $\geq p$.

Let $x = uvw$ be a split with the properties of the PL.

Then the word $x' = uv^2w$ is also in L . Since $|uv| \leq p$, uv consists only of symbols a and $x' = a^{|u|} a^{2|v|} a^{p-|uv|} b^p = a^{p+|v|} b^p$.

Since $|v| \geq 1$ it follows that $p + |v| \neq p$ and thus $x' \notin L$.

This is a contradiction to the PL. $\leadsto L$ is not regular. □

Pumping Lemma: Caveat

Caveat:

The pumping lemma is a **necessary condition** for a language to be regular, but not a **sufficient one**.

- ⇝ there are languages that satisfy the pumping lemma conditions but are **not** regular
- ⇝ for such languages, other methods are needed to show that they are not regular (e.g., the [Myhill-Nerode theorem](#))

Pumping Lemma: Another Example I

Example

The language $L = \{ab^n ac^{n+2} \mid n \in \mathbb{N}\}$ is not regular.

Proof.

Assume L is regular. Then let p be a pumping number for L .

The word $x = ab^p ac^{p+2}$ is in L and has length $\geq p$.

Let $x = uvw$ be a split with the properties of the PL.

From $|uv| \leq p$ and $|v| \geq 1$ we know that uv consists of one a followed by at most $p-1$ b s.

We distinguish two cases, $|u| = 0$ and $|u| > 0$

Pumping Lemma: Another Example II

Example

The language $L = \{ab^nac^{n+2} \mid n \in \mathbb{N}\}$ is not regular.

Proof (continued).

If $|u| = 0$, then word v starts with an a.

Hence, $uv^0w = b^{p-|v|+1}ac^{p+2}$ does not start with symbol a and is therefore not in L . This is a contradiction to the PL.

If $|u| > 0$, then word v consists only of bs.

Consider $uv^0w = ab^{p-|v|}ac^{p+2}$. As $|v| \geq 1$, this word does not contain two more cs than bs and is therefore not in language L . This is a contradiction to the PL.

We have in all cases a contradiction to the PL.

$\rightsquigarrow L$ is not regular.

□

Pumping Lemma: Exercise



This was an exam question in 2020:

Use the pumping lemma to prove that $L = \{a^m b^n \mid m \geq 0, n < m\}$ is not regular.

Summary

- ▶ The **pumping lemma** can be used to show that a language is **not regular**.