CSCI 3434: Theory of Computation
Lecture 5: Pumping Lemma

Ashutosh Trivedi

Department of Computer Science
University of Colorado Boulder
Find a DFA for the following languages:

- The set of strings having an equal number of 0’s and 1’s
- The set of strings with an equal number of occurrences of 01 and 10.
Some languages are not regular!

Let’s do mental computations again.

- The language $\{0^n1^n : n \geq 0\}$
- The set of strings having an equal number of 0’s and 1’s
- The language $\{ww : w \in \{0, 1\}^*\}$
- The language $\{w\overline{w} : w \in \{0, 1\}^*\}$
- The language $\{0^i1^j : i > j\}$
- The language $\{0^i1^j : i \leq j\}$
- The language of palindromes of $\{0, 1\}$
Some languages are not regular!

Let’s do mental computations again.

- The language \( \{0^n1^n : n \geq 0\} \)
- The set of strings having an equal number of 0’s and 1’s
- The language \( \{ww : w \in \{0, 1\}^*\} \)
- The language \( \{w\overline{w} : w \in \{0, 1\}^*\} \)
- The language \( \{0^i1^j : i > j\} \)
- The language \( \{0^i1^j : i \leq j\} \)
- The language of palindromes of \( \{0, 1\} \)

How do we prove that a language is not regular?
Theorem (Pumping Lemma for Regular Languages)

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that for every string $w \in L$ of length greater than $p$, there exists an infinite family of strings belonging to $L$.
**Pumping Lemma**

**Theorem (Pumping Lemma for Regular Languages)**

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that

for every string $w \in L$ of length greater than $p$,

there exists an infinite family of strings belonging to $L$. 

Why?

Think: Regular expressions, DFAs

Formalize our intuition!
Theorem (Pumping Lemma for Regular Languages)

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that for every string $w \in L$ of length greater than $p$, there exists an infinite family of strings belonging to $L$.

Why?
Theorem (Pumping Lemma for Regular Languages)

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that for every string $w \in L$ of length greater than $p$, there exists an infinite family of strings belonging to $L$.

Why? Think: Regular expressions, DFAs
Theorem (Pumping Lemma for Regular Languages)

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that
for every string $w \in L$ of length greater than $p$,
there exists an infinite family of strings belonging to $L$.

Why? Think: Regular expressions, DFAs Formalize our intuition!
Theorem (Pumping Lemma for Regular Languages)

For every regular language $L$ there exists a constant $p$ (that depends on $L$) such that for every string $w \in L$ of length greater than $p$, there exists an infinite family of strings belonging to $L$.

Why? Think: Regular expressions, DFAs Formalize our intuition!

If $L$ is a regular language, then there exists a constant (pumping length) $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, \text{ and } z$ s.t. $w = xyz$ such that

1. $|y| > 0,$
2. $|xy| \leq p,$ and
3. for all $i \geq 0$ we have that $xy^iz \in L.$
A simple observation about DFA

![Diagram of a DFA with states E and O, transitions labeled 0 and 1, and a computation shown for different strings.]

- **Computation**
  - Start: E
  - After 0: E
  - After 1: O
  - After 0: E

- **String**
  - 0:
  - 1:
  - 0:
  - 1:

- **Computation**
  - Start: E
  - After 0: E
  - After 1: O
  - After 0: E

- **String**
  - 0:
  - 1:
  - 0:
  - 1:
A simple observation about DFA

Let $A = (S, \Sigma, \delta, s_0, F)$ be a DFA.

For every string $w \in \Sigma^*$ of the length greater than or equal to the number of states of $A$, i.e. $|w| \geq |S|$, we have that

- the unique computation of $A$ on $w$ re-visits at least one state.
Pumping Lemma: Proof

Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, \text{ and } z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$. 

Proof.

– Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.

– Let $w = (a_1a_2...a_k) \in L$ be any string of length $\geq p$.

– Let $s_0a_1\rightarrow s_1a_2\rightarrow s_2...a_k\rightarrow s_k$ be the run of $w$ on $A$.

– Let $i$ be the index of first state that the run revisits and let $j$ be the index of second occurrence of that state, i.e. $s_i = s_j$,

– Let $x = a_1a_2...a_i$ and $y = a_i+1...a_j$, and $z = a_j+1...a_k$.

– notice that $|y| > 0$ and $|xy| \leq n$.

– Also, notice that for all $i \geq 0$ the string $xy^iz$ is also in $L$. 

Pumping Lemma: Proof

Theorem (Pumping Lemma for Regular Languages)
If L is a regular language, then there exists a constant p such that for every string
w ∈ L s.t. |w| ≥ p there exists a division of w in strings x, y, and z s.t. w = xyz
such that |y| > 0, |xy| ≤ p, and for all i ≥ 0 we have that xy^iz ∈ L.

Proof.

– Let A be the DFA accepting L and p be the set of states in A.
Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, \text{and } z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$.

Proof.

- Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.
- Let $w = (a_1a_2\ldots a_k) \in L$ be any string of length $\geq p$. 

Pumping Lemma: Proof

### Theorem (Pumping Lemma for Regular Languages)

If L is a regular language, then there exists a constant p such that for every string w ∈ L s.t. |w| ≥ p there exists a division of w in strings x, y, and z s.t. w = xyz such that |y| > 0, |xy| ≤ p, and for all i ≥ 0 we have that xy^iz ∈ L.

### Proof.

- Let A be the DFA accepting L and p be the set of states in A.
- Let w = (a_1a_2...a_k) ∈ L be any string of length ≥ p.
- Let s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 ... \xrightarrow{a_k} s_k be the run of w on A.
Pumping Lemma: Proof

Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$.

Proof.

- Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.
- Let $w = (a_1a_2\ldots a_k) \in L$ be any string of length $\geq p$.
- Let $s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \ldots \xrightarrow{a_k} s_k$ be the run of $w$ on $A$.
- Let $i$ be the index of first state that the run revisits and let $j$ be the index of second occurrence of that state, i.e. $s_i = s_j$. 
Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, \text{ and } z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$.

Proof.

- Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.
- Let $w = (a_1a_2 \ldots a_k) \in L$ be any string of length $\geq p$.
- Let $s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \ldots \xrightarrow{a_k} s_k$ be the run of $w$ on $A$.
- Let $i$ be the index of first state that the run revisits and let $j$ be the index of second occurrence of that state, i.e. $s_i = s_j$.
- Let $x = a_1a_2 \ldots a_i$ and $y = a_{i+1} \ldots a_j$, and $z = a_{j+1} \ldots a_k$. 
Pumping Lemma: Proof

Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y, \text{ and } z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$.

Proof.

- Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.
- Let $w = (a_1a_2\ldots a_k) \in L$ be any string of length $\geq p$.
- Let $s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \ldots \xrightarrow{a_k} s_k$ be the run of $w$ on $A$.
- Let $i$ be the index of first state that the run revisits and let $j$ be the index of second occurrence of that state, i.e. $s_i = s_j$.
- Let $x = a_1a_2\ldots a_i$ and $y = a_{i+1}\ldots a_j$, and $z = a_{j+1}\ldots a_k$.
- notice that $|y| > 0$ and $|xy| \leq n$
Pumping Lemma: Proof

Theorem (Pumping Lemma for Regular Languages)

If $L$ is a regular language, then there exists a constant $p$ such that for every string $w \in L$ s.t. $|w| \geq p$ there exists a division of $w$ in strings $x, y,$ and $z$ s.t. $w = xyz$ such that $|y| > 0$, $|xy| \leq p$, and for all $i \geq 0$ we have that $xy^iz \in L$.

Proof.

- Let $A$ be the DFA accepting $L$ and $p$ be the set of states in $A$.
- Let $w = (a_1a_2 \ldots a_k) \in L$ be any string of length $\geq p$.
- Let $s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \ldots \xrightarrow{a_k} s_k$ be the run of $w$ on $A$.
- Let $i$ be the index of first state that the run revisits and let $j$ be the index of second occurrence of that state, i.e. $s_i = s_j$.
- Let $x = a_1a_2 \ldots a_i$ and $y = a_{i+1} \ldots a_j$, and $z = a_{j+1} \ldots a_k$.
- notice that $|y| > 0$ and $|xy| \leq n$
- Also, notice that for all $i \geq 0$ the string $xy^iz$ is also in $L$. 
Theorem (Pumping Lemma for Regular Languages)

$L \in \Sigma^*$ is a regular language

$\implies$

there exists $p \geq 1$ such that

for all strings $w \in L$ with $|w| \geq p$ we have that

there exists $x, y, z \in \Sigma^*$ with $w = xyz$, $|y| > 0$, $|xy| \leq p$ such that

for all $i \geq 0$ we have that

$xy^iz \in L$. 
Applying Pumping Lemma

Theorem (Pumping Lemma for Regular Languages)

$L \in \Sigma^*$ is a regular language

$\implies$

there exists $p \geq 1$ such that

for all strings $w \in L$ with $|w| \geq p$ we have that

there exists $x, y, z \in \Sigma^*$ with $w = xyz$, $|y| > 0$, $|xy| \leq p$ such that

for all $i \geq 0$ we have that

$xy^iz \in L$.

Pumping Lemma (Contrapositive)

For all $p \geq 1$ we have that

there exists a string $w \in L$ with $|w| \geq p$ such that

for all $x, y, z \in \Sigma^*$ with $w = xyz$, $|y| > 0$, $|xy| \leq p$ we have that

there exists $i \geq 0$ such that

$xy^iz \notin L$

$\implies$

$L \in \Sigma^*$ is not a regular language.
Applying Pumping Lemma

Pumping Lemma (Contrapositive)

For all $p \geq 1$ we have that there exists a string $w \in L$ with $|w| \geq p$ such that for all $x, y, z \in \Sigma^*$ with $w = xyz$, $|y| > 0$, $|xy| \leq p$ we have that there exists $i \geq 0$ such that $xy^iz \not\in L$ implies $L \in \Sigma^*$ is not a regular language.

How to show that a language $L$ is non-regular.

1. Let $p$ be an arbitrary number (pumping length).
2. (Cleverly) Find a representative string $w$ of $L$ of size $\geq p$.
3. Try out all ways to break the string into $xyz$ triplet satisfying that $|y| > 0$ and $|xy| \leq n$. If the step 3 was clever enough, there will be finitely many cases to consider.
4. For every triplet show that for some $i$ the string $xy^iz$ is not in $L$, and hence it yields contradiction with pumping lemma.
Applying Pumping Lemma I

**Theorem**

Prove that the language \( L = \{0^n1^n\} \) is not regular.
Applying Pumping Lemma I

**Theorem**

Prove that the language $L = \{0^n1^n\}$ is not regular.

**Proof.**

1. State the contrapositive of Pumping lemma.
2. Let $p$ be an arbitrary number.
3. Consider the string $0^p1^p \in L$. Notice that $|0^p1^p| \geq p$. 

Hence $L$ is non-regular.
### Theorem

*Prove that the language $L = \{0^n1^n\}$ is not regular.*

### Proof.

1. State the contrapositive of Pumping lemma.
2. Let $p$ be an arbitrary number.
3. Consider the string $0^p1^p \in L$. Notice that $|0^p1^p| \geq p$.
4. Only way to break this string in $xyz$ triplets such that $|xy| \leq p$ and $y \neq \varepsilon$ is to choose $y = 0^k$ for some $1 \leq k \leq p$. 

Applying Pumping Lemma I

Theorem

Prove that the language \( L = \{0^n1^n\} \) is not regular.

Proof.

1. State the contrapositive of Pumping lemma.
2. Let \( p \) be an arbitrary number.
3. Consider the string \( 0^p1^p \in L \). Notice that \( |0^p1^p| \geq p \).
4. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
5. For each such triplet, there exists an \( i \) (say \( i = 0 \)) such that \( xy^iz \not\in L \).
6. Hence \( L \) is non-regular.
## Applying Pumping Lemma II

**Theorem**

Prove that the language

\[ M = \{ w : w \text{ has an equal number of 0 and 1} \} \]

is not regular.
Applying Pumping Lemma II

Theorem

Prove that the language

\[ M = \{ w : \text{w has an equal number of 0 and 1} \} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^p \in L \). Notice that \( |0^p1^p| \geq p \).
Applying Pumping Lemma II

Theorem

Prove that the language

\[ M = \{ w : \text{w has an equal number of 0 and 1} \} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^p \in L \). Notice that \( |0^p1^p| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
Applying Pumping Lemma II

**Theorem**

Prove that the language

\[ M = \{ w : w \text{ has an equal number of } 0 \text{ and } 1 \} \]

is not regular.

**Proof.**

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^p \in L \). Notice that \( |0^p1^p| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \epsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
4. For each such triplet, there exists an \( i \) (say \( i = 0 \)) such that \( xy^iz \notin L \).
5. Hence \( L \) is non-regular.
Applying Pumping Lemma III

Theorem

Prove that the language

\[ M = \{1^{n^2} : n \geq 0\} \]

is not regular.
Applying Pumping Lemma III

**Theorem**

*Prove that the language*

\[ M = \{ 1^{n^2} : n \geq 0 \} \]

*is not regular.*

**Proof.**

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 1^p^2 \in L \). Notice that \( |1^p^2| \geq p \).
Applying Pumping Lemma III

Theorem

Prove that the language

\[ M = \{1^{n^2} : n \geq 0\} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 1^p^2 \in L \). Notice that \( |1^p^2| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 1^k \) for some \( 1 \leq k \leq p \).
Theorem

Prove that the language

\[ M = \{1^n^2 : n \geq 0\} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 1^p^2 \in L \). Notice that \( |1^p^2| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 1^k \) for some \( 1 \leq k \leq p \).
4. Now consider \( 1^l1^k1^k1^p^2-l+k \) (pumping twice) and show that it is not perfect square.
5. Hence \( L \) is non-regular.
Theorem

Prove that the language

\[ M = \{0^i1^j : i > j\} \]

is not regular.
Applying Pumping Lemma IV

Theorem

Prove that the language

\[ M = \{ 0^i 1^j : i > j \} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p 1^{p+1} \in L \). Notice that \( |0^p 1^{p+1}| \geq p \).
Theorem

Prove that the language

\[ M = \{0^i1^j : i > j\} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^{p+1} \in L \). Notice that \( |0^p1^{p+1}| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
Applying Pumping Lemma IV

**Theorem**

Prove that the language

\[ M = \{0^i1^j : i > j\} \]

is not regular.

**Proof.**

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^{p+1} \in L \). Notice that \(|0^p1^{p+1}| \geq p\).
3. Only way to break this string in \( xyz \) triplets such that \(|xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
4. All pumping-ups are in the language!
Applying Pumping Lemma IV

Theorem

Prove that the language

\[ M = \{0^i1^j : i > j\} \]

is not regular.

Proof.

1. Let \( p \) be an arbitrary number.
2. Consider the string \( 0^p1^{p+1} \in L \). Notice that \( |0^p1^{p+1}| \geq p \).
3. Only way to break this string in \( xyz \) triplets such that \( |xy| \leq p \) and \( y \neq \varepsilon \) is to choose \( y = 0^k \) for some \( 1 \leq k \leq p \).
4. All pumping-ups are in the language!
5. Solution: pump-down.
6. Hence \( L \) is non-regular.
Proving Regularity

Pumping Lemma is necessary but not sufficient condition for regularity.

Consider the language

$L = \{ \#a^n b^n : n \geq 1 \} \cup \{ \#k w : k \neq 1, w \in \{a, b\}^* \}$

Verify that this language satisfies the pumping condition, but is not regular!
Proving Regularity

Pumping Lemma is necessary but not sufficient condition for regularity.

Consider the language

\[ L = \{ \#a^n b^n : n \geq 1 \} \cup \{ \#^k w : k \neq 1, w \in \{a, b\}^* \} \].

Verify that this language satisfies the pumping condition, but is not regular!