Foundations of Computing Lecture 6

Arkady Yerukhimovich

February 1, 2024

Outline

- 1 Lecture 5 Review
- 2 A Non-regular Language
- The Pumping Lemma for Regular Languages
- Using the Pumping Lemma

Lecture 5 Review

- Regular expressions
- Equivalence of regular expressions and NFAs/DFAs

Outline

- Lecture 5 Review
- 2 A Non-regular Language
- The Pumping Lemma for Regular Languages
- Using the Pumping Lemma

What We Know So Far

The following four things are equivalent:

- Regular languages
- 2 Languages recognized by a DFA
- Standard Languages recognized by an NFA
- Languages described by a regular expression

What We Know So Far

The following four things are equivalent:

- Regular languages
- 2 Languages recognized by a DFA
- Languages recognized by an NFA
- Languages described by a regular expression

Are all languages regular?

Today we will see that there are languages that are not regular

Important

In a Finite Automaton, the number of states is finite

Important

In a Finite Automaton, the number of states is finite

This means that:

• The number of states is fixed independently of the input size

Important

In a Finite Automaton, the number of states is finite

This means that:

- The number of states is fixed independently of the input size
- ullet An automaton must be able to process strings w s.t. |w|>|Q|

Important

In a Finite Automaton, the number of states is finite

This means that:

- The number of states is fixed independently of the input size
- ullet An automaton must be able to process strings w s.t. |w|>|Q|
- Thus, a finite automaton cannot store its whole input

A Nonregular Language

Consider the following language:

$$L = \{0^n 1^n | n \ge 0\}$$

Let's try to build a DFA for L:

A Nonregular Language

Consider the following language:

$$L = \{0^n 1^n | n \ge 0\}$$

Let's try to build a DFA for L:

The Problem

We need to count the number of 0s, but this is unbounded so can't have a state for each value

The Need for a Proof

What we just saw

Intuition: An NFA/DFA cannot count unbounded inputs

The Need for a Proof

What we just saw

Intuition: An NFA/DFA cannot count unbounded inputs

Why isn't this a proof?

The Need for a Proof

What we just saw

Intuition: An NFA/DFA cannot count unbounded inputs

Why isn't this a proof?

Consider the following language:

 $L = \{w|w \text{ has an equal number of occurrences of 01 and 10 as substrings}\}$

We will prove that a language L is not regular by contradiction

lacktriangle Assume L is regular – there is a NFA/DFA M accepting it

- **1** Assume L is regular there is a NFA/DFA M accepting it
- 2 Pick a string $w \in L$

- lacktriangle Assume L is regular there is a NFA/DFA M accepting it
- ② Pick a string $w \in L$
- **3** Show that if M(w) = 1 then there exists a string $w' \notin L$ s.t. M(w') = 1

- lacktriangle Assume L is regular there is a NFA/DFA M accepting it
- ② Pick a string $w \in L$
- ③ Show that if M(w) = 1 then there exists a string $w' \notin L$ s.t. M(w') = 1
- Onclude that L is not regular since any M that accepts all strings in L must also accept strings not in L

Outline

- 1 Lecture 5 Review
- 2 A Non-regular Language
- 3 The Pumping Lemma for Regular Languages
- Using the Pumping Lemma

Pumping Lemma

If L is a regular language, then there exists an integer p (the pumping length) where any string $w \in L$ such that $|w| \ge p$ can be divided into three pieces w = xyz satisfying:

Pumping Lemma

If L is a regular language, then there exists an integer p (the pumping length) where any string $w \in L$ such that $|w| \ge p$ can be divided into three pieces w = xyz satisfying:

• For each $i \ge 0$, $xy^i z \in L$

Pumping Lemma

If L is a regular language, then there exists an integer p (the pumping length) where any string $w \in L$ such that $|w| \ge p$ can be divided into three pieces w = xyz satisfying:

- For each $i \ge 0$, $xy^i z \in L$
- ② |y| > 0, and

Pumping Lemma

If L is a regular language, then there exists an integer p (the pumping length) where any string $w \in L$ such that $|w| \ge p$ can be divided into three pieces w = xyz satisfying:

- For each $i \ge 0$, $xy^i z \in L$
- ② |y| > 0, and
- $|xy| \leq p$

Pumping Lemma

If L is a regular language, then there exists an integer p (the pumping length) where any string $w \in L$ such that $|w| \ge p$ can be divided into three pieces w = xyz satisfying:

- For each $i \ge 0$, $xy^iz \in L$
- ② |y| > 0, and
- $|xy| \leq p$

Next steps:

- Prove the pumping lemma
- 2 Show how to use the pumping lemma to prove languages nonregular

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA that recognizes L and let p = |Q|

Let $M=(Q,\Sigma,\delta,q_1,F)$ be a DFA that recognizes L and let p=|Q|

• If for all $w \in L$, |w| < p, then we are done

Let $M=(Q,\Sigma,\delta,q_1,F)$ be a DFA that recognizes L and let p=|Q|

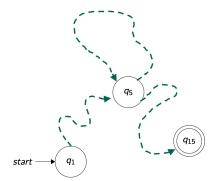
- If for all $w \in L$, |w| < p, then we are done
- Suppose $w \in L$ s.t. $|w| = n \ge p$

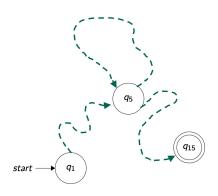
Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA that recognizes L and let p = |Q|

- If for all $w \in L$, |w| < p, then we are done
- Suppose $w \in L$ s.t. $|w| = n \ge p$
 - If we run M(w), it will go through n+1 states since it transitions on each symbol of w

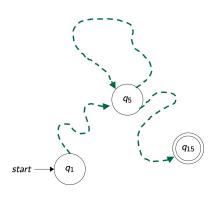
Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA that recognizes L and let p = |Q|

- If for all $w \in L$, |w| < p, then we are done
- Suppose $w \in L$ s.t. $|w| = n \ge p$
 - If we run M(w), it will go through n+1 states since it transitions on each symbol of w
 - Since n + 1 > p, there must be some state that is visited twice



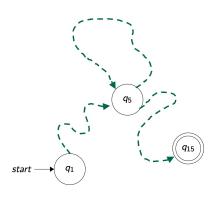


Divide w = xyz as follows:



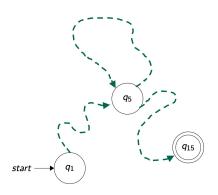
Divide w = xyz as follows:

• x is the part of w until M(w) visits q_5



Divide w = xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5



Divide w = xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

We show that the properties of the pumping lemma hold:

• For each $i \ge 0$, $xy^i z \in L$

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

We show that the properties of the pumping lemma hold:

• For each $i \ge 0$, $xy^iz \in L$ Proof: y takes M from q_5 back to q_5 . So, if you run M(xyyz), it would just run this cycle twice...

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

- For each $i \ge 0$, $xy^iz \in L$ Proof: y takes M from q_5 back to q_5 . So, if you run M(xyyz), it would just run this cycle twice...
- **2** |y| > 0

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

- For each $i \ge 0$, $xy^iz \in L$ Proof: y takes M from q_5 back to q_5 . So, if you run M(xyyz), it would just run this cycle twice...
- 2 |y| > 0Proof: \geq one character of w must be between the two visits of q_5

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

- For each $i \ge 0$, $xy^iz \in L$ Proof: y takes M from q_5 back to q_5 . So, if you run M(xyyz), it would just run this cycle twice...
- ② |y| > 0Proof: \geq one character of w must be between the two visits of q_5
- $|xy| \leq p$

We divided w into xyz as follows:

- x is the part of w until M(w) visits q_5
- y is the part of w between the two visits to q_5
- z is the part of w after the second visit to q_5

- For each $i \ge 0$, $xy^iz \in L$ Proof: y takes M from q_5 back to q_5 . So, if you run M(xyyz), it would just run this cycle twice...
- ② |y| > 0Proof: \geq one character of w must be between the two visits of q_5
- ③ $|xy| \le p$ Proof: if q_5 is the first repetition in M(w), then this repetition must occur in the first p+1 states, hence $|xy| \le p$

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA recognizing L

• Let $w = w_1 w_2 \dots w_n \in L$

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, ..., r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, \ldots, r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$
- In first p+1 elements of this sequence there must be a repeated element $r_j = r_k$, so $k \le p+1$

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, \ldots, r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$
- In first p+1 elements of this sequence there must be a repeated element $r_j = r_k$, so $k \le p+1$
- Let $x = w_1 \cdots w_{i-1}$, $y = w_i \cdots w_{k-1}$, and $z = w_k \cdots w_n$

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA recognizing L

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, \ldots, r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$
- In first p+1 elements of this sequence there must be a repeated element $r_i = r_k$, so $k \le p+1$
- Let $x = w_1 \cdots w_{i-1}$, $y = w_i \cdots w_{k-1}$, and $z = w_k \cdots w_n$

Observe that:

• x takes M from $r_1=q_1$ to r_j , y takes M from r_j to r_k , and z takes M from r_k to r_{n+1} , which is an accept state. So, M must accept xy^iz for $i\geq 0$

Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA recognizing L

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, \ldots, r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$
- In first p+1 elements of this sequence there must be a repeated element $r_i = r_k$, so $k \le p+1$
- Let $x = w_1 \cdots w_{i-1}$, $y = w_i \cdots w_{k-1}$, and $z = w_k \cdots w_n$

Observe that:

- x takes M from $r_1=q_1$ to r_j , y takes M from r_j to r_k , and z takes M from r_k to r_{n+1} , which is an accept state. So, M must accept xy^iz for $i \geq 0$
- $j \neq k$ so, |y| > 0



Let $M = (Q, \Sigma, \delta, q_1, F)$ be a DFA recognizing L

- Let $w = w_1 w_2 \dots w_n \in L$
- Let $r_1, r_2, \ldots, r_{n+1}$ be the sequence of states visited by M(w) $(r_{i+1} = \delta(r_i, w_i))$
- In first p+1 elements of this sequence there must be a repeated element $r_i = r_k$, so $k \le p+1$
- Let $x = w_1 \cdots w_{i-1}$, $y = w_i \cdots w_{k-1}$, and $z = w_k \cdots w_n$

Observe that:

- x takes M from $r_1=q_1$ to r_j , y takes M from r_j to r_k , and z takes M from r_k to r_{n+1} , which is an accept state. So, M must accept xy^iz for $i \geq 0$
- $j \neq k$ so, |y| > 0
- $k \le p+1$, so $|xy| \le p$



Outline

- 1 Lecture 5 Review
- 2 A Non-regular Language
- 3 The Pumping Lemma for Regular Languages
- 4 Using the Pumping Lemma

To use the pumping lemma to prove that L is not regular, we do the following:

Assume that L is regular

- Assume that L is regular
- ② By pumping lemma there exists pumping length p, s.t. all w with |w|>p can be pumped

- Assume that L is regular
- ② By pumping lemma there exists pumping length p, s.t. all w with |w|>p can be pumped
- **3** Choose a particular $w \in L$ with $|w| \ge p$

- Assume that L is regular
- ② By pumping lemma there exists pumping length p, s.t. all w with |w|>p can be pumped
- **3** Choose a particular $w \in L$ with $|w| \ge p$
- Demonstrate that w cannot be pumped:
 - For each possible division w = xyz, find an i such that $xy^iz \notin L$

- Assume that L is regular
- ② By pumping lemma there exists pumping length p, s.t. all w with |w|>p can be pumped
- **3** Choose a particular $w \in L$ with $|w| \ge p$
- \odot Demonstrate that w cannot be pumped:
 - For each possible division w = xyz, find an i such that $xy^iz \notin L$
- Contradiction!!!

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

Proof:

lacktriangle Assume L is regular, and let p be the pumping length this implies

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- ② Choose $w = 0^p 1^p$

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- **2** Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$
- lacktriangle Complete proof by considering all possible values for y

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$
- Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$
- Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$
 - y consists of only 1s then xyyz has more 0s than 1s, so $w \notin L$

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$
- Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$
 - y consists of only 1s then xyyz has more 0s than 1s, so $w \notin L$
 - y consists of both 0s and 1s then xyyz has 0s alternating with 1s more than once, so $w \notin L$

Consider $L = \{0^n 1^n | n \ge 0\}$, prove L is not regular.

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w can be split into xyz s.t. $xy^iz \in L$
- Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$
 - y consists of only 1s then xyyz has more 0s than 1s, so $w \notin L$
 - y consists of both 0s and 1s then xyyz has 0s alternating with 1s more than once, so $w \notin L$
- ⑤ Contradiction hence, L is not regular



Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

Proof:

lacksquare Assume L is regular, and let p be the pumping length this implies

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- ② Choose $w = 0^p 1^p$

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- **1** Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- **1** Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction
- **5** Solution: Use condition that $|xy| \le p$

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- **1** Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction
- **5** Solution: Use condition that $|xy| \le p$
 - Since $w = 0^p 1^p$ and $|xy| \le p$, we know that y must be in first p symbols

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- **1** Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction
- **5** Solution: Use condition that $|xy| \le p$
 - Since $w = 0^p 1^p$ and $|xy| \le p$, we know that y must be in first p symbols
 - But, this means that y must be all 0s

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- **1** Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction
- **5** Solution: Use condition that $|xy| \le p$
 - Since $w = 0^p 1^p$ and $|xy| \le p$, we know that y must be in first p symbols
 - But, this means that y must be all 0s
- Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- lacktriangle Assume L is regular, and let p be the pumping length this implies
- 2 Choose $w = 0^p 1^p$
- **3** By pumping lemma, w = xyz s.t. $xy^iz \in L$
- Problem: If $y = 0^m 1^m$, then w can be pumped no contradiction
- **5** Solution: Use condition that $|xy| \le p$
 - Since $w = 0^p 1^p$ and $|xy| \le p$, we know that y must be in first p symbols
 - But, this means that y must be all 0s
- \odot Complete proof by considering all possible values for y
 - y consists of only 0s then xyyz has more 0s than 1s, so $w \notin L$
- O Contradiction hence, L is not regular



Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

A simpler proof:

lacksquare We already proved that $L_1=\{0^n1^n|n\geq 0\}$ is nonregular

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- ① We already proved that $L_1 = \{0^n 1^n | n \ge 0\}$ is nonregular
- ② Observe that $L_1 = L \cap 0^*1^*$

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- **①** We already proved that $L_1 = \{0^n 1^n | n \ge 0\}$ is nonregular
- ② Observe that $L_1 = L \cap 0^*1^*$
- Easy to see that 0*1* is regular
- **3** Since regular languages are closed under \cap , if L is regular then L_1 must be regular

Consider $L = \{w | w \text{ has an equal number of 0s and 1s} \}$, prove L is not regular

- **①** We already proved that $L_1 = \{0^n 1^n | n \ge 0\}$ is nonregular
- ② Observe that $L_1 = L \cap 0^*1^*$
- Easy to see that 0*1* is regular
- **3** Since regular languages are closed under \cap , if L is regular then L_1 must be regular
- lacktriangle Since we know L_1 is nonregular, this means that L must be nonregular

Exercise

Prove that the following language is nonregular:

$$L = \{0^{i}1^{j}2^{i}3^{j}|i,j>0\}$$

What's Next?

- We will get plenty of practice with proving languages nonregular
- We will add (a small amount of) memory to our machines to recognize a richer class of languages