CM0081 Formal Languages and Automata
§ 4.1 Proving Languages Not to Be Regular

Andrés Sicard-Ramírez

Universidad EAFIT

Semester 2024-1

Preliminaries

Conventions

- The number and page numbers assigned to chapters, examples, exercises, figures, quotes, sections and theorems on these slides correspond to the numbers assigned in the textbook [Hopcroft, Motwani and Ullman (1979) 2007].
- The natural numbers include the zero, that is, $\mathbb{N} = \{0, 1, 2, \dots\}$.
- The power set of a set A, that is, the set of its subsets, is denoted by $\mathcal{P}A$.

Properties of Regular Languages

- Proving languages not to be regular
- Closure properties
- Decision properties
- Equivalence and minimization of automata

The Pumping Lemma

Introduction

- Is $L_1 = \{\, 0^m 1^n \mid m, n \geq 0 \,\}$ a regular language?

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^* \mathbf{1}^*)$.

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^*\mathbf{1}^*)$.
- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^* \mathbf{1}^*)$.
- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

  Yes! $L_2 = \mathrm{L}(\mathbf{0}^+ \mathbf{1}^+)$.

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^*\mathbf{1}^*)$.

- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

  Yes! $L_2 = \mathrm{L}(\mathbf{0}^+\mathbf{1}^+)$.

- Is $L_3 = \{ 0^m 1^n \mid m \geq 2, n \geq 4 \}$ a regular language?

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^* \mathbf{1}^*)$.

- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

  Yes! $L_2 = \mathrm{L}(\mathbf{0}^+ \mathbf{1}^+)$.

- Is $L_3 = \{ 0^m 1^n \mid m \geq 2, n \geq 4 \}$ a regular language?

  Yes! $L_3 = \mathrm{L}(\mathbf{000}^* \mathbf{11111}^*)$.

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^* \mathbf{1}^*)$.

- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

  Yes! $L_2 = \mathrm{L}(\mathbf{0}^+ \mathbf{1}^+)$.

- Is $L_3 = \{ 0^m 1^n \mid m \geq 2, n \geq 4 \}$ a regular language?

  Yes! $L_3 = \mathrm{L}(\mathbf{000}^* \mathbf{11111}^*)$.

- Is $L_4 = \{ 0^n 1^n \mid n \geq 1 \}$ a regular language?

The Pumping Lemma

Introduction

- Is $L_1 = \{ 0^m 1^n \mid m, n \geq 0 \}$ a regular language?

  Yes! $L_1 = \mathrm{L}(\mathbf{0}^* \mathbf{1}^*)$.

- Is $L_2 = \{ 0^m 1^n \mid m, n \geq 1 \}$ a regular language?

  Yes! $L_2 = \mathrm{L}(\mathbf{0}^+ \mathbf{1}^+)$.

- Is $L_3 = \{ 0^m 1^n \mid m \geq 2, n \geq 4 \}$ a regular language?

  Yes! $L_3 = \mathrm{L}(\mathbf{000}^* \mathbf{11111}^*)$.

- Is $L_4 = \{ 0^n 1^n \mid n \geq 1 \}$ a regular language?

  No! Informal proof (whiteboard).

The Pumping Lemma

Theorem 4.1 (Pumping Lemma for regular languages)

Let L be a regular language. Then there exists a positive integer n (which depends on L) such that for every string $w \in L$ such that $|w| \geq n$, we can break w into three strings, $w = xyz$, such that:

$$y \neq \varepsilon, \tag{1}$$

$$|xy| \leq n, \tag{2}$$

$$(\forall k \geq 0)(xy^k z \in L). \tag{3}$$

Formally,

$$(\exists n \in \mathbb{Z}^+)(\forall w \in L)(|w| \geq n \Rightarrow (\exists x)(\exists y)(\exists z)[w = xyz \wedge (1) \wedge (2) \wedge (3)]).$$

The Pumping Lemma

Proof

1. Suppose L is a regular language. Exist a DFA $A = (Q, \Sigma, \delta, q_0, F)$ with n states such that $\mathrm{L}(A) = L$.

The Pumping Lemma

Proof

1. Suppose L is a regular language. Exist a DFA $A = (Q, \Sigma, \delta, q_0, F)$ with n states such that $\mathrm{L}(A) = L$.
2. Let $w = a_1 \cdots a_m \in L$, $m \geq n$ and $q_i = \hat{\delta}(q_0, a_1 \cdots a_i)$.

The Pumping Lemma

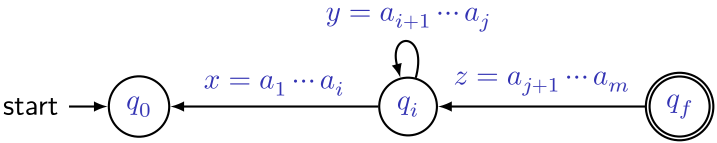
Proof

1. Suppose L is a regular language. Exist a DFA $A = (Q, \Sigma, \delta, q_0, F)$ with n states such that $\mathrm{L}(A) = L$.
2. Let $w = a_1 \cdots a_m \in L$, $m \geq n$ and $q_i = \hat{\delta}(q_0, a_1 \cdots a_i)$.
3. By the pigeonhole principle, exists i and j, with $0 \leq i < j \leq n$ such that $q_i = q_j$.

The Pumping Lemma

Proof

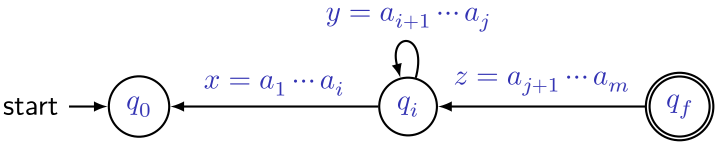
1. Suppose L is a regular language. Exist a DFA $A = (Q, \Sigma, \delta, q_0, F)$ with n states such that $\mathrm{L}(A) = L$.
2. Let $w = a_1 \cdots a_m \in L$, $m \geq n$ and $q_i = \hat{\delta}(q_0, a_1 \cdots a_i)$.
3. By the pigeonhole principle, exists i and j, with $0 \leq i < j \leq n$ such that $q_i = q_j$.
4. Let $w = xyz$ where



The Pumping Lemma

Proof

1. Suppose L is a regular language. Exist a DFA $A = (Q, \Sigma, \delta, q_0, F)$ with n states such that $\mathrm{L}(A) = L$.
2. Let $w = a_1 \cdots a_m \in L$, $m \geq n$ and $q_i = \hat{\delta}(q_0, a_1 \cdots a_i)$.
3. By the pigeonhole principle, exists i and j, with $0 \leq i < j \leq n$ such that $q_i = q_j$.
4. Let $w = xyz$ where



5. Then $(\forall k \geq 0)(xy^k z \in L)$. ■

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof schemata

Whiteboard.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Exercise 4.1.2.e

Let $\Sigma = \{0, 1\}$ be an alphabet and let $L = \{\, ww \mid w \in \Sigma^* \,\}$ be the so-called copy language. Prove that L is not regular.

(continued on next slide)

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^n 1 0^n 1 \in L$ and $|w| \geq n$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^n 1 0^n 1 \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^n 1 0^n 1 \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^n 1 0^n 1 \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$.
6. But, $xy^0 z \notin L$ which contradicts the Pumping Lemma.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^n 1 0^n 1 \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$.
6. But, $xy^0 z \notin L$ which contradicts the Pumping Lemma.
7. Therefore, L is not regular. ∎

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Exercise 4.1.2.a

Let L be the language

$$L = \{\, 0^n \mid n \text{ is a perfect square} \,\}.$$

Prove that L is not regular.

(continued on next slide)

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$ and $n^2 + 1 \leq |xyyz| \leq n^2 + n$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$ and $n^2 + 1 \leq |xyyz| \leq n^2 + n$.
6. Since the next perfect square after n^2 is $(n+1)^2 = n^2 + 2n + 1$, we know that $xyyz \notin L$ because $|xyyz|$ is strictly between the consecutive perfect squares n^2 and $(n+1)^2$.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$ and $n^2 + 1 \leq |xyyz| \leq n^2 + n$.
6. Since the next perfect square after n^2 is $(n+1)^2 = n^2 + 2n + 1$, we know that $xyyz \notin L$ because $|xyyz|$ is strictly between the consecutive perfect squares n^2 and $(n+1)^2$.
7. This a contradiction with the Pumping Lemma.

Application of the Pumping Lemma: Proving Languages Not to Be Regular

Proof

1. Suppose L is regular.
2. Let $n \in \mathbb{Z}^+$ be a constant (according to the Pumping Lemma).
3. Let $w = 0^{n^2} \in L$ and $|w| \geq n$.
4. For the Pumping Lemma should exist x, y and z such that $w = xyz$, $|xy| \leq n$, $y \neq \varepsilon$ and $(\forall k \geq 0)(xy^k z \in L)$.
5. For any x, y and z such that $w = xyz$, $|xy| \leq n$ and $y \neq \varepsilon$, we have that $y = 0^m$ with $0 < m \leq n$ and $n^2 + 1 \leq |xyyz| \leq n^2 + n$.
6. Since the next perfect square after n^2 is $(n+1)^2 = n^2 + 2n + 1$, we know that $xyyz \notin L$ because $|xyyz|$ is strictly between the consecutive perfect squares n^2 and $(n+1)^2$.
7. This a contradiction with the Pumping Lemma.
8. Therefore, L is not regular. ■

Other Methods for Proving Languages Not to Be Regular

Observation

Frishberg and Gasarch [2018] show other methods and some open problems when proving that a language is not regular. The open problem 3.2 is related to the pumping lemma.

Open problem

'Find a non-regular language that cannot be proven non-regular using the pumping theorem and reductions, or show such a language does not exist.'

References

Frishberg, D. and Gasarch, W. (2018). Open Problems Column. Different Ways to Prove a Language is Not Regular. SIGACT News 49.1, pp. 40–54. DOI: 10.1145/3197406.3197413 (cit. on p. 36).

Hopcroft, J. E., Motwani, R. and Ullman, J. D. [1979] (2007). Introduction to Automata Theory, Languages, and Computation. 3rd ed. Pearson Education (cit. on p. 2).