CS420

Introduction to the Theory of Computation

Lecture 6: The pumping lemma; non-regular languages

Tiago Cogumbreiro

Today we will learn...

- Regular languages
- The pumping lemma
- Non-regular languages
- Proving that a language is not regular with the Pumping lemma

> Section 1.4 Nonregular Languages
> Today's lecture is based on the excellent Prof. Emanuele Viola's slides.

What is a regular language?

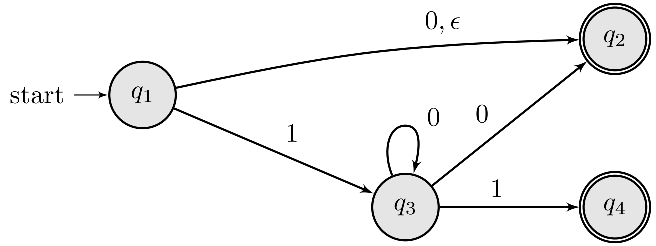
What is a regular language?

Definition 1.16

We say that L_1 is regular if there exists a DFA M such that $L(M) = L_1$.

Example 1

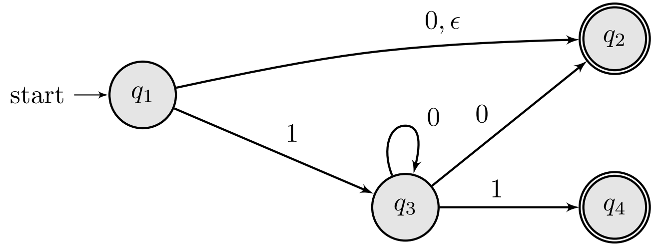
Let N_1 be the following NFA:



Is $L(N_1)$ regular?

Example 1

Let N_1 be the following NFA:

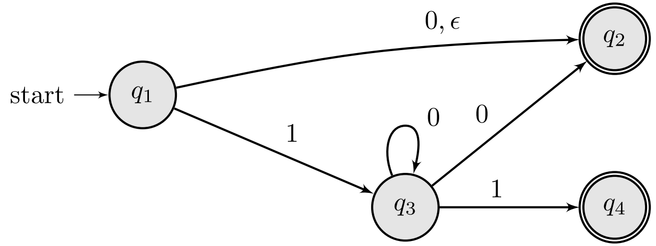


Is $L(N_1)$ regular?

Yes. **Proof:** we can convert N_1 into an equivalent DFA, which then satisfies Definition 1.16.

Example 1

Let N_1 be the following NFA:



Is $L(N_1)$ regular?

Yes. **Proof:** we can convert N_1 into an equivalent DFA, which then satisfies Definition 1.16.

Theorem

> We say that L_1 is regular, if there exits an NFA N such that $L(N) = L_1$

Example 2

Is $L(0 + 1^\star)$ regular?

Example 2

Is $L(0 + 1^\star)$ regular?

Yes. **Proof:** We have that $L(0 + 1^\star) = L(\mathrm{NFA}(0 + 1^\star))$, which is regular (from the previous theorem).

Example 2

Is $L(0 + 1^\star)$ regular?

Yes. **Proof:** We have that $L(0 + 1^\star) = L(\mathrm{NFA}(0 + 1^\star))$, which is regular (from the previous theorem).

Theorem

> We say that L_1 is regular, if there exits a regular expression R such that $L(R) = L_1$

What is a regular language?

1. A language is regular if there exists a DFA that recognizes it
2. A language is regular if there exists an NFA that recognizes it
3. A language is regular if there exists a Regex that recognizes it

Example

The language of strings that have a possibly empty sequence of n zeroes followed by a sequence of n ones.

$$B = \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Is this language *regular*?

Example

The language of strings that have a possibly empty sequence of n zeroes followed by a sequence of n ones.

$$B = \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Is this language *regular*?

How do we prove that a language is *not* regular?

Example

The language of strings that have a possibly empty sequence of n zeroes followed by a sequence of n ones.

$$B = \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Is this language *regular*?

How do we prove that a language is *not* regular?

The only way we know is by proving that there is no NFA/DFA/regex that can recognize such a language.

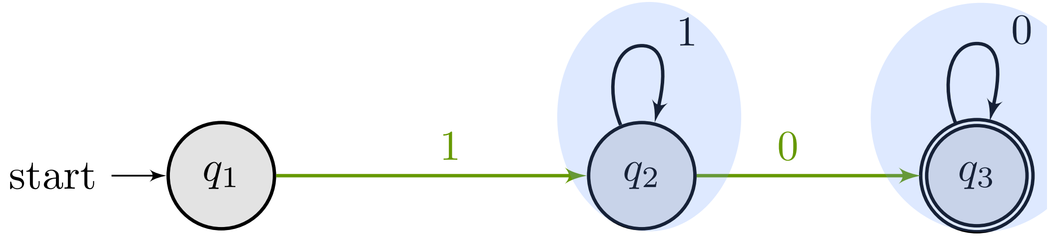
Pumping lemma

An intuition

> The pumping lemma tells us that **all** regular languages (that have a loop) have the following characteristics:

Every word in a regular language, $w \in L$, can be partitioned into three parts $w = xyz$:

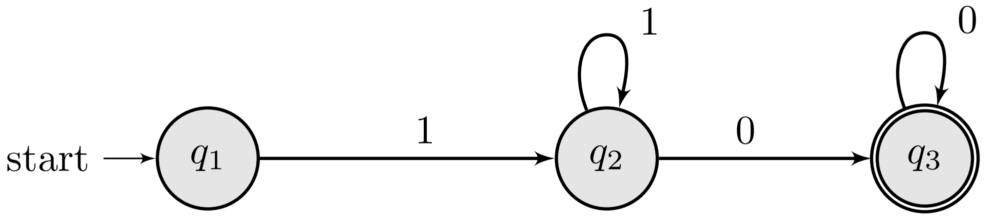
- a portion x before the first loop,
- a portion y that is one loop's iteration (nonempty), and
- a portion z that follows the first loop



Additionally, since y is a loop, then it may be omitted or replicated as many times as we want and that word will also be in the given language, that is $xy^iz \in L$

Pumping lemma

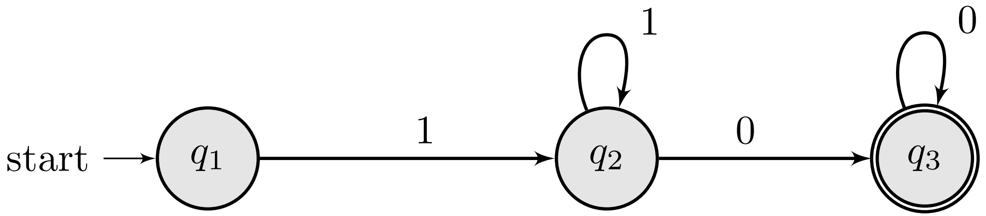
Pictorial intuition



You: Give me any string accepted by the automaton of at least size 3.

Pumping lemma

Pictorial intuition

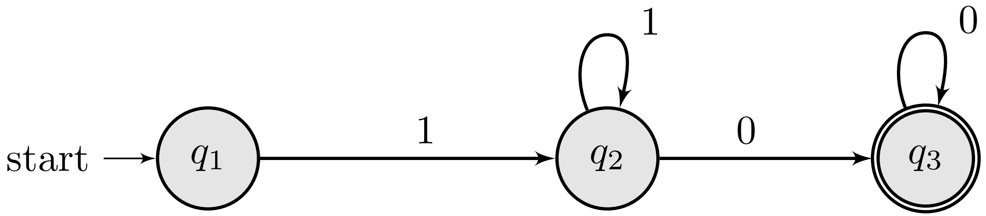


You: Give me any string accepted by the automaton of at least size 3.

Example: 100

Pumping lemma

Pictorial intuition



You: Give me any string accepted by the automaton of at least size 3.

Example: 100

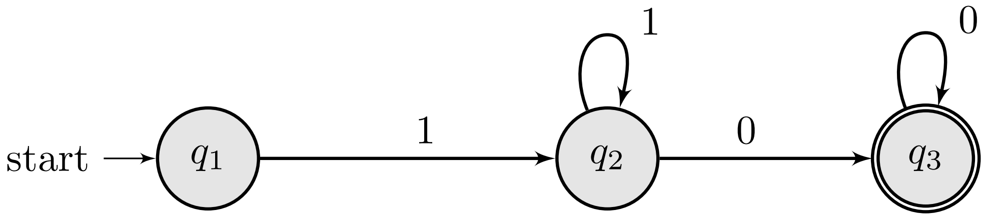
Me: I will partition 100 into three parts $100 = xyz$ such that $xy^i z$ is accepted for any i:

$$\underbrace{10}_{x}\underbrace{0}_{y}\underbrace{\epsilon}_{z}$$

- $xz = 10 \cdot \epsilon = 10$ is accepted
- $x\underline{yy}z = 10\underline{00}$ is accepted
- $x\underline{yyyy}z = 10\underline{0000}$ is accepted
- $x\underline{yyyyyy}z = 10\underline{000000}$ is accepted

Pumping lemma

Pictorial intuition



You: Give me a string of **size 4**.

Example: 1100

Me: I will partition 1100 into three parts $1100 = xyz$ such that $xy^i z$ is accepted for any i:

$$\underbrace{1}_{x}\underbrace{1}_{y}\underbrace{00}_{z}$$

- $xz = 100$ is accepted
- $x\underline{yy}z = 1\underline{11}00$ is accepted
- $x\underline{yyyy}z = 1\underline{1111}00$ is accepted
- $x\underline{yyyyyy}z = 1\underline{111111}00$ is accepted

Pumping lemma

If A is a regular language, then there exists a *pumping length* (a number) where if $s \in A$ and $|s| \geq p$, then there exist x, y, z such that

1. $s = xyz$
2. $\forall i \colon i \geq 0$ we have that $xy^i z \in A$
3. $|y| > 0$
4. $|xy| \leq p$

Nonregular languages

Recall the contrapositive

If $P \implies Q$, then $\neg Q \implies \neg P$

```
Theorem contrapositive:
  forall P Q: Prop, (P → Q) → (~Q → ~P).
Proof.
```

Recall the contrapositive

If $P \implies Q$, then $\neg Q \implies \neg P$

```
Theorem contrapositive:
  forall P Q: Prop, (P → Q) → (~Q → ~P).
Proof.

  intros.         (* introduce assumptions. *)
  unfold not.     (* open the definition of not, P → False *)
  intros.         (* introduce assumption P *)
  apply H in H1.  (* We have P→Q and P apply the former to the latter. *)
  contradiction.  (* We have Q and ~Q, so we reach a contradiction. *)
Qed.
```

Feel free to iterate through the proof using CoqIDE:

coq.inria.fr

Pumping Lemma and not regular languages

From the Pumping lemma we have

$$A \text{ is regular} \implies \exists p, \text{Pumping}(p, A)$$

Then, by the contrapositive, we have:

Pumping Lemma and not regular languages

From the Pumping lemma we have

$$A \text{ is regular} \implies \exists p, \text{Pumping}(p, A)$$

Then, by the contrapositive, we have:

$$\neg\left(\exists p, \text{Pumping}(p, A)\right) \implies \neg A \text{ is regular}$$

Pumping Lemma and not regular languages

From the Pumping lemma we have

$$A \text{ is regular} \implies \exists p, \text{Pumping}(p, A)$$

Then, by the contrapositive, we have:

$$\neg(\exists p, \text{Pumping}(p, A)) \implies \neg A \text{ is regular}$$

Thus,

$$\forall p, \neg\text{Pumping}(p, A) \implies A \text{ is regular} \implies \bot$$

Pumping Lemma and not regular languages

From the Pumping lemma we have

$$A \text{ is regular} \implies \exists p, \text{Pumping}(p, A)$$

Then, by the contrapositive, we have:

$$\neg(\exists p, \text{Pumping}(p, A)) \implies \neg A \text{ is regular}$$

Thus,

$$\forall p, \neg\text{Pumping}(p, A) \implies A \text{ is regular} \implies \perp$$

In other words, if we have $\neg\text{Pumping}(p, A)$ (next slide) and A is regular, then we can reach a contradiction.

Pumping Lemma and not regular languages

$\neg\text{Pumping}(p, A)$ and A is regular $\implies$ $\bot$ can be written as follows:

$H_0 : \forall p : p \geq 0$

$H_1 : \exists w : w \in A$ such that $|w| \geq p$

$H_2 : \forall x, y, z : w = xyz$ where $|y| > 0$ and $|xy| \leq p$

$H_3 : \exists i : i \geq 0$

$H_4 :$ A is regular

Goal: $\bot$

Pumping Lemma and not regular languages

$\neg\text{Pumping}(p, A)$ and A is regular $\implies$ $\bot$ can be written as follows:

$H_0 : \forall p : p \geq 0$

$H_1 : \exists w : w \in A$ such that $|w| \geq p$

$H_2 : \forall x, y, z : w = xyz$ where $|y| > 0$ and $|xy| \leq p$

$H_3 : \exists i : i \geq 0$

$H_4 :$ A is regular

Goal: $\bot$

Proof strategy

Proving that a language A is nonregular involves using the $\forall$ and $\exists$ quantifiers.

Proving can be seen as a game, concluding a proof means winning the game.

- The $\forall$ quantifier is picked by your adversary
- The $\exists$ quantifier is picked by you (the player)

Proof example (with existential)

Theorem: For any number, there exists a another number that is greater than the given number.

$H_0 : \forall a \colon a \geq 0$

Goal $\exists b \colon b > a$

- $\forall$: Your adversary can pick any number, including the biggest number they can think of
- $\exists$: But, because we can pick another number, by knowing what number was given we can just answer the successor

Proof. Pick $a + 1$.

```
Goal
  forall a, exists b,
  b > a.
Proof.
```

Proof example (with existential)

Theorem: For any number, there exists a another number that is greater than the given number.

$H_0 : \forall a \colon a \geq 0$

Goal $\exists b \colon b > a$

- $\forall$: Your adversary can pick any number, including the biggest number they can think of
- $\exists$: But, because we can pick another number, by knowing what number was given we can just answer the successor

Proof. Pick $a + 1$.

```
Goal
  forall a, exists b,
  b > a.
Proof.

  intros.
  exists (1 + a).
  auto with *.
Qed.
```

Proving that a language is not regular

1. **Adversary** picks p such that:
   $p \geq 0$
2. **You** pick some w so that:
   $w \in A$ and $|w| \geq p$
3. **Adversary** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. **You** pick some i such that:
   $i \geq 0$
5. **Goal: You** show that $xy^i z \notin A$

Tips

- **The accepted word:** usually that words has an exponent, in which case use the pumping length
- **How many times y repeats:** usually 0 or 2

$\{0^n 1^n \mid \forall n\colon n \geq 0\}$ is nonregular

Proving nonregular languages

Theorem $\{0^n 1^n \mid \forall n\colon n \geq 0\}$ is not regular.

Proof idea

1. **Adversary:** picks p such that $p \geq 0$

Proving nonregular languages

Theorem $\{0^n 1^n \mid \forall n \colon n \geq 0\}$ is not regular.

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^p 1^p$
   $w \in A$ and $|w| \geq p$ (trivially holds)

Proving nonregular languages

Theorem $\{0^n 1^n \mid \forall n\colon n \geq 0\}$ is not regular.

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^p 1^p$
   $w \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$

Proving nonregular languages

Theorem $\{0^n 1^n \mid \forall n\colon n \geq 0\}$ is not regular.

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^p 1^p$
   $w \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. You: Let us pick $i = 2$:
   $i \geq 0$ (trivially holds)

Proving nonregular languages

Theorem $\{0^n 1^n \mid \forall n\colon n \geq 0\}$ is not regular.

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^p 1^p$
   $w \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. You: Let us pick $i = 2$:
   $i \geq 0$ (trivially holds)
5. **Goal:** You: show that $xyyz \notin A$

Why?

- The final goal is to show that $w \notin A$; thus, to show that the exponent of 1 is different than the exponent of 0.
- By picking p as the exponent, we force the exponent of 1 to contain at least $|xy|$, meaning that z will be fixed.
- By selecting $i = 2$ we make the exponent of 1 bigger than that of 0.

Theorem $:_1 = \{0^n 1^n \mid \forall n \colon n \geq 0\}$ is not regular.

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^p 1^p$ and must show that
   - $w \in \{0^n 1^n \mid \forall n \colon n \geq 0\}$, which holds by replacing n by p.
   - $|w| \geq p$, which holds since $|w| = 2p \geq p$.

Theorem $:_1 = \{0^n 1^n \mid \forall n \colon n \geq 0\}$ is not regular.

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^p 1^p$ and must show that
   - $w \in \{0^n 1^n \mid \forall n \colon n \geq 0\}$, which holds by replacing n by p.
   - $|w| \geq p$, which holds since $|w| = 2p \geq p$.
2. Finally, given some x, y, z our assumptions are (H1) $w = xyz$, (H2) $|xy| \leq p$, and (H3) $|y| > 0$, we must prove that

$$\exists i, xy^i z \notin L_1$$

(We write in red what you need to prove)

Proof. (Continuation...)

Let $a + b = p$, where $xy = 0^a$ and $a, b \in \mathcal{N}_0$ (non-negative).

We can rewrite (H1) $w = xyz$ such that

$$(H_1) \quad w = \underbrace{0^p 1^p}_{xyz} = \underbrace{0^a}_{xy} \underbrace{0^b 1^{a+b}}_{z}$$

Proof. (Continuation...)

Let $a + b = p$, where $xy = 0^a$ and $a, b \in \mathcal{N}_0$ (non-negative).

We can rewrite (H1) $w = xyz$ such that

$$(H_1) \qquad w = \underbrace{0^p 1^p}_{xyz} = \underbrace{0^a}_{xy} \underbrace{0^b 1^{a+b}}_{z}$$

Or, simply,

$$(H_1) \qquad \underbrace{0^a}_{xy} \underbrace{0^b 1^{a+b}}_{z} = \underbrace{0^{|xy|}}_{xy} \underbrace{0^b 1^{|xy|+b}}_{z}$$

Proof. (Continuation...) We pick $i = 2$, so our goal is to show that

$$\underbrace{0^{|xyy|}}_{xyy}\underbrace{0^{b}1^{|xy|+b}}_{z} \notin \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Proof. (Continuation...) We pick $i = 2$, so our goal is to show that

$$\underbrace{0^{|xyy|}}_{xyy}\underbrace{0^b 1^{|xy|+b}}_{z} \notin \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Thus, it is equivalent to show that

$$|xyy| + b \neq |xy| + b$$

We can simplify it with,

Proof. (Continuation...) We pick $i = 2$, so our goal is to show that

$$\underbrace{0^{|xyy|}}_{xyy}\underbrace{0^b 1^{|xy|+b}}_{z} \notin \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Thus, it is equivalent to show that

$$|xyy| + b \neq |xy| + b$$

We can simplify it with,

$$|xyy| + b - (|xy| + b) \neq |xy| + b - (|xy| + b)$$

And,

$$|y| \neq 0$$

Proof. (Continuation...) We pick $i = 2$, so our goal is to show that

$$\underbrace{0^{|xyy|}}_{xyy}\underbrace{0^b 1^{|xy|+b}}_{z} \notin \{0^n 1^n \mid \forall n \colon n \geq 0\}$$

Thus, it is equivalent to show that

$$|xyy| + b \neq |xy| + b$$

We can simplify it with,

$$|xyy| + b - (|xy| + b) \neq |xy| + b - (|xy| + b)$$

And,

$$|y| \neq 0$$

Which is trivially true since (H3) $|y| > 0$

$\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Theorem $\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$

Theorem $\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick the same w as before
   $0^p 1^p \in A$ and $|w| \geq p$ (trivially holds)

Theorem $\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick the same w as before
   $0^p 1^p \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$

Theorem $\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick the same w as before
   $0^p 1^p \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. **You:** Let us pick $i = 2$:
   $i \geq 0$ (trivially holds)

Theorem $\{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick the same w as before
   $0^p 1^p \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. **You:** Let us pick $i = 2$:
   $i \geq 0$ (trivially holds)
5. **Goal:** **You:** show that $xyyz \notin A$

Why?

- We are responsible for picking w, which is the hardest part of the problem.
- By picking $0^p 1^p$, we replicate the proof we did in the previous exercise!

Theorem $L_2 = \{w \mid w \text{ has as many 0's as 1's}\}$ is not regular

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^p 1^p$ and must show that
   - $w \in L_2$, which holds since there are p 0's and p 1's.
   - $|w| \geq p$, which holds since $|w| = 2p \geq p$.
2. Finally, given some x, y, z our assumptions are (H1) $w = xyz$, (H2) $|xy| \leq p$, and (H3) $|y| > 0$, we must prove that

$$\exists i, xy^i z \notin L_2$$

(We write in red what you need to prove)

Proof. (Continuation...)

Let $p = a + b$ and $|xy| = a$. We pick $i = 2$ and show that

$$\underbrace{0^a}_{xy}\underbrace{0^{|y|}}_{y}\underbrace{0^b 1^{a+b}}_{z} \notin \{w \mid \forall n\colon n \text{ has as many 0's as 1's}\}$$

Proof. (Continuation...)

Let $p = a + b$ and $|xy| = a$. We pick $i = 2$ and show that

$$\underbrace{0^a}_{xy}\underbrace{0^{|y|}}_{y}\underbrace{0^b 1^{a+b}}_{z} \notin \{w \mid \forall n\colon n \text{ has as many 0's as 1's}\}$$

The goal below is equivalent:

$$a + |y| + b \neq a + b$$

Proof. (Continuation...)

Let $p = a + b$ and $|xy| = a$. We pick $i = 2$ and show that

$$\underbrace{0^a}_{xy}\underbrace{0^{|y|}}_{y}\underbrace{0^b 1^{a+b}}_{z} \notin \{w \mid \forall n\colon n \text{ has as many 0's as 1's}\}$$

The goal below is equivalent:

$$a + |y| + b \neq a + b$$

And can be simplified to

$$|y| \neq 0$$

Proof. (Continuation...)

Let $p = a + b$ and $|xy| = a$. We pick $i = 2$ and show that

$$\underbrace{0^a}_{xy}\ \underbrace{0^{|y|}}_{y}\ \underbrace{0^b 1^{a+b}}_{z} \notin \{w \mid \forall n\colon n \text{ has as many 0's as 1's}\}$$

The goal below is equivalent:

$$a + |y| + b \neq a + b$$

And can be simplified to

$$|y| \neq 0$$

Which is given by the hypothesis that $|y| > 0$.

$\{0^j 1^k \mid j > k\}$ is not regular

Theorem: $A = \{0^j 1^k \mid j > k\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$

Theorem: $A = \{0^j 1^k \mid j > k\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^{p+1}1^p$
   $0^{p+1}1^p \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$

Theorem: $A = \{0^j 1^k \mid j > k\}$ is not regular

Proof idea

1. **Adversary:** picks p such that $p \geq 0$
2. **You:** Let us pick $w = 0^{p+1}1^p$
   $0^{p+1}1^p \in A$ and $|w| \geq p$ (trivially holds)
3. **Adversary:** decomposes w in xyz such that:
   $|y| > 0$ and $|xy| \leq p$
4. **You:** Let us pick $i = 0$:
   $i \geq 0$ (trivially holds)
5. **Goal: You:** show that $xz \notin A$

Why?

- Ultimately, our goal is to show that $w \notin A$, thus that the exponent of 1 smaller or equal than the exponent of 0.
- Since the loop always appears on the left-hand side of the string, we should pick the smallest exponent possible that uses p and still $w \in A$. Thus, we pick $0^{p+1}1^p$.

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^{p+1}1^p \in A$. Let $|xy| + b = p$. We have $|xy| \leq p$ and that $w = 0^{p+1}1^p$.

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^{p+1}1^p \in A$. Let $|xy| + b = p$. We have $|xy| \leq p$ and that $w = 0^{p+1}1^p$.
2. We pick $i = 0$ and show that

$$xz \notin \{0^j1^k \mid j > k\}$$

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^{p+1}1^p \in A$. Let $|xy| + b = p$. We have $|xy| \leq p$ and that $w = 0^{p+1}1^p$.
2. We pick $i = 0$ and show that

$$xz \notin \{0^j 1^k \mid j > k\}$$

3. Thus,

$$0^{|xy|-|y|+b+1} 1^{|xy|+b} \notin \{0^j 1^k \mid j > k\}$$

Proof. We prove that the language above does not satisfy the pumping property, thus the language is not regular. Let p be the pumping length.

1. We pick $w = 0^{p+1}1^p \in A$. Let $|xy| + b = p$. We have $|xy| \leq p$ and that $w = 0^{p+1}1^p$.
2. We pick $i = 0$ and show that

$$xz \notin \{0^j 1^k \mid j > k\}$$

3. Thus,

$$0^{|xy|-|y|+b+1}1^{|xy|+b} \notin \{0^j 1^k \mid j > k\}$$

4. So, we have to show that

$$\begin{gathered} |xy| - |y| + b + 1 \leq |xy| + b \\ |x| + 1 \leq |xy| \\ |y| \geq 1 \quad \text{which holds, since} |y| > 0 \end{gathered}$$