Tianjin International Engineering Institute

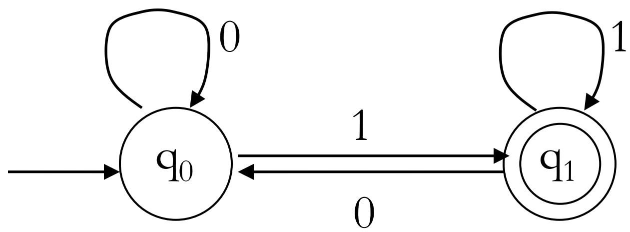
Formal Languages and Automata

Lesson 8: Limitations of finite automata

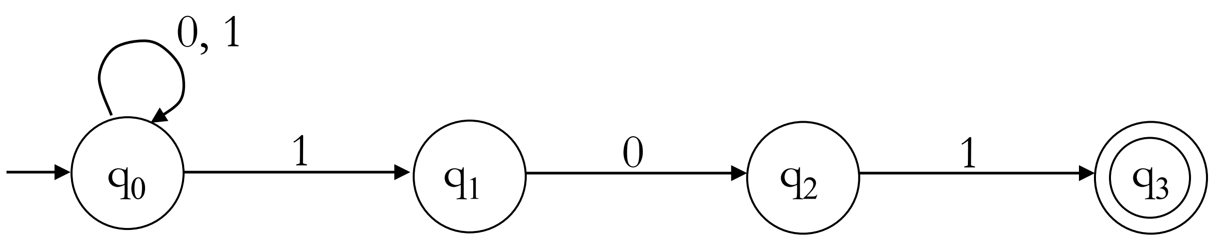
Marc Gaetano

Edition 2018

Many examples of regular languages



$$((0 + 1)0 + (0 + 1)(0 + 1)(0 + 1))^*$$



all strings not containing pattern 010 followed by 101

Is every language regular?

Which are regular?

$L_1 = \{0^n 1^m : n, m \geq 0\}$ $\Sigma = \{0, 1\}$

$L_2 = \{0^n 1^n : n \geq 0\}$

$L_3 = \{1^n : n \text{ is divisible by } 3\}$ $\Sigma = \{1\}$

$L_4 = \{1^n : n \text{ is prime}\}$

$L_5 = \{x : x \text{ has same number of 0s and 1s}\}$ $\Sigma = \{0, 1\}$

$L_6 = \{x : x \text{ has same number of patterns 01 and 10}\}$

Which are regular?

$$L_1 = \{0^n 1^m \colon n, m \geq 0\} = 0^*1^*$$, so regular

- How about:

$$L_2 = \{0^n 1^n \colon n \geq 0\} = \{\varepsilon, 01, 0011, 000111, \ldots\}$$

- Let's try to design a DFA for it

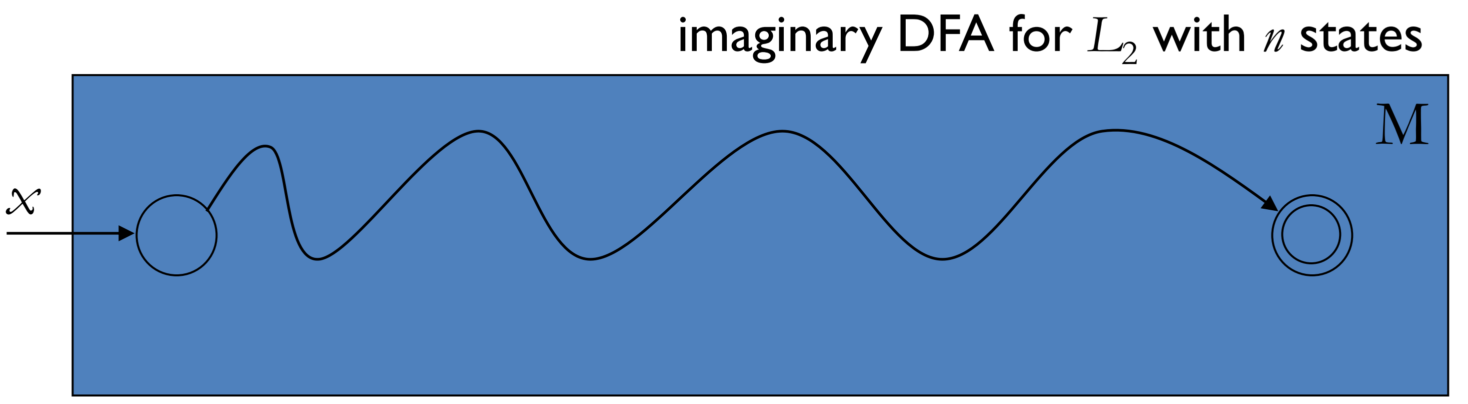
A non-regular language

- Theorem

> The language $L_2 = \{0^n 1^n : n \geq 0\}$ is not regular.

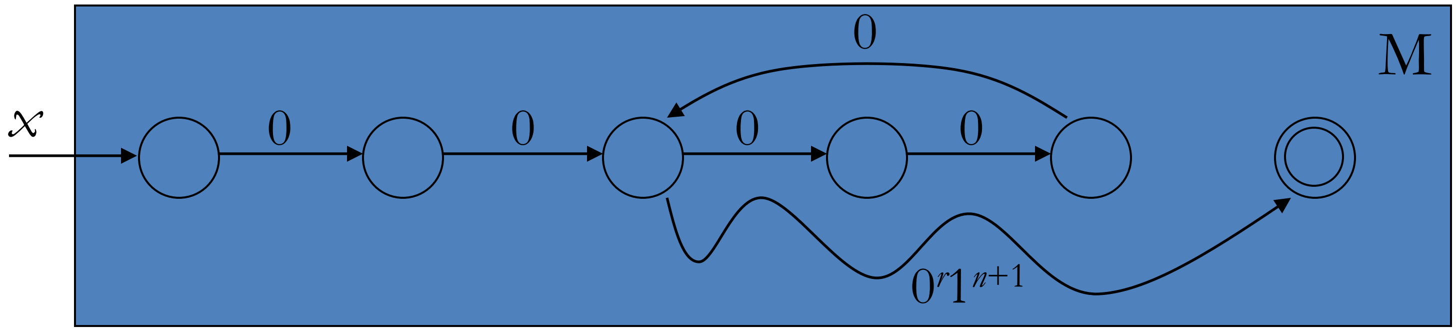
- To prove this, we argue by contradiction:
  - Suppose we have managed to construct a DFA M for L_2
  - We show something must be wrong with this DFA
  - In particular, M must accept some strings not in L_2

A non-regular language



- What happens when we run M on input $x = 0^{n+1}1^{n+1}$?
  - M better accept, because $x \in L_2$

A non-regular language



- What happens when we run M on input $x = 0^{n+1}1^{n+1}$?
  - M better accept, because $x \in L_2$
  - But since M has n states, it must revisit at least one of its states while reading 0^{n+1}

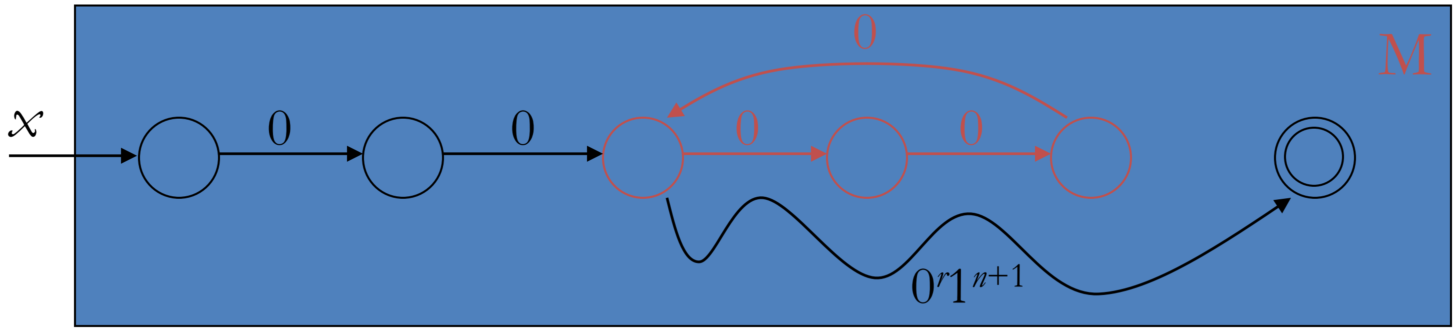
Pigeonhole principle

Suppose you are tossing $n + 1$ balls into n bins. Then two balls end up in the same bin.

- Here, balls are 0s, bins are states:

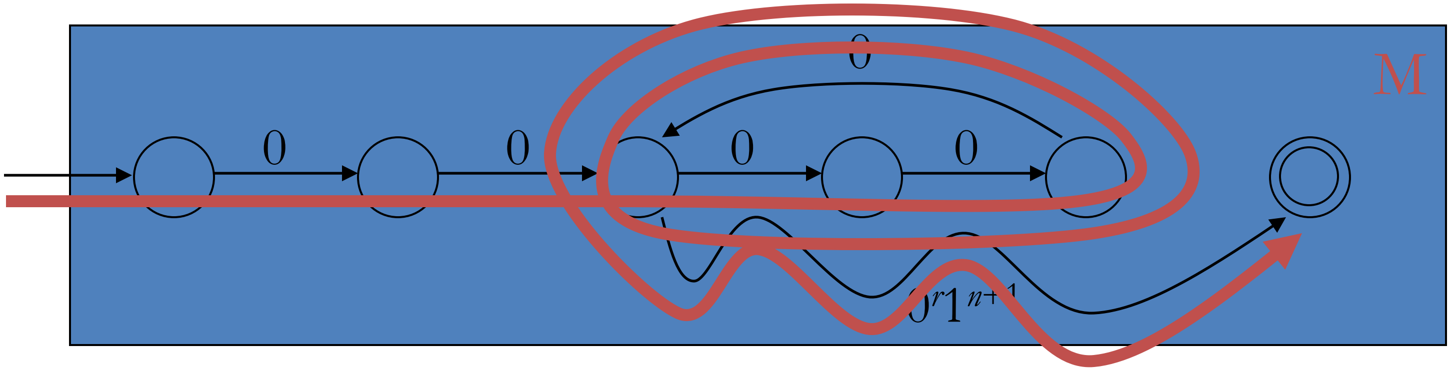
If you have a DFA with n states and it reads $n + 1$ consecutive 0s, then it must end up in the same state twice.

A non-regular language



- What happens when we run M on input $x = 0^{n+1}1^{n+1}$?
  - M better accept, because $x \in L_2$
  - But since M has n states, it must revisit at least one of its states while reading 0^{n+1}
  - But then the DFA must contain a loop with 0s

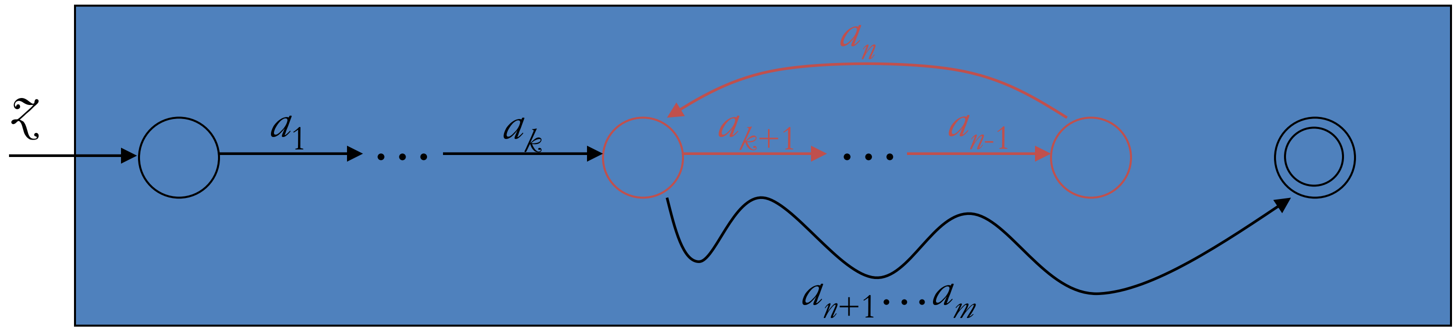
A non-regular language



- The DFA will then also accept strings that go around the loop multiple times
- But such strings have more 0s than 1s, so they are not in L_2!

General method for showing non-regularity

- Every regular language L has a property:



- For every sufficiently long input z in L, there is a "middle part" in z that, even if repeated several times, keeps the input inside L

Pumping lemma for regular languages

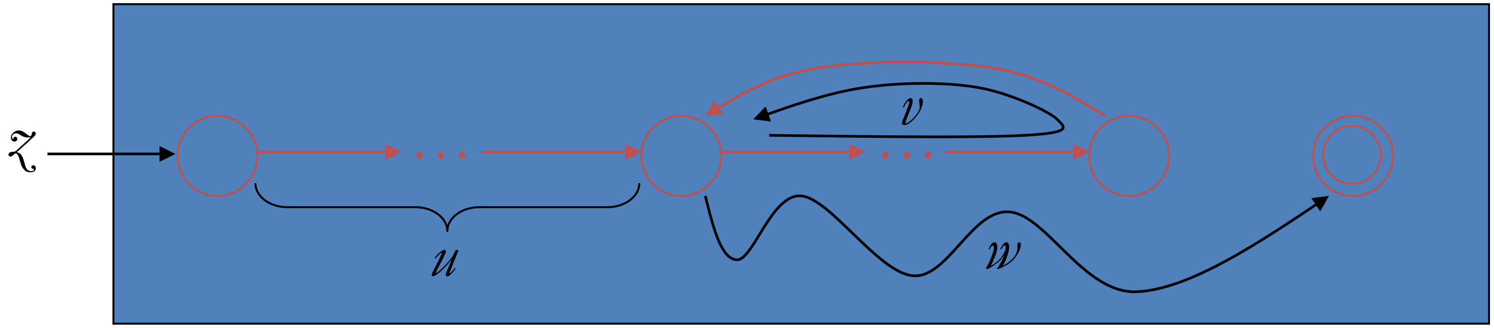
- Theorem: For every regular language L

There exists a number n such that for every string z in L, $|z| \geq n$ we can write $z = u\,v\,w$ where

① $|uv| \leq n$

② $|v| \geq 1$

③ For every $i \geq 0$, the string $u\,v^i\,w$ is in L.



Proving non-regularity

- If L is regular, then:

> There exists n such that for every z in L, $|z| \geq n$, we can write $z = u\,v\,w$ where ① $|uv| \leq n$, ② $|v| \geq 1$ and ③ For every $i \geq 0$, the string $u\,v^i\,w$ is in L.

- So to prove L is not regular, it is enough to show:

> For every n there exists z in L, $|z| \geq n$, such that for every way of writing $z = u\,v\,w$ where ① $|uv| \leq n$ and ② $|v| \geq 1$, the string $u\,v^i\,w$ is not in L for some $i \geq 0$.

Proving non regularity

For every n there exists z in L, $|z| \geq n$, such that for every way of writing $z = u\,v\,w$ where ① $|uv| \leq n$ and ② $|v| \geq 1$, the string $u\,v^i\,w$ is not in L for some $i \geq 0$.

- This is a game between you and an imagined adversary

	adversary	you
1	choose n	choose $z \in L$, $|z| \geq n$
2	write $z = uvw$ $(|uv| \leq n, |v| \geq 1)$	choose i
		you win if $uv^iw \notin L$

Proving non-regularity

- You need to give a strategy that, regardless of what the adversary does, always wins you the game

	adversary	you
1	choose n	choose $z \in L$, $\|z\| \geq n$
2	write $z = uvw$ $(\|uv\| \leq n, \|v\| \geq 1)$	choose i
		you win if $uv^iw \notin L$

Example

	adversary	you
1	choose n	choose $z \in L$, $\lvert z \rvert \geq n$
2	write $z = uvw$ $(\lvert uv \rvert \leq n, \lvert v \rvert \geq 1)$	choose i
		you win if $uv^iw \notin L$

- $L_2 = \{0^n1^n : n \geq 0\}$ $\Sigma = \{0, 1\}$

	adversary	you
1	choose n	$z = 0^{n+1}1^{n+1}$
2	$\underbrace{\text{write } z = uvw \ (\lvert uv \rvert \leq n, \lvert v \rvert \geq 1)}$	$i = 2$

$$u = 0^j,\ v = 0^k,\ w = 0^l1^{n+1}$$
$$j + k + l = n + 1$$

$$\begin{aligned} uv^iw &= 0^{j+2k+l}1^{n+1} \\ &= 0^{n+1+k}1^{n+1} \\ &\notin L \end{aligned}$$

More examples

$L_3 = \{1^n: n \text{ is divisible by } 3\}$ $\Sigma = \{1\}$

$L_4 = \{1^n: n \text{ is prime}\}$

$L_5 = \{x: x \text{ has same number of 0s and 1s}\}$ $\Sigma = \{0, 1\}$

$L_6 = \{x: x \text{ has same number of patterns 01 and 10}\}$

$L_7 = \{x: x \text{ has more 0s than 1s}\}$

$L_8 = \{x: x \text{ has different number of 0s and 1s}\}$