

BU CS 332 – Theory of Computation

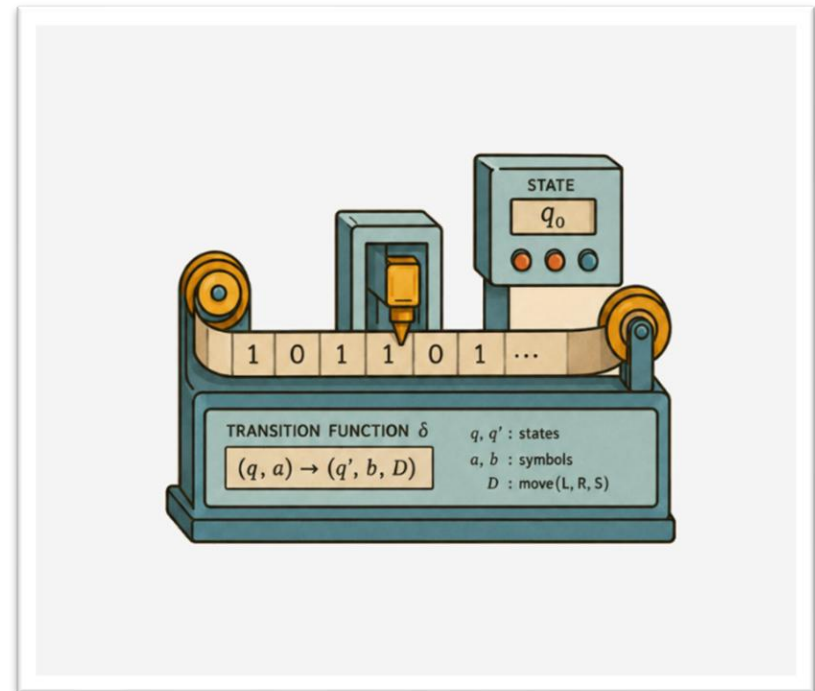
Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

“Myhill-Nerode” note

Test 1: Thursday, October 1



Alexander Poremba
September 24, 2026

Last Time

- Regular expressions characterize the regular languages
 - Every regex can be converted to an NFA recognizing its language
 - Every NFA can be converted to a regex generating its language

Last time: NFA $\rightarrow$ Regex

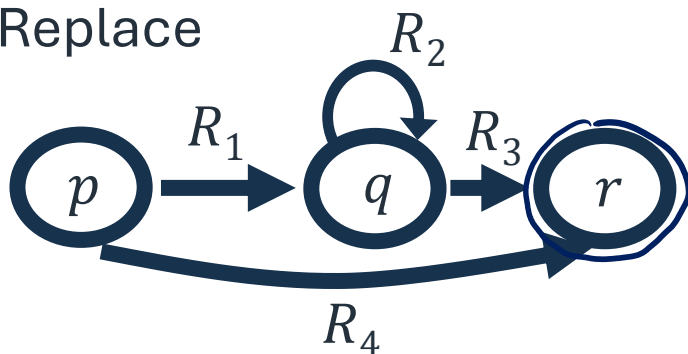
Theorem 2: Every NFA has an equivalent regular expression

Given NFA N :

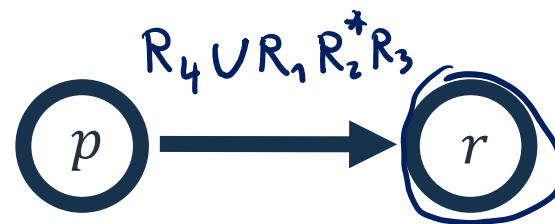
1. Make N into a **GNFA** by adding new start and accept state
2. For each non-start and non-accept state q :

For each pair of states p, r :

Replace

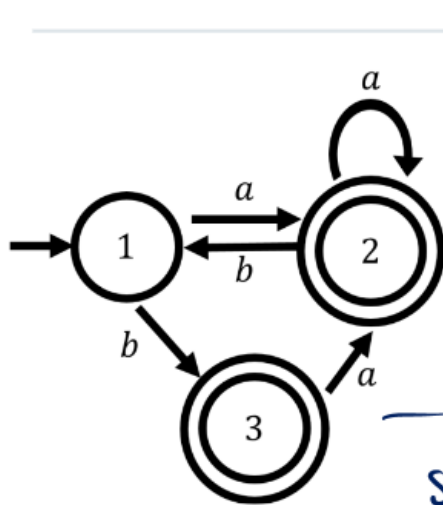


with

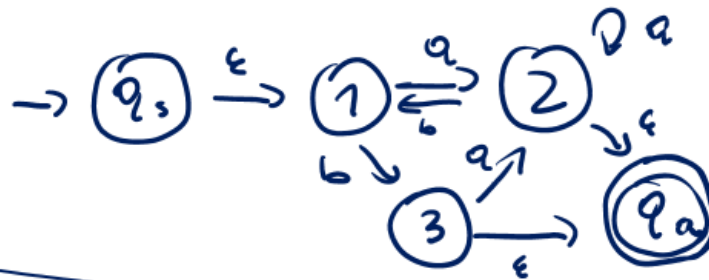


3. Output regex labeling transition from start to accept state

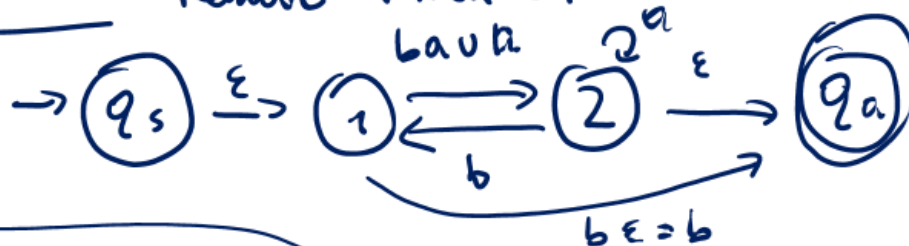
Last time: Example for NFA $\rightarrow$ Regex



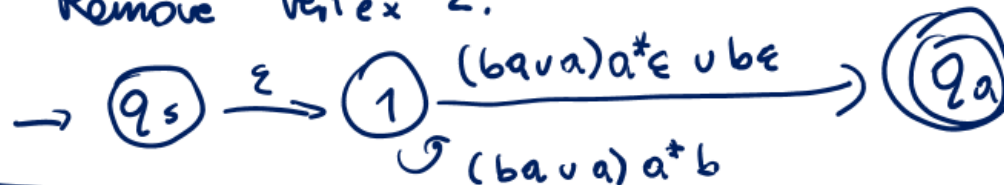
Step 1: Add q_s and q_a



Step 2: Remove vertex 3.

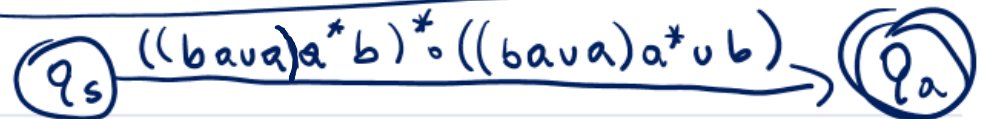


Step 3: Remove vertex 2.



final
regex

Step 4: Remove vertex 1.





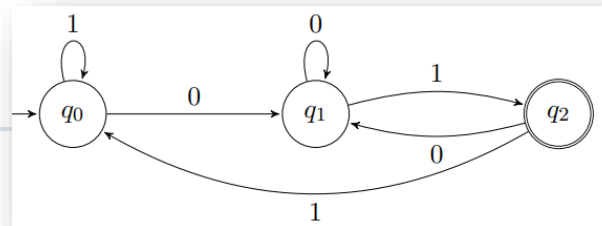
Limitations of Finite Automata

Motivating Questions

- We've seen several techniques for showing languages are **regular**
- How can we tell if we've found the **smallest DFA** recognizing a given language?
- Which languages are **not regular**? How can we prove so?

An Example

$M =$

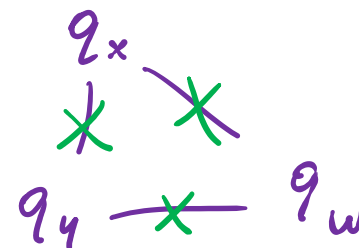


$$A = \{ w \in \{0, 1\}^* \mid w \text{ ends with } 01 \}$$

Claim: Every DFA recognizing A needs at least 3 states

Proof: Let M be any DFA recognizing A . Consider running M on each of $x = \varepsilon, y = 0, w = 01$

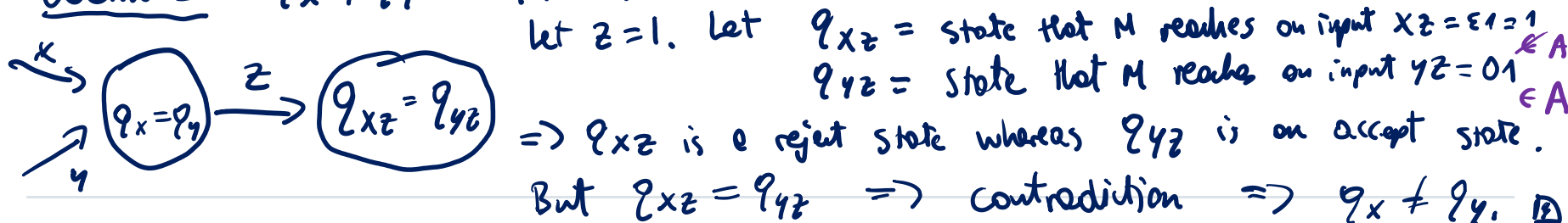
Let q_x = state that M reaches on input $x = \varepsilon$.
 q_y = state " " " on input $y = 0$.
 q_w = state " " " on input $w = 01$.



Claim: q_x, q_y, q_w are all distinct states.

Claim I: $q_w \neq q_x$ and $q_w \neq q_y$ because q_w is an accept state but q_x, q_y are reject states. ✓

Claim II: $q_x \neq q_y$. Why? Suppose for contradiction that $q_x = q_y$.



A General Technique

$$A = \{w \in \{0, 1\}^* \mid w \text{ ends with } 01\}$$

Definition: Strings x and y are **distinguishable** by L if there exists a “distinguishing extension” $z \in \Sigma^*$ such that exactly one of xz or yz is in L .

Ex. $x = \varepsilon$, $y = 0$ are distinguishable by A because
 $z = 1$ is a distinguishing extension!
 $xz = \varepsilon 1 = 1 \notin A$
 $yz = 01 \in A$ ✓

Definition: A set of strings S is **pairwise distinguishable** by L if every pair of distinct strings $x, y \in S$ is distinguishable by L .

Ex. $S = \{\varepsilon, 0, 01\}$

- $x = \varepsilon$ $y = 0 \Rightarrow z = 1$ ✓
 $xz = 0\varepsilon = 0 \notin A$
- $x = 0$ $y = 01 \Rightarrow z = \varepsilon$
 $xz = 0\varepsilon = 0 \notin A$
 $yz = 01\varepsilon = 01 \in A$ ✓
- $x = \varepsilon$ $y = 01 \Rightarrow z = \varepsilon$
 $xz = \varepsilon\varepsilon = \varepsilon \notin A$
 $yz = 01\varepsilon = 01 \in A$ ✓

A General Technique

Theorem: If S is **pairwise distinguishable** by L , then every DFA recognizing L must have at least $|S|$ states

Proof: Let M be a DFA with $< |S|$ states.

WTS: M cannot recognize the language L

Claim: There are distinct strings $x, y \in S$ such that M ends up in same state on x and y ... why? "Pigeonhole principle"

Pigeons: Strings in S

Assign "pigeon" x to "hole" q_x with

Holes: States of M

$q_x =$ state M reaches on input x .

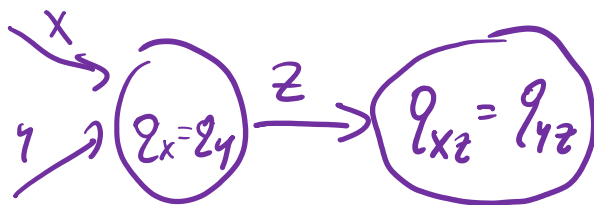
Let $x, y \in S$ be strings guaranteed by the claim. Since S is pairwise distinguishable by L , $\exists$ dist. extension z s.t. $xz \in L$ or $yz \in L$.

WTS: M gives the wrong answer on either xz or yz .

Let $q_x = q_y =$ state M reaches on x (or y)

$\Rightarrow q_{xz} = q_{yz}$... can't be both on except and reject state

$\Rightarrow$ contradiction



$\Rightarrow M$ requires $\geq |S|$ states!

Another Example

Claim: Every DFA that recognizes B must have ≥ 4 states

$$B = \{w \in \{0,1\}^* \mid |w| = 2\}$$

Theorem: If S is pairwise distinguishable by L , then every DFA recognizing L must have at least $|S|$ states

$$S = \{\epsilon, 0, 00, 000\}$$

Exercise: Check all of the different cases!

- $x = \epsilon \quad y = 0 \quad \boxed{z = 0}$

Then: $xz = \epsilon 0 = 0 \notin B$
 $yz = 00 \in B$ ✓

• ...

Intuition: A DFA recognizing B must:

- On input ϵ : "be prepared to reject iff it sees 2 more characters"
- On input 0 : "... 1 more character"
- On input 00 : "needs to accept"
- On input 000 : "needs to reject"

Question: Distinguishing Extension

Which of the following is a valid **distinguishing extension** for $x = 0$ and $y = 00$ for the language $B = \{w \in \{0, 1\}^* \mid |w| = 2\}$?

a) $z = \varepsilon$ ✓

b) $z = 0$ ✓

c) $z = 1$ ✓

d) $z = 00$ ✗ ←

$$xz = 000 \notin B$$

$$yz = 0000 \notin B$$

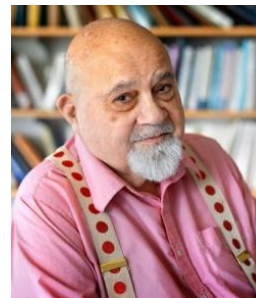


Historical Note

Converse to the distinguishing set method:

If L has **no** distinguishing set of size $> k$, then L is recognized by a DFA with k states

Myhill-Nerode Theorem (1958): L is recognized by a DFA with $\leq k$ states **if and only if** L does not have a distinguishing set of size $> k$



Non-Regularity

Theorem: If S is pairwise distinguishable by L , then every DFA recognizing L must have at least $|S|$ states

contrapositive form:

$\exists$ DFA recognizing L with $< K$ states $\Rightarrow$ there is no pairwise distinguishable set for L with $\geq K$ states.

Corollary: If S is an **infinite** set that is pairwise distinguishable by L , then no DFA recognizes L

contrapositive form:

$\exists$ DFA recognizing $L \Rightarrow L$ does not have an infinite pairwise distinguishable set

Proof that "contrapositive of Thm $\Rightarrow$ contrapositive of Corollary":

Let L be recognized by some DFA. Let K be the number of states.

(By contrapositive of Thm) $\Rightarrow$ there is no pairwise distinguishable set of size $\geq K+1$
 $\Rightarrow L$ does not have an infinite pairwise dist. set. $\square$

The Classic Example

Theorem: $A = \{0^n 1^n \mid n \geq 0\}$ is not regular

Proof: We construct an **infinite** pairwise distinguishable set

$$S = \{\epsilon, 0, 00, 000, \dots\} = \{0^n \mid n \geq 0\}$$

let $x, y \in S$ be arbitrary distinct strings.

$\Rightarrow$ we can write $x = 0^n$, $y = 0^m$ ($n, m \geq 0$, $n \neq m$)

Claim: $z = 1^n$ is a distinguishing extension.

$$xz = 0^n 1^n \in A \quad \checkmark$$

$$yz = 0^m 1^n \notin A \quad \times$$

Since S is an infinite pairwise dist. set for A

$\Rightarrow A$ is not regular $\Rightarrow$ no DFA/NFA can recognize A . $\square$



Next time: More Non-Regularity & Test Review