# BU CS 332 – Theory of Computation

https://forms.gle/vs8YHr7UJNdMGzjv8

For regex $R$: $R^+ = RR^*$

$\varepsilon \in L(R^*)$ for all $R$,

$\varepsilon \in L(RR^*) \iff \varepsilon \in L(R)$

## Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

"Myhill-Nerode" note

Mark Bun

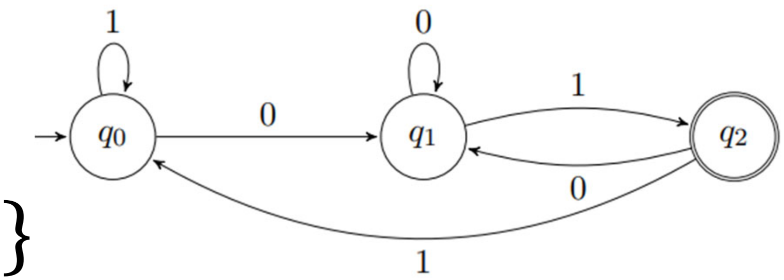
February 12, 2024

# Last Time

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

- Limits of Finite Automata
  - How can we tell if we've found the smallest DFA recognizing a language?
  - Are all languages regular? How can we prove that a language is not regular?

# An Example




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

Claim: *Every* DFA recognizing $A$ needs at least 3 states

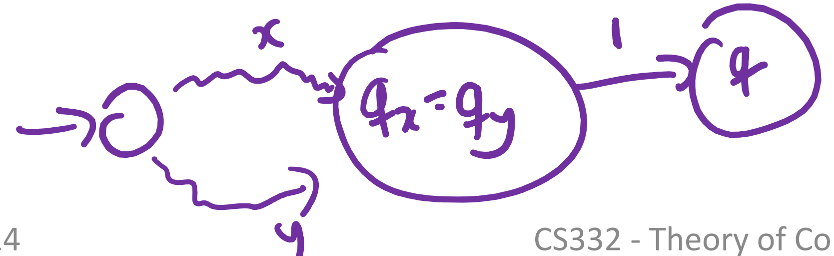
Proof: Let $M$ be any DFA recognizing $A$. Consider running $M$ on each of $x = \varepsilon, y = 0, w = 01$

Let $q_x$ = state M reaches after reading $x$
$q_y$ = " $y$
$q_w$ = " $w$

Goal: Prove that $q_x, q_y, q_z$ are all distinct.

Claim: $q_x \neq q_w$ and $q_y \neq q_w$ Why? $q_w$ is an accept state
$q_x, q_y$ are reject states

claim: $q_x \neq q_y$
Proof: suppose FTSOC $q_x = q_y$




M on input $x1 = 1$ must reject
M on input $y1 = 01$ must accept
* because $q$ can't be both an accept & reject state

# 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. $x = \varepsilon, \quad y = 0$

$z = 1$

$xz = \varepsilon 1 = 1 \notin A$

$yz = 01 \in A$

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

$x = \varepsilon \quad y = 0 \quad z = 1 \quad (\varepsilon 1 \notin A, \ 01 \in A)$

$x = \varepsilon \quad y = 01 \quad z = \varepsilon \quad (\varepsilon\varepsilon \notin A, \ 01\varepsilon \in A)$

$x = 0 \quad y = 01 \quad z = \varepsilon \quad (0 \notin A, \ 01 \in A)$

# A General Technique

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

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

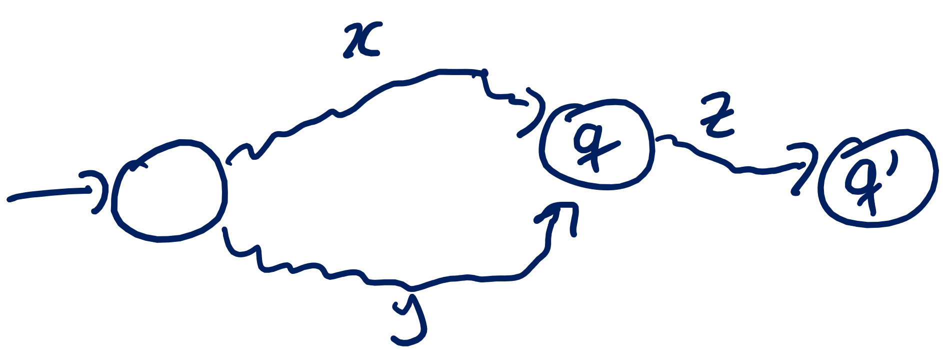
<u>Claim:</u> There are distinct strings $x, y \in S$ such that $M$ ends up in same state on $x$ and $y$   Why? Pigeonhole principle

<u>Proof of claim:</u>   Pigeons: strings in $S$   Holes: DFA states

Pigeon $x$ is assigned to hole (state) $q$ if $M$ lands in state $q$ when run on $x$

$\exists$ DFA state $q$ s.t. $M$ lands in $q$ when reading each of $x, y$

Let $z$ be distinguishing extension for $x, y$: exactly one of $xz, yz \in L$




$\exists$ state $q'$ s.t. $M$ ends in $q'$ when reading each of $xz$, $yz$

Regardless of whether $q'$ is an accept or reject state, $M$ must mess up on either $xz$ or $yz$ $\Rightarrow$ $M$ does not recognize $L$.

# Another Example

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

Claim: Every DFA recognizing B requires $\geq 4$ states

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

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

Claim: S is pairwise distinguishable

$x = \varepsilon \quad y = 0: \quad z = 0$

$xz = 0 \notin B \quad yz = 00 \in B$

$x = \varepsilon \quad y = 00: \quad z = \varepsilon$

$xz = \varepsilon \notin B \quad yz = 00 \in B$

$x = \varepsilon \quad y = 000: \quad z = 00$

$xz = 00 \in B \quad yz = 00000 \notin B$

$x = 0 \quad y = 00: \quad z = \varepsilon \quad xz = 0 \notin B \quad yz = 00 \in B$

Idea: Intuitively, any DFA for B needs to remember if prefix read so far has length:

0

1

2

$\geq 3$

$x = 0 \quad y = 000: \quad z = 0$

$x = 00 \quad y = 000: \quad z = \varepsilon$

# Distinguishing Extension

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

a) $z = \varepsilon$ $xz = 0 \notin B$ $yz = 00 \in B$

b) $z = 0$ $xz = 00 \in B$ $yz = 000 \notin B$

c) $z = 1$ $xz = 01 \in B$ $yz = 001 \notin B$

d) $z = 00$ $xz = 000 \notin B$ $yz = 0000 \notin B$

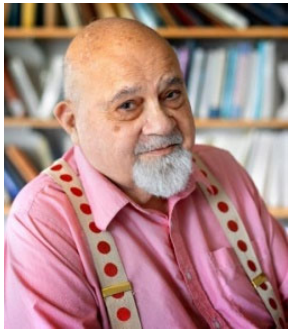
Remark: Want exactly one of $xz$ or $yz$ to be in $B$

# 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 $L$ needs at least $|S|$ states

Contrapositive: $\exists$ DFA for $L$ using $< k$ states $\Rightarrow$ no pairwise dist. set of size $k$

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

Contrapositive: $\exists$ DFA for $L$ $\Rightarrow$ no infinite pairwise dist. set

---

Proof of contrapositive of cor. from contrapositive of Thm:

$\exists$ DFA for $L$ $\Rightarrow$ $\exists$ $k$ s.t. $\exists$ a DFA of size $k$ for $L$

$\Rightarrow$ $L$ does not have a pairwise dist. set of size $k+1$

$\Rightarrow$ $L$ does not have an infinite pairwise dist. set.

# The Classic Example

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

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

Idea: Need to behave differently on $\varepsilon$, 0, 00, 000, ...
Because need to know how many 1's to wait for.

Let $S = L(0^*) = \{\varepsilon, 0, 00, 000, \ldots\}$

Claim: $S$ is an infinite pairwise dist set for $A$.

Proof: Let $x, y \in S$ distinct.
Suppose $x = 0^m$, $y = 0^n$ where $m \neq n$
Let $z = 1^m$. Then $xz = 0^m 1^m \in A$
$yz = 0^n 1^m \notin A$

# Palindromes

Theorem: $L = \{w \in \{0,1\}^* \mid w = w^R\}$ is not regular

Proof: We construct an infinite pairwise distinguishable set

Attempt 1: $S = \{0,1\}^*$

Let $x, y$ be arbitrary

Set $z = x^R$

Want: $xx^R \in L$ but $yx^R \notin L$

$x = 00$ $\quad xx^R = 0000$

$y = 000$ $\quad yx^R = 00000$

Attempt 2: $S = L(0^*1) = \{0^n1 \mid n \geq 0\}$

Let $x = 0^n1$, $y = 0^m1$ $\in S$ where $m \neq n$

Claim: $z = 0^n$ is a dist. extension

$xz = 0^n10^n \notin L$ $\qquad yz = 0^m10^m \notin L$.

# Now you try!


Use the distinguishing set method to show that the following languages are not regular

$L_1 = \{0^i 1^j \mid i > j \geq 0\}$ $= \{0, 00, 001, 000, 0001, 00011, 0000, \ldots\}$

Your job: Build an infinite set $S$ such that for all $x \neq y \in S$, there exists a $z$ such that exactly one of $xz$ and $yz$ is in $L$

$S = L(0^*) = \{0^n \mid n \geq 0\}$

Let $x \neq y \in S$: $x = 0^m$ $y = 0^n$ s.t. $n > m$ (WLOG, switch $x$ & $y$ if necessary)

$z = 1^{n-1}$

$xz = 0^m 1^{n-1} \notin L_1$ $yz = 0^n 1^{n-1} \in L_1$

because $n > m \Rightarrow m \leq n-1$


# Now you try!

Use the distinguishing set method to show that the following languages are not regular

$$L_2 = \{1^{n^2} \mid n \geq 0\}$$

$S = L_2 = \{1^{n^2} \mid n \geq 0\}$

let $x \neq y \in S$

$x = 1^{n^2}$ $\quad y = 1^{m^2}$ where $n > m$

$z = 1^{2m+1}$

$xz = 1^{n^2+2m+1} \notin L_2$

$yz = 1^{m^2} 1^{2m+1} = 1^{m^2+2m+1} = 1^{(m+1)^2} \in L_2$

because "next" perfect square after $n^2$ is $(n+1)^2 = n^2 + 2n+1 > n^2 + 2m+1$

# Reusing a Proof

Finding a distinguishing set can take some work…

Let's try to reuse that work!

How might we show that

$$BALANCED = \{w \mid w \text{ has an equal \# of 0s and 1s}\}$$

is not regular?

Not regular $\qquad\qquad = L(0^*1^*)$ $\quad$ regular

$$\{0^n1^n \mid n \geq 0\} = BALANCED \cap \{w \mid \text{all 0s in } w \text{ appear before all 1s}\}$$

Claim. BALANCED is not regular.

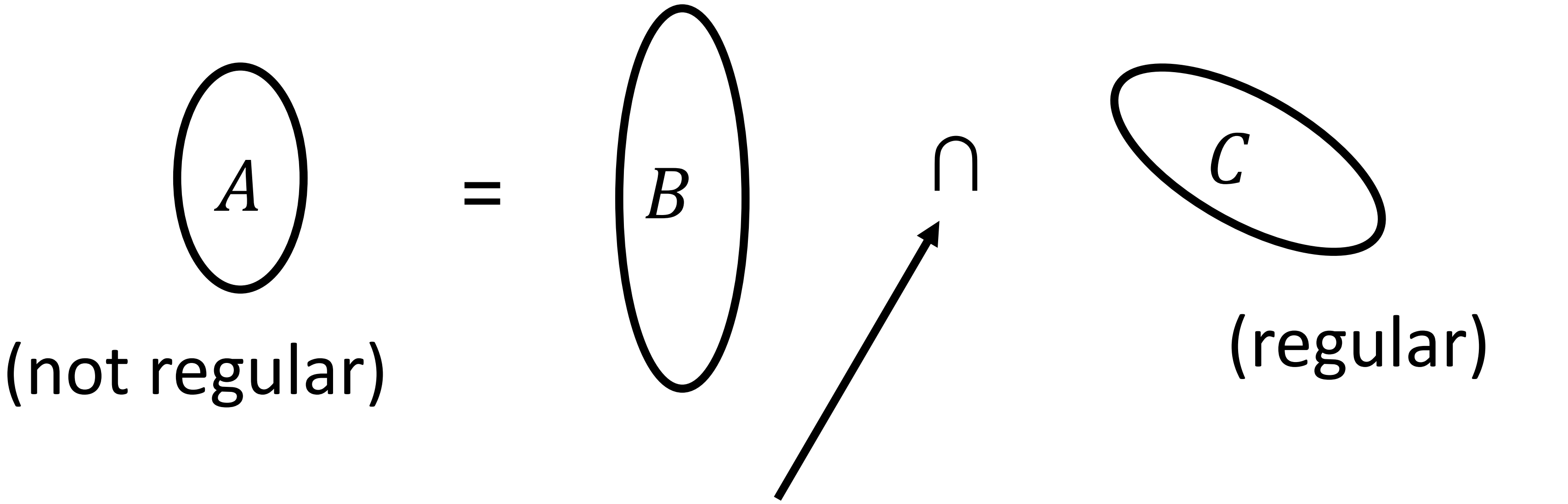
Proof. Assume for contradiction that BALANCE is regular

$\Rightarrow$ BALANCED $\cap$ $L(0^*1^*)$ is regular (because reg. langs. closed under $\cap$)

$\Rightarrow$ $\{0^n1^n \mid n \geq 0\}$ is regular ✗

# Using Closure Properties

If $A$ is not regular, we can show a related language $B$ is not regular




any of $\{\circ, \cup, \cap\}$ or, for one language, $\{\neg, {}^R, {}^*\}$

By contradiction: If $B$ is regular, then $B \cap C$ $(= A)$ is regular.

But $A$ is not regular so neither is $B$!