BU CS 332 – Theory of Computation

https://forms.gle/bhghDqPsgErUBYPw7

Lecture 7:

- Distinguishing sets
- Non-regular languages

Reading:

"Myhill-Nerode" note

Mark Bun

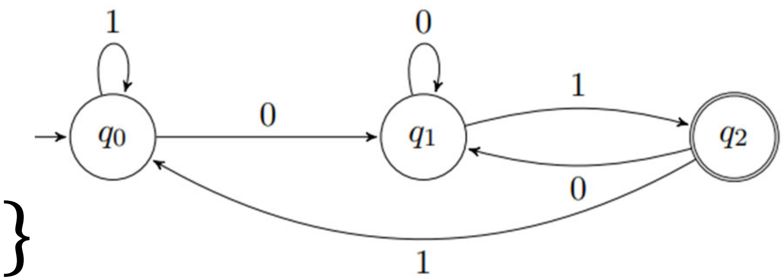
February 12, 2025

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

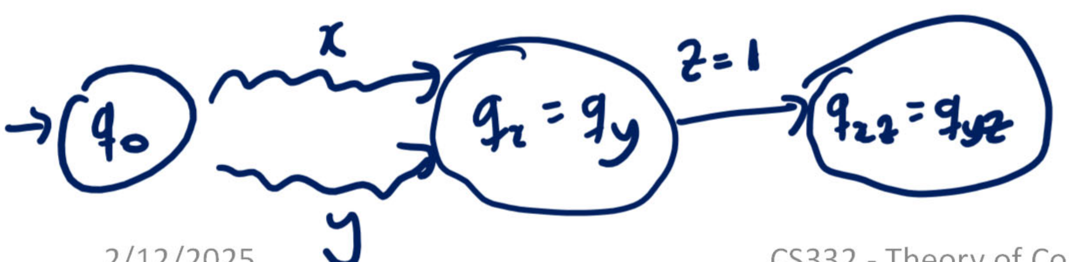
Proof: Let M be any DFA recognizing A. Consider running M on each of $x = \varepsilon, y = 0, w = 01$

Let q_x = state M reaches on input x
q_y = " " input y
q_w = " " input w

Goal: Show that q_x, q_y, q_w are all distinct states

$q_x \neq q_w$, $q_y \neq q_w$ why? $x{=}\varepsilon, y \notin A \Rightarrow q_x, q_y$ are reject states
$w \in A \Rightarrow q_w$ is an accept state

Now WTS $q_x \neq q_y$. Assume FTSOC that $q_x = q_y$. Let $z = 1$,



q_{xz} = state M reaches on input xz is a reject state
$\parallel$
q_{yz} = state M reaches on input yz is an accept 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. $x = \varepsilon, \quad y = 0$

$z = 1$ is a distinguishing extension because

$xz = \varepsilon 1 = 1 \notin A$
$yz = 01 \in A$

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

$x = \varepsilon \quad y = 0$: Let $z = 1$

$x = \varepsilon \quad y = 01$: Let $z = \varepsilon$. Then $xz = \varepsilon \notin L$
$yz = 01 \in L$

$x = 0 \quad y = 01$: Let $z = \varepsilon$. Then $xz = 0 \notin L$
$yz = 01 \in L$.

A General Technique

Eg. $A = \{w \mid w \text{ ends in } 01\}$
$S = \{\varepsilon, 0, 01\} \Rightarrow$ every DFA for A needs ≥ 3 states

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.

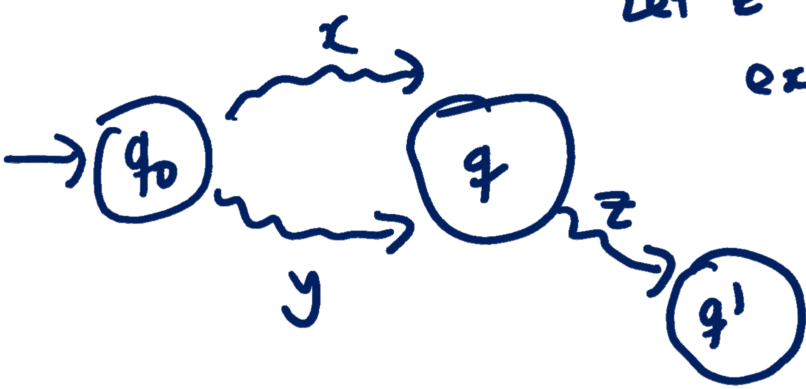
<u>Claim:</u> There are distinct strings $x, y \in S$ such that M ends up in same state on x and y Why?. Pigeonhole principle

Holes = states of M
Pigeons = strings in S

Map string x to state q_x = state M reaches when reading x

Let q be the state M reaches when reading x (= state reached on y)

Let z be the distinguishing extension for x, y, i.e. exactly one of $xz \in L$ or $yz \in L$



Let q' = state M reaches on xz
= state M reaches on yz

Since q' is either an accept state or reject state, $\Rightarrow$ must give wrong answer on either xz or yz

$\Rightarrow M$ does not recognize L.

Another Example

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

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

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

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

$x = \varepsilon \quad y = 00$: Let $z = \varepsilon$
$xz = \varepsilon \notin L$
$yz = 00 \in L$

$x = \varepsilon \quad y = 000$: Let $z = 00$
$xz = 00 \in L$
$yz = 000 \notin L$

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

Intuition:

On input ε, M needs to wait for exactly 2 more symbols
0 " 1 more symbol
00 " 0 more symbols
000 M should reject

$x = 00 \quad y = 000$: Let $z = \varepsilon$
$xz = 00 \in L$
$yz = 000 \notin L$

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

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

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

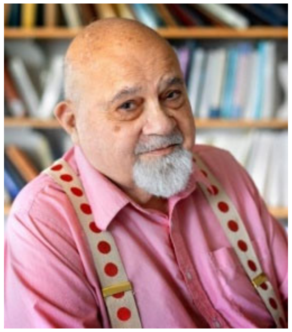
d) $z = 00$ $\quad xz = 000 \notin B \quad yz = 0000 \notin B$

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 needs at least $|S|$ states

Contrapositive: If $\exists$ a DFA for L w/ k states,
then L does not have a pair. dist. set of size $k+1$

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

Contrapositive: If $\exists$ a DFA for L
then L does not have an infinite pair. dist. set

Proof that (contrapositive of Thm) $\Rightarrow$ (Contrapositive of Cor.)

Suppose L is recognized by DFA M. Let $k = \#$ of states of M
By (contrapositive of Thm), L does not have a pair. dist. set of size $k+1$
$\Rightarrow$ L does not have an infinite pair. 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

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

Check S is pairwise distinguishable:

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

$\exists\ m \neq n \geq 0$ s.t. $x = 0^m$, $y = 0^n$.

Let $z = 1^m$. Then $xz = 0^m 1^m \in A$

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

So S is pairwise dist.

Palindromes

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

Proof: We construct an infinite pairwise distinguishable set

First attempt: Let $S = \{0,1\}^*$ (infinite)

Let $x \neq y \in S$. Let $z = x^R$. Then $xz = xx^R \in L$

$yz = yx^R$ ... maybe $\in L$

maybe $\notin L$?

e.g. $x = 000$ $y = 0000 \Rightarrow yx^R = 0000000 \in L$.

Let $S = \{0^n 1 \mid n \geq 0\}$.

Let $x \neq y \in S$. Then $\exists\, m \neq n$ s.t. $x = 0^m 1$, $y = 0^n 1$

Let $z = 0^m$. Then $xz = 0^m 1 0^m \in L$

$yz = 0^n 1 0^m \notin L$.

Now you try!

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

$L_1 = \{0^i 1^j \mid i > j \geq 0\}$

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

$S = \{0^n \mid n \geq 0\}$

If $x \neq y \in S$, WLOG $m > n$ st. $x = 0^m$ $y = 0^n$

Let $z = 1^n$ $\Rightarrow$ $xz \in L_1$

$yz \notin L_1$

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 $x \neq y \in S$, so $\exists\ \underline{m > n}$ s.t. $x = 1^{m^2}$ $y = 1^{n^2}$ (wLOG)

Let $z = 1^{2n+1}$

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

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

$m^2 < m^2+2n+1 < m^2+2m+1 = (m+1)^2$