BU CS 332 – Theory of Computation

https://forms.gle/L9BMDsJ48ezNaEW76

Lecture 8:

- More on non-regularity
- Test 1 review

Reading:

"Myhill-Nerode" note

HW3 due tomorrow (Friday) 11:59 PM

Test 1 next Thursday 2/19

Alexander Poremba & Mark Bun

February 12, 2026

Last Time: Distinguishing Set Method

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.

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.

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

Corollary: If language L has an infinite pairwise distinguishable set, then L is not regular.

Palindromes

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

Proof: We construct an infinite pairwise distinguishable set

Attempt 1

$S_1 = \{0^m 1^n \mid m, n \geq 0\}$

$S_2 = \{0^k 1^l \mid k, l \geq 0\}$

Not allowed to have two sets

$x = 0^m 1^n$ $\quad z = 1^n 0^m$

$y = 0^k 1^l$

$xz = 0^m 1^n 1^n 0^m \in L$

$yz = 0^k 1^l 1^n 0^m$ might accidentally be in lang. if $k = m$

Attempt 2

Let $S = \{0^n 1^n \mid n \geq 0\}$ which is an infinite set

To show S is pairwise distinguishable:

let $x = 0^n 1^n$

$y = 0^m 1^m$ for some $m \neq n$ be two arbitrary distinct strings.

Then $z = 1^n 0^n$ is a distinguishing extension because

$xz = 0^n 1^n 1^n 0^n \in L$

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

Now you try!

Use the distinguishing set method to show that the following language is not regular

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

<u>Your job:</u> 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

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

Let $x = 0^m$
$y = 0^n$ where $m \neq n$ be arbitrary distinct strings in S
WLOG $m > n$

Then $z = 1^n$ is a distinguishing extension since:

$xz = 0^m 1^n \in L$ $(m > n)$
$yz = 0^n 1^n \notin L$ (n is not $> n$)

So S is pairwise distinguishable

Perfect Squares

Use the distinguishing set method to show that the following language is not regular

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

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

To show S is pairwise distinguishable:

Let $x = 1^{m^2}$

$y = 1^{n^2}$ where WLOG $m > n$

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

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

Not a perfect square: Because $m^2+2n+1 < m^2+2m+1 = (m+1)^2$ which is the next perfect square past m^2.

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

Reusing a Proof

Finding a distinguishing set can take some work…

Let's try to reuse that work!

$\{0^n1^n\} \subseteq BALANCED \subseteq \{0,1\}^*$

How might we show that

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

is not regular?

known non-regular — known regular $L(0^*1^*)$

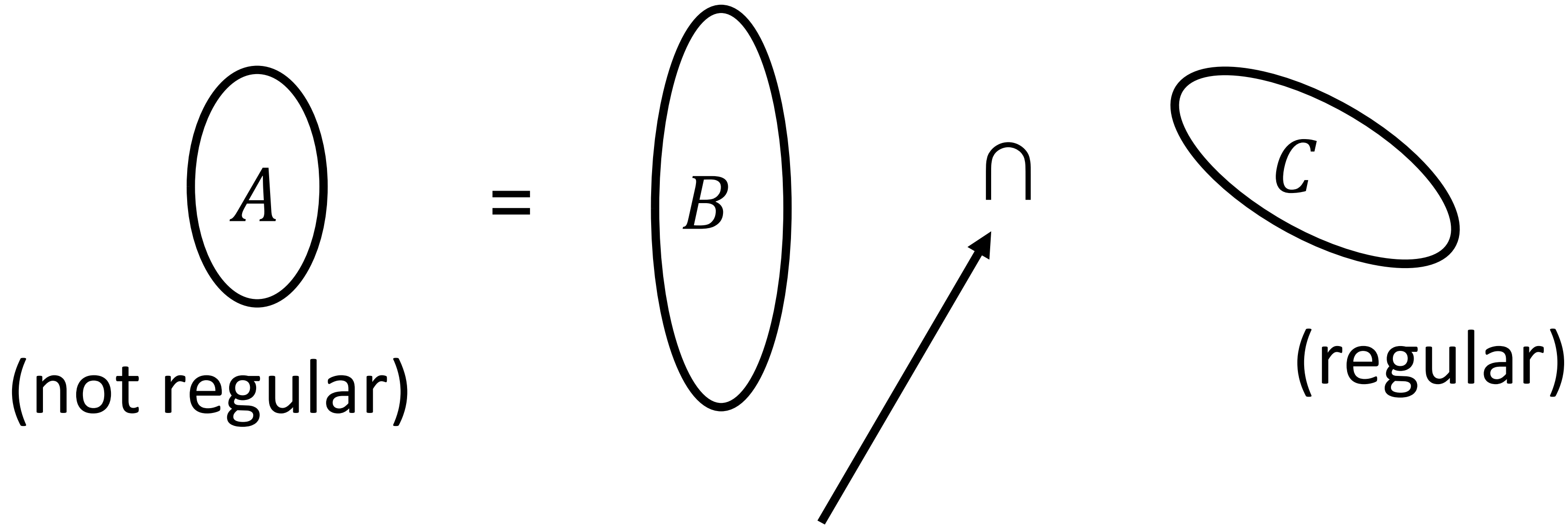
$\{0^n1^n \mid n \geq 0\} = BALANCED \cap \{w \mid$ all 0s in w appear before all 1s$\}$

WTS: BALANCED is not regular

Proof by contradiction: Assume FTSOC that BALANCED is regular. Then $\{0^n1^n \mid n \geq 0\}$ is the intersection of two regular languages, hence regular because the regular langs. are closed under intersection. But this contradicts our previous proof that $\{0^n1^n \mid n \geq 0\}$ was non-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, $\{\neg, {}^R, *\}$

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

But A is not regular so neither is B!

Example

$\bar{B} = \{0^i 1^j \mid i = j\} \cup \{w \mid$ not the case that all 0's come before all 1's$\}$

Prove $B = \{0^i 1^j | i \neq j\}$ is not regular using $C = L(0^*1^*)$

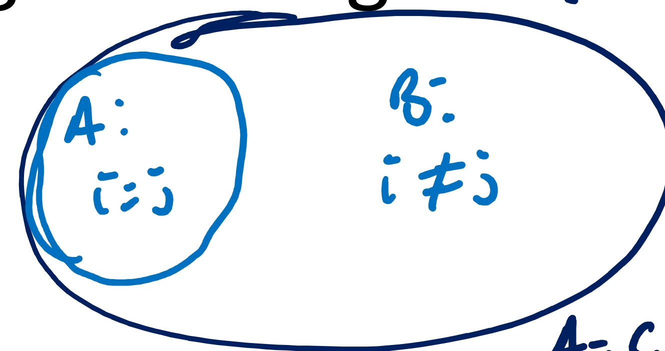
- Nonregular language

  $A = \{0^n 1^n | n \geq 0\}$ and

- Regular language

  $C = \{w \mid \text{all 0s in } w \text{ appear before all 1s}\}$ $= C \cap \bar{B}$

  $= L(0^*1^*)$



A = C

Which of the following expresses A in terms of B and C?

a) $A = B \cap C$

b) $A = \bar{B} \cap C$

c) $A = B \cup C$

d) $A = \bar{B} \cup C$

Proof that B is nonregular

Assume for the sake of contradiction that B is regular

We know: $A = \bar{B} \cap C$ ← regular

(A: known non-regular; $\bar{B}$: mystery language)

$\bar{B}$ is regular (closure under complement)

C is regular ($C = L(0^* 1^*)$)

$\Rightarrow \bar{B} \cap C$ is regular (closure under intersection)

$\Rightarrow A$ is regular $\times$ because A is non-regular

!DANGER!

Let $B = \{0^i 1^j | i \neq j\}$ and write $B = A \cup C$ where

- Nonregular language

  $A = \{0^i 1^j | i > j \geq 0\}$ and

- Nonregular language

  $C = \{0^i 1^j | j > i \geq 0\}$ and

Does this let us conclude B is nonregular?

No Because the union of two non-regular languages could be regular!

Ex: $L = \{0^n 1^n \mid n \geq 0\}$ non-regular

$\overline{L} = \{0^m 1^n \mid m \neq n\} \cup \{w \mid$ not all 0's come before all 1's$\}$ non-regular

(because if $\overline{L}$ were regular, $\overline{(\overline{L})} = L$ would also be regular)

$L \cup \overline{L} = \{0,1\}^*$ which is regular.

Test 1 Topics

Sets, Strings, Languages (0)

- Know the definition of a string and of a language (and the difference between them)
- Understand operations on strings: Concatenation, reverse

  $x \circ y = xy = x_1 \dots x_l y_1 \dots y_m$
- Understand operations on languages: Union, intersection, concatenation, reverse, star, complement

  $L_0 \circ L_1 = \{ xy \mid x \in L_0, y \in L_1 \}$
- Know the difference between $\emptyset$ and ε

Deterministic FAs (1.1)

- Given an English or formal description of a language L, draw the state diagram of a DFA recognizing L (and vice versa)
- Know the formal definition of a DFA (A DFA is a 5 tuple…) and convert between state diagram and formal description
- Know the formal definition of how a DFA computes
- Construction for closure of regular languages under complement

Nondeterministic FAs (1.2)

- Given an English or formal description of a language L, draw the state diagram of an NFA recognizing L (and vice versa)
- Know the formal definition of an NFA
- Know the subset construction for converting an NFA to a DFA
- Proving closure properties: Know the constructions for union, concatenation, star
- Know how to prove your own closure properties

Regular Expressions (1.3)

- Given an English or formal description of a language L, construct a regex generating L (and vice versa)
- Formal definition of a regex

Base: a, ϕ, ε

Inductive rules: $\cup$, $\circ$, $*$

- Know how to convert a regex to an NFA
- Know how to convert a DFA/NFA to a regex

Limitations of DFAs (Myhill-Nerode Note)

- Understand the statements of the distinguishing set method for proving DFA size lower bounds / non-regularity
- Understand the proof of why the distinguishing set method works, and be able to use it to prove similar statements
- Know how to apply the method to specific languages
- We won't ask you to prove that some language is non-regular, since you didn't have any homework problems on this yet

Test format

Problem 1: "Check your type checker"

E.g., Is aabba a string, language, or a regex?

How about {ab} U {aab}? Language

Problem 2: True/false with **justification**

Either provide a convincing explanation or a specific counterexample

Problems 3-5(?) Homework-style problems

Study tips

- Make sure you know how to solve the problems on the practice test and are familiar with the format. The format/length of the real test will be very similar.

- If you need more practice, there are lots of problems in the book. We're happy to talk about any of these problems in office hours.

- You may bring a page of notes (writing on both sides ok) to the test. Preparing this note sheet is a great study aid.

Test tips

- You may cite without proof any result…
  - Stated in lecture
  - Stated and proved in the main body of the text (Ch. 0-1.3 + Myhill-Nerode note)
  - These include worked-out examples of state diagrams, regexes
- Not included above: homework problems, discussion problems, (solved) exercises/problems in the text

- Showing your work / explaining your answers will help us give you partial credit
- Make sure you're interpreting quantifiers (for all / there exists) correctly and in the correct order

Practice Problems

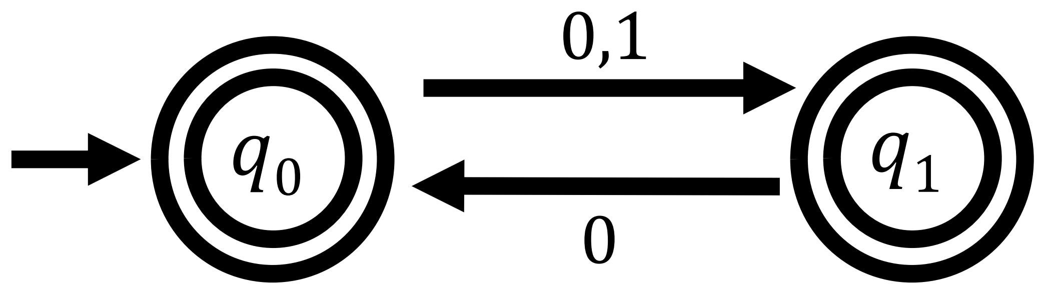
Name six operations under which the regular languages are closed

Prove or disprove: All finite languages are regular

Prove or disprove: The **non-**regular languages are closed under intersection

Give the state diagram of an NFA recognizing the language $(01 \cup 10)^* \circ 1$

Give an equivalent regular expression for the following NFA



For a language L over $\{0,1\}$, define the operation $\mathbf{split}(L) = \{x\#y \mid x, y \in L\}$. Show that the regular languages are closed under split

$\text{split}(L) \subseteq \{0, 1, \#\}^*$

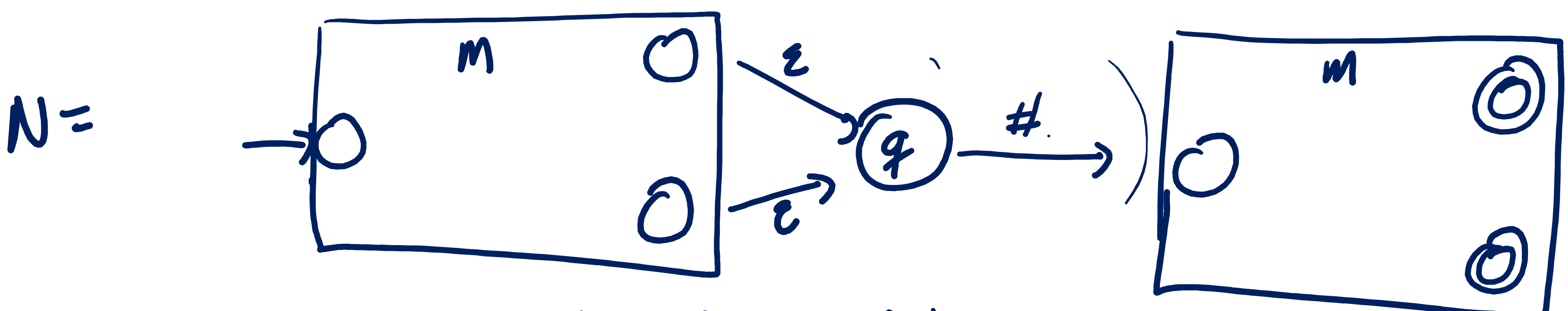
Solution 1: Let L be an arbitrary regular lang. generated by regex R

split(L) is generated by $R\#R$

$\Rightarrow$ split(L) is regular

Solution 2: Let L be an arbitrary regular language. Then L is recognized by some DFA M.

Construct a new NFA N:



i.e. connect accept states in first copy to q via ε-transitions, make those non-accepting

Then connect q to start state of second copy via # transition.

For a language L over alphabet Σ, define the operation $\mathbf{drop}(L) = \{xz \mid xyz \in L \text{ where } x, z \in \Sigma^*, y \in \Sigma\}$. Show that the regular languages are closed under $\mathbf{drop}$.

Is the following language regular?
$\{0^n1^n \mid 0 \leq n \leq 2026\}$

Is the following language regular? $\{a^n a^n \mid n \geq 0\}$

How many states does a DFA recognizing $\{0^n 1^n \mid 0 \le n \le 2026\}$ require?