CSCI 361 Lecture 6: Verifying Regularity

Shikha Singh

Announcements & Logistics

- **Recap/Questions**
  - Any questions from Chapter 1.3 (Regular expressions)?
- **HW 2** is due this **Wed at 10 pm** on Gradescope
  - Questions about regular expressions and how to convert between regular expressions and NFA/DFA
- Few reminders about assignment submissions:
  - Must select pages for each question
  - Must use the LaTeX template provided
  - Attaching images: please crop and attach ***in place***

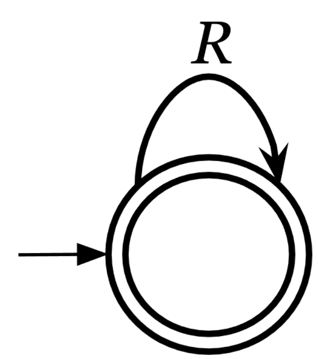
Overview So Far

- Simple model of computation: finite automata
  - DFA $\Longleftrightarrow$ NFA $\Longleftrightarrow$ Regular expression
- Simplest class of decision problems: regular languages
- A language L is regular iff
  - Exists a DFA M such that $L(M) = L$, or
  - Exists a NFA N such that $L(N) = L$, or
  - Exists a regular expression R such that $L(R) = L$

NFAs to Regular Expression

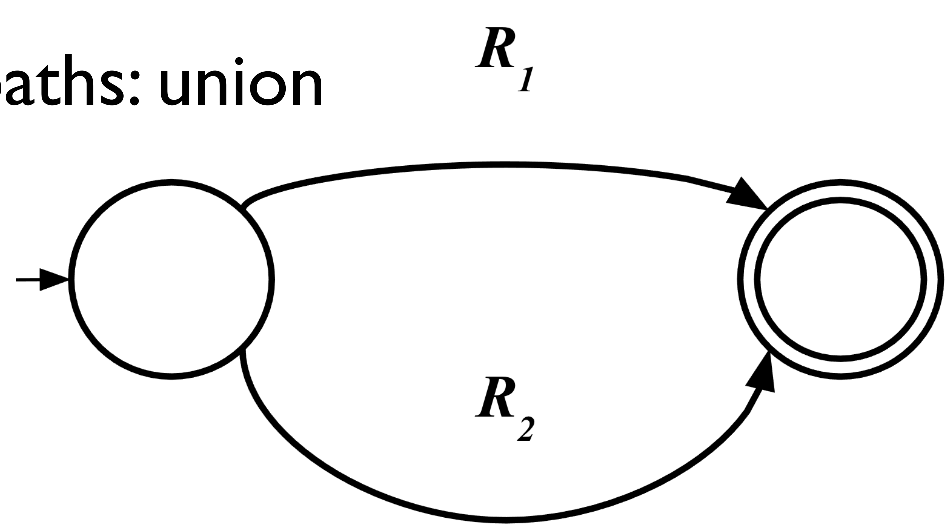
- Every NFA can be converted to an equivalent regular expression

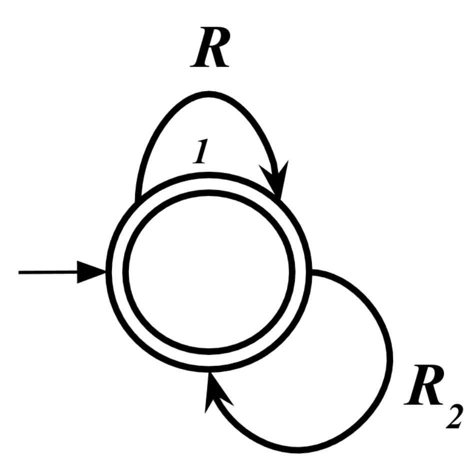
Self loops: Kleene star



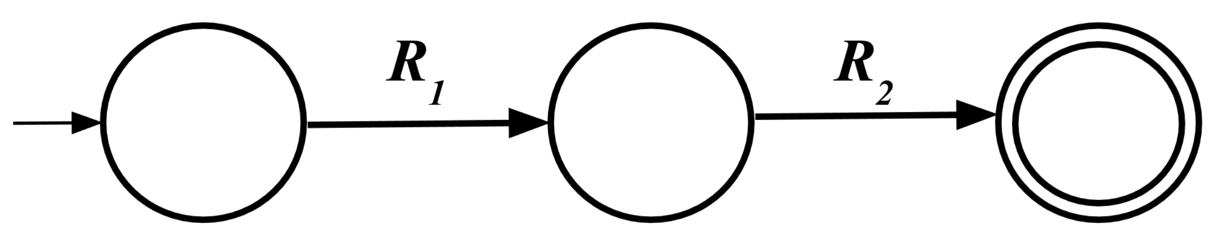
R^*

Alternate paths: union





$R_1 \cup R_2$



$R_1 \circ R_2$

Adjacent paths: concatenation

This Week

- Move on to **non-regular languages**
- Techniques to prove a language is not regular:
    - Myhill-Nerode theorem (not in textbook)
    - Pumping Lemma (textbook approach)

Not All Languages Are Regular

- Any language does not a DFA that recognizes it is not-regular
- How we do ***prove no such DFA*** exists?
  - First example of an impossibility results in this class
  - Many more to come
- Intuitively, any decision problem that requires ***finite*** memory "to solve" is regular
- **Question.** Are finite languages regular?
  - $L \subseteq \Sigma^*$ and $|L|$ is finite
  - All finite languages are regular
- Given an infinite language, how do we know its regular?

All Finite Languages are Regular

- **Theorem.** All finite languages are regular.
- $L = \{w_1, \ldots, w_n\}$ for some $n \in \mathbb{N}$
- Let $L_i = \{w_i\}$ for each $i \in \{1, \ldots, n\}$
- $L = \cup_{i=1}^{n} L_i$
- **Claim 1.** Each L_i is regular.
- **Claim 2.** A finite union of regular languages is regular.
- Using Claim 1 and 2, L is regular

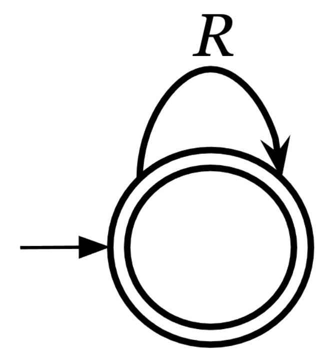
Infinite Regular Languages

- Have seen many infinite regular languages
- What do they have in common?

Structure of Infinite Regular Languages

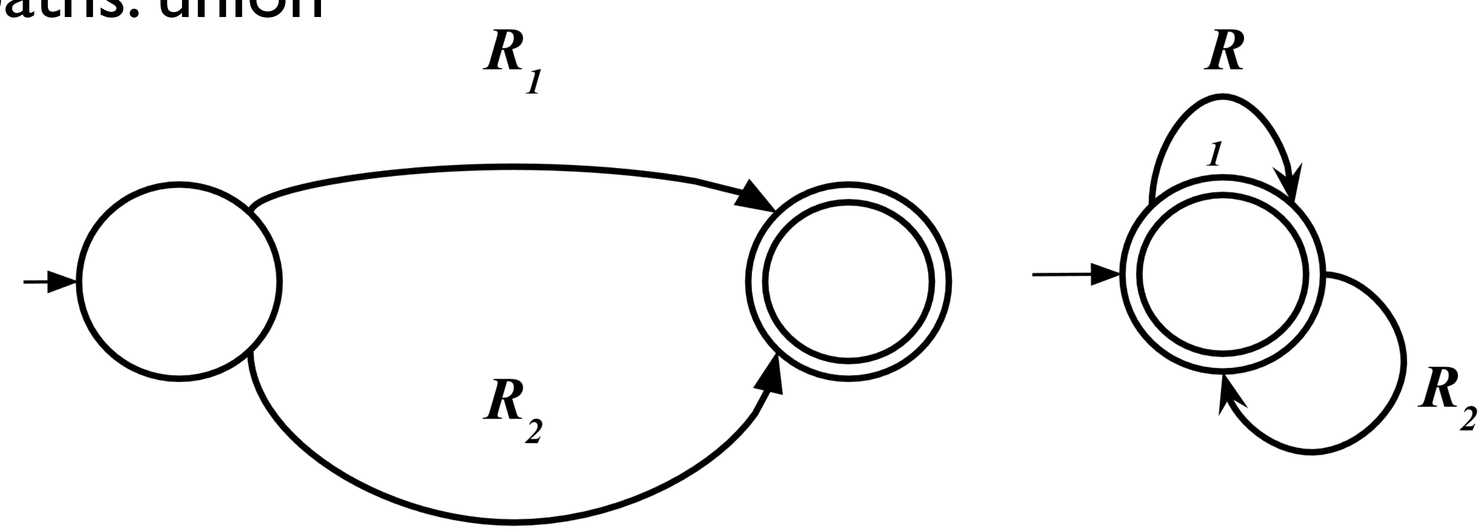
- Which of these are responsible for going from finite to infinite?

Self loops: Kleene star



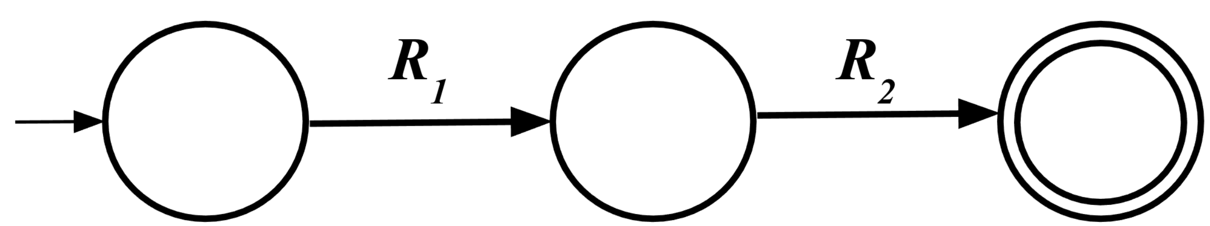
R^*

Alternate paths: union



$R_1 \cup R_2$

$R_1 \circ R_2$



Adjacent paths: concatenation

Loops in DFA: Intuition

- Consider the DFA $M's$ transitions on an input string w
- It enters some states $q_0, \ldots, q_1, q_2, \ldots, q_n$
- **Question.** If there is a "loop" what does that mean about the states visited?

- Now suppose two different strings x, y bring M to the same state q
- Consider attaching the same suffix z to both
- **Question.** What can we say about the state M is in after reading input string xz versus after reading input string yz?

Indistinguishability (DFA)

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Let x, y be any string over Σ.

Definition. x **indistinguishable to** y with respect to M, denoted $x \sim_M y$ if and only if $\delta^*(q_0, x) = \delta^*(q_0, y)$ (i.e., the state reached by M on x is the same as the state reached by M on y)

Corollary. If $x \sim_M y$ then for all $z \in \Sigma^*$, then
$xz \in L(M) \Longleftrightarrow yz \in L(M)$

Indistinguishability (Languages)

Let L be any language over an alphabet Σ.

Definition. x **indistinguishable to** y with respect to L, denoted $x \equiv_L y$ if and only if for all $z \in \Sigma^*$, we have that $xz \in L \iff yz \in L$

Problem 5 in HW 2: $\equiv_L$ is an equivalence relation over Σ^*

Thus, $\equiv_L$ ***partitions*** Σ^* into equivalence classes.

Distinguishing Suffixes

- Every string in the same equivalence class $[x]$ of $\equiv_L$ are indistinguishable with each other
- Two strings $x, y \in \Sigma^*$ are in different equivalence iff they are ***distinguishable***
  - Can find a suffix $z \in \Sigma^*$ that distinguishes them, that is, $xz \in L$ and $yz \notin L$ or $xz \notin L$ and $yz \in L$
- **Question.** Suppose $x \in L$ and $y \notin L$, are they distinguishable?

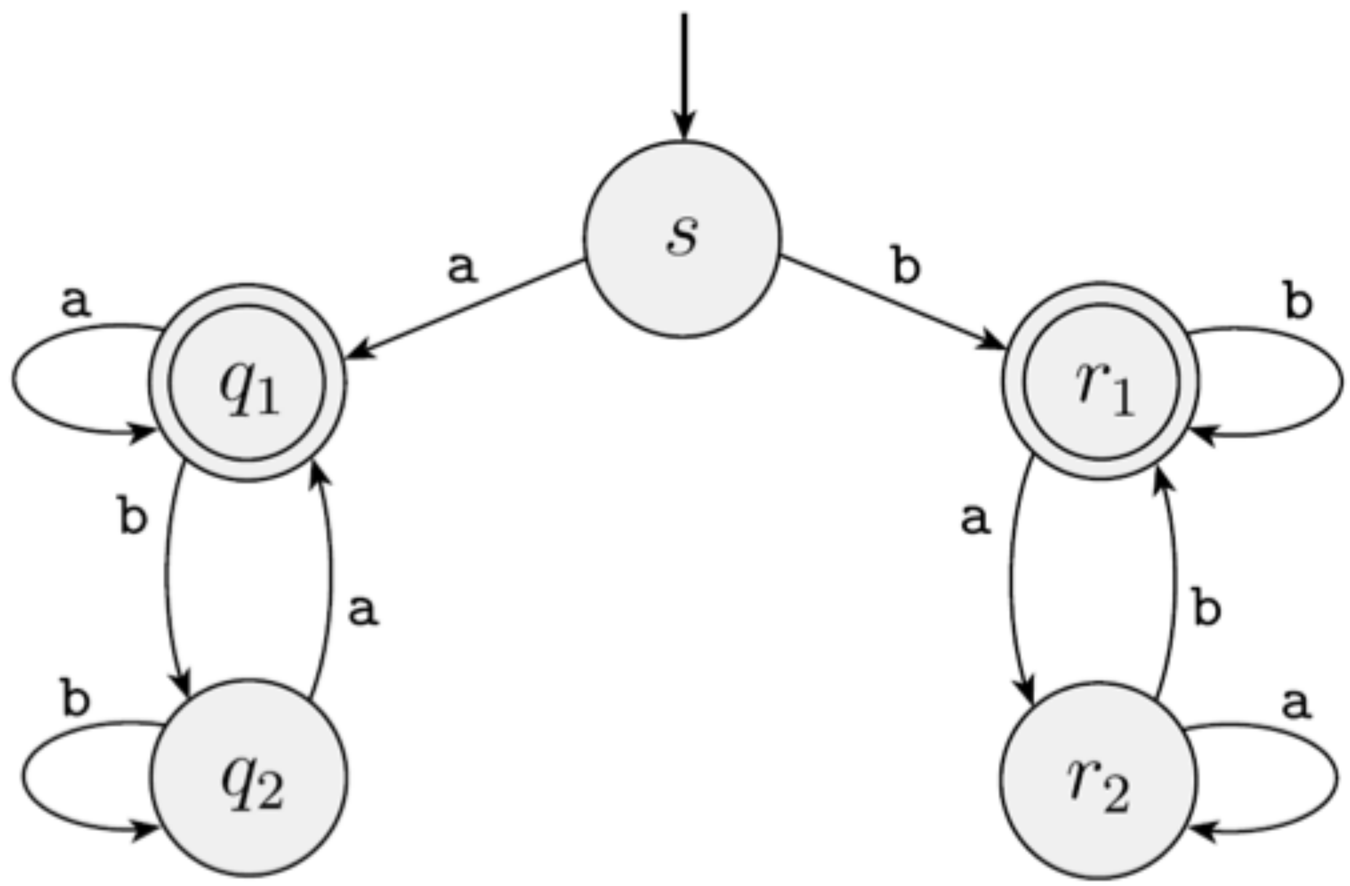
Indistinguishability (Languages)

- **Example.**
  $L = \{w \in \{a, b\}^* \mid w \text{ starts and ends with the same symbol}\}$
- **Question.** Regular expression for L?
- **Problem.** Find the equivalence classes of the relation $\equiv_L$.

Minimal DFA

- **Example.**

  $L = \{w \in \{a, b\}^* \mid w$ starts and ends with the same symbol$\}$



Indistinguishability DFA vs Languages

- **Claim.** If $x \sim_M y$, then $x \equiv_{L(M)} y$.

Minimal DFA

- **Claim.** If a language L over Σ has k equivalence classes defined by $\equiv_L$, then any DFA for L must have at least k states.
- **Corollary.** If a DFA M for L has number of states equal to the number of equivalence classes of $\equiv_L$ then such a DFA is minimal.

Myhill-Nerode Theorem

Let L be a language over Σ^*, then L is regular **if and only if** the relation $\equiv_L$ over Σ^* has a finite number of equivalence classes.

Myhill-Nerode Theorem

Let L be a language over Σ^*, then L is regular **if and only if** the relation $\equiv_L$ over Σ^* has a finite number of equivalence classes.

Necessary condition. For L to be regular, it must have finitely many equivalence classes. Equivalently, if $\equiv_L$ over Σ^* has an infinite number of equivalence classes, then L cannot be regular.

Sufficient condition. If $\equiv_L$ has finitely many equivalence classes, then L must be regular.

Proving Non Regularity

- Myhill-Nerode theorem says that any language that has infinitely many equivalence classes with respect to $\equiv_L$ is not regular
- Typically, we don't need to find all of equivalence classes
- Sufficient to find an infinite subset of strings that are mutually distinguishable

Fooling Sets

Definition. A set of strings $S \subseteq \Sigma^*$ is a **fooling set** with respect to a language $L \subseteq \Sigma^*$ if every pair of strings in S is distinguishable with respect to each other.

Example. $L = \{w \in \{a, b\}^* \mid w$ starts and ends with the same symbol$\}$

An example fooling set for L?

Question. Can the size of a fooling set be bigger than the number of equivalence classes?

- Max size of a fooling set for L = # of equivalence class of $\equiv_L$
- Size of any fooling set for $L \leq$ # of equivalence class of $\equiv_L$

Myhill-Nerode Theorem

Maximum fooling set size of L

$= \#$ equivalence classes of $\equiv_L$

$=$ minimum states of DFA for L

Takeaway. If we could prove that there exists an infinite number of distinguishable sets for a language, it would mean that even the smallest DFA for the language would require an infinite number of states. Therefore, no such DFA exists and the language cannot be regular.

Proving Non-Regularity

Problem. Prove that the language $L = \{a^i b^i \mid i \in \mathbb{N}\}$ is not regular.

Hint. Identify and prove that L has an infinite fooling set.

Proving Non-Regularity

Problem. Prove that the langueg

$L = \{a^n \mid n \in \mathbb{N}$ and n is a power of $2\}$ is not regular.

Hint. Identify and prove that L has an infinite fooling set.