CSCI 361 Lecture 5: Proving Non-Regularity

Shikha Singh

Announcements & Logistics

- **HW 3** will be released this afternoon, due Wed 24 at 10 pm
- Hand in Exercise #4, pick up Exercise #5
- Colloquium tomorrow: **2:35pm in Wege**
  - *Data driven algorithms for online decision making* (Roie Levin, Rutgers)

Not All Languages Are Regular

- Last time: all finite languages are regular.
- Today: Characterizing what type of infinite languages are regular?
- Intuitively, DFAs can only remember ***finitely*** many things
- Use the property that DFA cannot distinguish between two different strings that brings it to the same state
- Today: ways to prove a language is not regular
    - Myhill Nerode (not in the book)
    - Pumping lemma (Ch 1.4 in the book)
    - Closure properties and known non-regular languages

Indistinguishability (DFA)

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

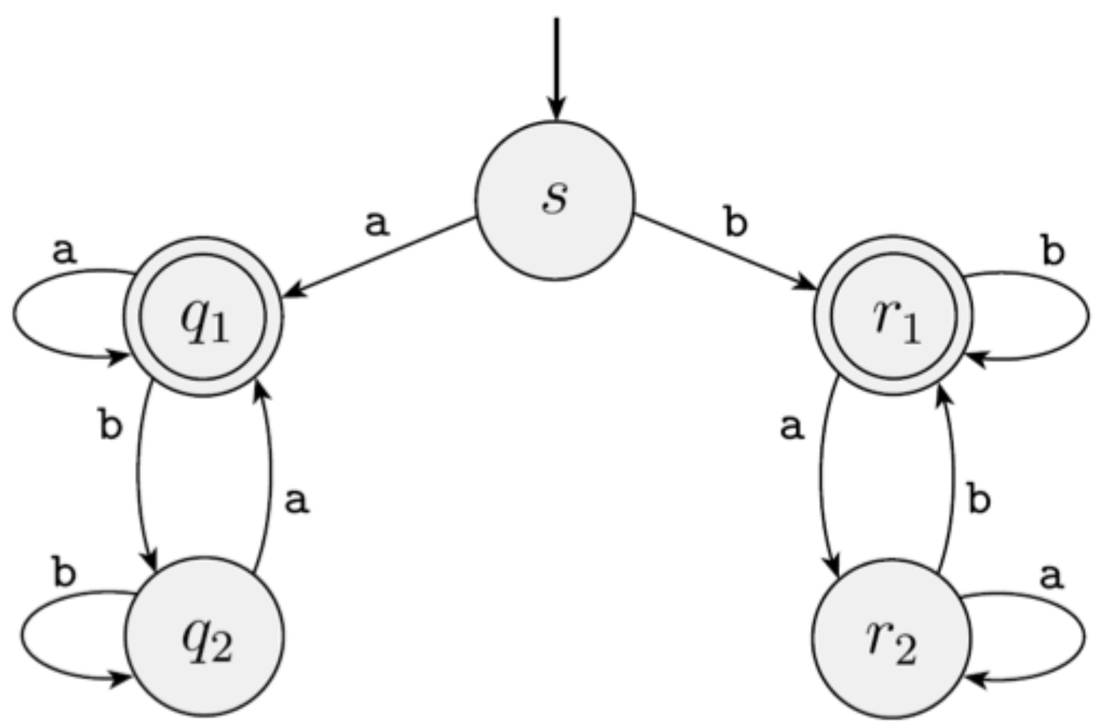
Definition. x **indistinguishable to** y with respect to a DFA M, denoted $x \sim_M y$ if and only if $\delta^*(q_0, x) = \delta^*(q_0, y)$ (i.e., the state reached by M on x is the same as the state reached by M on y)

Corollary. If $x \sim_M y$ then for all $z \in \Sigma^*$, then

$xz \in L(M) \iff yz \in L(M)$

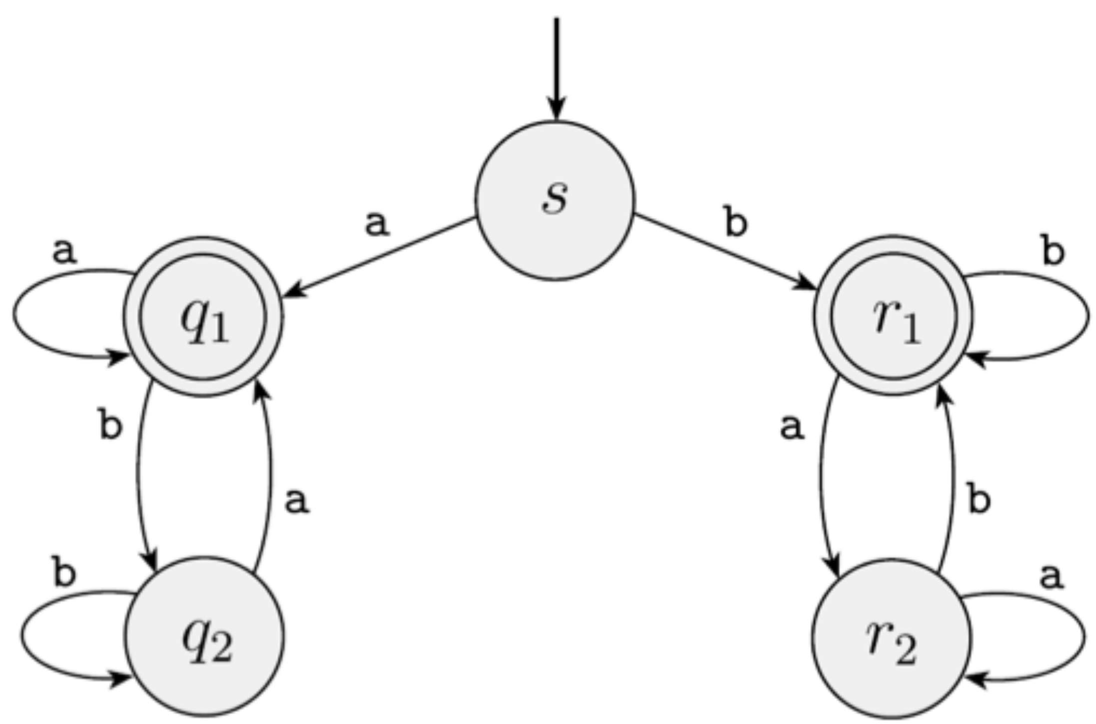
Class Exercise

- **Example.** $L = \{w \in \{a, b\}^* \mid w$ starts and ends with the same symbol$\}$
- **Definition.** ***x* indistinguishable to *y*** with respect to a DFA 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)
- **Question:** for each state in the DFA for L, write a regular expression characterizing all strings that bring the DFA to that state.



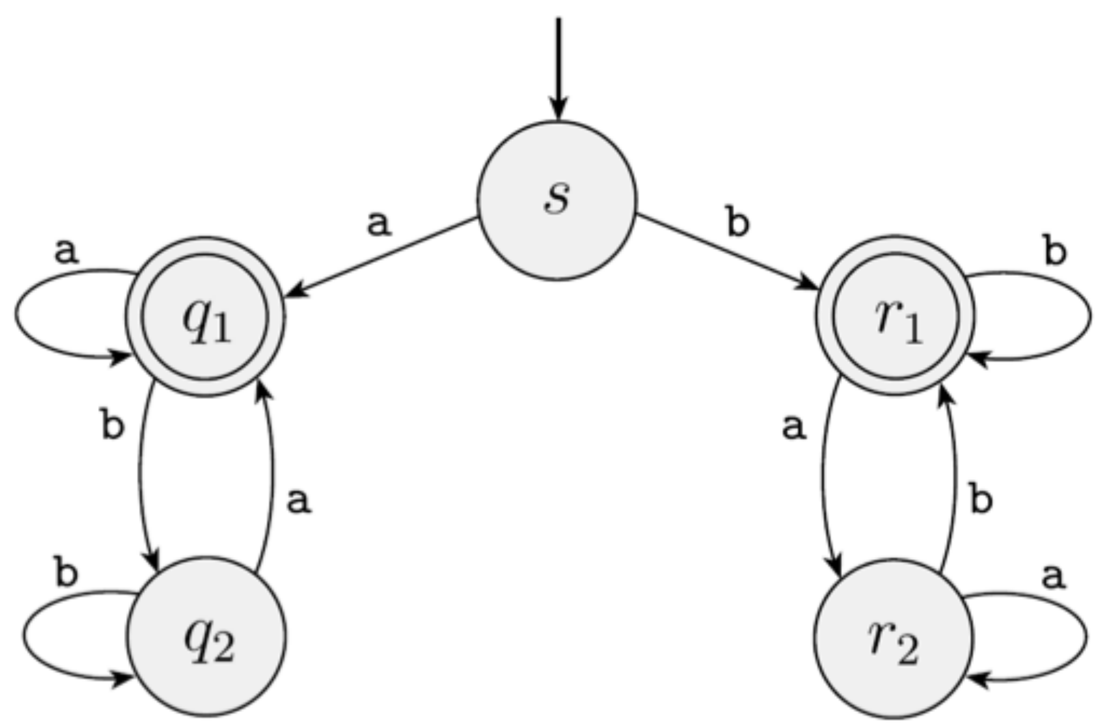
Solution

- State s : ε
- State q_1: all strings that start with a and end with a: $a\Sigma^*a$
- State q_2: all strings that start with a and end with b: $a\Sigma^*b$
- State r_1: all strings that start with b and end with b: $b\Sigma^*b$
- State r_2: all strings that start with b and end with a: $b\Sigma^*a$



Understanding the Partitions

- These five classes partition Σ^* : $\varepsilon, a\Sigma^*a, a\Sigma^*b, b\Sigma^*b, b\Sigma^*a$
- All strings in Σ^* is in exactly one of the these classes
- Union of these classes covers Σ^*
- Intuitively, to decide this language, we only must be able to distinguish between exactly these five cases



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 $xz \in L \iff yz \in L$

Observation: $\equiv_L$ is an equivalence relation over Σ^*

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

Distinguishing Suffixes

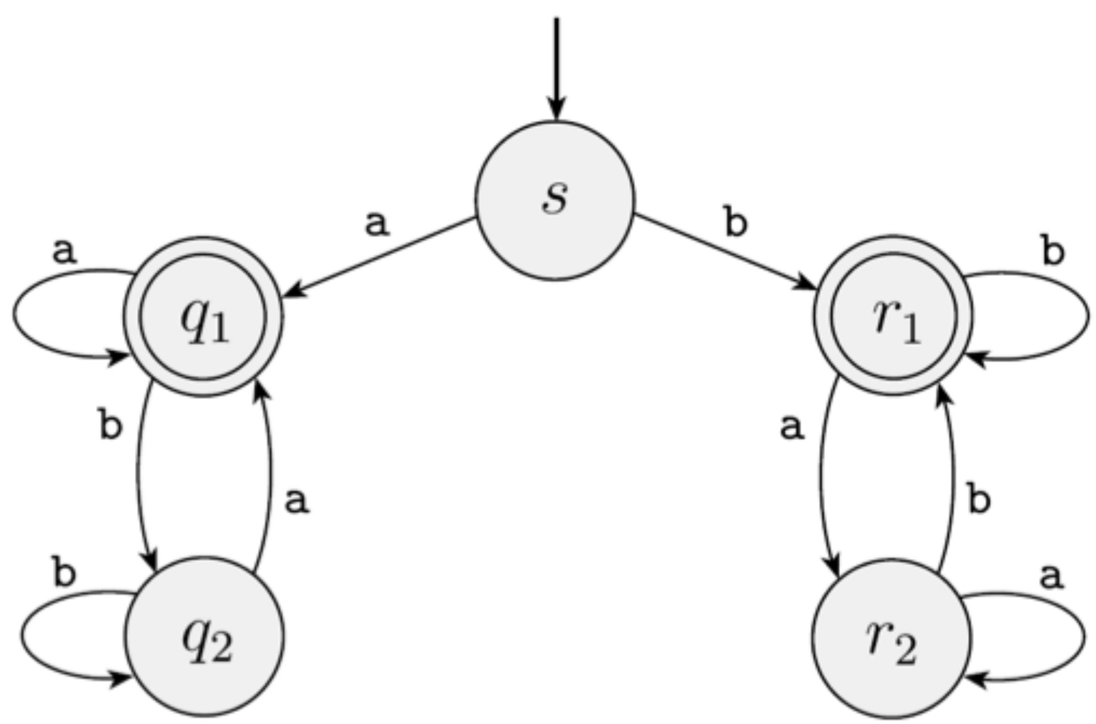
- Every string in the same equivalence class $[x]$ of $\equiv_L$ are indistinguishable with each other
- Two strings $x, y \in \Sigma^*$ are in different equivalence iff they are ***distinguishable***
  - Can find a suffix $z \in \Sigma^*$ that distinguishes them, that is, $xz \in L$ and $yz \notin L$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

Indistinguishability (Languages)

- **Example.**

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

- **Problem.** Find the equivalence classes of the relation $\equiv_L$.



Indistinguishability (Languages)

- **Example.** $L = \{w \in \{0,1\}^* \mid w \text{ ends in } 01\}$
- **Problem.** Find the equivalence classes of the relation $\equiv_L$.
- *Hint*: try to construct a minimal DFA for L and find the classes of strings that map to each state

Indistinguishability DFA vs Languages

- **Observation.** If $x \sim_M y$, then $x \equiv_{L(M)} y$.
- **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.
- How can we prove this?

Minimal DFA

- **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. (**HW 3** question proves this direction.)

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.

Exercises: Proving Non-Regularity

Problem 1. Prove that the language
$L = \{a^n \mid n \in \mathbb{N}$ and n is a power of $2\}$ is not regular.

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

Problem 2. Prove that the language $L = \{ww \mid w \in \{0,1\}^*\}$ is not regular.

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

Problem 3. Prove that the language
$L = \{w \in \{0,1\}^*$ has an equal number of 0s and1s$\}$ is not regular.

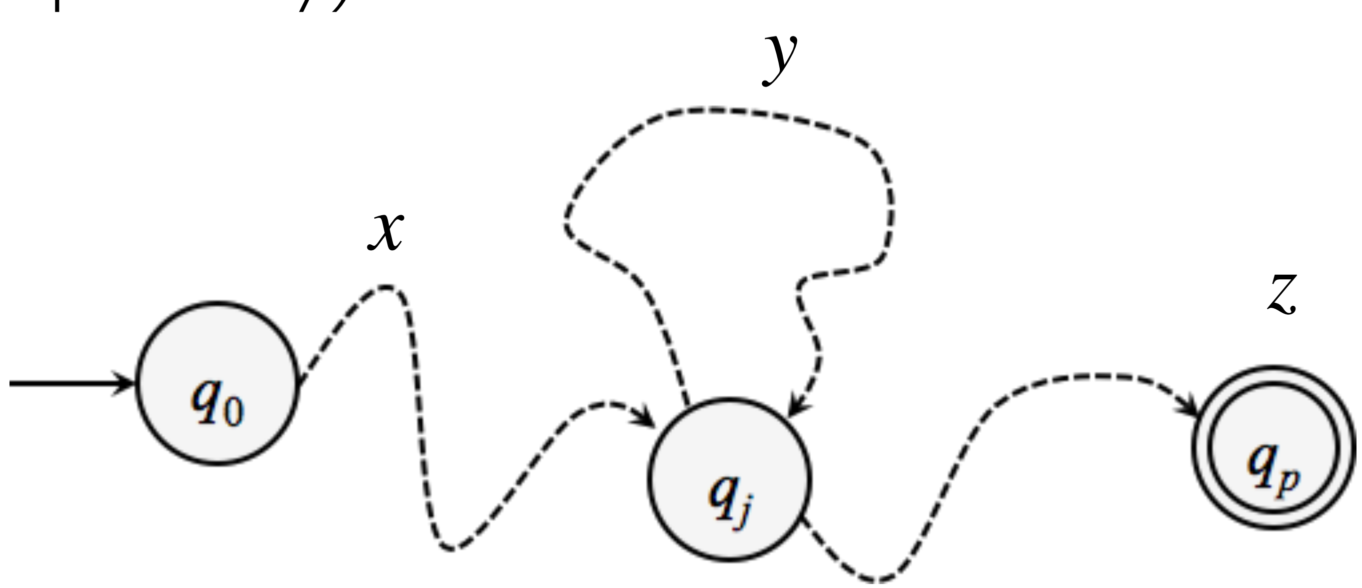
Hint. Use the fact that $L = \{0^i1^i \mid i \in \mathbb{N}\}$ *is not regular and closure properties of regular languages.*

Takeaways: Myhill Nerode

- Powerful characterization of regular languages
- Both upper and lower bound on number of states needed:
  - Can be used to prove that a DFA is minimal
  - Can be used to prove that no DFA exists for a language
- This method does not extend beyond regular languages
- Next method (**pumping lemma**) is weaker but generalizes to the next class of problems we will study

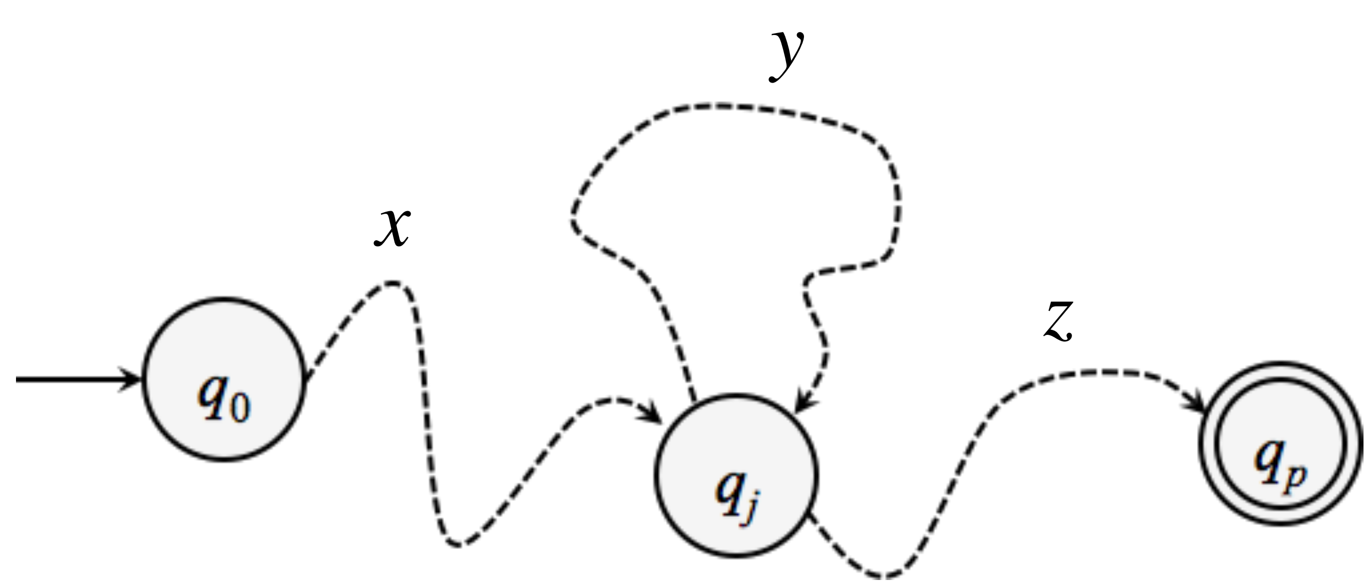
Pumping Lemma: Intuition

- If DFA M has p states then M visits a state more than once on any string with length at least p
  - Number of states visited = length of string $+ 1$
- Let $w = xyz$ be the string that is accepted such that y is component in between the first repeated state (q_j)
  - Then xy^iz should also be accepted (can "pump" the middle piece repeatedly)



Pumping Lemma: Proof

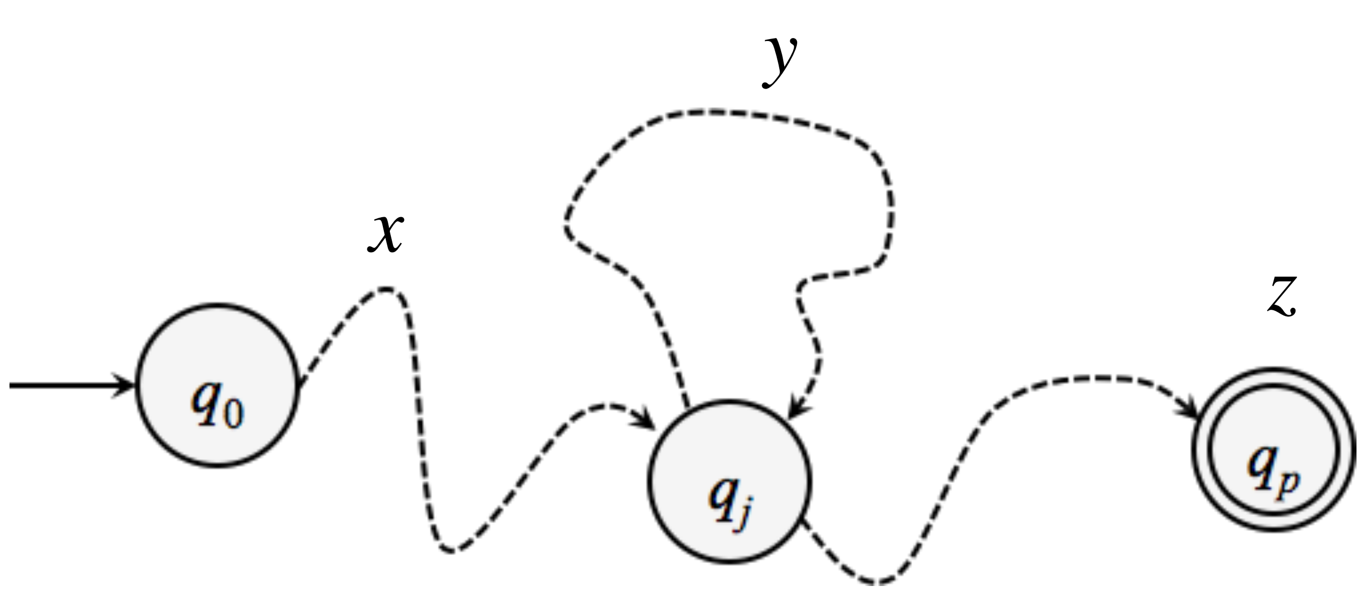
- Consider DFA M for L. Let p be the number of states in M
- Let s be a string of length $n \geq p$
- Then M's computation sequences enters $n + 1$ states on s
- By pigeonhole principle, there must be a repeated state q_j in the first $p + 1$ states of this sequence
- Let x be the substring that brings M from q_0 to first occurrence of q_j



Formal Statement

Pumping Lemma. If L is a regular language, then there exists a number p where if $w \in L$ is any string of length at least p, then w may be divided into three pieces $w = xyz$ such that:

1. $|y| > 0$
2. $|xy| \leq p$ (y must appear amongst the first p symbols)
3. for each $i \geq 0$, $xy^iz \in L$



Pumping Lemma: Game View

- Defender claims L satisfies pumping lemma
- Challenger claims L does not satisfy pumping lemma

Defender		**Challenger**
Pick pumping length p	$\xrightarrow{p}$	
	$\xleftarrow{z}$	Pick $s \in L$ s.t. $\lvert s \rvert \geq p$
Divide s into xyz s.t. $\lvert y \rvert > 0$ and $\lvert xy \rvert \leq p$	$\xrightarrow{x, y, z}$	
	$\xleftarrow{i}$	Pick i, such that $xy^iz \notin L$

Pumping Lemma: Game View

- If L is regular: defender has a winning strategy, challenger gets stuck
- If challenger has a winning strategy, L cannot be regular

Defender		**Challenger**
Pick pumping length p	$\xrightarrow{p}$	
	$\xleftarrow{z}$	Pick $s \in L$ s.t. $\lvert s \rvert \geq p$
Divide s into xyz s.t. $\lvert y \rvert > 0$ and $\lvert xy \rvert \leq p$	$\xrightarrow{x, y, z}$	
	$\xleftarrow{i}$	Pick i, such that $xy^iz \notin L$

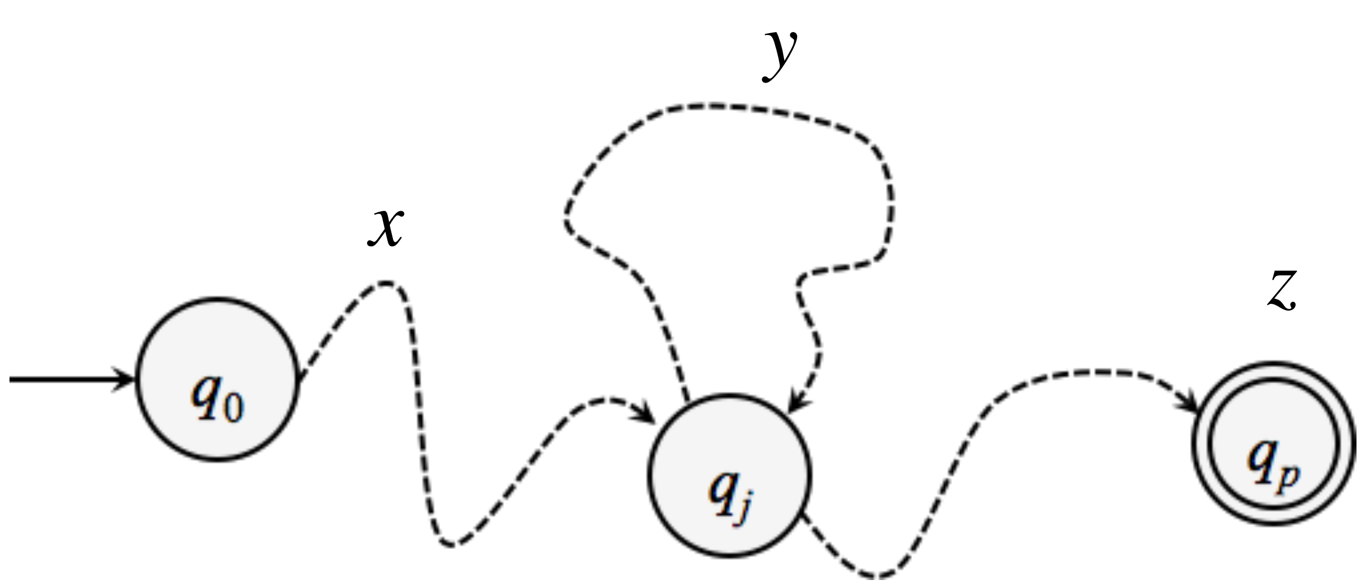
Questions

- Do all regular languages satisfy the pumping lemma?
- If a language satisfies the pumping lemma, does that mean it is regular?

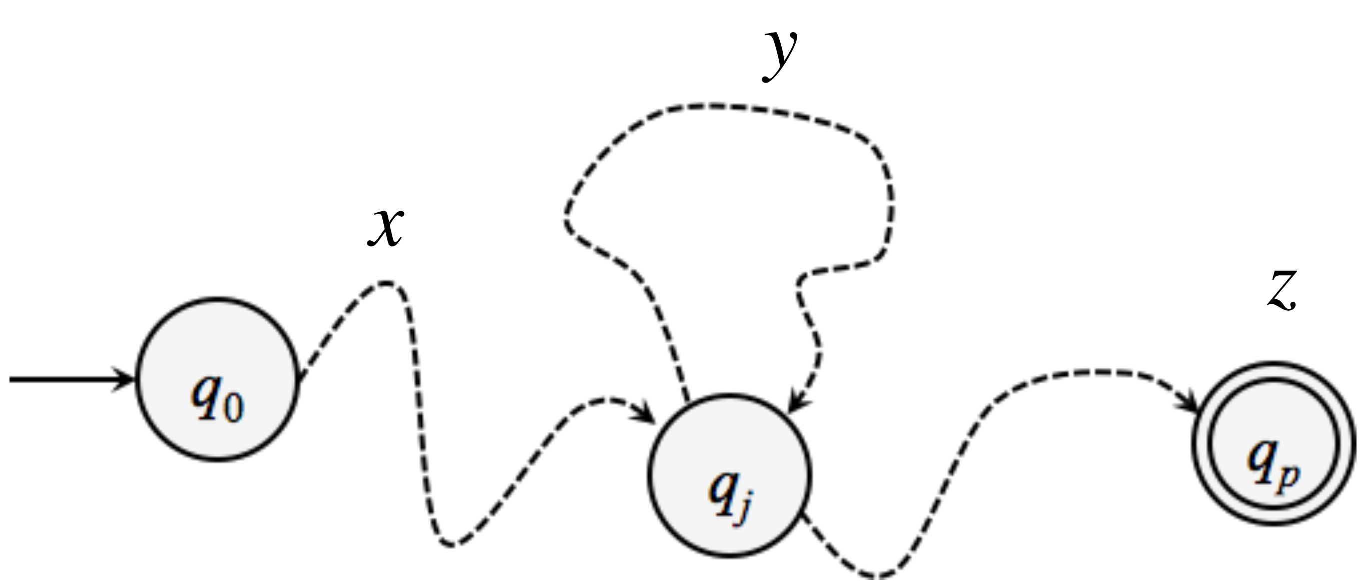
Pumping Lemma Proof

Proof. Let DFA M for L have p states. Let $w = w_1 \cdots w_n$ such that $n \geq p$ and $q_0, q_1, \ldots, q_n$ be the states entered by M on w. M must revisit a state in the first p symbols. Let q_j and q_k be the first and second occurrence of this state.

Let $x = w_1 w_2 \cdots w_{j-1}$, $y = w_j w_{j+1} \cdots w_k$ and $z = w_{k+1} \cdots w_n$ which satisfies the conditions (1) and (2). Condition (3) follows from the fact that the strings xy^i are all ***indistinguishable*** wrt M.




PUMP ALL THE STRINGS!
imgflip.com


y
x
z
q_0
q_j
q_p

Practice: Using Pumping Lemma

Problem 1. Prove that the language $L = \{a^i b^i \mid i \in \mathbb{N}\}$ is not regular using the pumping lemma.

Problem 2. Prove that $L = \{ww \mid w \in \{0,1\}^*\}$ is not regular.

Problem 3. Is the language $L = \{(ab)^i \circ (ab)^i \mid i \geq 0\}$ regular?

Problem 4. Prove that

$L = \{w \mid w \in \{0,1\}^*$ and w has equal number of 1s and 0s$\}$ is not regular using pumping lemma.