CS 383

Lecture 07 – Closure properties of regular languages

Stephen Checkoway

Spring 2024

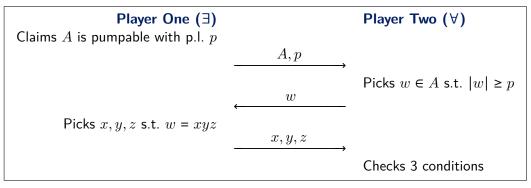
Last time: pumping lemma

Theorem

Pumping lemma for regular languages For every regular language A, there exists an integer p > 0 called the pumping length such that for every $w \in A$ there exist strings x, y, and z with w = xyz such that

- $1 xy^i z \in A \text{ for all } i \ge 0$
- **2** |y| > 0
- $3 |xy| \le p.$

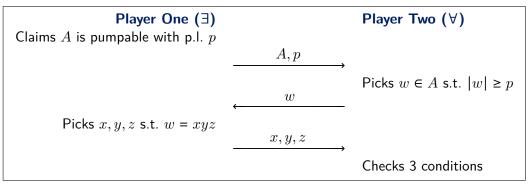
A two-player game



Player One "wins" if Can play as either Player One or Two

xyⁱz ∈ A for all i ≥ 0
 |y| > 0
 |xy| ≤ p

A two-player game



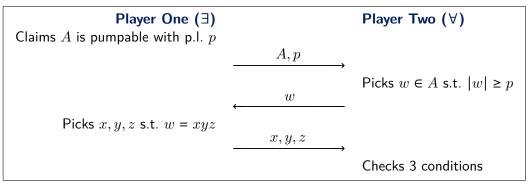
Player One "wins" if

- **1** $xy^i z \in A$ for all $i \ge 0$
- **2** |y| > 0
- $3 |xy| \le p$

Can play as either Player One or Two

• To show that A is pumpable, play as Player One You must consider all possible w and pick x, y, and z

A two-player game



Player One "wins" if

- **1** $xy^i z \in A$ for all $i \ge 0$
- 2 |y| > 0
- $3 |xy| \le p$

Can play as either Player One or Two

- To show that A is pumpable, play as Player One You must consider all possible w and pick x, y, and z
- To show that A is not pumpable, play as Player Two You must pick w and consider all possible x, y, and z

Last time: strategy for proving a language is not regular

To show that A is not regular, we assume it is and then find a string that cannot be "pumped"

Since we don't know the pumping length p, we have to construct a string w that depends on p

E.g., w might contain 0^p or $(aba)^p$

Usually, we want to construct w such that the condition $|xy| \leq p$ constrains the possible choices of x and y

Next, we consider all possible combination of x, y, and z such that $xyz = w, |xy| \le p$ and |y| > 0

Finally, for each combination, we find an $i \ge 0$ such that $xy^i z \notin A$

Duplicated strings

Prove that $A = \{xx \mid x \in \{0,1\}^*\}$ is not regular

Proof.

Assume A is regular with pumping length p.

What string should we pick?

Duplicated strings

Prove that $A = \{xx \mid x \in \{0,1\}^*\}$ is not regular

Proof.

Assume A is regular with pumping length p.

What string should we pick? Let $w = 0^p 1^p 0^p 1^p$.

For
$$x, y, z \in \{0, 1\}^*$$
 such that $xyz = w$, $|xy| \le p$, and $|y| > 0$,
we have $x = 0^m$, $y = 0^n$, and $z = 0^{p-m-n} 1^p 0^p 1^p$
Since $xy^0 z = 0^{p-n} 1^p 0^p 1^p \notin A$, A must not be regular.

An easier method (sometimes)

Assume the language A is regular and apply closure properties of regular languages to arrive at a language that isn't regular

We know regular languages are closed under

- union
- concatenation
- Kleene star
- reversal
- complement
- intersection
- . . .

If we, for example, intersect ${\cal A}$ with a regular language and end up with a nonregular language, then ${\cal A}$ is not regular

Same number of 0 and 1 $\,$

Prove that $B = \{w \mid w \in \{0,1\}^* \text{ and } w \text{ has the same number of 0s as 1s} \}$ is not regular

Proof.

If B is regular, then $B \cap \underline{0^*1^*} = \{0^n 1^n \mid n \ge 0\}$ is regular which is a contradiction so A must not be regular.

Prove that $C = \{xy \mid x, y \in \{0, 1\}^*, |x| = |y|, \text{ and } x \neq y\}$ is not regular Proof. Assume C is regular.

Prove that $C = \{xy \mid x, y \in \{0, 1\}^*, |x| = |y|, \text{ and } x \neq y\}$ is not regular Proof. Assume C is regular.

Then $\overline{C} = \{xx \mid x \in \{0,1\}^*\} \cup \{y \mid |y| \text{ is odd}\}$ is regular.

Prove that $C = \{xy \mid x, y \in \{0, 1\}^*, |x| = |y|, \text{ and } x \neq y\}$ is not regular Proof. Assume C is regular.

Then $\overline{C} = \{xx \mid x \in \{0,1\}^*\} \cup \{y \mid |y| \text{ is odd}\}$ is regular.

Therefore $\overline{C} \cap (\Sigma \Sigma)^* = \{xx \mid x \in \{0,1\}^*\}$ is regular

Prove that $C = \{xy \mid x, y \in \{0, 1\}^*, |x| = |y|, \text{ and } x \neq y\}$ is not regular Proof. Assume C is regular.

Then $\overline{C} = \{xx \mid x \in \{0,1\}^*\} \cup \{y \mid |y| \text{ is odd}\}$ is regular.

Therefore $\overline{C} \cap (\Sigma \Sigma)^* = \{xx \mid x \in \{0,1\}^*\}$ is regular which is a contradiction so C must not be regular.

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = a^k b^m c^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = \mathbf{a}^k \mathbf{b}^m \mathbf{c}^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

1 k = 0 and m > 0. Let y = b and $z = b^{m-1}c^n$. Thus $xy^i z = b^{m+i-1}c^n \in D$

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = a^k b^m c^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

1 k = 0 and m > 0. Let y = b and $z = b^{m-1}c^n$. Thus $xy^i z = b^{m+i-1}c^n \in D$ **2** k = 0 and m = 0. Let y = c and $z = c^{n-1}$. Thus $xy^i z = c^{n+i-1} \in D$

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = a^k b^m c^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

1 k = 0 and m > 0. Let y = b and $z = b^{m-1}c^n$. Thus $xy^i z = b^{m+i-1}c^n \in D$ 2 k = 0 and m = 0. Let y = c and $z = c^{n-1}$. Thus $xy^i z = c^{n+i-1} \in D$

3 k = 1. In this case, m = n. Let y = a and $z = b^n c^n$. Thus $xy^i z = a^i b^n c^n \in D$

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = \mathbf{a}^k \mathbf{b}^m \mathbf{c}^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

1 k = 0 and m > 0. Let y = b and z = b^{m-1}cⁿ. Thus xyⁱz = b^{m+i-1}cⁿ ∈ D
2 k = 0 and m = 0. Let y = c and z = cⁿ⁻¹. Thus xyⁱz = cⁿ⁺ⁱ⁻¹ ∈ D
3 k = 1. In this case, m = n. Let y = a and z = bⁿcⁿ. Thus xyⁱz = aⁱbⁿcⁿ ∈ D
4 k = 2. Since m need not equal n, we need to be careful that pumping down doesn't leave us with one a. Let y = a and z = b^mcⁿ. Thus xyⁱz = a²ⁱb^mcⁿ ∈ D

9/24

 $D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$ is pumpable with pumping length p = 2.

Consider a string $w = a^k b^m c^n \in D$ with $|w| \ge 2$. We need to partition w into xyz = w such that $xy^i z \in D$ for all $i \ge 0$, $|xy| \le 2$, and |y| > 0

There are five cases to consider and in all of them, let $x = \varepsilon$

1 k = 0 and m > 0. Let y = b and z = b^{m-1}cⁿ. Thus xyⁱz = b^{m+i-1}cⁿ ∈ D
2 k = 0 and m = 0. Let y = c and z = cⁿ⁻¹. Thus xyⁱz = cⁿ⁺ⁱ⁻¹ ∈ D
3 k = 1. In this case, m = n. Let y = a and z = bⁿcⁿ. Thus xyⁱz = aⁱbⁿcⁿ ∈ D
4 k = 2. Since m need not equal n, we need to be careful that pumping down doesn't leave us with one a. Let y = aa and z = b^mcⁿ. Thus xyⁱz = a²ⁱb^mcⁿ ∈ D
5 k ≥ 3. Let y = a and z = a^{k-1}b^mcⁿ. Then xyⁱz = a^{k+i-1}b^mcⁿ Since k ≥ 3, k + i - 1 ≥ 2 so it doesn't matter if m = n or not. Thus xyⁱz ∈ D
In each case, |xy| ≤ 2 and |y| > 0. Thus D is pumpable

$$D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\} \text{ is not regular}$$

Proof.

Assume D is regular and intersect with \underline{ab}^*c^* giving the language $E = \{ab^nc^n \mid n \ge 0\}.$

By assumption D is regular so E is regular with pumping length p.

Let $w = ab^{p}c^{p}$ and consider all partitions xyz = w with $|xy| \le p$ and |y| > 0.

If y contains a, then xy^0z does not start with a so it's not in E

If y does not contain a, then $x = ab^m$, $y = b^n$, and $z = b^{p-m-n}c^p$ for some m and n. Therefore, $xy^0z = ab^{p-n}c^p \notin E$.

Therefore E is not regular so D must not be regular.

Let $F = \{0^m 1^n \mid m \neq n\}$ Let $G = \{w \mid w \in \{0, 1\}^* \text{ and } w \text{ has an unequal number of 0s and 1s} \}$ Neither F nor G is regular

Easy proof via closure properties.

Note that $\overline{F} \cap \underline{0^*1^*} = \{0^n 1^n \mid n \ge 0\}$ and $\overline{G} \cap \underline{0^*1^*} = \{0^n 1^n \mid n \ge 0\}$. This is not regular so neither F nor G is regular.

Let $F = \{0^m 1^n \mid m \neq n\}$ Let $G = \{w \mid w \in \{0, 1\}^* \text{ and } w \text{ has an unequal number of 0s and 1s} \}$ Neither F nor G is regular

Hard proof via pumping lemma.

Assume F (resp. G) is regular with pumping length p.

What string w should we pick?

Let $F = \{0^m 1^n \mid m \neq n\}$ Let $G = \{w \mid w \in \{0, 1\}^* \text{ and } w \text{ has an unequal number of 0s and 1s} \}$ Neither F nor G is regular

Hard proof via pumping lemma.

Assume F (resp. G) is regular with pumping length p.

What string w should we pick? Let $w = 0^p 1^{p+p!}$. Consider all partitions of xyz = w such that $|xy| \le p$ and |y| > 0. $x = 0^a$, $y = 0^b$, and $z = 0^{p-a-b} 1^{p+p!}$ for some $a \ge 0$ and b > 0

Let $F = \{0^m 1^n | m \neq n\}$ Let $G = \{w | w \in \{0, 1\}^* \text{ and } w \text{ has an unequal number of 0s and 1s} \}$ Neither F nor G is regular

Hard proof via pumping lemma.

Assume F (resp. G) is regular with pumping length p.

What string w should we pick? Let $w = 0^p 1^{p+p!}$. Consider all partitions of xyz = w such that $|xy| \le p$ and |y| > 0. $x = 0^a$, $y = 0^b$, and $z = 0^{p-a-b} 1^{p+p!}$ for some $a \ge 0$ and b > 0

Set i = p!/b + 1 which is an integer because $b \le p$ so b divides $p! = p \cdot (p - 1) \cdots b \cdots 1$

$$xy^{i}z = 0^{a+i\cdot b+(p-a-b)}1^{p+p!}$$

= $0^{a+(p!+b)+(p-a-b)}1^{p+p!}$
= $0^{p+p!}1^{p+p!}$

Since $xy^i z \notin F$, F is not regular (resp. $xy^i z \notin G$ so G is not regular)

Complement and reversal of nonregular languages

Theorem

The class of nonregular languages is closed under complement and reversal.

Complement and reversal of nonregular languages

Theorem

The class of nonregular languages is closed under complement and reversal.

Proof.

Assume not. That is, assume language L is nonregular but \overline{L} (resp. $L^{\mathcal{R}}$) is regular. Since \overline{L} (resp. $L^{\mathcal{R}}$) is regular and regular languages are closed under complement (resp. reversal), $\overline{(\overline{L})} = L$ (resp. $(L^{\mathcal{R}})^{\mathcal{R}} = L$) is regular. This is a contradiction.

Union, intersection, and star of nonregular languages

Theorem

The class of nonregular languages is not closed under union, intersection, or Kleene star

What steps would you take to prove these?

Union, intersection, and star of nonregular languages

Theorem

The class of nonregular languages is not closed under union, intersection, or Kleene star

What steps would you take to prove these?

Pick concrete, nonregular languages, apply the operation in question, and show that the result is regular

Union of nonregular languages

Proof that the class of nonregular languages is not closed under union. Let $A = \{0^n 1^n \mid n \ge 0\}$. Since nonregular languages are closed under complement, \overline{A} is nonregular.

Since $A \cup \overline{A} = \Sigma^*$ is regular, the class of nonregular languages is not closed under union.

Question 1

Does this mean the union of any two nonregular languages is regular?

Question 1

Does this mean the union of any two nonregular languages is regular? No. If A is nonregular, then $A \cup A = A$ is nonregular.

Operations more generally

 A^* : $A \cup B$: $A \setminus B$ Regular Nonregular A Regular Regular Either Regular Regular Nonregular Either Either Nonregular Either \overline{A} : $A \cap B$: $A \setminus B$ Regular Nonregular A Either Regular Regular Regular Regular Nonregular Either Either Nonregular Nonregular $A^{\mathcal{R}}$: $A \circ B$: $A \setminus B$ Regular Nonregular A Regular Regular Either Regular Regular Nonregular Either Either Nonregular Nonregular

It's worth spending time thinking up examples for the "Either" cases

Prefix, suffix, and quotient

For a language A over Σ , define

PREFIX(A) = {w | for some $x \in \Sigma^*$, $wx \in A$ } SUFFIX(A) = {w | for some $x \in \Sigma^*$, $xw \in A$ }

Prefix, suffix, and quotient

For a language A over Σ , define

PREFIX(A) = {
$$w$$
 | for some $x \in \Sigma^*$, $wx \in A$ }
SUFFIX(A) = { w | for some $x \in \Sigma^*$, $xw \in A$ }

For a string $u \in \Sigma^*$, define right and left quotient of A by u as

$$Au^{-1} = \{w \mid w \in \Sigma^* \text{ and } wu \in A\}$$
$$u^{-1}A = \{w \mid w \in \Sigma^* \text{ and } uw \in A\}$$

Prefix, suffix, and quotient

For a language A over Σ , define

PREFIX(A) = {
$$w$$
 | for some $x \in \Sigma^*$, $wx \in A$ }
SUFFIX(A) = { w | for some $x \in \Sigma^*$, $xw \in A$ }

For a string $u \in \Sigma^*$, define right and left quotient of A by u as

$$Au^{-1} = \{w \mid w \in \Sigma^* \text{ and } wu \in A\}$$
$$u^{-1}A = \{w \mid w \in \Sigma^* \text{ and } uw \in A\}$$

We can generalize to a quotient of A by a language B over Σ

$$A/B = \{w \mid \text{ for some } x \in B, wx \in A\}$$

$$B \setminus A = \{w \mid \text{ for some } x \in B, xw \in A\}$$
 [Note, this is not $B \setminus A$]

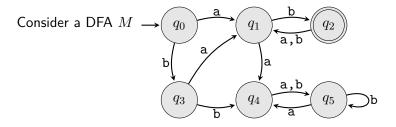
Prefix, suffix, and quotient

Theorem

The class of regular languages is closed under PREFIX, SUFFIX, and quotient.¹

¹We can make a stronger statement: If A is regular and B is any language, then A/B and $B\setminus A$ are regular.

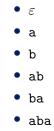
$$PREFIX(A) = \{ w \mid \text{ for some } x \in \Sigma^*, wx \in A \}$$



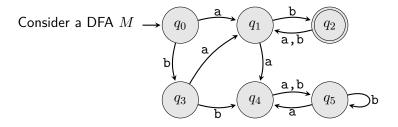
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in PREFIX(L(M))



$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$



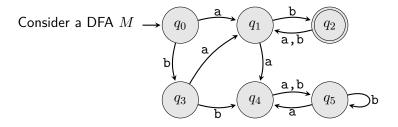
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in PREFIX(L(M))

ε x = ab
a
b
ab
ba
aba

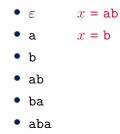
$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$



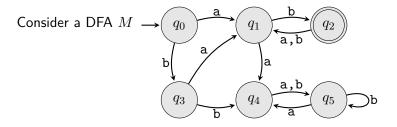
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in $\operatorname{PREFIX}(L(M))$



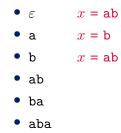
$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$



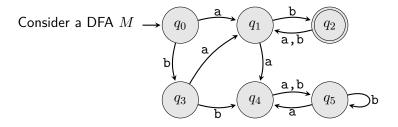
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in PREFIX(L(M))



$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$



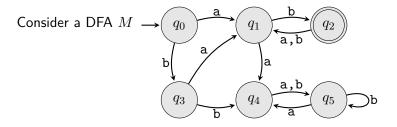
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in $\operatorname{PREFIX}(L(M))$

• ε x = ab• a x = b• b x = ab• ab $x = \varepsilon$ • ba• aba

$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$



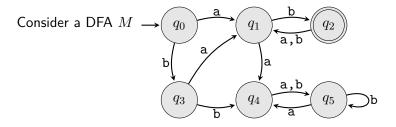
Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in $\operatorname{PREFIX}(L(M))$

• ε x = ab• a x = b• b x = ab• ab $x = \varepsilon$ • ba x = b• aba

$$PREFIX(A) = \{w \mid \text{ for some } x \in \Sigma^*, wx \in A\}$$

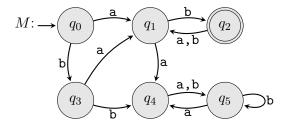


Some strings in L(M)

- ab
- bab
- abab
- babbb
- abbbab
- bababbb

Some strings in PREFIX(L(M))

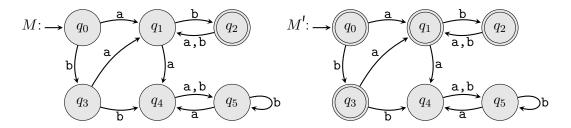
• ε	$x = \mathtt{a}\mathtt{b}$
• a	$x = \mathbf{b}$
• b	$x = \mathtt{a}\mathtt{b}$
• ab	$x = \varepsilon$
• ba	$x = \mathbf{b}$
• aba	x = b



We want to build a new DFA M' s.t. L(M') = PREFIX(L(M))

When M reads string w, it ends in some state q

w is a prefix of some string in ${\cal L}(M)$ if there is some path through M from q to an accept state



We want to build a new DFA M' s.t. L(M') = PREFIX(L(M))

When M reads string w, it ends in some state q

w is a prefix of some string in ${\cal L}(M)$ if there is some path through M from q to an accept state

This suggests a strategy: Build M' from M by making every state with a path to a state in F an accept state

Regular languages are closed under PREFIX

Proof.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A.

Construct $M' = (Q, \Sigma, \delta, q_0, F')$ where $F' = \{q \mid q \in Q \text{ and there is a path from } q \text{ to a state in } F\}$

 2 There may be multiple strings if one of the edges is labeled with multiple symbols (

Regular languages are closed under PREFIX

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A.

Construct $M' = (Q, \Sigma, \delta, q_0, F')$ where $F' = \{q \mid q \in Q \text{ and there is a path from } q \text{ to a state in } F\}$

Consider running M' on string w and ending in some state q

M' accepts $w \iff q \in F' \iff$ there is a path from q to some state in F. Let $x \in \Sigma^*$ be a string² corresponding to that path. Thus $wx \in A$

Therefore, M' accepts $w \iff w \in \text{PREFIX}(A)$ so PREFIX(A) is regular.

²There may be multiple strings if one of the edges is labeled with multiple symbols

Regular languages are closed under (left) quotient by a string

Proof.

We want to show that if A is a regular language over Σ and u is a string over Σ , then $u^{-1}A = \{x \mid x \in \Sigma^* \text{ and } ux \in A\}$ is regular

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A

We want to build an M' that acts on input x just like M does on input ux How should we build M'?

Regular languages are closed under (left) quotient by a string

Proof.

We want to show that if A is a regular language over Σ and u is a string over Σ , then $u^{-1}A = \{x \mid x \in \Sigma^* \text{ and } ux \in A\}$ is regular

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A

We want to build an M' that acts on input x just like M does on input uxHow should we build M'? Let q be the state M ends in after reading input uLet $M' = (Q, \Sigma, \delta, q, F)$. Regular languages are closed under (left) quotient by a string

Proof.

We want to show that if A is a regular language over Σ and u is a string over Σ , then $u^{-1}A = \{x \mid x \in \Sigma^* \text{ and } ux \in A\}$ is regular

Let M = $(Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A

We want to build an M' that acts on input x just like M does on input uxHow should we build M'? Let q be the state M ends in after reading input uLet $M' = (Q, \Sigma, \delta, q, F)$.

If r_0, r_1, \ldots, r_n are the states M goes through on input ux, then $r_{|u|}, r_{|u|+1}, \ldots, r_n$ are the states M' goes through on input x. Thus M' accepts $x \iff M$ accepts ux. \Box

Another proof of nonregularity

$$D = \{a^k b^m c^n \mid \text{if } k = 1, \text{ then } m = n\}$$
 is not regular

Proof.

 $a^{-1}(D \cap \underline{ab^*c^*}) = \{b^n c^n \mid n \ge 0\}$ which is not regular but regular languages are closed under intersection and quotient so D must not be regular.