Proving nonregularity

How can we prove that a language is nonregular?

A. Design a DFA to recognize it and prove that it doesn’t work (that it doesn’t actually recognize the language).
B. Prove that it's a strict subset of some regular language.
C. Prove that it's the union of two regular languages.
D. Prove that its complement is not regular.
E. I don't know.
Where we stand

• There exist nonregular languages.

• If we know that some languages are not regular, we can conclude others are also not regular judiciously reasoning using closure properties of class of regular languages.

• No example of a specific nonregular language ... yet.
Bounds on DFA

• In DFA, memory = states

• Automata can only "remember"...
  • …finitely far in the past
  • …finitely much information

• If a computation path visits the same state more than once, the machine can't tell the difference between the first time and future times it visited that state.
Example!

\{0^n1^n \mid n \geq 0\}

What are some strings in this set?
What are some strings not in this set?

Compare to \(L(0^*1^*)\)
Design a DFA? NFA?
Example!

\[ \{ 0^n1^n \mid n \geq 0 \} \]

What are some strings in this set?
What are some strings not in this set?

Compare to \( L(0^*1^*) \)
Design a DFA? NFA?
Pumping

- Focus on computation path through DFA
Pumping

- Focus on computation path through DFA
Pumping

- Focus on computation path through DFA

Idea: if one long string is accepted, then many other strings have to be accepted too.
Pumping Lemma

If $A$ is a regular language, then there is a number $p$ (the pumping length) where, if $s$ is any string in $A$ of length at least $p$, then $s$ may be divided into three pieces, $s = x \, y \, z$ such that

- $|y| > 0$, and
- for each $i \geq 0$, $xy^i z \in A$,
- $|xy| \leq p$. 

Sipser p. 78 Theorem 1.70
Pumping Lemma

If $A$ is a regular language, then there is a number $p$ (the pumping length) where, if $s$ is any string in $A$ of length at least $p$, then $s$ may be divided into three pieces, $s = x y z$ such that

- $|y| > 0$, and
- for each $i \geq 0$, $xy^i z \in A$,
- $|xy| \leq p$. 

# states in DFA recognizing $A$

Transition labels along loop
Using the Pumping Lemma

Claim: The set \( L = \{0^n1^n \mid n \geq 0\} \) is not regular.
Using the Pumping Lemma

Claim: The set \( L = \{0^n1^n \mid n \geq 0\} \) is not regular.

Proof: Assume, 
\textbf{towards a contradiction}, that \( L \) is regular.

Pumping Lemma gives property of \textbf{all} regular sets. Can we get a contradiction by assuming that the Pumping Lemma applies to this set?
Using the Pumping Lemma

Claim: The set $L = \{0^n1^n \mid n \geq 0\}$ is not regular.

Proof: Assume, towards a contradiction, that $L$ is regular. Therefore, the Pumping Lemma applies to $L$ and gives us some number $p$, the pumping length of $L$. In particular, this means that every string in $L$ that is of length $p$ or more can be "pumped".

...Idea: can we find some long string in $L$ that can't be?
Using the Pumping Lemma

Claim: The set \( L = \{0^n1^n \mid n \geq 0\} \) is not regular.

Proof: ...In particular, this means that every string in \( L \) that is of length \( p \) or more can be "pumped".

Goal: pick a string \( s \) in \( L \) of length greater than or equal to \( p \) such that any division of \( s \) as \( s =xyz \) with \(|y| > 0 \) and \(|xy| \leq p \) gives some value \( i \geq 0 \) with \( xy^iz \) not in \( L \)

So we have a contradiction, and \( L \) is not regular.
Claim: The set \( L = \{0^n1^n | n \geq 0\} \) is not regular.

Proof: …

Goal: pick a string \( s \) in \( L \) of length greater than or equal to \( p \) such that any division of \( s \) as \( s = xyz \) with \(|y| > 0 \) and \(|xy| \leq p \) gives some value \( i \geq 0 \) with \( xy^iz \) not in \( L \).

Choose \( s = 0^p1^p \). Consider any \( s = xyz \) with \(|y| > 0 \), \(|xy| \leq p \).
Using the Pumping Lemma

Claim: The set $L = \{0^n1^n \mid n \geq 0\}$ is not regular.

Proof: …

Goal: pick a string $s$ in $L$ of length greater than or equal to $p$ such that any division of $s$ as $s = xyz$ with $|y| > 0$ and $|xy| \leq p$ gives some value $i \geq 0$ with $xy^iz$ not in $L$.

Choose $s = 0^p1^p$. Consider any $s = xyz$ with $|y| > 0$, $|xy| \leq p$.

Since $|y| > 0$ and $|xy| \leq p$, $y$ is a nonempty string of 0’s.
Claim: The set $L = \{0^n1^n \mid n \geq 0\}$ is not regular.

Proof: ...

Goal: pick a string $s$ in $L$ of length greater than or equal to $p$ such that any division of $s$ as $s = xyz$ with $|y| > 0$ and $|xy| \leq p$ gives some value $i \geq 0$ with $xy^iz$ not in $L$.

Choose $s = 0^p 1^p$. Consider any $s = xyz$ with $|y| > 0$, $|xy| \leq p$. Since $|y| > 0$ and $|xy| \leq p$, $y$ is a nonempty string of 0’s. Picking $i = 0$: Then $xy^iz = xz$ has fewer 0’s than 1’s, not in $L$. 
Using the Pumping Lemma

Claim: The set \( L = \{0^n1^n \mid n \geq 0\} \) is not regular.

Proof: …

Goal: pick a string \( s \) in \( L \) of length greater than or equal to \( p \) such that any division of \( s \) as \( s = xyz \) with \( |y| > 0 \) and \( |xy| \leq p \) gives some value \( i \geq 0 \) with \( xy^i z \) not in \( L \)

Choose \( s = 0^p1^p \). Consider any \( s = xyz \) with \( |y| > 0 \), \( |xy| \leq p \).

Since \( |y| > 0 \) and \( |xy| \leq p \), \( y \) is a nonempty string of 0’s.

Picking \( i = 0 \): Then \( xy^i z = xz \) has fewer 0’s than 1’s, not in \( L \).

This contradicts the Pumping Lemma, so \( L \) must not be regular.
KEEP CALM AND TAKE A STEP BACK
General Structure of Proof

**Claim:** Language L is not regular.

**Proof:**
Assume towards a contradiction L is regular.

*So by Pumping Lemma,* L has a pumping length, call it p.

**FACT:** p is a pumping length for L (by definition).

**CLAIM:** p is not a pumping length for L.

Conclude: contradiction!
Key Ingredients in Proof

**Claim**: Language $L$ is not regular.

**Proof**: Assume, towards a contradiction, that $L$ is regular. By the Pumping Lemma, there is a pumping length $p$ for $L$. **Consider the string** $s = \ldots$ You must pick $s$ carefully: we want $|s| \geq p$ and $s$ in $L$. Now prove a contradiction with the statement "$s$ can be pumped". Consider an arbitrary choice of $x,y,z$ such that $s = xyz$, $|y| > 0$, $|xy| \leq p$. **This means that**... What do you know about $x,y,z$?

**Consider $i=\ldots$** In this case, $xy^iz = \ldots$, which is not in $L$ because ... This contradicts the Pumping Lemma, so $L$ is not regular.
"P is not a pumping length for L"
General Structure of Proof

**Claim:** Language L is not regular.

**Proof:**

**CLAIM:** \( p \) is not a pumping length for L.

To prove claim: Prove that

\[
\exists w \left( |w| \geq p \land w \in L \land \forall x \forall y \forall z \left( (w = xyz \land |y| > 0 \land |xy| \leq p) \rightarrow \exists i (xy^iz \not\in L) \right) \right)
\]
Another example

Claim: The set \( \{a^mb^na^n | m,n \geq 0 \} \) is not regular.

Proof: ...Consider the string \( s = \ldots \) You must pick \( s \) carefully: we want \( |s| \geq p \) and \( s \) in \( L \). Now prove a contradiction with the statement "\( s \) can be pumped"

Which choices of \( s \) cannot be used to complete the proof?

A. \( s = a^p b^p \)  
B. \( s = ab^p a \)  
C. \( s = a^p b^p a^p \)  
D. \( s = a^p b a^p \)  
E. None of the above (all of these choices work).
Another example

Claim: The set \( \{a^m b^m a^n \mid m, n \geq 0\} \) is not regular.

Proof: Consider the string \( s = a^p b a^p \) Check: \(|s| \geq p \) and \( s \) in \( L \).

Now prove a contradiction with the statement "s can be pumped".

Consider an arbitrary choice of \( x, y, z \) such that \( s = xyz \), \(|y| > 0\), \(|xy| \leq p\). This means that \( y \) is a nonempty string of \( a \)'s.

Consider \( i = 2 \). In this case, \( xy^i z = xy^2 z \), which is not in \( L \) because it has more \( a \)'s before the \( b \) than after the \( b \).

This contradicts the Pumping Lemma, so \( L \) is not regular.
And another

Claim: The set \{w w^R \mid w \text{ is a string over } \{0,1\} \} is not regular.

Proof: ...Consider the string \(s = \ldots\). You must pick \(s\) carefully: we want \(|s| \geq p\) and \(s\) in \(L\). Now prove a contradiction with the statement "\(s\) can be pumped".

Which \(s\) and \(i\) let us complete the proof?

A. \(s = 0^p0^p, i=2\)  
B. \(s = 0110, i=0\)  
C. \(s = 0^p110^p, i=1\)  
D. \(s = 1^p001^p, i=3\)  
E. I don't know
How do we choose i?

Claim: The set \( \{0^j1^k \mid j,k \geq 0 \text{ and } j \geq k \} \) is not regular.

Proof: ...Consider the string \( s = \ldots \). You must pick \( s \) carefully: we want \( |s| \geq p \) and \( s \) in \( L \). Now prove a contradiction with the statement "\( s \) can be pumped".

Which \( s \) and \( i \) let us complete the proof?

A. \( s = 0^p1^p, i=2 \)  
B. \( s = 0^p1^p, i=p \)  
C. \( s = 0^p1^p, i=1 \)  
D. \( s = 0^p1^p, i=0 \)  
E. I don't know
Regular sets: not the end of the story

• Many **nice / simple / important** sets are not regular
• Limitation of the finite-state automaton model
  • Can't "count"
  • Can only remember finitely far into the past
  • Can't backtrack
  • Must make decisions in "real-time"
• We know computers are more powerful than this model…

*Which conditions should we relax?*