Today's learning goals

- Decide whether or not a string is described by a given regular expression
- Design a regular expression to describe a given language
- Convert between regular expressions and automata
- Justify the existence of nonregular sets

Sipser Ch 1.1 – 1.3
Regular languages

To prove that a set of strings over the alphabet $\Sigma$ is regular,

• Build a DFA whose language is this set.
• Build an NFA whose language is this set.
• Use the closure properties of the class of regular languages to construct this set from others known to be regular.
  • Union
  • Intersection
  • Complementation
  • Concatenation
  • Flip bits
  • Kleene star
Inductive application of closure

R is a regular expression over $\Sigma$ if

1. $R = a$, where $a \in \Sigma$
2. $R = \varepsilon$
3. $R = \emptyset$
4. $R = (R_1 \cup R_2)$, where $R_1, R_2$ are themselves regular expressions
5. $R = (R_1 \circ R_2)$, where $R_1, R_2$ are themselves regular expressions
6. $(R_1^*)$, where $R_1$ is a regular expression.

$\Sigma$ is shorthand for $(0 \cup 1)$ if $\Sigma = \{0, 1\}$, Parentheses may be omitted, $R^*$ means $RR^*$, $R^k$ means $R$ concatenated with itself $k$ times

Watch out for overloaded symbols!
Syntax $\rightarrow$ Languages

The language described by a regular expression, $L(R)$:

- $L\left(\Sigma^*\Sigma^*\Sigma^*\Sigma^*\right) = \{\}$
- $L\left(1^*\emptyset\emptyset\right) = \{\}$
- $L\left(\epsilon\epsilon\right) = \{\}$

1. $R = a$, where $a \in \Sigma$
2. $R = \epsilon$
3. $R = \emptyset$
4. $R = (R_1 \cup R_2)$
5. $R = (R_1 \circ R_2)$
6. $(R_1^*)$
Which of the following strings is not in the language described by

\[ ( ((00)^*(11)) \cup 01 )^* \]

A. 00
B. 01
C. 1101
D. \( \varepsilon \)
E. I don't know
Let L be the language over \{a, b\} described by the regular expression

$$((a \cup \emptyset) b^*)^*$$

Which of the following is not true about L?

A. Some strings in L have equal numbers of a's and b's
B. L contains the string aaaaaaa
C. a's never follow b's in any string in L
D. L can also be represented by the regular expression \((ab^*)^*\)
E. More than one of the above.
Regular expressions in practice

• **Compilers**: first phase of compiling transforms Strings to Tokens *keywords, operators, identifiers, literals*
  • One regular expression for each token type

• **Other software tools**: grep, Perl, Python, Java, Ruby, …
"Regular = regular"

Theorem: A language is regular if and only if some regular expression describes it.

Lemma 1.55: If a language is described by a regular expression, then it is regular.

Lemma 1.60: If a language is regular, then it is described by some regular expression.
L(R) to NFA (to DFA)

- Idea: basic regular expressions are easy to implement as DFA, for inductive step of definition, use closure under regular operations.

- E.g.: build NFA recognizing the language described by 
  \((00 \cup 11)^*\)
DFA to regular expression

- Idea: use intermediate model **GNFA** whose labels are regular expressions

- E.g.: build regular expression describing language recognized by
All roads lead to … regular sets?

Are there any languages over \{0,1\} that are not regular?

A. Yes: a language that is recognized by an NFA but not any DFA.

B. Yes: there is some infinite language of strings over \{0,1\} that is not described by any regular expression.

C. No: all languages over \{0,1\} are regular because that's what it means to be a language.

D. No: for each set of strings over \{0,1\}, some DFA recognizes that set.

E. I don't know.
Counting

• **Fact:** a countable union of countable sets is countable.

• **Fact:** \(\{0,1\}^*\) is countably infinite. \(X^*\) is countably infinite when \(X\) is finite.

• **Fact:** the set of subsets of a countably infinite sets is uncountable.

• **Fact:** there are countably many DFA with \(\Sigma=\{0,1\}\)

• **Fact:** there are countably many regular languages over \(\{0,1\}\)
Counting

- Fact: a countable union of countable sets is countable.
- Fact: \({\{0,1\}}^*\) is countably infinite. \(X^*\) is countably infinite when \(X\) is finite.
- Fact: the set of subsets of a countably infinite sets is uncountable.
- Fact: there are countably many DFA with \(\Sigma = \{0,1\}\)
- Fact: there are countably many regular languages over \(\{0,1\}\)

Uncountably many languages over \(\{0,1\}\)

Countably many regular languages over \(\{0,1\}\)
Proving nonregularity

How can we prove that a set is non-regular?
A. Try to design a DFA that recognizes it and, if the first few attempts don't work, conclude there is none that does.
B. Prove that it's a strict subset of some regular set.
C. Prove that it's the union of two regular sets.
D. Prove that its complement is not regular.
E. I don't know.
Where we stand

• There exist non-regular sets.

• If we know that some sets 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 regular set ... 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?
# of states and lengths of words

What is the length of the shortest word accepted by this DFA?

What is the length of the longest word accepted by this DFA?

What is the language of this DFA?
What is the length of the longest word accepted by this DFA without visiting any state more than once?

A. 0
B. 1
C. 2
D. 3
E. 4
What feature do computation paths of longer words that are accepted by the DFA have?

A. They must visit $q_3$ at least once.
B. They must visit $q_1$ more than once.
C. More than one of the above.
D. None of the above.
E. I don't know.
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 $i \geq 0$, $xy^iz \in A$,
- $|xy| \leq p$. 

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 $i \geq 0$, $xy^i z \in A$,
- $|xy| \leq p$. 

Sipser p. 78 Theorem 1.70

# 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, towards a contradiction, that $L$ is regular.
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 \( L = \{0^n1^n \mid n \geq 0\} \) 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 \( |y| > 0 \), \( |xy| \leq p \).
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 $|xy| \leq p$, $x = 0^m$, $y = 0^n$, $z = 0^r1^p$ with $m+n+r = p$, $j > 0$. 

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 $|xy| \leq p$, $x = 0^m$, $y = 0^n$, $z = 0^r1^p$ with $m + n + r = p$, $j > 0$.

Picking $i = 0$: $xy^iz = xz = 0^m0^r1^p = 0^{m+r}1^p$, 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 $xy^iz$ not in $L$

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

Since $|xy| \leq p$, $x = 0^m$, $y = 0^n$, $z = 0^r1^p$ with $m+n+r = p$, $j > 0$.

Picking $i = 0$: $xy^iz = xz = 0^m0^r1^p = 0^{m+r}1^p$, not in $L$! This is a contradiction with the Pumping Lemma applied to $L$, so $L$ must not be regular.
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 = ....... You must pick s carefully: we want |s|≥p and s in L. *Confirm these facts as part of your proof*

Now we will 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| ≤p. **This means that...** What properties are guaranteed about x,y,z?

Consider i=... In this case, xy^iz = ...., which is not in L, a contradiction with the Pumping Lemma applying to L and so L is not regular.
For next time

Homework 3 due tonight (solutions released soon after)
Review session tomorrow 8pm-10pm PETER 108

Exam 1: Thursday!