Lecture 3:

- Deterministic Finite Automata
- Non-deterministic FAs

Reading:
Sipser Ch 1.1-1.2

Islam’s OH
T 5-6
Th 4-5

Mark Bun
September 9, 2021
Last Time

• Parts of a theory of computation: Model for machines, model for problems, theorems relating machines and problems

• Strings: Finite concatenations of symbols
• Languages: Sets $L$ of strings
• Computational (decision) problem: Given a string $x$, is it in the language $L$?
Deterministic Finite Automata
A (Real-Life?) Example

• **Example**: Kitchen scale
• $P =$ Power button ($\text{ON} / \text{OFF}$)
• $U =$ Units button (cycles through $g / oz / lb$)
  Only works when scale is $\text{ON}$, but units remembered when scale is $\text{OFF}$
• Starts $\text{OFF}$ in $g$ mode

• A computational problem: Does a sequence of button presses in $\{P, U\}^*$ leave the scale $\text{ON}$ in $oz$ mode?
Machine Models

• **Finite Automata (FAs):** Machine with a finite amount of unstructured memory

<table>
<thead>
<tr>
<th>Input</th>
<th>P</th>
<th>U</th>
<th>P</th>
<th>U</th>
<th>...</th>
</tr>
</thead>
</table>

Control scans left-to-right

1) How does the control “start”?

2) What are the different “states” that the control can be in?

3) When the control reads an input character, how does it transition between states?

4) How do I know if I’m in the desired state at the end?
A DFA for the Kitchen Scale Problem
A DFA Recognizing Parity

The language recognized by a DFA is the set of inputs on which it ends in an “accept” state

Parity: Given a string consisting of $a$’s and $b$’s, does it contain an even number of $a$’s?

$\Sigma = \{a, b\} \quad L = \{w \mid w \text{ contains an even number of } a\text{'s}\}$

Which state is reached by the parity DFA on input $aabab$?

a) “even”  
b) “odd”
Anatomy of a DFA

- States
- Start state
- "Accept" states or "final" states
- Transitions

$q_0 \rightarrow 0 \rightarrow q_1 \rightarrow 1 \rightarrow q_2 \rightarrow 1 \rightarrow q_3 \rightarrow 0 \rightarrow q_1 \rightarrow 0,1$
Some Tips for Thinking about DFAs

Given a DFA, what language does it recognize?
- Try experimenting with it on short strings. Do you notice any patterns?
- What kinds of inputs cause the DFA to get trapped in a state?

Given a language, construct a DFA recognizing it
- Imagine you are a machine, reading one symbol at a time, always prepared with an answer
- What is the essential information that you need to remember? Determines set of states.
What language does this DFA recognize?

\[ \exists w \mid w \text{ contains substring 0013} \]
Practice!

- Lots of worked out examples in Sipser
- Automata Tutor: https://automata-tutor.model.in.tum.de/
Formal Definition of a DFA

A finite automaton is a 5-tuple \( M = (Q, \Sigma, \delta, q_0, F) \)

- \( Q \) is the set of states
- \( \Sigma \) is the alphabet
- \( \delta: Q \times \Sigma \rightarrow Q \) is the transition function
- \( q_0 \in Q \) is the start state
- \( F \subseteq Q \) is the set of accept states
A DFA for Parity

**Parity:** Given a string consisting of $a$'s and $b$'s, does it contain an even number of $a$'s?

$$\Sigma = \{a, b\} \quad L = \{w \mid w \text{ contains an even number of } a\}'s\}$$

**State set** $Q = \{q_0, q_1, q_2\}$

**Alphabet** $\Sigma = \{a, b\}$

**Transition function** $\delta : Q \times \Sigma \to Q$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_0$</td>
<td>$q_1$</td>
<td>$q_0$</td>
</tr>
<tr>
<td>$q_1$</td>
<td>$q_0$</td>
<td>$q_1$</td>
</tr>
</tbody>
</table>

- **Start state** $q_0$
- **Set of accept states** $F = \{q_0, q_2\}$
Formal Definition of DFA Computation

A DFA $M = (Q, \Sigma, \delta, q_0, F)$ accepts a string $w = w_1w_2 \cdots w_n \in \Sigma^*$ (where each $w_i \in \Sigma$) if there exist $r_0, \ldots, r_n \in Q$ such that

1. $r_0 = q_0$
2. $\delta(r_i, w_{i+1}) = r_{i+1}$ for each $i = 0, \ldots, n - 1$, and
3. $r_n \in F$

$L(M) =$ the language of machine $M$

= set of all strings machine $M$ accepts

$M$ recognizes the language $L(M)$
Example: Computing with the Parity DFA

Let $w = \text{abba}$

Does $M$ accept $w$?

What is $\delta(r_2, w_3)$?

A DFA $M = (Q, \Sigma, \delta, q_0, F)$ accepts a string $w = w_1w_2 \cdots w_n \in \Sigma^*$ (where each $w_i \in \Sigma$) if there exist $r_0, \ldots, r_n \in Q$ such that

1. $r_0 = q_0$
2. $\delta(r_i, w_{i+1}) = r_{i+1}$ for each $i = 0, \ldots, n - 1$, and
3. $r_n \in F$
Regular Languages

**Definition:** A language is **regular** if it is recognized by a DFA

\[ L = \{ w \in \{a, b\}^* \mid w \text{ has an even number of } a's \} \text{ is regular} \]

\[ L = \{ w \in \{0,1\}^* \mid w \text{ contains 001} \} \text{ is regular} \]

\[ L = \{ w \in \{a,b\}^* \mid w = a^n b^n \text{ for } n \geq 0 \} \text{ is not regular} \]

Many interesting problems are captured by regular languages

- NETWORK PROTOCOLS
- COMPILERS
- GENETIC TESTING
- ARITHMETIC
Let $\text{TCPS} = \{ w \mid w \text{ is a complete TCP Session}\}$

**Theorem.** TCPS is regular
Comments:

Are delimited by /* */
Cannot have nested /* */
Must be closed by */
*/ is illegal outside a comment

$\text{COMMENTS} = \{\text{strings over } \{0,1, /, *\} \text{ with legal comments}\}$

Theorem. $\text{COMMENTS}$ is regular.
Genetic Testing

**DNA sequences** are strings over the alphabet \{A, C, G, T\}.

\[
\text{T A G A C A T}
\]

A **gene** \(g\) is a special substring over this alphabet.

\[
\text{g} = \text{G A C}
\]

A **genetic test** searches a DNA sequence for a gene.

Does \(g\) appear in T A G A C A T?

**GENETICTEST** \(_g\) = \{strings over \{A, C, G, T\} containing \(g\) as a substring\}

**Theorem.** GENETICTEST \(_g\) is regular for every gene \(g\).
Arithmetic

LET $\Sigma = \begin{Bmatrix} [ ] , [ ] , [ ] , [ ] , [ ] , [ ] , [ ] , [ ] , [ ] , [ ] , [ ] \end{Bmatrix}$

- A string over $\Sigma$ has three ROWS (ROW_1, ROW_2, ROW_3)
- Each ROW $b_0b_1b_2 ... b_N$ represents the integer
  $$b_0 + 2b_1 + ... + 2^N b_N.$$ 
- Let $ADD = \{ S \in \Sigma^* \mid ROW_1 + ROW_2 = ROW_3 \}$

**Theorem.** ADD is regular.
Non-deterministic Finite Automata
In a DFA, the machine is always in exactly one state upon reading each input symbol.

In a non-deterministic FA, the machine can try out many different ways of reading the same string:
- Next symbol may cause an NFA to “branch” into multiple possible computations.
- Next symbol may cause NFA’s computation to fail to enter any state at all.
Nondeterminism

A Nondeterministic Finite Automaton (NFA) accepts if there exists a way to make it reach an accept state.
Example: Does this NFA accept the string 1100?

Option 1: \[ q_0 \xrightarrow{1} q_0 \xrightarrow{1} q_0 \xrightarrow{0} q_0 \xrightarrow{0} q_0 \]

Option 2: \[ q_0 \xrightarrow{0} q_0 \xrightarrow{0} q_1 \xrightarrow{0} q_2 \xrightarrow{0} q_3 \] (accept)

Option 3: \[ q_0 \xrightarrow{0} q_3 \xrightarrow{1} \perp \] (fail)
Example: Does this NFA accept the string 11?

No: All paths of computation do not accept.
Some special transitions

- Transition for 0
- Transition for 0, 1
- Transition without reading next symbol
- Transition for ε
- Transition for 1