

## Week4 monday

Regular sets are 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 actual computers are more powerful than this model...

The **next** model of computation. Idea: allow some memory of unbounded size. How?

- To generalize regular expressions: **context-free grammars**
- To generalize NFA: **Pushdown automata**, which is like an NFA with access to a stack: Number of states is fixed, number of entries in stack is unbounded. At each step (1) Transition to new state based on current state, letter read, and top letter of stack, then (2) (Possibly) push or pop a letter to (or from) top of stack. Accept a string iff there is some sequence of states and some sequence of stack contents which helps the PDA process the entire input string and ends in an accepting state.

Is there a PDA that recognizes the nonregular language  $\{0^n 1^n \mid n \geq 0\}$ ?



The PDA with state diagram above can be informally described as:

Read symbols from the input. As each 0 is read, push it onto the stack. As soon as 1s are seen, pop a 0 off the stack for each 1 read. If the stack becomes empty and we are at the end of the input string, accept the input. If the stack becomes empty and there are 1s left to read, or if 1s are finished while the stack still contains 0s, or if any 0s appear in the string following 1s, reject the input.

Trace the computation of this PDA on the input string 01.

Trace the computation of this PDA on the input string 011.

A PDA recognizing the set  $\{ \}$  can be informally described as:

Read symbols from the input. As each 0 is read, push it onto the stack. As soon as 1s are seen, pop a 0 off the stack for each 1 read. If the stack becomes empty and there is exactly one 1 left to read, read that 1 and accept the input. If the stack becomes empty and there are either zero or more than one 1s left to read, or if the 1s are finished while the stack still contains 0s, or if any 0s appear in the input following 1s, reject the input.

Modify the state diagram below to get a PDA that implements this description:



## Week4 wednesday

**Definition** A **pushdown automaton** (PDA) is specified by a 6-tuple  $(Q, \Sigma, \Gamma, \delta, q_0, F)$  where  $Q$  is the finite set of states,  $\Sigma$  is the input alphabet,  $\Gamma$  is the stack alphabet,

$$\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon \rightarrow \mathcal{P}(Q \times \Gamma_\epsilon)$$

is the transition function,  $q_0 \in Q$  is the start state,  $F \subseteq Q$  is the set of accept states.

Draw the state diagram and give the formal definition of a PDA with  $\Sigma = \Gamma$ .

Draw the state diagram and give the formal definition of a PDA with  $\Sigma \cap \Gamma = \emptyset$ .

For the PDA state diagrams below,  $\Sigma = \{0, 1\}$ .

Mathematical description of language	State diagram of PDA recognizing language
	$\Gamma = \{\$, \#\}$ 
	$\Gamma = \{\odot, 1\}$ 

$$\{0^i 1^j 0^k \mid i, j, k \geq 0\}$$

*Note: alternate notation is to replace ; with  $\rightarrow$*

## Week3 monday

So far we have that:

- If there is a DFA recognizing a language, there is a DFA recognizing its complement.
- If there are NFA recognizing two languages, there is a NFA recognizing their union.
- If there are DFA recognizing two languages, there is a DFA recognizing their union.
- If there are DFA recognizing two languages, there is a DFA recognizing their intersection.

Our goals for today are (1) prove similar results about other set operations, (2) prove that NFA and DFA are equally expressive, and therefore (3) define an important class of languages.

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N_1$  such that  $L(N_1) = A_1$  and NFA  $N_2$  such that  $L(N_2) = A_2$ , then there is another NFA, let's call it  $N$ , such that  $L(N) = A_1 \circ A_2$ .

**Proof idea:** Allow computation to move between  $N_1$  and  $N_2$  “spontaneously” when reach an accepting state of  $N_1$ , guessing that we've reached the point where the two parts of the string in the set-wise concatenation are glued together.

**Formal construction:** Let  $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$  and  $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$  and assume  $Q_1 \cap Q_2 = \emptyset$ . Construct  $N = (Q, \Sigma, \delta, q_0, F)$  where

- $Q =$
- $q_0 =$
- $F =$
- $\delta : Q \times \Sigma_\varepsilon \rightarrow \mathcal{P}(Q)$  is defined by, for  $q \in Q$  and  $a \in \Sigma_\varepsilon$ :

$$\delta((q, a)) = \begin{cases} \delta_1((q, a)) & \text{if } q \in Q_1 \text{ and } q \notin F_1 \\ \delta_1((q, a)) & \text{if } q \in F_1 \text{ and } a \in \Sigma \\ \delta_1((q, a)) \cup \{q_2\} & \text{if } q \in F_1 \text{ and } a = \varepsilon \\ \delta_2((q, a)) & \text{if } q \in Q_2 \end{cases}$$

*Proof of correctness would prove that  $L(N) = A_1 \circ A_2$  by considering an arbitrary string accepted by  $N$ , tracing an accepting computation of  $N$  on it, and using that trace to prove the string can be written as the result of concatenating two strings, the first in  $A_1$  and the second in  $A_2$ ; then, taking an arbitrary string in  $A_1 \circ A_2$  and proving that it is accepted by  $N$ . Details left for extra practice.*

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N$  such that  $L(N) = A$ , then there is another NFA, let's call it  $N'$ , such that  $L(N') = A^*$ .

**Proof idea:** Add a fresh start state, which is an accept state. Add spontaneous moves from each (old) accept state to the old start state.

**Formal construction:** Let  $N = (Q, \Sigma, \delta, q_1, F)$  and assume  $q_0 \notin Q$ . Construct  $N' = (Q', \Sigma, \delta', q_0, F')$  where

- $Q' = Q \cup \{q_0\}$
- $F' = F \cup \{q_0\}$
- $\delta' : Q' \times \Sigma_\varepsilon \rightarrow \mathcal{P}(Q')$  is defined by, for  $q \in Q'$  and  $a \in \Sigma_\varepsilon$ :

$$\delta'((q, a)) = \begin{cases} \delta((q, a)) & \text{if } q \in Q \text{ and } q \notin F \\ \delta((q, a)) & \text{if } q \in F \text{ and } a \in \Sigma \\ \delta((q, a)) \cup \{q_1\} & \text{if } q \in F \text{ and } a = \varepsilon \\ \{q_1\} & \text{if } q = q_0 \text{ and } a = \varepsilon \\ \emptyset & \text{if } q = q_0 \text{ and } a \in \Sigma \end{cases}$$

*Proof of correctness would prove that  $L(N') = A^*$  by considering an arbitrary string accepted by  $N'$ , tracing an accepting computation of  $N'$  on it, and using that trace to prove the string can be written as the result of concatenating some number of strings, each of which is in  $A$ ; then, taking an arbitrary string in  $A^*$  and proving that it is accepted by  $N'$ . Details left for extra practice.*

**Application:** A state diagram for a NFA over  $\Sigma = \{a, b\}$  that recognizes  $L((a^*b)^*)$ :

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N$  such that  $L(N) = A$  then there is a DFA  $M$  such that  $L(M) = A$ .

**Proof idea:** States in  $M$  are “macro-states” – collections of states from  $N$  – that represent the set of possible states a computation of  $N$  might be in.

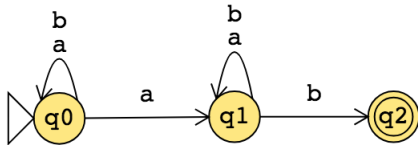
**Formal construction:** Let  $N = (Q, \Sigma, \delta, q_0, F)$ . Define

$$M = ( \mathcal{P}(Q), \Sigma, \delta', q', \{X \subseteq Q \mid X \cap F \neq \emptyset\} )$$

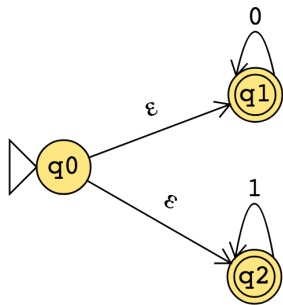
where  $q' = \{q \in Q \mid q = q_0 \text{ or is accessible from } q_0 \text{ by spontaneous moves in } N\}$  and

$\delta'((X, x)) = \{q \in Q \mid q \in \delta(r, x) \text{ for some } r \in X \text{ or is accessible from such an } r \text{ by spontaneous moves in } N\}$

Consider the state diagram of an NFA over  $\{a, b\}$ . Use the “macro-state” construction to find an equivalent DFA.



Consider the state diagram of an NFA over  $\{0, 1\}$ . Use the “macro-state” construction to find an equivalent DFA.



Note: We can often prune the DFAs that result from the “macro-state” constructions to get an equivalent DFA with fewer states (e.g. only the “macro-states” reachable from the start state).



## The class of regular languages

Fix an alphabet  $\Sigma$ . For each language  $L$  over  $\Sigma$ :

**There is a DFA over  $\Sigma$  that recognizes  $L$**   $\exists M$  ( $M$  is a DFA and  $L(M) = A$ )  
*if and only if*

**There is a NFA over  $\Sigma$  that recognizes  $L$**   $\exists N$  ( $N$  is a NFA and  $L(N) = A$ )  
*if and only if*

**There is a regular expression over  $\Sigma$  that describes  $L$**   $\exists R$  ( $R$  is a regular expression and  $L(R) = A$ )

A language is called **regular** when any (hence all) of the above three conditions are met.

We already proved that DFAs and NFAs are equally expressive. It remains to prove that regular expressions are too.

Part 1: Suppose  $A$  is a language over an alphabet  $\Sigma$ . If there is a regular expression  $R$  such that  $L(R) = A$ , then there is a NFA, let's call it  $N$ , such that  $L(N) = A$ .

**Structural induction:** Regular expression is built from basis regular expressions using inductive steps (union, concatenation, Kleene star symbols). Use constructions to mirror these in NFAs.

**Application:** A state diagram for a NFA over  $\{a, b\}$  that recognizes  $L(a^*(ab)^*)$ :

Part 2: Suppose  $A$  is a language over an alphabet  $\Sigma$ . If there is a DFA  $M$  such that  $L(M) = A$ , then there is a regular expression, let's call it  $R$ , such that  $L(R) = A$ .

**Proof idea:** Trace all possible paths from start state to accept state. Express labels of these paths as regular expressions, and union them all.

1. Add new start state with  $\varepsilon$  arrow to old start state.
2. Add new accept state with  $\varepsilon$  arrow from old accept states. Make old accept states non-accept.
3. Remove one (of the old) states at a time: modify regular expressions on arrows that went through removed state to restore language recognized by machine.

**Application:** Find a regular expression describing the language recognized by the DFA with state diagram



## Week2 friday

**Review:** The language recognized by the NFA over  $\{a, b\}$  with state diagram



is:

So far, we know:

- The collection of languages that are each recognizable by a DFA is **closed** under complementation.  
*Could we do the same construction with NFA?*

- The collection of languages that are each recognizable by a NFA is **closed** under union.  
*Could we do the same construction with DFA?*

Happily, though, an analogous claim is true!

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a DFA  $M_1$  such that  $L(M_1) = A_1$  and DFA  $M_2$  such that  $L(M_2) = A_2$ , then there is another DFA, let's call it  $M$ , such that  $L(M) = A_1 \cup A_2$ .  
*Theorem 1.25 in Sipser, page 45*

**Proof idea:**

**Formal construction:**

**Example:** When  $A_1 = \{w \mid w \text{ has an } a \text{ and ends in } b\}$  and  $A_2 = \{w \mid w \text{ is of even length}\}$ .



Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a DFA  $M_1$  such that  $L(M_1) = A_1$  and DFA  $M_2$  such that  $L(M_2) = A_2$ , then there is another DFA, let's call it  $M$ , such that  $L(M) = A_1 \cap A_2$ .  
*Footnote to Sipser Theorem 1.25, page 46*

**Proof idea:**

**Formal construction:**

## Week10 monday

In practice, computers (and Turing machines) don't have infinite tape, and we can't afford to wait unboundedly long for an answer. "Decidable" isn't good enough - we want "Efficiently decidable".

For a given algorithm working on a given input, how long do we need to wait for an answer? How does the running time depend on the input in the worst-case? average-case? We expect to have to spend more time on computations with larger inputs.

A language is **recognizable** if \_\_\_\_\_

A language is **decidable** if \_\_\_\_\_

A language is **efficiently decidable** if \_\_\_\_\_

A function is **computable** if \_\_\_\_\_

A function is **efficiently computable** if \_\_\_\_\_

Definition (Sipser 7.1): For  $M$  a deterministic decider, its **running time** is the function  $f : \mathbb{N} \rightarrow \mathbb{N}$  given by

$$f(n) = \max \text{ number of steps } M \text{ takes before halting, over all inputs of length } n$$

Definition (Sipser 7.7): For each function  $t(n)$ , the **time complexity class**  $TIME(t(n))$ , is defined by

$$TIME(t(n)) = \{L \mid L \text{ is decidable by a Turing machine with running time in } O(t(n))\}$$

An example of an element of  $TIME(1)$  is

An example of an element of  $TIME(n)$  is

Note:  $TIME(1) \subseteq TIME(n) \subseteq TIME(n^2)$

Definition (Sipser 7.12) :  $P$  is the class of languages that are decidable in polynomial time on a deterministic 1-tape Turing machine

$$P = \bigcup_k TIME(n^k)$$

Theorem (Sipser 7.8): Let  $t(n)$  be a function with  $t(n) \geq n$ . Then every  $t(n)$  time deterministic multitape Turing machine has an equivalent  $O(t^2(n))$  time deterministic 1-tape Turing machine.

Definitions (Sipser 7.1, 7.7, 7.12): For  $M$  a deterministic decider, its **running time** is the function  $f : \mathbb{N} \rightarrow \mathbb{N}$  given by

$$f(n) = \max \text{ number of steps } M \text{ takes before halting, over all inputs of length } n$$

For each function  $t(n)$ , the **time complexity class**  $TIME(t(n))$ , is defined by

$$TIME(t(n)) = \{L \mid L \text{ is decidable by a Turing machine with running time in } O(t(n))\}$$

$P$  is the class of languages that are decidable in polynomial time on a deterministic 1-tape Turing machine

$$P = \bigcup_k TIME(n^k)$$

Definition (Sipser 7.9): For  $N$  a nondeterministic decider. The **running time** of  $N$  is the function  $f : \mathbb{N} \rightarrow \mathbb{N}$  given by

$$f(n) = \max \text{ number of steps } N \text{ takes on any branch before halting, over all inputs of length } n$$

Definition (Sipser 7.21): For each function  $t(n)$ , the **nondeterministic time complexity class**  $NTIME(t(n))$ , is defined by

$$NTIME(t(n)) = \{L \mid L \text{ is decidable by a nondeterministic Turing machine with running time in } O(t(n))\}$$

$$NP = \bigcup_k NTIME(n^k)$$

**True or False:**  $TIME(n^2) \subseteq NTIME(n^2)$

**True or False:**  $NTIME(n^2) \subseteq TIME(n^2)$

**Every problem in NP is decidable with an exponential-time algorithm**

Nondeterministic approach: guess a possible solution, verify that it works.

Brute-force (worst-case exponential time) approach: iterate over all possible solutions, for each one, check if it works.

# Week10 wednesday

## Examples in $P$

*Can't use nondeterminism; Can use multiple tapes; Often need to be "more clever" than naïve / brute force approach*

$$PATH = \{\langle G, s, t \rangle \mid G \text{ is digraph with } n \text{ nodes there is path from } s \text{ to } t\}$$

Use breadth first search to show in  $P$

$$RELPRIME = \{\langle x, y \rangle \mid x \text{ and } y \text{ are relatively prime integers}\}$$

Use Euclidean Algorithm to show in  $P$

$$L(G) = \{w \mid w \text{ is generated by } G\}$$

(where  $G$  is a context-free grammar). Use dynamic programming to show in  $P$ .

## Examples in $NP$

*"Verifiable" i.e.  $NP$ , Can be decided by a nondeterministic TM in polynomial time, best known deterministic solution may be brute-force, solution can be verified by a deterministic TM in polynomial time.*

$$HAMPATH = \{\langle G, s, t \rangle \mid G \text{ is digraph with } n \text{ nodes, there is path from } s \text{ to } t \text{ that goes through every node exactly once}\}$$

$$VERTEX - COVER = \{\langle G, k \rangle \mid G \text{ is an undirected graph with } n \text{ nodes that has a } k\text{-node vertex cover}\}$$

$$CLIQUE = \{\langle G, k \rangle \mid G \text{ is an undirected graph with } n \text{ nodes that has a } k\text{-clique}\}$$

$$SAT = \{\langle X \rangle \mid X \text{ is a satisfiable Boolean formula with } n \text{ variables}\}$$

Problems in $P$	Problems in $NP$
(Membership in any) regular language	Any problem in $P$
(Membership in any) context-free language	
$A_{DFA}$	$SAT$
$E_{DFA}$	$CLIQUE$
$EQ_{DFA}$	$VERTEX - COVER$
$PATH$	$HAMPATH$
$RELPRIME$	$\dots$
$\dots$	

Notice:  $NP \subseteq \{L \mid L \text{ is decidable}\}$  so  $A_{TM} \notin NP$

Million-dollar question: Is  $P = NP$ ?

One approach to trying to answer it is to look for *hardest* problems in  $NP$  and then (1) if we can show that there are efficient algorithms for them, then we can get efficient algorithms for all problems in  $NP$  so  $P = NP$ , or (2) these problems might be good candidates for showing that there are problems in  $NP$  for which there are no efficient algorithms.



Definition (Sipser 7.29) Language  $A$  is **polynomial-time mapping reducible** to language  $B$ , written  $A \leq_P B$ , means there is a polynomial-time computable function  $f : \Sigma^* \rightarrow \Sigma^*$  such that for every  $x \in \Sigma^*$

$$x \in A \quad \text{iff} \quad f(x) \in B.$$

The function  $f$  is called the polynomial time reduction of  $A$  to  $B$ .

**Theorem** (Sipser 7.31): If  $A \leq_P B$  and  $B \in P$  then  $A \in P$ .

Proof:

Definition (Sipser 7.34; based in Stephen Cook and Leonid Levin's work in the 1970s): A language  $B$  is **NP-complete** means (1)  $B$  is in NP **and** (2) every language  $A$  in NP is polynomial time reducible to  $B$ .

**Theorem** (Sipser 7.35): If  $B$  is NP-complete and  $B \in P$  then  $P = NP$ .

Proof:

## Week10 friday

### NP-Complete Problems

**3SAT**: A literal is a Boolean variable (e.g.  $x$ ) or a negated Boolean variable (e.g.  $\bar{x}$ ). A Boolean formula is a **3cnf-formula** if it is a Boolean formula in conjunctive normal form (a conjunction of disjunctive clauses of literals) and each clause has three literals.

$$3SAT = \{\langle \phi \rangle \mid \phi \text{ is a satisfiable 3cnf-formula}\}$$

Example string in  $3SAT$

$$\langle (x \vee \bar{y} \vee \bar{z}) \wedge (\bar{x} \vee y \vee z) \wedge (x \vee y \vee z) \rangle$$

Example string not in  $3SAT$

$$\langle (x \vee y \vee z) \wedge (x \vee y \vee \bar{z}) \wedge (x \vee \bar{y} \vee z) \wedge (x \vee \bar{y} \vee \bar{z}) \wedge (\bar{x} \vee y \vee z) \wedge (\bar{x} \vee y \vee \bar{z}) \wedge (\bar{x} \vee \bar{y} \vee z) \wedge (\bar{x} \vee \bar{y} \vee \bar{z}) \rangle$$

**Cook-Levin Theorem**:  $3SAT$  is NP-complete.

*Are there other NP-complete problems?* To prove that  $X$  is NP-complete

- *From scratch*: prove  $X$  is in NP and that all NP problems are polynomial-time reducible to  $X$ .
- *Using reduction*: prove  $X$  is in NP and that a known-to-be NP-complete problem is polynomial-time reducible to  $X$ .

**CLIQUE:** A  $k$ -**clique** in an undirected graph is a maximally connected subgraph with  $k$  nodes.

$$CLIQUE = \{\langle G, k \rangle \mid G \text{ is an undirected graph with a } k\text{-clique}\}$$

Example string in *CLIQUE*

Example string not in *CLIQUE*

Theorem (Sipser 7.32):

$$3SAT \leq_P CLIQUE$$

Given a Boolean formula in conjunctive normal form with  $k$  clauses and three literals per clause, we will map it to a graph so that the graph has a clique if the original formula is satisfiable and the graph does not have a clique if the original formula is not satisfiable.

The graph has  $3k$  vertices (one for each literal in each clause) and an edge between all vertices except

- vertices for two literals in the same clause
- vertices for literals that are negations of one another

Example:  $(x \vee \bar{y} \vee \bar{z}) \wedge (\bar{x} \vee y \vee z) \wedge (x \vee y \vee z)$

Model of Computation	Class of Languages
<p><b>Deterministic finite automata:</b> formal definition, how to design for a given language, how to describe language of a machine? <b>Nondeterministic finite automata:</b> formal definition, how to design for a given language, how to describe language of a machine? <b>Regular expressions:</b> formal definition, how to design for a given language, how to describe language of expression? <i>Also:</i> converting between different models.</p>	<p><b>Class of regular languages:</b> what are the closure properties of this class? which languages are not in the class? using <b>pumping lemma</b> to prove nonregularity.</p>
<p><b>Push-down automata:</b> formal definition, how to design for a given language, how to describe language of a machine? <b>Context-free grammars:</b> formal definition, how to design for a given language, how to describe language of a grammar?</p>	<p><b>Class of context-free languages:</b> what are the closure properties of this class? which languages are not in the class?</p>
<p>Turing machines that always halt in polynomial time</p> <p>Nondeterministic Turing machines that always halt in polynomial time</p>	<p><math>P</math></p> <p><math>NP</math></p>
<p><b>Deciders</b> (Turing machines that always halt): formal definition, how to design for a given language, how to describe language of a machine?</p>	<p><b>Class of decidable languages:</b> what are the closure properties of this class? which languages are not in the class? using diagonalization and mapping reduction to show undecidability</p>
<p><b>Turing machines</b> formal definition, how to design for a given language, how to describe language of a machine?</p>	<p><b>Class of recognizable languages:</b> what are the closure properties of this class? which languages are not in the class? using closure and mapping reduction to show unrecognizability</p>

**Given a language, prove it is regular**

*Strategy 1:* construct DFA recognizing the language and prove it works.

*Strategy 2:* construct NFA recognizing the language and prove it works.

*Strategy 3:* construct regular expression recognizing the language and prove it works.

*“Prove it works” means ...*

**Example:**  $L = \{w \in \{0,1\}^* \mid w \text{ has odd number of 1s or starts with } 0\}$

Using NFA

Using regular expressions

**Example:** Select all and only the options that result in a true statement: “To show a language  $A$  is not regular, we can...”

- a. Show  $A$  is finite
- b. Show there is a CFG generating  $A$
- c. Show  $A$  has no pumping length
- d. Show  $A$  is undecidable

**Example:** What is the language generated by the CFG with rules

$$S \rightarrow aSb \mid bY \mid Ya$$

$$Y \rightarrow bY \mid Ya \mid \varepsilon$$

**Example:** Prove that the language  $T = \{\langle M \rangle \mid M \text{ is a Turing machine and } L(M) \text{ is infinite}\}$  is undecidable.

**Example:** Prove that the class of decidable languages is closed under concatenation.