## Week9 wednesday

Recall: A is **mapping reducible to** B, written  $A \leq_m B$ , means there is a computable function  $f: \Sigma^* \to \Sigma^*$  such that for all strings x in  $\Sigma^*$ ,

$$x \in A$$

if and only if

$$f(x) \in B$$
.

True or False:  $\overline{A_{TM}} \leq_m \overline{HALT_{TM}}$ 

True or False:  $HALT_{TM} \leq_m A_{TM}$ .

**Theorem** (Sipser 5.28): If  $A \leq_m B$  and B is recognizable, then A is recognizable.

**Proof**:

Corollary: If  $A \leq_m B$  and A is unrecognizable, then B is unrecognizable.

### Strategy:

- (i) To prove that a recognizable language R is undecidable, prove that  $A_{TM} \leq_m R$ .
- (ii) To prove that a co-recognizable language U is undecidable, prove that  $\overline{A_{TM}} \leq_m U$ , i.e. that  $A_{TM} \leq_m \overline{U}$ .

$$E_{TM} = \{ \langle M \rangle \mid M \text{ is a Turing machine and } L(M) = \emptyset \}$$

Example string in  $E_{TM}$  is \_\_\_\_\_\_\_. Example string not in  $E_{TM}$  is \_\_\_\_\_\_.

 $E_{TM}$  is decidable / undecidable and recognizable / unrecognizable .

 $\overline{E_{TM}}$  is decidable / undecidable and recognizable / unrecognizable .

Claim:  $\underline{\qquad} \leq_m \overline{E_{TM}}$ .

**Proof**: Need computable function  $F: \Sigma^* \to \Sigma^*$  such that  $x \in A_{TM}$  iff  $F(x) \notin E_{TM}$ . Define

F = "On input x,

- 1. Type-check whether  $x=\langle M,w\rangle$  for some TM M and string w. If so, move to step 2; if not, output
- 2. Construct the following machine  $M'_x$ :
- 3. Output  $\langle M'_x \rangle$ ."

### Verifying correctness:

Input string	Output string
$\langle M, w \rangle$ where $w \in L(M)$	
$\langle M, w \rangle$ where $w \notin L(M)$	
x not encoding any pair of TM and string	

## Week9 friday

Recall: A is mapping reducible to B, written $A \leq_n$	$_{i}$ B, means there is a computable function $f: \Sigma^{*} \to \Sigma^{*}$
such that for all strings $x$ in $\Sigma^*$ ,	

 $x \in A \qquad \text{if and only if} \qquad f(x) \in B.$   $EQ_{TM} = \{\langle M, M' \rangle \mid M \text{ and } M' \text{ are both Turing machines and } L(M) = L(M')\}$  Example string in  $EQ_{TM}$  is \_\_\_\_\_\_\_\_. Example string not in  $EQ_{TM}$  is \_\_\_\_\_\_\_.  $EQ_{TM} \text{ is decidable / undecidable and recognizable / unrecognizable .}$   $\overline{EQ_{TM}} \text{ is decidable / undecidable and recognizable / unrecognizable .}$  To prove, show that \_\_\_\_\_\_\_  $\leq_m \overline{EQ_{TM}}$  and that \_\_\_\_\_\_\_  $\leq_m \overline{EQ_{TM}}$ .

### Verifying correctness:

Input string	Output string
$\langle M, w \rangle$ where M halts on w	
$\langle M, w \rangle$ where $M$ loops on $w$	
x not encoding any pair of TM and string	

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.

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

 $f(n) = \max$  number of steps M 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:  $TIME(1) \subseteq TIME(n) \subseteq TIME(n^2)$ 

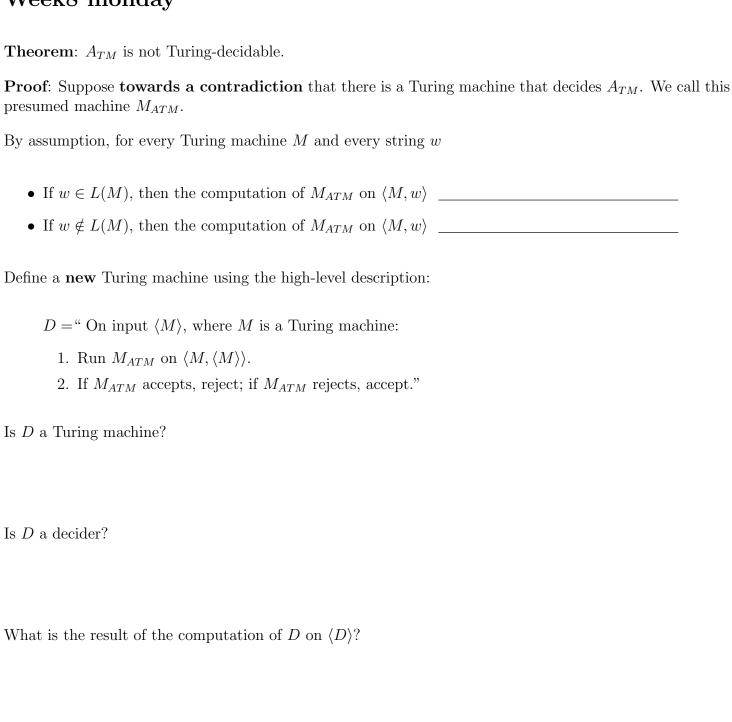
Definition (Sipser 7.12): P is the class of languages that are decidable in polynomial time on a deterministic 1-tape Turing machine

$$P = \bigcup_{k} TIME(n^k)$$

 $Compare\ to\ exponential\ time:\ brute-force\ search.$ 

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

### Week8 monday



**True** or **False**: The class of Turing-decidable languages is closed under complementation?

Definition: A language L over an alphabet  $\Sigma$  is called **co-recognizable** if its complement, defined as  $\Sigma^* \setminus L = \{x \in \Sigma^* \mid x \notin L\}$ , is Turing-recognizable.

Notation: The complement of a set X is denoted with a superscript  $c, X^c$ , or an overline,  $\overline{X}$ .

### Week8 wednesday

### Mapping reduction

Motivation: Proving that  $A_{TM}$  is undecidable was hard. How can we leverage that work? Can we relate the decidability / undecidability of one problem to another?

If problem X is **no harder than** problem Y

- $\dots$  and if Y is easy,
- $\dots$  then X must be easy too.

If problem X is **no harder than** problem Y

- $\dots$  and if X is hard,
- $\dots$  then Y must be hard too.

"Problem X is no harder than problem Y" means "Can answer questions about membership in X by converting them to questions about membership in Y".

Definition: A is **mapping reducible to** B means there is a computable function  $f: \Sigma^* \to \Sigma^*$  such that for all strings x in  $\Sigma^*$ ,

 $x \in A$  if and only if  $f(x) \in B$ .

Notation: when A is mapping reducible to B, we write  $A \leq_m B$ .

Intuition:  $A \leq_m B$  means A is no harder than B, i.e. that the level of difficulty of A is less than or equal the level of difficulty of B.

#### Computable functions

Definition: A function  $f: \Sigma^* \to \Sigma^*$  is a **computable function** means there is some Turing machine such that, for each x, on input x the Turing machine halts with exactly f(x) followed by all blanks on the tape

Examples of computable functions:

The function that maps a string to a string which is one character longer and whose value, when interpreted as a fixed-width binary representation of a nonnegative integer is twice the value of the input string (when interpreted as a fixed-width binary representation of a non-negative integer)

$$f_1: \Sigma^* \to \Sigma^*$$
  $f_1(x) = x0$ 

To prove  $f_1$  is computable function, we define a Turing machine computing it.

High-level description

"On input w

- 1. Append 0 to w.
- 2. Halt."

Implementation-level description

"On input w

- 1. Sweep read-write head to the right until find first blank cell.
- 2. Write 0.
- 3. Halt."

Formal definition ( $\{q0, qacc, qrej\}, \{0, 1\}, \{0, 1, \bot\}, \delta, q0, qacc, qrej$ ) where  $\delta$  is specified by the state diagram:

The function that maps a string to the result of repeating the string twice.

$$f_2: \Sigma^* \to \Sigma^* \qquad f_2(x) = xx$$

The function that maps strings that are not the codes of Turing machines to the empty string and that maps strings that code Turing machines to the code of the related Turing machine that acts like the Turing machine coded by the input, except that if this Turing machine coded by the input tries to reject, the new machine will go into a loop.

$$f_3: \Sigma^* \to \Sigma^* \qquad f_3(x) = \begin{cases} \varepsilon & \text{if } x \text{ is not the code of a TM} \\ \langle (Q \cup \{q_{trap}\}, \Sigma, \Gamma, \delta', q_0, q_{acc}, q_{rej}) \rangle & \text{if } x = \langle (Q, \Sigma, \Gamma, \delta, q_0, q_{acc}, q_{rej}) \rangle \end{cases}$$

where  $q_{trap} \notin Q$  and

$$\delta'((q,x)) = \begin{cases} (r,y,d) & \text{if } q \in Q, \ x \in \Gamma, \ \delta((q,x)) = (r,y,d), \ \text{and} \ r \neq q_{rej} \\ (q_{trap}, \neg, R) & \text{otherwise} \end{cases}$$

The function	on that r	maps strings	that are	not the	codes	of CFGs	to the	empty	string	and that	maps	strings
that code (	CFGs to	the code of a	a PDA tl	hat reco	gnizes t	he langu	iage ge	nerated	l by the	e CFG.		

Other examples?

## Week8 friday

Recall definition: A is **mapping reducible to** B means there is a computable function  $f: \Sigma^* \to \Sigma^*$  such that for all strings x in  $\Sigma^*$ ,

$$x \in A$$
 if and only if  $f(x) \in B$ .

Notation: when A is mapping reducible to B, we write  $A \leq_m B$ .

Intuition:  $A \leq_m B$  means A is no harder than B, i.e. that the level of difficulty of A is less than or equal the level of difficulty of B.

Example:  $A_{TM} \leq_m A_{TM}$ 

Example:  $A_{DFA} \leq_m \{ww \mid w \in \{0, 1\}^*\}$ 

Example:  $\{0^{i}1^{j} \mid i \geq 0, j \geq 0\} \leq_{m} A_{TM}$ 

**Theorem** (Sipser 5.22): If  $A \leq_m B$  and B is decidable, then A is decidable.

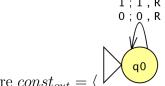
**Theorem** (Sipser 5.23): If  $A \leq_m B$  and A is undecidable, then B is undecidable.

### Halting problem

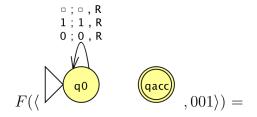
 $HALT_{TM} = \{\langle M, w \rangle \mid M \text{ is a Turing machine, } w \text{ is a string, and } M \text{ halts on } w\}$ 

Define  $F: \Sigma^* \to \Sigma^*$  by

$$F(x) = \begin{cases} const_{out} & \text{if } x \neq \langle M, w \rangle \text{ for any Turing machine } M \text{ and string } w \text{ over the alphabet of } M \\ \langle M', w \rangle & \text{if } x = \langle M, w \rangle \text{ for some Turing machine } M \text{ and string } w \text{ over the alphabet of } M. \end{cases}$$



where  $const_{out} = \langle V, \varepsilon \rangle$  and M' is a Turing machine that computes like M except, if the computation ever were to go to a reject state, M' loops instead.





To use this function to prove that  $A_{TM} \leq_m HALT_{TM}$ , we need two claims:

Claim (1): F is computable

Claim (2): for every  $x, x \in A_{TM}$  iff  $F(x) \in HALT_{TM}$ .

### Week10 monday

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

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

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

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

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

$$P = \bigcup_{k} TIME(n^k)$$

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

 $f(n) = \max$  number of steps N 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**:  $NTIME(n^2) \subseteq DTIME(n^2)$ 

#### Examples in P

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

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

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.

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

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

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

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

### 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.

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	• • •

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.

## Week10 wednesday

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  $f: \Sigma^* \to \Sigma^*$  such that for every  $x \in \Sigma^*$ 

$$x \in A$$
 iff  $f(x) \in B$ .

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

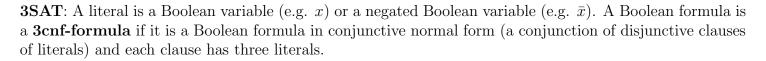
**Theorem** (Sipser 7.31): If  $A \leq_P B$  and  $B \in P$  then  $A \in P$ .

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:



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

Example strings in 3SAT

Example strings not in 3SAT

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.



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

Example strings in CLIQUE

Example strings 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:  $(x \lor \bar{y} \lor \bar{z}) \land (\bar{x} \lor y \lor z) \land (x \lor y \lor z)$ 

# Week10 friday

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

Given	a	language,	prove	it	is	regui	ar
Given	а	ianguage,	prove	16	15	regu	lai

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 t	hat the	class of	decidable	e languag	es is close	ed under	concatena	ation.	