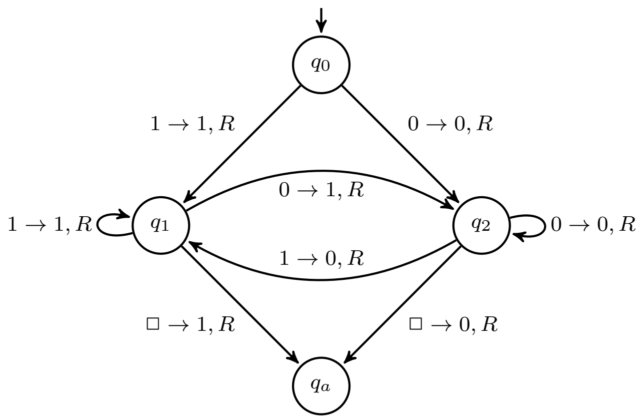
Final

Note

This is a closed-book exam. Each problem accounts for 10 points, unless otherwise marked.

Problems

1. Below is a formal description of a Turing machine that computes a function from $\{0,1\}^*$ to $\{0,1\}^*$. (Here, $\Box$ is used in place of $\sqcup$, because of a typesetting problem.) Explain in words what exactly the machine computes.



2. Show that single-tape TMs that cannot write on the portion of the tape containing the input string recognize only regular languages.

3. Let $\mathit{INFINITE}_{\text{PDA}} = \{\langle M\rangle \mid M \text{ is a PDA and } L(M) \text{ is infinite}\}$. Show that $\mathit{INFINITE}_{\text{PDA}}$ is decidable.

4. Let $\mathit{CONTAIN}_{\text{DFA\_PDA}} = \{\langle M_1, M_2\rangle \mid M_1 \text{ is a DFA and } M_2 \text{ is a PDA such that } L(M_1) \subseteq (M_2)\}$. Show that $\mathit{CONTAIN}_{\text{DFA\_PDA}}$ is undecidable.

5. Let $\mathit{AMBIG}_{\text{CFG}} = \{\langle G\rangle \mid G \text{ is an ambiguous CFG}\}$. Show that $\mathit{AMBIG}_{\text{CFG}}$ is undecidable. (Hint: use a reduction from PCP. Given an instance
$$P = \left\{ \left[\frac{t_1}{b_1}\right], \left[\frac{t_2}{b_2}\right], \cdots, \left[\frac{t_k}{b_k}\right] \right\}$$
of PCP, construct a CFG G with the rules:
$$\begin{array}{lcl} S & \to & T \mid B \\ T & \to & t_1Ta_1 \mid \cdots \mid t_kTa_k \mid t_1a_1 \mid \cdots \mid t_ka_k \\ B & \to & t_1Ba_1 \mid \cdots \mid t_kBa_k \mid t_1a_1 \mid \cdots \mid t_ka_k, \end{array}$$
where $a_1, \ldots, a_k$ are new terminal symbols. Prove that this reduction works.)

6. Discuss the applicability of Rice's Theorem for each of the following problems (languages). Please give the reasons why or why not.

   (a) $\{\langle M \rangle \mid M \text{ is a TM and } 1011 \subseteq L(M)\}$.

   (b) $W\_USELESS_{\mathrm{TM}} = \{\langle M \rangle \mid M \text{ is a TM with useless states}\}$. (A *useless state* in a Turing machine is one that is never entered on any input string.)

7. Prove that $HALT_{\mathrm{TM}} \leq_m \overline{E_{\mathrm{TM}}}$, where $HALT_{\mathrm{TM}} = \{\langle M, w \rangle \mid M \text{ is a TM and } M \text{ halts on } w\}$ and $E_{\mathrm{TM}} = \{\langle M \rangle \mid M \text{ is a TM and } L(M) = \emptyset\}$.

8. Show that NP is closed under *union* and *concatenation*. It is unknown if NP is also closed under *complement*. Can you explain why determining this closure property is hard?

9. In the proof of the Cook-Levin theorem, which states that SAT is NP-complete, we used 2×3 windows of cells to formulate the constraint that the configuration of each row (except the first one) in the $n^k \times n^k$ tableau legally follows the configuration of the preceding row. Consider the machine in Problem 1. Which of the following 2×3 windows of cells are illegal? Why?

q_0	1	0
1	q_1	0

1	q_1	0
1	0	q_2

1	0	q_2
1	0	1

1	0	1
q_2	0	1

1	1	0
1	1	0

1	0	q_a
1	0	$\square$

10. In the proof that the *3SAT* problem is polynomially reducible to the *VERTEX_COVER* problem, we convert an arbitrary boolean expression in 3CNF (input of the *3SAT* problem) to an input graph of the *VERTEX_COVER* problem.

    (a) Please illustrate the conversion by drawing the graph that will be obtained from the following boolean expression:

    $$(x + \overline{y} + z) \cdot (w + \overline{y} + \overline{z}) \cdot (\overline{w} + x + y).$$

    (b) The original boolean expression is satisfiable. As a demonstration of why the reduction is correct, please use the resulting graph to argue that it is indeed the case.

11. (You may substitute this problem from the midterm for one of the previous problems. Please identify the problem, if any, to be replaced. You should not attempt to solve the replaced problem.)

    Prove *by induction* that, if G is a CFG in Chomsky normal form, then for any string $w \in L(G)$ of length $n \geq 1$, exactly $2n - 1$ steps are required for any derivation of w.

Appendix

- $ALL_{\text{CFG}} = \{\langle G\rangle \mid G \text{ is a CFG and } L(G) = \Sigma^*\}$. ALL_{CFG} is undecidable.

- **Rice's Theorem** states that any problem P about Turing machines satisfying the following two conditions is undecidable:

  1. For any TMs M_1 and M_2, where $L(M_1) = L(M_2)$, we have $\langle M_1\rangle \in P$ iff $\langle M_2\rangle \in P$.
  2. P is nontrivial, i.e., there exist TMs M_1 and M_2 such that $\langle M_1\rangle \in P$ and $\langle M_2\rangle \notin P$.

- Language A is **mapping reducible** (many-one reducible) to language B, written $A \leq_m B$, if there is a computable function $f : \Sigma^* \longrightarrow \Sigma^*$, where for every w,

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

- $A \leq_m B$ is equivalent to $\overline{A} \leq_m \overline{B}$.

- $SAT = \{\langle\phi\rangle \mid \phi \text{ is a satisfiable Boolean formula}\}$. SAT is NP-complete (the Cook-Levin theorem).

- $3SAT = \{\langle\phi\rangle \mid \phi \text{ is a satisfiable 3CNF-formula}\}$. (A 3CNF-formula is a CNF-formula where all the clauses have three literals.) $3SAT$ is NP-complete.

- $VERTEX\_COVER = \{\langle G, k\rangle \mid G \text{ is an undirected graph that has a } k\text{-node vertex cover}\}$. (A *vertex cover* of an undirected graph G is a subset of the nodes where every edge of G touches one of those nodes.) $VERTEX\_COVER$ is NP-complete.

- A context-free grammar is in **Chomsky normal form** if every rule is of the form

$$\begin{array}{lcl} A & \rightarrow & BC \quad \text{or} \\ A & \rightarrow & a \end{array}$$

  where a is any terminal and A, B, and C are any variables—except that B and C may not be the start variable. In addition,

$$S \rightarrow \varepsilon$$

  is permitted if S is the start variable.