You may assume without proof any statement proved in class or assigned as homework.

(4 pts) **1** Let k be a fixed positive integer. Construct the smallest possible DFA for the language $L_k = (0 \cup 0^k 1)^*$ over the binary alphabet, and prove that your DFA is the smallest possible.

(3×2 pts) **2** Consider the grammar $S \to SS \mid aS \mid b$.

- **a.** Give an equivalent regular expression.
- **b.** Prove that the grammar is ambiguous.
- **c.** Construct an equivalent unambiguous grammar.

(4 pts) **3** Prove that the language $L \subseteq \{0, 1\}^*$ of strings of prime length is not context-free.

(4 pts) **4** Recall that L^R denotes the set of all strings in L written backwards. Prove that L^R is context-free whenever L is context-free.

(4 pts) **5** Let L^π denote the set of strings that can be obtained by permuting a string in L. For example, $\{\varepsilon, a, 123\}^\pi = \{\varepsilon, a, 123, 132, 213, 231, 312, 321\}$. Prove that context-free languages are not closed under the $^\pi$ operation.

(4 pts) **6** Prove or give a counterexample: if L is given by the context-free grammar (V, Σ, R, S), then L^* is given by the context-free grammar $(V, \Sigma, R \cup \{S \to SS \mid \varepsilon\}, S)$.

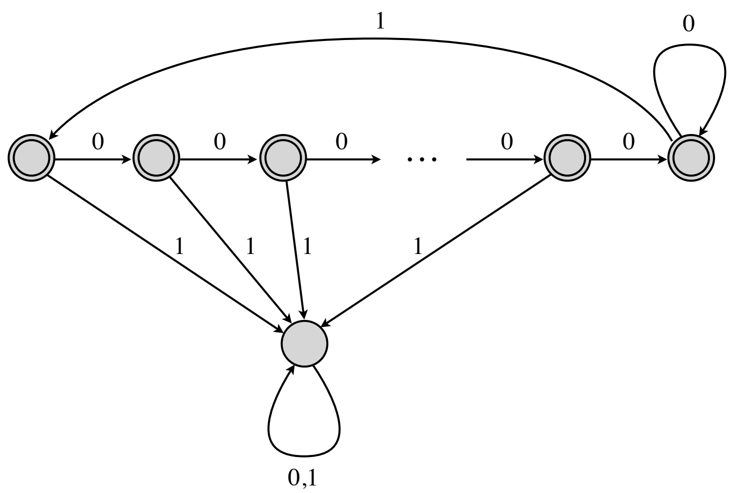
(4 pts) **7** Let D and D' be DFAs with k and k' states, respectively, and input alphabet Σ. Prove that if D and D' agree on every input string of length at most kk', then D and D' recognize the same language. Formally, assume that $w \in L(D) \Leftrightarrow w \in L(D')$ holds for every string w of length at most kk', and prove that $L(D) = L(D')$.

(7×2 pts) **8** Rigorously establish the decidability or undecidability of the following languages:

- **a.** $L = \{\langle D \rangle : D \text{ is a DFA that accepts a palindrome}\}$
- **b.** $L = \{\langle D \rangle : D \text{ is a DFA, and } L(D) \text{ is not recognized by a DFA smaller than } D\}$
- **c.** $L = \{\langle G, k \rangle : G \text{ is a CFG that generates exactly } k \text{ strings}\}$
- **d.** $L = \{\langle G, w \rangle : G \text{ is a CFG that generates a string containing } w \text{ as a substring}\}$
- **e.** $L = \{\langle M \rangle : M \text{ is a Turing machine that accepts } w^R \text{ whenever it accepts } w\}$
- **f.** $L = \{\langle M, k \rangle : M \text{ is a Turing machine that halts within } k \text{ steps on every input}\}$
- **g.** $L = \{\langle M, M' \rangle : M, M' \text{ are Turing machines that recognize the same language}\}$

SOLUTIONS

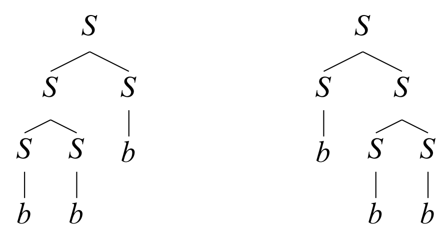
1 L_k is the language of strings in which every 1 is immediately preceded by k or more 0's. It can be recognized by the DFA with $k+2$ states shown below:



We now rule out a smaller DFA. The strings $\varepsilon, 0, 00, 000, \ldots, 0^k$ are pairwise distinguishable because for $i < j$, we have $0^i\boldsymbol{0^{k-j}1} \notin L_k$ but $0^j\boldsymbol{0^{k-j}1} \in L_k$. Moreover, the string $1 \notin L_k$ is by definition distinguishable from $\varepsilon, 0, 00, 000, \ldots, 0^k \in L_k$. Since L_k has at least $k+2$ pairwise distinguishable strings, any DFA for L_k has at least $k+2$ states.

2 **a.** $(a \cup b)^*b$

b. The string bbb has more than one parse tree:



c. $S \to aS \mid bS \mid b$

3 For the sake of contradiction, assume that L is context-free. Let p be any prime larger than the pumping length of L. Since $0^p \in L$, the pumping lemma shows that $0^p = uvxyz$ for some strings u, v, x, y, z such that $|v| + |y| \neq 0$ and $uv^ixy^iz \in L$ for all $i = 0, 1, 2, \ldots$. We arrive at a contradiction because the length of $uv^{p+1}xy^{p+1}z$ is the composite number $p + |v| \cdot p + |y| \cdot p = p(1 + |v| + |y|)$. Thus, L is not context-free.

4 Given a grammar for L, reverse the right-hand side of every rule. To illustrate, $X \to Y_1Y_2 \cdots Y_m$ becomes $X \to Y_mY_{m-1} \cdots Y_1$. The new grammar generates L^R because parse trees in the new grammar are precisely the mirror images of parse trees in the old grammar.

5 The language $L = (abc)^*$ is regular and thus context-free. But L^π is the language of strings with equal numbers of a's, b's, and c's. As proved in class, L^π is not context-free.

6 The claim is false. Consider the grammar $S \to aSb \mid \#$, which generates the language $L = \{a^n\#b^n : n \geqslant 0\}$. The grammar $S \to aSb \mid \# \mid SS \mid \varepsilon$ generates the string $ab \notin L^*$ and thus is not a correct grammar for L^*.

7 Let $L = L(D)$ and $L' = L(D')$. As shown in class, $L \cap \overline{L'}$ has a DFA of size kk'. Since that DFA rejects every string of length at most kk', it rejects all strings (Problem 1.64 from homework). Thus, $L \cap \overline{L'} = \varnothing$. Similarly, $\overline{L} \cap L' = \varnothing$. We conclude that $L = L'$.

8 Parts (a), (c), (d) use the fact that the intersection of a regular language and a CFL is a CFL.

- **a.** Decidable: construct a grammar G for $L(D) \cap \{w : w = w^R\}$ and accept iff $L(G) \neq \varnothing$.
- **b.** Decidable: accept iff $L(D') \neq L(D)$ for each of the (finitely many) DFAs D' smaller than D.
- **c.** Decidable. Let p be the pumping length of G. If $L(G) \cap \Sigma^p \Sigma^* = \varnothing$, we examine every string of length at most p and accept iff exactly k of them are in $L(G)$. If $L(G) \cap \Sigma^p \Sigma^* \neq \varnothing$, we reject right away because by the pumping lemma, $L(G)$ contains a string that can be pumped and hence $L(G)$ is infinite.
- **d.** Decidable: construct a grammar G' for $L(G) \cap \Sigma^* w \Sigma^*$ and accept iff $L(G') \neq \varnothing$.
- **e.** Let $\mathscr{C}$ be the set of Turing-recognizable languages A with $A = A^R$. Then $L = \{\langle M \rangle : L(M) \in \mathscr{C}\}$. Since $\mathscr{C}, \overline{\mathscr{C}} \neq \varnothing$, Rice's theorem shows that L is undecidable.
- **f.** Decidable: simulate M on inputs of length at most k, and verify in each case that M halts within k steps. If this verification succeeds, one need not consider longer inputs because the prefix of length k fully determines M's behavior and in particular prevents M from ever reaching the $(k+1)^{\text{st}}$ cell.
- **g.** If L were decidable, one could also decide the language $A = \{\langle M \rangle : L(M) = \varnothing\}$, by fixing M' to be the Turing machine that disregards its input and enters the reject state right away. By Rice's theorem, A undecidable. Therefore, so is L.