Midterm 1
CS 373: Theory of Computation

Date: Thursday, February 21, 2013.

Instructions:

- This is a closed book exam. No notes, cheat sheets, textbook, or printed material allowed.
- You have 90 minutes to solve this exam.
- This exam has 4 problems. Problems 1 and 4 are worth 10 points, while problems 2 and 3 are worth 15 points. The points are not a measure of the relative difficulty of the problems.
- Please write your name on the top of *every* page in the space provided.
- If your solution does not fit in the space provided, and continues onto one of the back sheets, please indicate clearly where we should look for the solution.
- Unless otherwise stated, "prove that", "show that" for a problem means you need to formally prove what you are claiming.
- You may use, without proof, any result that you were asked to prove in the homework or was proved in the lecture. If you use such a result, please explicitly state the result (like " 'Reverse of regular languages is regular' was proved in a homework", instead of "this was shown in a homework").

Name	SOLUTIONS
Netid	solutions

Discussion:	W 10:00–10:50	W 11:00–11:50	W 12:00–12:50
	W 2:00–2:50	W 3:00–3:50	W 4:00–4:50

Problem	Maximum Points	Points Earned	Grader
1	10		
2	15		
3	15		
4	10		
Total	50		

Problem 1. [Category: Comprehension] **True/False.** Decide for each statement whether it is true or false. Circle **T** if the statement is *necessarily true*; circle **F** if it it is not necessarily true. Each correct answer is worth **1 point**.

(a) For languages $L_1 = L(0^*11^*)$ and $L_2 = L(0^*)$, define $L_1/L_2 = \{w \mid \exists x.\ wx \in L_1 \text{ and } x \in L_2\}$. Then $010 \in L_1/L_2$.

False. L_1/L_2 consists of strings w such that there is some string x from L_2 such that $wx \in L_1$. All strings in L_1 end in 1, and all strings in L_2 end in 0. Thus, $010x$ (for $x \in L_2$ is going to end in 0 and so not be in L_1.

(b) Let Σ and Δ be two alphabets. For a set A, let $|A|$ denote the number of elements in A. Then, $|\Sigma^0| = |\Delta^0|$.

True. No matter what the alphabet Σ is, $\Sigma^0 = \{\epsilon\}$, and so the result holds.

(c) There are regular languages L, such that the smallest GNFA N recognizing L has at least 5 states.

False. Every regular language L (expressed by regular expression r) has a GNFA with exactly two states, where the unique transition (from initial state to final state) is labelled by r.

(d) There is a language L such that there is an NFA recognizing L but no NFA recognizing $\overline{L}$.

False. Regular languages are closed under complementation. Thus if L is regular (i.e., is recognized by an NFA) then $\overline{L}$ is also regular (i.e., is recognized by an NFA).

(e) For any language L, $L^* = L^*L^*$.

True. Since $\epsilon \in L^*$, $L^* = L^*\{\epsilon\} \subseteq L^*L^*$. On the other hand, if $u, v \in L^*$ then $uv \in L^*$. Thus, $L^*L^* \subseteq L^*$.

(f) If L_1 is regular and $L_2 \subseteq L_1$ then L_2 is regular.

False. Take $L_1 = \{0,1\}^*$ and $L_2 = \{0^n1^n \mid n \geq 0\}$.

(g) Let L_1 be a language described by regular expression R_1 and recognized by NFA N_1. Let L_2 be a language described by R_2 and recognized by NFA N_2. If R_1 and R_2 have the same size then N_1 and N_2 have the same number of states.

False. N_1 and N_2 could be have any number of "useless" states. If N_1 and N_2 were constructed using our translation from regular expressions to NFAs (which we are not told), then the result would hold.

(h) Suppose L is a language and h a homomorphism such that $h(L)$ is regular. Since regular languages are closed under inverse homomorphisms, $h^{-1}(h(L)) = L$ must be regular.

False. Take $L = \{0^n1^n \mid n \geq 0\}$ and $h(0) = h(1) = \epsilon$. Then $h(L) = \{\epsilon\}$.

(i) In homework 4, we showed that if L is regular L^R (reverse of L) is regular. This means that if L is not regular then L^R is also not regular.

True. If L^R is regular then $(L^R)^R = L$ is regular.

(j) If L satisfies the pumping lemma then L is regular.

False. See quiz 10 for an example.

Problem 2. [Category: Comprehension+Design+Proof] The language A over alphabet $\{0,1\}$ is defined inductively as follows:

- ϵ is in A
- If x is in A then $10x$ and $11x$ are both in A

(a) For each of the following strings determine if they belong to A.

(i) 1 $\quad$ $1 \notin A$ **[1 point]**

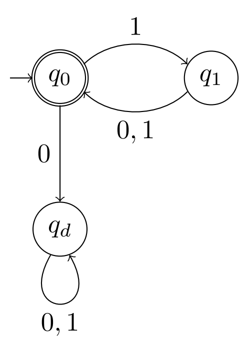
(ii) 1101 $\quad$ $1101 \notin A$ **[1 point]**

(iii) 111010 $\quad$ $111010 \in A$ **[1 point]**

(iv) 111110001010101 $\quad$ $111110001010101 \notin A$ **[1 point]**

(b) Design a DFA with at most 3 states recognizing A. You need not prove the correctness of your construction, but your construction should be clear. **[5 points]**

Observe that $A = \mathbf{L}((10 \cup 11)^*)$. Thus, the 3 states remember if the prefix read so far cannot be extended to a string in A, the input read so far is of even length (and can be the prefix of a string in A), and the string read so far is of odd length (and can be the prefix of a string in A). So the DFA is



(c) Prove that any DFA recognizing A must have at least 3 states. **[6 points]**

The proof, as in any lower bound proof, identifies 3 strings each of which must go to different states. The strings we consider are ϵ, 0 and 1. Let M be any DFA recognizing A with initial state q_0. Let p_ϵ, p_0, and p_1 be the states of M such that $\hat{\delta}_M(q_0, \epsilon) = p_\epsilon$, $\hat{\delta}_M(q_0, 0) = p_0$, and $\hat{\delta}_M(q_0, 1) = p_1$. We will show that p_ϵ, p_0 and p_1 must all be different states.

Case $p_\epsilon \neq p_0$ **and** $p_\epsilon \neq p_1$**:** Observe that $\epsilon \in A$ and $0, 1 \notin A$. Suppose (for contradiction) $p_\epsilon = p_0$ then $q_0 \xrightarrow{\epsilon}_M q_0 = p_\epsilon$. And $q_0 \xrightarrow{0}_M p_0 = p_\epsilon$, which means that either both ϵ and 0 are accepted or both are rejected, which gives us a contradiction. The proof showing $p_\epsilon \neq p_1$ is similar.

Case $p_0 \neq p_1$**:** Observe that $10 \in A$ but $00 \notin A$. Suppose (for contradiction) $p_0 = p_1$. Then, $q_0 \xrightarrow{0}_M p_0 = p_1 \xrightarrow{0}_M p$ (for some p). Also, $q_0 \xrightarrow{1}_M p_1 = p_0 \xrightarrow{0}_M p$ because M is deterministic. Thus, either both 00 and 10 are accepted by M or neither is, giving us the desired contradiction.

Problem 3. [Category: Comprehension+Design+Proof] Recall that $L_1 \setminus L_2 = \{w \mid w \in L_1 \text{ and } w \notin L_2\}$. We can show that regular languages are closed under set difference as follows: Given DFAs M_1 and M_2 recognizing L_1 and L_2, respectively, the DFA for L will run simultaneously both M_1 and M_2 (as in the cross-product construction for intersection) on input w, and accept if M_1 accepts w but M_2 does not.

Complete the following proof of this closure property based on the above intuition. Let $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ and $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$, with $L(M_1) = L_1$ and $L(M_2) = L_2$. The DFA recognizing $L = L_1 \setminus L_2$ is given by $M = (Q, \Sigma, \delta, q_0, F)$, where

(a) $Q = \underline{Q_1 \times Q_2}$ **[1 point]**

(b) $q_0 = \underline{(q_1, q_2)}$ **[1 point]**

(c) $F = \underline{F_1 \times (Q_2 \setminus F_2)}$ **[1 point]**

(d) δ is defined as **[2 points]**

$$\delta((p_1, p_2), a) = (\delta_1(p_1, a), \delta_2(p_2, a))$$

(e) The correctness of this construction can be established by proving **[1 point]**

$$q_1 \xrightarrow{w}_{M_1} p_1 \text{ and } q_2 \xrightarrow{w}_{M_2} p_2 \text{ iff } q_0 \xrightarrow{w}_{M} \underline{(p_1, p_2)}$$

(f) Prove by induction on the length of w, the statement in part (e).

- Prove the base case. **[2 points]**

  When $w = \epsilon$, we know that $q_1 \xrightarrow{\epsilon}_{M_1} q_1$ and $q_2 \xrightarrow{\epsilon}_{M_2} q_2$. Also $q_0 \xrightarrow{\epsilon}_{M} q_0 = (q_1, q_2)$; thus the base case is proved.

- State the induction hypothesis. **[1 point]**

  Suppose

  $$q_1 \xrightarrow{w}_{M_1} p_1 \text{ and } q_2 \xrightarrow{w}_{M_2} p_2 \text{ iff } q_0 \xrightarrow{w}_{M} (p_1, p_2)$$

  for all w such that $|w| \leq n$.

- Prove the induction step. **[4 points]**

  Let $w = ua$, where $|u| = n$ and $a \in \Sigma$. Now $q_1 \xrightarrow{u}_{M_1} p_1 \xrightarrow{a}_{M_1} p'_1$ for some $p_1, p'_1 in Q_1$. Similarly, $q_2 \xrightarrow{u}_{M_2} up_2 \xrightarrow{a}_{M_1} p'_2$ for some $p_2, p'_2 \in Q_2$. By induction hypothesis $q_0 \xrightarrow{u}_M (p_1, p_2)$ and by definition of δ, $(p_1, p_2) \xrightarrow{a}_M (p'_1, p'_2)$. Hence $q_0 \xrightarrow{w}_M (p'_1, p'_2)$.

(g) Prove that $\mathbf{L}(M) = L_1 \setminus L_2$. **[2 points]**

Observe that $w \in \mathbf{L}(M)$ iff $q_0 \xrightarrow{w}_M (p_1, p_2)$ with $(p_1, p_2) \in F$ (defn. of acceptance) iff $q_0 \xrightarrow{w}_M (p_1, p_2)$ with $p_1 \in F_1$ and $p_2 \notin F_2$ (defn. of F) iff $q_1 \xrightarrow{w}_{M_1} p_1$ and $q_2 \xrightarrow{w}_{M_2} p_2$ with $p_1 \in F_1$ and $p_2 \notin F_2$ (due to statement in part (e)) iff $w \in L(M_1)$ and $w \notin L(M_2)$ (defn. of acceptance) iff $w \in L(M_1) \setminus L(M_2) = L_1 \setminus L_2$.

Problem 4. [Category: Proof] Consider the language $B \subseteq \{a,b\}^*$ defined as

$$B = \{babaabaaab \cdots ba^{n-1}ba^nb \mid n \geq 1\}$$

Prove that B is not regular. If needed, you may use the fact (without proof) that the language $\{a^{n^2} \mid n \geq 0\}$ is not regular. **[10 points]**

Observe that the length of the string $z = babaabaaab \cdots ba^{n-1}ba^nb$ is $(n+1) + \frac{n(n+1)}{2}$ because there are $n+1$ bs, and $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ number of a's.

Closure Properties: Consider the homomorphism $h(a) = aa$ and $h(b) = a$. Then $h(babaab \cdots ba^{n-1}ba^nb) = a^{(n+1)+n(n+1)} = a^{(n+1)^2}$. Hence $L_1 = h(B) = \{a^{n^2} \mid n \geq 2\}$. Observe that $L_2 = \{\epsilon, a\}$ is a finite language and hence regular. Finally, $L_1 \cup L_2 = \{a^{n^2} \mid n \geq 0\}$ which we have proved to be not context-free. Hence B is not context-free.

Pumping Lemma: Suppose p is the pumping length. Consider the string $z = babaab \cdots ba^{2p-1}ba^{2p}b \in B$. Let u, v, w be such that $z = uvw$, $|v| \geq 1$ and $|uv| \leq p$.

Consider $z' = uv^2w$. Now

$$\begin{aligned}
(2p+1) + \tfrac{2p(2p+1)}{2} = (2p+1)(p+1) < |z'| \quad &< (2p+1)(p+1) + p \\
&< (2p+1)(p+1) + p + 1 \\
&= (2p+2)(p+1) \\
&< (2p+3)(p+1) = (2p+1+1) + \tfrac{(2p+1)(2p+1+1)}{2}.
\end{aligned}$$

Thus, $z' \notin B$.

Lower Bound proof: Suppose B is regular (for contradiction). Let M with initial state q_0 and transition function δ be some DFA recognizing B. Let $z_n = babaabaaab \cdots ba^{n-1}ba^nb$. We claim that for $i \neq j$, $\hat{\delta}_M(q_0, z_i) \neq \hat{\delta}_M(q_0, z_j)$; if we manage to show that then M has infinitely many states which contradicts the assumption that M is a DFA.

Suppose (for contradiction), there is $i \neq j$ such that $\hat{\delta}_M(q_0, z_i) = \hat{\delta}_M(q_0, z_j) = \{q\}$, for some q. We can assume without loss of generality that $i < j$. Consider the string $u = a^{j+1}b$. Observe that $z_ju \in B$ but $z_iu \notin B$. But, since $\hat{\delta}_M(q_0, z_i) = \hat{\delta}_M(q_0, z_j) = \{q\}$, $\hat{\delta}_M(q_0, z_iu) = \hat{\delta}_M(q, u) = \hat{\delta}_M(q_0, z_ju)$. And so either M accepts both z_iu and z_ju, or it rejects both z_iu and z_ju, which contradicts our assumption that M recognizes B.