Solutions for Problem Set 2
CS 373: Theory of Computation

Assigned: January 24, 2013 Due on: January 31, 2013

Problem 1. [Category: Design+Proof] Let $A_k \subseteq \{a,b\}^*$ be the collection of strings w where there is a position i in w such that the symbol at position i (in w) is a, and the symbol at position $i+k$ is b. For example, consider A_2 (when $k=2$). $baab \in A_2$ because the second position ($i=2$) has an a and the fourth position has a b. On the other hand, $bb \notin A_2$ (because there are no as) and $aba \notin A_2$ (because none of the as are followed by a b 2 positions away).

1. Design a DFA for language A_k. Your formal description (by listing states, transitions, etc. and not "drawing the DFA") will depend on the parameter k but should work no matter what k is; see lecture 2, last page for such an example. **[5 points]**
2. Prove that your DFA is correct when $k=2$. **[5 points]**

Solution:

1. The DFA for A_k will remember the last k symbols read from the input. When a b is read, if the symbol k positions before was an a then the DFA will move to an accept state, where it will stay no matter what the remaining symbols in the input are; this is because once we find a pair of a and b that are k positions apart, the input is in the language no matter what the other symbols are. This intuition is formalized in the following construct of a DFA $M_k = (Q_k, \{a,b\}, \delta_k, q_k, F_k)$ where
   - The set of states $Q_k = \{a,b\}^k \cup \{q_a\}$. That is, a state either remembers the last k symbols read (is a member of $\{a,b\}^k$) or is the state q_a which remembers that the input must be accepted.
   - The initial state is $q_k = bb\cdots b = b^k$
   - The set of final states is $F_k = \{q_a\}$
   - The transition function δ_k is given by

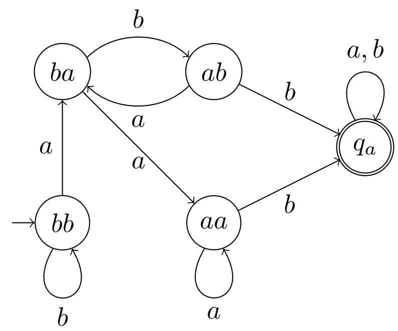
   $$\delta_k(q,c) = \begin{cases} w_2 w_3 \cdots w_k c & \text{if } q = w_1 w_2 \cdots w_k \text{ and either } c = a \text{ or } w_1 \neq a \\ q_a & \text{if } q = a w_2 \cdots w_k \text{ and } c = b \\ q_a & \text{if } q = q_a \end{cases}$$

   Thus, the DFA M_k has $2^k + 1$ states.

2. When $k=2$, the automaton M_2 can be drawn as follows. Given that the initial state is bb and the unique accept state is q_a, to prove correctness we need to show that

   $$\forall w \in \{a,b\}^*.\ bb \xrightarrow{w}_{M_2} q_a \text{ iff } w \in A_2$$

   However, as in other examples we have seen in the lecture notes, this statement needs to be strengthened if the standard proof by induction (on the length of w) is to succeed. The way to strengthen this proof is by characterizing the collection of strings that are accepted from *each* state, and not just the initial state.



Thus what we will prove by induction on the length of w is the following (stronger) statement.

$$\forall w \in \{a,b\}^*. \quad \begin{array}{l} bb \xrightarrow{w}_{M_2} q_a \text{ iff } w \in A_2 \\ ba \xrightarrow{w}_{M_2} q_a \text{ iff } baw \in A_2 \\ ab \xrightarrow{w}_{M_2} q_a \text{ iff } abw \in A_2 \\ aa \xrightarrow{w}_{M_2} q_a \text{ iff } aaw \in A_2 \\ q_a \xrightarrow{w}_{M_2} q_a \text{ iff } w \in \{a,b\}^* \end{array}$$

We will prove the above statement by induction on the length of w.

Base Case: When $w = \epsilon$, $q \xrightarrow{w}_{M_2} q$ for each $q \in Q_2$. Also, $\epsilon \not\in A_2$, $ba\epsilon \not\in A_2$, $ab\epsilon \not\in A_2$, and $aa\epsilon \not\in A$. Thus, when $w = \epsilon$, we have established each of the 5 conditions above.

Ind. Hyp.: We will assume that the statement we are trying to prove holds for all w, such that $|w| < k$. That is, we will assume that

$$\forall w \in \{a,b\}^*. |w| < k, \quad \begin{array}{l} bb \xrightarrow{w}_{M_2} q_a \text{ iff } w \in A_2 \\ ba \xrightarrow{w}_{M_2} q_a \text{ iff } baw \in A_2 \\ ab \xrightarrow{w}_{M_2} q_a \text{ iff } abw \in A_2 \\ aa \xrightarrow{w}_{M_2} q_a \text{ iff } aaw \in A_2 \\ q_a \xrightarrow{w}_{M_2} q_a \text{ iff } w \in \{a,b\}^* \end{array}$$

Ind. Step: Consider w such that $|w| = k$. Now, w can be in one of two forms: either $w = au$, or $w = bu$, where $|u| = k - 1$. We will consider the various cases, and show that the correctness statement we are trying to prove holds in the induction step.

- **Case bb:** First consider $w = au$. Then we have $bb \xrightarrow{w=au}_{M_2} q_a$ iff $ba \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(bb, a) = ba$) iff $bau \in A_2$ (ind. hyp.) iff $au \in A_2$ (definition of A_2) iff $w \in A_2$. Similarly, if $w = bu$ then $bb \xrightarrow{w=bu}_{M_2} q_a$ iff $bb \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(bb, b) = bb$) iff $u \in A_2$ (ind. hyp.) iff $bu \in A_2$ (definition of A_2) iff $w \in A_2$.
- **Case ba:** When $w = au$, we have $ba \xrightarrow{w=au}_{M_2} q_a$ iff $aa \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(ba, a) = aa$) iff $aau \in A_2$ (ind. hyp.) iff $ba(au) \in A_2$ (definition of A_2) iff $baw \in A_2$. Similarly, if $w = bu$ then $ba \xrightarrow{w=bu}_{M_2} q_a$ iff $ab \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(ba, b) = ab$) iff $abu \in A_2$ (ind. hyp.) iff $ba(bu) \in A_2$ (definition of A_2) iff $baw \in A_2$.
- **Case ab:** When $w = au$, we have $ab \xrightarrow{w=au}_{M_2} q_a$ iff $ba \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(ab, a) = ba$) iff $bau \in A_2$ (ind. hyp.) iff $ab(au) \in A_2$ (definition of A_2) iff $abw \in A_2$. Similarly, if $w = bu$ then $ab \xrightarrow{w=bu}_{M_2} q_a$ iff $q_a \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(ab, b) = q_a$) iff $u \in \{a,b\}^*$ (ind. hyp.) iff $ab(bu) \in A_2$ (definition of A_2) iff $abw \in A_2$.

- **Case** aa**:** When $w = au$, we have $aa \xrightarrow{w=au}_{M_2} q_a$ iff $aa \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(aa, a) = aa$) iff $aau \in A_2$ (ind. hyp.) iff $aa(au) \in A_2$ (definition of A_2) iff $aaw \in A_2$. Similarly, if $w = bu$ then $aa \xrightarrow{w=bu}_{M_2} q_a$ iff $q_a \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(aa, b) = q_a$) iff $u \in \{a, b\}^*$ (ind. hyp.) iff $aa(bu) \in A_2$ (definition of A_2) iff $aaw \in A_2$.
- **Case** q_a**:** Consider $w = cu$, where $c = a$ or $c = b$. We have $q_a \xrightarrow{w=cu}_{M_2} q_a$ iff $q_a \xrightarrow{u}_{M_2} q_a$ (because $\delta_2(q_a, a/b) = q_a$) iff $u \in \{a, b\}^*$ (ind. hyp.) iff $(cu) \in \{a, b\}^*$ iff $w \in \{a, b\}^*$.

Thus the correctness has been established by induction.

■

Problem 2. [Category: Comprehension] Consider the following NFA M_0 over the alphabet $\{0, 1\}$.

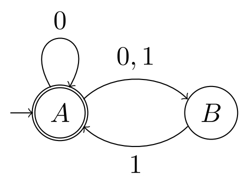


Figure 1: NFA M_0 for Problem 2

1. Describe formally what the following are for automaton M_0: set of states, initial state, final states, and transition function. **[4 points]**
2. What are $\hat{\delta}_{M_0}(A, 010)$, $\hat{\delta}_{M_0}(A, 101)$, $\hat{\delta}_{M_0}(A, 1101)$, and $\hat{\delta}_{M_0}(B, 10)$? **[4 points]**
3. What is $\mathbf{L}(M_0)$? You don't have to prove your answer. **[2 points]**

Solution:

1. States: $\{A, B\}$; Initial state: A; Final states: $\{A\}$; and transtions given by the following matrix

	0	1
A	$\{A, B\}$	$\{B\}$
B	$\emptyset$	$\{A\}$

2. Recall that $\hat{\delta}_M(q, w)$ gives the set of all states M could be in if it reads w starting at state q. So for this examples, we have: $\hat{\delta}_{M_0}(A, 010) = \{A, B\}$; $\hat{\delta}_{M_0}(A, 101) = \emptyset$; $\hat{\delta}_{M_0}(A, 1101) = \{A, B\}$; and $\hat{\delta}_{M_0}(B, 10) = \{A, B\}$.
3. $\mathbf{L}(M_0) = \{w \in \{0, 1\}^* \mid \text{the number of 1s before the first 0 in } w \text{ is even}\}$.

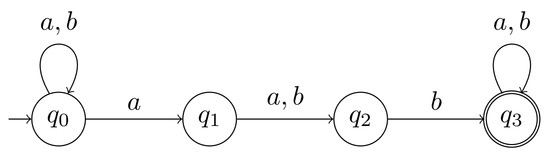
■

Problem 3. [Category: Design+Proof] Consider the language $A_2 \subseteq \{a, b\}^*$, from problem 1, which was defined to be the collection of strings w where there is a position i in w such that the symbol at position i (in w) is a, and the symbol at position $i + 2$ is b.

1. Design an NFA for language A_2 that has at most 4 states. You need not prove that your construction is correct, but the intuition behind your solution should be clear and understanable. **[5 points]**
2. Prove that any DFA recognizing A_2 has at least 5 states. **[5 points]**

Solution:

1. The NFA N_2 for the language A_2 as it reads the input, will "guess" some point when an a will be followed by a b two positions away. When it makes such a guess, it will check that the symbol two positions away is indeed a b and if so it will accept no matter what else follows. We could draw this NFA as follows. More generally, the same idea can be used to construct an NFA N_k with $k+2$ states that



recognizes A_k. The formal definition of such an NFA would be as follows. $N_k = (Q_k^N, \{a,b\}, \delta_k^N, q_0, F_k^N)$ where

- $Q_k^N = \{q_0, q_1, \ldots q_{k+1}\}$
- $F_k^N = \{q_{k+1}\}$
- And the transition function is given as

$$\delta(q,c) = \begin{cases} \{q_0, q_1\} & \text{if } q = q_0 \text{ and } c = a \\ \{q_0\} & \text{if } q = q_0 \text{ and } c = b \\ \{q_{i+1}\} & \text{if } q = q_i,\ 0 < i < k \\ \{q_{k+1}\} & \text{if } q = q_k \text{ and } c = b \\ \{q_{k+1}\} & \text{if } q = q_{k+1} \end{cases}$$

and $\delta(q,c) = \emptyset$ in all other cases.

2. Let $M = (Q, \{a,b\}, \delta, q_0, F)$ be any DFA that recognizes A_2. We will argue that $|Q| \geq 5$. Let $u_A = aa$, $u_B = ab$, $u_C = ba$, $u_D = bb$, and $u_E = aab$. Let us denote by A, B, C, D, and E the states of M such that $q_0 \xrightarrow{u_A}_M A$, $q_0 \xrightarrow{u_B}_M B$, $q_0 \xrightarrow{u_C}_M C$, $q_0 \xrightarrow{u_D}_M D$, and $q_0 \xrightarrow{u_E}_M E$. Our main claim is that each of the states A, B, C, D, E are distinct. The proof of this fact is going to be based on the following observation. Suppose for some $p, q \in \{A, B, C, D, E\}$ we have $p = q$. Then since $q_0 \xrightarrow{u_p}_M p = q$ and $q_0 \xrightarrow{u_q}_M q = p$, for any string w, $\hat{\delta}_M(q_0, u_p w) = \hat{\delta}_M(q_0, u_q w)$. Thus, $u_p w \in A_2$ iff $u_q w \in A_2$. So to show that $p \neq q$, we will find a string w such that $u_p w \in A_2$ and $u_q w \notin A_2$ (or vice versa). Thus, in the various cases below we just list the "witness" string w in each case.

- **Case** $\{E\} \cap \{A, B, C, D\} = \emptyset$: Take $w = \epsilon$. Observe that $aab\epsilon = aab \in A_2$, but $aa\epsilon = aa$, $ab\epsilon = ab$, $ba\epsilon = ba$, and $bb\epsilon = bb$ are not in A_2. Thus, $E \neq A$, $E \neq B$, $E \neq C$, and $E \neq D$.
- **Case** $\{A, B\} \cap \{C, D\} = \emptyset$: Take $w = b$. Observe that $u_A w = aab$, and $u_B w = abb$ are members of A_2, but $u_C w = bab$ and $u_D w = bbb$ are not in A_2. Thus, $A \neq C$, $A \neq D$, $B \neq C$, and $B \neq D$.
- **Case** $A \neq B$: Consider $w = ab$. Now $u_A w = aaab \in A_2$ but $u_B w = abab \notin A_2$. Hence $A \neq B$.
- **Case** $C \neq D$: Consider $w = bb$. $u_C w = babb \in A_2$, but $u_D w = bbbb \notin A_2$. Hence, $C \neq D$.

Based on the above cases, and the argument above, we can conclude that all the states A, B, C, D, E must be distinct, and M has at least 5 states.

■