Exam 1 solutions

1. The following questions are short-answer questions. For each question, there is a correct answer consisting of just a few written lines (or less).

   (a) Give a specific example of a language over the alphabet $\{0,1\}$ for each of the following categories.

   A finite language: $\varnothing$

   An infinite regular language: $\{0,1\}^*$

   A nonregular language: $\{0^n 1^n : n \in \mathbb{N}\}$

   (Of course these are just examples of acceptable answers, of which there are infinitely many.)

   (b) Give an example of a regular language $A \subseteq \{0,1\}^*$ for which there does *not* exist a DFA M for which both of these properties hold simultaneously: (i) $A = \mathrm{L}(M)$ and (ii) M has exactly one accept state.

   You do not need to prove that your answer is correct—it is enough to just describe the language (but do not hesitate to explain your answer if you think it will help).

Solution. The language $A = \{\varepsilon, 0\}$ has the required property. The start state must be an accept state because $\varepsilon \in A$, and from the start state we must transition on input 0 to an accept state. This transition must, however, be to a new state, for otherwise we would accept 0^n for all $n \in \mathbb{N}$.

   (c) For every regular language $A \subseteq \{0,1\}^*$, there exists an NFA N having *exactly one accept state* such that $A = \mathrm{L}(N)$. Give a brief, high-level proof that supports this claim.

Solution. Given an DFA M for which $\mathrm{L}(M) = A$, we can construct an NFA N with one accept state such that $\mathrm{L}(N) = A$ as follows. Let N include all states and transitions of M along with a new state, make this new state the only accept state of N, and add ε-transitions from every state that was accepting in M to this new state.

More precisely, if $M = (Q, \Sigma, \delta, q_0, F)$, then define

$$N = (Q \cup \{r\}, \Sigma, \eta, q_0, \{r\}),$$

where

$$\eta(q,\sigma) = \begin{cases} \{\delta(q,\sigma)\} & \text{if } q \in Q \text{ and } \sigma \in \{0,1\} \\ \{r\} & \text{if } q \in F \text{ and } \sigma = \varepsilon \\ \varnothing & \text{otherwise.} \end{cases}$$

2. Let $\Sigma = \{0,1\}$ and $\Gamma = \{0,1,X\}$ be alphabets. The symbol X may be an unusual choice for a symbol in an alphabet in this course, but for this problem it makes sense, as you will see.

Now, if we are given strings $w \in \Gamma^*$ and $x \in \Sigma^*$, we may imagine replacing every instance of the symbol X in w with the string x. For example, if

$$w = 00XX0X1 \quad \text{and} \quad x = 0100,$$

then by replacing every X in w by x we obtain the string

$$0001000100001001.$$

The following two problems concern this notion.

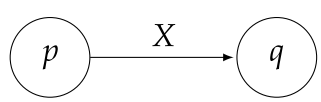
(a) Suppose a regular language $A \subseteq \Gamma^*$ and a string $x \in \Sigma^*$ are given. Prove that this language is regular:

$$B = \left\{ y \in \Sigma^* \;:\; \begin{array}{l} \text{there exists a string } w \in A \text{ such that } y \text{ is obtained} \\ \text{by replacing every } X \text{ in } w \text{ with } x \end{array} \right\}.$$

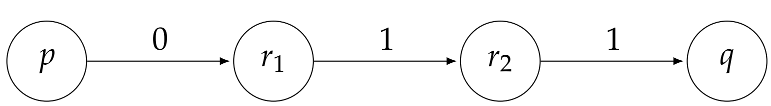
In words, B is the language of all strings you can get by picking any string from A and replacing all of the X symbols in that string with the string x.

Solution. Perhaps the simplest solution of this problem uses regular expressions. Given that A is regular, there must exist a regular expression R over the alphabet Γ such that $A = \mathrm{L}(R)$. Let S be the regular expression obtained from R by replacing every occurrence of the symbol X by the string x. It holds that $\mathrm{L}(S) = B$, and therefore B is regular.

One alternative is to argue that a DFA M for A can be converted to an NFA N for B by replacing each X transition in M by a string of transitions through a sequence of new auxiliary states, in accordance with the symbols in x. For example, if $x = 011$, then each transition



is replaced by the sequence



If you used this approach, it was important to include a separate copy of this sequence of transitions for each X transition—if you tried to connect a single sequence of transitions like this with ε-transitions from/to the states involved with each X transition, you would be adding too much freedom to jump between different states.

(b) Suppose a regular language $C \subseteq \Sigma^*$ and a string $x \in \Sigma^*$ are given. Prove that this language is regular:

$$D = \left\{ w \in \Gamma^* \;:\; \begin{array}{l} \text{by replacing every } X \text{ in } w \text{ with } x \text{, a string} \\ y \in C \text{ is obtained} \end{array} \right\}.$$

In words, a string $w \in \Gamma^*$ is contained in D if, after replacing every X symbol in w by x, you end up with a string in C.

Solution. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA such that $\mathrm{L}(M) = C$. Define a new DFA K as follows:

$$K = (Q, \Gamma, \eta, q_0, F)$$

where

$$\eta(q, \sigma) = \begin{cases} \delta(q, \sigma) & \text{if } \sigma \in \Sigma \\ \delta^*(q, x) & \text{if } \sigma = X. \end{cases}$$

In words, K works just like M, except that it also includes one transition out of each state corresponding to the symbol X. The state that this transition goes to is precisely the state that the string x leads to in M (through a number of transitions equalling the length of x). It holds that $\mathrm{L}(K) = D$, and therefore D is regular.

3. For this problem you will be asked to prove that two languages are nonregular, and you can do this through whatever method you choose. In case you choose to use the pumping lemma for regular languages to do this, here is a statement of that lemma for your convenience.

Pumping lemma: For every alphabet Σ and every regular language $A \subseteq \Sigma^*$, there exists a pumping length $n \geq 1$ for A for which the following statement holds: for every string $w \in A$ with $|w| \geq n$, there exist strings $x, y, z \in \Sigma^*$, with $w = xyz$, such that

1. $y \neq \varepsilon$,
2. $|xy| \leq n$, and
3. $xy^iz \in A$ for every $i \in \mathbb{N}$.

(a) Prove that the language

$$A = \left\{ 0^{\lfloor m/2 \rfloor} 1^m \;:\; m \in \mathbb{N} \right\}$$

is nonregular.

Solution. Assume toward contradiction that A is regular. By the pumping lemma, there exists a pumping length $n \geq 1$ for A that satisfies the property stated in that lemma. Let us fix such a choice of n for the remainder of the proof.

Let $w = 0^n 1^{2n}$. It holds that $w \in A$ and $|w| = 3n \geq n$, so it is possible to write $w = xyz$ for strings $x, y, z \in \{0, 1\}^*$ satisfying the three conditions listed in the pumping lemma.

Because $|xy| \leq n$, the prefix xy of w is not long enough to reach the 1s in that string, so conclude that $y = 0^k$ for some natural number k. Because $y \neq \varepsilon$, it follows that $k \geq 1$. Finally, because $xy^iz \in A$ for all $i \in \mathbb{N}$, we conclude that

$$xy^2z = 0^{n+k}1^{2n} \in A.$$

However, because $k \geq 1$, it is not the case that $0^{n+k}1^{2n} \in A$, because $\lfloor (2n)/n \rfloor = n \neq n + k$. Having obtained a contradiction, we conclude that A is nonregular, as required.

Note: if you chose $w = 0^{\lfloor n/2 \rfloor}1^n$ as the string leading to a contradiction, it could still work, but more effort is required. In particular, you cannot conclude that y takes the form 0^k for $k \geq 1$ in this case, so you have to consider that y could take one of the forms $y = 0^k$, $y = 0^k1^j$, or $y = 1^j$, and reason that a contradiction arises in all cases.

(b) Prove that the language

$$B = \{0^n1^m \,:\, n, m \in \mathbb{N},\ n \neq m\}$$

is nonregular.

Solution. Assume toward contradiction that B is regular. Because the regular languages are closed under complementation and intersection, it follows that the language $\overline{B} \cap \mathrm{L}(0^*1^*)$ is regular. However, it holds that

$$\overline{B} \cap \mathrm{L}(0^*1^*) = \{0^n1^n \,:\, n \in \mathbb{N}\},$$

which we already proved in lecture is not regular. Having obtained a contradiction, we conclude that B is not regular, as required.

Note that you could also prove that B is nonregular using the pumping lemma, but it requires you make a very good choice for w to get a contradiction. One choice that works is to take

$$w = 0^n1^{n!+n},$$

for n being a pumping length for B, which exists under the assumption that B is regular. Reasoning through the pumping lemma along similar lines to several examples we've seen, we conclude that there exists $k \in \{1, \ldots, n\}$ so that

$$0^{n+(i-1)k}1^{n!+n} \in B$$

for every $i \in \mathbb{N}$. However, taking

$$i = \frac{n!}{k} + 1,$$

which is a natural number because k must evenly divide $n!$, you find that

$$0^{n!+n}1^{n!+n} \in B,$$

which is a contradiction. (You could alternatively replace $n!$ with the least common multiple of the numbers $2, \ldots, n$, for instance, or simply argue the existence of a value to substitute for $n!$ using a system of congruences.)

4. This is intended to be a more challenging problem than the others on this exam. You are advised to answer all of the other problems first before attempting this one.

   Let $\Sigma = \{0,1\}$. Given two strings $v, w \in \Sigma^*$, it is said that v is obtained from w by a *substring reversal* if it is the case that

   $$w = xyz \qquad \text{and} \qquad v = xy^{\mathrm{R}}z$$

   for some choice of strings $x, y, z \in \Sigma^*$.

   For example, the string $v = 0110110$ is obtained from the string $w = 0111010$ by a substring reversal, as we may take $x = 011$, $y = 10$, and $z = 10$ in the definition above. The string w is also obtained from itself by a substring reversal, as we may take $x = w$, $y = \varepsilon$, and $z = \varepsilon$ (for instance); there is nothing in the definition that prohibits us from choosing x, y, or z to be the empty string.

   Suppose that $A \subseteq \Sigma^*$ is a given regular language, and $B \subseteq \Sigma^*$ is defined as

   $$B = \{v \in \Sigma^* \,:\, v \text{ is obtained from some string } w \in A \text{ by a substring reversal}\}.$$

   Prove that B is a regular language.

Solution. The language A is regular, and therefore there exists a DFA $M = (Q, \Sigma, \delta, q_0, F)$ such that $\mathrm{L}(M) = A$. For every choice of states $p, q \in Q$, define a new DFA

$$M_{p,q} = (Q, \Sigma, \delta, p, \{q\}),$$

and define $A_{p,q} = \mathrm{L}(M_{p,q})$. The language $A_{p,q}$ is regular, as it is the language accepted by the DFA $M_{p,q}$. (Intuitively speaking, the language $A_{p,q}$ consists of all strings that would cause M to transition from state p to state q.)

Now, the language B may be written as

$$B = \bigcup_{\substack{p,q \in Q \\ r \in F}} A_{q_0,p} A_{p,q}^R A_{q,r}.$$

We proved in lecture that the reverse of every regular language is regular, and therefore $A_{p,q}^R$ is regular for each $p, q \in Q$. As the regular languages are closed under finite unions and concatenations, it follows that B is regular.