You have 90 minutes to complete this exam. You may assume without proof any statement proved in class.

1. Find a regular expression for each of the following languages over \{0, 1\}:

(a) nonempty strings in which the first and last symbols are different; (1 pt)
(b) strings in which the number of 0s is even; (1 pt)
(c) strings not containing the substring 01; (2 pts)
(d) strings in which the number of 0s and the number of 1s are either both even or both odd. (3 pts)
2 Prove or disprove:

(2 pts) \( a. \) for any regular languages \( L_1 \subseteq L_2 \subseteq \cdots \subseteq L_n \subseteq \cdots \), the union \( \bigcup_{n=1}^{\infty} L_n \) is regular;

(3 pts) \( b. \) if \( L_1 \) and \( L_2 \) are two languages such that the equivalence classes of \( \equiv_{L_1} \) are exactly the same as those of \( \equiv_{L_2} \), then \( L_1 = L_2 \).

(3 pts) Construct a language that can be recognized by a DFA with 2015 states but not with 2014 states. Prove both claims.
4 For each of the following languages, determine whether it is regular, and prove your answer:

(2 pts) a. binary strings with five times as many 0s as 1s;

(2 pts) b. binary strings of the form $uvu$, where $u$ and $v$ are nonempty strings;

(3 pts) c. strings over the decimal alphabet {0, 1, 2, ..., 9} with characters in sorted order;

(3 pts) d. binary strings such that in every suffix, the number of 0s and the number of 1s differ by at most 2.
SOLUTIONS
You have 90 minutes to complete this exam. You may assume without proof any statement proved in class.

1. Find a regular expression for each of the following languages over \{0, 1\}:

   a. nonempty strings in which the first and last symbols are different; (1 pt)
   b. strings in which the number of 0s is even; (1 pt)
   c. strings not containing the substring 01; (2 pts)
   d. strings in which the number of 0s and the number of 1s are either both even or both odd. (3 pts)

Solution.

   a. \(0\Sigma^*1 \cup 1\Sigma^*0\)
   b. \(1^*(01*01^*)^*\)
   c. \(1^*0^*\)
   d. \((\Sigma \Sigma)^*\)
2 Prove or disprove:

(2 pts) a. for any regular languages \( L_1 \subseteq L_2 \subseteq \ldots \subseteq L_n \subseteq \ldots \), the union \( \bigcup_{n=1}^{\infty} L_n \) is regular;

(3 pts) b. if \( L_1 \) and \( L_2 \) are two languages such that the equivalence classes of \( \equiv_{L_1} \) are exactly the same as those of \( \equiv_{L_2} \), then \( L_1 = L_2 \).

Solution.

a. False. Let \( L_n = 01 \cup 0011 \cup \ldots \cup 0^n 1^n \). Then each \( L_n \) is finite and hence regular, whereas their union is the language \( \{0^n 1^n : n \geq 1\} \), which was shown in class to be nonregular.

b. False. The languages \( L_1 = \emptyset \) and \( L_2 = \Sigma^* \) have the same set of equivalence classes, namely, a single class \( \Sigma^* \).

3 Construct a language that can be recognized by a DFA with 2015 states but not with 2014 states. Prove both claims.

Solution. Let \( L \) be the language of binary strings whose length is a multiple of 2015. Then \( L \) is recognized by the following DFA with 2015 states: \( (\{0, 1, 2, \ldots, 2014\}, \{0, 1\}, \delta, 0, \{0\}) \), where \( \delta(q, \sigma) = (q + \sigma) \mod 2015 \). We will now show that no smaller DFA exists. For any distinct \( i, j \in \{0, 1, 2, \ldots, 2014\} \), we have \( 0^i 0^{2015-i} \in L \) and \( 0^i 0^{2015-i} \notin L \). As a result, the 2015 strings \( \varepsilon, 0, 00, 000, \ldots, 0^{2014} \) are each in a different equivalence class of \( \equiv_L \). By the Myhill-Nerode theorem, any DFA for \( L \) must have at least 2015 states.
For each of the following languages, determine whether it is regular, and prove your answer:

(2 pts)  
(a) binary strings with five times as many 0s as 1s;

(2 pts)  
(b) binary strings of the form $uuuv$, where $u$ and $v$ are nonempty strings;

(3 pts)  
(c) strings over the decimal alphabet \{0, 1, 2, \ldots, 9\} with characters in sorted order;

(3 pts)  
(d) binary strings such that in every suffix, the number of 0s and the number of 1s differ by at most 2.

Solution.

In each part, $L$ stands for the language in question.

(a) Nonregular. Let $p$ be arbitrary, and consider the string $0^{5p}1^p$. Let $x, y, z$ be any strings such that $y$ is nonempty, $|xy| \leq p$, and $xyz = 0^{5p}1^p$. Then $y$ is a nonempty string of 0s, and therefore the number of 0s in $xz$ is less than five times the number of 1s. Since $xz \notin L$, the language is nonregular by (the contrapositive of) the pumping lemma.

(b) Nonregular. We claim that the infinite collection of strings $\varepsilon, 0, 00, \ldots, 0^n, \ldots$ are each in a different equivalence class of $\equiv_L$. Indeed, for $n < N$, the language contains $0^n10^N1$ but not $0^N10^N1$. By the Myhill-Nerode theorem, $L$ is nonregular.

(c) Regular. This language is given by the regular expression $0^*1^*2^*3^* \cdots 9^*$.

(d) Regular. Observe that $L^R$, the reverse of $L$, is the language of strings with the property that in every prefix, the number of 0s and the number of 1s differ by at most 2. This language is regular because it is recognized by the following NFA:

Since regular languages are closed under the reverse operation, $L$ must be regular as well.