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

(a) strings that begin with 00 and do not end with 11; (2 pts)

(b) strings in which both the number of 0s and the number of 1s are even; (2 pts)

(c) strings containing no more than one occurrence of 00 (the string 000 has two occurrences). (2 pts)
(1 pt) 2 Find a string of minimum length not in the language $0^* (100^*)^* 1^*$.  

(3 pts) 3 Let $L$ be an infinite regular language. Prove that $L$ can be partitioned into two disjoint infinite regular languages.
Construct a DFA for $\Sigma^*1(\Sigma^*)^*1\Sigma^*$ with the smallest possible number of states, where $\Sigma = \{0, 1\}$. Prove that your DFA is the smallest possible.

5. True or false? Prove your answer.

(2 pts) a. If $A$, $A \cap B$, and $A \cup B$ are regular languages, then $B$ is regular as well.

(2 pts) b. For any nonregular languages $L_1 \subseteq L_2 \subseteq L_3 \subseteq \ldots$, the union $\bigcup_{n=1}^{\infty} L_n$ is nonregular.
For each of the following languages over $\Sigma = \{0, 1\}$, determine whether it is regular, and prove your answer:

- (2 pts) a. strings that contain exactly twice as many 1s as 0s;
- (2 pts) b. strings that do not contain a palindrome of length 2016 or shorter as a substring;
- (2 pts) c. strings that can be made into a palindrome by removing fewer than 2016 symbols;
- (2 pts) d. odd-length strings in which the middle symbol also occurs elsewhere in the string.
SOLUTIONS
You have 90 minutes to complete this exam. You may state without proof any fact taught in class or assigned as homework.

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

   (2 pts) a. strings that begin with 00 and do not end with 11;
   
   (2 pts) b. strings in which both the number of 0s and the number of 1s are even;
   
   (2 pts) c. strings containing no more than one occurrence of 00 (the string 000 has two occurrences).

\textbf{Solution.}

a. \( 00 \cup 00 \Sigma \cup 00 \Sigma^*(00 \cup 01 \cup 10) \);

b. \( (00 \cup 11 \cup (01 \cup 10)(00 \cup 11)^*(01 \cup 10))^* \)

c. \( 1^*(01^+)^*(\varepsilon \cup 0)1^*(01^+)^*(\varepsilon \cup 0) \)

In all three parts, you could solve the problem in a roundabout manner by constructing an NFA and converting it to a regular expression. But there are easier solutions, as follows. For \textbf{a.}, it helps to subdivide into cases and handle each of them separately; think back to my snowdrift metaphor. For \textbf{b.}, the problem gets much easier if you group the symbols in pairs. For \textbf{c.}, it helps to note that the language is precisely \( LL \), where \( L \) is the set of all strings that do not contain 00.
(1 pt) 2 Find a string of minimum length not in the language $0^*(100^*)^*1^*$.

**Solution.** 110

(3 pts) 3 Let $L$ be an infinite regular language. Prove that $L$ can be partitioned into two disjoint infinite regular languages.

**Solution.** By the pumping lemma, there is $p \geq 0$ such that every string $w \in L$ of length at least $p$ can be written as $w = xyz$, where $y$ is nonempty and $xy^iz \in L$ for all $i = 0, 1, 2, 3, \ldots$.

So, fix an arbitrary string $w \in L$ of length at least $p$ (it exists because $L$ is infinite). Let $w = xyz$ be a decomposition guaranteed by the pumping lemma. Partition $L = A \cup (L \setminus A)$, where $A = \{xy^iz : i = 0, 2, 4, 8, \ldots\}$.

- **DISJOINTNESS:** $A$ and $L \setminus A$ are disjoint by definition.
- **INFINITENESS:** $\{xy^iz : i = 0, 2, 4, 8, \ldots\} \subseteq A$ and $\{xy^iz : i = 1, 3, 5, 7, \ldots\} \subseteq L \setminus A$.
- **REGULARITY:** $A$ is regular because it is given by a regular expression, $x(yy)^*z$, which makes $L \setminus A$ regular as well by the closure properties.
Construct a DFA for $\Sigma^*1(\Sigma^*)^*1\Sigma^*$ with the smallest possible number of states, where $\Sigma = \{0, 1\}$. Prove that your DFA is the smallest possible.

**Solution.** The following DFA with four states recognizes $L = \Sigma^*1(\Sigma^*)^*1\Sigma^*$:

![DFA Diagram](image)

By the Myhill–Nerode theorem, no smaller DFA exists because each of the four strings $\varepsilon, 1, 10, 11$ is in a different equivalence class of $\equiv_L$. Their distinguishing suffixes are as follows:

<table>
<thead>
<tr>
<th></th>
<th>$\varepsilon$</th>
<th>1</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>$\varepsilon$</td>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>01</td>
<td>01</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
</tr>
</tbody>
</table>

5 True or false? Prove your answer.

(2 pts) **a.** If $A, A \cap B,$ and $A \cup B$ are regular languages, then $B$ is regular as well.

(2 pts) **b.** For any nonregular languages $L_1 \subseteq L_2 \subseteq L_3 \subseteq \ldots$, the union $\bigcup_{n=1}^{\infty} L_n$ is nonregular.

**Solution.**

**a.** True. One can obtain $B$ from the regular languages $A, A \cap B,$ and $A \cup B$ using set difference: $B = (A \cup B) \setminus (A \setminus (A \cap B))$. Since regular languages are closed under set difference, $B$ must be regular as well.

**b.** False. Define $L_n = (\Sigma \cup \varepsilon)^n \cup \{0^m1^m : m \geq 0\}$. We claim that:

- each $L_n$ is nonregular;
- $L_1 \subseteq L_2 \subseteq L_3 \subseteq \ldots$;
- the union $\bigcup_{n=1}^{\infty} L_n$ is the regular language $\Sigma^*$.

The second and third claims are obvious. For the first claim, the nonregular language $\{0^m1^m : m \geq 0\}$ is the set difference of $L_n$ and a finite (hence regular!) language. This means that each $L_n$ is nonregular—otherwise, the closure properties would force the regularity of $\{0^m1^m : m \geq 0\}$. 


For each of the following languages over $\Sigma = \{0, 1\}$, determine whether it is regular, and prove your answer:

(2 pts) **a.** strings that contain exactly twice as many 1s as 0s;
(2 pts) **b.** strings that do not contain a palindrome of length 2016 or shorter as a substring;
(2 pts) **c.** strings that can be made into a palindrome by removing fewer than 2016 symbols;
(2 pts) **d.** odd-length strings in which the middle symbol also occurs elsewhere in the string.

**Solution.**

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

**a.** Nonregular. For any positive integers $i \neq j$, we have $0^i 1^{2i} \in L$ but $0^j 1^{2i} \notin L$. Therefore, each of the strings $0, 00, 000, \ldots, 0^n, \ldots$ is in a different equivalence class of $\equiv_L$. Since there are infinitely many equivalence classes, $L$ is nonregular by the Myhill–Nerode theorem.

**b.** Regular. Let $F$ denote the finite (hence regular) language of palindromes of length at most 2016. Then $L$ can be built up from $F$ using the Kleene star, concatenation, and complement operations: $L = \Sigma^* F \Sigma^*$. Since regular languages are closed under these operations, $L$ must be regular as well.

**c.** Nonregular. For any positive integers $i \neq j$, we have $0^{2016i} 1^{20160^{2016i}} \in L$ but $0^{2016j} 1^{20160^{2016j}} \notin L$. Therefore, each of the strings $0^{2016}, 0^{4032}, \ldots, 0^{2016n}, \ldots$ is in a different equivalence class of $\equiv_L$. Since there are infinitely many equivalence classes, $L$ is nonregular by the Myhill–Nerode theorem.

**d.** Nonregular. For any positive integers $i \neq j$, we have $0^{2i} 10^{2i} \notin L$ but $0^{2j} 10^{2i} \in L$. Therefore, each of the strings $0^2, 0^4, \ldots, 0^{2n}, \ldots$ is in a different equivalence class of $\equiv_L$. Since there are infinitely many equivalence classes, $L$ is nonregular by the Myhill–Nerode theorem.