1. Prove the following assertions, where $\Sigma = \{0, 1\}$.

(a) $\Sigma^* \setminus (0^* \cup 1^*) = 0^* \cup 1^*$

(b) $0^*(10^+)^*(\varepsilon \cup 1) = (\varepsilon \cup 1)(01 \cup 0)^*$

Solution.

(a) $\Sigma^* \setminus (0^* \cup 1^*) = \Sigma^* \cap \overline{0^* \cup 1^*} = \Sigma^* \cup \overline{0^* \cup 1^*} = \emptyset \cup 0^* \cup 1^* = 0^* \cup 1^*$.

(b) The left-hand side corresponds to binary strings in which every 1, with the possible exception of the last 1, is immediately followed by a 0. The right-hand side corresponds to binary strings in which every 1, with the possible exception of the first 1, is immediately preceded by a 0. In both cases, the language is $\Sigma^* \Pi \Sigma^*$. 

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

(a) even-length strings that contain 01;

(b) strings in which every 1 is adjacent to a 0.

Solution.

(a) $(\Sigma \Sigma)^*01(\Sigma \Sigma)^* \cup \Sigma(\Sigma \Sigma)^*01\Sigma(\Sigma \Sigma)^*$

(b) $(0 \cup 01 \cup 10 \cup 101)^*$
3. Let $L$ be a nonempty finite language in which the longest string has length $n$. Prove that any DFA for $L$ must have at least $n + 1$ states.

*Solution.* For the sake of contradiction, assume that $L$ has a DFA $D$ with at most $n$ states. Let $w \in L$ be a string of length $n$. Clearly, $D$ accepts $w$. Moreover, by the same argument as in the pumping lemma, $D$ must visit some state at least twice while processing $w$. The segment of $w$ between those visits can be repeated as many times as desired without affecting $D$’s output. This gives an infinite family of strings accepted by $D$, a contradiction to the fact that $L$ is finite.

4. Construct a DFA for the language $0\Sigma^* \cup \Sigma^*11$ over the binary alphabet, using the smallest possible number of states. Prove that your DFA is the smallest possible.

*Solution.* The language $L = 0\Sigma^* \cup \Sigma^*11$ consists of binary strings that begin with 0 or end with 11. It is recognized by the following DFA:

![DFA Diagram](image)

By the Myhill–Nerode theorem, no smaller DFA exists for $L$ because each of the five strings $\varepsilon, 0, 1, 10, 11$ is in a different equivalence class of $\equiv_L$. The distinguishing suffixes are as follows:

<table>
<thead>
<tr>
<th></th>
<th>$\varepsilon$</th>
<th>0</th>
<th>1</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$\varepsilon$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$0$</td>
<td>$\varepsilon$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$0$</td>
<td>$\varepsilon$</td>
<td>$1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>$\varepsilon$</td>
<td>$0$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td></td>
</tr>
</tbody>
</table>
Prove or disprove:

a. if \( L \) is a nonregular language and \( w \) a string, then the concatenation \( wL \) is nonregular; (2 pts)

b. if \( L \) is a nonregular language, then prefix(\( L \)) is also nonregular; (2 pts)

c. if \( L \) is a regular language, then the language \( L' \) of even-length strings whose first half is in \( L \) is also regular. (2 pts)

**Solution.**

a. True. If \( wL \) were regular with a DFA \( D = (Q, \Sigma, \delta, q_0, F) \), then \( L \) would be recognized by the DFA \( (Q, \Sigma, \delta, q, F) \), where \( q \) is the end state of \( D \) after processing \( w \).

b. False. Consider the language \( L \) of strings that contain as many 0s as 1s. We showed in class that \( L \) is nonregular. But prefix(\( L \)) = \( \Sigma^* \), which is a regular language because it is given by a regular expression.

c. False. Consider the regular language \( L = 0^* \). Then \( L' \) is the set of even-length binary strings whose first half does not contain a 1. For any nonnegative integers \( i < j \), we have \( 0^i1^j \notin L' \) but \( 0^i1^j \in L' \). Therefore, each of the strings \( \varepsilon, 0, 00, 000, \ldots \) is in a different equivalence class of \( \equiv_{L'} \). Since there are infinitely many equivalence classes, \( L' \) is nonregular by the Myhill–Nerode theorem.
For each of the following languages \( L \) over the binary alphabet, determine whether it is regular and prove your answer:

(a) even-length strings whose first half contains as many 0s as the second half; (2 pts)
(b) strings \( w \) such that every prefix of \( w \) is equal to some suffix of \( w \); (2 pts)
(c) strings whose length, when expressed as a decimal integer, uses no digits other than 0 and 1. (2 pts)

**Solution.**

(a) Nonregular. For any nonnegative integers \( i \neq j \), we have \( 0^i1^{i+j}0^j \notin L \) but \( 0^j1^{i+j}0^i \in L \). Therefore, each of the strings \( \varepsilon, 0, 00, 000, \ldots \) 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, with regular expression \( L = 0^* \cup 1^* \). Indeed, \( L \) by definition contains every string in \( 0^* \cup 1^* \). Conversely, let \( w = w_1w_2\ldots w_n \) be any string in \( L \). Then the prefix \( w_1w_2\ldots w_{n-1} \) must be equal to some suffix of \( w \). But \( w \)'s only suffix of length \( n-1 \) is \( w_2w_3\ldots w_n \). This means that \( w_1w_2\ldots w_{n-1} = w_2w_3\ldots w_n \), which simplifies to \( w_1 = w_2 = \cdots = w_n \) and thus \( w \in 0^* \cup 1^* \).

(c) Nonregular. Consider the following family of strings whose length is a power of ten: \( 0^1, 010, 0100, 01000, \ldots \) Each of them is in a different equivalence class of \( \equiv_L \). Indeed, for any nonnegative integers \( i \neq j \), we have \( 0^{10^i}0^{10^j} = 0^{2 \cdot 10^i} \notin L \) but \( 0^{10^j}0^{10^i} = 0^{10^i+10^j} \in L \). Since there are infinitely many equivalence classes, \( L \) is nonregular by the Myhill–Nerode theorem.