Problem 1. [25 points] The reverse of a string $\alpha$ over an alphabet $\Sigma$, denoted $\alpha^\top$, is defined inductively by $\varepsilon^\top = \varepsilon$ and, for $a \in \Sigma$ and $w \in \Sigma^*$, by $(aw)^\top = w^\top a$. The reverse of a language $L$ over an alphabet $\Sigma$ is the language $L^\top = \{ w^\top : w \in L \}$.

Is it always true that the reverse of a regular language is regular? What of context-free languages: is $L^\top$ always context-free whenever $L$ is context-free?

Prove that your answers are correct.

Solution: Yes, both regular and context-free languages are closed under the reverse operation. Before we prove this, however, observe that whenever $\alpha, \beta \in \Sigma^*$, we have $(\alpha\beta)^\top = \beta^\top\alpha^\top$, which can easily be shown by induction on $\alpha$. One can also easily show by induction on $n$ that $(\alpha_0 \cdots \alpha_{n-1})^\top = \alpha_{n-1}^\top \cdots \alpha_0^\top$ for all $n \in \mathbb{N}$.

Let’s start by proving that regular languages are closed under the reverse operation. We show by induction on $R$ that, for any regular expression $R$, there is a regular expression $R^\top$ such that $L(R^\top) = L(R)^\top$ (note that $R^\top$ has not been defined and is just notation, so any other name, such as $R'$, would do). As $R$ is an inductively defined data structure, the proof by induction analyses all possible ways $R$ could have been defined:

- If $R = \emptyset$, then $L(R) = \emptyset$, so simply let $R^\top = \emptyset$;
- If $R = w$, where $w \in \Sigma^*$, $L(R) = \{w\}$, so let $R^\top = w^\top$ as then $L(R^\top) = \{w^\top\} = L(R)^\top$;
- If $R = R_0 | R_1$, where $R_0, R_1$ are regular expressions, we use the induction hypothesis to obtain regular expressions $R_0^\top, R_1^\top$ such that $L(R_0^\top) = L(R_0)^\top$ and $L(R_1^\top) = L(R_1)^\top$. We now simply let $R^\top = R_0^\top | R_1^\top$. This gives us

$$L(R^\top) = L(R_0^\top) \cup L(R_1^\top) = L(R_0)^\top \cup L(R_1)^\top = (L(R_0) \cup L(R_1))^\top = L(R)^\top;$$

- If $R = R_0 R_1$, where $R_0, R_1$ are regular expressions, we obtain $R_0^\top$ and $R_1^\top$ just as in the previous case. We then let $R^\top = R_1^\top R_0^\top$, so that

$$L(R^\top) = L(R_1^\top) L(R_0^\top) = \{w_1 w_0 : w_0 \in L(R_0^\top), w_1 \in L(R_1^\top) \} = \{\beta^\top \alpha^\top : \alpha \in L(R_0), \beta \in L(R_1) \} = \{((\alpha\beta)) : \alpha \in L(R_0), \beta \in L(R_1) \} = \{\alpha \beta : \alpha \in L(R_0), \beta \in L(R_1) \} = (L(R_0) L(R_1))^\top = L(R)^\top.$$


• Lastly, if $R = R_0^*$, where $R_0$ is a regular expression, then, by the induction hypothesis, we obtain a regular expression $R_0^T$ with $L(R_0^T) = L(R_0)^\top$. Then, by letting $R^\top = (R_0^* )^\top$,

$$L(R^\top) = \{w_0 \cdots w_{n-1} : \forall i, w_i \in L(R_0^\top)\} = \{x_0 \cdots x_{n-1} : \forall i, x_i \in L(R_0)\} = \{(x_{n-1} \cdots x_0)^\top : \forall i, x_i \in L(R_0)\} = \{x_0 \cdots x_0 : \forall i, x_i \in L(R_0)^\top\} = \{(R_0)^*\}^\top = L(R)^\top.$$

Therefore, since every regular language $L$ has a regular expression $R$ with $L(R) = L$, the language $L(R^\top) = L(R)^\top = L^\top$ is also regular.

Note: this problem can also be solved through the usual DFA methodology. Given a DFA recognizing $L$, we can build a NFA recognizing $L^\top$. Can you see how?

If we now take a context-free language $L$, we this time have a context-free grammar $G = (V, \Sigma, R, S)$ such that $L = L(G) = \{w \in \Sigma^* : S \xrightarrow{G} w\}$. Construct then the context-free grammar $G^\top = (V, \Sigma, R^\top, S)$ where $R^\top = \{X \rightarrow \alpha^\top : (X \rightarrow \alpha) \in R\}$. We show by induction on $k$ that, for all $k \in \mathbb{N}$ and all $\omega \in (V \cup \Sigma)^*$, $S \xrightarrow{k \ G^\top} \omega$ if, and only if, $S \xrightarrow{k \ G} \omega^\top$. The base case ($k = 0$) should be clear as $S$ is the only string in $(V \cup \Sigma)^*$ derivable from $S$ in zero steps.

By definition, $S \xrightarrow{k+1 \ G^\top} \omega$ if, and only if, there are strings $\alpha, \gamma \in (V \cup \Sigma)^*$ and a rule $(X \rightarrow \beta) \in R$ such that $\omega = \alpha \beta \gamma$ and $S \xrightarrow{G} \alpha X \gamma$. As $(x^\top)^\top = x$ for all $x \in (V \cup \Sigma)^*$, this is equivalent to there being strings $\alpha, \gamma \in (V \cup \Sigma)^*$ and a rule $(X \rightarrow \beta^\top) \in R^\top$ such that $\omega^\top = (\alpha \beta \gamma)^\top = \gamma^\top \beta^\top \alpha^\top$ and, by the inductive hypothesis, $S \xrightarrow{k \ G^\top} (\alpha X \gamma)^\top = \gamma^\top X \alpha^\top$. However, this is in turn equivalent to there being strings $x, z \in (V \cup \Sigma)^*$ and a rule $(X \rightarrow y) \in R^\top$ such that $w = xyz$ and $S \xrightarrow{G^\top} x X z$. However, this is the definition of $S \xrightarrow{k+1 \ G^\top} w$.

Since $S \xrightarrow{k \ G} w$ if, and only if, $S \xrightarrow{k \ G^\top} w^\top$ for all $k \in \mathbb{N}$ and $w \in \Sigma^*$, we have that $S \xrightarrow{G} w$ if, and only if, $S \xrightarrow{G^\top} w^\top$ for all $w \in \Sigma^*$. As each $w \in \Sigma^*$ is then in $L(G)$ if, and only if, $w^\top$ is in $L(G^\top)$, it follows that $L(G^\top) = L(G)^\top = L^\top$, showing that $L^\top$ is context-free.
Problem 2. [50 points] For each of the following languages, determine whether it is regular and prove that your answer is correct:

- \( \{ww^T : w \in \{0,1\}^*\} \);
- \( \{0^a1^b0^c : a + b = c\} \);
- The language of strings over \( \{0,1\} \) that contain an equal number of occurrences of 01 and 10 or, more formally,
  \[
  \left\{ a_0 \cdots a_{n-1} \in \{0,1\}^* : |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 01\}| = |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 10\}| \right\};
  \]
- The language of strings over \( \{0,1,2\} \) that contain an equal number of occurrences of 01 and 10 or, more formally,
  \[
  \left\{ a_0 \cdots a_{n-1} \in \{0,1,2\}^* : |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 01\}| = |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 10\}| \right\};
  \]
- \( \{0^n1^n : m = n\} \).

The language \( \{ww^T : w \in \{0,1\}^*\} \) is not regular: if it were, by the Pumping Lemma, it would have a pumping length \( p \in \mathbb{N} \setminus \{0\} \). The string \( 0^p1^p0^p \) could then be written as \( xyz \) satisfying the lemma’s conditions. As \( |xy| \leq p \), \( y = 0^k \) for some \( k \in \mathbb{N} \). As \( y \neq \varepsilon \), \( k > 0 \). However, the string \( xy^0z = xz = 0^{p-k}1^p0^p \) would belong to the language, a clear contradiction.

The language \( \{0^a1^b0^c : a + b = c\} \) is not regular: again by the Pumping Lemma, if it were regular, it would have a pumping length \( p \in \mathbb{N} \setminus \{0\} \), so the string \( 0^p1^p0^p \), which is in the language, would be writable as \( xyz \) obeying the lemma’s conditions. As \( |xy| \leq p \), \( y = 0^k \) for some \( k \). As \( y \neq \varepsilon \), \( k > 0 \). But then \( xy^0z = xz = 0^{p-k}1^p0^p \) would also be in the language despite \( p - k + p = 2p - k < 2p \).

The language \( \left\{ a_0 \cdots a_{n-1} \in \{0,1\}^* : |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 01\}| = |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 10\}| \right\} \) is regular: it can be recognized by the following DFA.

![DFA Diagram]

To prove this, we start by defining the debt \( D(w) \) of a string \( w = a_0 \cdots a_{n-1} \in \{0,1\}^* \) as the number of occurrences of 01 minus the number of occurrences of 10 or, in other words,

\[
D(w) = |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 01\}| - |\{i \in \{0,\ldots,n-2\} : a_ia_{i+1} = 10\}|.
\]

We then show by induction on \( w \) that \( D(w) \in \{-1,0\} \) if \( w \) ends with a 0, that \( D(w) \in \{0,1\} \) if \( w \) ends with a 1 and that \( \delta(q_{\varepsilon 0},w) = q_{a,D(w)} \) if \( w \) ends with character \( a \in \Sigma \). The result is obvious for \( w = \varepsilon \) as there is nothing to prove. When \( w = a \in \Sigma \), we have \( D(w) = 0 \), so indeed \( \delta(q_{\varepsilon 0},w) = \delta(q_{\varepsilon 0},a) = q_{a,D(w)} \). Suppose then that \( w = w'ab \) with \( w' \in \Sigma^* \) and \( a,b \in \Sigma \). The induction hypothesis gives us that \( D(w'a) \) is in \( \{-1,0\} \) if \( a = 0 \) and in \( \{0,1\} \) if \( a = 1 \). Furthermore, we get \( \delta(q_{\varepsilon 0},w'a) = q_{a,D(w'a)} \). As \( \delta(q_{\varepsilon 0},w) = \delta(\delta(q_{\varepsilon 0},w'a),b) = \delta(q_{a,D(w'a)},b) \), we consider two possible cases:

- If \( a = b \), then no new 01 or 10 sequences can involve the last character of \( w \), so \( D(w) = D(w'a) \). However, from the diagram, we get \( \delta(q_{a,D(w'a)},b) = q_{a,D(w'a)} = q_{b,D(w)} \), showing the inductive step.
- If \( a \neq b \), we have two subcases:
- If $a = 0$ and $b = 1$, then $w$ has one more occurrence of $01$ than $w'a$, so $D(w) = D(w'a) + 1$. Since $D(w'a) \in \{-1, 0\}$, we get $D(w) \in \{0, 1\}$. Furthermore, the diagram shows that $\delta(q_0, D(w'a), 1) = q_1, D(w'a) + 1 = q_1, D(w)$.

- If $a = 1$ and $b = 0$, then $w$ has one more occurrence of $10$ than $w'a$, which implies $D(w) = D(w'a) - 1$. Since $D(w'a) \in \{0, 1\}$, we have $D(w) \in \{-1, 0\}$. However, the diagram shows that $\delta(q_1, D(w'), 0) = q_0, D(w') - 1 = q_0, D(w)$, concluding the proof of the inductive step.

With this proof by induction complete, we have that the DFA accepts $\varepsilon$ (as it should) and satisfies $\delta(q_\varepsilon, w) = q_\varepsilon, D(w)$ for every non-empty $w \in \Sigma^*$ with last character $a$. Since the only accepting states are $q_\varepsilon, q_00$ and $q_10$, exactly the strings of debt 0 are accepted.

**The language** $\{a_0 \cdots a_{n-1} \in \{0, 1\}^* : |\{i \in \{0, \ldots, n-2\} : a_ia_{i+1} = 01\}| = |\{i \in \{0, \ldots, n-2\} : a_ia_{i+1} = 10\}|\}$

**is not regular:** suppose a DFA $M = (Q, \Sigma, \delta, q_0, F)$ recognized the language. Then define $q_i = \delta(q_0, (012)^i)$ for every $i \in \mathbb{N}$ (note that this agrees with $q_0$). Since $(012)^i(210)^j$ is in the language exactly when $i = j$, we have that, whenever $i \neq j$, that $\overline{\delta}(q_i, (210)^i) \in F$ while $\overline{\delta}(q_j, (210)^i) \notin F$, so $q_i \neq q_j$. As all $q_i$ for $i \in \mathbb{N}$ are in $Q$, the set $Q$ is infinite, which is a contradiction.

**The language** $\{0^m1^n : m = n\}$ **is not regular:** if it were it would have a pumping length $p$. We would then be able to write $0^p1^p = xyz$ observing the conditions of the Pumping Lemma. As $|xy| \leq p$, $y = 0^{|y|}$ and, as $y \neq \varepsilon$, $|y| > 0$. However, then the string $xy^0z = xz = 0^{p-|y|}1^p$ is in the language, a clear contradiction.
Problem 3. [25 points] Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Prove that:

- $L(M) = \emptyset$ if, and only if, there is no $w \in L(M)$ with $|w| < |Q|$; and
- $L(M)$ is an infinite set if, and only if, there is some $w \in L(M)$ with $|Q| \leq |w| < 2|Q|$.

**Solution:** If $L(M) = \emptyset$ then obviously there is no $w \in L(M)$ with $|w| < |Q|$. Suppose then that there is no $w \in L(M)$ with $|w| < |Q|$ and, for the purpose of a proof by contradiction, that $L(M) \neq \emptyset$. The well-ordering principle for the natural numbers then allows us to take a word $w \in L(M)$ of minimum length. However, we know that $|w| \geq |Q|$, which is a pumping length for $L(M)$ (as $M$ is a DFA recognizing $L(M)$). By the Pumping Lemma, we can write $w = xyz$ with $x, y, z \in \Sigma^*$ satisfying the Pumping Lemma conditions and, in particular, $|y| > 0$ and $xz = xy^i z \in L(M)$. However, this is a contradiction as $|xz| = |w| - |y| < |w|$ and $w$ is a string in $L(M)$ of minimum length.

If $L(M)$ is infinite, it needs to have a word $w$ of length at least $|Q|$ as there are only a finite amount of strings of length less than $|Q|$. The well-ordering principle for the natural numbers then allows us to pick a word $w \in \{x \in L(M) : |x| \geq |Q|\}$ of minimum length. In other words, there is a word $w \in L(M)$ of minimum length among all the words in $L(M)$ of length at least $|Q|$. Since $|Q|$ is a pumping length for $L(M)$, we can write $w = xyz$ with $x, y, z \in \Sigma^*$ satisfying the conditions of the Pumping Lemma. As $|y| > 0$ and $xz = xy^0 z \in L(M)$, we need to have $|xz| < |Q|$ as otherwise $xz$ is a word in $L(M)$ of length at least $|Q|$ but shorter than $w$. However, since $|xy| \leq |Q|$ (one of the Pumping Lemma conditions),

$$|Q| \leq |w| = |xz| + |y| < |Q| + |y| \leq |Q| + |Q| = 2|Q|.$$ 

Conversely, suppose there is a word $w \in L(M)$ with $|Q| \leq |w| < 2|Q|$. As again $|Q|$ is a pumping length for $L(M)$, we can write $w = xyz$ with $x, y, z \in \Sigma^*$ satisfying the conditions of the Pumping Lemma. In particular, we have that $xy^iz \in L(M)$ for every $i \in \mathbb{N}$ and, since $|y| > 0$ and $|xy^iz| = |xz| + i|y|$, that $L(M)$ has strings of arbitrarily large length, so it is infinite.