Problem 1. In class, we have seen that if $A$ and $B$ are regular languages over an alphabet $\Sigma$, then so is the language $A \cup B$. Which of the following languages are guaranteed to be regular as well? Prove that your answer is correct.

- $\overline{A} = \Sigma^* \setminus A = \{w \in \Sigma^*: w \notin A\}$;
- $A \cap B = \{w \in \Sigma^*: w \in A \text{ and } w \in B\}$;
- $A \setminus B = \{w \in \Sigma^*: w \in A \text{ but } w \notin B\}$; and
- $A \oplus B = \{w \in \Sigma^*: w \text{ is in exactly one of } A \text{ and } B\}$.

Solution: These are all guaranteed to be regular.

Since language $A$ is regular, there is a DFA $M = (Q, \Sigma, \delta, q_0, F)$ such that $L(M) = A$. In the solution to Problem 2, we showed that the DFA $M' = (Q, \Sigma, \delta, q_0, Q \setminus F)$ satisfies $L(M') = \overline{A}$, so $\overline{A}$ is also regular. As language $A$ is arbitrary, the complement of any regular language is regular.

Since $A$ and $B$ are regular, we obtain that both $\overline{A}$ and $\overline{B}$ are regular. From class, we know then that $\overline{A} \cup \overline{B}$ is regular. The complement of this language, $\overline{\overline{A} \cup \overline{B}}$, is then also regular, but, by De Morgan's laws, that is just $A \cap B$.

We also have $A \setminus B = A \cap \overline{B}$, an intersection of two regular languages, which is, as we have just shown, regular. Similarly, $A \oplus B = (A \cap \overline{B}) \cup (\overline{A} \cap B)$, which is a union of intersections of regular languages and therefore also regular.
Problem 2. Let \( M = (Q, \Sigma, \delta, q_0, F) \) be a DFA and consider the DFA \( M' = (Q, \Sigma, \delta, q_0, Q \setminus F) \). Is it always true that \( L(M') = L(M) = \Sigma^* \setminus L(M) \)?

What if one took a NFA \( N = (Q, \Sigma, \delta, q_0, F) \) and built a NFA \( N' = (Q, \Sigma, \delta, q_0, Q \setminus F) \)? Would it always be true that \( L(N') = L(N) \)?

Prove that your answers are correct.

Solution: We do have

\[
L(M') = \{ w \in \Sigma^* : \delta(q_0, w) \in Q \setminus F \} = \Sigma^* \setminus \{ w \in \Sigma^* : \delta(q_0, w) \in F \} = \Sigma^* \setminus L(M) = L(M).
\]

However, the same relation does not necessarily hold between \( N \) and \( N' \): if \( N \) is as in Figure 1a, then \( N' \) is as in Figure 1b, but both \( N \) and \( N' \) clearly accept the string \( \varepsilon \).

![Figure 1](image-url)
**Problem 3.** Consider the following scheme to describe chess positions. We use the letters P, B, N, R, Q, and K to denote, respectively, white pawns, bishops, knights, rooks, queens and kings; their black counterparts are denoted by the corresponding lower-case letter; empty board cells are denoted by the underscore character (_); and a chess position is denoted by the string of characters that is obtained by traversing the board in row-major order. For instance, the position depicted below is denoted by the string

```
___r_____b______R__________pkn_______p____BP___q_K___P__________Q
```

Consider some standard rule set of chess and the language of all strings over the alphabet \{P, B, N, R, Q, K, p, b, n, r, q, k, _\} that, under our established scheme, represent a legal chess position from which, if White were to start the game and play perfectly, White would be assured victory. Is this language regular? Prove that your answer is correct.

**Solution:** All strings in the language have length 64, which limits them to \(|\Sigma|^{64}\) in number (where \(\Sigma\) is the alphabet). The language is thus finite and, as we saw in class, regular.

That all finite languages are regular also follows by induction from the fact that the union of two regular languages is regular and the fact that any language with a single string is regular.
Problem 4. One way to define a balanced sequence of parentheses is as a string over the alphabet $Σ = \{\_, \)\}$ such that: its total number of ( characters equals its total number of ) characters; and in each of its prefixes, there are at least as many ( characters as there are ) characters.

The debt of a prefix of a balanced sequence of parentheses is the number of ( characters in the prefix minus the number of ) characters in it. The depth of a balanced sequence of parentheses is the maximum debt among all its prefixes.

For a given $d \in \mathbb{N}$, we can define the language $P_d$ of all balanced sequences of parentheses having depth at most $d$. We can also define the language

$$P = \bigcup_{d \in \mathbb{N}} P_d = P_0 \cup P_1 \cup \cdots$$

of all balanced sequences of parentheses.

Later in the course we shall see that $P$ is not a regular language. This will be shown in a systematic manner, but it is possible to prove this fact using only material covered in class. For now this is left as a challenging exercise but is not part of the assignment.

In this problem you must answer the following question and prove that your answer is correct. Which values of $d \in \mathbb{N}$ are such that $P_d$ is a regular language?

**Detailed Solution (worth 25 points):** $P_d$ is regular for all $d \in \mathbb{N}$. To show this, given $d \in \mathbb{N}$, we build the DFA $M = (Q, Σ, δ, q_0, \{q_d\})$ illustrated below where: $Q = \{\bot, q_0, \ldots, q_d\}$ is comprised of $d + 1$ counting states $q_0, \ldots, q_d$ and of an error state $\bot$; and $δ : Q × Σ → Q$ is given by

$$δ(q, a) = \begin{cases} \bot, & q = \bot \\ \bot, & q = q_0 \text{ and } a = ) \\ \bot, & q = q_d \text{ and } a = ( \\ q_{i-1}, & q = q_i \text{ with } 0 < i \leq d \text{ and } a = ) \\ q_{i+1}, & q = q_i \text{ with } 0 \leq i < d \text{ and } a = ( \end{cases}$$

\[ \begin{array}{c}
\text{ ( ) } \text{ ( )} \\
\text{ ( ( ) ) } \text{ ( ( ) )} \\
\rightarrow \text{ } \bot \text{ } \rightarrow \\
q_0 \text{ } q_1 \text{ } \cdots \\
\rightarrow \text{ } \rightarrow \text{ } \rightarrow \text{ } \rightarrow \\
q_d \text{ } q_{d-1}
\end{array} \]

Let us generalize the definition of debt by defining $D(w)$, the debt of a string $w ∈ Σ^*$, as the number of ( characters in $w$ minus the number of ) characters in $w$. Using this notation, we have that a string $w ∈ Σ^*$ is in $P_d$ if, and only if, $D(w) = 0$ and $0 ≤ D(α) ≤ d$ for every prefix $α$ of $w$.

We now show by induction on $w ∈ Σ^*$ that $δ(q_0, w) = q_d(w)$ if $0 ≤ D(α) ≤ d$ for every prefix $α$ of $w$ (including $w$ itself) and that $δ(q_0, w) = \bot$ otherwise. This should be clear for $w = ε$ as $D(ε) = 0$. Consider then the case $w = w’a$ where $w’ ∈ Σ^*$ and $a ∈ Σ$. If any prefix $α$ of $w’$ has $D(α) < 0$ or $D(α) > d$, then $δ(q_0, w’) = \bot$ by the inductive hypothesis, but then $δ(q_0, w) = δ(δ(q_0, w’), a) = δ(\bot, a) = \bot$ as it should be since $α$ is also a prefix of $w$.

Suppose now that every prefix of $w’$ has debt between 0 and $d$. The induction hypothesis then states that $δ(q_0, w’) = q_d(w’)$, so $δ(q_0, w) = δ(δ(q_0, w’), a) = δ(q_d(w’), a)$. We also know that every prefix of $w’$ except for $w$ itself has debt between 0 and $d$ since every prefix of $w’$ is a prefix of $w$. Note however that $D(w) = D(w’ − 1) + 1$ if $α = )$ and $D(w) = D(w’) + 1$ if $a = ($. Given this, the only way we can have $D(w) < 0$ is if $D(w’ = 0$ and $a = ($. Similarly the only way we can have $D(w) > d$ is if $D(w’ = d$ and $a = )$. Then indeed have $δ(q_0, w) = δ(q_d(w’), a) = \bot$ due to the second and third cases of the definition of $δ$.

If neither of these cases occur, i.e., if $0 ≤ D(w) ≤ d$, then the fourth and fifth cases in the definition of $δ$ do set $δ(q_d(w’), a)$ to $D(w’ − 1)$ if $α = )$ and to $D(w’) + 1$ if $a = ($. In other words, we have that $δ(q_0, w) = δ(q_d(w’), a) = q_d(w’)$, as desired.

To conclude the proof, observe that any $w ∈ P_d$ is in $L(M)$, as all prefixes of $w$ have debt between 0 and $d$, and so $δ(q_0, w) = q_d(w) = q_0$, which is the final state. Conversely, any string $w ∈ Σ^* \setminus P_d$ either has a prefix with debt below 0 or above $d$, in which case we know $δ(q_0, w) = \bot$, or else it has overall non-zero debt, in which case we know $δ(q_0, w) = q_d(w) ≠ q_0$. Either way $w ∉ L(M)$ since $q_0$ is the only final state.
Incomplete solution (worth 10 points): $P_d$ is regular for all $d \in \mathbb{N}$ since the DFA below recognizes $P_d$.

Note: The “…” notation in the picture above makes the definition imprecise. Note how the previous solution formally defines the DFA and uses the picture strictly as an illustration. This problem would not occur if the DFA were represented in full, but unfortunately we can not do this here as $d$ is a variable. A precise definition of the DFA would add 5 points to this answer’s mark. The remaining 10 points were taken because it was not argued why the DFA recognizes $P_d$.

Less detailed but still fully correct solution (worth 25 points): $P_d$ is regular for all $d \in \mathbb{N}$. To show this, given $d \in \mathbb{N}$, we build the DFA $M = (Q, \Sigma, \delta, q_0, \{q_0\})$ illustrated below where: $Q = \{\bot, q_0, \ldots, q_d\}$ is comprised of $d + 1$ counting states $q_0, \ldots, q_d$ and of an error state $\bot$; and $\delta : Q \times \Sigma \rightarrow Q$ is given by

\[
\delta(q, a) = \begin{cases} 
\bot, & q = \bot \\
\bot, & q = q_0 \text{ and } a = ) \\
\bot, & q = q_d \text{ and } a = ( \\
q_i, & q = q_i \text{ with } 0 < i \leq d \text{ and } a = ) \\
q_i, & q = q_i \text{ with } 0 \leq i < d \text{ and } a = ( . 
\end{cases}
\]

Let us generalize the definition of debt by defining $D(w)$, the debt of a string $w \in \Sigma^*$, as the number of ( characters in $w$ minus the number of ) characters in $w$. Using this notation, we have that a string $w \in \Sigma^*$ is in $P_d$ if, and only if, $D(w) = 0$ and $0 \leq D(\alpha) \leq d$ for every prefix $\alpha$ of $w$.

It can be shown by induction on $w \in \Sigma^*$ that $D(q_0, w) = q_{D(w)}$ if $0 \leq D(\alpha) \leq d$ for every prefix $\alpha$ of $w$ (including $w$ itself) and that $D(q_0, w) = \bot$ otherwise.

Observe then that any $w \in P_d$ is in $L(M)$, as all prefixes of $w$ have debt between 0 and $d$, and so $D(q_0, w) = q_{D(w)} = q_0$, which is the final state. Conversely, any string $w \in \Sigma^* \setminus P_d$ either has a prefix with debt below 0 or above $d$, in which case we know $D(q_0, w) = \bot$, or else has overall non-zero debt, in which case we know that $D(q_0, w) = q_{D(w)} \neq q_0$. Either way $w \notin L(M)$ since $q_0$ is the only final state.

Note: the sub-proof by induction that is omitted in this proof is recognized in the solution to be trivial and is indeed trivial, therefore the omission is not penalized.