INSTRUCTIONS

Upload a single file to Gradescope for each group. All group members’ names and PIDs should be on each page of the submission.

Your assignments in this class will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should always explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.

READING Sipser Chapter 1.3-1.4

KEY CONCEPTS Regular Expressions, Regular and Non-regular languages, Pumping Lemma
1. **(10 points)** Let $\Sigma = \{a, b\}$. For each regular expression, describe the language of the regular expression by choosing one of the sets named below, or saying none if the language of the regular expression is not any of the given sets. Sets may be used more than once or not at all.

Sets:

- $A = \{ w \in \Sigma^* \mid w \text{ doesn’t contain the substring } bb \}$
- $B = \{ w \in \Sigma^* \mid w \text{ contains the substring } bb \}$
- $C = \{ w \in \Sigma^* \mid w \text{ starts or ends with } bb \}$
- $D = \{ w \in \Sigma^* \mid w \text{ starts and ends with } bb \}$
- $E = \{ w \in \Sigma^* \mid w \text{ does not end with } bb \}$
- $F = \{ w \in \Sigma^* \}$

Regular Expressions:

1. $bb(a \cup b)^*bb$
   
   **Solution:** none

2. $(ba \cup a)^* \cup (ab)^*$
   
   **Solution:** none

3. $(ba^* \cup a)^*$
   
   **Solution:** $F$

4. $bb(a \cup b)^* \cup (a \cup b)*bb$
   
   **Solution:** $C$

5. $bb(a \cup b)^*bb \cup bb \cup bbb$
   
   **Solution:** $D$

6. $b^*b^*(a \cup b)^* \cup (a \cup b)^*b^*b^*$
   
   **Solution:** $F$

7. $(bb)^*$
   
   **Solution:** none

8. $(a \cup ab)^* \cup (ba)^*$
   
   **Solution:** none

9. $(a \cup b)^*(a \cup ab) \cup \varepsilon \cup b$
   
   **Solution:** $E$

10. $(a^* \cup b^*)^*bb(a \cup b)^*$
    
    **Solution:** $B$
2. (10 points) A common misconception about regular languages is that if a language is regular, then its subsets or supersets must be too. In this question, you will show that this is false. Let \( \Sigma = \{0, 1\} \).

(a) Give an example of a regular language \( X \) that is a subset of all nonregular languages over \( \Sigma \). Briefly justify your answer. (3 points)

**Solution:** Let \( X = \emptyset \). This is a regular set (it is described by the regular expression \( \emptyset \), and every set described by a regular expression is regular). Since the empty set is a subset of every set, \( X \) is a subset of all the nonregular sets.

(b) Give an example of a regular language \( A \) and a nonregular language \( B \) such that \( A \subseteq B \). For this part, you must choose \( A \) and \( B \) that are neither equal to \( \emptyset \) nor to \( \Sigma^* \). (3 points)

**Solution:** Let \( A = \{01\} \). This is a finite set and hence is regular. Let \( B = \{0^n1^n \mid n \geq 0\} \), which is proved in textbook to be nonregular. \( A \subseteq B \) because \( 01 = 0^11^1 \) with \( n = 1 \).

(c) Give an example of a nonregular language \( C \) and a regular language \( D \) such that \( C \subseteq D \). For this part, you must choose \( C \) and \( D \) that are neither equal to \( \emptyset \) nor to \( \Sigma^* \). (4 points)

**Solution:** Let \( C = \{0^n1^n \mid n \geq 0\} \), which is proved in textbook to be nonregular. Let \( D = \{0^m1^n \mid m, n \geq 0\} \), which is regular because it is described by the regular expression \( 0^*1^* \). \( C \subseteq D \) because in the set definition for \( D \), it’s possible to have \( m = n \).

For each part of this problem, justify any claims you make about certain sets being regular or nonregular either by proving the claim from definitions or citing a fact proved in class or in the textbook.

3. (10 points) Consider the language \( L \) of odd length strings over the alphabet \( \{a, b, c\} \) such that the middle symbol is \( b \). For example, \( acbca \), \( b \), \( aaabca \) are in \( L \) but \( bb \), \( acb \), \( a \) are not in \( L \). Fill in the missing parts of the following proof to show that \( L \) is not regular.

Assume (towards a contradiction) that \( L \) is regular. Then the Pumping Lemma applies to \( L \). Let \( p \) be the pumping length of \( L \). Choose \( s \) to be the string \( a^pba^p \) Since this string is in \( L \) and has length greater than or equal to \( p \), the Pumping Lemma guarantees \( s \) can be divided into parts \( s = xyz \) such that for any \( i \geq 0 \), \( xy^iz \) is in \( L \), and that \( |y| > 0 \) and \( |xy| \leq p \). But if we let \( i = 2 \), we get a string \( xy^2z \) that is not in \( L \) because

| it has more a’s before the b than after. We see that this must be the case because \( |y| > 0 \) and \( |xy| \leq p \) implies that \( y \) must be a nonempty string of \( a \)’s, since the first \( p \) symbols of our string \( s \) are all \( a \)’s. Then when we pump up with \( i = 2 \), our string has more \( a \)’s before the \( b \) than after, which means that the one and only \( b \) is no longer the middle symbol. So the pumped string is not in \( L \). |

This is a contradiction. Therefore the assumption is false, and \( L \) is not regular.
4. **(10 points)** Consider the language $L = \{tu \mid t$ and $u$ are strings over $\{0,1\}$ with the same number of $1$’s}. Explain and correct the error below:

**Proof that $L$ is not regular using the Pumping Lemma:**
Assume (towards a contradiction) that $L$ is regular. Then the Pumping Lemma applies to $L$. Let $p$ be the pumping length of $L$. Choose $s$ to be the string $1^p0^p1^p0^p$, which is in $L$ because $t = 1^p0^p$ and $u = 1^p0^p$ each have $p$ $1$’s. Since this string is in $L$ and has length greater than or equal to $p$, the Pumping Lemma guarantees $s$ can be divided into parts $s = xyz$ such that for any $i \geq 0$, $xy^i z$ is in $L$, and that $|y| > 0$ and $|xy| \leq p$. Since the first $p$ characters of $s$ are all $1$’s and we have $|y| > 0$ and $|xy| \leq p$, we know that $y$ must be nonempty and made up of all $1$’s. But if we let $i = 2$, we get a string $xy^2z$ that is not in $L$ because repeating $y$ twice adds $1$’s to $t$ but not to $u$, and strings in $L$ are required to have the same number of $1$’s in $t$ as in $u$. This is a contradiction. Therefore the assumption is false, and $L$ is not regular.

**Solution:**
The error is that actually when $s = 1^p0^p1^p0^p$ and $i = 2$, we can still have $xy^2z$ be in the language $L$. It is correct that $y$ is nonempty and made up of all $1$’s, but it is still possible that the string $xy^2z$ can be written as the concatenation of two strings with the same number of $1$’s. For example, if $y = 11$, then the pumped string $xy^2z = 1^{p+2}0^p1^p0^p$ which is still in $L$ because it can be written as $tu$ where $t = 1^{p+1}$ and $u = 10^p1^p0^p$, which both have the same number of $1$’s.

In fact, the language $L$ is regular, because any string that can be written as the concatenation of two strings with the same number of $1$’s must have an even number of $1$’s. Similarly, any string with an even number of $1$’s can be written as the concatenation of two strings with an equal number of $1$’s. So $L$ is just the language of strings over $\{0,1\}$ with an even number of $1$’s. This is easily shown to be regular by the following DFA that recognizes $L$. 

![DFA Diagram]
5. (20 points) Prove that the following languages are not regular. The format of proof should be similar to problem 3.

1. \( L = \{ wtw^R \mid w \in \{0, 1\}^*, t \in \{1\}^*, \text{ and both } w, t \text{ are nonempty} \} \). (10 points)

   **Proof:** Assume by way of contradiction that this set was indeed regular. The pumping lemma for regular languages then tells us that we can find a positive integer \( p \) such that we can write any \( s \in L \) with \( |s| \geq p \) as \( s = xyz \) where \( |xy| \leq p \), \( |y| \geq 1 \) for some choice of \( x, y, z \) such that \( xy^iz \in L \) for any integer \( i \geq 0 \). Applying this now to this language, we assume such a \( p \) exists. Consider the string \( s = 0^p1^p0^p \in L \) i.e. where \( w = 0^p \) and \( t = 1^p \). Then whenever we express \( s = xyz \) (for the appropriate constraints) then \( x = 0^i \) and \( y = 0^j \) with \( j \geq 1 \) and \( i + j \leq p \) and \( z \) is the rest of \( s \). Considering \( xy^2z \), this string starts with more than \( p \) zeros. Therefore, by definition of the language, in order to be included, such a string would also have to end in more than \( p \) zeros (since \( t \) cannot contain any). However, by construction, there are only \( p \) zeros in the suffix of \( z \). Therefore, we’ve reached a contradiction and our original assumption about the set being regular must be false. Hence, the language is not regular.

2. \( L = \{ a^{2^n} \mid n \geq 0 \} \). (10 points)

   **Proof:** Suppose \( L \) is regular. Let \( p \) be the pumping length given by the pumping lemma. Choose \( s = a^{2^p} \). By the lemma, \( |xy| \leq p \) and \( |y| > 0 \), therefore \( p \geq 0 \) and \( s \in L \). Clearly, \( |s| = 2^p \geq p \), thus \( s = xyz \) for some \( x, y, z \). Let us write \( x = a^a \), \( y = a^b \) and \( z = a^c \). The number of \( a \)'s in \( s \) is \( a + b + c = 2^p \). Let \( i = 2 \) and \( s' = xy^iz = xyyz \). The number of \( a \)s in \( s' \), denoted \( \#_a(s') \), is \( a + 2b + c = 2^p + b \). Since \( |y| > 0 \) and \( |y| = b, b > 0 \). From \( 2p = a + b + c < a + 2b + c \), we conclude \( 2^p < \#_a(s') \). Substituting for \( b \) on the right-hand side of \( a + 2b + c = 2^p + b \), we find \( a + 2b + c = 2^p + 2^p - a - c \). Since \( |xy| \leq p, c = |xyz| - |xy| \geq 2^p - p > 0 \), we have \( a + 2b + c < 2^{p+1} \). Thus \( \#_a(s') < 2^{p+1} \). Because \( 2^p < \#_a(s') < 2^{p+1} \), \( \#_a(s') \) is not an even power of 2 and \( s' \notin L \), a contradiction. Therefore \( L \) is non-regular.