1. Disprove the following statements about regular languages by finding a counterexample to each.

a. Suppose that \( L_1 \cap L_2 \) and \( L_1 \) are regular. Then \( L_2 \) is regular.

Let \( L_1 = \{a,b\}^* \) and \( L_2 = \) your favorite non-regular language over \( \{a,b\}^* \). Then \( L_1 \cap L_2 = L_1 \) so both \( L_1 \) and \( L_1 \cap L_2 \) are both regular, but \( L_2 \) is not.

b. If \( L_1 \) and \( L_2 \) are both non-regular languages, then \( L_1 \cap L_2 \) is also non-regular.

Let \( L_1 \) be your favorite non-regular language and \( L_2 = \) the complement of \( L_1 \). Then both are non-regular (if \( L_2 \) were regular, then so would be its complement, \( L_1 \)), yet the union is everything, which is regular.

c. If \( R \neq \emptyset \), \( S \neq e \), \( T \neq e \), and \( RS = RT \), then \( S = T \).

Let \( R = \{a,b\}^* \), \( S = a^* \), \( T = b^* \). Then \( RS = \{a,b\}^* = RT \), but \( S \neq T \).

2. a. Write down a ndfa for \( L = \{x \in \{a,b\}^* | \) x contains \( ab \} \).

```
q1    q2    q3    q4
\( a \rightarrow \) a \( \rightarrow b \rightarrow a \rightarrow a,b \)
```

b. Use the subset construction to find a dfa for \( L = \{x \in \{a,b\}^* | \) x contains \( ab \} \).

\( Q_1 = \{q_1\} \), \( Q_2 = \{q_1,q_2\} \), \( Q_3 = \{q_1,q_3\} \), \( Q_4 = \{q_1,q_2,q_4\} \), \( Q_5 = \{q_1,q_3,q_4\} \), \( Q_6 = \{q_1,q_4\} \)

```
Q1    Q2    Q3    Q4    Q5    Q6
\( a \rightarrow \) a \( \rightarrow b \rightarrow a \rightarrow a \rightarrow b \)
```

c. Use the algorithm in the text to construct a minimal dfa equivalent to the one found by the subset construction in part b.

\( Q_4 \), \( Q_5 \), and \( Q_6 \) collapse to the same state:
d. Write down a DFA for the complement of $L$, 
$L' = \{ x \in \{a,b\}^* | x \text{ does not contain } aba \}$.

![DFA Diagram]

3. Let $\square$, $\sqsubset$, and $\sqsupset$ be regular expressions and suppose we are given that $L(\square) \sqsubset L(\square)$. Prove carefully that $L(\square*) \sqsubset L(\square)$.

(Hint: Prove by induction on the number of occurrences of strings from $L(\square)$ in $w \sqsubset L(\square*)$). That is, if $w \sqsubset L(\square*)$, then $w = v_1...v_nz$ for $v_i \sqsubset L(\square)$. Your proof will be by induction on the $n$ telling the number of $v_i$. Be careful, it is easy to get skewed up in this problem and confuse regular expressions $\square$ with the set of elements they represent, $L(\square)$, and write down something which is meaningless.)

Prove by induction on $n \geq 0$ that if $w = v_1...v_nz$ for $v_i \sqsubset L(\square)$ and $z \sqsubset L(\square)$, then $w \sqsubset L(\square)$.

**Base case:** $n = 0$. Then $w = z$ where $z \sqsubset L(\square)$. Therefore, $w \sqsubset L(\square) \sqsubset L(\square) \sqsubset L(\square*) \sqsubset L(\square)$. Therefore $w \sqsubset L(\square)$.

**Induction hypothesis:** Suppose that if $w = v_1...v_nz$ for $v_i \sqsubset L(\square)$ and $z \sqsubset L(\square)$, then $w \sqsubset L(\square)$. We need to show this for $n+1$.

Let $w = v_1...v_nv_{n+1}z$ for $v_i \sqsubset L(\square)$ and $z \sqsubset L(\square)$. Let $w' = v_2...v_nv_{n+1}z$. Therefore, by induction $w' \sqsubset L(\square)$. But then $w = v_1w' \sqsubset L(\square) \sqsubset L(\square) \sqsubset L(\square) \sqsubset L(\square)$. Therefore $w \sqsubset L(\square)$.

Thus $L(\square*) \sqsubset L(\square)$.
4. The shuffle operation is important in the theory of concurrent systems. If \( x, y \in \mathcal{S}^* \), we write \( x \shuffle y \) for the set of all strings that can be obtained by shuffling strings \( x \) and \( y \) together like a deck of cards; for example,

\[
ab \shuffle cd = \{abcd, acbd, acdb, cabd, cadb, cdab\}
\]

The set \( x \shuffle y \) can be defined formally by induction:

\[
e \shuffle y = \{y\}
\]

\[
x \shuffle e = \{x\}
\]

\[
xa \shuffle yb = (x \shuffle yb) \cdot \{a\} \sqcup (xa \shuffle y) \cdot \{b\}
\]

The shuffle of two languages \( A \) and \( B \), denoted \( A \shuffle B \), is the set of all strings obtained by shuffling a string from \( A \) with a string from \( B \):

\[
A \shuffle B = \{ x \shuffle y \mid x \in A, y \in B \}
\]

For example, \( \{ab\} \shuffle \{cd, f\} = \{abf, afb, fab, abcd, acbd, acdb, cabd, cadb, cdab\} \)

(a) What is \( (01)^* \shuffle (10)^* \)?

This set can be characterized in many ways. For example, it is the set of all strings with the same number of 0's and 1's where the difference between the number of 0's and 1's is no greater than one in any prefix. However, it can also be written as \( (01 \shuffle 10)^* \).

(b) Show that if \( A \) and \( B \) are regular sets then so is \( A \shuffle B \).

**Hint:** If \( M_A \) and \( M_B \) are DFA's for \( A \) and \( B \), design a machine whose states are pairs of states from \( M_A \) and \( M_B \) - like the intersection machines, but have it choose non-deterministically as it is computing whether to act like \( M_A \) or \( M_B \). Provide a very convincing argument that the new machine accepts what it is supposed to (and no more).

Define \( M = (K_A \times K_B, \delta, \emptyset, <s_A, s_B>, F_A \times F_B) \) where for all \( s \) in \( K_A \), \( q \) in \( K_B \), and \( a \) in \( \mathcal{S} \),

\[
\delta_A(s, a) = t \implies <s, q, a, <t, q>> \sqcup \emptyset, \quad \text{and} \quad \delta_B(q, a) = r \implies <s, q, a, <t, r>> \sqcup \emptyset (\text{and those are the only elements in } \emptyset).
\]

Thus each move of \( M \) corresponds to either a move of the first automaton based on the first state or a move of the second automaton based on the second state.

Suppose \( w \) in \( A \shuffle B \). Then \( w = c_1 \ldots c_k \) where \( w \) can be decomposed into \( a = c_{a1} \ldots c_{aj} \) and \( b = c_{b1} \ldots c_{bm} \) where \( j + m = k \), \( a \) in \( A \), \( b \) in \( B \), and the subscripts \( a1, \ldots, ak \) and \( b1, \ldots, bk \) are each in increasing order (this follows because \( w \) is in the shuffle). When running \( M \) on \( w \), each time a character from \( a \) is read, take the transition corresponding to \( M_A \), while each time a character from \( b \) is read, take a transition corresponding to \( M_B \). Because \( a \) is accepted by \( M_A \), running \( M_A \) on \( w \) ends in a final state \( f_a \) of \( M_A \), and because moves corresponding to \( M_B \) don't change the first component of the state, after reading \( w \), \( M \) will end up in a state whose first component is \( f_a \). By a similar argument on \( b \), when running \( M \) on \( w \), the second component will end up in a final state \( f_b \) of \( M_B \). Thus, \( w \) will be accepted by \( M \).

Now suppose \( w \) is accepted by \( M \). Look at the accepting computation, and mark all of the characters of \( w \) which, when read, used a transition where the state that changed was from \( M_A \). Remove those characters from \( w \) to obtain \( w_A \). Let the remaining characters form \( w_B \).
Because running $M$ on $w$ ended in state $<f_A,f_B>$, where $f_A$ in $F_A$ and $f_B$ in $F_B$, running $M_A$ on $w_A$ results in state $f_A$, while running $M_B$ on $w_B$ results in state $f_B$ (because all the marked transitions took moves from $M_A$, while all of the unmarked took moves from $M_B$). Hence $w$ is a shuffle of $w_A$ and $w_B$ where $w_A$ in $A$ and $w_B$ in $B$. I.e., $w$ in $A \| B$.

5. Prove that $L = \{a^{n^2} \mid n \geq 0\}$ is not regular. Hint: Think about the difference between successive perfect squares.

Play the pumping game. They choose a number $k$. I choose $w = a^{k^2}$. They break $w$ into $xyz$ where $|xy| \leq k$ and $y \neq e$. Thus $x = a^j$, $y = a^m$, and $z = a^{k^2-j-m}$.

I pick $i = 2$ and show that $xy^2z$ is not in $L$. $xy^2z = a^i a^{2m} a^{k^2-j-m} = a^{k^2+m}$. But $0 < m \leq k$ because $y \neq e$ and $|xy| \leq k$. Therefore $k^2 < k^2 + m \leq k^2 + k < k^2 + 2k + 1 = (k+1)^2$. Because there are no primes between $k^2$ and $(k+1)^2$, $k^2 + m$ cannot be a square, and hence $xy^2z$ is not in $L$.

Because we won the pumping game, $L$ is not regular.