Solutions of homework 2

1. Sipser problem 1.41.

Solution Outline: (15 points)

Interpreting the number of occurrences of the substrings 01 and 10 as the number of transitions from 0 to 1 and from 1 to 0 respectively, it is easy to see that \( w \in D \) iff \( w \) starts and ends with the same symbol. Hence, \( D \) is described by the regular expression \((0 \circ (0 \cup 1)^* \circ 0) \cup (1 \circ (0 \cup 1)^* \circ 1)\), and is therefore regular.

2. A string is a palindrome if it reads the same way forward and backward, like radar, or 1101011. Show that for every alphabet \( \Sigma \) (with \( |\Sigma| \geq 2 \)), the language of palindromes over \( \Sigma \) is not regular.

Solution Outline: (15 points)

Assume to the contrary that the language \( L \) of palindromes is regular. Let \( p \) be the pumping length given by the pumping lemma. Now, pick any two distinct elements \( a, b \) in \( \Sigma \), and consider the palindrome \( s = a^p bba^p \in L \). The pumping lemma guarantees that \( s \) can be split into 3 pieces \( s = xyz \), where \( |xy| \leq p \). Hence, \( y = a^i \) for some \( i \geq 1 \). Then, \( xy^2z = a^{p+i} bba^p \in L \), but is not a palindrome, a contradiction.

3. (a) Let \( A \) be the set of strings over \( \{0, 1\} \) that can be written in the form \( 1^k y \) where \( y \) contains at least \( k \) 1s, for some \( k \geq 1 \). Show that \( A \) is a regular language.

[Note that the same string could fit the definition for more than one value of \( k \). For example 1101010 can be seen as 1 followed by the string \( y = 101010 \), which contains at least one 1, or as 11 followed by 01010. On the other hand, the string 100, for example, is not in \( A \) because there is no value of \( k \) for which the definition applies.]

(b) Let \( B \) be the set of strings over \( \{0, 1\} \) that can be written in the form \( 1^k 0 y \) where \( y \) contains at least \( k \) 1s, for some \( k \geq 1 \). Show that \( B \) is not a regular language.

(c) Let \( C \) be the set of strings over \( \{0, 1\} \) that can be written in the form \( 1^k z \) where \( z \) contains at most \( k \) 1s, for some \( k \geq 1 \). Show that \( C \) is not a regular language.

Solution Outline: (15, 15, 15 points)

(a) It is easy to see that any string in \( A \) must start with a 1, and contain at least one other 1 (in the matching \( y \) segment). Conversely, any string that starts with a 1 and contains at least one other 1 matches the description for \( k = 1 \). Hence, \( A \) is described by the regular expression \( 1 \circ 0^* \circ 1 \circ (0 \cup 1)^* \), and is therefore regular.

(b) Assume to the contrary that \( B \) is regular. Let \( p \) be the pumping length given by the pumping lemma. Consider the string \( s = 1^p 0^p 1^p \in B \). The pumping lemma guarantees that \( s \) can be split into 3 pieces \( s = abc \), where \( |ab| \leq p \). Hence, \( y = a^i \) for some \( i \geq 1 \). Then, by the pumping lemma, \( ab^2c = 1^{p+i} 0^p 1^p \in B \), but cannot be written in the form specified, a contradiction.
(c) Assume to the contrary that $C$ is regular. Let $p$ be the pumping length given by the pumping lemma. Consider the string $s = 1^p0^p1^p \in B$. The pumping lemma guarantees that $s$ can be split into 3 pieces $s = abz$, where $|ab| \leq p$. Hence, $b = 1^i$ for some $i \geq 1$. Then, by the pumping lemma, $ac = 1^p - i0^p1^p \in C$, but cannot be written in the form specified, a contradiction.

REMARK: Note that we can describe $A$ using

\[ \bigcup_{k=1}^{\infty} \left( 1^k \{0,1\}^* 1 \cdots \{0,1\}^* 1 \{0,1\}^* \right)^k \]

However, this expression has infinite length, and does not show that $A$ is regular. Similarly, we can describe $B$ using

\[ \bigcup_{k=1}^{\infty} \left( 1^k 0 \{0,1\}^* 1 \cdots \{0,1\}^* 1 \{0,1\}^* \right)^k \]

and this does not contradict the fact that $B$ is not regular.

4. Let $k$ be a positive integer. Let $\Sigma = \{0,1\}$, and $L$ be the language consisting of all strings over $\{0,1\}$ containing a 1 in the $k$th position from the end (in particular, all strings of length less than $k$ are not in $L$).

(a) Prove that any DFA that recognizes $L$ has at least $2^k$ states.

(b) Prove that any NFA that recognizes $L$ has at least $k$ states.

SOLUTION OUTLINE: (15, 10 points)

(a) Suppose on the contrary that there is a DFA $M$ with at most $2^k - 1$ states that recognizes $L$. Then, by the Pigeonhole Principle, amongst the $2^k$ strings in $\{0,1\}^k$, there are 2 distinct strings $x, y$ such that on input $x$ and on input $y$, $M$ ends up at the same state. Pick any $i$ such that $x, y$ differ in the $i$th position from the end, so that exactly one of $x0^{k-i}$ and $y0^{k-i}$ contains a 1 in the $k$th position from the end. Now, $M$ on input $x0^{k-i}$ and $y0^{k-i}$ end up in the same state, but $M$ is supposed to accept exactly one of the two strings, a contradiction.

(b) METHOD I Suppose on the contrary that there is a NFA $N$ with at most $k - 1$ states that recognizes $L$. Then, we may derive from $N$ a DFA with at most $2^{k-1}$ states that recognizes $L$, a contradiction to part (a).

METHOD II (due to John Firebaugh) We will prove a strong lower bound of $k+1$ states, that matches the upper bound from problem set 1. Again, suppose on the contrary that there is a NFA $N$ with at most $k$ states that recognizes $L$. In particular, $N$ accepts some string of length $k$ (take $1^k$ say). Now, consider the $k + 1$ strings $\epsilon, 1, 11, \ldots, 1^k$. As in the proof of the pumping lemma, the accepting path will revisit some state twice, and deleting that loop, we obtain a string of length less than $k$ that $N$ accepts, a contradiction.