Due: beginning of class on Tue. Feb. 5, 2002

In the first homework assignment (problem 4), you were given five languages $L_1, \ldots, L_5$, two of which ($L_1$ and $L_5$) were regular, and the remaining three were not. You can find the proof that $L_2$ is not regular in the textbook (at the end of chapter 1). In this problem set you will show that the remaining languages are not regular.

Problems 1

Prove, using the pumping lemma, that

$$L_4 = \{ w \in \Sigma^* : w \text{ contains more 0's than 1's} \}$$

is not a regular language.

(P.S.: you will be asked to prove that $L_3$ is not regular in problem 3, using the closure properties of regular languages. But first you will prove that regular languages are closed under two new operations. If you wish, you can read the definition of these operations from problem 2, and then move directly to the solution of problem 3, before solving problems 2.)

Problem 2

There are two possible ways to encode signals: level transitions or pulses. When using pulses, you normally send a 0, and when you want to send a signal you send a 1. E.g., if you want to send signals at times 3,6,7, you would send string 001001100.... The same signals can be encoded using level transitions by the string 001101011111111.... This time you alternate the transmission of 0's with the transmission of 1's, with changes from one symbol to the other in correspondence to the times when you want to send the signals.

Both ways to encode signals are useful, and actually used in different applications, and sometime one needs to translate from one to the other. In this problem you will prove that a set of signals can be recognized by a finite state automaton independently from the encoding chosen for the signals.

Part (a) Let $T$ be the translation function from impulses to levels. For example $T(000100100100) = 000111000111$, $T(011111) = 010101$, $T(111) = 101$. The translation function is applied to languages in the obvious way: $T(L) = \{ T(w) : w \in L \}$.

Prove that if $L$ is regular, then also $T(L)$ is regular.
Part (b) Define the inverse translation function $I(w) = T^{-1}(w)$, from level transitions to pulses. E.g., $I(0001110101) = 0001001111$ $I(11111) = 100000$, $I(\epsilon) = \epsilon$.

Prove that if $L$ is regular, then $I(L)$ is also regular.

Problem 3

Prove that the language

$$L_3 = \{ w \in \Sigma^* : \text{the length of } w \text{ is odd and the middle character of } w \text{ is a 1} \}$$

is not regular, using the non regularity of any of the following languages studied in class

- $L_\approx = \{ w : w \text{ contains the same number of 0's as 1's} \}$
- $L_\succ = \{ w : w \text{ contains more 0 than 1's} \}$
- $L_{nn} = \{ 0^n1^n : n \geq 0 \}$

and the closure of regular languages under the regular operations (union, concatenation, star), and the translation operation $T$ and $I$ defined in problem 2. (You may also use the fact that languages $\{0\}$ and $\{1\}$ are regular, and therefore any language denoted by a regular expression is regular.)

Problem 4

For each of the following languages, say whether it is regular or not, and prove your answer. You can use any (combination) of the techniques we studied to prove that languages are regular or non regular. (I.e., for regular languages you can either give a DFA, NFA, RE, or use the closure properties, while for non-regular languages you can use the pumping lemma and/or the closure properties of regular languages.)

- The set of all strings of the form $a^m b^n a^{m+n}$ where $m, n \geq 0$.
- The set of all strings of the form $xwx^R$, where $x, w$ are arbitrary non-empty strings over the alphabet $\Sigma = \{a, b\}$, and $w^R$ is the reverse of $w$, i.e., word $w$ spelled backward. (E.g., $(aabbba) = babbba$.)
- The set of all strings $xx^Rw$, where $x, w$ are arbitrary non-empty strings over the alphabet $\Sigma = \{a, b\}$. 