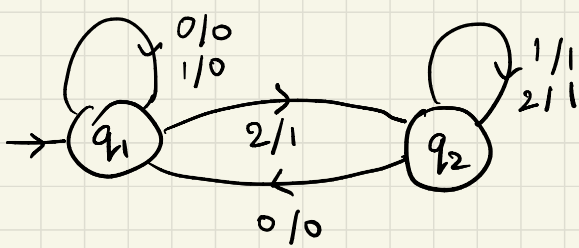
Assignment 1

Due: 5pm on 29^{th} Aug.

Instructions

- Discussion is allowed and infact encouraged
- Answers **must** be written by yourself.
- All sources (including discussions) that are used to reach the solution must be mentioned.

① A finite state transducer (FST) is a type of deterministic finite automaton whose output is a string and not just accept or reject. The following is an example:



T.

Each transition of an FST is labelled with two symbols, one designating the input symbol for that transition and the other designating the output symbol. The two symbols are written with a slash separating them (i.e. input/output). Some transitions may have multiple input - output pairs. When an FST gets an input $w = w_1 w_2 \cdots w_n$, it starts at the initial state and follows the transitions by matching the input labels with the sequence of symbols $w_1 w_2 \cdots w_n$. Every time it goes along a transition, it outputs the corresponding output symbol. For example, on input 2212011, T enters the sequence of states $q_1, q_2, q_2, q_2, q_2, q_1, q_1, q_1$ and produces output 1111000.

Using the above informal definition of an FST, give the state diagram of an FST with the following behaviour and prove its correctness.

- Input & output alphabets are $\{0,1\}$
- Output string is identical to the input string on the even positions and inverted $(0 \leftrightarrow 1)$ on the odd positions.

For example, on input 0000111, it should output 1010010.

[4+4]

(2) Let L be a language and X be a set of strings. We say that for $x, y \in X$,

x and y are distinguishable by L if $\exists z \in X$ s.t. either $xz \in L$ or $yz \in L$ but not both. If this is not the case, then x and y are said to be indistinguishable by L. That is, x, y are indistinguishable by L if $\forall z \in X$, $xz \in L$ if and only if $yz \in L$.

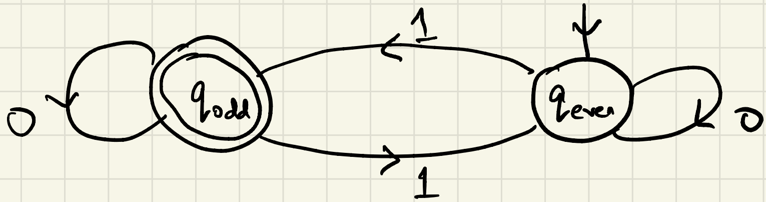
X is said to be pairwise distinguishable by L if $\forall x, y \in X$ if $x \neq y$, then x and y are distinguishable by L.

Define the index of L to be the max. no. of elements in any set of strings that is pairwise distinguishable by L.

Note that the index of L may be finite or infinite.

i) If x and y are indistinguishable by L, then we write $x \equiv_L y$. Show that $\equiv_L$ is an equivalence relation.

ii) Let L be the language recognised by the following DFA:



What is the index of L? Why?

iii) Guess what a general statement about the index of a language and the size of the smallest DFA recognising it might be.

[4+6+2]

③ Construct an NFA recognising the language corr. to $(01 \cup 001 \cup 010)^*$ using at most 5 states. Prove correctness.

[Hint: Describe the lang. in words] [4]

④ Is the following language regular? Prove your claim:

PAREN $\doteq \{ w \in \{(,)\}^* :$ w is a balanced string of paranthesis $\}$. [4]

A seq. of paranthesis is said to be balanced if it "makes sense". That is, every 'open paranthesis' has a corr. 'closing paranthesis' after it and every 'closing paranthesis' has a corr. 'open paranthesis' before it. For example,

$(())\in$ PAREN, but $())(\notin$ PAREN.

(5) (a) Given a DFA for a language $A \subseteq \Sigma^*$, construct a DFA for $A^c = \Sigma^* \setminus A$. Prove correctness.

(b) Fill in the blanks:

(a) shows that ________ are ________ under taking complements.

[3+1]

Something to think about: Why do you think ^{c}, $\cap$ operations are not used while creating regular expressions?

(6) A homomorphism is a function $f : \Sigma \to \Gamma^*$ from one alphabet to strings over another alphabet. We can extend f to operate on strings by defining

$$f(w) = f(w_1) \cdot f(w_2) \cdot \cdots \cdot f(w_n)$$

where $w = w_1 w_2 \ldots w_n$, and each $w_i \in \Sigma$.

We further extend f to operate on languages by defining

$$f(A) = \{ f(w) : w \in A \}$$ for any language A.

(a) Given an example of a language A and a function f such that A is not regular, but $f(A)$ is regular.

(b) Fill in the blanks:

(a) shows that ________ are ________ under homomorphisms.

[2+1]

(7) Let G be a CFG in Chomsky Normal Form that contains b variables.

a) Suppose a word is derived using G in k steps. Prove that k must be odd. What is the number of leaves in the parse tree corresponding to this derivation?

b) Suppose we have a tree with l leaves. What is the minimum height of the tree?

c) How many vertices are present in the longest root to leaf path of a tree of height h?

d) Show that if there is a word in $L(G)$ which requires $\geq 2^b$ steps to derive, then $L(G)$ is infinite.

[3 + 1 + 1 + 2]

(8) Let $A \setminus B = \{w : wx \in A \text{ for some } x \in B\}$. Show that if A is context-free and B is regular, then $A \setminus B$ is context-free. [4]

(9) Give an example to show that the intersection of context-free languages need not be context-free. That is, give examples of languages L_1 and L_2 such that

L_1 and L_2 are context free but

$L_1 \cap L_2 = \{w : w \in L_1 \cap L_2\}$ is not. [4]

x ———————— x