# Problem Set 4

**Fall 11**

**Due:** Thursday, 20th October, 2011, 11:00 am before class begins
Please follow the homework format guidelines posted on the class web page:
`http://www.cs.illinois.edu/class/fa11/cs373/`
Also, note that Problem 5 is an extra credit question.

1. [**Category**: Non-regularity, **Points**: 30]

Prove the following languages are not regular, from first principles, using the Myhill-Nerode Theorem or using the pumping lemma.

You cannot assume the non-regularity of *any* language to solve this problem.

(a) $L = \{w \in \{a,b\}^* \mid N_b(w) = 2N_a(w)\}$ is non-regular. ($N_a(w)$ denotes the number of $a$'s in $w$; hence $L$ contains all words over $\{a,b\}$ in which the number of $a$'s is precisely twice the number of $b$'s.)

## Solution:

**Proof using the pumping lemma**

Assume $L$ is regular. Then the pumping lemma applies to $L$. Let $p$ be the number that satisfying the pumping lemma's consequence. Consider the string $s = a^p b^{2p}$. It is clear that $s \in L$ and $|s| > p$. Now let $x, y, z \in \Sigma^*$ be any three words such that $s = xyz$, $|xy| \leq p$, and $|y| > 0$. Then both $x$ and $y$ consist of $a$s; let $y = a^j (j \geq 1)$. Set $n = 0$. Then $xy^n z = xy^0 z = a^{p-j} b^{2p} \notin L$ as as $N_b(xy^0z) = 2p > 2(p-j) = 2N_a(xy^0z)$. This contradicts the pumping lemma. The contradiction shows that the assumption that $L$ is regular is false. Hence $L$ is not regular.

**Proof using Myhill-Nerode Theorem**

Let $S = \{a^i \mid i \geq 0\}$. Clearly $S$ is infinite. Let $x = a^i$ and $y = a^j$ be two different elements in $S$. Without loss of generality, assume that $i < j$. Now choose $z = b^{2j}$. Then $xz = a^i b^{2j} \notin L$ but $yz = a^j b^{2j} \in L$. Hence $[L/x]$ does not contain $z$ but $[L/y]$ contains $z$. Hence the suffix languages $[L/x]$ and $[L/y]$ are different, for every $x, y \in S$, $x \neq y$. Hence there are infinitely many suffix languages for $L$, and by Myhill-Nerode Theorem, $L$ is not regular.

(b) $L = \{0^{i^2+1} | i \geq 0\}$

## Solution:

**Proof using the pumping lemma**

Assume $L$ is regular.
Then the pumping lemma applies to $L$.

Let $p$ be the number that satisfying the pumping lemma's consequence.
Consider the string $s = 0^{p^2+1}$. It is clear that $s \in L$ and $|s| > p$.
Now let $x, y, z$ be any three words such that $s = xyz$, $|xy| \leq p$, and $|y| \geq 1$.
Let $|x| = i$, $|y| = j$, and $|z| = k$. Then $p^2 + 1 = i + j + k$ and $j > 1$ and $i + j < p$.
Set $n = 2$. Then the pumping lemma consequence says that $xy^2z \in L$.
Clearly, the next longer word in $L$ (shortest word in $L$ that is longer than $|s|$) is of length $(p+1)^2 + 1 = p^2 + 2p + 2$.
Now, $|xy^2z| = i + 2j + k = (p^2 + 1) + j < p^2 + 2p + 1$ since $j < p$.
Also, $|xy^2z| > p^2 + 1$, since $j > 1$.
Hence the length of $|xy^2z|$ falls strictly between $p^2 + 1$ and $(p+1)^2 + 1$.
So $|xy^2z|$ cannot be of the form $h^2 + 1$, and hence $xy^2z \not\in L$. This contradicts the pumping lemma.
So $L$ cannot be regular.

**Proof using Myhill-Nerode Theorem**

Let $S = \{0^{i^2+1} \mid i \geq 0\}$. Clearly $S$ is infinite.
Let $x = 0^{i^2+1}$ and $y = 0^{j^2+1}$ be two different elements in $S$.
Without loss of generality, assume that $i < j$.
Now choose $z = 0^{2i+1}$.
Then $xz = 0^{i^2+1}0^{2i+1} = 0^{(i+1)^2+1} \in L$ but $yz = 0^{j^2+1}0^{2i+1} = 0^{j^2+2i+2} \notin L$, since $j^2 + 1 < j^2 + 2i + 2 < (j+1)^2 + 1$.
Hence $z \in [L/x]$ and $z \notin [L/y]$.
Hence the suffix languages $[L/x]$ and $[L/y]$ are different.
Hence there are infinitely many suffix languages for $L$ and, by MNT $L$ is not regular.

(c) Consider the following language $L$ over the alphabet $\{0, 1\}$:

$$L = \{wtw \mid w, t \in \{0, 1\}^+\}$$

(Recall, $\{0, 1\}^+$ is the set of all binary strings not including $\epsilon$; i.e. $\{0, 1\}^+ = \{0, 1\}.\{0, 1\}^*$.) Thus, $111011 \in L$, because $111011 = 11.10.11$. On the other hand, $10100 \notin L$ because there is no length $i > 0$ such that the first $i$ symbols are the same as the last $i$ symbols.

# Solution:

**Proof using the pumping lemma**

Assume $L$ is regular.
Then the pumping lemma applies to $L$.
Let $p$ be the number that satisfying the pumping lemma's consequence.
Consider the string $s = 0^p1^p10^p1^p$. It is clear that $s \in L$ because $0^p1^p10^p1^p = (0^p1^p)1(0^p1^p)$. By the pumping lemma, we know that there exist a division $s = xyz$, $|xy| \leq p$, and $|y| \geq 1$. Hence both $x$ and $y$ consist of 0s, say $x = 0^i$,

$y = 0^j (j \geq 1)$. By the pumping lemma, $xy^2z = 0^{p+j}1^p10^p1^p$ also belongs to $L$. But this is not true, because for arbitrary $0 < k \leq p$, the first $k$ symbols are $0^k$ while the last $k$ symbols are $1^k$; for arbitrary $k > p$, the first $k$ symbols start with more than $p$ 0s, but in the last $k$ symbols, the longest sequence of 0s is only $0^p$. The contradiction concludes the proof.

**Proof using Myhill-Nerode Theorem**

Let $S = \{0^i1^i1 \mid i > 0\}$. Clearly $S$ is infinite. Let $x = 0^i1^i1$ and $y = 0^j1^j1$ be two different elements in $S$. Without loss of generality, assume that $i < j$. Now choose $z = 0^i1^i$. Then $xz = 0^i1^i10^i1^i \in L$ but $yz = 0^j1^j10^i1^i \notin L$, because it is not of the form $wtw$ for any $w, t \in \Sigma^+$. The latter is true because for arbitrary $0 < k \leq i$, the first $k$ symbols of $yz$ are $0^k$ while the last $k$ symbols are $1^k$ (hence leaving no choice for $w$); and for arbitrary $k > i$, the first $k$ symbols start with more than $i$ 0s, but in the last $k$ symbols, the longest sequence of 0s is only $0^i$.
Hence $x$ and $y$ are distinguishable w.r.t. $L$.
I.e. $z \in [L/x]$ and $z \notin [L/y]$.
Hence $[L/x]$ and $[L/y]$ are different, for every $x, y \in S$, $x \neq y$.
Hence there are infinitely many suffix languages for $L$. and by MNT, we $L$ is not regular.

2. [**Category**: Non-regularity using closure properties, **Points**: 20]

Prove that the following languages are not regular using *only* closure properties (you cannot use the Pumping Lemma or the MNT technique). You may assume that regular languages are closed under union, intersection, concatenation, Kleene-*, complement, and reverse. The only language you can assume non-regular is $\{0^n1^n \mid n \geq 0\}$.

(a) $L_{eq} = \{w \in \{0,1\}^* \mid N_0(w) = N_1(w)\}$ ($N_0(w)$ denotes the number of 0's in $w$ and $N_1(w)$ denotes the number of 1's in $w$)

## Solution:

Assume $L_{eq}$ is regular. Then by closure properties $A = L_{eq} \cap L(0^*1^*)$ is regular too. Note that all the strings in $A$ must be of the form $0^n1^n$, where $n \geq 0$. Hence $A = \{0^n1^n \mid n \geq 0\}$. But we know that this set is not regular and therefore we have got a contradiction. So $L_{eq}$ is not regular.

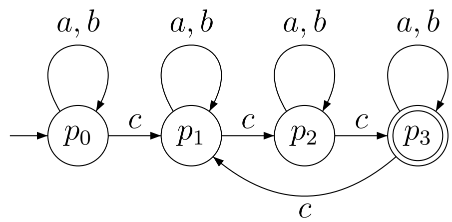
(b) $L_2 = \{0^{2n}1^{2n} \mid n \geq 0\}$

## Solution:

Assume $L_2$ is regular. Then by closure properties $A = L_2 \cup (\{0\}L_2\{1\})$ is regular too. Now note that all the strings of the form $0^n1^n$, where $n \geq 0$ are in $A$ (when $n$ is even $0^n1^n \in L_2$ and when $n$ is odd $0^n1^n \in (\{0\}L_2\{1\})$), therefore

$A = \{0^n 1^n \mid n \geq 0\}$. But we know that this last set is not regular and therefore we have got a contradiction. So $L_2$ is not regular.

3. [**Category**: Suffix Language, **Points**: 20]




(a) For each state $q$ in the above DFA, give the language accepted from $q$ (i.e. $L_q$). Also, for each state $q$, give a string $x$ such that $L_q = [\![L/x]\!]$.

(b) Prove that all the languages defined by the states are different: for each pair of states $q, q'$, show that $L_q \neq L_{q'}$, by giving a string that belongs to one language but not the other. Note that you need to give 6 such strings, one for each pair of languages.

(c) Is this DFA a minimal DFA for the language it accepts? Why?

## Solution:

Rubric: 1.5 points for each answer in part a and b; 5 points for c, no partial credit.

(a) $[\![L/\epsilon]\!] = L_{p_0} = (a+b)^*c(a+b)^*c(a+b)^*c(a+b)^*$
$[\![L/c]\!] = L_{p_1} = (a+b)^*c(a+b)^*c(a+b)^*$
$[\![L/cc]\!] = L_{p_2} = (a+b)^*c(a+b)^*$
$[\![L/ccc]\!] = L_{p_3} = (a+b)^*$

(b) $ccc \in L_{p_0}$ *but* $ccc \notin L_{p_1}$
$ccc \in L_{p_0}$ *but* $ccc \notin L_{p_2}$
$ccc \in L_{p_0}$ *but* $ccc \notin L_{p_3}$
$cc \in L_{p_1}$ *but* $ccc \notin L_{p_2}$
$cc \in L_{p_1}$ *but* $ccc \notin L_{p_3}$
$c \in L_{p_2}$ *but* $ccc \notin L_{p_3}$

(c) Yes, since all suffix languages are disjoint.