CS/ECE 374 A (Spring 2022)
Homework 1 Solutions

Problem 1.1: Let $L \subseteq \{0,1\}^*$ be the language defined recursively as follows:

- The empty string ε is in L.
- For any string x in L, the strings $0101x$ and $1010x$ are also in L.
- For any strings x, y such that xy is in L, the strings $x00y$ and $x11y$ are also in L. (In other words, inserting two consecutive 0's or two consecutive 1's anywhere to a string in L yields another string in L.)
- The only strings in L are those that can be obtained by the above rules.

Define $L_{ee} = \{x \in \{0,1\}^* : x \text{ has an even number of 0's and an even number of 1's}\}$.

(a) Prove that $L \subseteq L_{ee}$, by using induction. (You should use *strong* induction.)

(b) Conversely, prove that $L_{ee} \subseteq L$, by using induction.

Solution: Let $\#_0(x)$ denote the number of 0's in x, and $\#_1(x)$ denote the number of 1's in x.

(a) It suffices to prove the following claim:

Claim 1. *For every string $w \in L$, the numbers $\#_0(w)$ and $\#_1(w)$ are both even.*

Proof. The proof is by (strong) induction on the length $|w|$.

Base case: $|w| = 0$. Here, $w = \varepsilon$ and $\#_0(w) = \#_1(w) = 0$, so the claim is trivially true.

Induction hypothesis. Suppose $n \geq 1$. Assume that for every string $w' \in L$ with $|w'| < n$, the numbers $\#_0(w')$ and $\#_1(w')$ are both even.

Induction step. Let $w \in L$ with $|w| = n$. We want to prove that $\#_0(w)$ and $\#_1(w)$ are both even.

By the recursive definition of L, we know that one of the following cases must hold:

- CASE 1: $w = 0101x$ for some string $x \in L$. Since $|x| = |w| - 2 < n$, by the induction hypothesis, $\#_0(x)$ and $\#_1(x)$ are both even. So, $\#_0(w) = \#_0(x) - 2$ and $\#_1(w) = \#_1(x) - 2$ must be both even.
- CASE 2: $w = 1010x$ for some string $x \in L$. This case is similar to Case 1 (with 0's and 1's swapped).
- CASE 3: $w = x00y$ for some x, y with $xy \in L$. Since $|xy| = |w| - 2 < n$, by the induction hypothesis, $\#_0(xy)$ and $\#_1(xy)$ are both even. So, $\#_0(w) = \#_0(xy) - 2$ and $\#_1(w) = \#_1(xy)$ are both even.
- CASE 4: $w = x11y$ for some x, y with $xy \in L$. This case is similar to Case 3 (with 0's and 1's swapped).

□

(b) It suffices to prove the following claim:

Claim 2. *For every string $w \in \{0,1\}^*$ such that $\#_0(w)$ and $\#_1(w)$ are both even, we must have $w \in L$.*

Proof. The proof is by (strong) induction on the length $|w|$.

Base case: $|w| = 0$. Here, $w = \varepsilon$ and by definition of L, we have $w \in L$.

Induction hypothesis. Suppose $n \geq 1$. Assume that for every string $w' \in \{0,1\}^*$ of length smaller than n such that $\#_0(w')$ and $\#_1(w')$ are both even, we must have $w' \in L$.

Induction step. Let $w \in \{0,1\}^*$ with $|w| = n$ such that $\#_0(w)$ and $\#_1(w)$ are even. We want to prove that $w \in L$.

One of the following three cases must be true:

- Case 1: w contains 00 as a substring. Then $w = x00y$ for some string $x, y \in \{0,1\}^*$. Since $\#_0(xy) = \#_0(w) - 2$ is even and $\#_1(xy) = \#_1(w)$ is even and $|xy| = |w| - 2 < n$, we have $xy \in L$ by the induction hypothesis. So, $w \in L$ by the recursive definition of L.
- Case 2: w contains 11 as a substring. Similar to Case 1.
- Case 3: w does not contain 00 nor 11 as a substring. Then w must alternate between 0 and 1; more precisely, (i) if the first symbol in w is 0, then the second symbol must be a 1, the third must be a 0, etc.; (ii) if the first symbol in w is 1, then the second must be a 0, the third must be a 1, etc. In subcase (i), w begins with 0101 and can be written as $w = 0101x$ for some string $x \in \{0,1\}^*$. Since $\#_0(x) = \#_0(w) - 2$ and $\#_1(x) = \#_1(w) - 2$ are even, we have $w \in L$ by the recursive definition of L. In subcase (ii), the argument is similar. $\square$

Problem 1.2: Let $L = \{x \in \{0,1,\ldots,9\}^* : x \text{ does not contain 374 as a substring}\}$.

Obviously, the number of strings in $\{0,1,\ldots,9\}^*$ of length n is equal to 10^n.

Prove that the number of strings in L of length n is at most $2 \cdot 9.992^n$, by using induction.

[Hint: consider two cases: x does not start with 3, or starts with 3. In the second case, consider two subcases: the second symbol is not 7, or is 7.]

[We may give bonus points for a proof of an upper bound better than $O(9.990^n)$.]

Solution: By induction on n.

Base cases: $n \in \{0,1,2\}$. The number of strings in L of length n is 0 for $n = 0$ and 10 for $n = 1$ and 100 for $n = 2$, and $0 \leq 2 \cdot 9.992^0$, $10 \leq 2 \cdot 9.992^1$, and $100 \leq 2 \cdot 9.992^2$.

Induction hypothesis. Suppose $n \geq 3$. Assume that for all $m < n$, the number of strings in L of length m is at most $2 \cdot 9.992^m$.

Induction step. If x is a string in L of length n, then one of the following cases must hold:

- Case 1: the first symbol of x is not 3. Then $x = ay$ for some symbol $a \in \{0,1,\ldots,9\} \setminus \{3\}$ and some string y. Note that y has length $n-1$ and cannot contain 374 as a substring and so is in L. Thus, there are at most $2 \cdot 9.992^{n-1}$ choices for y, and there are 9 choices for a. So, the number of strings x in Case 1 is at most $9 \cdot 2 \cdot 9.992^{n-1}$ by the induction hypothesis.

- Case 2: the first symbol of x is 3.
  - Subcase 2.1: the second symbol of x is not 7. Then $x = 3bz$ for some symbol $b \in \{0, 1, \ldots, 9\} \setminus \{7\}$ and some string z. Note that z has length $n - 2$ and cannot contain 374 as a substring and so is in L. Thus, there are at most $2 \cdot 9.992^{n-2}$ choices for z, and there are 9 choices for b. So, the number of strings x in Subcase 1.2 is at most $9 \cdot 2 \cdot 9.992^{n-2}$ by the induction hypothesis.
  - Subcase 2.2: the second symbol of x is 7. Since $x \in L$, x cannot contain 374 as a substring and so the third symbol of x cannot be 4. Thus, $x = 37cw$ for some symbol $c \in \{0, 1, \ldots, 9\} \setminus \{4\}$ and some string w. Note that z has length $n - 3$ and cannot contain 374 as a substring and so is in L. Thus, there are at most $2 \cdot 9.992^{n-3}$ choices for w, and there are 9 choices for c. So, the number of strings x in Subcase 1.2 is at most $9 \cdot 2 \cdot 9.992^{n-3}$ by the induction hypothesis.

Therefore, the total number of strings $x \in L$ of length n is at most

$$
\begin{aligned}
9 \cdot 2 \cdot 9.992^{n-1} + 9 \cdot 2 \cdot 9.992^{n-2} + 9 \cdot 2 \cdot 9.992^{n-3} \quad &\leq \quad 2 \cdot 9.992^n \cdot \left(\frac{9}{9.992} + \frac{9}{9.992^2} + \frac{9}{9.992^3} \right) \\
&< \quad 2 \cdot 9.992^n \cdot 0.9998864 \\
&< \quad 2 \cdot 9.992^n.
\end{aligned}
$$

Remark: Alternatively, one can explicitly set up a recurrence for the number f_n of strings in L of length n. The above argument yields $f_n \leq 9f_{n-1} + 9f_{n-2} + 9f_{n-3}$, with base cases $f_0 = 1$, $f_1 = 10$, and $f_2 = 100$...

Bonus Solution (sketch): Intuitively, there is room for improvement, for example, in Subcase 2.1: if $b = 3$, not only do we know that y cannot contain 374 as a substring, but also that z cannot start with 74. There are similar potential room for improvement in Subcase 2.2.

More precisely, let f_n be the number of strings in L of length n. Let g_n be the number of strings in L of length n that does not start with 74. Let h_n be the number of strings in L of length n that does not start with 4.

Arguments similar to above yield the following system of recurrences:

$$
\begin{aligned}
f_n &= 9f_{n-1} + g_{n-1} \\
g_n &= 8f_{n-1} + g_{n-1} + h_{n-1} \\
h_n &= 8f_{n-1} + g_{n-1}
\end{aligned}
$$

with the base cases $f_0 = g_0 = h_0 = 1$. (There is actually a systematic way to generate these recurrences by looking at the DFA for L...)

One way to solve this system of recurrences is to use the following induction hypotheses: $f_n \leq c\alpha^n$ and $g_n \leq c\beta\alpha^n$ and $h_n \leq c\gamma\alpha^n$ for some appropriate choice of constants α, β, γ, c. I will not write out the details formally, but one can check that the induction proof goes through for a sufficiently large c if we choose $\alpha, \beta, \gamma > 0$ to satisfy

$$
\begin{aligned}
\alpha^n &= 9\alpha^{n-1} + \beta\alpha^{n-1} \\
\beta\alpha^n &= 8\alpha^{n-1} + \beta\alpha^{n-1} + \gamma\alpha^{n-1} \\
\gamma\alpha^n &= 8\alpha^{n-1} + \beta\alpha^{n-1},
\end{aligned}
$$

i.e.,

$$\begin{aligned}
\alpha &= 9+\beta \\
\beta\alpha &= 8+\beta+\gamma \\
\gamma\alpha &= 8+\beta.
\end{aligned}$$

These conditions are equivalent to $\beta = \alpha - 9$, and $\gamma = (\alpha-1)/\alpha$, and $(\alpha-9)\alpha = \alpha - 1 + (\alpha-1)/\alpha$. The latter simplifies to the cubic equation $\alpha^3 - 10\alpha^2 + 1 = 0$, which has root $\alpha = 9.9899799\cdots < 9.990$.