COMPSCI 2AC3, Automata and Computability
Assignment 3 Solution

Alireza Fathollah Pour

March 28, 2023

Question 1

We know that the first bit of n is always different than the first bit of $n+1$. If n ends with a 0 then $n+1$ ends with a bit of 1 and the rest of the bits are the same for n and $n+1$. On the other hand, if n ends with 1 then $n+1$ ends with 0 and also the second bit to the last of $n+1$ should be added by 1 and thus differs from that of n. We can use the same reasoning for other bits and break the grammar in two cases: When the ith bit of n and $n+1$ are similar and when they are different. This leads to the following grammar.

$$\begin{aligned}
&S \to 0A1 \mid 1B0 \mid 1C0\\
&A \to 0A0 \mid 1A1 \mid 1\$1\\
&B \to 1B0 \mid 0A1 \mid 1\$10\\
&C \to \$1
\end{aligned}$$

Intuitively, if the last bit of n is 0, we produce $0A1$ and the rest of the bits between n and $n+1$ are similar. Therefore, we either generate $1A1$ or $0A0$. Another case is where the last bit of n is 1. In this case, we should generate $1B0$. Then, from B we can generate $1B0$ up to the point that the bits differ between n and $n+1$. When we reach at a point where the rest of the bits are going to be similar, we generate $0A1$ and the rest would be similar to the previous case. The $1C0$ case is for generating $rev(f(1))\$f(2)$, i.e., $1\$10$.

Question 2

2a.

We prove that this set is not a CFL using pumping lemma. For any $k \geq 0$ consider $z = a^p$ where $p \geq k$ is the smallest prime number that is larger than or equal to k. It is obvious that $|z| \geq k$. Let $z = uvwxy = a^{t_1}a^{t_2}a^{t_3}a^{t_4}a^{t_5}$ be a way of breaking z such that $vx \neq \epsilon$ and $\sum_i t_i = p$. This means that $t_2 + t_4 \geq 1$ or equivalently at least one of them is larger than 0. Choose $i = p$. Then we have

$$uv^pwx^py = a^{t_1}a^{t_2+p}a^{t_3}a^{t_4+p}a^{t_5} = a^{3p}$$

if both t_2 and t_4 are larger than 0. Otherwise, if either $v = \epsilon$ or $x = \epsilon$ then $uv^pwx^py = a^{2p}$. Note that as mentioned above, it is not possible that both t_2 and t_4 are equal to 0. Since neither $2p$ nor $3p$ is a prime number we conclude that $uv^pwx^py \notin B$ and thus B is not CFL.

2b.

We prove that that C is a CFL. Particularly, we prove that C is generated by the following CFL, which we denote by G.

$$
\begin{aligned}
S &\to LC \mid AR \\
L &\to aLb \mid aA \mid bB \\
R &\to bRc \mid bB \mid cC \\
A &\to aA \mid \epsilon \\
B &\to bB \mid \epsilon \\
C &\to cC \mid \epsilon
\end{aligned}
$$

We first prove that anything generated by G is in set C, i.e., $L(G) \subseteq C$. Starting from S, we can either derive LC or AR. We analyze these two cases separately.

Case 1 $S \to LC$. We use induction on the length of derivation. Denote by γ the sentenial that is either C or ϵ. We conclude that after n derivation, we end up in one of the following cases.

$$
x = \begin{cases} a^i L b^i c^k \gamma & \text{where } n = 2i + k, \\ a^i \alpha b^j c^k \gamma & \text{where } n = i + j + k,\ i > j \text{ and } \alpha \text{ is either } A \text{ or } \epsilon, \\ a^i b^j \beta b^i c^k \gamma & \text{where } n = i + j + k,\ j > 0 \text{ and } \beta \text{ is either } B \text{ or } \epsilon \end{cases} \tag{1}
$$

It can be verified that the only way that x becomes a string with non-terminals is when $\alpha = \beta = \gamma = \epsilon$, which would be either $a^i b^j c^k$, $i > j$ or $a^i b^{i+j} c^k$, $j > 0$, both of which are in $L(a^* b^* c^*) - \{a^n b^n c^n | n \geq 0\}$.

Basis. It is easy to verify that from $S \to LC$, if we apply one production we can derive one of the following sentenials: $LC \to L$, $LC \to LcC$, $LC \to aLbC$, $LC \to aAC$, and $LC \to bBC$, all of which fall in one of the cases defined above for different values of i, j, and k.

Induction step. We assume by our inductive hypothesis that after n productions applied to LC, we generate one of the cases in 2. We now prove that the same holds for $n + 1$. Let η be the sentential form derived after $n + 1$ productions and ζ be the sentential from immediatly preceding η, i.e., $LC \xrightarrow[G]{n} \zeta \xrightarrow[G]{1} \eta$. From inductive hypothesis we know that ζ has one of the forms in 2. We analyze them in the following.

Case 1.1. If $\zeta = a^i L b^i c^k \gamma$, then

$$
\eta = \begin{cases} a^{i+1} L b^{i+1} c^k \gamma & \text{if } L \to aLb \text{ is applied,} \\ a^{i+1} A b^i c^k \gamma = a^{i+1} \alpha b^i c^k \gamma & \text{if } L \to aA \text{ is applied,} \\ a^i b B b^i c^k \gamma = a^i b \beta b^i c^k \gamma & \text{if } L \to bB \text{ is applied,} \\ a^i L b^i c^{k+1} \gamma & \text{if } C \to cC \text{ is applied and } \gamma = C, \\ a^i L b^i c^k \epsilon = a^i L b^i c^k \gamma & \text{if } C \to \epsilon \text{ is applied and } \gamma = C. \end{cases}
$$

Case 1.2. If $\zeta = a^i \alpha b^j c^k \gamma$ such that $i > j$, then

$$
\eta = \begin{cases} a^{i+1} A b^j c^k \gamma = a^{i+1} \alpha b^j c^k \gamma & \text{if } A \to aA \text{ is applied and } \alpha = A, \\ a^i b^j c^k \gamma = a^i \alpha b^j c^k \gamma & \text{if } A \to \epsilon \text{ is applied and } \alpha = A, \\ a^i \alpha b^j c^{k+1} \gamma & \text{if } C \to cC \text{ is applied and } \gamma = C, \\ a^i \alpha b^j c^k \epsilon = a^i \alpha b^j c^k \gamma & \text{if } C \to \epsilon \text{ is applied and } \gamma = C. \end{cases}
$$

In the first case above, since from induction hypothesis we know that $i > j$ then $i + 1 > j$ as well.

Case 1.3. If $\zeta = a^i b^j \beta b^i c^k \gamma$ such that $j > 0$, then

$$
\eta = \begin{cases} a^i b^{j+1} B c^k \gamma = a^i b^{j+1} \beta c^k \gamma & \text{if } B \to bB \text{ is applied and } \beta = B, \\ a^i b^{i+j} c^k \gamma = a^i b^i \beta b^j c^k \gamma & \text{if } B \to \epsilon \text{ is applied and } \beta = B, \\ a^i b^j \beta b^i c^{k+1} \gamma & \text{if } C \to cC \text{ is applied and } \gamma = C, \\ a^i b^j \beta b^i c^k \epsilon = a^i b^j \beta b^i c^k \gamma & \text{if } C \to \epsilon \text{ is applied and } \gamma = C. \end{cases}
$$

Case 2 $S \to AR$. Similar to Case 1, we use induction on the length of derivation. This time we set α to be the sentenial that is either A or ϵ for any $i \geq 0$. We then conclude that after n derivation, we end up in one of the following cases.

$$x = \begin{cases} a^i \alpha b^j R c^j & \text{where } n = i + 2j, \\ a^i \alpha b^j \beta c^k \gamma & \text{where } n = i + j + k,\ j > k \text{ and } \beta \text{ is either } B \text{ or } \epsilon, \\ a^i \alpha b^j c^k \gamma c^j & \text{where } n = i + j + k,\ k > 0 \text{ and } \gamma \text{ is either } C \text{ or } \epsilon \end{cases} \tag{2}$$

Again, it is easy to conclude the only way that x becomes a string with non-terminals is when $\alpha = \beta = \gamma = \epsilon$, which would be either $a^i b^j c^k$, $j > k$ or $a^i b^j c^{k+j}$, $k > 0$, both of which are in $L(a^* b^* c^*) - \{a^n b^n c^n | n \geq 0\}$. The proof of the induction is similar to Case 1 where we have to analyze the new sentenial after a new production based on the sentenial immediately preceding it.

So far, we have proved that $L(G) \subseteq C$. We now prove $C \subseteq L(G)$ by showing that for any $a^i b^j c^k \in C$, the CFL G can derive this string. We know that $a^i b^j c^k \in C$ implies either $i \neq j$ or $j \neq k$. We analyze these two cases in the following.

Case 1 $i \neq j$. First, we derive $S \xrightarrow[G]{1} LC$. Then with another $k+1$ productions we derive $S \xrightarrow[G]{1} LC \xrightarrow[G]{k} Lc^k C \xrightarrow[G]{1} Lc^k$. Note that if $k = 0$ with 1 production from non-terminal C we derive $C \to \epsilon$. Without loss of generality, let $i < j$. Then we can derive $Lc^k \xrightarrow[G]{i} a^i L b^i c^k \xrightarrow[G]{1} a^i b B b^i c^k \xrightarrow[G]{j-i-1} a^i b^{j-i} B b^i c^k \xrightarrow[G]{1} a^i b^j c^k$. We can, therefore, conclude that $S \xrightarrow[G]{*} a^i b^j c^k$.

Case 2 $j \neq k$. This is similar to the previous case, except we first derive $S \xrightarrow[G]{1} AR$. Then with another $i+1$ productions we derive $S \xrightarrow[G]{1} AR \xrightarrow[G]{i} a^i AR \xrightarrow[G]{1} a^i R$. Again, if $i = 0$ then with 1 production from non-terminal A we derive $A \to \epsilon$. Without loss of generality, let $j < k$. Then we can generate $a^i R \xrightarrow[G]{j} a^i b^j R c^j \xrightarrow[G]{1} a^i b^j c C c^j \xrightarrow[G]{k-j-1} a^i b^j c^{k-j} C c^j \xrightarrow[G]{1} a^i b^j c^k$. We can thus conclude that $S \xrightarrow[G]{*} a^i b^j c^k$.

We proved that $L(G) \subseteq C$ and $C \subseteq L(G)$. This concludes that $L(G) = C$.