CS 332: *Theory of Computation* — Prof. Ran Canetti
Boston University — September 29, 2020

Homework 4 – Due Saturday, October 3rd, 2020 <u>before</u> 11:59PM

Reminder Collaboration is permitted, but you must write the solutions *by yourself without assistance*, and be ready to explain them orally to the course staff if asked. You must also identify your collaborators. Getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.

Exercises Please practice on exercises and solved problems in Chapter 2. The material they cover may appear on exams. Pay particular attention to **Problem 2.18**. It will be helpful for one of the problems on this homework and you can use the result of this problem without proof.

Problems There are 2 mandatory problems and one bonus problem.

1. (**CFGs**) Give CFGs that generate the following languages with at most 2 variables. Unless specified otherwise, the alphabet is $\Sigma = \{0, 1\}$. As always, explanations can help you earn partial credits.
   (a) (**10 points**) $L_1 = \{w | \ w \in \{\texttt{0},\texttt{1}\}^*$ and w contains a number of 1s that is divisible by 3$\}$.
   (b) (**10 points**) $L_3 = \{w | \ w$ is a balanced string of parentheses and brackets$\}$. The alphabet here is $\Sigma = \{\texttt{(},\texttt{)},\texttt{[},\texttt{]}\}$.
   (c) (**10 points**) L_4 is the collection of all strings that contain at least one 1 in their second half (if the string is of odd length, we exclude the middle symbol to construct the second "half"). In other words, $L_4 = \{uv \mid u \in \Sigma^*, v \in \Sigma^*\texttt{1}\Sigma^*$ and $|u| \geq |v|\}$.
2. (**Non-CFLs**) Prove that the following languages are not context-free.
   (a) (**20 points**) The language $A = \{ww \mid w \in \{0, 1\}^*\}$.
   (b) (**20 points**) The following language over the alphabet $\{\texttt{a}, \texttt{b}\}$:
   $C = \{\texttt{a}^i\texttt{b}^j | \ i, j \geq 0$ and if $i = 1$ then j is a prime$\}$.
   (*Careful:* C satisfies the pumping lemma for CFLs! Make sure you understand why, but you don't need to write it down.)
3. **Bonus problem, no collaboration is allowed**
   Recall that the definition of a context-free grammar is a 4-tuple $G = (V, \Sigma, R, S)$, and $\Sigma_\varepsilon := \Sigma \cup \{\varepsilon\}$.
   (a) (**Chomsky reduced form**) We call a context-free grammar to be in Chomsky normal form, if every production rule in R is either $A \to BC$ or $A \to a$ for some $A, B, C \in V$ and $a \in \Sigma_\varepsilon$. Prove that the set of possible languages generated by this class of grammar is equivalent to context-free languages (in other word, every context-free grammar has an equivalent Chomsky reduced form).
   (b) (**Right-regular grammar**) We call a context-free grammar to be right-regular, if every production rule in R is either $A \to aB$ or $A \to a$ for some $A, B \in V$ and $a \in \Sigma_\varepsilon$. Prove that the set of possible languages generated by right-regular grammar is equivalent to regular languages.

(c) (**Linear grammar**) We call a context-free grammar to be linear, if every production rule in R is either $A \to aBc$ or $A \to a$ for some $A, B \in V$ and $a, c \in \Sigma_\varepsilon$. Let the set of languages recognized by linear grammar to be linear languages.

i. Prove that regular languages is a strict subset of linear languages.

ii. Prove the following pumping lemma for linear languages. For every linear language $L \subseteq \Sigma^*$, there exists a constant $p \geq 1$, such that for every string $w \in L$ with $|w| \geq p$, there exists a partition $w = xuyvz$, where $|uv| > 0$ and $|xuvz| \leq p$, such that for any $i \geq 0$, $xu^iyv^iz \in L$.

iii. Use the pumping lemma above to show that the Dyck language generated by CFG below is not linear.

$$S \to (S)S \mid \varepsilon$$