CSE 105 - Theory of Computation (Winter 2026)
Homework 4: Pushdown Automata and Context-free grammars
Due Wednesday, February 4, 11:59pm

Instructions: There are two parts to this homework:

- Concept check (Question 1): Every student must complete this individually on Gradescope.
- Written Homework (Question 2-4): You may work individually or in a team of up to 3 people. Please ensure your name(s) and PID(s) are clearly visible on the first page of your submission, and then upload the PDF to Gradescope. If working in a group, submit only one submission per group: one partner uploads the submission through their Gradescope account and then adds the other group member to the Gradescope submission by selecting their name in the "Add Group Members" dialog box. You will need to re-add your group member every time you resubmit a new version of your assignment.

It is highly recommended (though not required) that you type your answers. It is your responsibility to make any handwriting clear and legible for grading. A LaTeX template for the homework is provided on Canvas. For ease of grading, please start each new problem on a separate page.

We will only be grading some of the problems below for correctness. However, because all of the concepts are important, we will not reveal which problems are being graded for correctness until after the assignment has been submitted. The remaining problems will be graded for completeness (i.e., does it look like there was a good-faith effort to solve the problem?).

Reading and extra practice problems: Sipser Sections 2.1, 2.2. Chapter 2 exercises 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.9, 2.10, 2.11, 2.12, 2.13, 2.16, 2.17.

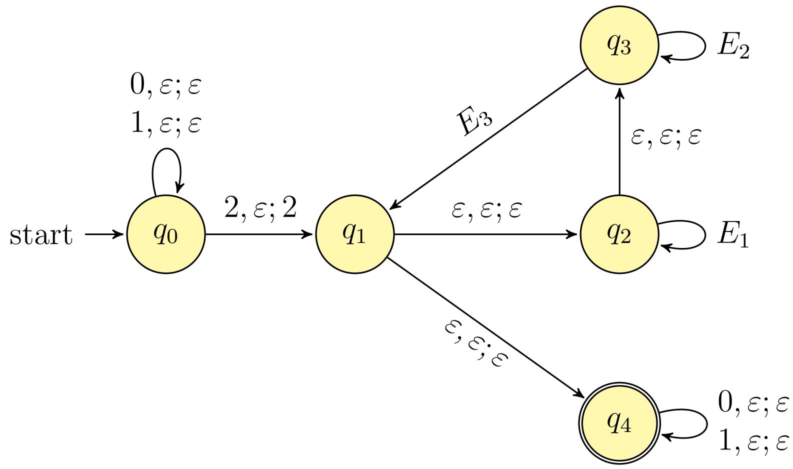
Problems:

1. **Concept check**

   Complete the assignment "Homework 4 - Concept Check" on Gradescope.

2. **A PDA with multiple possibilities**

   Consider the PDA with input and stack alphabet $\Gamma = \{0, 1, 2\}$ whose "unfinished" state diagram is given below:



There are three labels (E_1, E_2, and E_3) on the edges that are unspecified. To be precise, each E_i is of the form "$x, y; z$" where $x, y, z \in \Gamma_\varepsilon$ (recall $\Gamma_\varepsilon = \Gamma \cup \{\varepsilon\}$).

(a) Prove that (no matter how the labels E_1, E_2, E_3 are specified), the language recognized by this PDA is infinite. A complete solution will include a precise description of an infinite collection of strings each of which is accepted by the PDA, with a precise and clear description of the accepting computation of the PDA on each of these strings.

(b) Prove/Disprove: Over all the possible choices for the labels E_1, E_2, E_3, this PDA can only recognize finitely many languages. Justify your solution by referring back to the relevant definitions.

(c) Determine how to set the labels E_1, E_2, E_3 so that the language of the PDA is

$$\{x2yz \in \Gamma^* \mid x, z \in \{0,1\}^* \text{ and } y = 0^n 1^n 2 \text{ for } n \geq 0\}$$

In addition to specifying each E_i, a complete justification will include a precise description of why this choice of E_i means that the PDA recognizes the language indicated.

(d) Let $\Sigma = \{0, 1\}$ and let $A = \{0^n 1^n 2 \mid n \geq 0\}$. Determine how to set the labels E_1, E_2, E_3 so that the language of the PDA is

$$\Sigma^* \circ \{2\} \circ A^* \circ \Sigma^*$$

Since this question is similar to the previous one, you need only define E_1, E_2, and E_3. No justification is required.

3. **Grammar practice**

For each of the languages listed below, define a context-free grammar $G = (V, \Sigma, R, S)$ that generates the language. Instead of formally justifying your grammar, illustrate it by giving **two examples** of strings in the language and their derivations using your grammar and **one example** of a string not in the language with an explanation of why it cannot appear on the right side of any derivation in your grammar. Choose your examples so they are different enough to illustrate the role of as many of the variables in your grammar as possible.

(a) The language from problem 2(c):

$$\{x2yz \in \Gamma^* \mid x, z \in \{0,1\}^* \text{ and } y = 0^n 1^n 2 \text{ for } n \geq 0\}$$

(b) $\{1^n = 1^a + 1^b \in \{1, =, +\}^* \mid a, b, n \geq 1 \text{ such that } a + b = n\}$

4. **Substrings and regularity**

For this problem, we fix the alphabet $\Gamma = \{0, 1, 2\}$. Recall the definition of the SUBSTRING function: for all languages $K \subseteq \Gamma^*$

$$\textsc{Substring}(K) := \{w \in \Gamma^* \mid \text{there exist } a, b \in \Gamma^* \text{ such that } awb \in K\}.$$

(a) Prove that $\textsc{Substring}(\{0^n 1^n \mid n \geq 0\})$ is regular. A complete solution will include a precise definition of a DFA, NFA, or regular expression that recognizes or describes it, along with a brief justification of your construction by explaining the role each state plays in the machine or referring back to relevant definitions.

(b) Prove that $\textsc{Substring}(\{0^n 1^n 2^n \mid n \geq 0\})$ is not regular.

(c) Is $\textsc{Substring}(\{0^n 1^n 2^n \mid n \geq 0\})$ context-free? Informally justify your answer, referring to class discussions and/or the textbook.