Homework 3

COMS W3261, Summer A 2022

This homework is due **Tuesday, 6/14/2022, at 11:59pm EST**. Submit to GradeScope (course code: 2KGDW8).

Grading policy reminder: LaTeX is preferred, but neatly typed or handwritten solutions are acceptable. I recommend using the .tex file for the homework as a template to write up your answers. Your TAs may dock points for indecipherable writing.

Proofs should be complete; that is, include enough information that a reader can clearly tell that the argument is rigorous.

If a question is ambiguous, please state your assumptions. This way, we can give you credit for correct work. (Even better, post on Ed so that we can resolve the ambiguity.)

LaTeX resources.

- The website Overleaf (essentially Google Docs for LaTeX) may make compiling and organizing your .tex files easier. Here's a quick tutorial.

- Detexify is a nice tool that lets you draw a symbol and returns the LaTeX codes for similar symbols.

- The tool Table Generator makes building tables in LaTeX much easier.

- The tool Finite State Machine Designer may be useful for drawing automata. See also this example (PDF) (.tex) of how to make fancy edges (courtesy of Eumin Hong).

- The website mathcha.io allows you to draw diagrams and convert them to LaTeX code.

- To use the previous drawing tools (and for most drawing in LaTeX), you'll need to use the package Tikz (add the command "\usepackage{tikz}" to the preamble of your .tex file to import the package).

- This tutorial is a helpful guide to positioning figures.

1 Problem 1 (12 points)

1. (3 points). What is the language of the grammar G_1 below? Express this language as a set or with a sentence and explain your reasoning.

   (Recall that we can express a CFG briefly by writing the rules. When we do this, we interpret the variable at top left as the start variable, read the other variables off the left-hand side, and infer the terminal alphabet from the other symbols listed in the rules. For example: in this case, the start variable is S, the variable set is $\{S, A, B\}$, and the terminal alphabet is $\{x, y, \&, @\}$.)

$$S \to xAy$$
$$A \to xAy \mid B$$
$$B \to \&\& \mid @@$$

2. (3 points). What is the language of the grammar G_2 below? Express this language as a set or with a sentence and explain your reasoning.

$$S \to 0A \mid 1B \mid 2C$$
$$A \to 1F \mid 2E$$
$$B \to 0F \mid 2D$$
$$C \to 0E \mid 1D$$
$$D \to 0$$
$$E \to 1$$
$$F \to 2$$

3. (3 points). Design a grammar for the language

$$D = \{0^{2n}1^n0^m1^{2m} \mid m, n \geq 0\}$$

   and explain why your grammar produces D. (This language includes such strings as 001011, 000011000011111111, 000000111011, 001, etc.) You may use the brief representation of grammars (i.e., rules-only) or write out the full 4-tuple.

4. (3 points). Design a grammar for the language represented by the regular expression

$$R_1 = (01)^* \cup (10)^*$$

   and explain why your grammar produces the same language. You may use the brief representation of grammars (i.e., rules-only) or write-out the full 4-tuple.

Rationale: The goal of this question is to practice interpreting and building context-free grammars.

References: Sipser p. 102 and Lightning Review 5 (CFG definition and deriving strings), Sipser p.105 (figuring out the language of a CFG), and Sipser p.106-107 (tips for building CFGs).

2   Problem 2 (12 points)

1. (6 points.) Prove that the language

$$A = \{a^n b^{2n} a^n \mid n \geq 0\}$$

over the alphabet $\Sigma = \{a, b\}$ is nonregular. Hint: we know two ways of proving nonregularity: using the pumping lemma and proof by contradiction and reasoning from the closure of the regular languages under regular operations.

2. (6 points.) Prove that the language

$$B = \{ww \mid w \in \{0, 1\}^* \text{ and } w \text{ contains at least one 0 and at least one 1}\}$$

over the alphabet $\Sigma = \{0, 1\}$ is nonregular. You may use the pumping lemma and/or closure properties.

Rationale: The goal of this question is to practice using the pumping lemma and closure properties to show that languages are nonregular.

References: Sipser p. 78-79 and Lightning Review 4 (the pumping lemma), Sipser p.80-82 and Lightning Review 5 (using the pumping lemma).

3   Problem 3 (6 points)

1. (6 points). Recall the conversion procedure that we used in class to remove a state from a GNFA without changing its function: for every state pair (q_i, q_j) distinct from our removal state q_{rip}, we rerouted traffic by transforming the first picture below into the second picture:
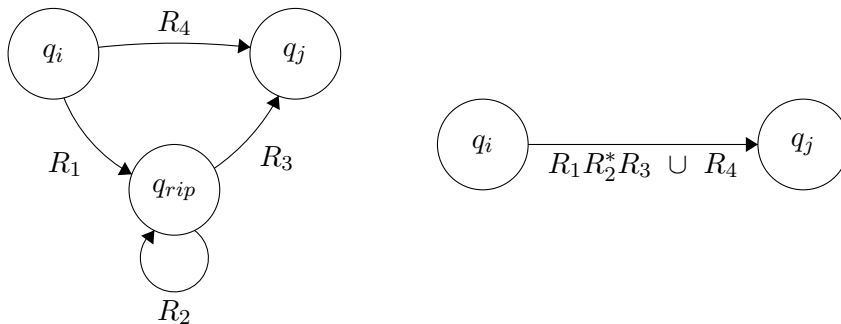
Using our procedure, remove state b from the state diagram below and show the resulting state diagram. If you like, you can use the provided table to compute $R_1 R_2^* R_3 \cup R_4$.
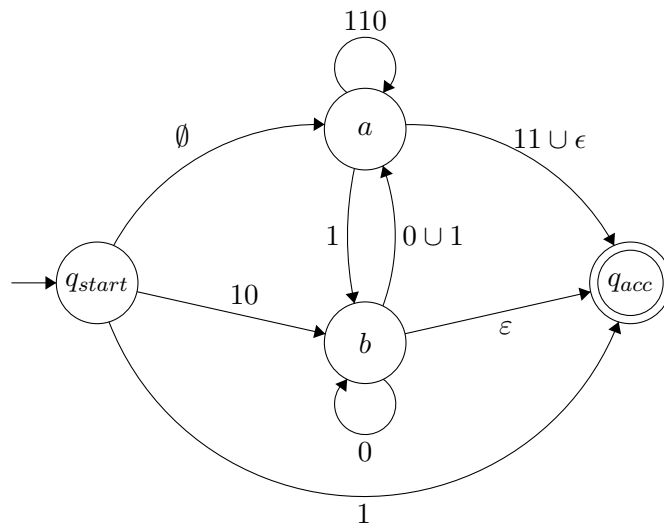
State pair (q_i, q_j)	Regular expression $R_1 R_2^* R_3 \cup R_4$

4 Problem 4 (7 points)

1. (4 points) In general, applying regular operations to nonregular languages is not guaranteed to result in a nonregular language.

   Suppose A is a nonregular language. Is the complement $\overline{A}$ also nonregular? Explain why or why not. [Hint: think about our proof that the regular languages are closed under complement.]

2. (3 points) Read the statement of the pumping lemma carefully. Does the pumping lemma guarantee that nonregular languages can't be pumped (i.e., that we can show any nonregular language is nonregular by demonstrating that it doesn't satisfy the pumping conditions?)

Rationale: The goal of this question is to practice logical thinking and careful interpretation of theorems.

References: Our proof that the regular languages are closed under complement: "If a DFA D recognizes L, then converting all accept states into reject states and vice versa creates a new DFA D' that recognizes $\overline{L}$". Sipser p. 78-79 and Lightning Review 4 (the pumping lemma).

5   Problem 5 (2 Extra Credit Points)

The Myhill-Nerode theorem says that a language L is a regular language if and only if L has a finite number of equivalence classes (i.e., L would not be a regular language if it had an infinite number of equivalence classes).

Consider the language L, defined over the alphabet $\Sigma = \{0\}$:

$$L = \{w | \text{ length of } w \text{ is divisible by } 3\}$$

so strings in the language include $000, 000000, \varepsilon$; and strings not in the language include $0, 00000$.

Is L regular or nonregular? Prove your claim using the Myhill-Nerode theorem. (Hint: You should define all the equivalence classes for L in terms of distinguishing strings, or prove that there are an infinite number of equivalence classes.)